

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE SISTEMAS

UNIDAD DE TITULACIÓN

TÍTULO DEL TRABAJO DE TITULACIÓN

IMPLEMENTAR UNA SOLUCIÓN INFORMÁTICA BASADA EN PANELES DE INTELIGENCIA EMPRESARIAL, PARA LA EVALUACIÓN, SEGUIMIENTO Y CONTROL DE PROYECTOS DE INGENIERÍA CIVIL, VIAL Y SERVICIOS PETROLEROS.

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCION DEL GRADO DE MAGISTER EN SISTEMAS DE INFORMACION MENCION INTELIGENCIA DE NEGOCIOS Y ANALITICA DE DATOS MASIVOS

FRANKLIN GUILLERMO ESPÍN MUÑOZ

franklin.espin01@epn.edu.ec

DIRECTOR: MSc. HENRY PAZ ARIAS

henry.paz@epn.edu.ec

2024

APROBACIÓN DEL DIRECTOR

Como director del trabajo de titulación implementar una solución informática basada en paneles de inteligencia empresarial, para la evaluación, seguimiento y control de proyectos de ingeniería civil, vial y servicios petroleros desarrollado por Franklin Guillermo Espín Muñoz, estudiante de la Maestría En Sistemas Informáticos Con Mención Inteligencia De Negocios Y Analítica De Datos Masivos, habiendo supervisado la realización de este trabajo y realizado las correcciones correspondientes, doy por aprobada la redacción final del documento escrito para que prosiga con los trámites correspondientes a la sustentación de la Defensa oral.

MSc. HENRY PAZ ARIAS
DIRECTOR

DECLARACIÓN DE AUTORÍA

Yo, Franklin Guillermo Espín Muñoz, 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.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

FRANKLIN GUILLERMO ESPÍN MUÑOZ

DEDICATORIA

Dedico este logro a mis queridos familiares, cuyo apoyo incondicional fue fundamental en cada paso de este proceso. Obtener este nuevo título representa un significativo avance tanto en mi desarrollo personal como profesional. A todos ustedes, mi profundo agradecimiento.

AGRADECIMIENTO

Agradezco profundamente a mis respetados docentes por impartirme sus conocimientos con entrega y dedicación. Especialmente, mi gratitud se dirige hacia mi tutor, Henry Paz, y mi codirectora, Maria Pérez, cuya orientación y apoyo fueron fundamentales en la culminación del presente proyecto de titulación.

Contenido

LISTA DE FIGURAS	VI
LISTA DE TABLAS	VIII
LISTA DE ANEXOS	IX
RESUMEN	X
ABSTRACT	XI
1. INTRODUCCIÓN	1
1.1. Pregunta de investigación	2
1.2. Objetivo General	2
1.3. Objetivos Específicos	2
1.4. Alcance	3
2. MARCO TEÓRICO	4
2.1. BI en Proyectos de Construcción e Ingeniería.	4
2.2. Aplicación de BI en Proyectos de Construcción.....	4
2.3. Business Intelligence.....	4
2.4. Dashboard.....	5
2.5. Minería de Datos	17
2.6. RapidMiner	17
2.7. Regresión Lineal en Machine Learning	19
2.8. Oracle.....	20
2.9. Data Warehouse.....	20
2.10. Data Mart.....	21
2.11. Modelo Estrella del DWH.....	21
2.12. Metodología HEFESTO	22
3. DESARROLLO E IMPLEMENTACIÓN.	43
3.1. Implementación de Fases de HEFESTO.....	43
3.2. Análisis de Requerimientos.....	44
3.3. Análisis de los OLTP.....	49
3.4. Modelo Lógico del DWH.	67
3.5. Integración de datos.	74
3.6. Resultados y despliegue.....	77
3.7. Entrenamiento del modelo de regresión lineal y la respectiva predicción de datos.	77
3.8. Evaluación y Usabilidad.....	88
4. CONCLUSIONES	90
5. RECOMENDACIONES	92

LISTA DE FIGURAS

Figura 2. Modelo conceptual power BI.....	9
Figura 3. Cuadrante de Gartner.	15
Figura 4. Comparativo Principales Características de Herramientas BI.....	16
Figura 5. Resultado Total De La Evaluación Herramientas BI.....	17
Figura 6. Fases Metodología Hefesto.	23
Figura 7. Caso práctico, indicadores y perspectivas.	26
Figura 8. Modelo Conceptual.	27
Figura 9. Caso práctico, Modelo Conceptual.	28
Figura 10. Ejemplo práctico, establecer correspondencia.....	30
Figura 11. Modelo Conceptual ampliado.....	33
Figura 12. Caso práctico, Modelo Conceptual ampliado.....	33
Figura 13. Diseño de tablas de dimensiones.	35
Figura 14. Caso práctico, tabla de dimensión” PRODUCTO”.	35
Figura 15. Caso práctico, tabla de dimensión” CLIENTE”.	35
Figura 16. Caso práctico, tabla de dimensión” FECHA”.	35
Figura 17. Tabla de hechos.....	36
Figura 18. Caso práctico, diseño de la tabla de hechos.....	37
Figura 19. Caso práctico, uniones.[19][20].....	38
Figura 20. Escala SUS.	42
Figura 21 Modelo Conceptual Monitoreo Económico	48
Figura 22 Modelo Conceptual Monitoreo de Contratos	48
Figura 23 Modelo Conceptual Monitoreo de Rubro.....	49
Figura 24 Correspondencia de Monitoreo Económico	55
Figura 25 Correspondencia Monitoreo Proyectos Centro de Costos.....	56
Figura 26 Correspondencia Monitoreo Rubros	57
Figura 27 Monitoreo Económico	64
Figura 28 Monitoreo de Contratos.....	65
Figura 29 Monitoreo de Rubros.....	66
Figura 30. Estructura C.E.E.....	67
Figura 31. Clase Centro de Costos.....	68
Figura 32. Localidad	68

Figura 33. Periodos	68
Figura 34. Obra.....	68
Figura 35. Capitulo	69
Figura 36. Rubro	69
Figura 37. Monitoreo Económico	70
Figura 38. Monitoreo por Contrato.....	71
Figura 39. Monitoreo Por Rubro	72
Figura 40. Modelo Completo BI.....	73
Figura 41. Listado de Datos Data Warehouse	74
Figura 42. Carga inicial Extracción, Tranformación,Carga	75
Figura 43. Dashboard Final Monitoreo Económico	76
Figura 44. Dashboard Final Monitoreo Por Contrato	76
Figura 45. Dashboard Final Monitoreo por Rubro	77
Figura 46. Información del contenido del DataFrame.	78
Figura 47. Muestra resumen estadístico de columnas del DataFrame.	79
Figura 48 Correlación de Pearson Inicial	80
Figura 49. Correlación de Pearson Modificado.....	81
Figura 50. Verificación de valores nulos.	81
Figura 51. Histogramas por variables.	82
Figura 52 Verificación de Datos Faltantes en columnas del DataFrame.....	83
Figura 53. Verificación de Datos Faltantes en las filas del DataFrame.....	83
Figura 54. Gráfico de Dispersión de Variables.	84
Figura 55. Separar Características del valor a predecir.....	84
Figura 56. Separar los datos de entrenamiento y testeo.	85
Figura 57. Código crear red neuronal	85
Figura 58. Ejecución Red neuronal.....	86
Figura 59. Grafica curva de aprendizaje.	87
Figura 60. Error medio absoluto.....	87
Figura 61. Eliminar variable a predecir del modelo inicial.	88
Figura 62. Margen operacional Predicho.	88
Figura 63. Generación del archivo para posterior puesta en marcha.	88
Figura 64. Resultados Ponderados de Usabilidad BI.....	89

