

# **ESCUELA POLITÉCNICA NACIONAL**

## **FACULTAD DE INGENIERÍA DE SISTEMAS**

### **DESARROLLO DE UN DATA MART PARA EL MONITOREO DE INDICADORES DE ACREDITACIÓN UNIVERSITARIA.**

#### **TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN SISTEMAS INFORMÁTICOS Y DE COMPUTACIÓN**

**ORTIZ YUMISACA LEONARDO ALEXANDER**

leonardo.ortiz01@epn.edu.ec

**DIRECTOR: PhD. MARÍA ASUNCIÓN HALLO CARRASCO**

maria.hallo@epn.edu.ec

**Quito, octubre 2018**

## DECLARACIÓN

Yo, Leonardo Alexander Ortiz Yumisaca, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

---

Leonardo Alexander Ortiz Yumisaca

## **CERTIFICACIÓN**

Certifico que el presente trabajo fue desarrollado por Leonardo Alexander Ortiz Yumisaca, bajo mi supervisión.

---

**PhD. María Hallo**  
**DIRECTOR DE PROYECTO**

## AGRADECIMIENTOS

Quiero agradecer a Dios por guiar mi camino y acompañarme durante toda mi vida, brindándome la salud y la capacidad de asimilar satisfactoriamente los conocimientos hasta ahora obtenidos.

Así, quiero mostrar mi gratitud a cada uno de los que son parte de mi familia, a mi padre Rubén Ortiz, a mi madre Nancy Yumisaca y a mis hermanos Adrián y Andrea Ortiz.

Agradezco profundamente a mi tía Mariana Ortiz, por apoyarme incondicionalmente mientras cumplía esta meta profesional.

Un agradecimiento especial a la PhD. María Hallo, mi tutora del proyecto, quien con dedicación y esfuerzo me supo supervisar y orientar durante el desarrollo de mi proyecto de titulación.

Leonardo

## DEDICATORIA

A Dios, por estar conmigo en cada paso que doy.

A mis padres, por su sacrificio y por todo el apoyo brindado en cada paso de mi vida.

A mis hermanos, por su amor incondicional con palabras de aliento, sus ocurrencias y sus sonrisas.

A todos mis amigos, por estar siempre conmigo en las buenas y en las malas apoyándome incondicionalmente y demostrándome su cariño.

Leonardo

# CONTENIDO

|   |      |
|---|------|
| ÍNDICE DE FIGURAS .....   | xi   |
| ÍNDICE DE TABLAS.....   | xiii |
| RESUMEN .....   | xv   |
| ABSTRACT .....  | xvi  |
| CAPÍTULO 1 .....  | 16   |
| 1. INTRODUCCIÓN.....  | 16   |
| 1.1. Planteamiento del problema.....  | 17   |
| 1.1.1. Descripción del problema .....   | 17   |
| 1.2. Objetivos .....  | 17   |
| 1.2.1. Objetivo general .....   | 17   |
| 1.2.2. Objetivos específicos.....   | 18   |
| 1.3. Alcance .....  | 18   |
| 1.4. Base teórica .....   | 19   |
| 1.4.1. Consejo de Evaluación, Acreditación y Aseguramiento de la<br>Calidad de la Educación Superior (CEAACES)..... | 19   |
| 1.4.2. Estructura del modelo de evaluación incorporado por el<br>CEAACES.....                                       | 19   |
| 1.4.3. Metodología de valoración .....  | 21   |
| 1.4.3.1. Variables .....  | 21   |
| 1.4.3.2. Indicadores .....  | 22   |
| 1.4.3.3. Estándares.....  | 22   |
| 1.4.3.4. Evidencias.....  | 22   |
| 1.4.3.5. Pesos de indicadores en el modelo .....  | 23   |
| 1.4.3.6. Funciones de utilidad .....  | 25   |
| 1.4.4. Inteligencia de negocios .....   | 26   |
| 1.4.4.1. Proceso de BI .....  | 26   |
| 1.4.4.2. Ventajas y desventajas de implementar herramientas de BI<br>en una organización .....                      | 27   |
| 1.4.5. Análisis de plataformas de BI .....  | 27   |
| 1.4.5.1. Oracle Business Intelligence.....  | 27   |
| 1.4.5.2. MicroStrategy.....   | 28   |
| 1.4.5.3. Pentaho Business Analytics.....  | 28   |

|            |   |    |
|------------|---|----|
| 1.4.6.     | Proceso ETL.....  | 29 |
| 1.5.       | Definición de términos.....                                   | 30 |
| 1.5.1.     | OLTP (On-Line Transactional Processing).....                  | 30 |
| 1.5.2.     | OLAP (On Line Analytic Processing).....                       | 30 |
| 1.5.2.1.   | MOLAP (Multidimensional On Line Analytical Processing).....   | 31 |
| 1.5.2.2.   | ROLAP (Relational On Line Analytic Processing).....           | 31 |
| 1.5.3.     | Características distintivas entre OLAP y OLTP .....           | 31 |
| 1.5.4.     | Almacén de datos.....   | 32 |
| 1.5.5.     | <i>Data mart</i> .....  | 33 |
| 1.5.6.     | Enfoque <i>Top-Down</i> .....                                 | 34 |
| 1.5.7.     | Enfoque <i>Bottom-Up</i> .....                                | 34 |
| 1.5.8.     | Tablas de dimensiones.....                                    | 35 |
| 1.5.9.     | Métricas.....   | 35 |
| 1.5.10.    | Tabla de hechos ( <i>fact table</i> ).....                    | 35 |
| 1.5.11.    | Esquema estrella.....   | 35 |
| 1.5.12.    | Área de preparación ( <i>staging area</i> ).....              | 36 |
| 1.6.       | La metodología de Kimball para el desarrollo .....            | 36 |
| 1.6.1.     | Planificación del proyecto.....                               | 38 |
| 1.6.2.     | Definición de los requerimientos del negocio .....            | 38 |
| 1.6.3.     | Modelo dimensional.....                                       | 39 |
| 1.6.3.1.   | Identificar el proceso de negocio .....                       | 39 |
| 1.6.3.2.   | Establecer el nivel de granularidad .....                     | 39 |
| 1.6.3.3.   | Identificar las dimensiones.....                              | 39 |
| 1.6.3.4.   | Identificar medidas y hechos.....                             | 40 |
| 1.6.4.     | Diseño físico.....  | 40 |
| 1.6.5.     | Construcción del sistema ETL.....                             | 40 |
| 1.6.6.     | Diseño de la arquitectura técnica .....                       | 40 |
| 1.6.7.     | Definición de las herramientas para la plataforma de BI ..... | 41 |
| 1.6.8.     | Especificación de aplicaciones de BI .....                    | 41 |
| 1.6.9.     | Desarrollo de aplicaciones de BI .....                        | 41 |
| 1.6.10.    | Despliegue .....  | 42 |
| 1.6.11.    | Mantenimiento y crecimiento.....                              | 42 |
| 1.6.12.    | Administración del proyecto .....                             | 42 |
| CAPÍTULO 2 | .....   | 43 |
| 2.         | METODOLOGÍA .....   | 43 |

|          |   |    |
|----------|---|----|
| 2.1.     | Planificación del proyecto.....                               | 43 |
| 2.1.1.   | Definición del proyecto .....                                 | 43 |
| 2.1.2.   | Objetivos y alcance del proyecto .....                        | 43 |
| 2.1.3.   | Situación actual .....  | 43 |
| 2.1.4.   | Estrategia de implementación .....                            | 45 |
| 2.1.5.   | Seleccionar la metodología de desarrollo.....                 | 45 |
| 2.1.6.   | Recurso humano .....  | 46 |
| 2.1.7.   | Inversión.....  | 46 |
| 2.2.     | Definición de requerimientos de la institución .....          | 47 |
| 2.2.1.   | Requerimientos de negocio.....                                | 47 |
| 2.2.2.   | Requerimientos de la solución .....                           | 48 |
| 2.2.2.1. | Requerimientos funcionales.....                               | 48 |
| 2.2.3.   | Requerimientos no funcionales .....                           | 50 |
| 2.2.4.   | Pesos y funciones de utilidad .....                           | 51 |
| 2.2.4.1. | Funciones de utilidad para indicadores cualitativos .....     | 51 |
| 2.2.4.2. | Tasa de retención de grado. ....                              | 52 |
| 2.2.4.3. | Tasa de graduación de grado. ....                             | 52 |
| 2.2.4.4. | Tasa de graduación de posgrado. ....                          | 53 |
| 2.2.5.   | Archivos fuente.....  | 54 |
| 2.3.     | Modelo dimensional .....                                      | 55 |
| 2.3.1.   | Definición del proceso de negocio.....                        | 55 |
| 2.3.2.   | Establecer los niveles de granularidad .....                  | 55 |
| 2.3.3.   | Elección de las dimensiones .....                             | 55 |
| 2.3.3.1. | Dimensiones encontradas.....                                  | 56 |
| 2.3.3.2. | Dimensiones y sus atributos .....                             | 57 |
| 2.3.4.   | Tabla de hechos fact_modelo_acreditación .....                | 57 |
| 2.4.     | Definición de la arquitectura técnica.....                    | 58 |
| 2.4.1.   | Definición de las herramientas para la plataforma de BI ..... | 59 |
| 2.5.     | Diseño físico.....  | 60 |
| 2.5.1.   | Dimensión criterio_dim .....                                  | 60 |
| 2.5.2.   | Dimensión subcriterio_dim .....                               | 61 |
| 2.5.3.   | Dimensión indicador_dim .....                                 | 61 |
| 2.5.4.   | Dimensión tiempo_dim .....                                    | 61 |
| 2.5.5.   | Tabla de hechos modelo acreditación.....                      | 62 |
| 2.6.     | Construcción del sistema ETL.....                             | 62 |



|          |  |    |
|----------|--|----|
| 2.7.     | Procesos ETL.....  | 63 |
| 2.7.1.   | Extracción de los datos fuentes.....                           | 63 |
| 2.7.1.1. | Descripción de las fuentes.....                                | 63 |
| 2.7.2.   | Procesos ETL para la carga de datos a las dimensiones.....     | 63 |
| 2.7.2.1. | Carga de la dimensión criterio_dim.....                        | 63 |
| 2.7.2.2. | Carga de la dimensión subcriterio_dim.....                     | 64 |
| 2.7.2.3. | Carga de la dimensión indicador_dim.....                       | 65 |
| 2.7.2.4. | Carga de la dimensión tiempo_dim.....                          | 66 |
| 2.7.2.5. | Proceso Inicio_ETL_data_mart_EPN.....                          | 67 |
| 2.7.3.   | Procesos ETL para carga de datos a la tabla de hechos.....     | 68 |
| 2.7.3.1. | Carga de tabla temporal de variables.....                      | 69 |
| 2.7.3.2. | Carga de la tabla temporal de cálculos.....                    | 73 |
| 2.7.3.3. | Carga de la tabla de hechos fact_modelo_acreditación.....      | 76 |
| 2.7.3.4. | Proceso_ETL_DATA_MART_EPN.....                                 | 76 |
| 2.8.     | Especificación de aplicaciones de BI.....                      | 77 |
| 2.8.1.   | Selección de requerimientos.....                               | 78 |
| 2.8.2.   | Definición de roles de usuario.....                            | 79 |
| 2.9.     | Desarrollo de aplicaciones de BI.....                          | 79 |
| 2.9.1.   | Desarrollo del cubo OLAP “DATA MART CEAACES”.....              | 79 |
| 2.9.1.1. | Dimensión Criterio_dim.....                                    | 79 |
| 2.9.1.2. | Dimensión Subcriterio_dim.....                                 | 80 |
| 2.9.1.3. | Dimensión Indicador_dim.....                                   | 81 |
| 2.9.1.4. | Dimensión Tiempo_dim.....                                      | 82 |
| 2.9.1.5. | Tabla de hechos fact modelo acreditación.....                  | 83 |
| 2.9.1.6. | Diagrama del cubo OLAP DATA MART CEAACES.....                  | 83 |
| 2.9.2.   | Desarrollo de aplicativo web para desplegar reportes.....      | 84 |
| 2.9.2.1. | Obtención de requerimientos.....                               | 85 |
| 2.9.2.2. | Diseño conceptual.....   | 88 |
| 2.9.2.3. | Diseño navegacional.....                                       | 94 |
| 2.9.2.4. | Diseño de interfaz abstracta.....                              | 95 |
| 2.9.3.   | Desarrollo de reportes iniciales.....                          | 96 |
| 2.9.3.1. | Cuadro de mando: Resultados obtenidos por criterio.....        | 96 |
| 2.9.3.2. | Reporte peso obtenido por criterio.....                        | 97 |
| 2.9.3.3. | Reporte del peso obtenido y por obtener según el criterio..... | 98 |

|  |     |
|--|-----|
| 2.9.3.4. Cuadro de mando: Resultados obtenidos por subcriterio .....               | 98  |
| 2.9.3.5. Reporte del peso obtenido por subcriterio .....                           | 100 |
| 2.9.3.6. Reporte peso obtenido y por obtener de los subcriterios.....              | 100 |
| 2.9.3.7. Cuadro de mando: Resultados obtenidos por indicador .....                 | 101 |
| 2.9.3.8. Cuadro de mando: Pesos obtenidos y por obtener de los<br>indicadores..... | 101 |
| 2.9.3.9. Cuadro de mando: Utilidad obtenida por indicador .....                    | 102 |
| 2.9.3.10. Reporte del valor, peso y utilidad de los indicadores .....              | 104 |
| 2.9.3.11. Reporte peso obtenido y por obtener de los indicadores .                 | 105 |
| 2.9.3.12. Reporte de la utilidad obtenida por indicador .....                      | 105 |
| 2.9.4. Implementación del aplicativo web .....                                     | 107 |
| 2.10. Despliegue .....   | 108 |
| 2.10.1. Pruebas Alfa.....  | 108 |
| 2.10.2. Pruebas Beta.....  | 109 |
| 2.11. Pruebas.....   | 109 |
| 2.11.1. Pruebas de validación de la información .....                              | 109 |
| 2.11.2. Pruebas de funcionalidad .....   | 109 |
| 2.11.2.1. Definición de los casos de pruebas.....                                  | 109 |
| 2.11.3. Pruebas de integridad .....  | 115 |
| CAPÍTULO 3 .....   | 117 |
| 3. RESULTADOS Y DISCUSIÓN .....  | 117 |
| 3.1. Resultados de las pruebas de validación .....                                 | 117 |
| 3.2. Resultados de las pruebas de funcionalidad.....                               | 118 |
| 3.3. Resultados de las pruebas de integridad .....                                 | 118 |
| CAPÍTULO 4 .....   | 121 |
| 4. CONCLUSIONES Y RECOMENDACIONES.....   | 121 |
| 4.1. Conclusiones.....   | 121 |
| 4.2. Recomendaciones.....  | 122 |
| REFERENCIAS BIBLIOGRÁFICAS .....   | 123 |
| ANEXOS .....   | 126 |

## ÍNDICE DE FIGURAS

|  |    |
|--|----|
| Figura 1. Descripción general del modelo. ....                                     | 20 |
| Figura 2. Fases del proceso BI.....  | 26 |
| Figura 3. Top-Down.....  | 34 |
| Figura 4. Bottom-up.....   | 34 |
| Figura 5. Esquema Estrella Estándar.....   | 36 |
| Figura 6. Diagrama del ciclo de vida de la metodología de Kimball. ....            | 37 |
| Figura 7. Proceso de levantamiento y entrega de información.....                   | 44 |
| Figura 8. Proceso para asegurar la calidad de datos. ....                          | 45 |
| Figura 9. Función de utilidad de indicadores cualitativos. ....                    | 51 |
| Figura 10. Función de utilidad del indicador tasa de retención de grado.....       | 52 |
| Figura 11. Función de utilidad del indicador tasa graduación de grado.....         | 53 |
| Figura 12. Función de utilidad del indicador tasa titulación posgrado. ....        | 54 |
| Figura 13. Análisis dimensional del proceso de negocio. ....                       | 56 |
| Figura 14. Tabla de hechos fact_modelo_acreditación. ....                          | 58 |
| Figura 15. Arquitectura técnica de la solución.....                                | 59 |
| Figura 16. Diseño físico.....  | 60 |
| Figura 17. Tabla de origen para la carga de la dimensión criterio. ....            | 64 |
| Figura 18. Estructura de la carga de información de la dimensión criterio. ....    | 64 |
| Figura 19. Tabla de origen para la carga de la dimensión criterio. ....            | 65 |
| Figura 20. Estructura de la carga de información de la dimensión subcriterio. .... | 65 |
| Figura 21. Tabla de origen para la carga de la dimensión indicador. ....           | 66 |
| Figura 22. Estructura de la carga de información de la dimensión indicador. ....   | 66 |
| Figura 23. Tabla de origen para la carga de la dimensión tiempo. ....              | 67 |
| Figura 24. Estructura de la carga de información de la dimensión tiempo. ....      | 67 |
| Figura 25. Proceso Inicio_ETL_DATA_MART_EPN. ....                                  | 68 |
| Figura 26. Proceso de carga a tabla temporal de variables. ....                    | 69 |
| Figura 27. Consulta SQL para obtener la variable NTEA. ....                        | 70 |
| Figura 28. Consulta SQL para obtener la variable NEMA. ....                        | 70 |
| Figura 29. Consulta SQL para obtener la variable NECp. ....                        | 71 |
| Figura 30. Consulta SQL para obtener la variable NEGp.....                         | 71 |
| Figura 31. Consulta SQL para obtener la variable NEGP. ....                        | 72 |
| Figura 32. Consulta SQL para obtener la variable NECP.....                         | 72 |
| Figura 33. Consulta SQL para obtener la cohorte. ....                              | 72 |
| Figura 34. Consulta SQL para obtener el período académico.....                     | 73 |
| Figura 35. Proceso de carga a tabla temporal de cálculos.....                      | 73 |
| Figura 36. Función de utilidad tasa retención de grado.....                        | 75 |
| Figura 37. Función de utilidad tasa titulación de grado. ....                      | 75 |
| Figura 38. Función de utilidad tasa titulación de posgrado.....                    | 75 |
| Figura 39. Proceso de carga de tabla de hechos. ....                               | 76 |
| Figura 40. Proceso_ETL_DATA_MART_EPN.....  | 77 |
| Figura 41. Atributos de la dimensión Criterio_dim. ....                            | 80 |
| Figura 42. Origen de datos de la dimensión Criterio_dim.....                       | 80 |
| Figura 43. Atributos de la dimensión Subcriterio_dim. ....                         | 81 |
| Figura 44. Origen de datos de la dimensión Subcriterio_dim. ....                   | 81 |
| Figura 45. Atributos de la dimensión Indicador_dim.....                            | 81 |

|  |     |
|--|-----|
| Figura 46. Vista de origen de datos de dimensión Indicador_dim. ....                         | 82  |
| Figura 47. Atributos de dimensión Tiempo_dim. ....   | 82  |
| Figura 48. Vista de origen de datos de dimensión Tiempo_dim. ....                            | 83  |
| Figura 49. Tabla de hechos fact_modelo_acreditación. ....                                    | 83  |
| Figura 50. Diagrama del cubo OLAP “DATA MART CEAACES”.....                                   | 84  |
| Figura 51. Fases de la metodología OOHDM. ....   | 85  |
| Figura 52. Diagrama de casos de inicio de sesión. ....                                       | 86  |
| Figura 53. Diagrama de casos administrador. ....   | 87  |
| Figura 54. Diagrama de casos usuario. ....   | 87  |
| Figura 55. Diagrama de casos data mart. ....   | 88  |
| Figura 56. Diagrama de clases.....   | 88  |
| Figura 57. Diagrama de secuencia ingresar a la plataforma del administrador. ....            | 89  |
| Figura 58. Diagrama de secuencia ingresar a la plataforma del usuario. ....                  | 89  |
| Figura 59. Diagrama de secuencia registrar usuario. ....                                     | 90  |
| Figura 60. Diagrama de secuencia actualizar usuario.....                                     | 90  |
| Figura 61. Diagrama de secuencia dar de baja usuario. ....                                   | 91  |
| Figura 62. Diagrama de secuencia crear nivel. ....   | 91  |
| Figura 63. Diagrama de secuencia modificar nivel. ....                                       | 92  |
| Figura 64. Diagrama de secuencia eliminar nivel. ....  | 92  |
| Figura 65. Diagrama de secuencia desplegar reportes del administrador.....                   | 93  |
| Figura 66. Diagrama de secuencia desplegar reportes del usuario.....                         | 93  |
| Figura 67. Diagrama de navegación.....   | 94  |
| Figura 68. Diseño de interfaz abstracta ingreso al sistema. ....                             | 95  |
| Figura 69. Diseño de interfaz abstracta administrador. ....                                  | 95  |
| Figura 70. Diseño de interfaz abstracta usuario. ....  | 96  |
| Figura 71. Cuadro de mando: Resultados obtenidos por criterio. ....                          | 97  |
| Figura 72. Reporte peso obtenido por criterio. ....  | 98  |
| Figura 73. Reporte comparativo entre el peso obtenido y por obtener según el criterio. ....  | 99  |
| Figura 74. Cuadro de mando: Resultados obtenidos por subcriterio.....                        | 99  |
| Figura 75. Reporte del peso obtenido por subcriterio.....                                    | 100 |
| Figura 76. Reporte comparativo entre el peso obtenido y por obtener de los subcriterios..... | 101 |
| Figura 77. Cuadro de mando: Resultados obtenidos por indicador.....                          | 102 |
| Figura 78. Cuadro de mando: Pesos obtenidos y por obtener de los indicadores.....            | 103 |
| Figura 79. Cuadro de mando: Utilidad obtenida por indicador. ....                            | 104 |
| Figura 80. Reporte del valor, peso y utilidad de los indicadores. ....                       | 105 |
| Figura 81. Reporte del peso obtenido y por obtener de los indicadores.....                   | 106 |
| Figura 82. Reporte de la utilidad obtenida por indicador. ....                               | 106 |
| Figura 83. Diagrama de despliegue del aplicativo web. ....                                   | 107 |
| Figura 84. Reporte obtenido con datos del data mart.....                                     | 117 |
| Figura 85. Diagrama del cubo OLAP “DATA MART CEAACES”.....                                   | 120 |
| Figura 86. Diseño del modelo de base de datos para crear el cubo.....                        | 119 |

## ÍNDICE DE TABLAS

|  |     |
|--|-----|
| Tabla 1. Indicadores con sus respectivos pesos. ....         | 23  |
| Tabla 2. Características distintivas entre OLAP y OLTP. .... | 31  |
| Tabla 3. Recurso humano de la solución de BI. ....           | 46  |
| Tabla 4. Costo de inversión de Hardware. ....                | 46  |
| Tabla 5. Costo de inversión de software. ....                | 46  |
| Tabla 6. Costo recurso humano. ....                          | 47  |
| Tabla 7. Costo de otras inversiones. ....                    | 47  |
| Tabla 8. Cuadro de requerimiento de negocio N°01. ....       | 47  |
| Tabla 9. Cuadro de requerimiento de negocio N°02. ....       | 48  |
| Tabla 10. Cuadro de requerimiento funcional N°01. ....       | 48  |
| Tabla 11. Cuadro de requerimiento funcional N°02. ....       | 48  |
| Tabla 12. Cuadro de requerimiento funcional N°03. ....       | 49  |
| Tabla 13. Cuadro de requerimiento funcional N°04. ....       | 49  |
| Tabla 14. Cuadro de requerimiento funcional N°05. ....       | 49  |
| Tabla 15. Cuadro de requerimiento funcional N°06. ....       | 49  |
| Tabla 16. Cuadro de requerimiento funcional N°07. ....       | 49  |
| Tabla 17. Cuadro de requerimiento funcional N°08. ....       | 50  |
| Tabla 18. Cuadro de requerimiento funcional N°09. ....       | 50  |
| Tabla 19. Cuadro de requerimiento no funcional N°01. ....    | 50  |
| Tabla 20. Cuadro de requerimiento no funcional N°02. ....    | 50  |
| Tabla 21. Cuadro de requerimiento no funcional N°03. ....    | 51  |
| Tabla 22. Dimensiones y sus atributos. ....                  | 57  |
| Tabla 23. Dimensión criterio_dim. ....                       | 60  |
| Tabla 24. Dimensión subcriterio_dim. ....                    | 61  |
| Tabla 25. Dimensión indicador_dim. ....                      | 61  |
| Tabla 26. Dimensión tiempo_dim. ....                         | 61  |
| Tabla 27. Tabla de hechos modelo acreditación. ....          | 62  |
| Tabla 28. Sistemas fuente de la solución. ....               | 63  |
| Tabla 29. Herramientas para elaboración de reportes. ....    | 78  |
| Tabla 30. Identificación de actores. ....                    | 85  |
| Tabla 31. Características del servidor de pruebas. ....      | 108 |
| Tabla 32. Caso de prueba CP-01. ....                         | 109 |
| Tabla 33. Caso de prueba CP-02. ....                         | 110 |
| Tabla 34. Caso de prueba CP-03. ....                         | 110 |
| Tabla 35. Caso de prueba CP-04. ....                         | 111 |
| Tabla 36. Caso de prueba CP-05. ....                         | 111 |
| Tabla 37. Caso de prueba CP-06. ....                         | 111 |
| Tabla 38. Caso de prueba CP-07. ....                         | 112 |
| Tabla 39. Caso de prueba CP-08. ....                         | 112 |
| Tabla 40. Caso de prueba CP-09. ....                         | 113 |
| Tabla 41. Caso de prueba CP-10. ....                         | 113 |
| Tabla 42. Caso de prueba CP-11. ....                         | 114 |
| Tabla 43. Caso de prueba CP-12. ....                         | 114 |
| Tabla 44. Caso de prueba CP-13. ....                         | 115 |
| Tabla 45. Caso de prueba PI-01. ....                         | 115 |

|   |     |
|---|-----|
| Tabla 46. Caso de prueba PI-02.....           | 116 |
| Tabla 47. Reporte de indicadores EPN. ....    | 117 |
| Tabla 48. Resultados pruebas funcionales..... | 118 |
| Tabla 49. Resultados pruebas integridad. .... | 119 |

## RESUMEN

El Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES) en el Ecuador ejecuta procesos de evaluación con fines de acreditación a las Instituciones de Educación Superior. Como resultado de este proceso varias universidades fueron clausuradas y otro número considerable han quedado descreditadas. Con el fin de apoyar al proceso de acreditación la Escuela Politécnica Nacional auspició el desarrollo de un *data mart* que permite obtener información oportuna para su rápido entendimiento y gestión, evitando que los datos requeridos para la acreditación universitaria, estén dispersos y no permitan un análisis integrado para procesos de toma de decisiones. La implementación se realizó inicialmente con los indicadores del criterio estudiantes según el Modelo de Evaluación Institucional de Universidades y Escuelas Politécnicas incorporado por el CEAACES y en el futuro se puedan añadir indicadores para otros criterios de acreditación. El proyecto se desarrolló empleando la metodología de Kimball y como herramientas de *Business Intelligence* (BI) *Open Source* se empleó la suite de Pentaho. Para la correcta implementación e integración del *data mart* se ejecutaron procesos de extracción, transformación y carga a partir de dos fuentes de datos. La implementación del *data mart* permite visualizar la información a través de herramientas de procesamiento analítico en línea. Para un mejor despliegue de la información se desarrolló reportes *ad-hoc* y cuadros de mando parametrizados. La combinación de estos elementos da como resultado una solución de BI que permite el monitoreo eficiente de indicadores previos a la acreditación universitaria y reducir el tiempo de respuesta y recursos en la generación de reportes.

## ABSTRACT

The CAACES (Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior), develops evaluation processes to accredit Higher Education Institutions of Ecuador. As a result of this process several universities were closed and others have been discredited. In order to support the accreditation process we propose an analytical data mart to a) obtain suitable information for its fast understanding and management, b) avoid that the dispersion of the data required for the university accreditation. The implementation was initially developed with the indicators of the students criterion according to Institutional Evaluation Model of Universities and Polytechnic Schools incorporated by CEAACES and in the future, indicators can be added for other accreditation criteria. The project was developed using the Kimball methodology and an open source tool of Business Intelligence (BI), the Pentaho suite. For the correct implementation and integration of the data mart, we executed the process for extraction, transformation and data load from two data sources. The developed data mart allows to visualize the information through online analytical processing tools. For a better display of the information, ad-hoc reports and parametrized dashboards were developed. The combination of these elements results in a BI solution that allows the efficient monitoring of indicators prior to university accreditation, reducing the response time and resources in the generation of reports.

**Keywords:** Business Intelligence, data mart, OLAP cubes, ETL, university accreditation.



# CAPÍTULO 1

## 1. INTRODUCCIÓN

El desarrollo de un *data mart*, subconjunto de un almacén de datos (*data warehouse*) es una solución para organizar adecuadamente la información que se encuentra dispersa facilitando la consulta por diferentes criterios.

En el artículo “*The future is Bright*” escrito por Ralph Kimball se menciona que: “Un almacén de datos nunca fue más valioso e interesante de lo que es ahora. Tomar decisiones basadas en datos es fundamental y obvia, que la actual generación de usuarios de negocios y diseñadores de almacenamiento, no pueden imaginar un mundo sin acceso a los datos” [1].

La utilización de un almacén de datos mejora la eficiencia y efectividad de la organización, la habilidad para diseminar el conocimiento dentro de la empresa y con los socios comerciales, mejorando la toma de decisiones y la competitividad de la organización [2]. Adicionalmente permite realizar reingeniería dentro de la organización, consolidar la información, medir los resultados del negocio de una manera más efectiva y dar un mejor servicio [3].

La razón principal detrás de la Inteligencia de Negocios, es el deseo de mejorar la toma de decisiones y lograr mejores resultados organizacionales, ya que estas permiten descubrir los problemas y oportunidades a tiempo, y además realizar análisis más detallados [4].

Un almacén de datos permite que la organización tenga un historial completo, no solamente la información transaccional y operacional, que se almacenará en una base de datos diseñada para ayudar al análisis y divulgación eficiente de datos.

En cuanto a los *data marts* se los define como una sub-parte del almacén de datos, o como un almacén de datos pequeño que se relaciona con un departamento específico de una organización. Por otra parte para facilitar el proceso de toma de decisiones es necesario el desarrollo de un componente de visualización [5].

El desarrollo del *data mart* se lo realizará orientado a procesos, es decir se buscará la integración de (BI) con los sistemas que soportan los diferentes procesos para evitar la dispersión y redundancia de datos [6].

## **1.1. Planteamiento del problema**

### **1.1.1. Descripción del problema**

En el Ecuador el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES) ejecuta procesos de evaluación quinquenal con fines de acreditación a todas las universidades y escuelas politécnicas del Sistema de Educación Superior del Ecuador, para determinar la categorización de las instituciones según lo determina la Ley Orgánica de Educación Superior (LOES) [7]. Como resultado de este proceso varias universidades han sido clausuradas y otro número considerable han quedado descreditadas.

Para llevar adelante el proceso de evaluación de la calidad es necesario, determinar las condiciones de la institución, carrera o programa académico, mediante la recopilación sistemática de datos cuantitativos y cualitativos que permitan emitir un juicio o diagnóstico, analizando sus componentes, funciones, procesos, a fin de que sus resultados sirvan para reformar y mejorar el programa de estudios, carrera o institución [8].

Por este motivo, en este trabajo se propone el desarrollo de un *data mart* que permita obtener información oportuna para su rápido entendimiento y gestión, evitando que los datos requeridos para la acreditación universitaria, estén dispersos y no permitan un análisis integrado para procesos de toma de decisiones.

## **1.2. Objetivos**

### **1.2.1. Objetivo general**

Desarrollar un *data mart* para el manejo de indicadores de acreditación universitaria.

## 1.2.2. Objetivos específicos

- Desarrollar un *data mart* para el manejo de indicadores de acreditación universitaria para el caso de estudio utilizando la metodología de Raph Kimball. Se implementará con al menos el 20% de los indicadores requeridos.
- Desarrollar e implementar un componente de visualización de datos que permita monitorear los indicadores en el tiempo.
- Probar la funcionalidad del *data mart* con datos reales de la Escuela Politécnica Nacional.

## 1.3. Alcance

El *data mart* para el manejo de indicadores de acreditación universitaria se lo implementará en la EPN. La implementación se realizará inicialmente con los indicadores del criterio estudiantes según el Modelo de Evaluación Institucional de Universidades y Escuelas Politécnicas incorporado por el CEAACES. Actualmente la EPN tiene la información referente a los indicadores de acreditación universitaria en bases de datos relacionales, hojas Excel y reportes, esta información se encuentra dispersa y fluye como reportes que son presentados como informe a las autoridades para la respectiva toma de decisiones, teniendo como resultado decisiones basadas en información que no está actualizada.

Al implementar un *data mart* para el manejo de indicadores de acreditación universitaria, se pretende brindar un sistema para monitorear internamente los indicadores previos a la acreditación de la universidad, además se reduciría el tiempo y recursos para obtener informes que permitan el análisis de los mismos.

La solución permitirá integrar toda la información que resida en cualquier otra plataforma que sea propiedad de la institución; logrando integrar con bases de datos SQL Server y PostgreSQL, además de otros datos que provengan de archivos planos como CSV, Excel, XML y txt.

Para extraer los datos de diversas fuentes, realizar depuraciones y prepararlos para luego cargarlos en un almacén de datos, la plataforma de BI contará con métodos de extracción, transformación y carga (ETL).

El enfoque de BI se basa en la elaboración de modelos multidimensionales para el monitoreo específico de los indicadores de acreditación que son evaluados por parte del CEAACES, permitiendo obtener información oportuna para su rápido entendimiento y gestión por parte de la universidad.

La arquitectura para la solución propuesta será escalable, flexible e integrable para poder adaptar nuevos requerimientos que el CEAACES pide a las Instituciones de Educación Superior (IES) para las respectivas acreditaciones. Por ejemplo, implementar nuevos reportes analíticos.

## **1.4. Base teórica**

### **1.4.1. Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES)**

El CEAACES es un organismo técnico, público y autónomo que se encarga de ejercer la rectoría política para la evaluación, acreditación y el aseguramiento de la calidad de las Instituciones de Educación Superior del Ecuador, sus programas y carreras. Para lograrlo, realizan procesos continuos de evaluación y acreditación que evidencien el cumplimiento de las misiones, fines y objetivos de estas [7].

### **1.4.2. Estructura del modelo de evaluación incorporado por el CEAACES**

La estructura del modelo de evaluación incorporado por el CEAACES [9], se organiza en torno a seis criterios de evaluación que consideran aspectos amplios de la calidad, y están relacionados con las funciones sustantivas de las universidades y escuelas politécnicas, así como los procesos, las condiciones y los recursos que permiten la ejecución adecuada de las mismas. El modelo de evaluación está concebido para la ejecución de un proceso de evaluación externa.

El proceso de evaluación externa según el artículo 100 de la LOES “es el proceso de verificación que el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior CEAACES realiza a través de pares académicos de la totalidad o de las actividades institucionales o de una carrera o programa para determinar que su desempeño cumple con las características y estándares de calidad de las IES y que sus actividades se realizan en concordancia con la misión, visión, propósitos y objetivos

institucionales o de carrera, de tal manera que pueda certificar ante la sociedad la calidad académica y la integridad institucional LOES, Art. 100” [10].

La consecuencia lógica de esta definición es que cualquier evaluación institucional externa debe contemplar la misión y los objetivos, de tal manera que a través de la planificación, la organización, los procesos, los recursos y las condiciones internas se garantice el cumplimiento de los mismos. De ello se desprende que la estructura general del modelo de evaluación se organice en un primer nivel de acuerdo a los criterios que se detallan en la Figura 1.

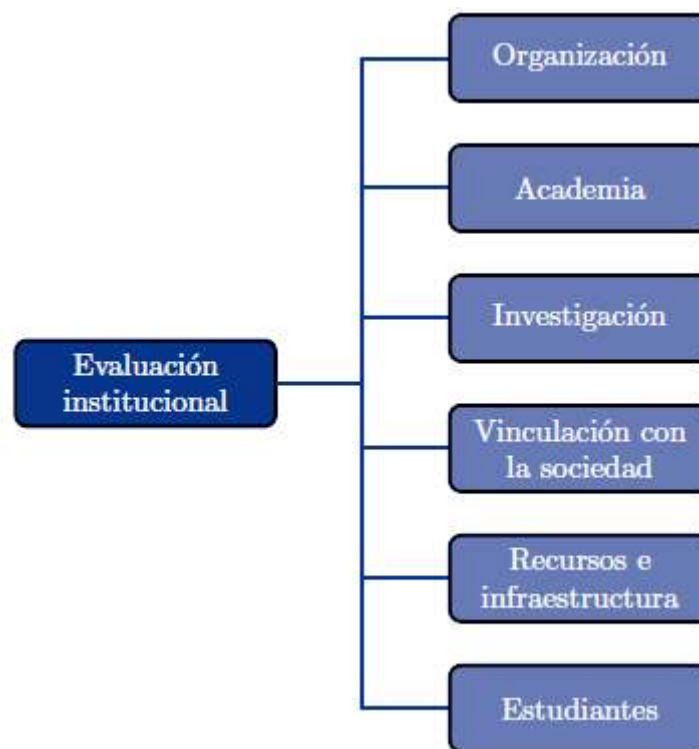


Figura 1. Descripción general del modelo [9].

En un segundo nivel, la estructura del modelo de evaluación considera subcriterios que abordan aspectos conceptuales más específicos y atributos de los criterios; los subcriterios son susceptibles de ser medidos a través de indicadores de evaluación. Los indicadores de evaluación están concebidos de manera que representan atributos específicos que son susceptibles de ser evaluados en términos de los estándares de evaluación que representan las cualidades deseables en las características de los procesos, la organización, la disponibilidad de recursos, y los resultados obtenidos.

### **1.4.3. Metodología de valoración**

En la evaluación del entorno de aprendizaje el CEAACES asigna valoraciones de desempeño (utilidades) a cada uno de los indicadores del modelo; esta asignación se efectúa de acuerdo con ciertas funciones de utilidad. Para el caso de los indicadores cualitativos y en base a la resolución N°. 024-CEAACES-SO-06-2017 de fecha 22 de marzo de 2017, donde se aprueba la modificación de la escala de valoración de los indicadores cualitativos (ver Anexo I), la escala es la siguiente:

- Altamente satisfactorio (1).- Alcanza el estándar.
- Satisfactorio (0,667).- Presenta debilidades no estructurales que pueden ser solventadas a través de la consolidación o mejora de los procesos ya implementados.
- Poco satisfactorio (0,333).- No alcanza el estándar evidenciando debilidades estructurales que comprometen la consecución de los objetivos, sin embargo, existen procesos viables a ser implementados.
- Deficiente (0).- No alcanza el estándar evidenciando debilidades estructurales que comprometen la consecución de los objetivos y/o la información presentada no permite el análisis.

Para los indicadores cuantitativos la valoración se realiza mediante una función matemática que adjudica un valor de desempeño entre 0 y 1. El desempeño de cada IES en el entorno de aprendizaje se obtiene a través de una suma ponderada de los desempeños de todos los indicadores; los pesos o ponderaciones utilizados son los que recibe cada indicador en el modelo, y que reflejan su importancia relativa en el mismo.

A continuación se presenta una descripción de los elementos constitutivos de la metodología de valoración; estos son: variables, indicadores, funciones de utilidad, estándares, evidencias, pesos y el criterio de aprobación de la evaluación del entorno de aprendizaje.

#### **1.4.3.1. Variables**

La primera etapa de cuantificación del desempeño se realiza a nivel de variables. Posteriormente se definen los indicadores. Estas variables son calculadas a partir de datos reportados por las IES (en el sistema GIES del CEAACES) [9], datos extraídos de

la información documental remitida al CEAACES o recolectados durante el proceso de evaluación in situ. Para el caso de los indicadores cualitativos la variable y el indicador coinciden. Con los datos validados se calculan las variables necesarias para construir los indicadores del modelo.

#### **1.4.3.2. Indicadores**

Los indicadores del modelo son de dos tipos: cualitativo y cuantitativo. El modelo tiene 19 indicadores cualitativos y 25 indicadores cuantitativos.

Los indicadores cualitativos se determinan automáticamente a partir de su variable. Para el caso de los indicadores cuantitativos, una vez que se cuenta con las variables del modelo validadas, se procede con su cálculo.

#### **1.4.3.3. Estándares**

Para los indicadores cualitativos, la definición de sus estándares se muestra de manera explícita en el modelo de evaluación, por ejemplo; el estándar para admisión a estudios de posgrado: “La institución de educación superior cuenta con políticas, procedimientos y estándares claros que se aplican en todos los programas de posgrado, considerando sus particularidades, a todos los postulantes de su oferta académica” [9].

Para los indicadores cuantitativos, los estándares se encuentran implícitos en su función de utilidad. Las funciones de utilidad de los indicadores cuantitativos se han definido con base en la situación y tendencia deseada del sistema de educación superior en el país, lo que convierte al estándar de calidad en un referente intrínseco del sistema.

#### **1.4.3.4. Evidencias**

Las evidencias son fuentes de información de carácter documental que respaldan la ejecución de un proceso o actividad académica; en particular, permiten justificar los valores de ciertas variables entregados por las instituciones de educación superior.

### 1.4.3.5. Pesos de indicadores en el modelo

El desempeño global de la IES se determina mediante la suma ponderada de los desempeños obtenidos en cada indicador. No todos los indicadores pesan igual, pues no tienen la misma importancia relativa en el modelo. De partida, los seis criterios no tienen la misma ponderación. El peso global de un indicador viene determinado por los pesos del criterio y del subcriterio al que pertenece y, dentro de estos últimos, por la ponderación recibida con respecto a los otros indicadores. Los pesos de los criterios, subcriterios e indicadores fueron socializados como parte del proceso de evaluación. Los pesos correspondientes a cada indicador se detallan en la Tabla 1.

Tabla 1. Indicadores con sus respectivos pesos [9].

| <b>Criterio</b>       | <b>Subcriterio</b>          | <b>Nombre</b>                  | <b>Tipo</b>        | <b>Peso</b>  |        |
|-----------------------|-----------------------------|--------------------------------|--------------------|--------------|--------|
| Organización          | Planificación institucional | Planificación Estratégica      | Cualitativo        | 0,0100       |        |
|                       |                             | Planificación Operativa        | Cualitativo        | 0,0100       |        |
|                       | Ética institucional         | Rendición de cuentas           | Cualitativo        | 0,0100       |        |
|                       |                             | Ética y responsabilidad        | Cualitativo        | 0,0100       |        |
|                       | Gestión de calidad          | Políticas y procedimientos     | Cualitativo        | 0,0100       |        |
|                       |                             | Sistemas de información        | Cualitativo        | 0,0080       |        |
|                       |                             | Oferta Académica               | Cualitativo        | 0,0080       |        |
|                       |                             | Información para la evaluación | Cuantitativo       | 0,0140       |        |
|                       | Academia                    | Posgrado                       | Formación posgrado | Cuantitativo | 0,0800 |
|                       |                             |                                | Doctores TC        | Cuantitativo | 0,0650 |
| Posgrado en formación |                             |                                | Cuantitativo       | 0,0160       |        |
| Dedicación            |                             | Estudiantes por docente TC     | Cuantitativo       | 0,0270       |        |



|                             |                              |                                   |              |        |
|-----------------------------|------------------------------|-----------------------------------|--------------|--------|
|                             |                              | Titularidad TC                    | Cuantitativo | 0,0200 |
|                             |                              | Horas clase TC                    | Cuantitativo | 0,0140 |
|                             |                              | Horas clase MT/TP                 | Cuantitativo | 0,0120 |
|                             | Carrera docente              | Titularidad                       | Cuantitativo | 0,0130 |
|                             |                              | Evaluación docente                | Cualitativo  | 0,0070 |
|                             |                              | Dirección mujeres                 | Cuantitativo | 0,0080 |
|                             |                              | Docencia mujeres                  | Cuantitativo | 0,0080 |
|                             |                              | Remuneración TC                   | Cuantitativo | 0,0720 |
|                             | Remuneración MT/TP           | Cuantitativo                      | 0,0180       |        |
| Investigación               | Institucionalización         | Planificación de la investigación | Cualitativo  | 0,0300 |
|                             |                              | Gestión de recursos investigación | Cualitativo  | 0,0100 |
|                             | Resultados                   | Producción científica             | Cuantitativo | 0,0900 |
|                             |                              | Producción regional               | Cuantitativo | 0,0200 |
|                             |                              | Libros revisados por pares        | Cuantitativo | 0,0600 |
| Vinculación con la sociedad | Institucionalización         | Planificación de la vinculación   | Cualitativo  | 0,0150 |
|                             |                              | Gestión de recursos vinculación   | Cualitativo  | 0,0050 |
|                             | Resultados de la vinculación | Proyectos de vinculación          | Cualitativo  | 0,0100 |
| Recursos e infraestructura  | Infraestructura              | Calidad aulas                     | Cuantitativo | 0,0300 |
|                             |                              | Espacios de bienestar             | Cualitativo  | 0,0300 |
|                             |                              | Oficinas a TC                     | Cuantitativo | 0,0300 |

|             |                    |                               |              |        |
|-------------|--------------------|-------------------------------|--------------|--------|
|             | TIC                | Salas MT-TP                   | Cuantitativo | 0,0120 |
|             |                    | Conectividad                  | Cuantitativo | 0,0180 |
|             |                    | Plataforma gestión académica  | Cualitativo  | 0,0200 |
|             | Bibliotecas        | Gestión biblioteca            | Cualitativo  | 0,0150 |
|             |                    | Libros por estudiante         | Cuantitativo | 0,0250 |
|             |                    | Espacio estudiantes           | Cuantitativo | 0,0200 |
| Estudiantes | Condiciones        | Admisión estudios de posgrado | Cualitativo  | 0,0150 |
|             |                    | Bienestar estudiantil         | Cualitativo  | 0,0200 |
|             |                    | Acción Afirmativa             | Cualitativo  | 0,0150 |
|             | Programa académico | Tasa de retención             | Cuantitativo | 0,0250 |
|             |                    | Tasa de titulación pregrado   | Cuantitativo | 0,0250 |
|             |                    | Tasa de titulación posgrado   | Cuantitativo | 0,0200 |

#### 1.4.3.6. Funciones de utilidad

En general, los indicadores cuantitativos no tienen un rango unificado de valores. Para que sean comparables y aditivos, los indicadores se transforman a un mismo rango de valores. Esta transformación se realiza a través de funciones matemáticas que se conocen como funciones de utilidad. Para este proceso, el rango unificado va desde cero a uno.

Las funciones de utilidad cumplen dos propósitos dentro de la valoración del desempeño de las IES. El primer propósito, como ya se mencionó, consiste en la transformación del rango de valores posibles de un indicador a un rango de valores unificado; sin embargo, esta transformación no siempre es proporcional, lo que implica que existan distintos tipos de funciones de utilidad con diferentes formas gráficas. El segundo propósito es el de definir estándares de calidad como referentes del modelo.

#### 1.4.4. Inteligencia de negocios

Según Peña [11], el término Inteligencia de Negocios en inglés (BI), procura caracterizar una amplia variedad de tecnologías, plataformas de software, especificaciones de aplicaciones y procesos. El objetivo primario de la Inteligencia de Negocios es contribuir a tomar decisiones que mejoren el desempeño de la empresa y promover su ventaja competitiva en el mercado. En resumen, la Inteligencia de Negocios faculta la organización a tomar mejores decisiones y más rápidas. Este concepto se requiere analizar desde tres perspectivas: Hacer mejores decisiones más rápido, convertir datos en información, y usar una aplicación relacional para la administración.

En relación con la conversión de datos en información, la Inteligencia de Negocios se orienta a establecer el puente que una las grandes cantidades de datos y la información que los tomadores de decisiones requieren cotidianamente [11].

##### 1.4.4.1. Proceso de BI

En la Figura 2 se ilustra el proceso de BI con sus diferentes fases que serán explicadas a continuación:



Figura 2. Fases del proceso BI [12].

Según Bernabeu [12], el proceso está compuesto por 5 fases que se detallan a continuación:

Fase 1. Dirigir y planear: En esta fase es donde se recolectan los requerimientos de información específicos.

Fase 2. Recolección de información: Se extrae la información de la empresa de diferentes fuentes, tanto internas como externas.

Fase 3. Procesamiento de datos: Se integran y cargan los datos en crudo en un formato utilizable para el análisis.

Fase 4. Análisis y producción: Se trabaja sobre los datos extraídos e integrados, utilizando herramientas y técnicas propias de la tecnología BI, para crear inteligencia. Al final de esta fase se entregará la información mediante reportes, cuadros de mando, etc.

Fase 5. Difusión: Se les entrega a los usuarios las herramientas necesarias para que puedan explorar los datos de manera sencilla.

#### **1.4.4.2. Ventajas y desventajas de implementar herramientas de BI en una organización**

Implementar herramientas de BI dentro de la organización permite soportar las decisiones que se toman; al nivel interno ayuda en la gestión de la organización [13] y del lado externo produce ventajas sobre sus competidores [14].

Existen ocasiones en las cuales no se pueden lograr todos los beneficios que tiene BI; debido al proceso que lleva consigo implementar un proyecto de estas características, se puede cometer errores en la definición del planteamiento de las necesidades de conocimiento de la empresa; el no determinar la magnitud de los problemas de información a solucionar generalmente repercute en el fracaso del proyecto [15].

#### **1.4.5. Análisis de plataformas de BI**

Existen diversas herramientas para implementar sistemas de BI, por lo tanto se hizo el análisis de tres plataformas de procesamiento, almacenamiento y presentación de información y de esta manera conocer ventajas y desventajas de cada una, a continuación se listan las plataformas:

##### **1.4.5.1. Oracle Business Intelligence**

Es una herramienta completa que cuenta con varios paneles intuitivos que proporcionan enlaces para las actividades más comunes. La presentación de componentes visuales permite que la presentación de datos sea intuitiva, relevante y fácil de entender. Cuenta con funciones integradas de análisis avanzado para predecir y optimizar los resultados.

Dentro de las características que tiene la herramienta, a continuación se mencionan las más relevantes [16]:

- Permite el empoderamiento de los tomadores de decisiones clave para encontrar rápidamente respuestas a preguntas predictivas y estadísticas.
- Mejora la eficiencia de implementación y rendimiento de consultas más rápido.
- Permite que los líderes empresariales tengan acceso seguro y exploren datos, sin importar dónde se encuentren, a través de dispositivos móviles y la nube.

#### **1.4.5.2. MicroStrategy**

La empresa Microstrategy ofrece un producto *Enterprise Business Intelligence Suites* (EBIS) denominado Microstrategy Business Intelligence [17]. Según Iñigo [18] las principales características de la herramienta son las siguientes:

- Permite transformar grandes volúmenes de datos en paneles e informes intuitivos orientados a sectores empresariales.
- Tiene bases de datos analíticas avanzadas y predictivas.
- Ofrece una gran escalabilidad y orientación a *Big Data*.

#### **1.4.5.3. Pentaho Business Analytics**

Herramienta para inteligencia de negocios de *Open Source* (código abierto), que integra, fusiona y analiza todos los datos que causan impacto en los resultados empresariales. La innovación es continua, logrando que la plataforma de análisis sea flexible, unificada y moderna que ayuda a las organizaciones a mejorar la canalización de datos analíticos [12][19]. Entre sus productos destacan *Pentaho Data Integration* (PDI) y *Pentaho Business Analytics*.

*Pentaho Data Integration* permite que las organizaciones pueden acceder a datos de fuentes complejas y heterogéneas para combinarlas con datos relacionales existentes para producir información de alta calidad y lista para analizar. A continuación se mencionan las principales características [19]:

- Ofrece procesos ETL de alto rendimiento gracias a un poderoso motor de transformaciones multiproceso.
- Tiene una interfaz intuitiva de arrastrar y soltar para simplificar la creación de procesos ETL.
- Ofrece una amplia conectividad con prácticamente cualquier fuente de datos, ya sea en archivos planos, RDBMS, Big Data, APIs o en la nube.
- Permite la integración con bases de datos transaccionales que incluyen Oracle, DB2, PostgreSQL, MySQL y otros.

*Pentaho Business Analytics* permite que los usuarios de negocios descubran y exploren prácticamente cualquier dato. Los usuarios pueden crear informes y tableros, así como visualizar y analizar datos en múltiples dimensiones. Se lista a continuación las principales características [19]:

- Permite realizar análisis y visualización ad hoc.
- Ofrece la generación de paneles de control e informes interactivos basados en la web para usuarios empresariales.
- Ofrece la posibilidad de realizar análisis predictivo.

#### **1.4.6. Proceso ETL**

Un proceso ETL, que significa extracción, transformación y carga, consiste en realizar procesos que impliquen la obtención de datos, la transformación de dichos datos con el fin de generar conocimiento mediante la información almacenada en los sistemas de información que posee la organización y la carga de dichos datos en sistemas de almacenamiento dedicados para ser proveídos a los sistemas de consulta de información basados en la inteligencia de negocios [20].

Según Cano [21], el proceso ETL se divide en 5 subprocesos que se describen a continuación:

- Extracción: Este proceso recupera los datos físicamente de las distintas fuentes de información. En este momento disponemos de los datos en bruto.
- Limpieza: Este proceso recupera los datos en bruto y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos y completa los valores vacíos, es decir se transforman los datos, siempre que sea

posible, para reducir los errores de carga. En este momento disponemos de datos limpios y de alta calidad.

- Transformación: Este proceso recupera los datos limpios y de alta calidad y los estructura en los distintos modelos de análisis. El resultado de este proceso es la obtención de datos limpios, consistentes y útiles.
- Integración: Valida que los datos que cargamos en el almacén de datos son consistentes con las definiciones y formatos del almacén de datos; los integra en los distintos modelos de las distintas áreas de negocio que hemos definido en el mismo. Estos procesos pueden ser complejos.
- Actualización: Este proceso es el que nos permite añadir los nuevos datos al almacén de datos.

## **1.5. Definición de términos**

### **1.5.1. OLTP (On-Line Transactional Processing)**

Morales [22] define al sistema transaccional OLTP como bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico, y puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales. El acceso a los datos está optimizado para tareas frecuentes de lectura y escritura (por ejemplo, la enorme cantidad de transacciones que tienen que soportar las BD de bancos o hipermercados diariamente).

### **1.5.2. OLAP (On Line Analytic Processing)**

OLAP es la sigla en inglés de *on line analytic processing*, o procesamiento de análisis en línea, el cual es un proceso analítico de datos en línea que genera la posibilidad de seleccionar y extraer la información desde el enfoque que lo requiera. Por medio de un análisis OLAP, cualquier miembro de la dirección de una organización podría analizar los resultados de la información para la cual está configurado el análisis OLAP. Por ejemplo, si se ha creado una configuración de análisis OLAP para las ventas, los directivos de la organización pueden visualizar sus ventas e ir filtrando información de acuerdo con criterios como zona, equipos de ventas, vendedores, etc., hasta encontrar la información relevante para la toma de decisiones [23].

### 1.5.2.1. MOLAP (Multidimensional On Line Analytical Processing)

Los sistemas MOLAP utilizan verdaderas bases de datos multidimensionales para el almacenamiento de datos [24].

### 1.5.2.2. ROLAP (Relational On Line Analytic Processing)

Un sistema ROLAP utiliza una base de datos relacional para el almacenamiento de datos pero está mapeada de modo que una herramienta OLAP ve los datos como multidimensionales [24].

La simulación de un modelo multidimensional se realiza a partir de una base de datos relacional basada en un esquema estrella o copo de nieve.

### 1.5.3. Características distintivas entre OLAP y OLTP

Tabla 2. Características distintivas entre OLAP y OLTP [25].

| <b>Características distintivas</b>        | <b>OLTP</b>   | <b>OLAP</b>  |
|---|---|--|
| Los usuarios y la orientación del sistema | Orientado al cliente y se utiliza para el procesamiento de transacciones y consultas por parte de empleados, clientes y profesionales de la tecnología de la información. | Orientado al mercado y se utiliza para el análisis de datos por parte de trabajadores del conocimiento, incluidos gerentes, ejecutivos y analistas.  |
| Contenido de los datos                    | Gestiona los datos actuales en un formato demasiado detallado.  | Maneja grandes cantidades de datos históricos, proporciona facilidades para el resumen y la agregación. Además, la información se almacena y administra en diferentes niveles de granularidad, hace que los datos sean |



|                            |  |  |
|----------------------------|--|--|
|                            |  | más fáciles de usar en la toma de decisiones.  |
| Diseño de la base de datos | Generalmente adopta un modelo de datos de relación entre entidades y un diseño de base de datos orientado a aplicaciones.    | Adopta un modelo de estrella o de copo de nieve y un diseño de base de datos orientado a temas.  |
| Vistas                     | Se enfoca principalmente en los datos actuales sin hacer referencia a datos históricos o datos en diferentes organizaciones. | Abarca múltiples versiones de un esquema de base de datos, debido al proceso evolutivo de una organización. Debido a su enorme volumen, los datos OLAP se comparten en múltiples medios de almacenamiento. |
| Patrones de acceso         | Consisten principalmente en transacciones atómicas cortas. Requiere concurrencia, control y mecanismos de recuperación.      | Son principalmente operaciones de solo lectura, aunque muchas podrían ser consultas complejas.   |

#### 1.5.4. Almacén de datos

Ralph Kimball define el almacén de datos como “una copia de las transacciones de datos específicamente estructurada para la consulta y el análisis” [26].

Por otra parte, W.H. Inmon define el almacén de datos como “un sistema orientado al usuario final, integrado, con variaciones en el tiempo para respaldar las decisiones de la administración. El almacén de datos contiene datos corporativos granulares” [27]. Esta definición es conceptualizada en términos de las características del almacén de datos:

- Orientado a temas

Los datos que están almacenados deben estar orientados a un tema y organizados según un mismo evento, objeto de la vida real o un área en particular. Por ejemplo, todos los

datos sobre ventas pueden ser consolidados en una única tabla del almacén de datos, logrando que las peticiones de información de ventas sean más fáciles de responder dado que toda esta información se encuentra en el mismo lugar.

- Integrado

Todos los datos de los sistemas que están en operación dentro de una organización están almacenados en el almacén de datos, por lo tanto la estructura debe ser consistente, es decir se debe eliminar inconsistencias existentes entre los diversos sistemas operacionales. La información se estructura en diferentes niveles de detalle y acorde a las necesidades del usuario.

- Histórico

La información que reside en el almacén de datos es solicitada en cualquier momento y sin importar que se acceda a dicha información en diferentes tiempos, esta debe seguir siendo exacta. La diferencia con el ambiente operacional radica en que los datos solicitados tienden a ser valores exactos a partir del momento de acceso.

- No volátil

La información contenida en un almacén de datos es de solo lectura, no puede ser modificada ni eliminada cuando ya se realizó el respectivo almacenamiento, lo que se convierte en información usada únicamente para consultas.

### **1.5.5. Data mart**

Un *data mart* es una base de datos especializada, departamental, orientada a satisfacer las necesidades específicas de un grupo particular de usuarios [28], en otras palabras son subconjuntos de los almacenes de datos diseñados para satisfacer necesidades específicas de un área de una institución.

Un *data mart* puede ser de dos tipos, los dependientes que obtienen sus datos del almacén de datos es decir de un repositorio empresarial centralizado. Y los independientes que no van a depender de un almacén de datos central, sino que obtienen sus datos de fuentes separadas.

### 1.5.6. Enfoque *Top-Down*

Con este enfoque se define primeramente el almacén de datos para luego desarrollar, construir y cargar los *data marts* a partir del mismo.

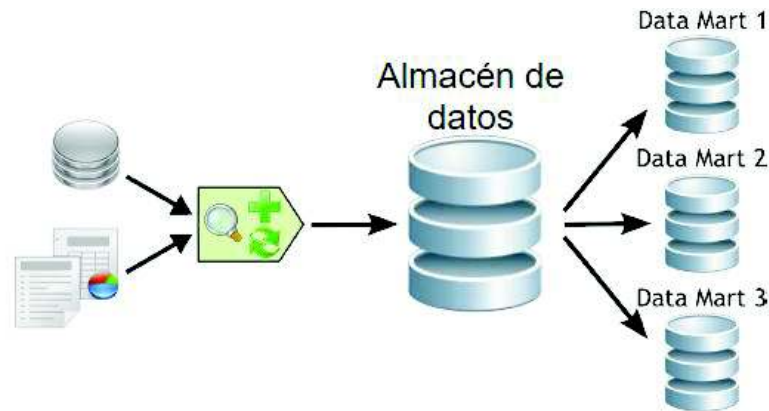


Figura 3. Top-Down [12].

En la Figura 3 se muestra como el almacén de datos es cargado mediante procesos ETL para después alimentar a diferentes *data marts* dependiendo el tema o departamento que corresponda. Una ventaja de este enfoque es que no se debe incurrir en sincronizaciones complicadas de hechos. Como desventaja se tiene que la inversión es grande y el tiempo de construcción es elevado.

### 1.5.7. Enfoque *Bottom-Up*

En este enfoque se definen previamente los *data marts* para después integrarlos en un almacén de datos centralizado. Este enfoque se ilustra en la Figura 4.

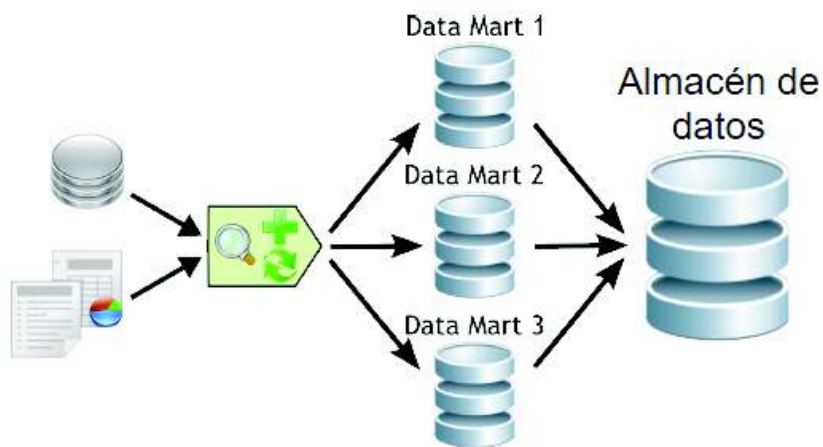


Figura 4. Bottom-up [12].

En la Figura 4 se muestra que para cargar los *data marts* se utilizan procesos ETL, los cuales suministrarán la información adecuada a cada uno de ellos. En muchas ocasiones, los *data marts* son implementados sin que exista el almacén de datos, ya que tienen sus mismas características pero con la particularidad de que están enfocados en un tema específico. Luego de que hayan sido creados y cargados todos los *data marts*, se procederá a su integración con el depósito.

Este enfoque tiene como ventaja que cada *data mart* se crea y pone en funcionamiento en un corto lapso de tiempo y se puede tener una pequeña solución a un costo no tan elevado. La desventaja es tener que sincronizar los hechos al momento de la consolidación en el depósito.

#### **1.5.8. Tablas de dimensiones**

Son componentes integrales de una tabla de hechos. Las tablas de dimensiones son categorías que describen el contexto en el cual se analizan las medidas. Cada dimensión se define mediante una única clave primaria, que sirve como base para la integridad referencial con cualquier tabla de hechos a la que se una [26].

#### **1.5.9. Métricas**

Son valores de tipo numéricos que describen el hecho que se está analizando. Las métricas son valores que se generan al realizar operaciones o transacciones del negocio.

#### **1.5.10. Tabla de hechos (*fact table*)**

Las tablas de hechos contienen, precisamente los hechos que serán utilizados por analistas de negocio para apoyar el proceso de toma de decisiones. Los Hechos son datos instantáneos en el tiempo, que son filtrados, agrupados y explorados a través de condiciones definidas en las tablas de dimensiones [12].

#### **1.5.11. Esquema estrella**

Los esquemas en estrella son estructuras dimensionales implementadas en un sistema de administración de base de datos relacionales. El esquema estrella consiste en tablas

de hechos vinculadas a tablas de dimensiones asociadas a través de relaciones de clave primaria/foránea [26]. En la Figura 5 se puede apreciar un esquema estrella estándar.

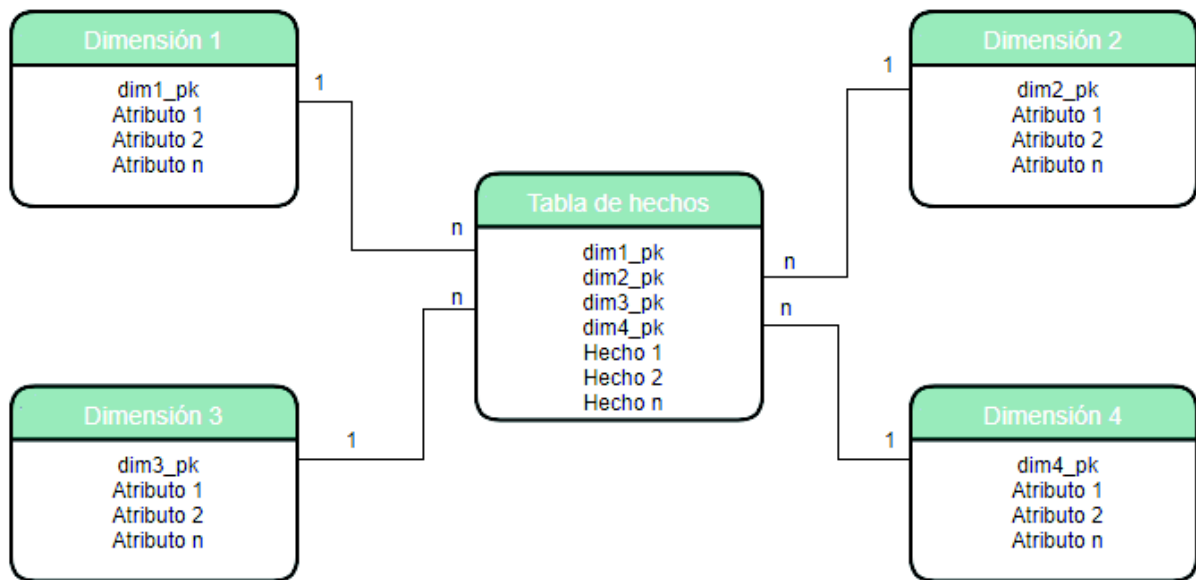


Figura 5. Esquema Estrella Estándar<sup>1</sup>.

### 1.5.12. Área de preparación (*staging area*)

Es un área temporal donde se almacenarán los datos que se necesitan de los sistemas de origen para su posterior análisis. Se recogen los datos estrictamente necesarios para las cargas, y se aplica el mínimo de transformaciones a los mismos. Estos datos al ser temporales se borran al cumplir su función por lo que no se utilizan para creación de reportes [12].

## 1.6. La metodología de Kimball para el desarrollo

La metodología se basa en el ciclo de vida dimensional de un negocio y se basa en 4 principios listados a continuación [26]:

- Concentrarse en el negocio.
- Construir una infraestructura de información adecuada.
- Realizar entregas en incrementos significativos.
- Ofrecer la solución completa.

<sup>1</sup> Fuente: Elaboración propia

El diagrama del ciclo de vida de la metodología de Kimball se ilustra en la Figura 6.

El ciclo comienza con la planificación del proyecto, en este módulo: se evalúa la preparación de la organización para una iniciativa *Data Warehousing / Business Intelligence (DW/BI)*, se establece el alcance preliminar y justificación, se obtiene los recursos y se lanza el proyecto.

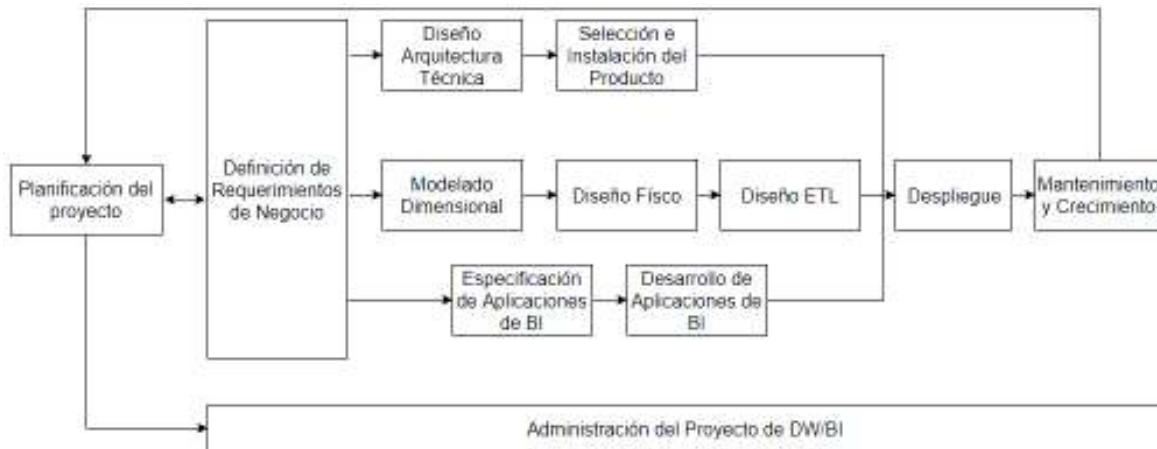


Figura 6. Diagrama del ciclo de vida de la metodología de Kimball [26].

La segunda tarea principal se centra en la definición de requerimientos del negocio. A continuación, hay tres rutas o caminos que se enfocan en 3 diferentes áreas:

- Tecnología: El diseño de la arquitectura técnica establece el marco general para apoyar la integración de diferentes tecnologías. Se obtiene una lista de compras de productos definidos en el diseño y a continuación, se evalúa y se procede a seleccionar los productos específicos.
- Datos: En esta ruta se diseña e implementa el modelo dimensional. Luego se realiza el diseño físico que se enfoca en estrategias de ajuste de rendimiento como la agregación, indexación y el particionamiento. Por último se diseña y desarrolla el sistema de Extracción, Transformación y carga de datos (ETL).
- Aplicaciones de inteligencia de negocios: En esta ruta se encuentran las tareas para diseñar y desarrollar las aplicaciones de negocios para el usuario final. Las aplicaciones de BI se desarrollan en forma de plantillas parametrizadas que cubren las necesidades analíticas definidas por los usuarios del negocio.

### **1.6.1. Planificación del proyecto**

La planificación del proyecto determina el propósito de la implementación del almacén de datos, sus objetivos específicos, el alcance del mismo y un acercamiento inicial a las necesidades de información.

Kimball [26] expone que es prudente evaluar la disposición de la organización antes de proceder. Existen tres factores principales que determinan el éxito del proyecto:

- Tener un fuerte *sponsor* ejecutivo de negocios ya que deben tener una visión clara del impacto del sistema en la organización.
- Tener una fuerte motivación comercial convincente para realizar el proyecto de BI.
- El tercer factor es la viabilidad, la técnica, la de recursos y la más importante, la viabilidad de los datos, es decir, si la organización está recopilando datos reales que respalden los requisitos del negocio. El desarrollo del proyecto se vería afectado si los datos fuentes no están siendo recopilados de manera limpia en la granularidad correcta.

### **1.6.2. Definición de los requerimientos del negocio**

Para tener éxito en el desarrollo de un proyecto de BI, es fundamental interpretar correctamente los requerimientos formulados por los usuarios. Los usuarios finales y sus requerimientos siempre tendrán un impacto en las implementaciones, ya que serán quienes especifiquen las funciones y características que deberá tener el almacén de datos.

Los requerimientos del negocio deben determinar el alcance del almacén de datos, esto se refiere a los datos que debe contener, cómo debe estar organizado, cada cuánto debe actualizarse, quiénes y desde dónde accederán, etc.

Para comprender los requisitos y garantizar la aceptación de la organización, es necesario colaborar con los usuarios empresariales. Se debe aprender tanto como se pueda sobre el negocio. Un buen punto de partida es entrevistar a personal que realmente estará inmerso en el desarrollo del proyecto, esto implica examinar el organigrama de la organización o investigar cómo está estructurado el departamento

donde se implementará una solución de BI, con el fin de identificar a las personas que influirán directamente en el desarrollo del proyecto.

### **1.6.3. Modelo dimensional**

Kimball propone 4 etapas para el proceso de diseño dimensional:

- Identificar el proceso de negocio.
- Establecer el nivel de granularidad.
- Identificar las dimensiones.
- Identificar medidas y tablas de hechos.

#### **1.6.3.1. Identificar el proceso de negocio**

En esta etapa básicamente se determinan los procesos de negocio que serán cubiertos por el *data mart*. Se entiende como proceso de negocio a las actividades empresariales que se realizan en la organización y que normalmente cuenta con sistemas de recolección de datos.

Los procesos de negocio que formarán parte del modelo son elegidos por la dirección de la organización y dependerá del análisis de los requerimientos.

#### **1.6.3.2. Establecer el nivel de granularidad**

Significa especificar el nivel de detalle que requiere alcanzar el modelo del almacén de datos. La granularidad se determinará a partir de los requerimientos y de los datos actuales con los que cuenta la organización.

#### **1.6.3.3. Identificar las dimensiones**

Las dimensiones deben ser definidas en la etapa donde se establece la granularidad. Las dimensiones son donde se recogen los datos de la tabla de hechos, por lo tanto son la base de la misma. Por lo general las dimensiones tienen un conjunto de atributos generalmente de tipo textual, brindando una perspectiva o forma de análisis sobre una medida en la tabla de hechos.



#### **1.6.3.4. Identificar medidas y hechos**

Como última etapa del proceso de diseño dimensional se debe establecer las medidas que surgen de los procesos del negocio. La medida es el atributo de una tabla que se requiera analizar, agrupando sus datos usando los criterios de corte conocidos como dimensión.

Los hechos se usan para definir qué es lo que se quiere medir. Kimball menciona que los hechos candidatos en un diseño deben ser fieles a la granularidad definida, y los hechos que pertenezcan claramente a otra granularidad deben separarse en una tabla de hechos diferente. También aconseja que por lo general los hechos a almacenar en una tabla de hechos sean de tipo numérico, enteros o reales.

#### **1.6.4. Diseño físico**

El modelo dimensional debe traducirse en un diseño físico. En este paso se debe desarrollar nombres y estándares de bases de datos. Algunos elementos de este proceso son: establecer los nombres de las tablas y columnas, el tipo de los datos, permisibilidad de valores nulos y declaraciones de clave.

#### **1.6.5. Construcción del sistema ETL**

Desarrollar el sistema de extracción, transformación y carga consume una gran parte del tiempo y esfuerzo del desarrollo del almacén de datos.

Hay que considerar que el proceso de transformación tiene los procesos adecuados que permitirán convertir o recodificar los datos de las fuentes con el fin de realizar la carga adecuada del modelo físico.

#### **1.6.6. Diseño de la arquitectura técnica**

Kimball expresa lo siguiente: “Un buen conjunto de planos, como cualquier buena documentación, nos ayudará más tarde cuando sea tiempo de remodelar o hacer incorporaciones”.

Al igual que un plano de una casa, la arquitectura técnica es el modelado de los servicios técnicos y la infraestructura del entorno del almacén de datos. El plan de arquitectura sirve como un marco de organización para admitir la integración de tecnologías y aplicaciones. La arquitectura permite detectar problemas a priori y trata de disminuir al iniciar el proyecto los problemas que pudieran surgir durante el desarrollo del mismo.

#### **1.6.7. Definición de las herramientas para la plataforma de BI**

El diseño de la arquitectura es similar a una lista de compras para seleccionar productos que se adaptan al marco del diseño. Antes de la compra de nuevos productos se debe comprender los procesos internos de la organización para comprar *hardware* y *software*. Además, se debe tener claro el costo y el tiempo que se invertirá para el aprendizaje de las herramientas.