LISTA DE TABLAS

Tabla 1. Tipos de Dashboards.	6
Tabla 2. Comparativa de Herramientas BI.	15
Tabla 3 identificar Preguntas.	45
Tabla 4 Indicadores y perspectivas	46
Tabla 5. Indicadores de Monitoreo Económico.	50
Tabla 6. Monitoreo de Contratos	52
Tabla 7 Monitoreo de Rubro	53
Tabla 8. Estructura CEE.	58
Tabla 9. Clase Centro de Costos.	58
Tabla 10. Localidad	58
Tabla 11. Periodos	59
Tabla 12. Fecha de Corte.	59
Tabla 13. Rubro	59
Tabla 14. Capitulo.	60
Tabla 15. Unidad.	60
Tabla 16. Centro de Costos	60
Tabla 17. Contrato	61
Tabla 18. Obra	61
Tabla 19. Cronograma.	61
Tabla 20. Planillaje	62
Tabla 21. Producción	62
Tabla 22. Costos	63

LISTA DE ANEXOS

Anexo I Acuerdo De Confidencialidad De La Información Del CEE

Anexo II Base de datos del DWH

Anexo III Encuesta de usabilidad SUS

Anexo IV Diagrama de la BDD del DWH

Anexo VI Proceso de Carga Principal Package Dwk Produccion

Anexo VII Proceso de Carga Cuerpo PACKAGE BODY Dwk Produccion

Anexo VIII Tableros de control BI Producción

RESUMEN

El proyecto consistirá en implementar una Solución Informática basada en Business Intelligence (BI) para mejorar la Gestión de Datos del Cuerpo de Ingenieros del Ejército (C.E.E.), responsable de operaciones militares y construcción. Se realizará la extracción manual y análisis de datos, proponiendo paneles visuales y métricas estratégicas. Su objetivo primordial es optimizar la interpretación de datos y la toma de decisiones eficientes para lograr un control y evaluación óptimos de proyectos.

La metodología empleada implicará el proceso ETL con Oracle SQL Developer, para alimentar el Data Warehouse, evaluando la eficiencia en costos, ingresos-gastos y avance de proyectos. Se utilizará técnicas de Machine Learning, específicamente regresión lineal, para analizar datos históricos del C.E.E. y predecir ingresos futuros con mayor precisión.

El proyecto culminará con presentaciones Gerenciales utilizando Power BI, mejorando la toma de decisiones en el C.E.E. gracias a la implementación de BI y la proyección del margen operacional con Machine Learning basado en datos históricos de ingresos y egresos del C.E.E. Este enfoque integral y tecnológico permitirá una gestión más eficaz de datos y un análisis más profundo para la toma de decisiones estratégicas del C.E.E.

Palabras clave: C.E.E., BI, Machine Learning, Data Warehouse, Power BI, ETL.

ABSTRACT

The project will consist of implementing a Business Intelligence (BI) based IT solution to enhance the Data Management of the Army Corps of Engineers (C.E.E.), responsible for military operations and construction. Manual data extraction and analysis will be carried out, proposing visual dashboards and strategic metrics. Its primary objective is to optimize data interpretation and enable efficient decision-making to achieve optimal project control and evaluation.

The methodology employed will involve the ETL process with Oracle SQL Developer to feed the Data Warehouse, evaluating cost efficiency, income-expenses, and project progress. Machine Learning techniques, specifically linear regression, will be used to analyze historical C.E.E. data and predict future revenues with greater accuracy.

The project will culminate in Management presentations on the web using Power BI, improving decision-making at the C.E.E. through the implementation of BI and revenue projection with Machine Learning based on historical C.E.E. income and expenditure data. This comprehensive and technological approach will enable more effective data management and deeper analysis for C.E.E.'s strategic decision-making

Keywords: C.E.E., BI, Machine Learning, Data Warehouse, Power BI, ETL.

1. INTRODUCCIÓN

El Cuerpo de Ingenieros del Ejército se ubica en la parroquia La Magdalena, Cantón Quito, Provincia de Pichincha. Es una entidad dedicada a realizar operaciones de ingeniería militar en apoyo a las Fuerzas Terrestre. Además, se dedica a la construcción, mantenimiento y rehabilitación de carreteras, estructuras civiles y proyectos petroleros, al tiempo que proporciona servicios a la comunidad global. Con un compromiso con la transparencia, la efectividad y la credibilidad Institucional, contribuye significativamente a la Seguridad Nacional y al Desarrollo del país.

Entre sus actividades de ingeniería militar, la ejecución de construcciones constituye el proceso central en la cadena de valor institucional del Cuerpo de Ingenieros del Ejército. Este proceso comienza con la adquisición de proyectos de entidades contratantes. Una vez que se adjudica un contrato, la ejecución del proyecto comienza con el apoyo de diversos procesos automatizados complementarios. Estos abarcan procesos financieros que capturan datos presupuestarios, contables y de tesorería. Además, la gestión de recursos humanos integra procesos automatizados para la contratación de personal, administración y nómina. La gestión logística comprende módulos clave como adquisiciones, control de almacenes, activos fijos y órdenes de mantenimiento. Estos módulos de soporte se sincronizan con el proceso central de desarrollo de construcciones del Cuerpo de Ingenieros, proporcionando información fundamental para el estudio.

El proceso de desarrollo de construcciones incluye la administración técnica, que facilita el registro de cronogramas, la planificación de proyectos, el seguimiento del avance y la gestión de nóminas, garantizando el control técnico y la evaluación del proyecto. Además, los componentes suplementarios del proyecto incluyen la sección de cobranzas que representa los ingresos institucionales y la producción, que comprende el registro de actividades diarias de maquinaria y mano de obra, subcontratación, consumo de materiales y actividades de talleres internos y externos. Estos elementos constituyen partes integrales de la estimación de costos primarios del proyecto.

Este flujo de trabajo institucional sirve como fuente de datos fundamental y punto focal para este proyecto de investigación, ofreciendo una visión profunda de las complejidades e interdependencias del proceso de desarrollo de construcciones del Cuerpo de Ingenieros del Ejército (C.E.E.).

1.1. Pregunta de investigación

¿Cómo la implementación de paneles de inteligencia empresarial segmentados por procesos institucionales a través de un datamart puede mejorar la evaluación, seguimiento y control de los proyectos de ingeniería civil, vial y servicios petroleros en el Cuerpo de Ingenieros del Ejército, optimizando la gestión de costos y recursos?

¿Cuál es el impacto de la centralización de datos en un repositorio único sobre la calidad de la toma de decisiones estratégicas en proyectos de ingeniería civil, vial y servicios petroleros?

¿De qué manera los procesos de integración, transformación y carga de datos (ETL) mejoran la coherencia y calidad de la información utilizada para la evaluación y seguimiento de los proyectos en el Cuerpo de Ingenieros del Ejército?

1.2. Objetivo General

Diseñar e implementar una solución integral de Business Intelligence (BI) que centralice y visualice los datos del Cuerpo de Ingenieros del Ejército (C.E.E.) mediante un datamart y paneles de control especializados. El propósito es optimizar el análisis, seguimiento y gestión de proyectos de ingeniería civil, vial y servicios petroleros, mejorando así la eficiencia en la evaluación de costos, recursos y toma de decisiones estratégicas.

1.3. Objetivos Específicos

- Definir los requisitos necesarios para el diseño de los diferentes visualizadores en el área de costos de producción.
- Recopilar, almacenar y mantener los datos de las diferentes fuentes relevantes en un repositorio de datos centralizado.

- Asegurar la calidad de los datos mediante procesos de integración, extracción, transformación y carga, con el propósito de mantener la información procesada coherente y de alta calidad.
- Identificar las métricas críticas para la toma de decisiones estratégicas en proyectos de construcción de obras civiles, viales y petroleras, a través del análisis, evaluación, seguimiento y control de la información, con el objetivo de optimizar la gestión en la organización.
- Crear paneles de control y reportes de visualización con el objetivo de simplificar la toma de decisiones en el área de construcción de obras civiles, viales y petroleras.