Una vez que se realiza la respectiva evaluación y selección de los productos que se adaptan al diseño de la arquitectura se procede con la instalación y pruebas pertinentes de los mismos.

#### **1.6.8. Especificación de aplicaciones de BI**

Siguiendo los requerimientos de la organización, se debe proporcionar aplicaciones al usuario final con el fin de generar informes con la información que quieran obtener. Es importante mencionar que no todos los usuarios deben tener el mismo nivel de acceso a la información, por lo tanto, en este proceso es importante determinar roles o perfiles de usuarios.

#### **1.6.9. Desarrollo de aplicaciones de BI**

En el desarrollo de aplicaciones de BI es importante enfocarse en convenciones de nomenclatura, cálculos, bibliotecas y estándares de codificación los cuales deben establecerse para minimizar retrabajos futuros.

La actividad de desarrollo de la aplicación puede comenzar cuando se completa el diseño de la base de datos, se instalan las herramientas de BI y se carga un subconjunto de datos históricos. Las especificaciones de la plantilla de aplicación de BI deben revisarse

para tener en cuenta los inevitables cambios en el modelo desde que se completaron las especificaciones.

#### **1.6.10. Despliegue**

El despliegue representa la convergencia de tecnología, los datos y las aplicaciones de usuarios finales para el usuario del negocio. Hay factores extras que determinan el correcto funcionamiento de todos estos elementos, entre los cuales se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de *feedback*. Todas estas áreas deben ser tomadas en cuenta antes de que cualquier usuario pueda tener acceso al almacén de datos.

#### **1.6.11. Mantenimiento y crecimiento**

Kimball afirma lo siguiente: “Si el trabajo fue realizado correctamente, inevitablemente habrá una demanda de crecimiento, ya sea para nuevos usuarios, nuevos datos, nuevas aplicaciones de BI o mejoras importantes a los entregables existentes”.

Una vez terminada la implementación, se debe continuar con la administración del entorno existente invirtiendo recursos en las siguientes áreas: soporte al usuario, educación, soporte técnico y soporte al programa de BI.

A diferencia de las iniciativas de desarrollo de sistemas tradicionales, los cambios en el desarrollo deben verse como un signo de éxito, no de fracaso. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

#### **1.6.12. Administración del proyecto**

La administración del proyecto cerciora que las actividades del ciclo de vida se realicen de manera sincronizada, por lo tanto, la administración del proyecto está presente en todo el ciclo de vida. Las actividades principales son: monitorear el estado del proyecto y la articulación de los requerimientos del negocio y restricciones de los sistemas de información para poder manejar correctamente las expectativas en ambos sentidos.

## CAPÍTULO 2

### 2. METODOLOGÍA

#### 2.1. Planificación del proyecto

##### 2.1.1. Definición del proyecto

Se determinó la necesidad de implementar una solución de BI en la EPN con el fin de evitar que la información referente a los indicadores de acreditación universitaria se encuentre dispersa y de esta manera evitar que las decisiones se tomen en base a información que no está actualizada.

##### 2.1.2. Objetivos y alcance del proyecto

Los objetivos y el alcance del proyecto se detallan en el capítulo 1 del presente documento, donde además se propone la solución de problema que ya se definió.

La solución consiste en elaborar un *data mart* que permita el monitoreo interno de los indicadores de acreditación universitaria, reduciendo el tiempo y recursos para la obtención de informes para el análisis de los mismos. Para ello se utilizará un ejemplo práctico sobre un *data mart* específico el cual guiará todo el proceso propuesto.

Para el caso práctico se utilizará el criterio estudiantes que está conformado por dos subcriterios. El primer subcriterio es condiciones con sus respectivos indicadores que son: admisión a estudios de posgrado, bienestar estudiantil y acción afirmativa. El segundo subcriterio es programa académico y sus indicadores son: tasa de retención, tasa de titulación de pregrado y tasa de titulación de posgrado.

##### 2.1.3. Situación actual

Actualmente la información referente a los indicadores de acreditación universitaria se encuentra dispersa y fluye como reportes cuando las autoridades pertinentes solicitan información para la respectiva toma de decisiones.

La Figura 7 ilustra la estructura actual del proceso que se lleva a cabo cuando las autoridades solicitan información para la respectiva toma de decisiones.

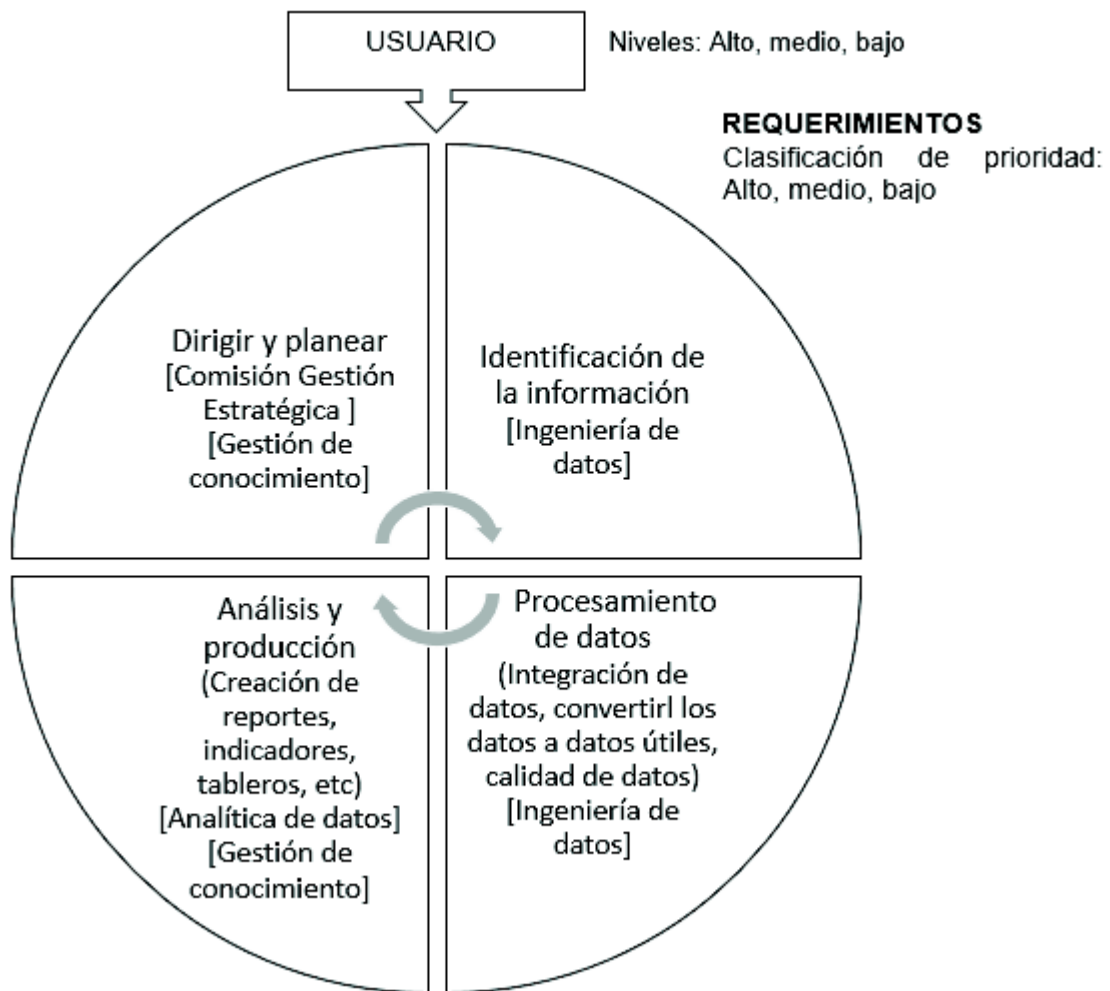


Figura 7. Proceso de levantamiento y entrega de información<sup>2</sup>.

Para el aseguramiento de la calidad de datos en la EPN, la Dirección de Gestión de la Información y Procesos (DGIP) adoptó el *framework Executing Data Quality Projects: Ten steps to Quality Data and Trusted Information*, el cual proporciona un enfoque sistemático para mejorar y crear datos e información de calidad. La Figura 8 describe los pasos para implementar los conceptos clave para evitar problemas de calidad de los datos, para evaluarlos y tratarlos.

<sup>2</sup> Fuente: DGIP

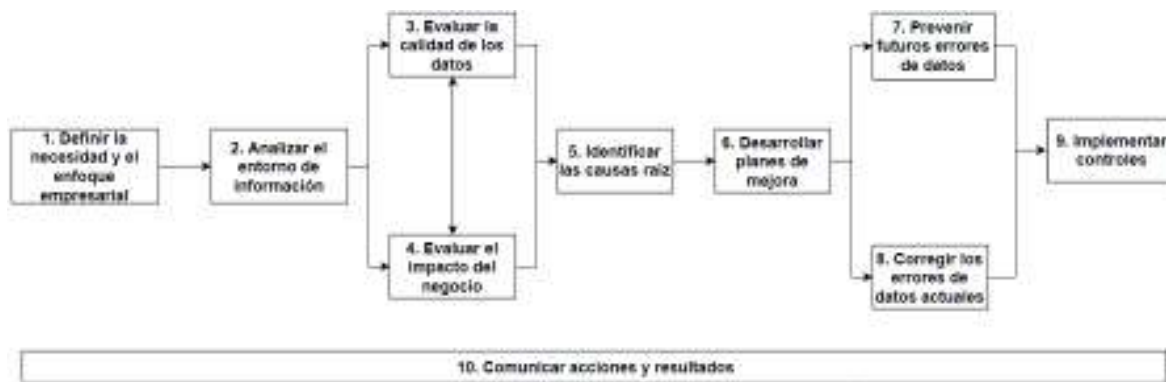


Figura 8. Proceso para asegurar la calidad de datos<sup>3</sup>.

#### 2.1.4. Estrategia de implementación

La implementación de la solución de BI se realizará utilizando herramientas *open Source* de Pentaho Community como son: Pentaho Data Integration para los procesos de ETL y Pentaho User Console para crear informes y tableros, así como visualizar y analizar datos en múltiples dimensiones.

#### 2.1.5. Seleccionar la metodología de desarrollo

Para llevar a cabo el desarrollo del proyecto se ha seleccionado la metodología de Kimball. El diseño se lo realizará con un enfoque *bottom up* (de abajo hacia arriba). Esta orientación se aplica bien al proyecto propuesto porque se puede implementar *data marts* en áreas específicas e irlos integrando en un almacén de datos.

La metodología de Kimball permitirá construir una arquitectura que se adapte fácilmente al cambio ya que a medida que avanza el tiempo, se presentan nuevos requerimientos, por lo tanto es necesario dejar abierta la posibilidad de que en el futuro se desarrollen nuevos *data marts* de acuerdo a las necesidades de la organización.

<sup>3</sup> Fuente: DGIP

## 2.1.6. Recurso humano

Tabla 3. Recurso humano de la solución de BI.

| Talento Humano             | Nombre y Apellido                 | Cargo  |
|----------------------------|-----------------------------------|--|
| Especialista de datos      | Geovanna Marisol Chela Tipantaxi  | Especialista de tecnologías de la información y comunicaciones |
| Especialista de datos      | Tania Guadalupe Gualli Culqui     | Analítica de datos   |
| Administrador del proyecto | Leonardo Alexander Ortiz Yumisaca | Tesista  |

## 2.1.7. Inversión

El desarrollo del proyecto propuesto, con una duración de 6 meses se estima con los siguientes costos:

Tabla 4. Costo de inversión de Hardware.

| Hardware  | Costo            |
|---|------------------|
| HP Envy 14 PC<br>Procesador: Intel Core i5<br>Disco duro: 500 GB<br>Ram: 8 GB | \$ 850.00        |
| <b>TOTAL</b>  | <b>\$ 850.00</b> |

Tabla 5. Costo de inversión de software.

| Software                   | Costo          |
|----------------------------|----------------|
| Ubuntu 18.04 LTS           | \$ 0.00        |
| PostgreSQL                 | \$ 0.00        |
| Pentaho Business Analytics | \$ 0.00        |
| Pentaho Data Integration   | \$ 0.00        |
| Pentaho Schema Workbench   | \$ 0.00        |
| <b>TOTAL</b>               | <b>\$ 0.00</b> |

Tabla 6. Costo recurso humano.

| <b>Recurso humano</b> | <b>Valor mensual</b> | <b>Valor total</b> |
|-----------------------|----------------------|--------------------|
| Desarrollador         | \$ 1200.00           | \$ 7200.00         |
| <b>TOTAL</b>          |                      | <b>\$ 7200.00</b>  |

Tabla 7. Costo de otras inversiones.

| <b>Otras inversiones</b> | <b>Costo</b>     |
|--------------------------|------------------|
| Servicio de internet     | \$ 132.00        |
| Energía eléctrica        | \$ 72.00         |
| Disco externo            | \$ 80.00         |
| Impresora                | \$ 200.00        |
| Resmas de hojas          | \$ 14.00         |
| Transporte               | \$ 60.00         |
| <b>TOTAL</b>             | <b>\$ 558.00</b> |

## **2.2. Definición de requerimientos de la institución**

Esta etapa se centra en definir las necesidades de la organización y conocer el negocio, esto se logra mediante la revisión de informes, reuniones con los directivos y personal que estará inmerso en el desarrollo del proyecto.

En el anexo II se presentan los resultados obtenidos en la entrevista realizada al equipo de apoyo tecnológico en los procesos de acreditación institucional.

A continuación se describen los requerimientos para el desarrollo de la solución.

### **2.2.1. Requerimientos de negocio**

Tabla 8. Cuadro de requerimiento de negocio N°01.

|                |       |            |                                       |
|----------------|-------|------------|---------------------------------------|
| Identificador: | RN-01 | Nombre:    | Reducir tiempo para obtener informes. |
| Verificable:   | Si    | Fecha:     | 05/20/2018                            |
| Prioridad:     | Alta  | Necesidad: | Si                                    |



|              |  |
|--------------|--|
| Descripción: | El sistema deberá reducir el tiempo para obtener informes que permitan el análisis de los indicadores de acreditación universitaria. |
|--------------|--|

Tabla 9. Cuadro de requerimiento de negocio N°02.

|                |   |            |  |
|----------------|---|------------|--|
| Identificador: | RN-02   | Nombre:    | Centralizar la información de los indicadores de acreditación universitaria. |
| Verificable:   | Si  | Fecha:     | 05/20/2018   |
| Prioridad:     | Alta  | Necesidad: | Si   |
| Descripción:   | El sistema deberá centralizar la información útil de los indicadores de acreditación universitaria a fin de realizar análisis integrados para procesos de toma de decisiones. |            |  |

## 2.2.2. Requerimientos de la solución

### 2.2.2.1. Requerimientos funcionales

Tabla 10. Cuadro de requerimiento funcional N°01.

|                |   |            |                                   |
|----------------|---|------------|-----------------------------------|
| Identificador: | RF-01   | Nombre:    | Valor obtenido de los indicadores |
| Verificable:   | Si  | Fecha:     | 05/20/2018                        |
| Prioridad:     | Alta  | Necesidad: | Si                                |
| Descripción:   | La solución deberá proporcionar el valor obtenido de los indicadores agrupados por criterio, subcriterio y periodo académico. |            |                                   |

Tabla 11. Cuadro de requerimiento funcional N°02.

|                |  |            |                                  |
|----------------|--|------------|----------------------------------|
| Identificador: | RF-02  | Nombre:    | Peso obtenido de los indicadores |
| Verificable:   | Si   | Fecha:     | 05/20/2018                       |
| Prioridad:     | Alta   | Necesidad: | Si                               |
| Descripción:   | La solución deberá proporcionar el peso obtenido de los indicadores agrupados por criterio, subcriterio y periodo académico. |            |                                  |

Tabla 12. Cuadro de requerimiento funcional N°03.

|                |  |            |                                      |
|----------------|--|------------|--------------------------------------|
| Identificador: | RF-03  | Nombre:    | Utilidad obtenida de los indicadores |
| Verificable:   | Si   | Fecha:     | 05/20/2018                           |
| Prioridad:     | Alta   | Necesidad: | Si                                   |
| Descripción:   | La solución deberá proporcionar la utilidad obtenida de los indicadores agrupados por criterio, subcriterio y periodo académico. |            |                                      |

Tabla 13. Cuadro de requerimiento funcional N°04.

|                |   |            |                               |
|----------------|---|------------|-------------------------------|
| Identificador: | RF-04   | Nombre:    | Peso obtenido por subcriterio |
| Verificable:   | Si  | Fecha:     | 05/20/2018                    |
| Prioridad:     | Alta  | Necesidad: | Si                            |
| Descripción:   | La solución deberá proporcionar el peso obtenido de los subcriterios agrupados por periodo académico. |            |                               |

Tabla 14. Cuadro de requerimiento funcional N°05.

|                |   |            |                                   |
|----------------|---|------------|-----------------------------------|
| Identificador: | RF-05   | Nombre:    | Utilidad obtenida por subcriterio |
| Verificable:   | Si  | Fecha:     | 05/20/2018                        |
| Prioridad:     | Alta  | Necesidad: | Si                                |
| Descripción:   | La solución deberá proporcionar la utilidad obtenida de los subcriterios agrupados por periodo académico. |            |                                   |

Tabla 15. Cuadro de requerimiento funcional N°06.

|                |  |            |                            |
|----------------|--|------------|----------------------------|
| Identificador: | RF-06  | Nombre:    | Peso obtenido por criterio |
| Verificable:   | Si   | Fecha:     | 05/20/2018                 |
| Prioridad:     | Alta   | Necesidad: | Si                         |
| Descripción:   | La solución deberá proporcionar el peso obtenido de los criterios agrupados por periodo académico. |            |                            |

Tabla 16. Cuadro de requerimiento funcional N°07.

|                |       |            |                                |
|----------------|-------|------------|--------------------------------|
| Identificador: | RF-07 | Nombre:    | Utilidad obtenida por criterio |
| Verificable:   | Si    | Fecha:     | 05/20/2018                     |
| Prioridad:     | Alta  | Necesidad: | Si                             |

|              |   |
|--------------|---|
| Descripción: | La solución deberá proporcionar la utilidad obtenida de los criterios agrupada por periodo académico. |
|--------------|---|

Tabla 17. Cuadro de requerimiento funcional N°08.

|                |  |            |                    |
|----------------|--|------------|--------------------|
| Identificador: | RF-08  | Nombre:    | Desplegar reportes |
| Verificable:   | Si   | Fecha:     | 05/20/2018         |
| Prioridad:     | Alta   | Necesidad: | Si                 |
| Descripción:   | La solución deberá desplegar reportes de los datos contenidos en el <i>data mart</i> . |            |                    |

Tabla 18. Cuadro de requerimiento funcional N°09.

|                |   |            |                          |
|----------------|---|------------|--------------------------|
| Identificador: | RF-09   | Nombre:    | Generar cuadros de mando |
| Verificable:   | Si  | Fecha:     | 05/20/2018               |
| Prioridad:     | Alta  | Necesidad: | Si                       |
| Descripción:   | La solución deberá contener cuadros de mando con controles de parámetros de entrada para una mayor comprensión de la información. |            |                          |

### 2.2.3. Requerimientos no funcionales

Tabla 19. Cuadro de requerimiento no funcional N°01.

|                |   |            |                                    |
|----------------|---|------------|------------------------------------|
| Identificador: | RNF-01  | Nombre:    | Base de datos del <i>data mart</i> |
| Verificable:   | Si  | Fecha:     | 05/20/2018                         |
| Prioridad:     | Alta  | Necesidad: | Si                                 |
| Descripción:   | El <i>data mart</i> se construirá sobre una base de datos PostgreSQL. |            |                                    |

Tabla 20. Cuadro de requerimiento no funcional N°02.

|                |  |            |                   |
|----------------|--|------------|-------------------|
| Identificador: | RNF-02   | Nombre:    | Acceso al sistema |
| Verificable:   | Si   | Fecha:     | 05/20/2018        |
| Prioridad:     | Alta   | Necesidad: | Si                |
| Descripción:   | El acceso al sistema deberá estar restringido por el uso de claves asignadas a cada uno de los usuarios. |            |                   |

Tabla 21. Cuadro de requerimiento no funcional N°03.

|                |   |            |   |
|----------------|---|------------|---|
| Identificador: | RNF-03  | Nombre:    | Acceso a funcionalidades del <i>data mart</i> |
| Verificable:   | Si  | Fecha:     | 05/20/2018                                    |
| Prioridad:     | Alta  | Necesidad: | Si  |
| Descripción:   | Las funcionalidades del <i>data mart</i> deberán estar accesibles solo para usuarios encargados del monitoreo de los indicadores de acreditación universitaria. |            |   |

## 2.2.4. Pesos y funciones de utilidad

Los pesos de los indicadores se encuentran detallados en la tabla 1 del presente documento.

### 2.2.4.1. Funciones de utilidad para indicadores cualitativos

Para el caso práctico los indicadores cualitativos son los siguientes: admisión a estudios de posgrado, bienestar estudiantil y acción afirmativa. Tomando como referencia la normativa que modifica la escala de valoración de los indicadores cualitativos (Ver Anexo I), la Figura 9 ilustra la función de utilidad según la nueva resolución establecida para este tipo de indicadores.

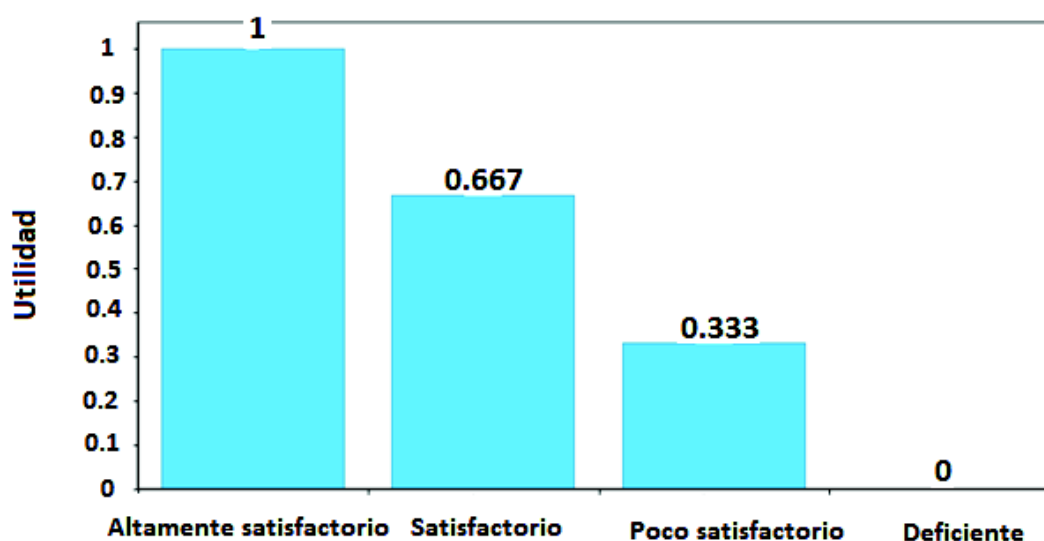


Figura 9. Función de utilidad de indicadores cualitativos<sup>4</sup>.

<sup>4</sup> Fuente: Elaboración propia

### 2.2.4.2. Tasa de retención de grado.

La tasa de retención de grado es un indicador de tipo cuantitativo. El porcentaje de retención mínimo esperado es del 80%. La Figura 10 muestra la función de utilidad de la tasa de retención de grado.

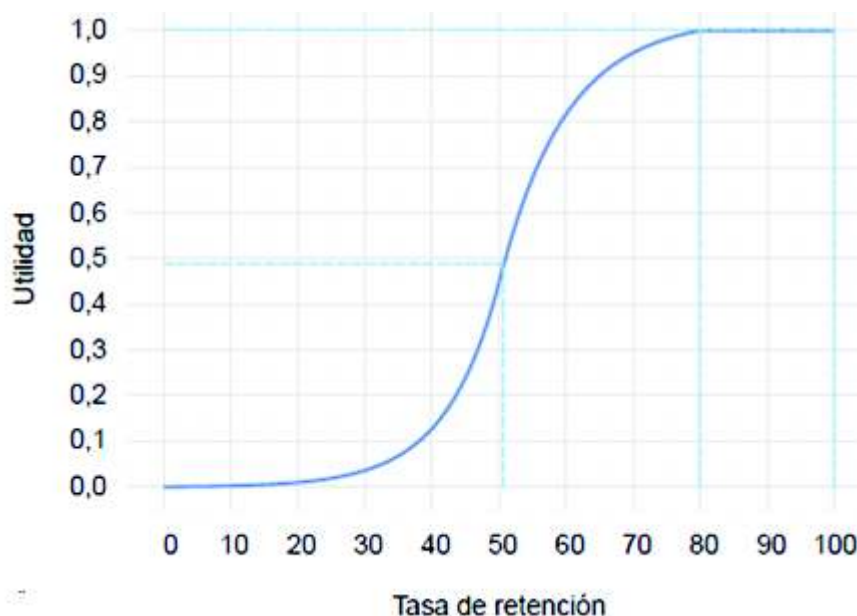


Figura 10. Función de utilidad del indicador tasa de retención de grado [29].

La tasa de retención se calcula en función de la siguiente fórmula [29]:

$$TRG = \frac{NEMA}{NTEA}$$

Donde:

TRG= Tasa de retención de grado.

NEMA = Número de estudiantes matriculados durante el período académico ordinario en el que se efectúa la evaluación de la institución, que fueron admitidos dos años antes.

NTEA = Número total de estudiantes que fueron admitidos en una carrera de la IES dos años antes.

### 2.2.4.3. Tasa de graduación de grado.

La tasa de graduación de grado es un indicador de tipo cuantitativo. La tasa de graduación mínima esperada es del 80%. La Figura 11 muestra la función de utilidad de la tasa de graduación de grado.

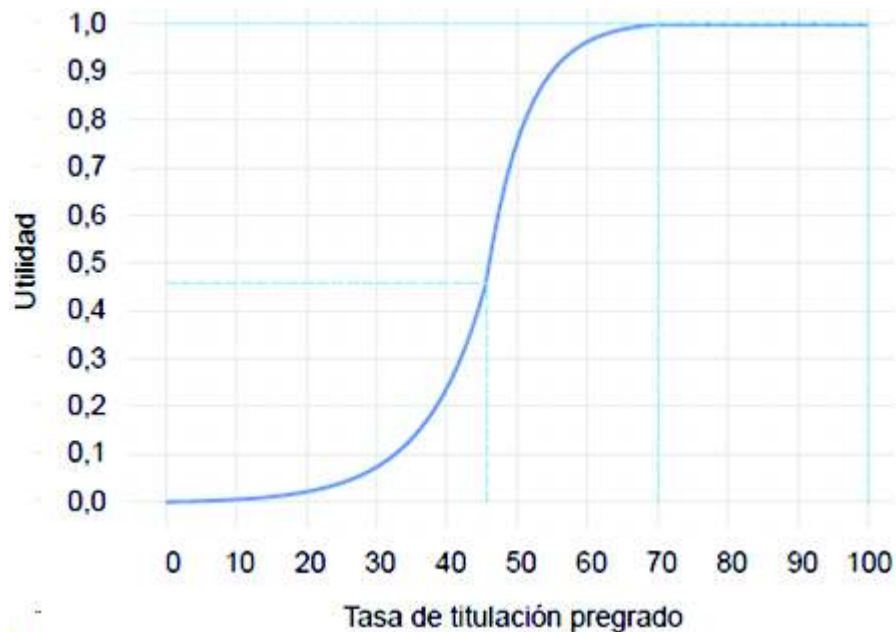


Figura 11. Función de utilidad del indicador tasa graduación de grado [29].

La tasa de titulación de grado se calcula en función de la siguiente fórmula [29]:

$$TGG = \frac{NEG}{NEC}$$

Donde:

TGG= Tasa de graduación de grado.

NEG = Número de estudiantes de grado que ingresaron en las cohortes definidas y se graduaron hasta el final del último período académico regular concluido antes de la evaluación.

NEC = Número de estudiantes de grado que ingresaron en las cohortes definidas.

#### 2.2.4.4. Tasa de graduación de posgrado.

La tasa de graduación de posgrado es un indicador de tipo cuantitativo. La tasa de graduación mínima esperada es del 80%. La Figura 12 muestra la función de utilidad de la tasa de graduación de posgrado.

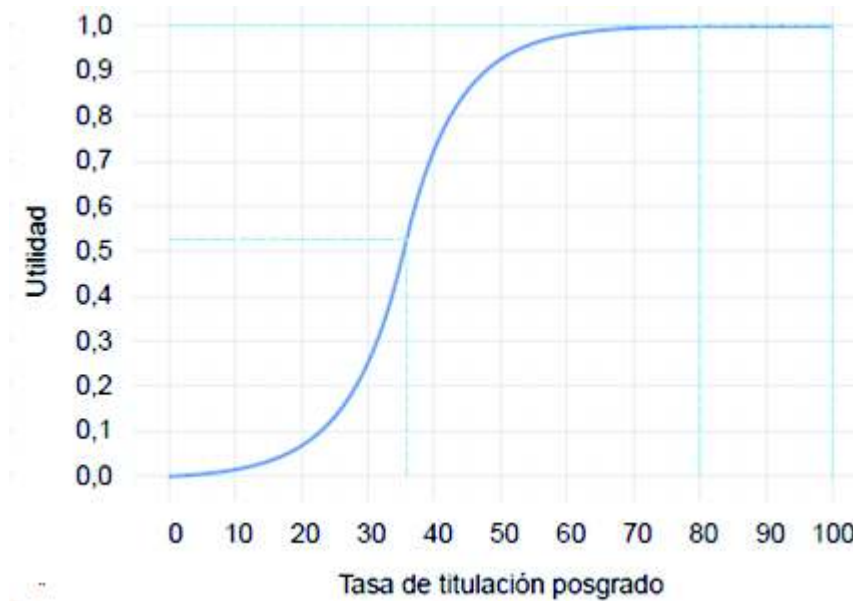


Figura 12. Función de utilidad del indicador tasa titulación posgrado [29].

La tasa de graduación de posgrado se calcula en función de la siguiente fórmula [29]:

$$TGP = \frac{NEGP}{NECP}$$

Donde:

TGG= Tasa de graduación de posgrado.

NEG = Número de estudiantes de posgrado que ingresaron en las cohortes definidas y se graduaron hasta el final del último período académico regular concluido antes de la evaluación.

NEC = Número de estudiantes de posgrado que ingresaron en las cohortes definidas.

### 2.2.5. Archivos fuente

Para realizar el cálculo de los indicadores cuantitativos la base de datos que proporcionará los datos necesarios será la del sistema académico transaccional SAEW que es propiedad de la Escuela Politécnica Nacional.

Para los indicadores cualitativos la institución no cuenta con una fuente que proporcione los datos necesarios para implementar la solución. Debido a este problema se procedió con la elaboración de un prototipo de base de datos en PostgreSQL, la cual servirá como fuente de indicadores cualitativos para alimentar el sistema. El script de la base de datos se detalla en el anexo III.

## **2.3. Modelo dimensional**

Una vez realizado el análisis de los requerimientos, se procede a identificar el proceso de negocio, las medidas y dimensiones que se orientan al análisis de la información en sus diferentes niveles.

### **2.3.1. Definición del proceso de negocio**

El proceso de negocio es: análisis de indicadores de acreditación universitaria, este proceso permitirá evaluar los diferentes indicadores según el periodo académico, criterio y subcriterio, por lo tanto, éste será el proceso que se va a desarrollar.

### **2.3.2. Establecer los niveles de granularidad**

Para establecer la granularidad, según lo recomendado por la metodología de Kimball, se debe tener como base una granularidad baja a partir del proceso de negocio. Esto quiere decir, que el modelo cuente con información lo más detallada posible, a tal punto que no se pueda desglosar. Partiendo de esta recomendación, la granularidad más baja es donde los usuarios analizarán la trazabilidad de los indicadores según el criterio, subcriterio y tiempo.

### **2.3.3. Elección de las dimensiones**

Debido a que la definición de la granularidad es clara, las dimensiones se pueden identificar fácilmente. Hay que recordar que el *data mart* es directamente proporcional a la calidad y profundidad de los atributos de cada dimensión.

A continuación en la Figura 13 se muestra el análisis dimensional del proceso de negocio identificado anteriormente.



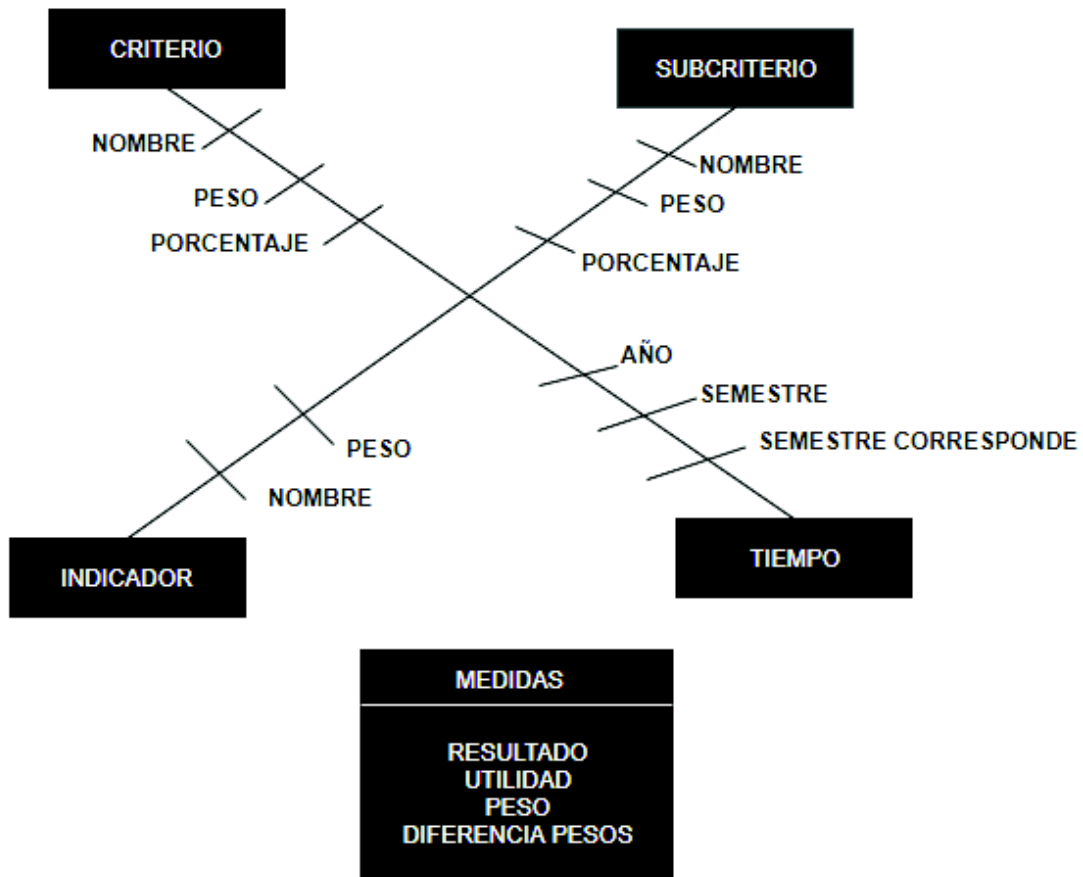


Figura 13. Análisis dimensional del proceso de negocio<sup>5</sup>.

### 2.3.3.1. Dimensiones encontradas

Según el análisis realizado en base a los requerimientos, las dimensiones son:

- Criterio\_dim.- La dimensión abarca el nombre, peso y porcentaje del criterio.
- Subcriterio\_dim.- La dimensión abarca el nombre, peso y porcentaje del subcriterio.
- Indicador\_dim.- La dimensión abarca el nombre y peso del indicador.
- Tiempo\_dim.- La dimensión abarca el periodo correspondiente al que pertenece el proceso de evaluación de la IES.

<sup>5</sup> Fuente: Elaboración propia

### 2.3.3.2. Dimensiones y sus atributos

Tabla 22. Dimensiones y sus atributos<sup>6</sup>.

| Dimensiones     | Atributos  |                   |                   |
|-----------------|------------|-------------------|-------------------|
|                 | Atributo 1 | Atributo 2        | Atributo 3        |
| Criterio_dim    | Porcentaje | Peso              | Nombre            |
| Subcriterio_dim | Porcentaje | Peso              | Nombre            |
| Indicador_dim   | Peso       | Nombre            |                   |
| Tiempo_dim      | Año        | Semestre carácter | Semestre numérico |

### 2.3.4. Tabla de hechos fact\_modelo\_acreditación

La tabla de hechos fact\_modelo\_acreditación representa el resultado obtenido, utilidad, peso y diferencia de pesos de los indicadores que se monitorearán para la acreditación universitaria.

Las medidas de la tabla de hechos son las siguientes:

- Resultado obtenido
- Utilidad
- Peso
- Diferencia de pesos

En la Figura 14, la tabla central “fact\_modelo\_acreditación” es la tabla de hechos de un esquema dimensional y está rodeada de las dimensiones criterio\_dim, subcriterio\_dim, indicador\_dim y tiempo\_dim.

La tabla de hechos almacena la clave primaria de cada dimensión y las medidas identificadas anteriormente.