1.4. Alcance

Esta investigación se centrará en la gestión de la data existente de los procesos constructivos, cuya interpretación rápida y efectiva es actualmente dificultosa, impactando negativamente en la toma de decisiones y generando pérdidas de tiempo y recursos. Además, no se dispone de una fuente única de información a nivel gerencial para controlar los proyectos. Este trabajo permitirá realizar el análisis, extracción, transformación, carga y presentación de los datos en una solución integral, sólida, confiable y robusta de Business Intelligence. De este modo, el área directiva, administrativa y de nivel medio podrá acceder a información de manera más comprensible, resumida y gráfica para el control y evaluación de proyectos de construcción civil, vial y servicios petroleros. Se logrará mediante tableros de visualización (dashboards) estandarizados, optimizados y clasificados según las diferentes áreas de análisis (económicos, control y evaluación de proyectos), permitiendo a la gerencia optimizar los recursos económicos y humanos y fomentar la mejora continua de la Institución.

2. MARCO TEÓRICO

2.1. BI en Proyectos de Construcción e Ingeniería.

El uso de Business Intelligence (BI) en proyectos de construcción e ingeniería ha demostrado ser una herramienta crucial para mejorar la eficiencia y efectividad en la gestión de proyectos. La implementación de soluciones BI permite a las organizaciones del sector evaluar, monitorear y controlar sus proyectos con mayor precisión y oportunidad. A continuación, se presentan estudios previos y ejemplos relevantes que ilustran la aplicación práctica de BI en estos sectores.[1].

2.2. Aplicación de BI en Proyectos de Construcción.

En el ámbito de la construcción, BI se utiliza para gestionar grandes volúmenes de datos provenientes de diversas fuentes, como presupuestos, cronogramas, informes de avance y costos.

La implementación de un sistema de Business Intelligence (BI) en una empresa constructora permitió una mejora significativa en la visibilidad de los costos del proyecto. El sistema integró datos de múltiples fuentes, facilitando la creación de dashboards interactivos que ofrecían una vista consolidada de los gastos y desviaciones en tiempo real. Como resultado, la empresa logró reducir los costos de sobre presupuesto en un 15% y mejorar la puntualidad de las entregas en un 20%.[2].

2.3. Business Intelligence.

El Business Intelligence (BI) es un conjunto de métodos y prácticas que se enfocan en gestionar la información para ayudar a las empresas a tomar decisiones más acertadas. A través del BI, las empresas pueden combinar y analizar datos de diferentes fuentes, obteniendo así una visión integral y actualizada. Esto les permite anticipar eventos futuros tanto en su entorno como dentro de la organización.[3].

la Inteligencia de Negocios facilita la recolección, limpieza y transformación de datos provenientes de sistemas transaccionales.

Convierte información desorganizada de diversas fuentes internas y externas en datos estructurados, que pueden ser utilizados para almacenar, analizar y elaborar informes sobre el desempeño y la evolución de la organización.[3].

2.4. Dashboard

Los dashboards o tableros de control son una herramienta que facilita la representación visual de indicadores y datos, tanto numéricos como alfanuméricos. Estos se emplean para evaluar los resultados de una tarea, acción o estrategia específica en relación con metas predefinidas. Los indicadores proporcionan una medida del éxito de estas actividades.[4].

Estas herramientas son muy útiles para identificar visualmente tendencias, patrones y anomalías, permitiendo un análisis más efectivo y la toma de decisiones más acertadas. Por lo cual, es fundamental diseñar estas herramientas aprovechando al máximo las capacidades visuales humanas.[4].

Tipos de Dashboards

Existen tres tipos comunes de dashboards, cada uno con un propósito particular:

- Operacional
- Estratégico/Ejecutivo
- Analítico

La Tabla 1 presenta una comparación entre estos diferentes tipos de dashboards.

Tabla 1. Tipos de Dashboards.

	Operacional	Táctico	Estratégico
Finalidad	Operaciones de monitoreo	Mide el progreso	Aplica estrategias
Usuarios	Jefes de área y especialistas	Administradores, analistas	Directivos, administrativos
Estrategia	Operacional	Departamental	Empresas
Tipo de información	Detallada	Detallada y resumida	Detallada y resumida
Frecuencia de actualizaciones	Diaria	Diaria o semanal	Mensual y cada cuatro meses
Énfasis	Monitoreo	Análisis	Administrativo

Dashboards Operativos. - Los dashboards operativos se enfocan más en el monitoreo que en el análisis y la gestión, según Kerzner (2013). Este tipo de dashboard presenta datos clave para las operaciones comerciales, permitiendo hacer un seguimiento diario de los procesos y áreas de la organización. Facilitan la toma de medidas correctivas cuando es necesario. El objetivo principal de estos dashboards es proporcionar la información necesaria para actuar y tomar decisiones operativas rápidamente. [5].

Dashboards Tácticos. - Un dashboard táctico se encarga de supervisar los procesos y proyectos departamentales que son importantes para una parte específica de la organización o un grupo reducido de personas. Se utiliza para comparar el rendimiento de dichas áreas o proyectos., así como para evaluar planes de presupuesto, pronósticos o resultados anteriores, según Kerzner (2013). Además, este tipo de dashboard permite explorar los datos en profundidad, proporcionando información adicional. Ofrece tanto información interna como externa necesaria para comprender la situación y evitar sorpresas desagradables que podrían afectar el posicionamiento estratégico y a largo plazo de la empresa. [5]

Dashboards Estratégicos. - Los dashboards estratégicos suelen mostrar los KPIs (Indicadores Clave de Rendimiento) que el equipo ejecutivo de la organización revisa regularmente, ya sea a diario, semanal o mensualmente. Su objetivo es alinear a toda la organización con los objetivos estratégicos y asegurar que todos los grupos avancen en la misma dirección, como señala Kerzner (2013). Estos dashboards permiten monitorear los resultados generales de la empresa y de sus áreas clave. El monitoreo se realiza aproximadamente una vez al mes. Pueden incluir indicadores de todos los sectores para los ejecutivos clave o estar segmentados para un ejecutivo específico.[5] .

Elementos generales de un Dashboard

- **Reporte o Pantalla:** Proporciona la información clave necesaria para el diagnóstico de una organización, según el formato y la configuración utilizados.[5] .
- **Apertura:** La manera en que se presentan los contenidos y se organiza la información, entre otros aspectos.[4] .
- **Frecuencia de actualización:** Este componente establece el intervalo entre las actualizaciones de los datos, ya que la frecuencia puede variar dependiendo del proyecto, siendo diaria, semanal, mensual o incluso anual. [4].
- **Indicadores de rendimiento (KIP):** Los KPIs, o indicadores clave de rendimiento, son métricas y datos estadísticos que ofrecen información para evaluar cómo un proceso está cumpliendo con una meta específica. Es crucial entender que los KPIs no miden directamente los resultados, sino que representan indicadores de éxito que reflejan objetivos o resultados esperados. Además, no prescriben acciones a seguir, sino que muestran lo que ya se ha logrado hasta el momento. [4].
- **Indicadores de eficiencia:** Los indicadores de eficiencia evalúan la velocidad con la que se lograron los indicadores de rendimiento, ya que se manejan en una extensa gama de actividades y se comparan

con los tiempos de entrega establecidos en un cronograma u operaciones requeridas por el proyecto.[4]

- **Periodo del Indicador:** Debe mantener un registro del tiempo necesario para cumplir con un indicador específico para facilitar la evaluación de la eficiencia.[4]
- **Referencia:** Este componente posibilita la comparación del historial de datos para analizar el avance de los indicadores y evaluar el logro de los objetivos. En cuanto al reporte, este aspecto debe presentar de forma clara los datos relevantes al momento de la consulta, facilitando así el análisis de la información de acuerdo con el formato y la configuración necesarios.[4]
- **Parámetro de alarma:** Este aspecto notifica cuando uno de los indicadores ha excedido los valores o el tiempo establecido en el cronograma. Debe presentar una alerta visual como referencia, lo que permite tomar las medidas necesarias o, si es preciso, ajustar los límites de los valores o las fechas de entrega.[4]
- **Productividad:** Es un indicador que mide cómo eficientemente se produce utilizando recursos, con el fin de optimizar el rendimiento máximo. Además, según González y Carro, mejorar la productividad implica perfeccionar el proceso de producción para lograr una comparación favorable entre los recursos empleados y la cantidad de bienes y servicios generados.[4]
- **Gráfico:** Se refiere a la representación visual de los indicadores mediante gráficos, los cuales pueden ser de diferentes tipos como barras, pastel, líneas, entre otros.[4]

Análisis comparativo de herramientas Business

Intelligence

En esta parte se llevará a cabo una comparación entre las diversas herramientas de Inteligencia de Negocios que se utilizan para mejorar la toma de decisiones en distintos contextos.

Herramientas De Inteligencia De Negocios

1. Microsoft Power BI

Power BI es un software de licencia propietaria desarrollado por Microsoft. Este software transforma los datos en gráficos y elementos visuales, permitiendo a los usuarios concentrarse en la información más relevante para la toma de decisiones. Según Microsoft, Power BI constituye un conjunto de aplicaciones para análisis empresarial que facilita la interpretación de datos y el intercambio de información. Sus paneles ofrecen a los usuarios una visión integral de sus métricas clave en un único lugar. La información se actualiza en tiempo real y está disponible en todos los dispositivos del usuario.[6].

A continuación, se muestra el Modelo conceptual power BI en la Figura 2.

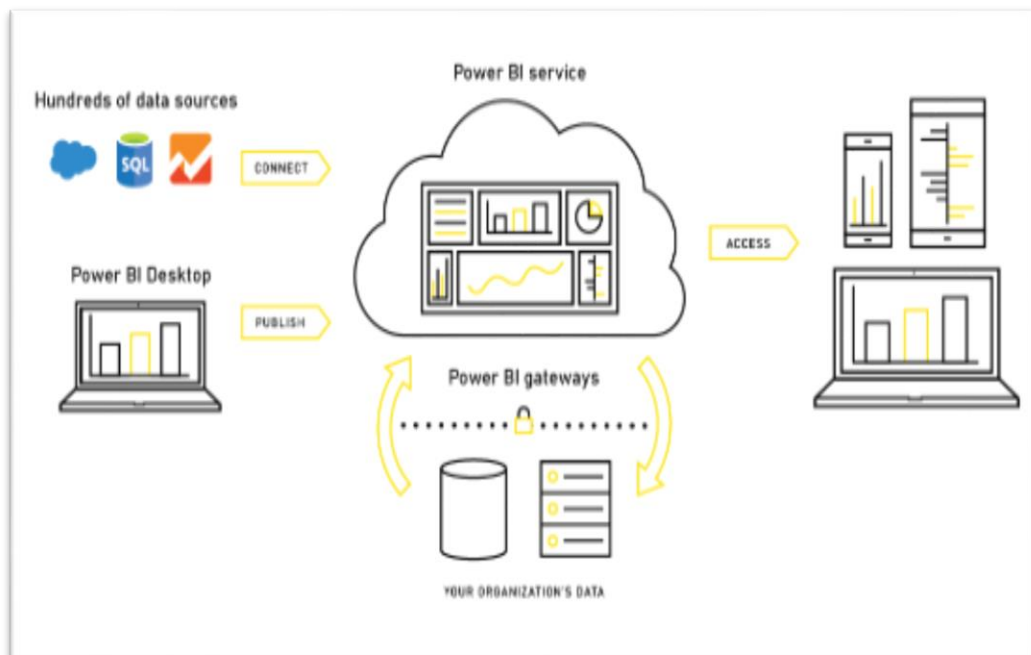


Figura 1. Modelo conceptual power BI.

Ventajas de Power BI

- **Integración con diferentes fuentes de datos:** Power BI posibilita la conexión y extracción de datos desde una amplia gama de fuentes, incluyendo bases de datos, archivos Excel, servicios en la nube y otros. Esto simplifica la integración de información procedente de diversos sistemas.

- **Facilidad de uso:** Power BI proporciona una interfaz amigable e intuitiva que no necesita de habilidades avanzadas de programación. Con solo arrastrar y soltar, los usuarios pueden crear visualizaciones interactivas y personalizadas.
- **Amplia gama de visualizaciones:** Power BI ofrece una amplia gama de opciones de presentación, que van desde gráficos básicos hasta tablas dinámicas y mapas interactivos. Esto posibilita la presentación clara y comprensible de los datos, lo que facilita la interpretación y comunicación de los resultados.
- **Actualización en tiempo real:** Con Power BI, es posible establecer conexiones con datos en tiempo real, lo que implica que las visualizaciones se actualizan automáticamente conforme cambian los datos subyacentes. Esta función facilita la toma de decisiones basadas en información actualizada y pertinente.
- **Capacidad de colaboración:** Power BI proporciona capacidades de colaboración que permiten compartir informes y tableros con otros usuarios, lo que facilita la colaboración en tiempo real y la toma de decisiones en equipo.
- **Integración con otras herramientas de Microsoft:** Power BI se incorpora de manera nativa con otras herramientas de Microsoft, como Excel y Azure, lo que facilita la exportación de datos desde estas herramientas y la utilización de funciones adicionales, como análisis avanzados o almacenamiento en la nube.[7].

En resumen, Power BI presenta beneficios notables en lo que respecta a la integración de datos, su manejo sencillo, la visualización de información, actualización en tiempo real, colaboración y su compatibilidad con otras herramientas de Microsoft. Estas cualidades hacen de Power BI una elección sólida para el análisis de datos en una variedad de entornos empresariales.[7].

Desventajas de Power BI

- **No maneja de forma adecuada grandes fuentes de datos:** Si te encuentras ante un extenso conjunto de datos que requiere análisis, es posible que Microsoft Power BI no sea la opción más adecuada. Se podrían enfrentar dificultades al conectar e importar grandes volúmenes de datos, así como a un rendimiento lento y largos tiempos de espera.
- **Procesamiento de información forzado:** Otro inconveniente es que se requiere un considerable esfuerzo para procesar la información y ajustarla a los conjuntos de datos pequeños del panel de control. Esto naturalmente limita la flexibilidad y la capacidad dinámica necesaria para analizar variables de manera más detallada. La interfaz puede resultar sobrecargada, dado que la ventana de ayuda y la barra lateral bloquean la visualización esencial de la pantalla, lo que provoca una interfaz abarrotada. Por lo tanto, navegar por el panel de control mediante desplazamiento requeriría un esfuerzo adicional para configurarlo o adaptarlo a las necesidades individuales.
- **Intercambio limitado de datos:** Los reportes y tableros pueden ser compartidos exclusivamente con usuarios que pertenezcan al mismo dominio de correo electrónico, así como con aquellos cuyos dominios de correo electrónico estén registrados en Office 365.[8].

2. Qlikview

Es una plataforma de inteligencia comercial contemporánea que ofrece a los usuarios una aplicación efectiva para analizar y descubrir datos con el objetivo de aumentar el valor de la empresa. Facilita el desarrollo y despliegue ágil de aplicaciones. Permite descubrir tendencias ocultas y tomar decisiones fundamentadas al promover el análisis sin restricciones de datos. Con un almacén de datos totalmente en memoria y herramientas de BI integradas, QlikView presenta un nuevo enfoque en la Inteligencia de Negocios.