---

<sup>6</sup> Fuente: Elaboración propia

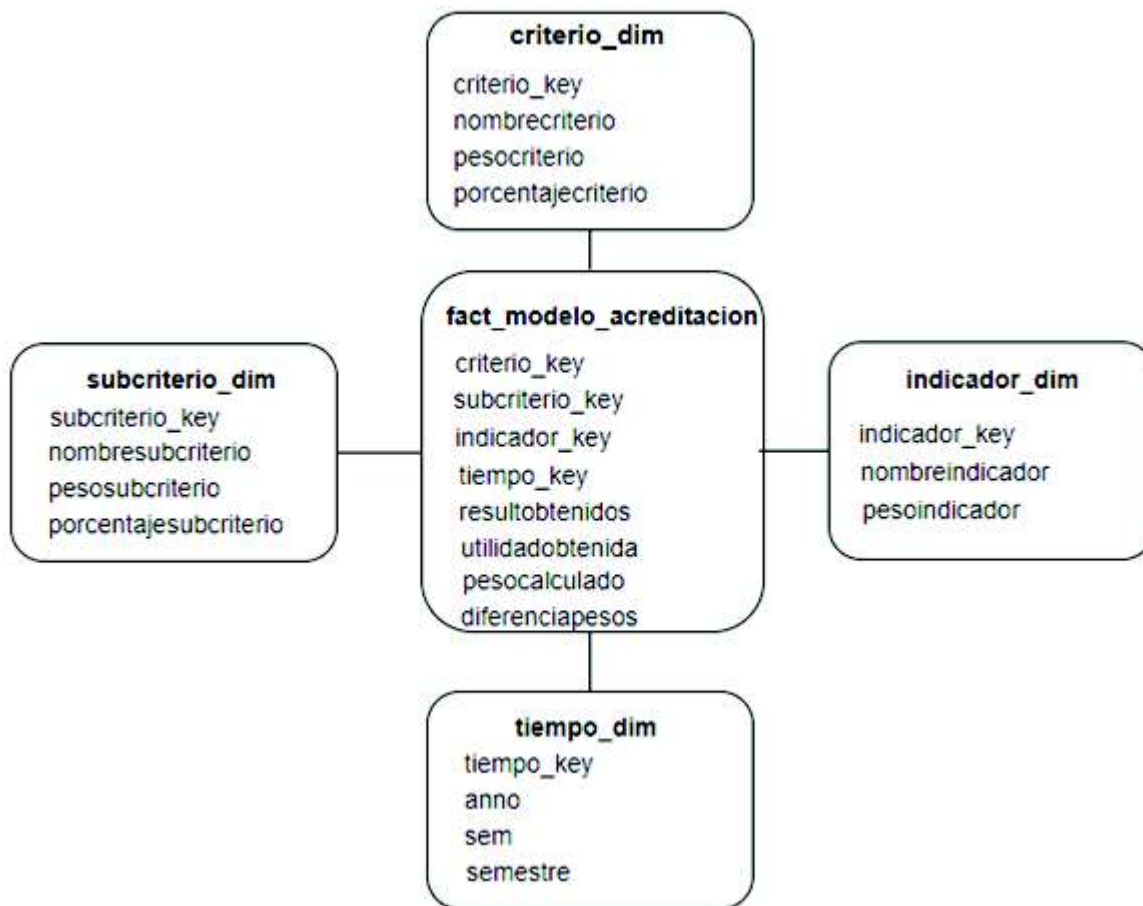


Figura 14. Tabla de hechos fact\_modelo\_acreditación<sup>7</sup>.

## 2.4. Definición de la arquitectura técnica

La Figura 15 muestra la arquitectura técnica de la solución en la cual se puede observar que hay dos fuentes de información, la primera se encuentra sobre un motor de base de datos SQL Server donde se encuentran los datos necesarios para el análisis de indicadores cuantitativos. La segunda fuente se encuentra sobre un motor de base de datos PostgreSQL que contiene los datos necesarios para el análisis de indicadores cualitativos.

Se ha considerado trabajar con un área temporal, donde se van a almacenar los datos necesarios de las fuentes de información. Los datos almacenados en el área temporal no alimentarán a ninguna aplicación de reportes, por lo tanto, una vez que los datos cumplan con su función serán eliminados.

<sup>7</sup> Fuente: Elaboración propia

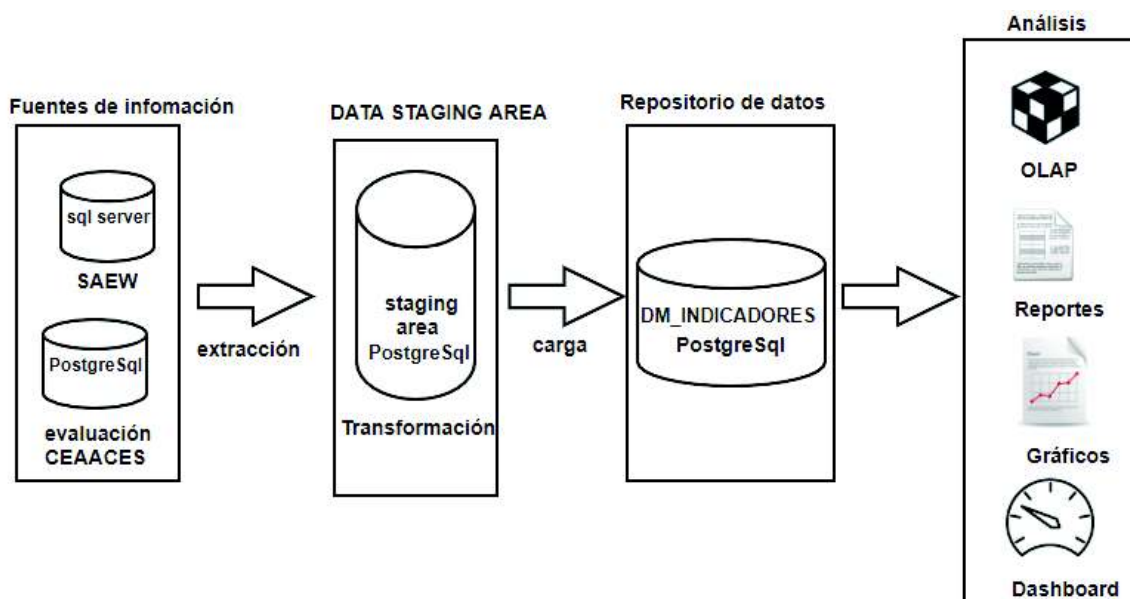


Figura 15. Arquitectura técnica de la solución<sup>8</sup>.

Una vez que se encuentren almacenados los datos en el *data mart*, estos podrán ser usados por herramientas de procesamiento analítico en línea OLAP para realizar análisis multidimensional. Para poder llevar a cabo el procesamiento de datos se utiliza el mecanismo ROLAP que permite simular un modelo multidimensional a partir de una base de datos relacional en PostgreSQL.

Para mejorar la visualización de la información del negocio se utilizan otras técnicas que facilitan el análisis de los datos del *data mart* como son: reportes de datos y cuadros de mando (*dashboards*). El uso de cuadros de mando permite visualizar el estado actual de los indicadores de acreditación universitaria en tiempo real.

#### 2.4.1. Definición de las herramientas para la plataforma de BI

Posterior a la definición de la arquitectura técnica, se procede a seleccionar e instalar los productos que se ajusten debidamente con cada fase del plan de la arquitectura.

El Consejo de Educación Superior en su reglamento de sanciones, artículo 11, numeral 9, determina como infracción leve, “incumplir la obligación de incorporar en la IES el uso de programas informáticos de software libre” [30]. Es por ello que se utilizarán herramientas de software libre en el desarrollo de BI (ver Anexo IV).

<sup>8</sup> Fuente: Elaboración propia

Para el desarrollo e implementación del proyecto se usarán las herramientas de la plataforma Pentaho, además de PostgreSQL como motor de base de datos.

## 2.5. Diseño físico

La Figura 15 muestra el diseño físico de la base de datos implementada para el *data mart* que permitirá el monitoreo de indicadores de acreditación universitaria.

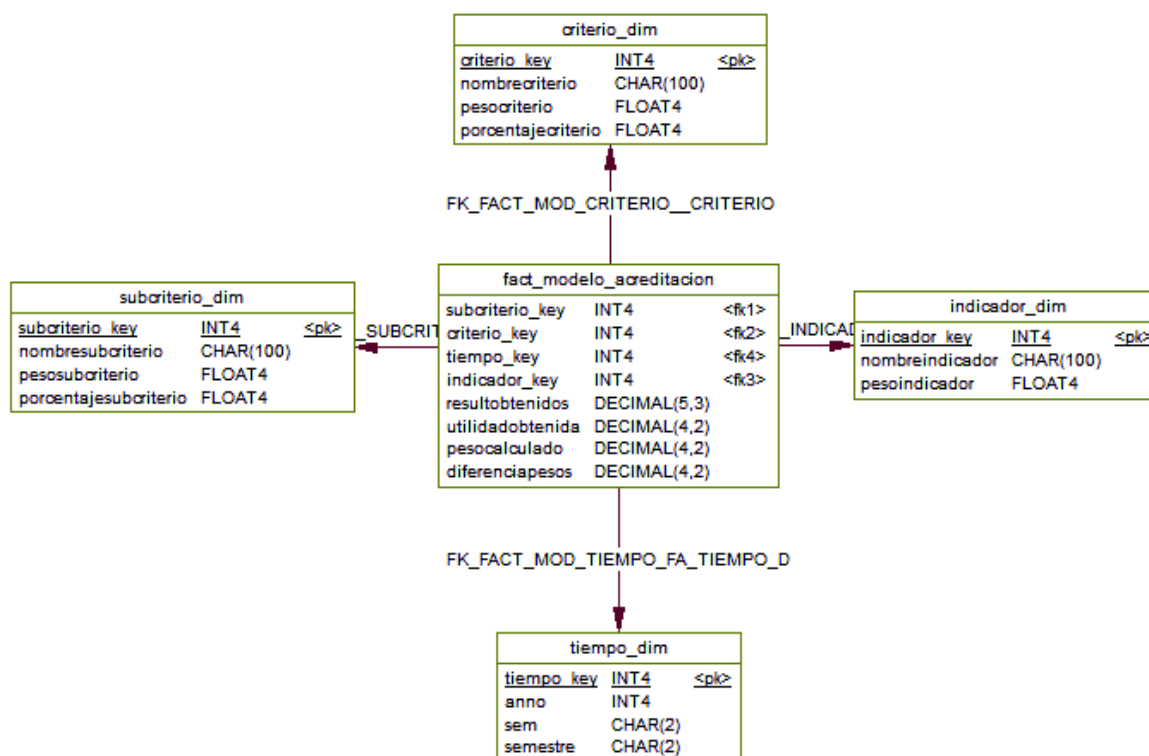


Figura 16. Diseño físico<sup>9</sup>.

### 2.5.1. Dimensión criterio\_dim

La dimensión criterio\_dim está conformada por:

Tabla 23. Dimensión criterio\_dim<sup>10</sup>.

| Campo          | Tipo | Longitud | Descripción         |
|----------------|------|----------|---------------------|
| criterio_key   | int  |          | Clave primaria      |
| nombrecriterio | char | 100      | Nombre del criterio |

<sup>9</sup> Fuente: Elaboración propia

<sup>10</sup> Fuente: Elaboración propia

|                    |       |  |                         |
|--------------------|-------|--|-------------------------|
| pesocriterio       | float |  | Peso del criterio       |
| porcentajecriterio | float |  | Porcentaje del criterio |

### 2.5.2. Dimensión subcriterio\_dim

La dimensión subcriterio\_dim está conformada por:

Tabla 24. Dimensión subcriterio\_dim<sup>11</sup>.

| Campo                 | Tipo  | Longitud | Descripción                |
|-----------------------|-------|----------|----------------------------|
| subcriterio_key       | int   |          | Clave primaria             |
| nombresubcriterio     | char  | 100      | Nombre del subcriterio     |
| pesosubcriterio       | float |          | Peso del subcriterio       |
| porcentajesubcriterio | float |          | Porcentaje del subcriterio |

### 2.5.3. Dimensión indicador\_dim

La dimensión indicador\_dim está conformada por:

Tabla 25. Dimensión indicador\_dim<sup>12</sup>.

| Campo           | Tipo  | Longitud | Descripción          |
|-----------------|-------|----------|----------------------|
| indicador_key   | int   |          | Clave primaria       |
| nombreindicador | char  | 100      | Nombre del indicador |
| pesoindicador   | float |          | Peso del indicador   |

### 2.5.4. Dimensión tiempo\_dim

La dimensión tiempo\_dim está conformada por:

Tabla 26. Dimensión tiempo\_dim<sup>13</sup>.

| Campo      | Tipo | Longitud | Descripción    |
|------------|------|----------|----------------|
| tiempo_key | int  |          | Clave primaria |

<sup>11</sup> Fuente: Elaboración propia

<sup>12</sup> Fuente: Elaboración propia

<sup>13</sup> Fuente: Elaboración propia

|          |      |   |   |
|----------|------|---|---|
| anno     | int  |   | Año del periodo de evaluación             |
| sem      | char | 2 | Semestre del periodo de evaluación        |
| semestre | char | 2 | Letra correspondiente al semestre (A o B) |

### 2.5.5. Tabla de hechos modelo acreditación

Tabla 27. Tabla de hechos modelo acreditación<sup>14</sup>.

| Campo            | Tipo    | Longitud | Descripción                              |
|------------------|---------|----------|--|
| criterio_key     | int     |          | Clave primaria dimensión criterio_dim    |
| subcriterio_key  | int     |          | Clave primaria dimensión subcriterio_dim |
| indicador_key    | Int     |          | Clave primaria dimensión indicador_dim   |
| tiempo_key       | int     |          | Clave primaria dimensión tiempo_dim      |
| resultobtenidos  | Decimal | (5,3)    | Medida resultado obtenido del indicador  |
| utilidadobtenida | Decimal | (4,2)    | Medida utilidad obtenida                 |
| pesocalculado    | Decimal | (4,2)    | Medida peso calculado                    |
| diferenciapesos  | Decimal | (4,2)    | Valor de la diferencia de pesos          |

## 2.6. Construcción del sistema ETL

Para realizar el proceso ETL se utiliza la herramienta Pentaho Data Integration (PDI) de la *suite* de Petaho, también conocida como Kettle.

Para el correcto desarrollo de los procesos ETL se debe realizar una lista de las principales actividades que son: extracción, limpieza, carga y entrega; a continuación se detallan las actividades que se realizaron en cada actividad.

<sup>14</sup> Fuente: Elaboración propia

## 2.7. Procesos ETL

### 2.7.1. Extracción de los datos fuentes

La extracción es la selección de datos orígenes de los sistemas. Para la extracción se tomó en cuenta información que será obtenida a partir de las fuentes que se detallan en la Tabla 28.

Tabla 28. Sistemas fuente de la solución<sup>15</sup>.

| Fuente  | Tipo       | Conexión |
|---|------------|----------|
| Sistema de gestión académica SAEW.                                      | SQL Server | JNDI     |
| Base de datos propuesta para la evaluación de indicadores cualitativos. | PostgreSQL | JNDI     |

#### 2.7.1.1. Descripción de las fuentes

- Sistema de gestión académica SAEW: contiene la información necesaria para realizar el análisis de los indicadores cuantitativos.
- Base de datos propuesta para evaluación de indicadores cualitativos: contiene información para el análisis de los indicadores cualitativos.

### 2.7.2. Procesos ETL para la carga de datos a las dimensiones

Este proceso gestiona la carga ordenada de los datos extraídos desde las fuentes hacia las tablas de dimensiones que se utilizarán para la creación del cubo de información. A continuación se detallan los procesos creados para la carga de las dimensiones.

#### 2.7.2.1. Carga de la dimensión criterio\_dim

El proceso de carga de la dimensión criterio se ejecuta una sola vez. La Figura 17 muestra la tabla origen con la respectiva información para la carga de la dimensión.

---

<sup>15</sup> Fuente: Elaboración propia



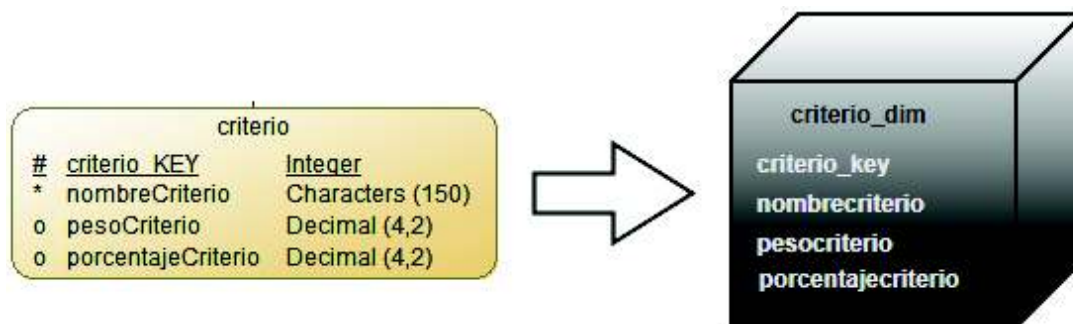


Figura 17. Tabla de origen para la carga de la dimensión criterio<sup>16</sup>.



Figura 18. Estructura de la carga de información de la dimensión criterio<sup>17</sup>.

La Figura 18 muestra la estructura de la carga de información de la dimensión criterio. “Entrada criterio” lee información de la tabla criterio que se encuentra en la base de datos propuesta para la evaluación de indicadores cualitativos. “Salida dim\_Criterio” carga los datos en la dimensión criterio del *data mart*. A continuación se muestra la consulta realizada para extraer la información de la fuente:

```
SELECT criterio_key, nombrecriterio, pesocriterio, porcentajecriterio FROM public.criterio;
```

### 2.7.2.2. Carga de la dimensión subcriterio\_dim

El proceso de carga de la dimensión subcriterio se ejecuta una sola vez. La Figura 19 muestra la tabla origen con la respectiva información para la carga de la dimensión.

<sup>16</sup> Fuente: Elaboración propia

<sup>17</sup> Fuente: Elaboración propia

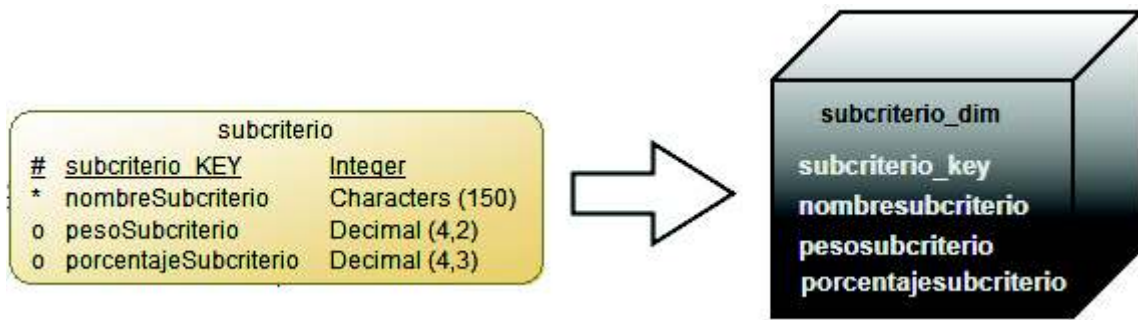


Figura 19. Tabla de origen para la carga de la dimensión criterio<sup>18</sup>.

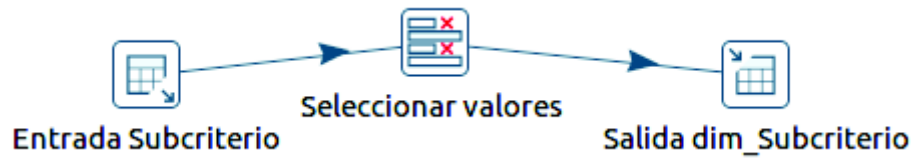


Figura 20. Estructura de la carga de información de la dimensión subcriterio<sup>19</sup>.

La Figura 20 muestra la estructura de la carga de información de la dimensión subcriterio. El primer paso “Entrada Subcriterio” lee la información de la tabla subcriterio que se encuentra en la base de datos propuesta para la evaluación de indicadores cualitativos. El segundo paso “Seleccionar valores” toma de los datos fuente únicamente los necesarios para cargarlos en la dimensión. El último paso “Salida dim\_Subcriterio” carga los datos en la dimensión criterio\_dim del *data mart*. A continuación se muestra la consulta realizada para extraer la información de la fuente:

```

SELECT  subcriterio_key,  criterio_key,  nombresubcriterio,  pesosubcriterio,
porcentajesubcriterio FROM public.subcriterio;
  
```

### 2.7.2.3. Carga de la dimensión indicador\_dim

El proceso de carga de la dimensión indicador se ejecuta una sola vez. La Figura 21 muestra la tabla origen con la respectiva información para la carga de la dimensión.

<sup>18</sup> Fuente: Elaboración propia

<sup>19</sup> Fuente: Elaboración propia

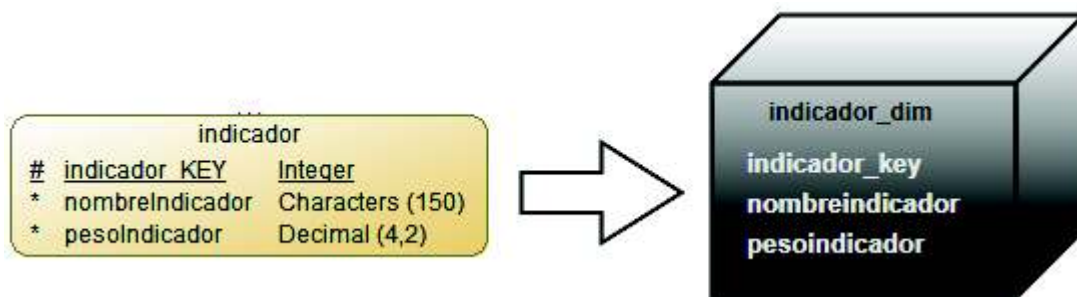


Figura 21. Tabla de origen para la carga de la dimensión indicador<sup>20</sup>.

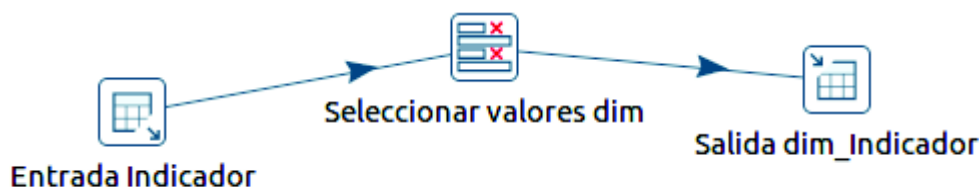


Figura 22. Estructura de la carga de información de la dimensión indicador<sup>21</sup>.

La Figura 22 muestra la estructura de la carga de información de la dimensión indicador. El primer paso “Entrada Indicador” lee la información de la tabla indicador que se encuentra en la base de datos propuesta para la evaluación de indicadores cualitativos. El segundo paso “Seleccionar valores dim” toma de los datos fuente únicamente los necesarios para cargarlos en la dimensión. El último paso “Salida dim\_Indicador” carga los datos en la dimensión indicador\_dim del *data mart*. A continuación se muestra la consulta realizada para extraer la información de la fuente:

```
SELECT indicador_key, subcriterio_key, nombreindicador, pesoindicador FROM public.indicador;
```

#### 2.7.2.4. Carga de la dimensión tiempo\_dim

La Figura 23 muestra la tabla origen con la respectiva información para la carga de la dimensión.

<sup>20</sup> Fuente: Elaboración propia

<sup>21</sup> Fuente: Elaboración propia

| SAEPERIODON       |               |
|-------------------|---------------|
| Nombre de columna | Tipo de datos |
| NIVELA            | numeric(2, 0) |
| ANOMAT            | numeric(4, 0) |
| SEMMAT            | numeric(2, 0) |
| MESINI            | numeric(2, 0) |
| ANOINI            | numeric(4, 0) |
| MESFIN            | numeric(2, 0) |
| ANOFIN            | numeric(4, 0) |
| FECHAE1           | datetime      |
| FECHAE2           | datetime      |
| FECHAE3           | datetime      |
| FECHAE4           | datetime      |
| FECMAT            | datetime      |
| FECLMAT           | datetime      |

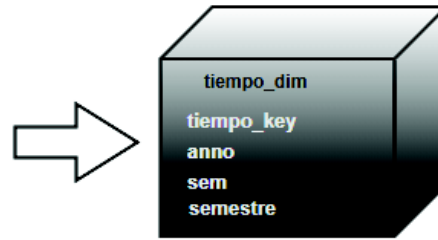


Figura 23. Tabla de origen para la carga de la dimensión tiempo<sup>22</sup>.

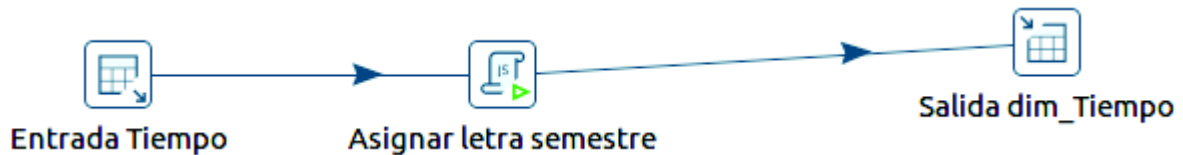


Figura 24. Estructura de la carga de información de la dimensión tiempo<sup>23</sup>.

La Figura 24 muestra la estructura de la carga de información de la dimensión tiempo. El primer paso “Entrada Tiempo” lee la información de la tabla SAEPERIODON que se encuentra en la base de datos del sistema de gestión académica SAEW. El segundo paso “Asignar letra semestre” es un script que crea la variable letra a la cual se le asigna A o B dependiendo del semestre que viene dado por la variable semmat de la tabla fuente. El último paso “Salida dim\_tiempo” carga los datos en la dimensión tiempo\_dim del *data mart*. A continuación se muestra la consulta realizada para extraer la información de la fuente:

```
SELECT anomat,semmat FROM saeperiodon WHERE nivela=3 and anomat>2009 and (semmat in ('1','2') or (anommat=2012 and semmat=4));
```

### 2.7.2.5. Proceso Inicio\_ETL\_DATA\_MART\_EPN

Las actividades de ETL se coordinan como se ilustra en la Figura 25.

<sup>22</sup> Fuente: Elaboración propia

<sup>23</sup> Fuente: Elaboración propia

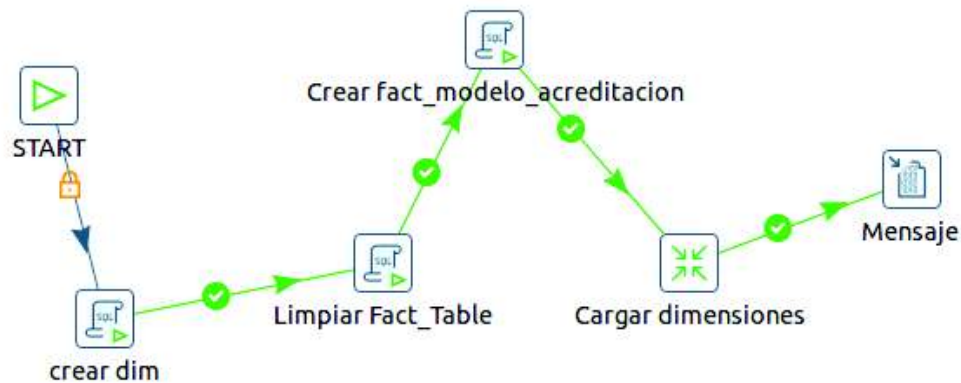


Figura 25. Proceso Inicio\_ETL\_DATA\_MART\_EPN<sup>24</sup>.

El proceso “Inicio\_ETL\_data\_mart\_EPN” se ejecuta una sola vez y comienza en el nodo *start* para luego continuar con la ejecución del *script* SQL “crear dim” para crear las tablas de dimensiones del *data mart*, después continúa la ejecución del *script* SQL “Limpiar Fact\_Table” que elimina cualquier tabla que tenga el mismo nombre de la tabla de hechos, luego se ejecuta el *script* SQL “Crear fact\_modelo\_acreditación” para crear la tabla de hechos. El siguiente paso “Cargar dimensiones” cuyos procesos fueron detallados anteriormente y su objetivo principal es cargar los datos de las fuentes a las dimensiones. Finalmente si el trabajo fue realizado con éxito, se escribe un mensaje en el sistema de registro de PDI comunicando al usuario que la ejecución del proceso se realizó satisfactoriamente.

### 2.7.3. Procesos ETL para carga de datos a la tabla de hechos

Este proceso permite realizar la carga ordenada de los datos extraídos desde las fuentes hacia la tabla de hechos que se utiliza para la creación del cubo de información. A continuación se detallan los procesos creados para la carga de la tabla de hechos.

<sup>24</sup> Fuente: Elaboración propia

### 2.7.3.1. Carga de tabla temporal de variables

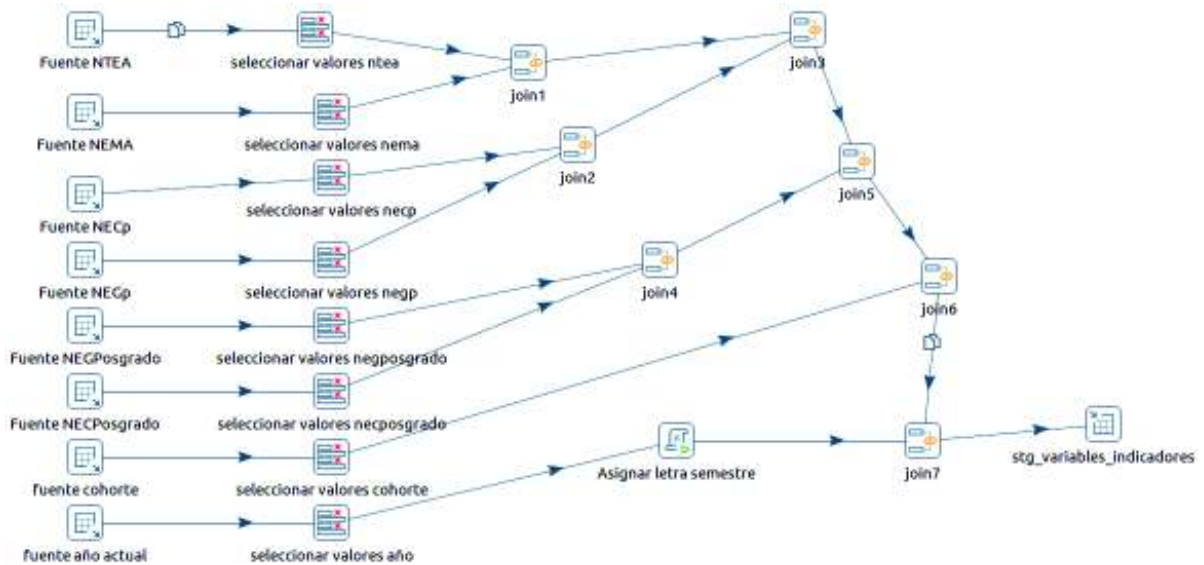


Figura 26. Proceso de carga a tabla temporal de variables<sup>25</sup>.

La Figura 26 muestra el proceso para extraer la información referente a las variables que se usan para el cálculo de los indicadores cuantitativos cuya fuente es la base de datos del sistema de gestión académica SAEW. Luego se selecciona la información que será útil para el respectivo cálculo de los indicadores y se carga en la tabla temporal.

Las consultas SQL realizadas para obtener las variables de los indicadores cuantitativos son las siguientes:

- La Figura 27 muestra la consulta SQL para obtener la variable número total de estudiantes que fueron admitidos en la carrera dos años antes del período de evaluación (NTEA).

<sup>25</sup> Fuente: Elaboración propia

```

SELECT count(codest) as NTEA FROM saematri nn join saecarrera ca on
(nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)+convert(varchar(2),semmat)
=(SELECT top 1 per FROM (SELECT top 5 convert(varchar(4),anomat)
+ convert(varchar(2),semmat) as per FROM saeperiodon where nivela=3
and (semmat in (1,2) or (anomat=2012 and semmat=4)) order by convert
(varchar(4),anomat)+ convert(varchar(2),semmat)desc ) as a order by per asc)
and ff.nivela=3 and nn.codcar not in ('128','129','182','156') and not exists
(select * from saematri where codcar =nn.codcar and codest=nn.codest and convert
(varchar(4),anomat)+convert(varchar(2),semmat)<convert(varchar(4),nn.anomat)
+convert(varchar(2),nn.semmat) and anulado is null) and not exists
(select * from saemater where codcar =nn.codcar and codest=nn.codest and isnull
(estatum,'')='E') and not exists (SELECT * FROM saematerr where codcar =nn.codcar
and codest=nn.codest)

```

Figura 27. Consulta SQL para obtener la variable NTEA.

- La Figura 28 muestra la consulta SQL para obtener la variable número de estudiantes matriculados durante el período académico ordinario en el que se efectúa la evaluación de la institución, que fueron admitidos dos años antes (NEMA).
- La Figura 29 muestra la consulta SQL para obtener la variable número de estudiantes de grado que ingresaron en la(s) cohorte(s) definidas (NECp).
- La Figura 30 muestra la consulta SQL para obtener la variable número de estudiantes de grado que ingresaron en la (s) cohorte(s) definidas y se graduaron hasta el final del último período académico regular concluido antes del inicio del proceso de evaluación (NEGp).

```

SELECT count(codest) as NEMA FROM saematri nn join saecarrera ca on
(nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)
+convert(varchar(2),semmat)=(SELECT top 1 per FROM ( SELECT top 5 convert
(varchar(4),anomat)+ convert(varchar(2),semmat) as per FROM saeperiodon
WHERE nivela=3 and (semmat in (1,2) or (anomat=2012 and semmat=4))
order by convert(varchar(4),anomat)+ convert(varchar(2),semmat)desc )
as a order by per asc) and ff.nivela=3 and nn.codcar not in
('128','129','182','156') and not exists(SELECT * FROM saematri WHERE
codcar =nn.codcar and codest=nn.codest and convert(varchar(4),anomat)
+convert(varchar(2),semmat)<convert(varchar(4),nn.anomat)
+convert(varchar(2),nn.semmat) and anulado is null) and not exists
(SELECT * FROM saemater WHERE codcar =nn.codcar and codest=nn.codest
and isnull(estatum,'')='E') and not exists (SELECT * FROM saematerr
WHERE codcar =nn.codcar and codest=nn.codest) and exists (SELECT * FROM saematri
WHERE codcar =nn.codcar and codest=nn.codest and anomat=(SELECT top 1 anomat
FROM saeperiodon where nivela=3 and anomat>2009 and (semmat in ('1','2') or
(anomat=2012 and semmat=4)) order by anomat desc) and semmat=(SELECT top 1 semmat
FROM saeperiodon WHERE nivela=3 and anomat>2009 and (semmat in ('1','2') or
(anomat=2012 and semmat=4)) order by anomat desc) and anulado is null)

```

Figura 28. Consulta SQL para obtener la variable NEMA.

```

SELECT count(codest) as NECp FROM saematri nn join saecarrera ca on
(nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)+convert(varchar(2),semmat)
=(SELECT top 1 per FROM ( SELECT top 14 convert(varchar(4),anomat)+
convert(varchar(2),semmat) as per FROM saeperiodon WHERE nivela=3 and
(semmat in (1,2) or (anomat=2012 and semmat=4)) order by convert(varchar(4),anomat)
+ convert(varchar(2),semmat)desc ) as a order by per asc) and ff.nivela=3
and nn.codcar not in ('128','129','182','156') and not exists(SELECT * FROM saematri
WHERE codcar =nn.codcar and codest=nn.codest and convert(varchar(4),anomat)
+convert(varchar(2),semmat)<convert(varchar(4),nn.anomat)+convert
(varchar(2),nn.semmat) and anulado is null) and not exists (SELECT * FROM saemater
WHERE codcar =nn.codcar and codest=nn.codest and isnull(estatum,'')='E')
and not exists (SELECT * FROM saematerr where codcar =nn.codcar and codest=nn.codest)

```

Figura 29. Consulta SQL para obtener la variable NECp.

```

SELECT count(codest) as NEGp FROM saematri nn join saecarrera ca
on (nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)+convert(varchar(2),semmat)=
(SELECT top 1 per FROM (SELECT top 14 convert(varchar(4),anomat)
+ convert(varchar(2),semmat) as per FROM saeperiodon WHERE nivela=3
and (semmat in (1,2) or (anomat=2012 and semmat=4)) order by
convert(varchar(4),anomat)+ convert(varchar(2),semmat)desc ) as a order by per asc)
and ff.nivela=3 and nn.codcar not in ('128','129','182','156')
and not exists(SELECT * FROM saematri WHERE codcar =nn.codcar and codest=nn.codest
and convert(varchar(4),anomat)+convert(varchar(2),semmat)<convert
(varchar(4),nn.anomat)+convert(varchar(2),nn.semmat) and anulado is null)
and not exists (SELECT * FROM saemater WHERE codcar =nn.codcar and codest=nn.codest
and isnull(estatum,'')='E') and not exists (SELECT * FROM saematerr
WHERE codcar =nn.codcar and codest=nn.codest) and exists (SELECT * FROM saeestcarr
WHERE codcar = nn.codcar and codest=nn.codest and isnull (estatu,'') = 'GRADUADO'
and convert(varchar(4), anomatg) + convert(varchar(2), semmatg) <= (SELECT top 1
convert( varchar(4), anomat) + convert(varchar(2), semmat) as per FROM saeperiodon
WHERE nivela=3 and anomat > 2009 and (semmat in ('1','2') or (anomat=2012
and semmat=4)) order by anomat desc))

```

Figura 30. Consulta SQL para obtener la variable NEGp.