Este modelo se basa en una Arquitectura Asociativa, que desafía el enfoque tradicional de exploración de BI al permitir conexiones automáticas a través de campos con nombres idénticos en diferentes tablas, sin necesidad de seguir una jerarquía de datos predeterminada. En comparación con las arquitecturas de BI convencionales, esto facilita la identificación de patrones y conexiones.[9].

Ventajas de Qlikview

- Proporciona funcionalidades de autoservicio que simplifican su utilización.
- Usando tecnología asociativa de memoria, los usuarios pueden analizar y procesar datos de manera rápida.
- Todas las relaciones entre los campos analizados se mantienen en su arquitectura asociativa.
- Permite la implementación inmediata sin esperar mucho para ver los resultados.[9]
- Facilitan el proceso de toma de decisiones al evitar limitaciones en los informes estáticos.
- Permite la exploración de datos instantánea y en tiempo real.
- Transforma grandes cantidades de datos complejos provenientes de múltiples fuentes en gráficos interactivos fáciles de entender y listos para usar.
- Proporcionan a los usuarios visualizaciones interactivas mediante el uso de avanzadas técnicas web.
- Para buscar datos complejos y acelerar el descubrimiento, utilizan la búsqueda natural.
- Ofrecen experiencias analíticas guiadas que ayudan a los usuarios a descubrir conceptos y a tomar decisiones.
- Tienen la capacidad de manejar enormes volúmenes de registros para análisis ad hoc.
- Al permitir la incorporación de datos en el motor de datos indexados en memoria de Qlik, no es necesario un almacén de datos.

- El procesamiento en memoria comprime los datos y optimiza el rendimiento.[9].

Desventajas Qlikview

- Limitaciones cuando se trata de separar los datos de la estructura propietaria.
- Problemas con la coherencia de datos durante la creación de modelos múltiples.
- Dependencia de scripts durante las cargas iniciales.
- El costo de mantenimiento resulta elevado.
- La lógica asociativa de QlikView:
 - La lógica asociativa de QlikView: QlikView ofrece capacidades de lógica asociativa en memoria, permitiendo que la aplicación realice cálculos en tiempo real y facilite la navegación por la información de manera más intuitiva que otras soluciones.

La principal distinción entre QlikView y otras soluciones de BI radica en la experiencia de usuario asociativa. Mientras que en las soluciones de BI tradicionales se utilizan caminos predefinidos para navegar y explorar los datos, QlikView no limita a jerarquías ni nociones preconcebidas. Los usuarios tienen la libertad de elegir cualquier ruta para realizar el análisis, lo que QlikTech describe como "trabajar de manera similar a la mente humana". La exploración se lleva a cabo de forma libre mediante selecciones interactivas y búsquedas de palabras clave, permitiendo hacer preguntas en cualquier dirección sin restricciones. [9].

3. Tableau

Fue creado con la intención de brindar un software que facilite un análisis fácil de entender, de fácil uso y que permita una comprensión clara de los datos. Brindar una experiencia de inteligencia empresarial que les permita visualizar sus propios datos con un mínimo de entrenamiento es su

enfoque estratégico. Además, la herramienta ofrece capacidades de análisis y funciones estadísticas previamente configuradas para los analistas de negocios. Esta solución es muy popular en entornos de inteligencia empresarial de autoservicio debido a su interfaz de usuario fácil de entender, inteligencia integrada y uso de memoria principal para optimizar el rendimiento. Además, Tableau se ha enfocado en la conectividad con múltiples fuentes de datos, ofreciendo adaptadores nativos para una variedad de sistemas de gestión de bases de datos relacionales, archivos locales, fuentes de datos en la nube y almacenes de datos multidimensionales.[9].

Ventajas de Tableau

- Ofrece capacidades para modelar y analizar datos con acceso a una amplia gama de fuentes de datos.
- Su interfaz de usuario es fácil de usar, intuitiva y cuenta con una alta aceptación por parte de los usuarios.
- Proporciona análisis visual con inteligencia integrada dirigida a usuarios empresariales, sin necesidad de poseer amplios conocimientos de programación.
- Permite la publicación de informes en la web corporativa o en cualquier sistema empresarial basado en arquitectura web.
- Simplifica la integración analítica de información proveniente de diversas fuentes.
- Ofrece soporte para API REST y JavaScript, lo que agiliza la integración a través de plataformas de análisis y reduce la necesidad de programación personalizada.[9].

Desventajas de Tableau

- La integración y modelado de datos se lleva a cabo en la aplicación de escritorio del cliente.
- Existe una carencia de soporte para la navegación guiada en paneles y aplicaciones de inteligencia empresarial (BI).

- Su uso principal se observa en despliegues individuales y departamentales, lo que puede generar un riesgo de "explosión" de informes al emplear Tableau como solución de BI empresarial.[9]

En la figura 3, se visualiza la comparativa de diferentes plataformas de BI consideradas como líderes. [10].

Con el cual se ha tomado como referencia el Cuadrante de Gartner de febrero del 2019.



Figura 2. Cuadrante de Gartner.

Tabla 2. Comparativa de Herramientas BI.

Característica	QlikView	Power BI	Tableau
Integración de datos	90	90	85
Facilidad de uso	85	95	75
Visualización de datos	80	95	90
Actualización en tiempo real	80	70	85
Capacidad de colaboración	90	75	85
Compatibilidad con otras herramientas	75	80	90
Manejo de grandes volúmenes de datos	70	70	60
Desempeño en procesamiento	70	75	65

Característica	QlikView	Power BI	Tableau
Interfaz para el usuario	80	75	70
Compartir informes fuera del dominio	60	75	70
Total	780	800	775

A continuación, en la figura 4, se muestra el comparativo de las principales características de las herramientas de BI.

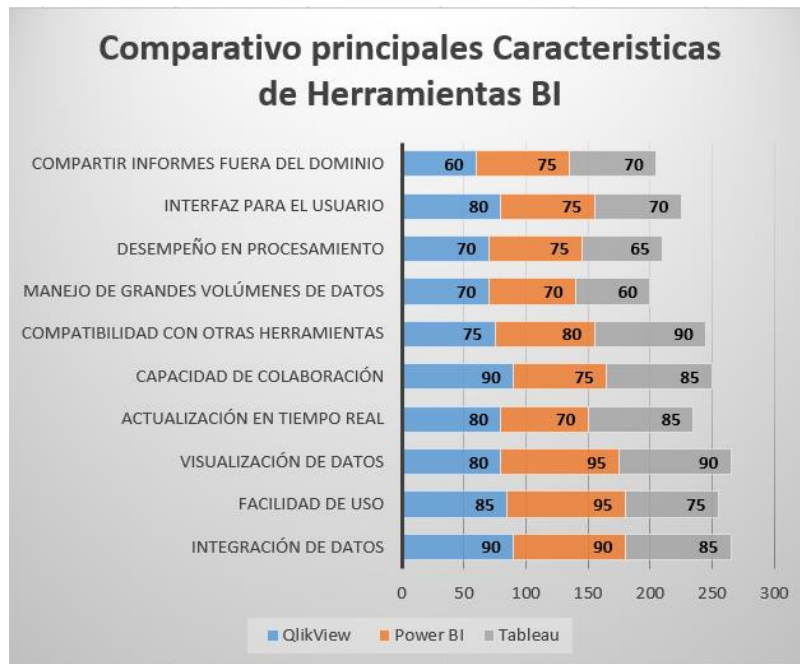


Figura 3. Comparativo Principales Características de Herramientas BI.

En la figura 5, se muestra en el gráfico del resultado total de la evaluación de herramientas de BI, donde se observa que la mejor opción es Power BI con 800 puntos.