- La Figura 31 muestra la consulta SQL para obtener la variable número de estudiantes de posgrado que ingresaron en la (s) cohorte(s) definidas y se graduaron hasta el final del último período académico regular concluido antes del inicio del proceso de evaluación (NEGP).
- La Figura 32 muestra la consulta SQL para obtener la variable número de estudiantes de posgrado que ingresaron en la(s) cohorte(s) definidas (NECP).
- La Figura 33 muestra la consulta SQL para obtener la cohorte.



```

SELECT count(codest) as NEGPosgrado FROM saematri nn join saecarrera ca
on (nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)+convert(varchar(2),semmat)=
(SELECT top 1 per FROM ( SELECT top 8 convert(varchar(4),anomat)+
convert(varchar(2),semmat) as per FROM saeperiodon WHERE nivela=3
and (semmat in (1,2) or (anomat=2012 and semmat=4)) order by convert
(varchar(4),anomat)+ convert(varchar(2),semmat)desc ) as a order by per asc)
and ff.nivela=4 and nn.codcar not in ('128','129','182','156') and not exists
(SELECT * FROM saematri WHERE codcar =nn.codcar and codest=nn.codest
and convert(varchar(4),anomat)+convert(varchar(2),semmat)<convert
(varchar(4),nn.anomat)+convert(varchar(2),nn.semmat) and anulado is null)
and not exists (SELECT * FROM saemater WHERE codcar =nn.codcar
and codest=nn.codest and isnull(estatum,'')='E') and not exists
(SELECT * FROM saematerr WHERE codcar =nn.codcar and codest=nn.codest)
and exists (SELECT * FROM saeestcarr WHERE codcar =nn.codcar and codest=nn.codest
and isnull (estatuie,'') = 'GRADUADO' and convert(varchar(4), anomatg) +
convert(varchar(2), semmatg) <= ( SELECT top 1 convert(varchar(4), anomat) +
convert(varchar(2), semmat) as per FROM saeperiodon WHERE nivela=3
and anomat > 2009 and (semmat in ('1','2') or (anomat=2012 and semmat=4))
order by anomat desc))

```

Figura 31. Consulta SQL para obtener la variable NEGP.

```

SELECT count(codest) as NECPosgrado FROM saematri nn join saecarrera ca
on (nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)+convert(varchar(2),semmat)=
(SELECT top 1 per FROM ( SELECT top 8 convert(varchar(4),anomat)+
convert(varchar(2),semmat) as per FROM saeperiodon WHERE nivela=3
and (semmat in (1,2) or (anomat=2012 and semmat=4)) order by
convert(varchar(4),anomat)+ convert(varchar(2),semmat)desc ) as a order by per asc)
and ff.nivela=4 and nn.codcar not in ('128','129','182','156')
and not exists(SELECT * FROM saematri WHERE codcar =nn.codcar
and codest=nn.codest and convert(varchar(4),anomat)+convert(varchar(2),semmat)<
convert(varchar(4),nn.anomat)+convert(varchar(2),nn.semmat) and anulado is null)
and not exists (SELECT * FROM saemater WHERE codcar =nn.codcar and codest=nn.codest
and isnull(estatum,'')='E') and not exists (SELECT * FROM saematerr
WHERE codcar =nn.codcar and codest=nn.codest)

```

Figura 32. Consulta SQL para obtener la variable NECP.

```

SELECT top 1 anomat,semmat FROM saematri nn join saecarrera ca
on (nn.codcar=ca.codcar) join dbo.SAENIVAC ff on (ff.nivela=ca.nivela)
WHERE anulado is null and convert(varchar(4),anomat)+convert(varchar(2),semmat)=
(SELECT top 1 per FROM ( SELECT top 8 convert(varchar(4),anomat)+
convert(varchar(2),semmat) as per FROM saeperiodon WHERE nivela=3
and (semmat in (1,2) or (anomat=2012 and semmat=4)) order by
convert(varchar(4),anomat)+ convert(varchar(2),semmat)desc ) as a order by per asc)
and ff.nivela=4 and nn.codcar not in ('128','129','182','156')
and not exists(SELECT * FROM saematri WHERE codcar =nn.codcar
and codest=nn.codest and convert(varchar(4),anomat)+convert(varchar(2),semmat)<
convert(varchar(4),nn.anomat)+convert(varchar(2),nn.semmat) and anulado is null)
and not exists (SELECT * FROM saemater WHERE codcar =nn.codcar and codest=nn.codest
and isnull(estatum,'')='E') and not exists (SELECT * FROM saematerr
WHERE codcar =nn.codcar and codest=nn.codest)

```

Figura 33. Consulta SQL para obtener la cohorte.

- La Figura 34 muestra la consulta SQL para obtener el período académico en el cual se está realizando el proceso de acreditación institucional.

```
SELECT top 1 anomat,semmat FROM saeperiodon WHERE nivela=3
and anomat>2009 and (semmat in ('1','2') or (anomat=2012 and semmat=4))
order by anomat desc
```

Figura 34. Consulta SQL para obtener el período académico.

### 2.7.3.2. Carga de la tabla temporal de cálculos

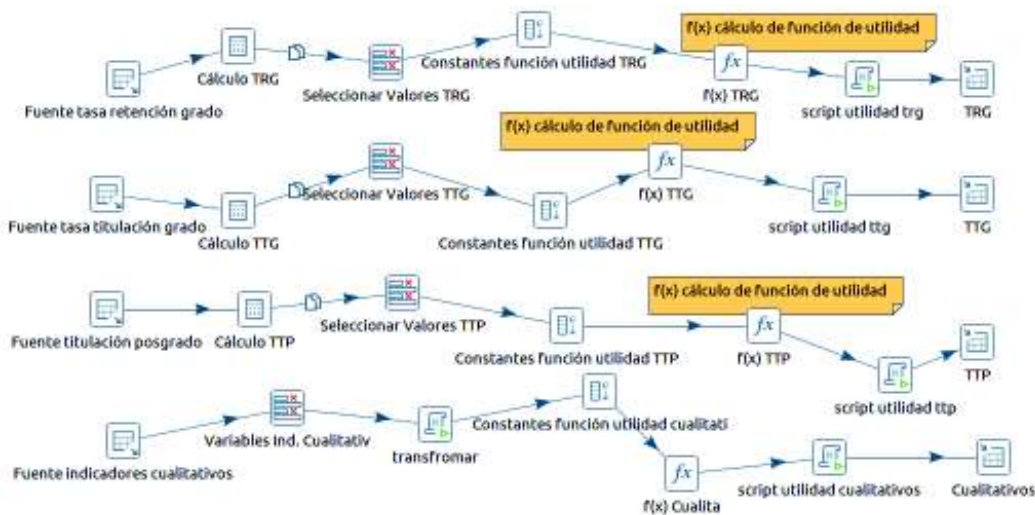


Figura 35. Proceso de carga a tabla temporal de cálculos<sup>26</sup>.

La Figura 35 ilustra el proceso de carga de la tabla temporal de cálculos que extrae la información de la tabla temporal de variables para realizar el cálculo de los indicadores cuantitativos y de sus respectivas funciones de utilidad. Una vez que se realizan las operaciones los resultados se cargan en la tabla temporal de cálculos.

Los indicadores cuantitativos son calculados a partir de una fórmula matemática dada para cada caso del modelo, la fórmula es ingresada al componente *calculator* de kettle que se encargará de realizar los respectivos cálculos.

Los indicadores cualitativos no tienen fórmulas matemáticas sino son evaluados en función a una escala de valores establecidos de acuerdo al modelo. Los valores asignados a estos indicadores dependen de las evidencias documentales y del cumplimiento mínimo de las acciones establecidas según sea el caso. Una vez realizada

<sup>26</sup> Fuente: Elaboración propia

la evaluación, los valores son almacenados en la base de datos propuesta para la evaluación de este tipo de indicadores, por lo que pueden ser extraídos directamente de esta base para cargarlos en la tabla temporal.

Para realizar el cálculo de las utilidades se tiene una función que marca los valores estándar para cada indicador. El CEAACES en su modelo de evaluación institucional proporciona las gráficas de las funciones de utilidad pero no proporciona la función que le correspondería a cada gráfica.

Por lo tanto, usando el modelo matemático de regresión polinómica se encontraron las funciones que mejor se ajusten a los datos dados, es decir, una función cuya representación gráfica sea una curva que se acerque de la mejor manera a los datos dados en las gráficas de las funciones de utilidad que el CEAACES expresa en su modelo de evaluación.

El análisis de las regresiones polinómicas fue desarrollado en hojas de cálculo del LibreOffice. En el Anexo V se detalla el análisis de las regresiones para la obtención de las funciones y el coeficiente de determinación.

El coeficiente de determinación establece la calidad del modelo y oscila entre 0 y 1, cuanto más se acerque a 1 mayor será la precisión, es decir el error tiende a disminuir. Debido a que las funciones de utilidad no son lineales y los valores están en una zona donde se puede cometer errores de apreciación, existe un porcentaje de error que es pequeño y que pudiera disminuir si el CEAACES facilitara directamente a las IES las funciones para el cálculo de las utilidades.

Las funciones obtenidas son ingresadas al componente *step formula* de kettle para calcular las utilidades de los respectivos indicadores. En las Figuras 36, 37 y 38 se detallan los componentes *step* con las funciones de utilidad que se obtuvieron como resultado de aplicar las regresiones polinómicas. En el campo "Formula" se ingresan las funciones obtenidas al realizar las regresiones polinómicas.



Figura 36. Función de utilidad tasa retención de grado<sup>27</sup>.



Figura 37. Función de utilidad tasa titulación de grado<sup>28</sup>.



Figura 38. Función de utilidad tasa titulación de posgrado<sup>29</sup>.

<sup>27</sup> Fuente: Elaboración propia

<sup>28</sup> Fuente: Elaboración propia

<sup>29</sup> Fuente: Elaboración propia

### 2.7.3.3. Carga de la tabla de hechos fact\_modelo\_acreditación

El proceso de carga de la tabla de hechos permite que la información de la tabla temporal cálculos que se cruza con las dimensiones de este modelo se cargue en la tabla “fact\_modelo\_acreditación”. El detalle de este proceso se ilustra en la Figura 39.

El *script* “Java Script calc pesos” contiene fórmulas matemáticas que permiten realizar el respectivo cálculo de los pesos de los indicadores.

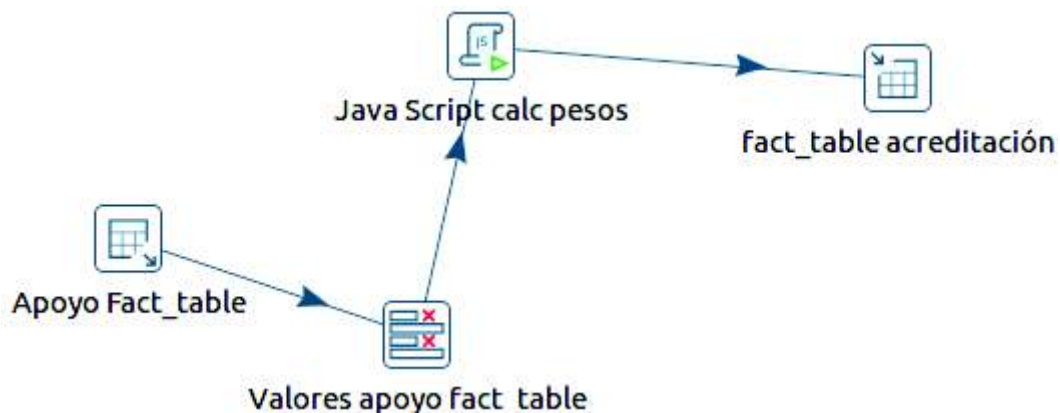


Figura 39. Proceso de carga de tabla de hechos<sup>30</sup>.

### 2.7.3.4. Proceso\_ETL\_DATA\_MART\_EPN

La carga de las tablas temporales y la tabla de hechos se realiza como se ilustra en la Figura 40.

El “Proceso\_ETL\_DATA\_MART\_EPN” comienza su ejecución en el nodo *start* para luego continuar con la ejecución de los *scripts* SQL “*staging area 1*” y “*staging area 2*”.

La ejecución del *script* “*staging area 1*” crea una tabla temporal en el *data mart* que contiene solo información de las variables necesarias para el cálculo de los indicadores cuantitativos. La ejecución del *script* “*staging area 2*” crea una tabla temporal en el *data mart* que contiene solo información de los resultados obtenidos de los indicadores.

<sup>30</sup> Fuente: Elaboración propia

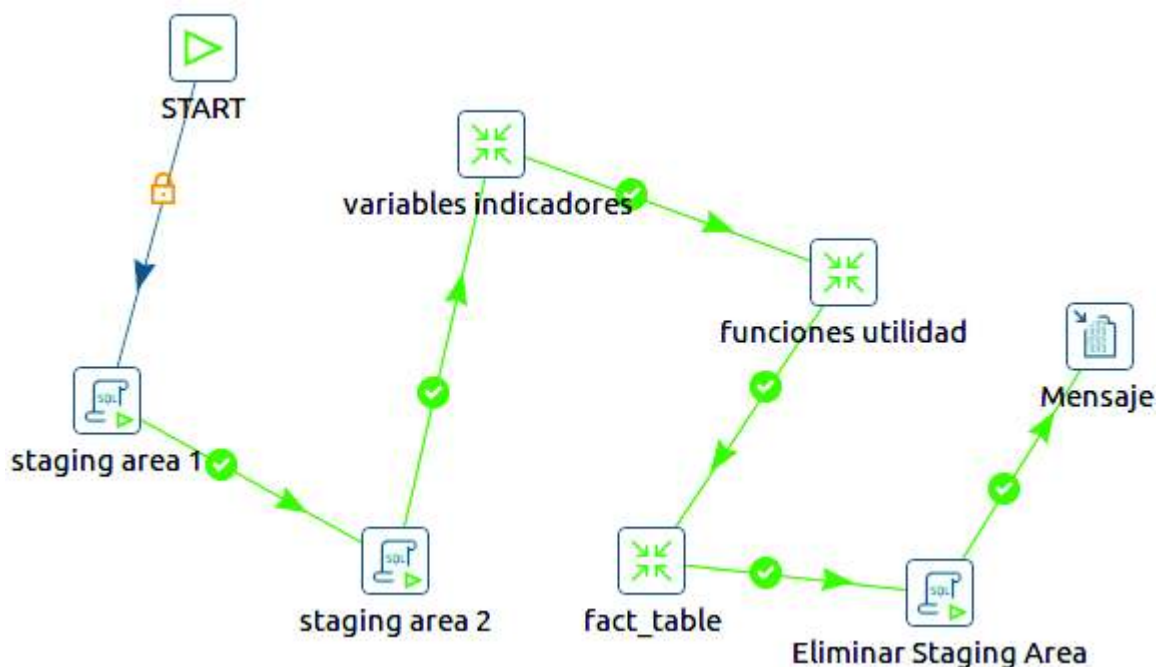


Figura 40. Proceso\_ETL\_DATA\_MART\_EPN<sup>31</sup>.

Los siguientes pasos “variables indicadores”, “funciones utilidad” y “*fact\_table*” cuyos procesos fueron detallados anteriormente, tienen como objetivo cargar los datos en las tablas temporales y tabla de hechos respectivamente.

Una vez que las tablas temporales cumplen su función deben ser eliminadas para lo cual se ejecuta el *script* “eliminar *staging area*”.

Finalmente si el trabajo fue realizado con éxito, se escribe un mensaje en el sistema de registro de PDI comunicando al usuario que el proceso fue ejecutado satisfactoriamente.

## 2.8. Especificación de aplicaciones de BI

Dentro del proyecto uno de los problemas a resolver es la generación de reportes creados a partir de la información registrada en el *data mart*, a fin de que los reportes sean utilizados para la toma de decisiones en el proceso de acreditación institucional.

Las herramientas seleccionadas para la elaboración de reportes forman parte de la plataforma Pentaho y se detallan en la Tabla 29.

<sup>31</sup> Fuente: Elaboración propia

Tabla 29. Herramientas para elaboración de reportes.

| Herramienta                               | Descripción   |
|---|---|
| CDE ( <i>Community Dashboard Editor</i> ) | Permite el desarrollo y despliegue de cuadros de mando avanzados a través de una interfaz gráfica e intuitiva.                |
| Saiku analytics                           | Permite el desarrollo y despliegue de vistas para el análisis de datos en forma de tablas pivotantes o gráficos estadísticos. |

### 2.8.1. Selección de requerimientos

- RQ01. Mostrar mediante un cuadro de mando el resultado, el peso y el peso por obtener de los criterios en un determinado periodo académico.
- RQ02. Mostrar mediante una vista de análisis de datos el peso obtenido de los criterios en un determinado periodo académico.
- RQ03. Mostrar mediante una vista de análisis de datos el peso obtenido y por obtener de los criterios en un determinado periodo académico.
- RQ04. Mostrar mediante un cuadro de mando el resultado, el peso y el peso por obtener de los subcriterios en un determinado periodo académico.
- RQ05. Mostrar mediante una vista de análisis de datos el peso obtenido de los subcriterios en un determinado periodo académico.
- RQ06. Mostrar mediante una vista de análisis de datos el peso obtenido y por obtener de los subcriterios en un determinado periodo académico.
- RQ07. Mostrar mediante un cuadro de mando el resultado obtenido de los indicadores en un determinado periodo académico.
- RQ08. Mostrar mediante un cuadro de mando el peso obtenido y por obtener de los indicadores en un determinado periodo académico.
- RQ09. Mostrar mediante un cuadro de mando la utilidad obtenida de los indicadores en un determinado periodo académico.
- RQ10. Mostrar mediante una vista de análisis de datos el valor, peso y utilidad de los indicadores agrupados por criterio y subcriterio en un determinado periodo académico.

- RQ11. Mostrar mediante una vista de análisis de datos el peso obtenido y por obtener de los indicadores agrupados por criterio y subcriterio en un determinado periodo académico.
- RQ12. Mostrar mediante una vista de análisis de datos la utilidad de los indicadores agrupada por criterio y subcriterio en un determinado periodo académico.

## **2.8.2. Definición de roles de usuario**

En esta fase es importante establecer que usuarios tendrán acceso a los reportes. Para este proyecto el usuario final es el siguiente:

Responsable de acreditación universitaria: el perfil de este usuario es de consulta. El usuario tendrá acceso a los reportes solicitados según sus requerimientos.

## **2.9. Desarrollo de aplicaciones de BI**

### **2.9.1. Desarrollo del cubo OLAP “DATA MART CEAACES”**

El cubo OLAP permitirá presentar datos de las dimensiones y hechos de forma intuitiva, con el fin de mejorar el rendimiento de las consultas.

Al desarrollar el cubo OLAP se crearon jerarquías en las dimensiones para satisfacer los requerimientos expresados por los usuarios.

A continuación se detallan los atributos y las vistas del origen de datos de las respectivas dimensiones.

#### **2.9.1.1. Dimensión Criterio\_dim**

En la Figura 41 se puede apreciar la dimensión criterio con sus atributos asociados, incluido el atributo de clave “Criterio Key”.



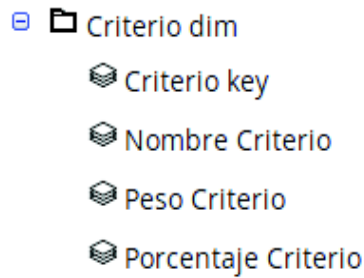


Figura 41. Atributos de la dimensión Criterio\_dim<sup>32</sup>.

La Figura 42 muestra la tabla “criterio\_dim” que forma parte del origen de datos de la dimensión criterio. La tabla cuenta con 4 columnas y su función principal es permitir explorar los datos.

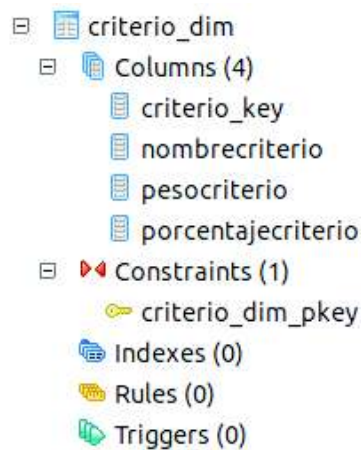


Figura 42. Origen de datos de la dimensión Criterio\_dim<sup>33</sup>.

### 2.9.1.2. Dimensión Subcriterio\_dim

En la Figura 43 se puede apreciar la dimensión subcriterio con sus atributos asociados, incluido el atributo de clave “Subcriterio Key”.

La Figura 44 muestra la tabla “subcriterio\_dim” que forma parte del origen de datos de la dimensión subcriterio. La tabla cuenta con 4 columnas y su función principal es permitir explorar los datos.

---

<sup>32</sup> Fuente: Elaboración propia

<sup>33</sup> Fuente: Elaboración propia

- [-] Subcriterio dim
  - [-] Subcriterio key
  - [-] Nombre Subcriterio
  - [-] Peso Subcriterio
  - [-] Porcentaje Subcriterio

Figura 43. Atributos de la dimensión Subcriterio\_dim<sup>34</sup>.

- [-] subcriterio\_dim
  - [-] Columns (4)
    - [-] subcriterio\_key
    - [-] nombresubcriterio
    - [-] pesosubcriterio
    - [-] porcentajesubcriterio
  - [-] Constraints (1)
  - [-] Indexes (0)
  - [-] Rules (0)
  - [-] Triggers (0)

Figura 44. Origen de datos de la dimensión Subcriterio\_dim<sup>35</sup>.

### 2.9.1.3. Dimensión Indicador\_dim

En la Figura 45 se puede apreciar la dimensión indicador con sus atributos asociados, incluido el atributo de clave “Indicador Key”.

- [-] Indicador dim
  - [-] Indicador key
  - [-] Nombre Indicador
  - [-] Peso Indicador

Figura 45. Atributos de la dimensión Indicador\_dim<sup>36</sup>.

<sup>34</sup> Fuente: Elaboración propia

<sup>35</sup> Fuente: Elaboración propia

<sup>36</sup> Fuente: Elaboración propia

La Figura 46 muestra la tabla “indicador\_dim” que forma parte del origen de datos de la dimensión indicador. La tabla cuenta con 3 columnas y su función principal es permitir explorar los datos.

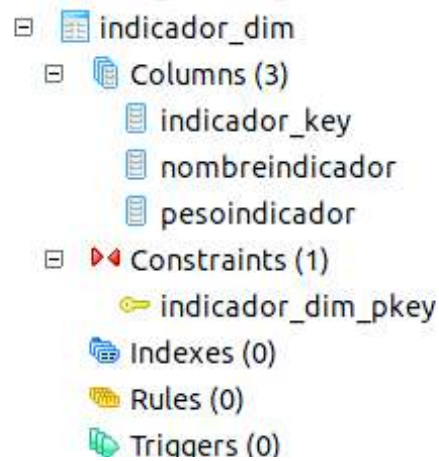


Figura 46. Vista de origen de datos de dimensión Indicador\_dim<sup>37</sup>.

#### 2.9.1.4. Dimensión Tiempo\_dim

En la Figura 47 se puede apreciar la dimensión tiempo con sus atributos asociados, incluido el atributo de clave “Tiempo Key”.

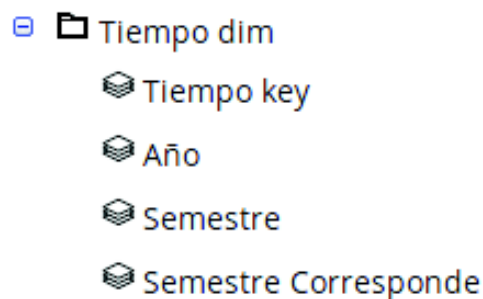


Figura 47. Atributos de dimensión Tiempo\_dim<sup>38</sup>.

La Figura 48 muestra la tabla “tiempo\_dim” que forma parte del origen de datos de la dimensión tiempo. La tabla cuenta con 4 columnas y su función principal es permitir explorar los datos.

<sup>37</sup> Fuente: Elaboración propia

<sup>38</sup> Fuente: Elaboración propia

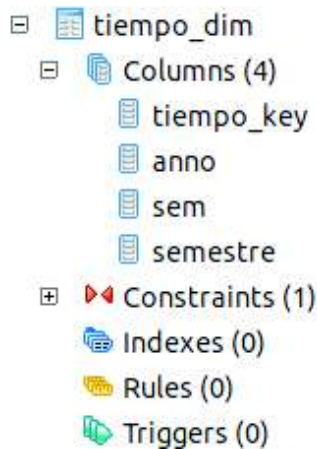


Figura 48. Vista de origen de datos de dimensión Tiempo\_dim<sup>39</sup>.

### 2.9.1.5. Tabla de hechos fact modelo acreditación

La tabla de hechos “Fact modelo acreditación” que se ilustra en la Figura 49 contiene las claves subrogadas de las dimensiones criterio, subcriterio, indicador y tiempo. La clave primaria de la tabla de hechos es la combinación de las claves primarias de sus dimensiones. Adicionalmente, la tabla de hechos incluye las medidas, en este caso, el valor obtenido, utilidad, peso obtenido y diferencia de pesos que representan los valores que se desean analizar.



Figura 49. Tabla de hechos fact\_modelo\_acreditación<sup>40</sup>.

### 2.9.1.6. Diagrama del cubo OLAP DATA MART CEAACES

La Figura 50 muestra el diagrama del cubo OLAP, en el que la tabla de hechos “fact\_modelo\_acreditación” se relaciona con 4 dimensiones, criterio\_dim, subcriterio\_dim, indicador\_dim y tiempo\_dim. Las columnas criterio\_key, subcriterio\_key,

<sup>39</sup> Fuente: Elaboración propia

<sup>40</sup> Fuente: Elaboración propia

indizador\_key y tiempo\_key en la tabla de hechos “fact\_modelo\_acreditacion” definen una relación de clave externa con las columnas de clave primaria de las respectivas dimensiones.

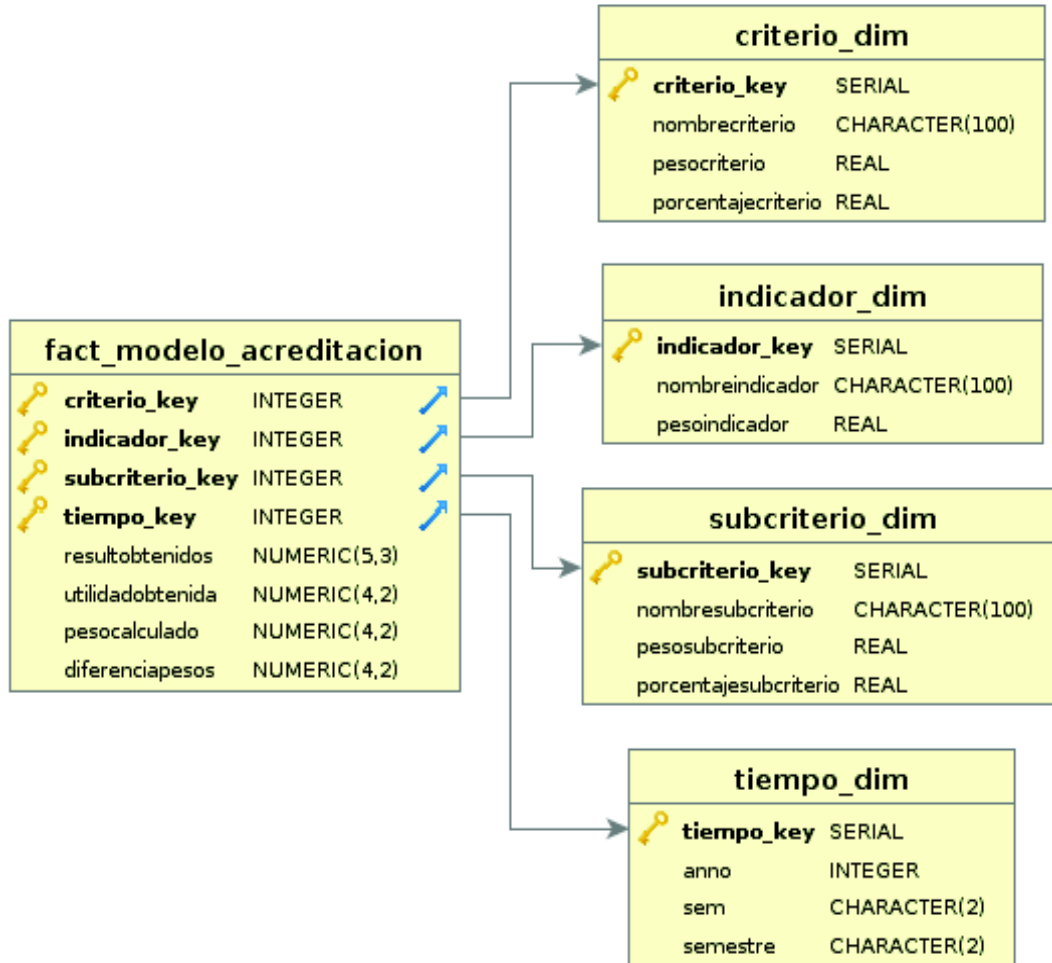


Figura 50. Diagrama del cubo OLAP “data mart CEAACES”<sup>41</sup>.

### 2.9.2. Desarrollo de aplicativo web para desplegar reportes

Esta aplicación tiene como propósito permitir que los usuarios puedan realizar consultas interactivas para obtener información referente a los indicadores de acreditación universitaria. Los usuarios mediante el aplicativo web podrán desplegar reportes analíticos *ad-hoc* y cuadros de mando parametrizados.

<sup>41</sup> Fuente: Elaboración propia

La metodología elegida para el desarrollo del aplicativo es OOHDM (*Object Oriented Hypermedia Design Model*) [31]. Esta metodología propone el diseño de aplicaciones web a través de 5 fases que se detallan en la Figura 51.

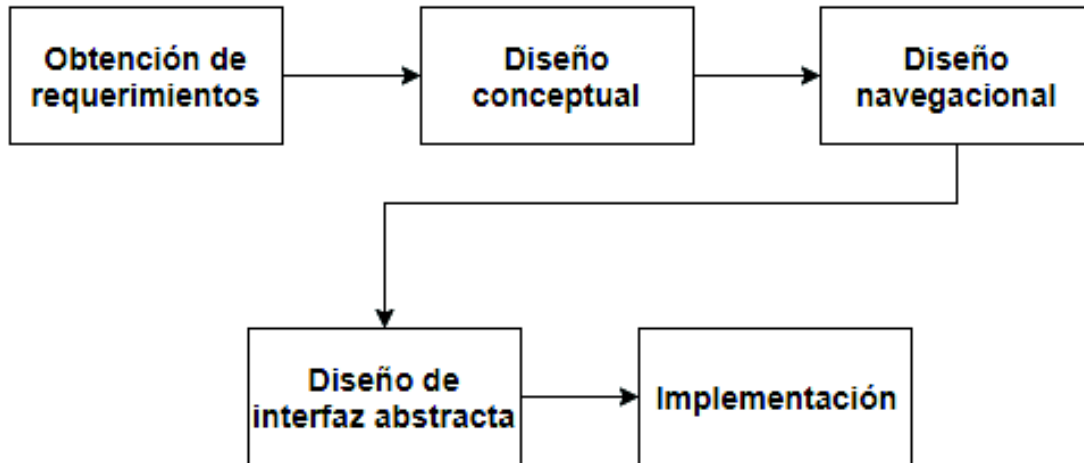



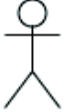

Figura 51. Fases de la metodología OOHDM [32].

### 2.9.2.1. Obtención de requerimientos

La herramienta en la cual se fundamenta esta fase son los diagramas de los casos de uso. En esta actividad es necesario identificar los actores y las tareas que cada uno debe realizar. La Identificación de actores se detalla en la Tabla 30.

Tabla 30. Identificación de actores.

| Actores   | Función   |
|---|---|
|  | <ul style="list-style-type: none"> <li>- Iniciar sesión en el sistema</li> <li>- Administrar usuarios               <ul style="list-style-type: none"> <li>Mostrar todos</li> <li>Crear</li> <li>Modificar</li> <li>Dar de baja</li> </ul> </li> <li>- Gestionar niveles               <ul style="list-style-type: none"> <li>Listar</li> <li>Crear</li> <li>Modificar</li> <li>Eliminar</li> </ul> </li> </ul> |

|  |  |
|--|--|
|  | <ul style="list-style-type: none"> <li>- Administrar reportes</li> <li>  Desplegar reportes</li> <li>  Consultar</li> <li>- Cerrar sesión</li> </ul> |
| <br>Usuario   | <ul style="list-style-type: none"> <li>- Iniciar sesión en el sistema</li> <li>- Desplegar reportes</li> <li>- Cerrar sesión</li> </ul>              |
| <br>Data Mart | <ul style="list-style-type: none"> <li>- Crear información a mostrar</li> </ul>  |

Las Figuras 52, 53, 54 y 55 muestran los diagramas de casos de uso del sistema.

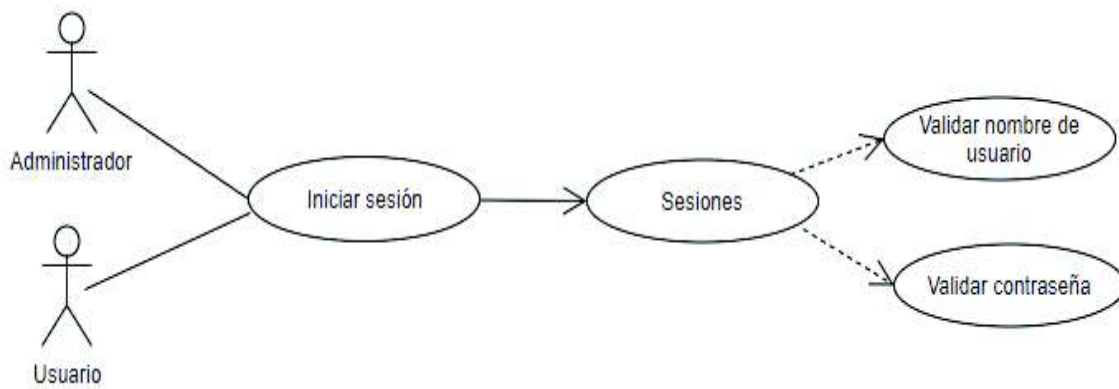


Figura 52. Diagrama de casos de inicio de sesión<sup>42</sup>.

<sup>42</sup> Fuente: Elaboración propia

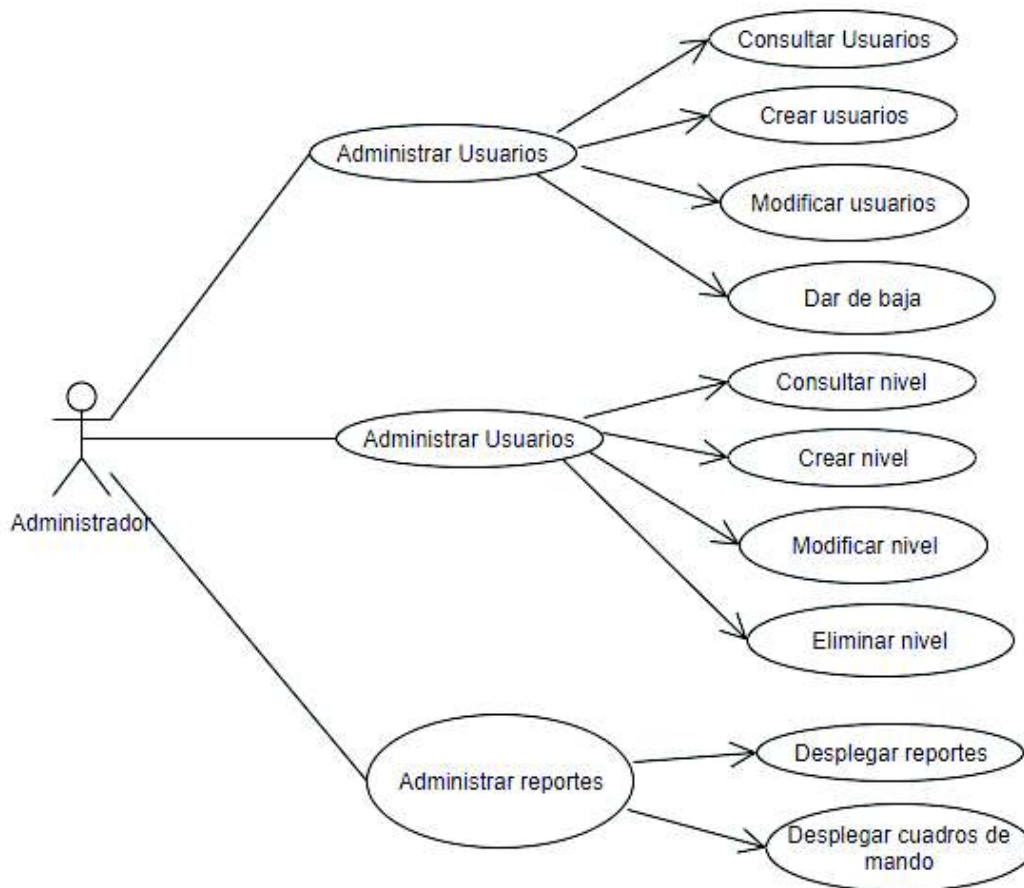


Figura 53. Diagrama de casos administrador<sup>43</sup>.

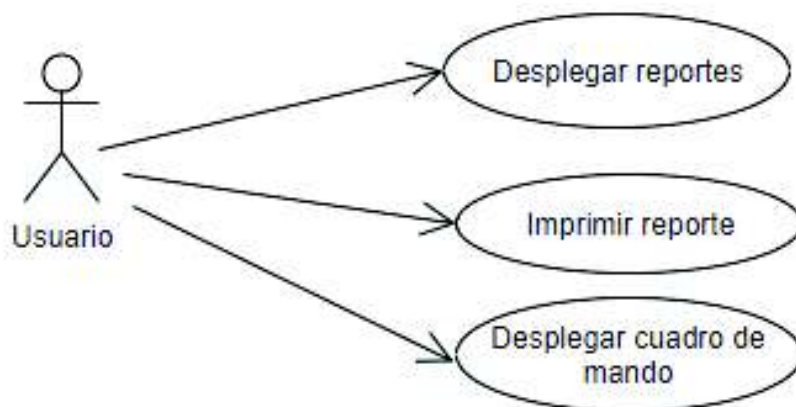


Figura 54. Diagrama de casos usuario<sup>44</sup>.