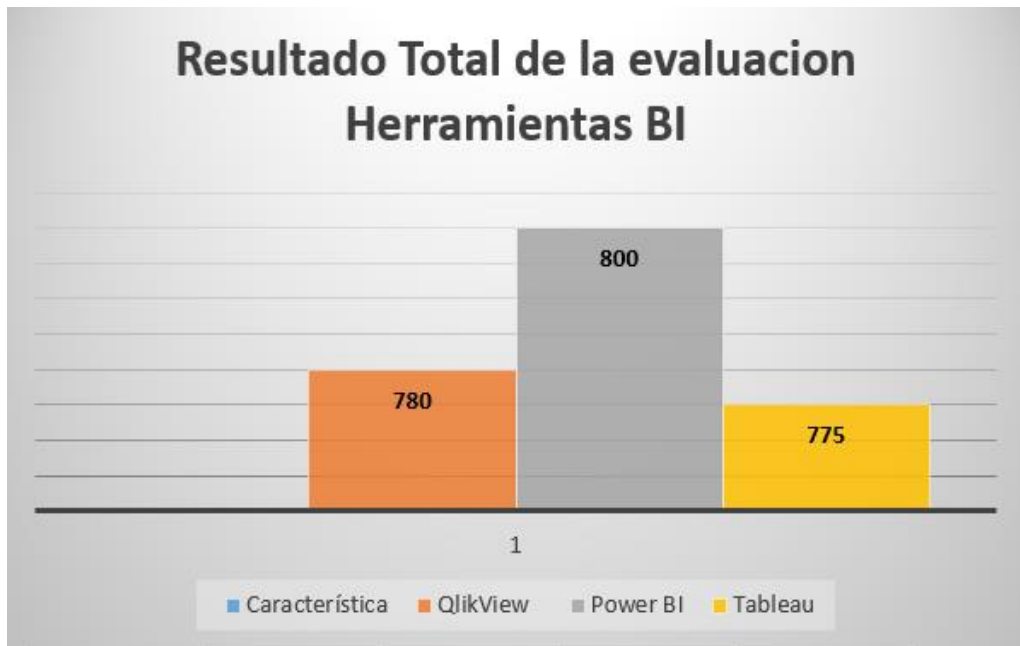


Figura 4.Resultado Total De La Evaluación Herramientas BI.

2.5. Minería de Datos

La minería de datos implica la identificación de patrones y conocimientos útiles a partir de grandes cantidades de datos. Estas fuentes de datos pueden incluir bases de datos y almacenes de datos, así como otros repositorios de información o datos que se reciben de manera dinámica.[11]

2.6. RapidMiner

Es una herramienta que simplifica el análisis predictivo mediante un método de programación arrastrar y soltar, con la opción de generar código opcionalmente. Este enfoque acelera la creación y validación de modelos, así como el procesamiento y desarrollo de datos. RapidMiner, escrita en Java, cuenta con una amplia gama de más de 500 operadores que ofrecen diversos enfoques para visualizar las relaciones en los datos. Incluye opciones para la minería de datos, textos y web, así como para el análisis de sentimientos y opiniones.[12].

Machine Learning - aprendizaje automático

Es una rama de la IA que permite a las máquinas aprender sin ser programadas para hacerlo, una habilidad esencial para la creación de

sistemas capaces de identificar patrones en los datos para hacer predicciones.

Esta tecnología se encuentra en una variedad de aplicaciones, como las recomendaciones de Netflix o Spotify, las respuestas inteligentes de Gmail o el habla de Siri y Alexa. Finalmente, el aprendizaje automático reconoce patrones y puede convertir una muestra de datos en un programa informático que puede hacer inferencias de nuevos conjuntos de datos.[13].

Esta fase del proyecto se enfocará en la construcción de un modelo predictivo para estimar el margen operacional de los proyectos de construcción. Para llevar a cabo este proceso, se utilizará la herramienta Collaborate de Google, que proporciona un entorno colaborativo para trabajar con datos y modelos de Machine Learning de manera eficiente.

- Preparación de los Datos:

Se comenzará cargando los datos en un DataFrame, que es una estructura de datos tabular utilizada en Python para almacenar y manipular conjuntos de datos.

- Se realizará un Análisis de Pearson:

Permitirá evaluar la correlación entre las diferentes variables y el margen operacional.

- Validación de datos faltantes en la data del DataFrame

Se utilizará la herramienta `msno.bar()` para visualizar los datos faltantes en el conjunto de datos y `msno.matrix()` para identificar patrones de datos perdidos.

- Graficaremos histogramas

Se graficarán histogramas para comprender la distribución y dispersión de las variables.

- Construcción del Modelo Predictivo

Se separarán los datos en conjuntos de entrenamiento y prueba, utilizando una proporción recomendada de 70% para entrenamiento y 30% para pruebas.

- Se Utilizará un modelo de regresión neuronal

Una técnica de Machine Learning inspirada en el funcionamiento del cerebro humano, para realizar predicciones precisas del margen operacional.

- Optimizaremos nuestro modelo

Se optimizará el modelo utilizando el algoritmo Adam y se definirá la función de pérdida como el error cuadrático medio (MSE), que permite evaluar la precisión de las predicciones.

- Evaluación del Modelo:

Se entrenará el modelo utilizando los datos de entrenamiento y se evaluará su rendimiento utilizando los datos de prueba. Se calculará el error absoluto medio para medir la discrepancia entre las predicciones del modelo y los valores reales del margen operacional.

- Predicción del Margen Operacional:

Una vez que el modelo esté entrenado y optimizado, se utilizará para realizar predicciones sobre nuevos conjuntos de datos. Se aplicarán técnicas de escalado estándar para preparar los datos de entrada y asegurar que estén en el formato adecuado para el modelo. Finalmente, se utilizará el modelo para predecir el margen operacional en diferentes escenarios de proyectos de construcción.

Al implementar este enfoque, esperamos obtener insights valiosos que nos ayuden a tomar decisiones informadas y mejorar la rentabilidad de nuestros proyectos de construcción.[13][14]

2.7. Regresión Lineal en Machine Learning

La regresión lineal, un algoritmo utilizado en el aprendizaje automático para el análisis y la predicción de resultados, se emplea para prever valores basados en una variable denominada variable dependiente, típicamente representada como 'y', en función de una o varias variables independientes generalmente designadas como 'x'.

En este contexto, la variable dependiente adopta valores numéricos, como los precios de las viviendas, las alturas o las distancias de los planetas al sol. En su forma más básica, cuando solo tenemos una variable independiente "x", la regresión lineal representa una línea recta en un plano bidimensional que mejor se ajusta a los valores de los datos. [14].

2.8. Oracle

Los principales actores en el mercado de sistemas de gestión de bases de datos (RDBMS) son Oracle Database, SAP HANA, Microsoft SQL Server e IBM Db2. De acuerdo con DB Engines, Oracle, fundada por Lawrence J. Ellison en 1977, lanzó su primera versión al mercado en 1979. Actualmente, ofrece la versión 19c como opción a largo plazo y la versión 21c como innovación más reciente (a partir de octubre de 2021).

El entorno informático empresarial se basa en Oracle Database. Las bases de datos pueden ser jerárquicas, en red, orientadas a objetos o basadas en documentos, dependiendo del modelo de estructura utilizado. Para su base de datos Oracle, Oracle utiliza un modelo de base de datos relacional, que permite almacenar y organizar los datos de la empresa y sus clientes en conjuntos de datos estructurados. Los datos están organizados en tablas, columnas y filas y están conectados por atributos. La capacidad de estructurar y presentar grandes cantidades de datos de manera eficiente e intuitiva es una de las principales ventajas de Oracle Database. Las empresas también pueden usar Oracle Database en entornos locales o en la nube.[15].