<sup>43</sup> Fuente: Elaboración propia

<sup>44</sup> Fuente: Elaboración propia



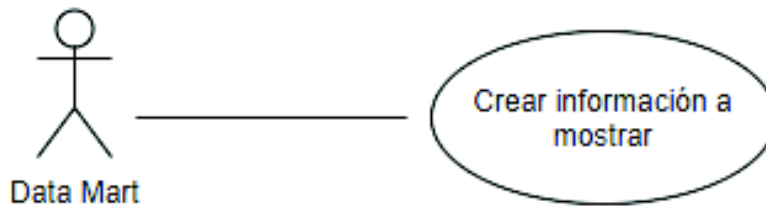


Figura 55. Diagrama de casos data mart<sup>45</sup>.

### 2.9.2.2. Diseño conceptual

En esta actividad se diseña un esquema conceptual representado por los objetos, las relaciones y colaboraciones establecidas entre ellos. La Figura 56 ilustra el diagrama de clases del sistema, donde las clases son usadas durante el diseño navegacional para derivar nodos y las relaciones para construir enlaces.

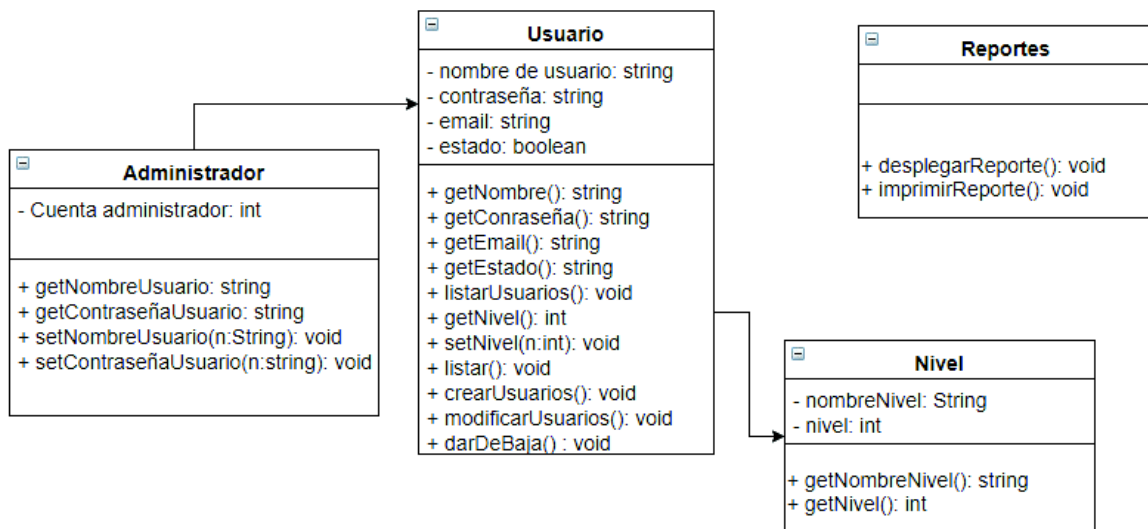


Figura 56. Diagrama de clases<sup>46</sup>.

Las Figuras 57, 58, 59, 60, 61, 62, 63, 64, 65 y 66 muestran los diagramas de secuencia que modelan la interacción de un conjunto de objetos de una aplicación a través del tiempo. Los diagramas de secuencia indican que clases o módulos van a formar parte del aplicativo, las respectivas llamadas que se hace a cada uno de ellos para realizar una tarea y que mensajes disparan esas llamadas.

<sup>45</sup> Fuente: Elaboración propia

<sup>46</sup> Fuente: Elaboración propia

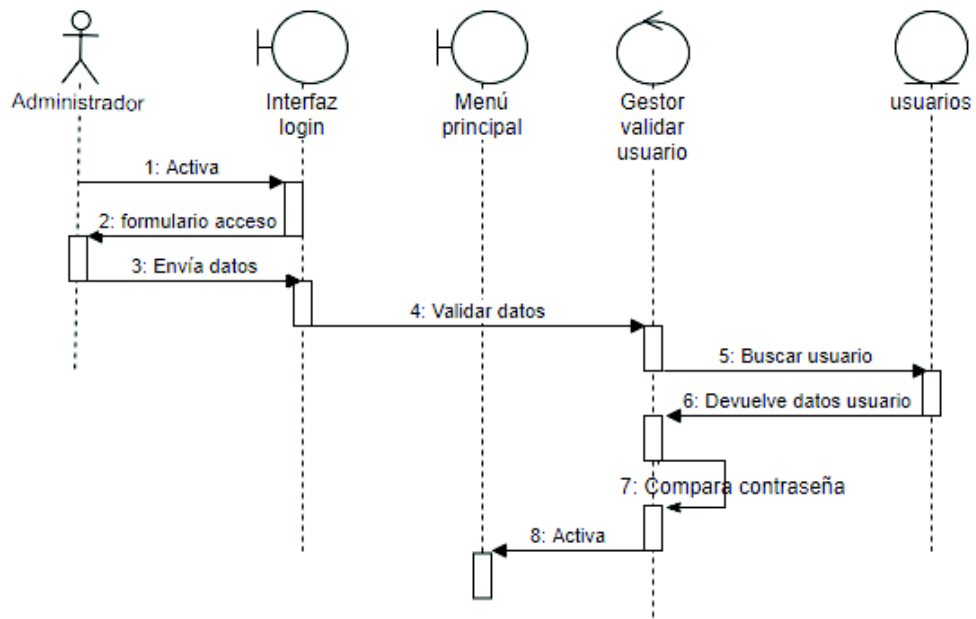


Figura 57. Diagrama de secuencia ingresar a la plataforma del administrador<sup>47</sup>.

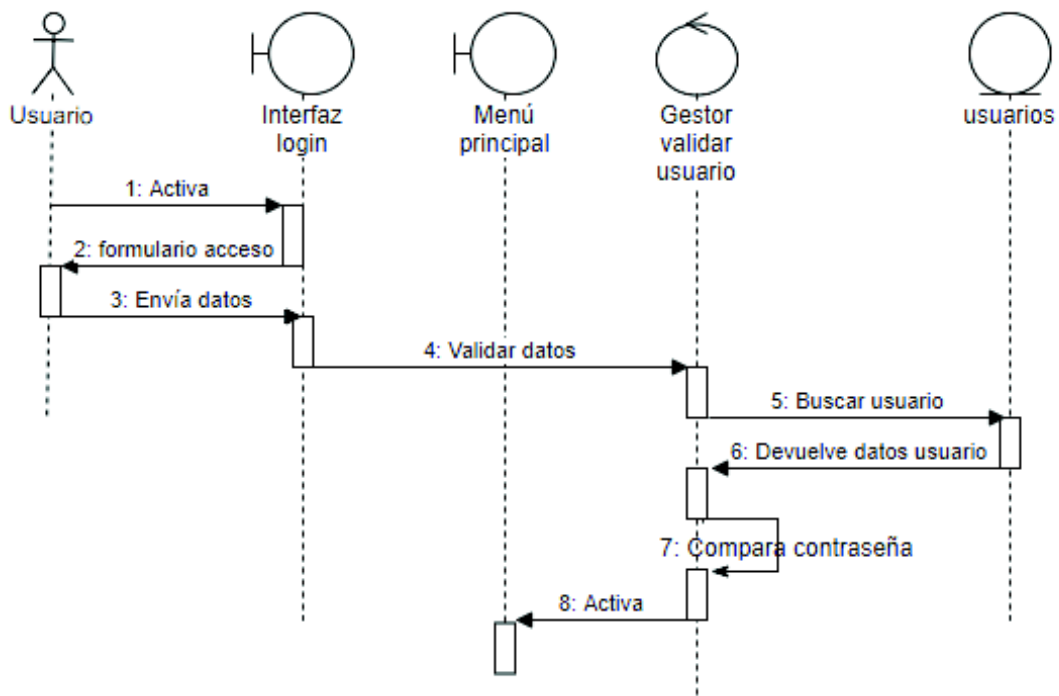


Figura 58. Diagrama de secuencia ingresar a la plataforma del usuario<sup>48</sup>.

<sup>47</sup> Fuente: Elaboración propia

<sup>48</sup> Fuente: Elaboración propia

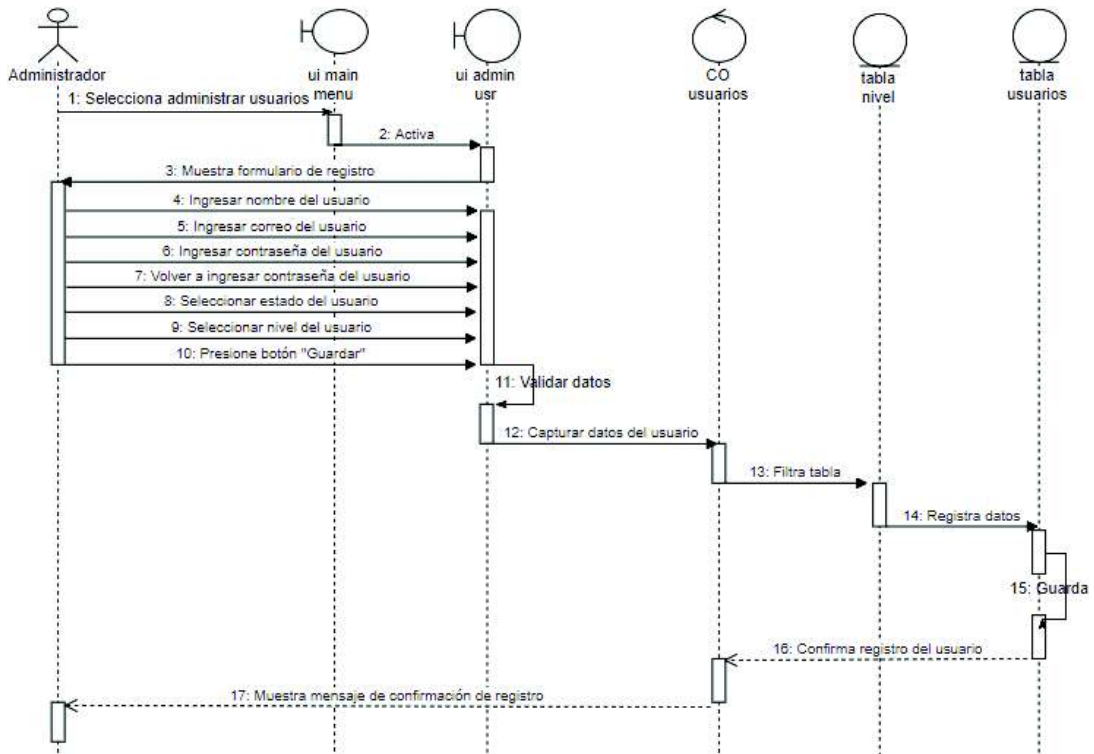


Figura 59. Diagrama de secuencia registrar usuario<sup>49</sup>.

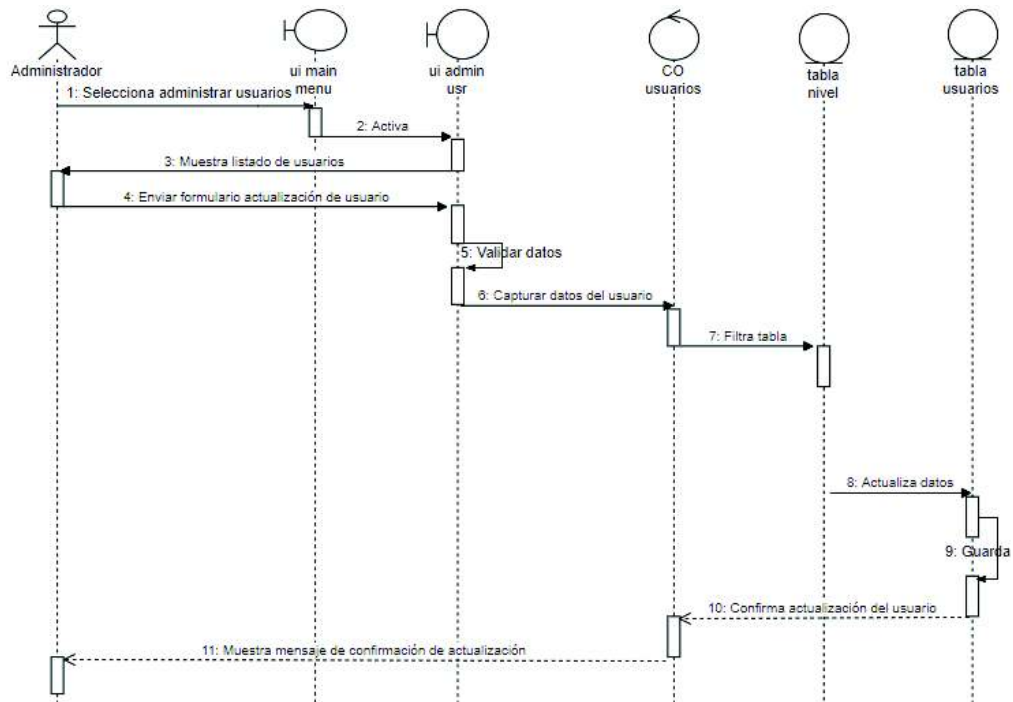


Figura 60. Diagrama de secuencia actualizar usuario<sup>50</sup>.

<sup>49</sup> Fuente: Elaboración propia

<sup>50</sup> Fuente: Elaboración propia

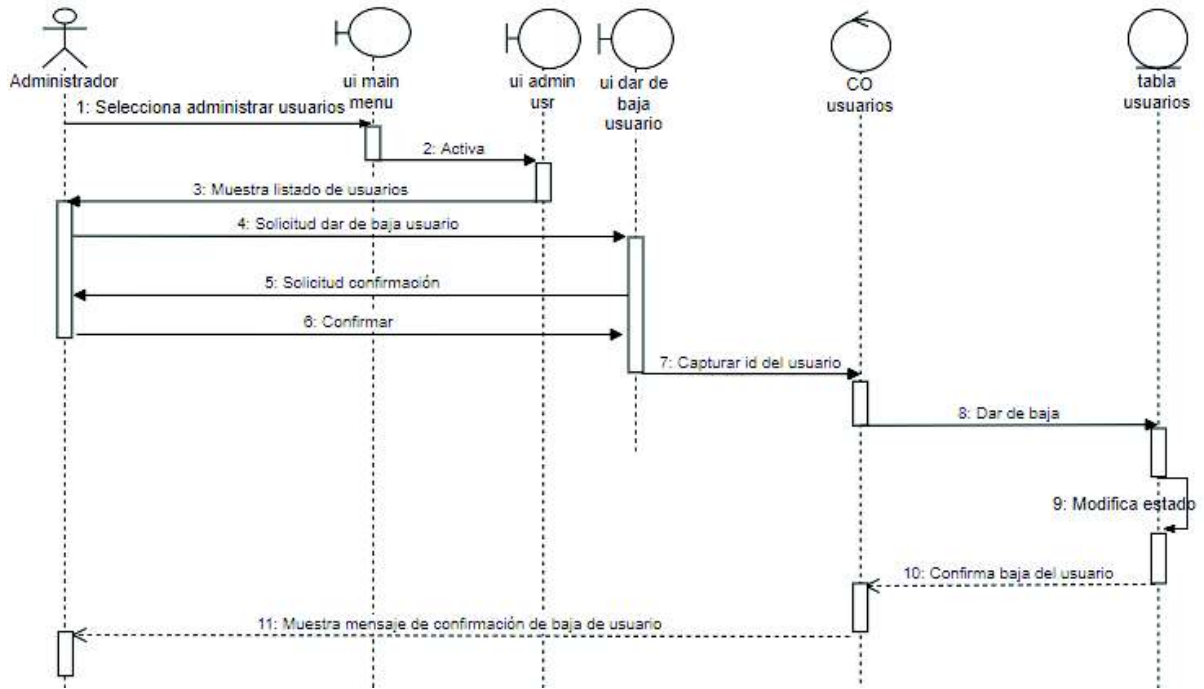


Figura 61. Diagrama de secuencia dar de baja usuario<sup>51</sup>.

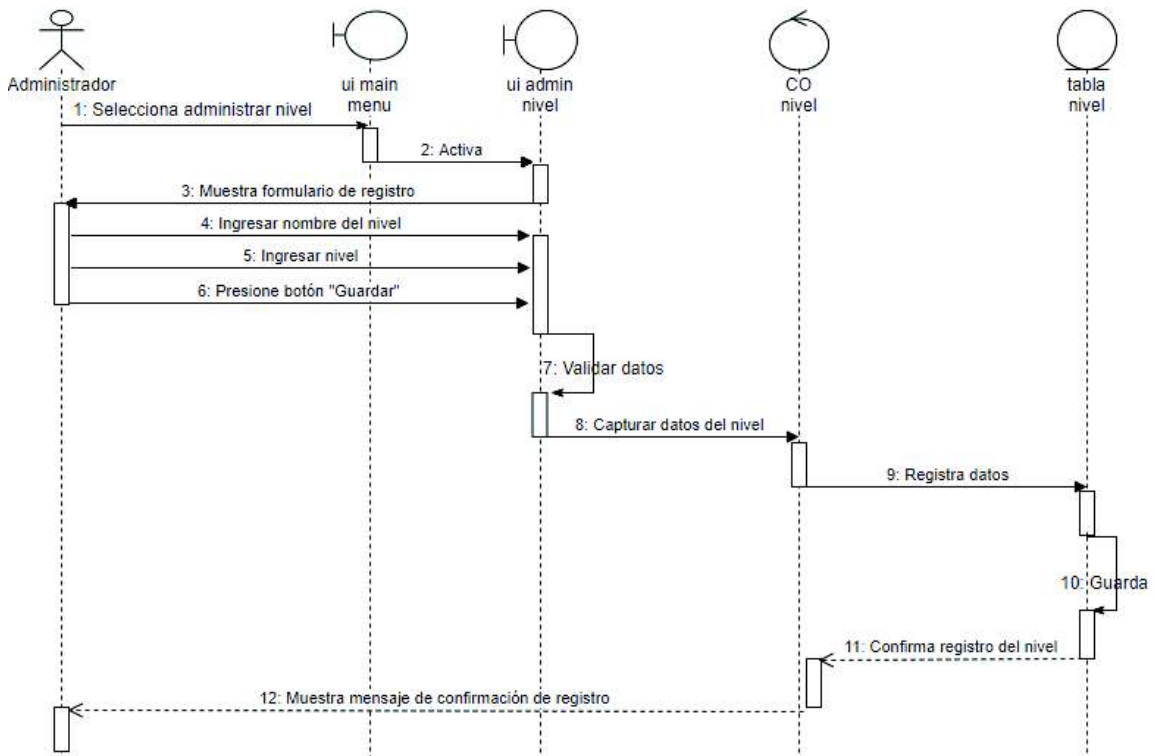


Figura 62. Diagrama de secuencia crear nivel<sup>52</sup>.

<sup>51</sup> Fuente: Elaboración propia

<sup>52</sup> Fuente: Elaboración propia

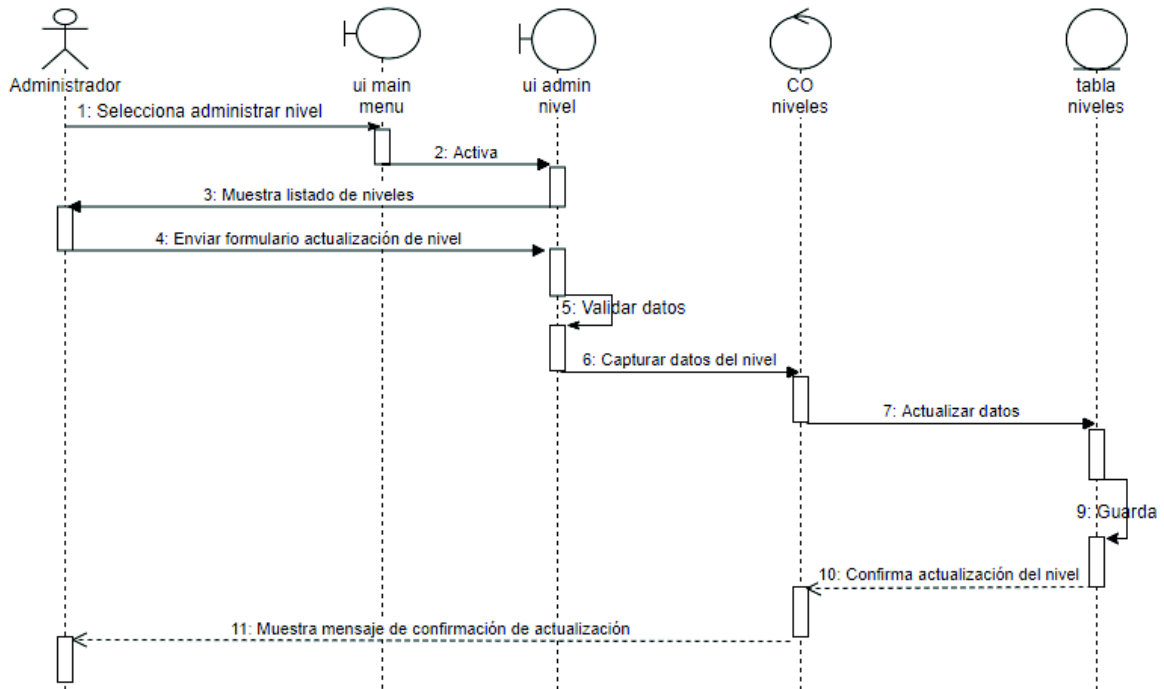


Figura 63. Diagrama de secuencia modificar nivel<sup>53</sup>.

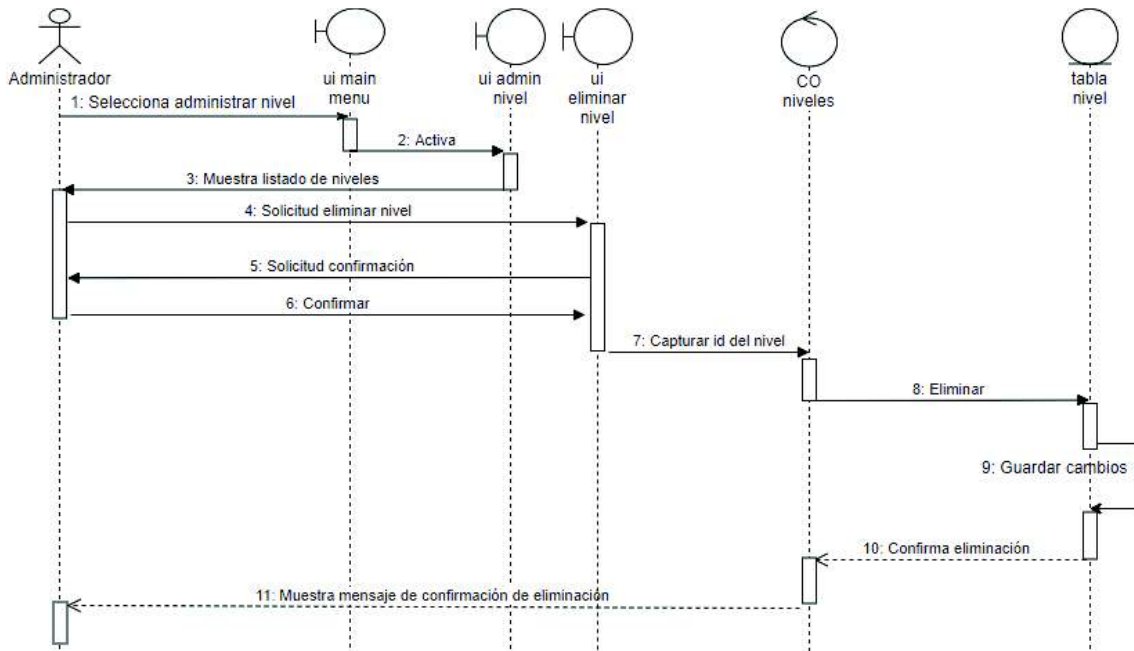


Figura 64. Diagrama de secuencia eliminar nivel<sup>54</sup>.

<sup>53</sup> Fuente: Elaboración propia

<sup>54</sup> Fuente: Elaboración propia

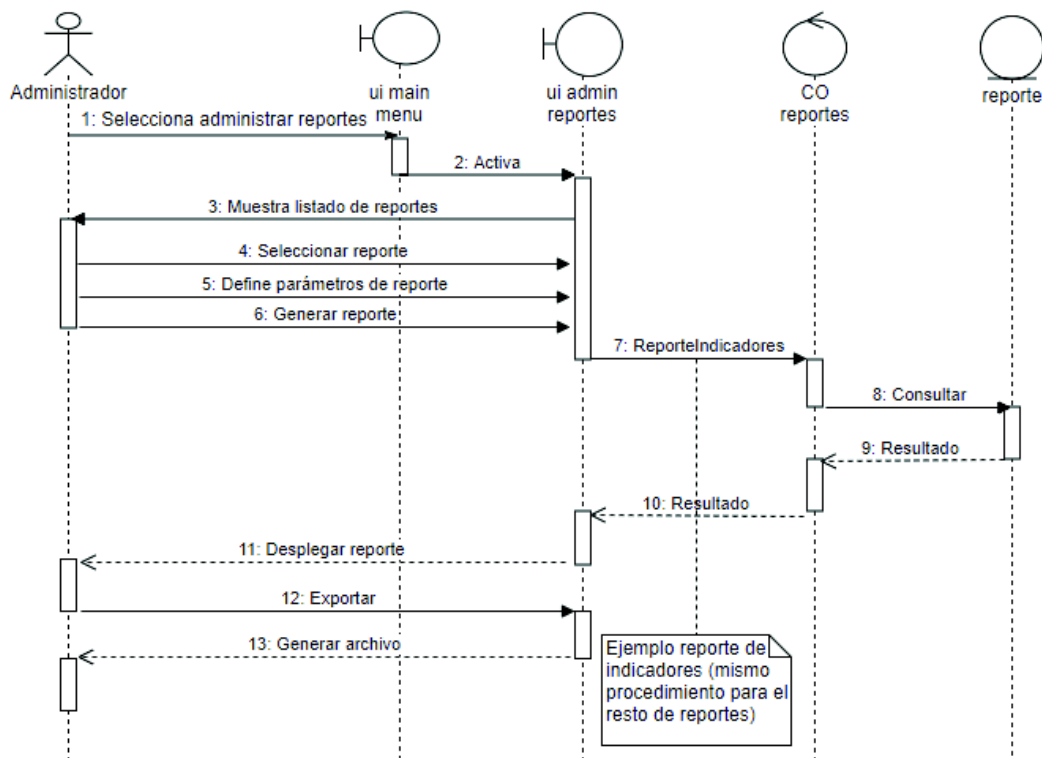


Figura 65. Diagrama de secuencia desplegar reportes del administrador<sup>55</sup>.

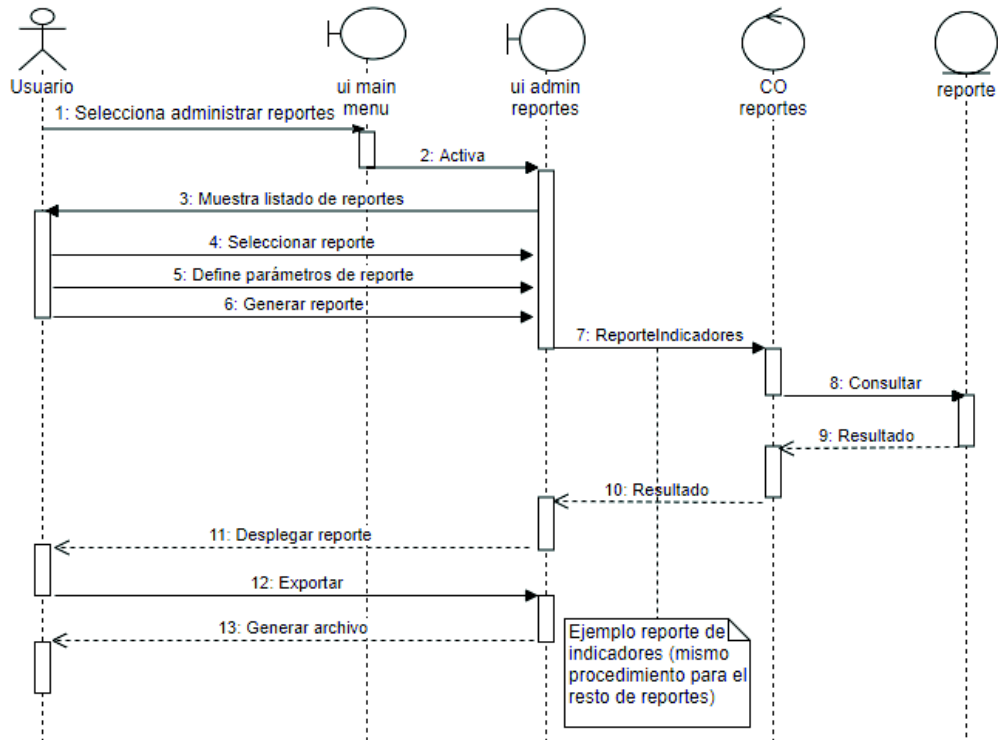


Figura 66. Diagrama de secuencia desplegar reportes del usuario<sup>56</sup>.

<sup>55</sup> Fuente: Elaboración propia

<sup>56</sup> Fuente: Elaboración propia

### 2.9.2.3. Diseño navegacional

La Figura 67 muestra el diagrama de navegación de la aplicación en el cual se define el conjunto de nodos y enlaces que forman parte de una vista navegacional de la aplicación web. Los nodos y enlaces se definen en el diseño conceptual.

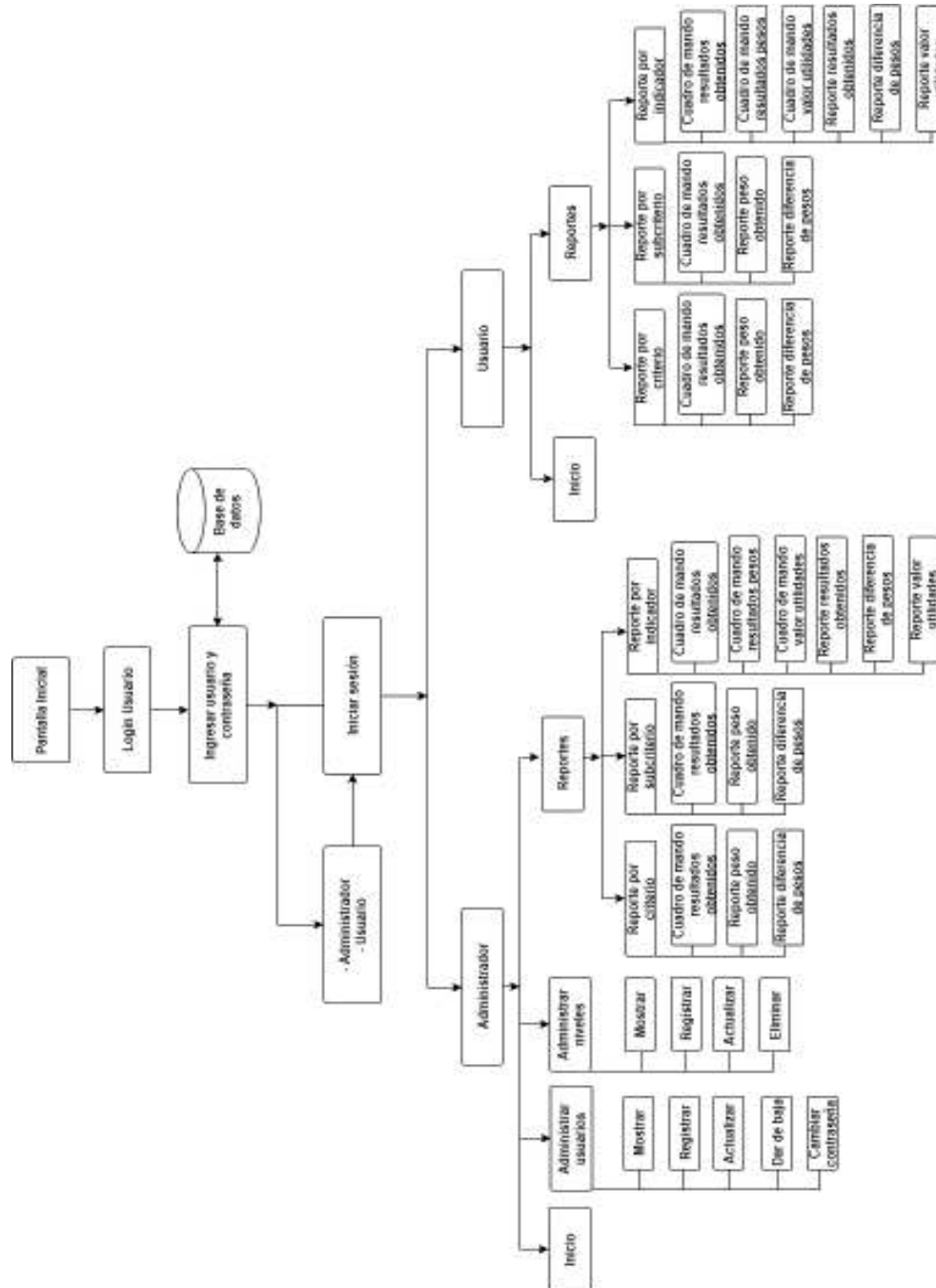


Figura 67. Diagrama de navegación<sup>57</sup>.

<sup>57</sup> Fuente: Elaboración propia

### 2.9.2.4. Diseño de interfaz abstracta

Las Figuras 68, 69 y 70 muestran los diseños de las interfaces abstractas que permiten especificar las diferentes interfaces de la aplicación. Para realizar el diseño de interfaces abstractas se definen las ADV (*Abstract Data View*) que especifican el comportamiento y organización de la interfaz.



Figura 68. Diseño de interfaz abstracta ingreso al sistema<sup>58</sup>.

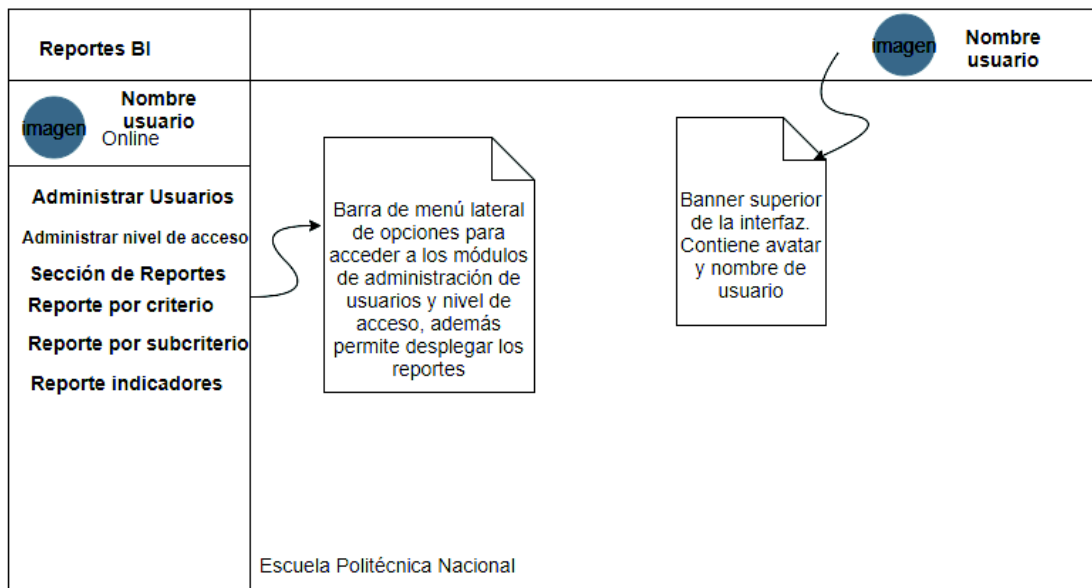


Figura 69. Diseño de interfaz abstracta administrador<sup>59</sup>.

<sup>58</sup> Fuente: Elaboración propia

<sup>59</sup> Fuente: Elaboración propia



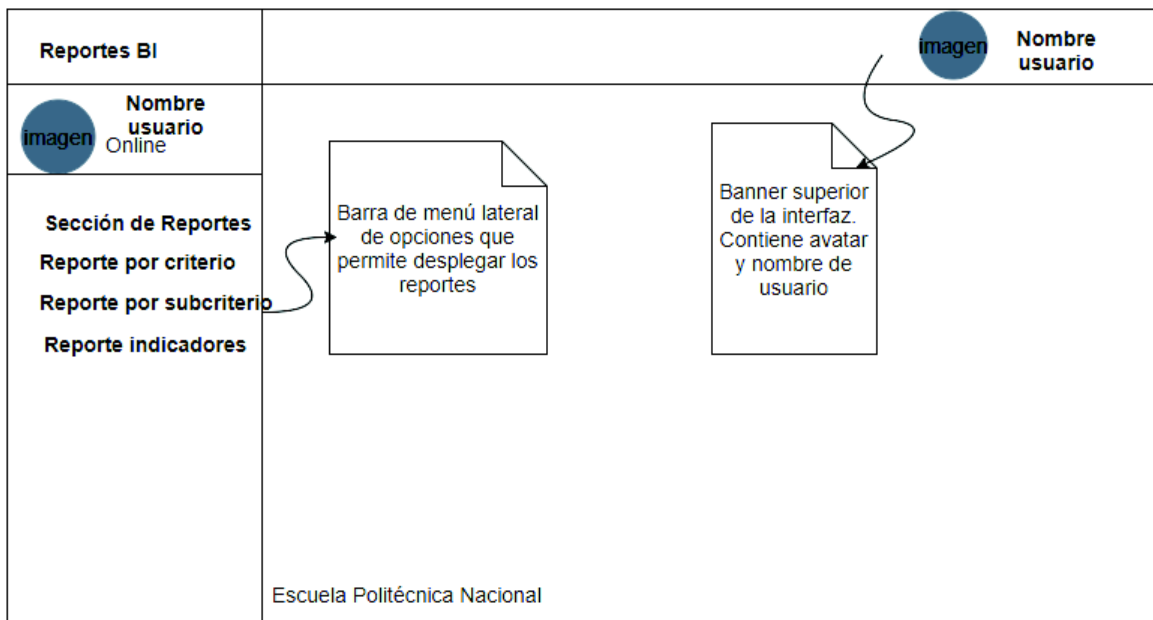


Figura 70. Diseño de interfaz abstracta usuario<sup>60</sup>.

### 2.9.3. Desarrollo de reportes iniciales

Los reportes permiten que el usuario final conozca la información consolidada y agregada para la toma de decisiones. Existen diferentes tipos de reportes: a) estáticos que tienen un formato preestablecido fijo, b) parametrizados que presentan parámetros de entrada y permiten múltiples consultas y, c) ad-hoc que permite que los usuarios finales puedan construir fácilmente sus propios informes y modificar los existentes con poca o ninguna capacitación.

#### 2.9.3.1. Cuadro de mando: Resultados obtenidos por criterio

En la Figura 71 se puede apreciar un cuadro de mando, el mismo que consta de 4 paneles, cada uno muestra información relevante agrupada por indicador para la toma de decisiones. Por ejemplo, en el primer panel se puede apreciar que el criterio de estudiantes en el periodo académico 2018-A, obtiene un valor 8.15 sobre 12. Los parámetros de entrada del cuadro de mando son: a) criterio y b) período académico.

<sup>60</sup> Fuente: Elaboración propia



Figura 71. Cuadro de mando: Resultados obtenidos por criterio<sup>61</sup>.

### 2.9.3.2. Reporte peso obtenido por criterio

El reporte que se puede apreciar en la Figura 72 es de tipo *ad-hoc* y muestra al usuario información del peso obtenido del criterio agrupado por periodo académico. La información es consumida del cubo.

<sup>61</sup> Fuente: Elaboración propia

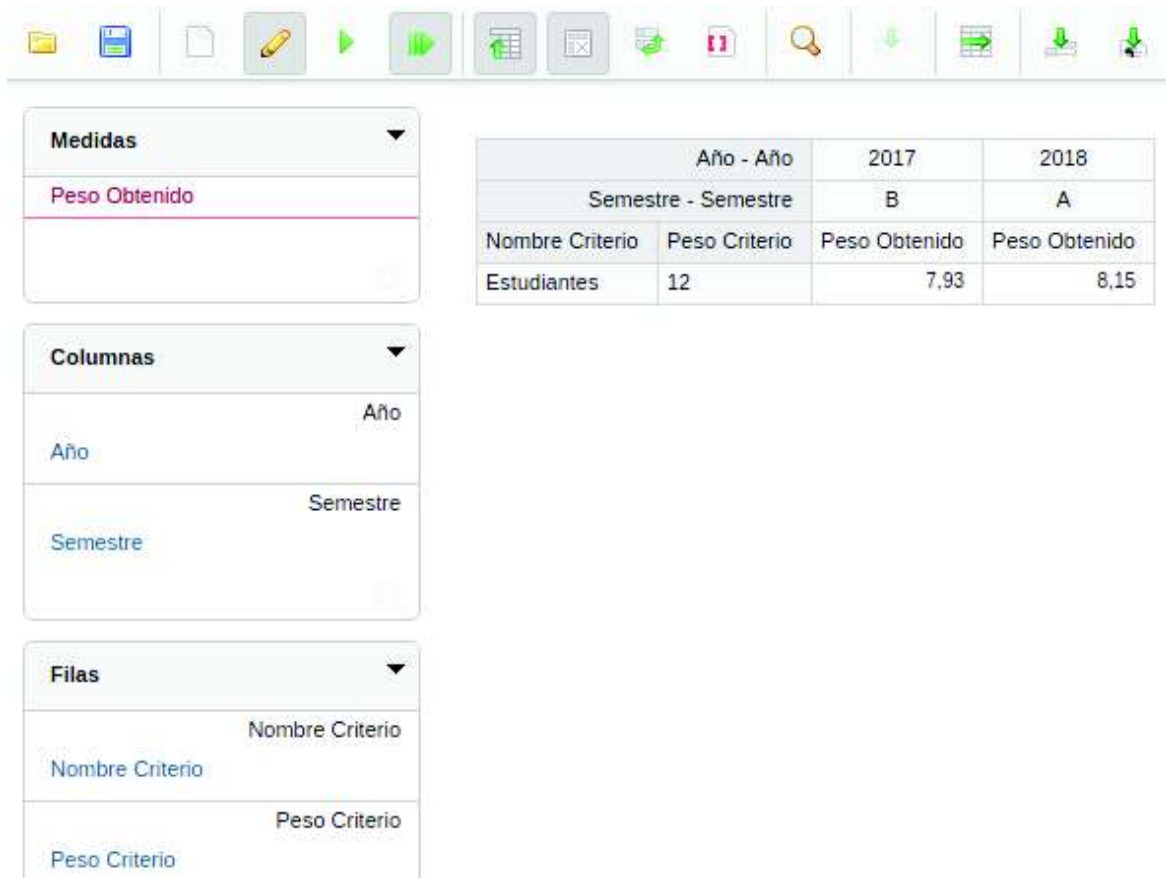


Figura 72. Reporte peso obtenido por criterio<sup>62</sup>.

### 2.9.3.3. Reporte del peso obtenido y por obtener según el criterio

El reporte que se puede apreciar en la Figura 73 es de tipo *ad-hoc* y muestra al usuario información del peso que puede alcanzar, del peso obtenido y cuanto le faltaría obtener al criterio para llegar al valor ideal. La información es consumida del cubo.

### 2.9.3.4. Cuadro de mando: Resultados obtenidos por subcriterio

En la Figura 74 se puede apreciar un cuadro de mando, el mismo que consta de 4 paneles, cada uno muestra información relevante agrupada por subcriterio para la toma de decisiones. Los parámetros de entrada son: a) período académico y b) subcriterio.

<sup>62</sup> Fuente: Elaboración propia

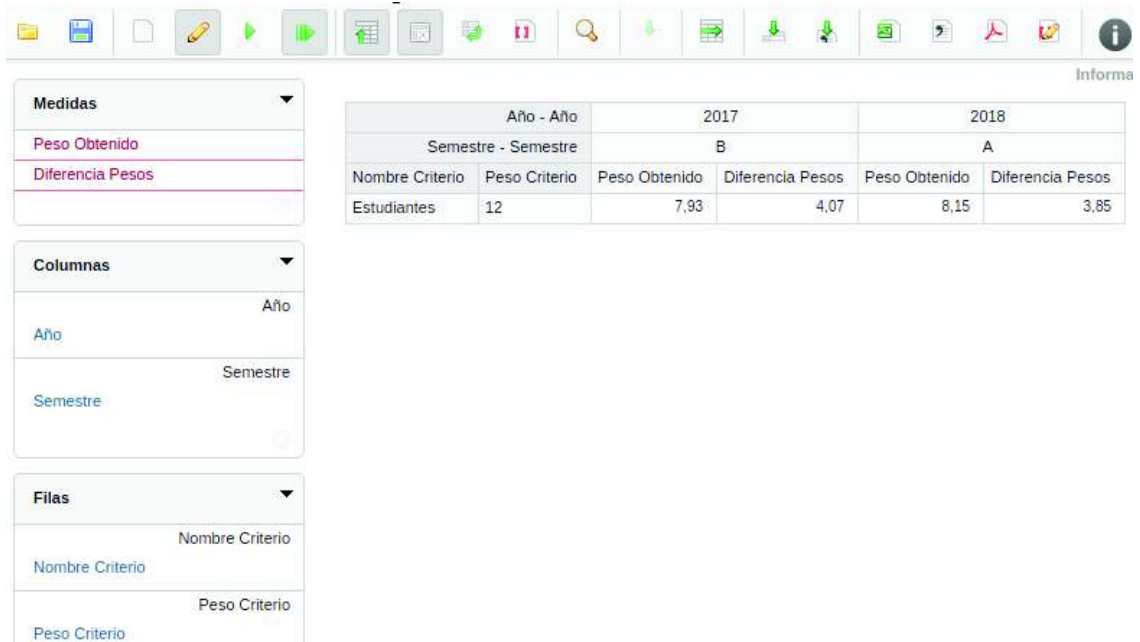


Figura 73. Reporte comparativo entre el peso obtenido y por obtener según el criterio<sup>63</sup>.



Figura 74. Cuadro de mando: Resultados obtenidos por subcriterio<sup>64</sup>.

<sup>63</sup> Fuente: Elaboración propia

<sup>64</sup> Fuente: Elaboración propia

### 2.9.3.5. Reporte del peso obtenido por subcriterio

El reporte que se puede apreciar en la Figura 75 es de tipo *ad-hoc* y muestra al usuario información del peso obtenido por subcriterio agrupado por periodo académico. La información es consumida del cubo.

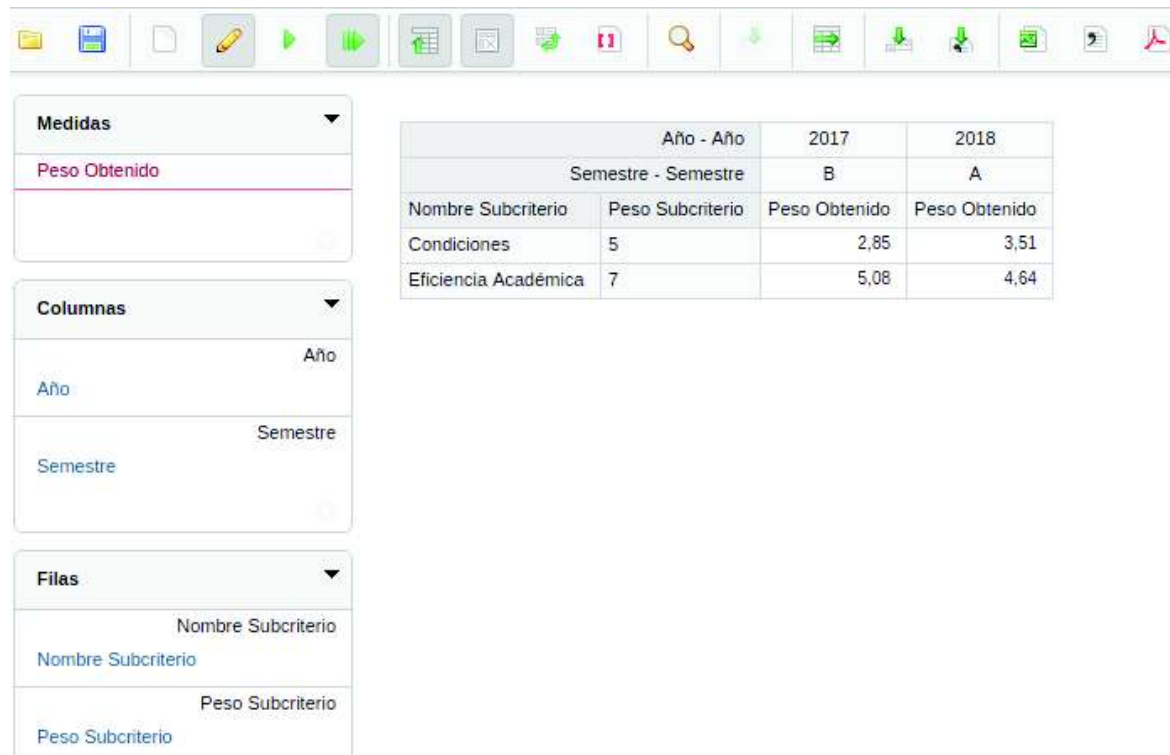


Figura 75. Reporte del peso obtenido por subcriterio<sup>65</sup>.

### 2.9.3.6. Reporte del peso obtenido y por obtener de los subcriterios

El reporte que se puede apreciar en la Figura 76 es de tipo *ad-hoc* y muestra al usuario información del peso que puede alcanzar, del peso obtenido y cuanto le faltaría obtener a cada subcriterio para llegar al valor ideal. La información es consumida del cubo.

<sup>65</sup> Fuente: Elaboración propia

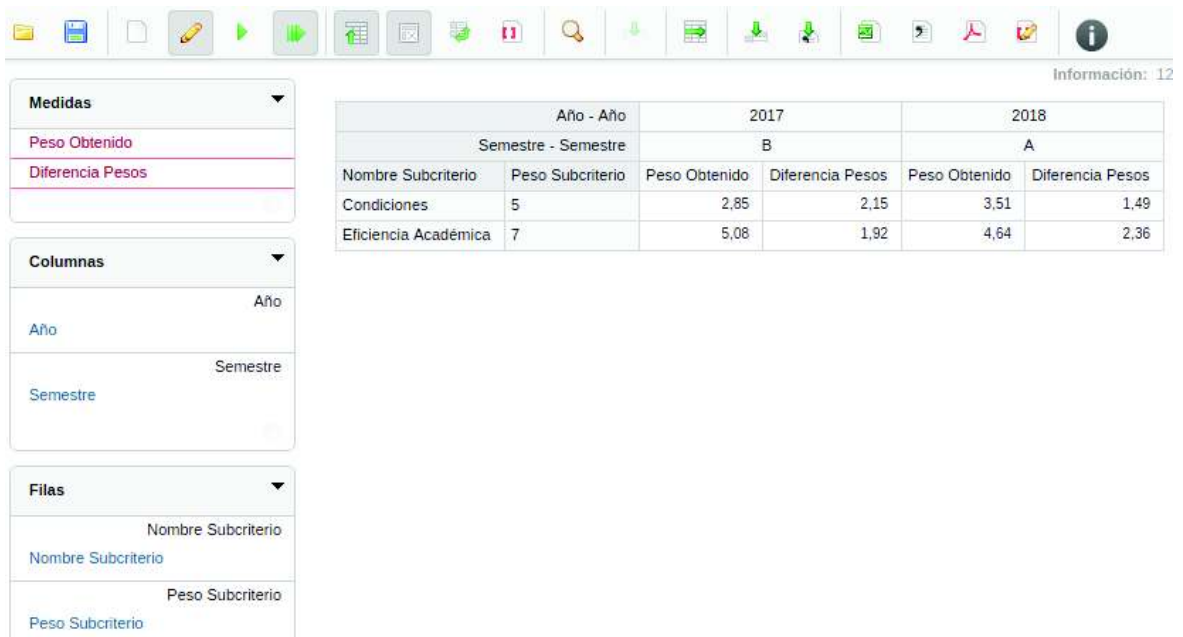


Figura 76. Reporte comparativo entre el peso obtenido y por obtener de los subcriterios<sup>66</sup>.

### 2.9.3.7. Cuadro de mando: Resultados obtenidos por indicador

En la Figura 77 se puede apreciar un cuadro de mando, el mismo que consta de 4 paneles, cada uno muestra información del valor obtenido de los indicadores cuantitativos y cualitativos para la toma de decisiones. Los parámetros de entrada son: a) indicador y b) período académico.

### 2.9.3.8. Cuadro de mando: Pesos obtenidos y por obtener de los indicadores

En la Figura 78 se puede apreciar un cuadro de mando, el mismo que consta de 6 paneles. Los dos primeros paneles contienen el componente *dial chart* que es un gráfico de medidor que muestra información del peso obtenido de los indicadores. Los 4 paneles restantes detallan información del peso obtenido y por obtener de los indicadores cuantitativos y cualitativos. Los parámetros de entrada son: a) indicador y b) período académico.

<sup>66</sup> Fuente: Elaboración propia

### 2.9.3.9. Cuadro de mando: Utilidad obtenida por indicador

En la Figura 79 se puede apreciar un cuadro de mando, el mismo que consta de 4 paneles, cada uno muestra información de la utilidad obtenida por indicador de los indicadores cuantitativos y cualitativos para la toma de decisiones. Los parámetros de entrada son: a) indicador y b) período académico.



Figura 77. Cuadro de mando: Resultados obtenidos por indicador<sup>67</sup>.

<sup>67</sup> Fuente: Elaboración propia



Figura 78. Cuadro de mando: Pesos obtenidos y por obtener de los indicadores<sup>68</sup>.



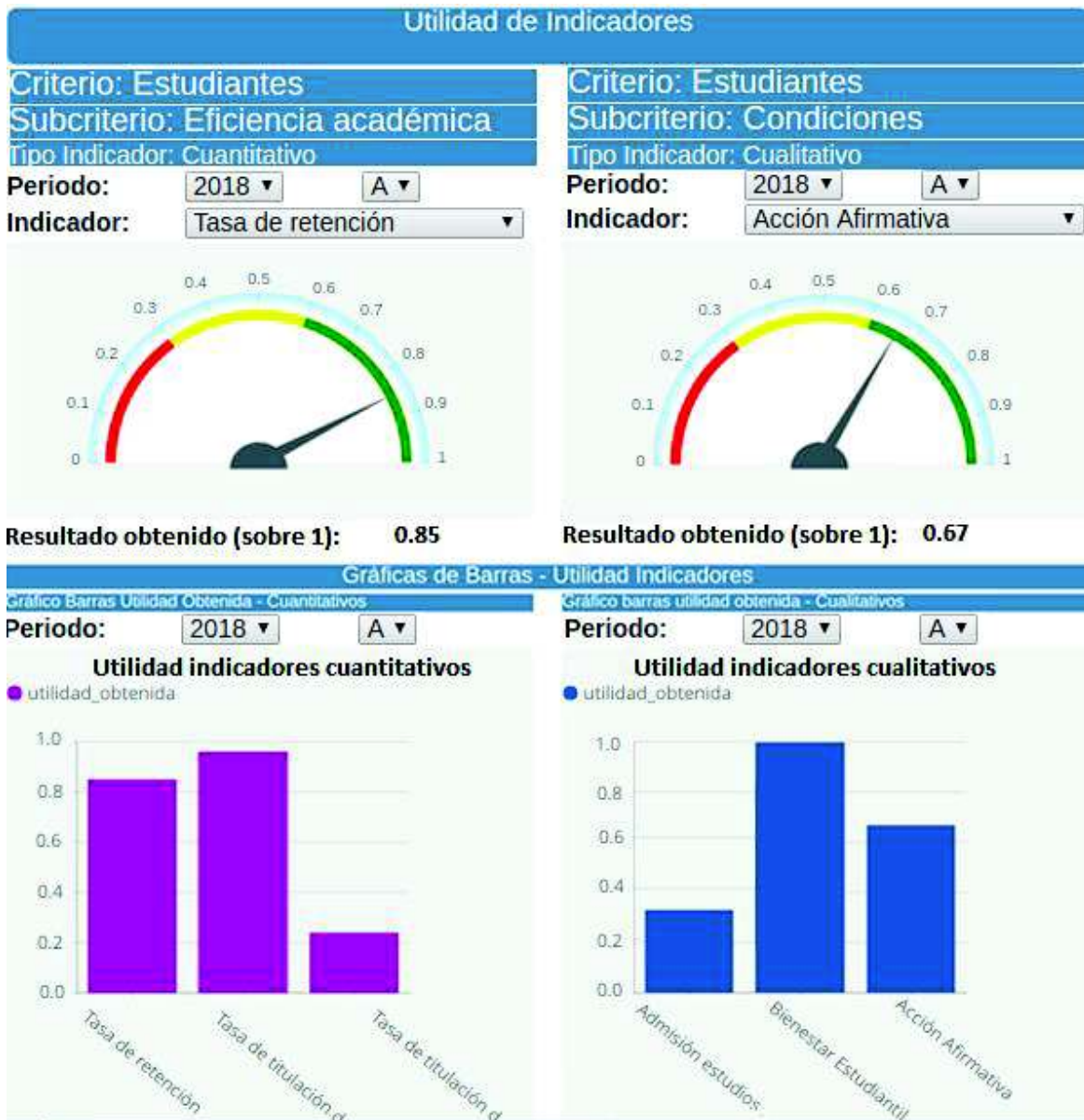


Figura 79. Cuadro de mando: Utilidad obtenida por indicador<sup>69</sup>.

### 2.9.3.10. Reporte del valor, peso y utilidad de los indicadores

El reporte detallado en la Figura 80 es de tipo *ad-hoc* y muestra al usuario información del valor, peso y utilidad de los indicadores. Los valores están agrupados por criterio y subcriterio al que corresponde cada indicador en un determinado periodo académico. La información es consumida del cubo.

<sup>68</sup> Fuente: Elaboración propia

<sup>69</sup> Fuente: Elaboración propia

| Año - Año           |                    |                                | 2017           |               |          | 2018           |               |          |
|---------------------|--------------------|--------------------------------|----------------|---------------|----------|----------------|---------------|----------|
| Semestre - Semestre |                    |                                | B              |               |          | A              |               |          |
| Nombre Criterio     | Nombre Subcriterio | Nombre Indicador               | Valor Obtenido | Peso Obtenido | Utilidad | Valor Obtenido | Peso Obtenido | Utilidad |
| Estudiantes         | Condiciones        | Acción Alternativa             | 0.67           | 1.01          | 0.67     | 0.97           | 1.01          | 0.67     |
|                     |                    | Admisión estudios de posgrado  | 0.33           | 0.5           | 0.33     | 0.33           | 0.5           | 0.33     |
|                     |                    | Bienestar Estudiantil          | 0.67           | 1.34          | 0.67     | 1              | 2             | 1        |
|                     | Eficacia Académica | Tasa de retención              | 66.54          | 2.03          | 0.81     | 62.08          | 2.12          | 0.85     |
|                     |                    | Tasa de titulación de posgrado | 62.45          | 2             | 1        | 54.73          | 1.92          | 0.96     |
|                     |                    | Tasa de titulación de pregrado | 42.78          | 1.05          | 0.42     | 37.77          | 0.6           | 0.24     |

Figura 80. Reporte del valor, peso y utilidad de los indicadores<sup>70</sup>.

### 2.9.3.11. Reporte del peso obtenido y por obtener de los indicadores

El reporte que se puede apreciar en la Figura 81 es de tipo *ad-hoc* y muestra al usuario información del peso que puede alcanzar, del peso obtenido y cuanto le faltaría obtener a cada indicador para llegar al valor ideal.

Los valores están agrupados por criterio y subcriterio al que corresponde cada indicador en un determinado periodo académico. La información es consumida del cubo.

### 2.9.3.12. Reporte de la utilidad obtenida por indicador

El reporte que se puede apreciar en la Figura 82 es de tipo *ad-hoc* y muestra al usuario información de la utilidad obtenida por indicador.

Los valores están agrupados por criterio y subcriterio al que corresponde cada indicador en un determinado periodo académico. La información es consumida del cubo.

<sup>70</sup> Fuente: Elaboración propia

|                 |                      | Año - Año                      |                |               | 2017             |               | 2018             |  |
|-----------------|----------------------|--------------------------------|----------------|---------------|------------------|---------------|------------------|--|
|                 |                      | Semestre - Semestre            |                |               | B                |               | A                |  |
| Nombre Criterio | Nombre Subcriterio   | Nombre Indicador               | Peso Indicador | Peso Obtenido | Diferencia Pesos | Peso Obtenido | Diferencia Pesos |  |
| Estudiantes     | Condiciones          | Acción Afirmativa              | 1.5            | 1.01          | 0.48             | 1.01          | 0.49             |  |
|                 |                      | Admisión estudios de posgrado  | 1.5            | 0.5           | 1                | 0.5           | 1                |  |
|                 |                      | Bienestar Estudiantil          | 2              | 1.34          | 0.66             | 2             | 0                |  |
|                 | Eficiencia Académica | Tasa de retención              | 2.5            | 1.03          | 0.47             | 1.12          | 0.38             |  |
|                 |                      | Tasa de titulación de posgrado | 2              | 2             | 0                | 1.92          | 0.08             |  |
|                 |                      | Tasa de titulación de pregrado | 2.5            | 1.05          | 1.45             | 0.8           | -1.9             |  |

Figura 81. Reporte del peso obtenido y por obtener de los indicadores<sup>71</sup>.

|                 |                      | Año - Año                      |                |          | 2017     |          | 2018     |  |
|-----------------|----------------------|--------------------------------|----------------|----------|----------|----------|----------|--|
|                 |                      | Semestre - Semestre            |                |          | B        |          | A        |  |
| Nombre Criterio | Nombre Subcriterio   | Nombre Indicador               | Peso Indicador | Utilidad | Utilidad | Utilidad | Utilidad |  |
| Estudiantes     | Condiciones          | Acción Afirmativa              | 1.5            | 0.67     | 0.67     |          |          |  |
|                 |                      | Admisión estudios de posgrado  | 1.5            | 0.33     | 0.33     |          |          |  |
|                 |                      | Bienestar Estudiantil          | 2              | 0.67     | 1        |          |          |  |
|                 | Eficiencia Académica | Tasa de retención              | 2.5            | 0.81     | 0.85     |          |          |  |
|                 |                      | Tasa de titulación de posgrado | 2              | 1        | 0.96     |          |          |  |
|                 |                      | Tasa de titulación de pregrado | 2.5            | 0.42     | 0.24     |          |          |  |

Figura 82. Reporte de la utilidad obtenida por indicador<sup>72</sup>.

<sup>71</sup> Fuente: Elaboración propia

<sup>72</sup> Fuente: Elaboración propia

## 2.9.4. Implementación del aplicativo web

Para la fase de implementación del aplicativo web se utiliza Laravel que es un *framework* de código abierto que permite desarrollar aplicaciones con PHP.

Para la creación de las interfaces web con CSS y JavaScript se utiliza Bootstrap. Gracias a este *framework* la interfaz del sitio web se adapta al tamaño del dispositivo en el que se visualice.

Para el almacenamiento de información, el aplicativo web utiliza PostgreSQL como motor de base de datos.

El desarrollo consiste en la implementación de las clases identificadas durante la fase del diseño conceptual y de las interfaces de usuario. En la Figura 83 se ilustra el diagrama de despliegue del aplicativo web para la visualización de los reportes detallados anteriormente. El código del aplicativo se detalla en el Anexo VI.

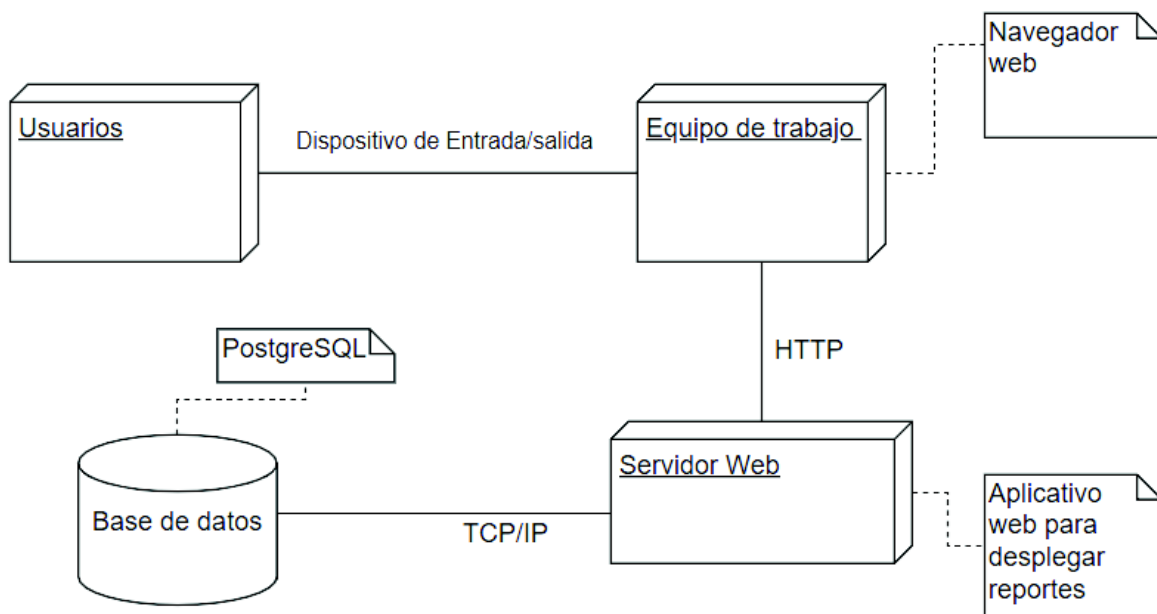


Figura 83. Diagrama de despliegue del aplicativo web.

## 2.10. Despliegue

El despliegue es la convergencia de tecnología, datos, y aplicaciones, es la integración de los componentes que se han ido desarrollando al transcurso del ciclo de vida del data mart, con el objetivo de que los usuarios puedan comenzar a sacar provecho [26].

En esta etapa los usuarios deben ser capacitados y preparados para poder integrar cada uno de los componentes de la solución de BI. En el Anexo VII se detalla el contenido, los objetivos específicos y los datos de los participantes de la capacitación técnica llevada a cabo con personal de la DGIP de la EPN.

El *data mart* para el monitoreo de indicadores de acreditación universitaria se implementó en un servidor de pruebas de la EPN. Las características del servidor son las siguientes:

Tabla 31. Características del servidor de pruebas.

| Software                 | Versión   |
|--------------------------|-----------|
| Ubuntu                   | 18.04.1   |
| Apache                   | 8.5       |
| PostgreSQL               | 10.5      |
| PHP                      | 7.2.10    |
| Java JDK                 | 1.8.0_181 |
| Pentaho Data Integration | 8.0       |
| Pentaho User Console     | 8.0       |

Antes de poner en producción el *data mart* en el servidor de pruebas de la EPN, se debe desarrollar pruebas alfa y pruebas beta.

### 2.10.1. Pruebas Alfa

Estas pruebas tienen la finalidad de verificar el correcto funcionamiento de los componentes del data mart. Estas pruebas incluyen probar la infraestructura, los procesos ETL y las aplicaciones de usuarios.

Las pruebas se realizan con una copia de la fuente de dato que está en producción, para este proyecto la fuente principal es la base de datos del sistema de gestión académica

SAEW. Una fuente secundaria es la base de datos propuesta para la gestión de indicadores cualitativos.

Al realizar estas pruebas se asegura la calidad de los datos y las aplicaciones de los usuarios finales. Estas pruebas pueden extenderse hasta que se asegure la calidad de la solución de BI.

### 2.10.2. Pruebas Beta

En este periodo de pruebas un determinado grupo de usuarios tiene acceso al data mart para comprobar la calidad e integridad de datos, la funcionalidad de las aplicaciones y para determinar si la solución cumple con los requerimientos especificados por la organización.

## 2.11. Pruebas

### 2.11.1. Pruebas de validación de la información

Las pruebas de validación de la información consisten en comparar los resultados de los indicadores de acreditación universitaria obtenidos por el data mart con los resultados oficiales que actualmente tiene la EPN.

### 2.11.2. Pruebas de funcionalidad

#### 2.11.2.1. Definición de los casos de pruebas

Tabla 32. Caso de prueba CP-01.

|                   |   |
|-------------------|---|
| Número de prueba: | CP-01   |
| Caso de prueba:   | Carga de <i>data mart</i>   |
| Propósito:        | Comprobar que la carga del <i>data mart</i> se realiza satisfactoriamente.                                    |
| Ambiente:         | <i>Data mart</i>  |
| Pasos:            | <ol style="list-style-type: none"> <li>1. Ejecutar ETL</li> <li>2. Cargar datos de las dimensiones</li> </ol> |

|                       |  |
|-----------------------|--|
|                       | <ol style="list-style-type: none"> <li>3. Cargar datos de la tabla de hechos</li> <li>4. Abrir base de datos</li> <li>5. Verificar datos consistentes</li> </ol> |
| Resultados esperados: | <p>Se realiza satisfactoriamente la carga del <i>data mart</i>.</p> <p>La carga del <i>data mart</i> está en un rango de tiempo aceptable.</p>                   |

Tabla 33. Caso de prueba CP-02.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-02  |
| Caso de prueba:       | Ejecución del cuadro de mando de los resultados obtenidos por criterio.  |
| Propósito:            | Verificar el comportamiento del cuadro mando.  |
| Ambiente:             | CDE ( <i>Community Dashboard Editor</i> )  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar al editor de cuadros de mando</li> <li>2. Ejecutar el cuadro de mando</li> <li>3. Ingresar los parámetros para generar los reportes</li> <li>4. Interactuar con las gráficas</li> </ol> |
| Resultados esperados: | <p>Presentación de datos en gráficas y la interacción de éstas.</p> <p>El tiempo de generación del cuadro de mando es aceptable.</p>   |

Tabla 34. Caso de prueba CP-03.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-03  |
| Caso de prueba:       | Ejecución del cuadro de mando de los resultados obtenidos por subcriterio.   |
| Propósito:            | Verificar el comportamiento del cuadro mando.  |
| Ambiente:             | CDE ( <i>Community Dashboard Editor</i> )  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar al editor de cuadros de mando</li> <li>2. Ejecutar el cuadro de mando</li> <li>3. Ingresar los parámetros para generar los reportes</li> <li>4. Interactuar con las gráficas</li> </ol> |
| Resultados esperados: | <p>Presentación de datos en gráficas y la interacción de éstas.</p> <p>El tiempo de generación del cuadro de mando es aceptable.</p>   |

Tabla 35. Caso de prueba CP-04.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-04  |
| Caso de prueba:       | Ejecución del cuadro de mando de los resultados obtenidos por indicador.   |
| Propósito:            | Verificar el comportamiento del cuadro mando.  |
| Ambiente:             | CDE ( <i>Community Dashboard Editor</i> )  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar al editor de cuadros de mando</li> <li>2. Ejecutar el cuadro de mando</li> <li>3. Ingresar los parámetros para generar los reportes</li> <li>4. Interactuar con las gráficas</li> </ol> |
| Resultados esperados: | <p>Presentación de datos en gráficas y la interacción de éstas.</p> <p>El tiempo de generación del cuadro de mando es aceptable.</p>   |

Tabla 36. Caso de prueba CP-05.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-05  |
| Caso de prueba:       | Ejecución del cuadro de mando de los pesos obtenidos y por obtener de los indicadores.   |
| Propósito:            | Verificar el comportamiento del cuadro mando.  |
| Ambiente:             | CDE ( <i>Community Dashboard Editor</i> )  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar al editor de cuadros de mando</li> <li>2. Ejecutar el cuadro de mando</li> <li>3. Ingresar los parámetros para generar los reportes</li> <li>4. Interactuar con las gráficas</li> </ol> |
| Resultados esperados: | <p>Presentación de datos en gráficas y la interacción de éstas.</p> <p>El tiempo de generación del cuadro de mando es aceptable.</p>   |

Tabla 37. Caso de prueba CP-06.

|                   |  |
|-------------------|--|
| Número de prueba: | CP-06  |
| Caso de prueba:   | Ejecución del cuadro de mando de la utilidad obtenida por indicador. |
| Propósito:        | Verificar el comportamiento del cuadro mando.                        |
| Ambiente:         | CDE ( <i>Community Dashboard Editor</i> )                            |



|                       |  |
|-----------------------|--|
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar al editor de cuadros de mando</li> <li>2. Ejecutar el cuadro de mando</li> <li>3. Ingresar los parámetros para generar los reportes</li> <li>4. Interactuar con las gráficas</li> </ol> |
| Resultados esperados: | <p>Presentación de datos en gráficas y la interacción de éstas.</p> <p>El tiempo de generación del cuadro de mando es aceptable.</p>   |

Tabla 38. Caso de prueba CP-07.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-07  |
| Caso de prueba:       | Ejecución de la vista de análisis de datos del peso obtenido de los criterios.   |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.  |
| Ambiente:             | Saiku Analytics  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol>  |
| Resultados esperados: | <p>Presentación de datos en tablas pivotantes y la interacción de éstas.</p> <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |

Tabla 39. Caso de prueba CP-08.

|                       |   |
|-----------------------|---|
| Número de prueba:     | CP-08   |
| Caso de prueba:       | Ejecución de la vista de análisis de datos del peso obtenido y por obtener de los criterios.  |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.   |
| Ambiente:             | Saiku Analytics   |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol> |
| Resultados esperados: | Presentación de datos en tablas pivotantes y la interacción de éstas.   |

|  |   |
|--|---|
|  | <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |
|--|---|

Tabla 40. Caso de prueba CP-09.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-09  |
| Caso de prueba:       | Ejecución de la vista de análisis de datos del peso obtenido de los subcriterios.  |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.  |
| Ambiente:             | Saiku Analytics  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol>  |
| Resultados esperados: | <p>Presentación de datos en tablas pivotantes y la interacción de éstas.</p> <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |

Tabla 41. Caso de prueba CP-10.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-10  |
| Caso de prueba:       | Ejecución de la vista de análisis de datos del peso obtenido y por obtener de los subcriterios.  |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.  |
| Ambiente:             | Saiku Analytics  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol>  |
| Resultados esperados: | <p>Presentación de datos en tablas pivotantes y la interacción de éstas.</p> <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |

Tabla 42. Caso de prueba CP-11.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-11  |
| Caso de prueba:       | Ejecución de la vista de análisis de datos del valor, peso y utilidad de los indicadores.  |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.  |
| Ambiente:             | Saiku Analytics  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol>  |
| Resultados esperados: | <p>Presentación de datos en tablas pivotantes y la interacción de éstas.</p> <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |

Tabla 43. Caso de prueba CP-12.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-12  |
| Caso de prueba:       | Ejecución de la vista de análisis de datos del peso obtenido y por obtener de los indicadores.   |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.  |
| Ambiente:             | Saiku Analytics  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol>  |
| Resultados esperados: | <p>Presentación de datos en tablas pivotantes y la interacción de éstas.</p> <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |

Tabla 44. Caso de prueba CP-13.

|                       |  |
|-----------------------|--|
| Número de prueba:     | CP-13  |
| Caso de prueba:       | Ejecución de la vista de análisis de datos de la utilidad de los indicadores.  |
| Propósito:            | Verificar el comportamiento de la vista de análisis de datos.  |
| Ambiente:             | Saiku Analytics  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Ingresar a Saiku Analytics</li> <li>2. Ejecutar la vista de análisis de datos</li> <li>3. Interactuar con las gráficas</li> </ol>  |
| Resultados esperados: | <p>Presentación de datos en tablas pivotantes y la interacción de éstas.</p> <p>Presentación de datos en gráficos estadísticos y la interacción de éstos.</p> <p>El tiempo de generación de la vista es aceptable.</p> |

### 2.11.3. Pruebas de integridad

Estas pruebas verifican que el *data mart* y el sistema de gestión académica SAEW trabajen conjuntamente sin problemas. Estas pruebas son importantes ya que se debe comprobar que los datos que se obtienen de la base de datos del SAEW son cargados al *data mart* sin inconvenientes.