2.9. Data Warehouse

Una subdivisión especializada de un almacén de datos adaptada a las necesidades de una línea de negocios específica se conoce como data mart. Los datos en este repositorio se mantienen durante el tiempo necesario para satisfacer los requisitos de análisis de la organización.

Es un centro de datos centralizado que ayuda en la toma de decisiones comerciales e independiente de los sistemas operativos. Los datos se almacenan en el almacén de datos desde una perspectiva histórica.

La información del almacén de datos se recopila de diferentes fuentes, se verifica, se depura y luego se integra en el sistema del almacén de datos. Una gran capacidad de almacenamiento rápido es necesaria para la gestión eficiente del almacén de datos. Esta plataforma responde a consultas de datos complejas y utiliza la información para guiar la toma de decisiones.[16].

2.10. Data Mart

Una subdivisión especializada de un almacén de datos que se adapta a las necesidades de una línea de negocios específica se conoce como data mart. Estos almacenes de datos están diseñados para recopilar información breve que sea relevante para un departamento específico, como el departamento de ventas. Se alimentan de una variedad de fuentes de datos, incluidos los sistemas internos, el almacén de datos de la empresa y otras fuentes externas. Su objetivo principal es facilitar la consulta de datos, lo que significa que la organización interna de la información es clara. Para su estructura, suelen utilizar modelos dimensionales de estrella o copo de nieve.

La data mart está configurada para satisfacer las necesidades únicas de un grupo específico de usuarios de la empresa. Esta es la razón por la que también se les considera almacenes de datos de departamento. Además de esta diferencia, las marjas de datos tienen otra ventaja importante: son más rápidos y fáciles de usar porque trabajan con conjuntos de datos más pequeños. [16].

2.11. Modelo Estrella del DWH

El esquema estrella empleado en el estudio consiste en una tabla principal de hechos denominada 'Hechos', que ocupa una posición central en el modelo y se relaciona con las correspondientes tablas de dimensiones.

Estas tablas de dimensiones representan cada uno de los ejes dentro de un espacio multidimensional. La tabla de hechos constituye el elemento central de análisis y contiene atributos de hechos, que son de naturaleza cuantitativa y cuyos valores se derivan mediante la aplicación de funciones estadísticas que resumen un conjunto de valores en un solo valor.[17].

2.12. Metodología HEFESTO

a) Definición

Desde el principio, la metodología Hefesto se presenta como una estrategia efectiva y adaptable para la creación de un Data Warehouse (DW). Su enfoque práctico se ajusta a varios ciclos de vida del software, centrándose en el análisis de los requisitos y fuentes de datos necesarios para la implementación del DW. Esta metodología funciona para el diseño de Data Warehouse y Data Mart. Se destaca por su claridad y por tener pasos claros con objetivos y resultados esperados.

La capacidad de involucrar a los usuarios finales en cada etapa del proceso es un aspecto importante de la metodología Hefesto; esto ayuda a reducir su resistencia al cambio al permitirles influir en el comportamiento y las funciones de DW.

Los modelos conceptuales y lógicos son fáciles de entender, y la independencia entre la metodología y las herramientas tecnológicas utilizadas facilita una adaptación ágil.

Los resultados obtenidos en cada fase sirven como punto de partida para las siguientes etapas, lo que permite una progresión continua en el desarrollo del DW. [18].

El diagrama siguiente resume la metodología HEFESTO. Ver Figura 6



Figura 5.Fases Metodología Hefesto.

b) Descripción de la fase Metodología Hefesto.

Se puede apreciar que el proceso, de forma resumida, cumple las siguientes fases:

Análisis de requerimientos: Deben identificarse los requerimientos mediante preguntas al usuario.

Análisis de OLTP: Se analizarán las fuentes de OLTP (On-Line Transaction Processing) las cuales nos ayudarán a determinar cómo debemos calcular los indicadores y establecer las respectivas correspondencias entre el modelo conceptual creado en el paso anterior y las fuentes de datos.

Modelo Lógico del DW: Se realiza el modelo lógico de la estructura del DW, tomando como base el modelo conceptual.

Integración de datos: son las reglas y políticas para su respectiva actualización.[19][20]

c) Análisis de Requerimientos

El proceso comienza con preguntas que definen los objetivos de la organización para determinar los requisitos de los usuarios. Después, se realiza un análisis exhaustivo de estas preguntas para determinar los indicadores y perspectivas que se tomarán en cuenta durante la construcción del Data Warehouse (DW). Finalmente, se crea un modelo conceptual que representa visualmente los hallazgos de esta etapa inicial. Es importante destacar que la metodología HEFESTO puede usarse simultáneamente para construir tanto un Data Warehouse como un Data Mart. En caso de que sea necesario crear dos Data Marts, la metodología debe aplicarse dos veces, una para cada uno. Además, si se están evaluando dos áreas de interés empresarial, como "Ventas" y "Compras", la metodología debe aplicarse dos veces para abordar cada una de ellas de manera específica. [19][20].

a. Identificar preguntas

El proceso de entender las necesidades de información implica la utilización de una variedad de técnicas, como entrevistas, cuestionarios y observaciones, cada una con sus propias características específicas.

El análisis de los requisitos de los diferentes usuarios constituye el primer paso vital en esta metodología, ya que son ellos quienes deben orientar la investigación hacia un desarrollo que refleje claramente las expectativas del almacén de datos en términos de funciones y características.

El objetivo principal de esta fase es identificar y comprender las necesidades de información clave a nivel estratégico que son esenciales para lograr los objetivos y estrategias de la empresa, lo que facilitará la toma de decisiones.

Es fundamental asegurarse de que esta información respalde el desarrollo de los pasos siguientes, por lo que es esencial prestar especial atención al recopilar los datos.

Para garantizar un análisis efectivo, es necesario que los resultados reflejen claramente los objetivos estratégicos establecidos por la empresa en estudio. Otra estrategia para guiar el proceso de relevamiento es centrarse en las necesidades de información relacionadas con los principales procesos de la empresa en cuestión.

La idea central es formular preguntas complejas sobre el negocio, incorporando variables de análisis consideradas relevantes, ya que estas permitirán examinar la información desde diversas perspectivas.

Un aspecto crucial a tener en cuenta es que la información debe estar respaldada de alguna manera mediante un sistema OLTP, ya que de lo contrario, no será posible desarrollar el Data Warehouse (DW).[19][20]

b. Identificar indicadores y perspectivas

Una vez que se han definido las preguntas comerciales, se deben desglosar para determinar los indicadores y perspectivas de análisis que se utilizarán. Es importante tener en cuenta que los indicadores suelen ser valores numéricos que representan específicamente lo que se desea analizar, como saldos, promedios, cantidades, sumatorias, fórmulas, entre otros.

Por otro lado, las perspectivas se refieren a las entidades mediante las cuales se pretende examinar los indicadores, con el propósito de responder a las interrogantes planteadas. Estas entidades pueden ser clientes, proveedores, sucursales, países, productos, categorías, entre otros. Es importante destacar que el tiempo suele ser una perspectiva comúnmente utilizada en este proceso. [19][20]. Ver Figura 7.