La definición del caso de prueba se detalla en la siguiente tabla:

Tabla 45. Caso de prueba PI-01.

|                   |   |
|-------------------|---|
| Número de prueba: | PI-01   |
| Caso de prueba:   | Comprobar que la carga de datos que se obtienen de la base de datos del SAEW hacia el <i>data mart</i> se realizó de manera correcta. |
| Propósito:        | Verificar la carga de datos desde las fuentes hacia el <i>data mart</i> .   |
| Pasos:            | <ol style="list-style-type: none"> <li>1. Sentencias SQL en la base de datos origen</li> </ol>  |

|                       |  |
|-----------------------|--|
|                       | <ol style="list-style-type: none"> <li>2. Sentencias SQL en la base de datos que contiene el cubo</li> <li>3. Comparación de los resultados</li> </ol> |
| Resultados esperados: | Los datos cargados desde las fuentes hacia el <i>data mart</i> son correctos.  |

Tabla 46. Caso de prueba PI-02.

|                       |   |
|-----------------------|---|
| Número de prueba:     | PI-02   |
| Caso de prueba:       | Comprobar que la estructura de la base de datos que contiene el cubo sea la correcta según el modelo.                   |
| Propósito:            | Verificar las estructura de la base de datos que contiene el cubo.  |
| Pasos:                | <ol style="list-style-type: none"> <li>1. Comparar el modelo diseñado con la estructura de la base de datos.</li> </ol> |
| Resultados esperados: | La base de datos funciona según lo diseñado.  |

## CAPÍTULO 3

### 3. RESULTADOS Y DISCUSIÓN

#### 3.1. Resultados de las pruebas de validación

Las pruebas de validación se realizaron comparando los resultados del data mart con la información del último reporte solicitado por las autoridades de la EPN para el monitoreo de los indicadores de acreditación universitaria del periodo académico 2018 A.

Tabla 47. Reporte de indicadores EPN.

| Indicador                      | Valores obtenidos manualmente |
|--------------------------------|-------------------------------|
| Tasa de retención de grado     | 62,67                         |
| Tasa de titulación de grado    | 44,51                         |
| Tasa de titulación de posgrado | 57,43                         |

| Año - Año           |                      |                                | 2017           |               |          | 2018           |               |          |
|---------------------|----------------------|--------------------------------|----------------|---------------|----------|----------------|---------------|----------|
| Semestre - Semestre |                      |                                | B              |               |          | A              |               |          |
| Nombre Criterio     | Nombre Subcriterio   | Nombre Indicador               | Valor Obtenido | Peso Obtenido | Utilidad | Valor Obtenido | Peso Obtenido | Utilidad |
| Estudiantes         | Condiciones          | Acción Afirmativa              | 0,67           | 1,01          | 0,67     | 0,67           | 1,01          | 0,67     |
|                     |                      | Admisión estudios de posgrado  | 0,33           | 0,5           | 0,33     | 0,33           | 0,5           | 0,33     |
|                     |                      | Bienestar Estudiantil          | 0,67           | 1,34          | 0,67     | 1              | 2             | 1        |
|                     | Eficiencia Académica | Tasa de retención              | 60,54          | 2,03          | 0,81     | 62,08          | 2,12          | 0,85     |
|                     |                      | Tasa de titulación de posgrado | 62,45          | 2             | 1        | 54,73          | 1,92          | 0,96     |
|                     |                      | Tasa de titulación de pregrado | 42,76          | 1,05          | 0,42     | 37,77          | 0,6           | 0,24     |

Figura 84. Reporte obtenido con datos del data mart<sup>73</sup>.

Al comparar los resultados generados por la EPN detallados en Tabla 47 y los resultados del data mart detallados en la Figura 84, se puede apreciar que los valores obtenidos de los indicadores se aproximan en gran manera, por lo tanto, se puede dar como válido el contenido del data mart. Los valores no son iguales ya que el reporte generado por la EPN se lo realizó sobre información que se encuentra dispersa y sin realizar ninguna depuración de las fuentes de información.

<sup>73</sup> Fuente: Elaboración propia

### 3.2. Resultados de las pruebas de funcionalidad

Los resultados de las pruebas funcionales se detallan en la siguiente tabla:

Tabla 48. Resultados pruebas funcionales.

| Caso de prueba | Fecha ejecución | Resultado |
|----------------|-----------------|-----------|
| CP-01          | 2018-08-27      | Si cumple |
| CP-02          | 2018-08-27      | Si cumple |
| CP-03          | 2018-08-27      | Si cumple |
| CP-04          | 2018-08-27      | Si cumple |
| CP-05          | 2018-08-27      | Si cumple |
| CP-06          | 2018-08-27      | Si cumple |
| CP-07          | 2018-08-27      | Si cumple |
| CP-08          | 2018-08-27      | Si cumple |
| CP-09          | 2018-08-27      | Si cumple |
| CP-10          | 2018-08-27      | Si cumple |
| CP-11          | 2018-08-27      | Si cumple |
| CP-12          | 2018-08-27      | Si cumple |
| CP-13          | 2018-08-27      | Si cumple |

Para corroborar el resultado de las pruebas, se puede observar los cuadros de mando y vistas de análisis de datos en la sección 2.9.3 de este documento.

### 3.3. Resultados de las pruebas de integridad

Al realizar las pruebas de integridad se verificó que los datos del sistema de gestión académica SAEW se cargan de manera correcta en el *data mart*.

También se verificó que la estructura de la base de datos que contiene el cubo esté acorde al modelo diseñado.

Los resultados de las pruebas de integridad se detallan en la Tabla 49.

Tabla 49. Resultados pruebas integridad.

| Caso de prueba | Fecha ejecución | Resultado |
|----------------|-----------------|-----------|
| PI-01          | 2018-08-30      | Si cumple |
| PI-02          | 2018-08-30      | Si cumple |

La Figura 85 ilustra el diagrama físico del cubo OLAP implementado. La Figura 86 muestra el modelo diseñado para la elaboración de la base de datos que contiene el cubo. Al comparar las dos figuras se puede observar que la base de datos que contiene el cubo está acorde al modelo diseñado.

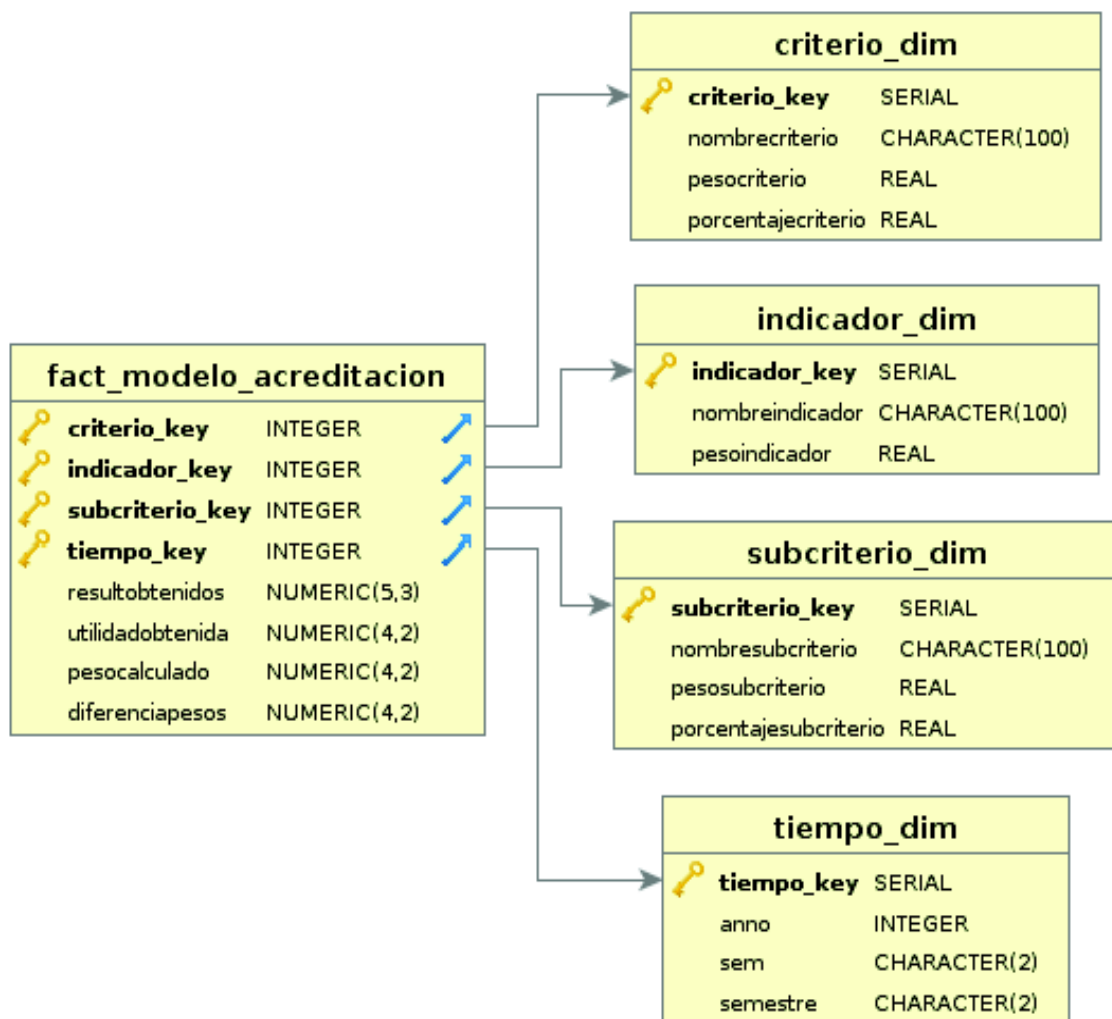


Figura 85. Diagrama del cubo OLAP "DATA MART CEAACES".



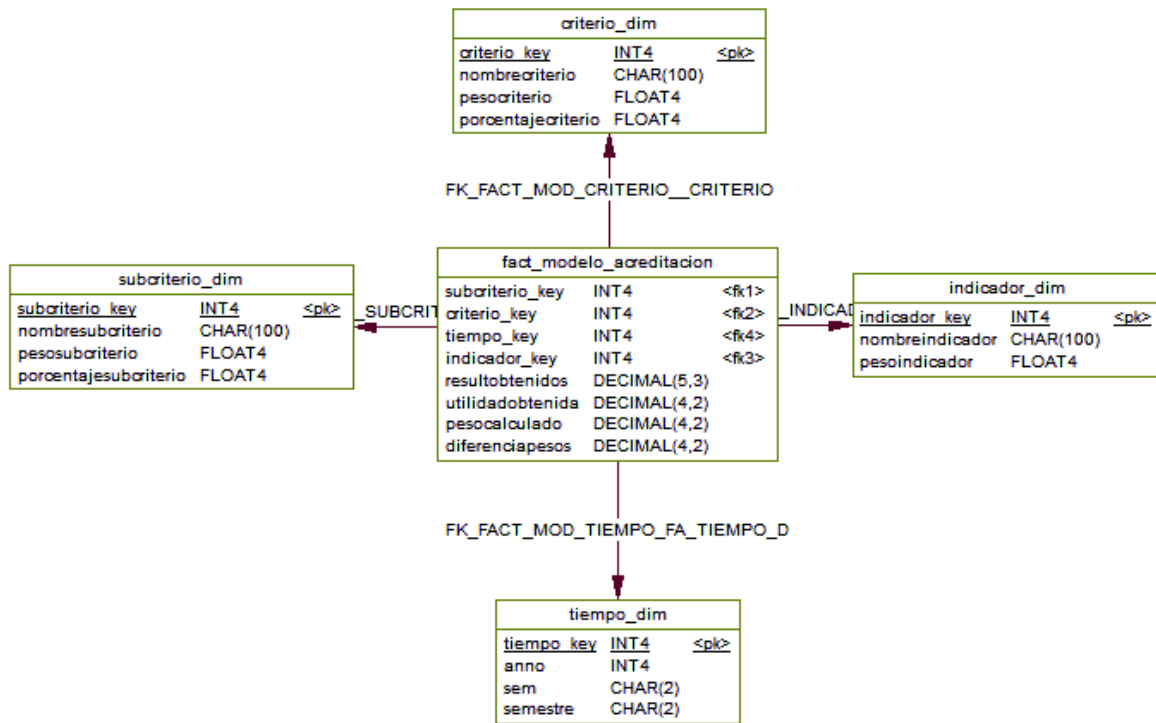


Figura 86. Diseño del modelo de base de datos para crear el cubo.

## CAPÍTULO 4

### 4. CONCLUSIONES Y RECOMENDACIONES

#### 4.1. Conclusiones

Lo expuesto a lo largo de este trabajo permite arribar a las siguientes conclusiones:

- Es posible desarrollar un *data mart* para el manejo de indicadores de acreditación universitaria aplicando la metodología de Kimball y con Pentaho como herramienta de BI de software libre, cumpliendo de esta manera la disposición del CES en su Reglamento de Sanciones, Art. 11, numeral 9, que establece el uso de programas informáticos de software libre.
- El *data mart* mejora la obtención de información para su rápido entendimiento y gestión, logrando que los datos relacionados a la acreditación universitaria ya no se encuentren dispersos y de esta manera permitir un análisis integrado para los procesos de toma de decisiones.
- Los procesos ETL son cruciales en la integración de datos, por lo tanto es primordial contar con fuentes de información confiables que aseguren la calidad de datos y de esta manera obtener información útil y valiosa.
- Se logró desarrollar e implementar un componente de visualización de datos, permitiendo monitorear los indicadores de acreditación universitaria en el tiempo. El componente de visualización permite la generación de reportes de manera sencilla e intuitiva al usuario final, y en márgenes de tiempo menores a los actuales.
- El uso de cubos OLAP y cuadros de mando produce una mayor agilidad y rapidez en el manejo de los datos del *data mart*. Además, las opciones de imprimir o exportar los reportes generados a formato Excel o PDF permiten una mayor portabilidad por parte del usuario.

## 4.2. Recomendaciones

Lo expuesto a lo largo de este trabajo permite arribar a las siguientes recomendaciones:

- La institución debería contar con las fuentes de datos digitalizadas para extraer la información necesaria para el monitoreo de todos los criterios establecidos en el modelo propuesto por el CEAACES.
- Se recomienda mantener una información continua con los usuarios finales que usan la solución de BI, debido a que pueden surgir nuevos requerimientos que no se identificaron en la fase de análisis de requerimientos.
- Antes de comenzar un proyecto de almacén de datos o *data mart* es importante conocer si existe la demanda y de donde proviene. Además, se debe contar con el apoyo de alta gerencia o de algún nivel gerencial influyente e interesado en sacar adelante el proyecto de BI, ya que el patrocinador es el intermediario para obtener la aprobación y los recursos necesarios para el desarrollo. Si no se tiene un patrocinador consistente y usuarios interesados es mejor suspender el proyecto.
- La etapa de ETL implica la mayor parte de esfuerzo en la elaboración del *data mart*, es por ello que se debe tener cuidado al momento de establecer los tiempos en el cronograma de actividades.

## REFERENCIAS BIBLIOGRÁFICAS

- [1] R. Kimball, "The Future Is Bright", © Kimball Group, [En línea]. Disponible en: <http://www.kimballgroup.com/2015/12/design-tip-180-the-future-is-bright/>. [Último acceso: 04 Julio 2017].
- [2] M. Parzinger y M. Frolick, "Creating Competitive Advantage through Data Warehousing", *The Executive's Journal*, vol. 17, nº 4, pp. 10-15, 2001.
- [3] B. Waldo, "Decision Support and Data Warehousing Tools Boost Competitive", *Nursing Economics*, vol. 16, nº 2, pp. 91-93, 1998.
- [4] L. Sang, H. Soongoo, y K. Pairin, "Impact of Data Warehousing on Organizational Performance of Retailing Firms", *International Journal of Information Technology & Decision Making*, vol. 3, nº 1, pp. 61-79, 2004.
- [5] L. Brynne, A. Bresell, y N. Sjögren, "Effective visualization of integrated knowledge and data to enable informed decisions in drug development and translational medicine", *Journal of translational medicine*, vol. 11, nº 1, p. 250, 2013.
- [6] T. Bucher y A. Gericke, "Process-centric business intelligence", *Business Process Management Journal*, vol. 15, nº 3, pp. 408-429, 2009.
- [7] CEAACES, "Proceso de evaluación institucional", [En línea]. Disponible en: <http://www.ceaaces.gob.ec/>. [Último acceso: 07 Mayo 2018].
- [8] J. Rojas, "Reforma universitaria en el Ecuador. Etapa de transición", *Innovación Educativa*, vol. 11, nº 57, pp. 59-67, 2011.
- [9] CEAACES, *Modelo de evaluación institucional de universidades y escuelas politécnicas*. Ecuador, 2015.
- [10] Asamblea Nacional, *Ley Orgánica de Educación Superior LOES*, Ecuador, 2010
- [11] A. Peña, *Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones*. México: Instituto Politécnico Nacional, 2006.
- [12] R. Bernabeu, *Hefesto—Data warehousing: Investigación y sistematización de conceptos-Hefesto: Metodología para la construcción de un data warehouse*. Argentina, 2010.
- [13] S. Sharma, J. Sharma y A. Devi, "Corporate Social Responsibility: The Key Role of Human Resource Management", *Sayco UK Business Intelligence Journal*, vol. 2, nº 1, pp. 205-213, 2009.

- [14] L. Maureen, *La gestión del valor de la cartera de clientes y su efecto en el valor global de la empresa*. Madrid: Universidad Complutense de Madrid, Servicio de Publicaciones, 2007.
- [15] A. Rosado y D. Rico, "Inteligencia de negocios: estado del arte", *Scientia et Technica*, vol. 1, nº 44, pp. 321-326, 2010.
- [16] Oracle, "Oracle business intelligence 12c – features", [En línea]. Disponible en: <https://www.oracle.com/es/solutions/businessanalytics/business-intelligence/features.html/>. [Último acceso: 08 Agosto 2018].
- [17] Microstrategy Business Intelligence, "Enterprise analytics and mobility", [En línea]. Disponible en: <https://www.microstrategy.com/es/>. [Último acceso: 08 Agosto 2018].
- [18] C. M. Iñigo, "Estudio comparativo de las herramientas de business intelligence: Empoderando el criterio de selección a las pymes", ResearchGate, 2013.
- [19] Hitachi Vantara, "Pentaho Business Analytics Platform", [En línea]. Disponible en: <https://www.hitachivantara.com/en-us/pdf/datasheet/pentaho-business-analytics-platform-datasheet.pdf>. [Último acceso: 08 Agosto 2018].
- [20] H. Flores, "Inteligencia de negocios como apoyo a la toma de decisiones en la gerencia", *Vínculos*, vol. 9, nº 2, pp. 11-23, 2012.
- [21] J. Cano, *Business Intelligence: Competir con Información*. España: Universidad Ramon Llull, ESADE Business School, 2007.
- [22] F. Morales, R. Cueva, y J. Martínez, "Procesamiento analítico con minería de datos", *Revista Iberoamericana de las Ciencias Computacionales e Informática*, vol. 5, nº 9, pp. 22-43, 2016.
- [23] C. Hurtado y A. Mendelzon, "OLAP Dimension Constraints", *Proc. PODS*, pp. 169-179. 2002.
- [24] M. Pasumansky et al., "Data cube script development and debugging systems and methodologies", U.S. Patente 7694278, 06 de abril, 2010.
- [25] G. Satyanarayana, R. Abandi, M. Poorna y S. Reddy, "Data warehousing, data mining, olap and oltp technologies are essential elements to support decision-making process in industries", *International Journal on Computer Science and Engineering*, vol. 2, nº 9, pp. 2865-2873, 2010.
- [26] R. Kimball y M. Ross, *The Data Warehouse Lifecycle Toolkit*. Indianapolis: Wiley, 2013.

- [27] W. H. Inmon, *Building the data warehouse*. Indianapolis: Wiley, 2011.
- [28] M. M. Martínez, *Analítica Web para empresas*. Barcelona: UOC, 2011.
- [29] CEAACES, *Instrumentos de cuantificación del desempeño institucional*. Ecuador, 2015.
- [30] Consejo de Educación Superior, *Reglamento de sanciones*, Ecuador, 2015
- [31] D. Schwabe y G. Rossi, "The Object-Oriented Hypermedia Design Model (OOHDM)", *Communications of the ACM*, vol. 38, nº 8, pp. 45-46, 1995.
- [32] J. Molina, M. Zea, M. Contento y F. García, "Comparación de metodologías en aplicaciones web", *3C Tecnología: glosas de innovación aplicadas a la pyme*, vol. 7, nº 1, pp. 1-19, 2018.

## ANEXOS

### ANEXO I: RESOLUCIÓN N°. 024-CEAACES-SO-06-2017



**Que** el Pleno del CEAACES analizó la propuesta de modificación de la escala de valoración de los indicadores cualitativos de los modelos de evaluación de este Consejo, considerando pertinente su aprobación; y,

En ejercicio de las atribuciones que le confiere la Constitución de la República del Ecuador, la Ley Orgánica de Educación Superior, el Reglamento Interno del CEAACES y el Reglamento de Evaluación, Acreditación y Categorización de Carreras de las Instituciones de Educación Superior,

#### RESUELVE:

**Artículo único.-** Aprobar la modificación de la escala de valoración de los indicadores cualitativos de los modelos de evaluación del CEAACES, de conformidad con el siguiente detalle:

1. **Deficiente (0):** No alcanza el estándar evidenciando debilidades estructurales que comprometen la consecución de los objetivos y/o la información presentada no permite el análisis.
2. **Poco satisfactorio (0.35):** No alcanza el estándar evidenciando debilidades estructurales que comprometen la consecución de los objetivos, sin embargo, existen procesos viables a ser implementados.
3. **Cuasi satisfactorio (0.7):** Presenta debilidades no estructurales que pueden ser solventadas a través de la consolidación o mejora de los procesos ya implementados.
4. **Satisfactorio (1):** Alcanza el estándar.

## ANEXO II: ENTREVISTAS

|   |  |
|---|--|
| <b>Entrevistados:</b> Ing. Geovanna Chela<br>Ing. Tania Gualli<br><b>Entrevistador:</b> Leonardo Alexander Ortiz<br>Yumisaca  | <b>Fecha:</b> 26-03-2018<br><b>Área:</b> DGIP - Tecnología |
| <b>Objetivo:</b> Conocer las dificultades en la integración, generación y análisis de los resultados de los indicadores de acreditación universitaria a fin de proponer una solución de <i>business intelligence</i> para la toma de decisiones.  |  |
| <b>Preguntas y respuestas:</b><br><br><b>1. ¿Considera eficiente el proceso que lleva la Escuela Politécnica Nacional en los procesos de evaluación institucional?</b><br><br>El proceso de evaluación institucional no es eficiente.<br><br><b>2. ¿El Sistema Académico bajo que Gestor de Base de Datos funciona?</b><br><br>El sistema académico SAEW funciona bajo el gestor de base de datos SQL Server.<br><br><b>3. ¿Qué herramienta utiliza para el análisis de datos de los indicadores de acreditación institucional?</b><br><br>Qlik.- se utiliza esta herramienta para reportería descriptiva. No tiene todos los indicadores de acreditación universitaria.<br><br><b>4. ¿Si usted necesita realizar consultas y/o reportes de la información de los indicadores de acreditación institucional, tiene alguna complicación o dificultad? Si la respuesta es afirmativa, detalle las complicaciones o dificultades.</b><br><br>Las complicaciones o dificultades son las siguientes:<br>a) Calidad de datos<br>b) Fuentes dispersas<br>c) No disponibilidad de datos requeridos<br>d) Afinamiento de criterios<br>e) No hay un ente estratégico que gestione la información<br>f) No hay gobernanza de datos<br><br><b>5. ¿Cuáles son las quejas constantes en cuanto a la solicitud de reportes?</b><br><br>Las quejas son en relación a la calidad de datos y al tiempo de respuesta.<br><br><b>6. ¿Cuentan con herramientas para generar reportes de manera que la información pueda ser procesada de diferentes maneras?</b> |  |



Se cuenta con la herramienta Qlik que permite procesar la información de diferentes maneras permitiendo realizar reportería descriptiva.

**7. ¿Cuál es el tiempo promedio que les lleva la elaboración de un reporte de indicadores de acreditación institucional solicitado por las autoridades?**

La elaboración de un reporte depende del nivel de complejidad y disponibilidad de la información.

- a) Baja complejidad.- de 1 a 2 días
- b) Mediana complejidad.- 1 semana
- c) Alta complejidad.- 2 semanas

**8. ¿Cuáles son los reportes referentes a indicadores de acreditación institucional que se solicitan con más frecuencia?**

Los reportes referentes a indicadores de acreditación institucional son: la tasa de retención de grado, la tasa de titulación de grado y la tasa de titulación de posgrado. Estos reportes se solicitan 1 o 2 veces al mes.

**9. ¿Con qué fuentes de información cuentan (Sistemas operacionales)?**

SAEW.- Sistema de gestión académica.  
SII.- Contiene módulos de apoyo (talento humano, alumni).  
Alfresco.- Sistema de gestión documental.

**10. ¿Qué le parece la posibilidad de implementar un Data Mart?**

La implementación de un *data mart* permitiría centralizar la información, evitando tener fuentes dispersas. Por lo tanto implementar esta solución de *business intelligence* es una necesidad que actualmente tiene la institución.

**ANEXO III: SCRIPT BASE DE DATOS PARA INDICADORES CUALITATIVOS (EN EL CD)**

El archivo que contiene el detalle del Script se llama “Script\_evaluarCualitativos” y se encuentra en la carpeta “DB Cualitativos” en el CD “Data mart Anexos”.

## ANEXO IV: ARTÍCULO 9 DEL REGLAMENTO DE SANCIONES EXPEDIDO POR EL CES

---

### TÍTULO III DE LAS INFRACCIONES Y LAS SANCIONES

#### CAPÍTULO I DE LAS INFRACCIONES

**Artículo 10.- De las infracciones.-** Las infracciones son los actos u omisiones establecidas como tales por el incumplimiento de las obligaciones prescritas en la LOES, su Reglamento General y demás normativa aplicable expedida por el CES.

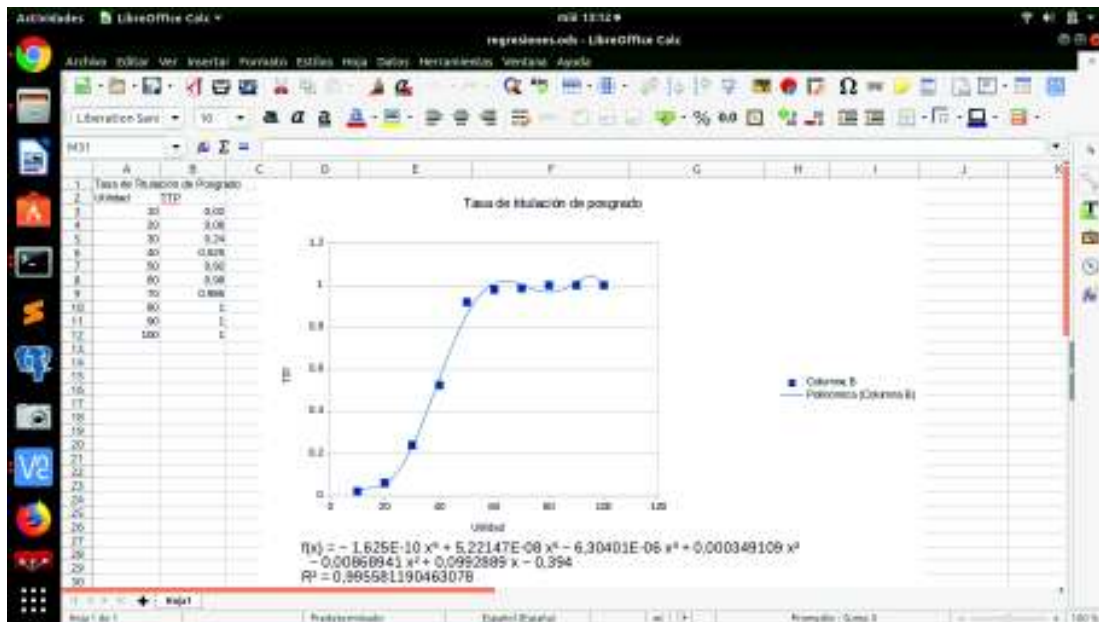
**Artículo 11.- Infracciones leves.-** Son infracciones leves:

1. Remitir extemporáneamente o no remitir anualmente los estados financieros auditados de las IES a la Secretaría de Educación Superior, Ciencia y Tecnología e Innovación (SENESCYT);
2. Abstenerse de publicar en el portal electrónico de las IES, las remuneraciones de sus autoridades, profesores, investigadores, servidores y trabajadores;
3. Entregar extemporáneamente o no entregar a la SENESCYT, el reporte final de los proyectos de investigación;
4. Remitir fuera del tiempo otorgado para el efecto o no remitir la información solicitada por las Coordinaciones, Procuraduría o Secretaría General del CES y del CEAACES;
5. Inhibirse de instrumentar el sistema de seguimiento a las y los graduados;
6. Abstenerse de desarrollar e integrar sistemas interconectados de bibliotecas;
7. Entregar extemporáneamente o no entregar los trabajos de titulación digitalizados al Sistema Nacional de Información de la Educación Superior (SNIESE);
8. Abstenerse de publicar en el portal electrónico de las IES los aranceles vigentes; y,
9. Incumplir la obligación de incorporar en las IES el uso de programas informáticos de software libre.

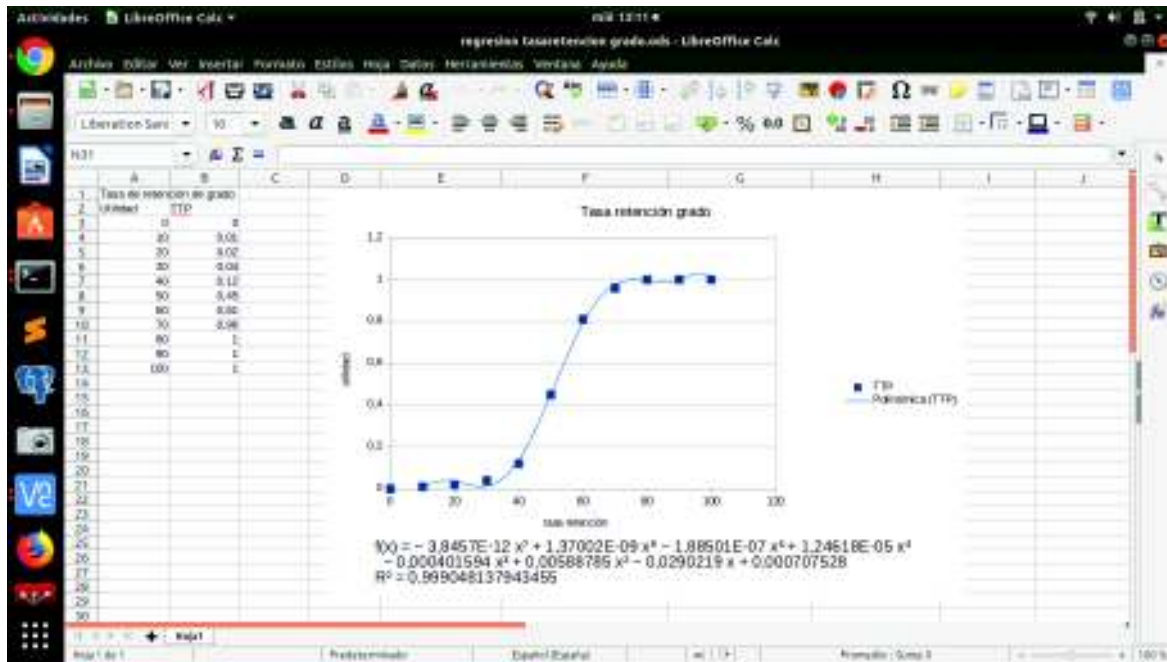
## ANEXO V: ANÁLISIS DE LAS REGRESIONES POLINÓMICAS DESARROLLADO EN HOJAS DE CÁLCULO DEL LIBREOFFICE

Los archivos LibreOffice de las regresiones polinómicas se encuentran almacenados en la carpeta “Regresiones Polinómicas” en el CD “Data mart Anexos”.

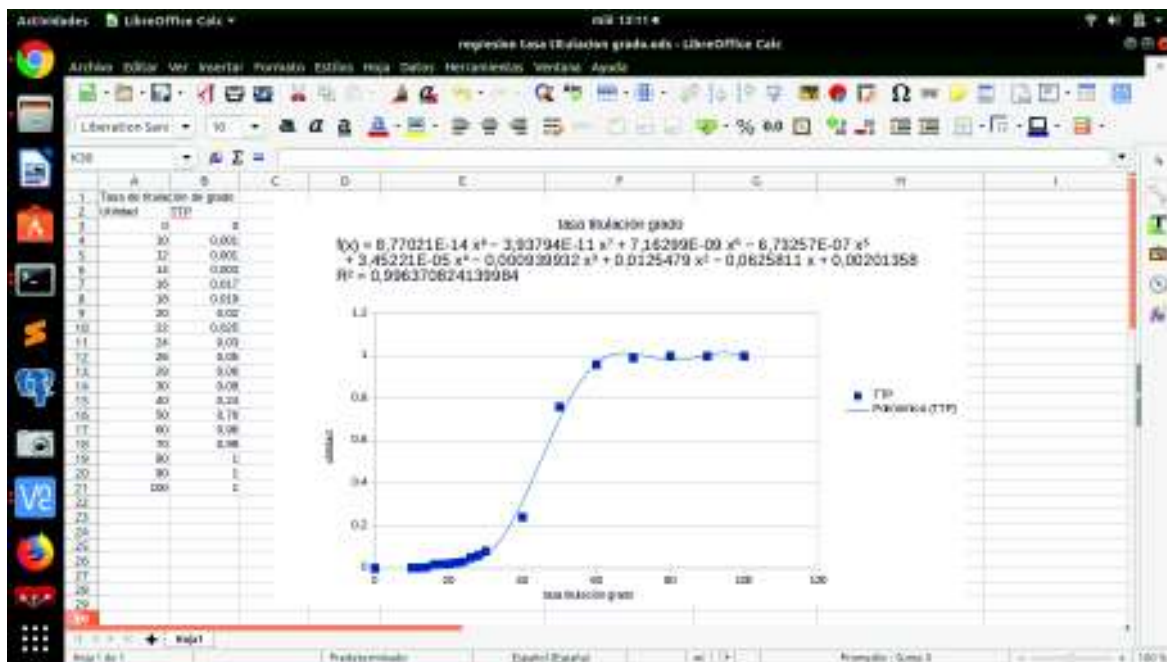
Regresión polinómica para obtener la función de utilidad del indicador tasa de titulación de posgrado.



Regresión polinómica para obtener la función de utilidad del indicador tasa de retención de grado.



Regresión polinómica para obtener la función de utilidad del indicador tasa de titulación de grado.



## ANEXO VI: CÓDIGO DEL APLICATIVO PARA DESPLEGAR REPORTES (EN EL CD)

EL código de la aplicación se encuentra en la carpeta “Aplicativo Reportes” en el CD “Data mart Anexos”.

## ANEXO VII: CAPACITACIÓN TÉCNICA

### Participantes:

| Nombre                              | Cargo  | Departamento |
|-------------------------------------|--|--------------|
| Geovanna Marisol Chela Tipantaxi    | Especialista de tecnologías                                    | DGIP-EPN     |
| Tania Guadalupe Gualli Culqui       | Analítica de datos   | DGIP-EPN     |
| María Daniela Gutierrez Castillo    | Especialista de tecnologías de la información y comunicaciones | DGIP-EPN     |
| Alexander Wladimir Pinchao Fuelagan | Apoyo analítica de datos                                       | DGIP-EPN     |

### Capacitación técnica

| Contenido            | Objetivos específicos   |
|----------------------|---|
| Diseño dimensional   | <ul style="list-style-type: none"><li>• Granularidad</li><li>• Fuentes de datos</li><li>• Modelo estrella</li></ul>   |
| Arquitectura técnica | <ul style="list-style-type: none"><li>• Motores de base de datos de las fuentes</li><li>• Prototipo de base de datos para el manejo de indicadores cualitativos</li><li>• Motor de base de datos del <i>data mart</i></li><li>• Herramientas <i>Open Source</i> de BI de la <i>suite</i> de Pentaho</li></ul> |
| ETL                  | <ul style="list-style-type: none"><li>• Entrada - salida de datos. Fuentes de datos</li></ul>   |

|                                    |   |
|------------------------------------|---|
|                                    | <ul style="list-style-type: none"> <li>• Funciones de utilidad mediante el modelo matemático de regresión polinomial</li> <li>• Conexiones JNDI</li> <li>• Tareas de extracción</li> <li>• Tareas de carga</li> <li>• Ejecución e interpretación de resultados</li> </ul>                                     |
| Cubos OLAP                         | <ul style="list-style-type: none"> <li>• Creación del cubo dimensional</li> <li>• Atributos de las dimensiones</li> <li>• Consultas MDX</li> <li>• Despliegue de información a través del navegador</li> <li>• Medidas del cubo dimensional</li> </ul>  |
| Aplicación para desplegar reportes | <ul style="list-style-type: none"> <li>• Despliegue de reportes</li> <li>• Creación de usuarios</li> <li>• Manejo de reportes <i>ad-hoc</i> en Saiku Analytics</li> <li>• Exportar reportes a PDF y Excel</li> <li>• Manejo de cuadros de mando parametrizados en CDE (Community Dashboard Editor)</li> </ul> |