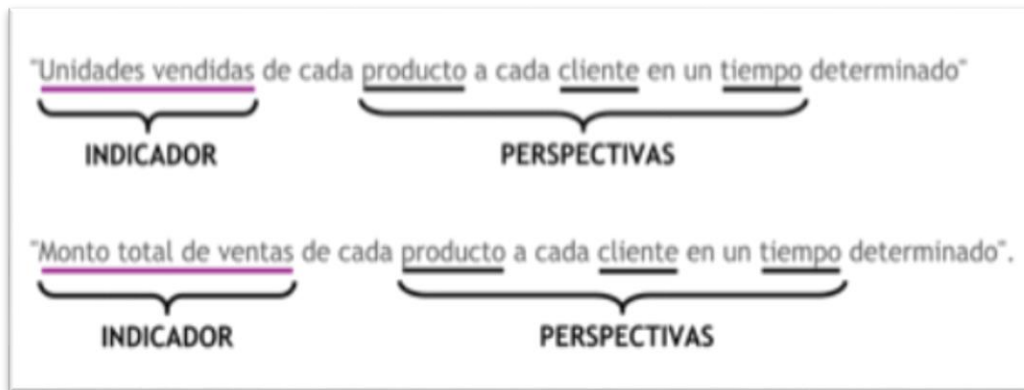


Figura 6.Caso práctico, indicadores y perspectivas.

En síntesis, los indicadores identificados son:

- Unidades vendidas.
- Monto total de ventas.

Y las perspectivas de análisis son:

- Clientes.
- Productos.
- Tiempo.

c. Modelo Conceptual

Se desarrollará un modelo conceptual en esta etapa basado en los indicadores y puntos de vista identificados en el paso anterior. El modelo conceptual proporciona una descripción detallada de la estructura de la base de datos, donde los objetos, las relaciones y las características representan la información.

Este modelo proporciona una visión clara de los objetivos del proyecto, lo que facilita el trabajo posterior. Además, al tener una definición detallada de los datos, puede presentarse fácilmente ante los usuarios y explicarse con claridad.[19][20]

El siguiente gráfico muestra la representación visual del modelo conceptual. Ver figura 8:

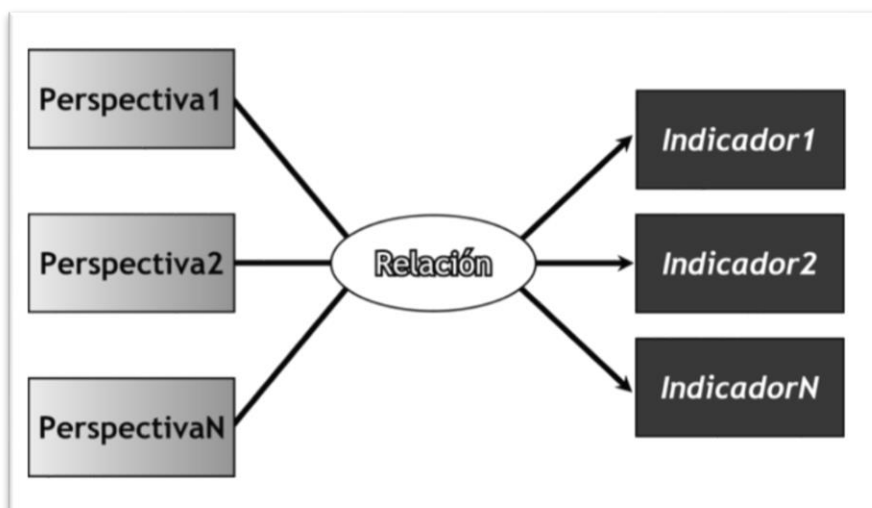


Figura 7. Modelo Conceptual.

En la representación gráfica, las perspectivas seleccionadas se ubican a la izquierda y se conectan mediante flechas a un óvalo central que representa las relaciones entre ellas, denominadas según el proceso o área de estudio seleccionada. Los indicadores, ubicados a la derecha del esquema, se extraen de esta relación.

El modelo conceptual proporciona una visión clara y fácil de entender, lo que permite comprender fácilmente los resultados esperados, las variables utilizadas para analizarlos y cómo se relacionan entre sí, como se muestra en el gráfico anterior.[19][20].

Caso práctico: El modelo conceptual resultante de los datos que se han recolectado, es el siguiente: Ver Figura 9.

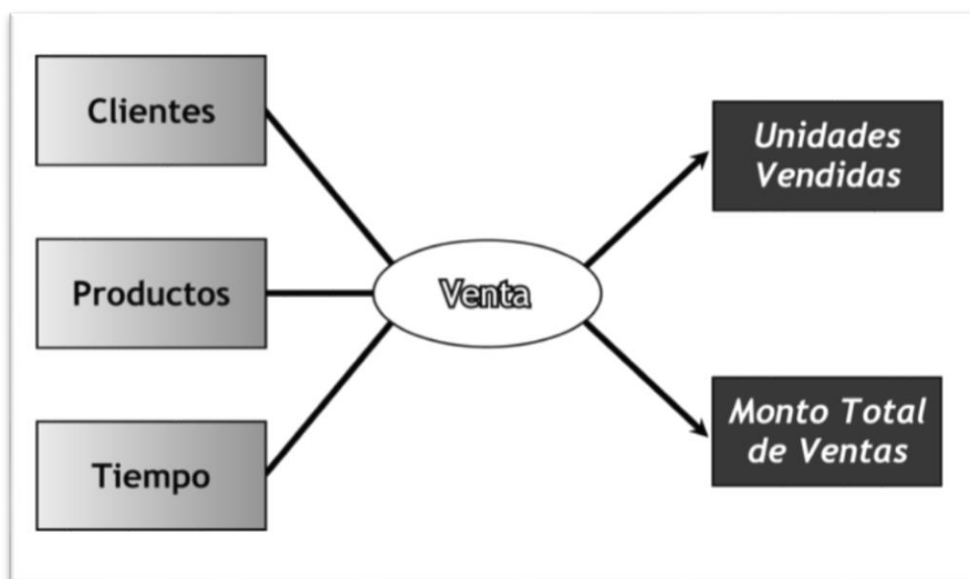


Figura 8.Caso práctico, Modelo Conceptual.

Se puede notar que la conexión entre las diversas perspectivas, que genera los indicadores solicitados por los usuarios, se identifica específicamente como "Venta".

d. Análisis de los OLPT

Para definir los métodos de cálculo de los indicadores y establecer las relaciones correspondientes entre las fuentes de datos y el modelo conceptual creado previamente, se examinarán las fuentes OLTP. Posteriormente, se detallarán los campos que se incluirán en cada perspectiva. Finalmente, los datos recopilados en esta fase mejorarán el modelo conceptual.[19][20]

a. Elaboración de Indicadores y/o métricas.

En esta etapa, es necesario detallar cómo se calcularán los indicadores, definiendo los siguientes aspectos para cada uno de ellos:

- *Hecho/s que lo componen, con su respectiva fórmula de cálculo. Por ejemplo: Hecho1 + Hecho2.*
- *Función de sumarización que se utilizará para su agregación. Por ejemplo: SUM, AVG, COUNT, etc.*

Ejemplo práctico:

Los indicadores se calcularán de la siguiente manera:

- "Unidades Vendidas":
 - Hechos: Unidades Vendidas.
 - Función de sumarización: SUM.

Nota: el indicador "Unidades Vendidas" representa la sumatoria de las unidades que se han vendido de un producto en particular.

- "Monto Total de Ventas":
 - Hechos: (Unidades Vendidas) * (Precio de Venta).
 - Función de sumarización: SUM.

Aclaración: el indicador "Monto Total de Ventas" representa la sumatoria del monto total que se ha vendido de cada producto, y se obtiene al multiplicar las unidades vendidas, por su respectivo precio.[19][20]

b. Establecer correspondencias

Para establecer relaciones entre el modelo conceptual y las fuentes de datos, en esta fase se busca revisar las bases de datos OLTP disponibles que contengan los datos necesarios, así como sus características. El objetivo es garantizar que todas las partes del modelo conceptual estén representadas correctamente en las bases de datos operacionales.

Ejemplo práctico:

Un diagrama relacional, utilizando entidades, relaciones, cardinalidades, claves, atributos y jerarquías de generalización, se utiliza en la base de datos operacional (OLTP) de la empresa examinada para representar el proceso de venta.[19][20]. Ver figura 10.

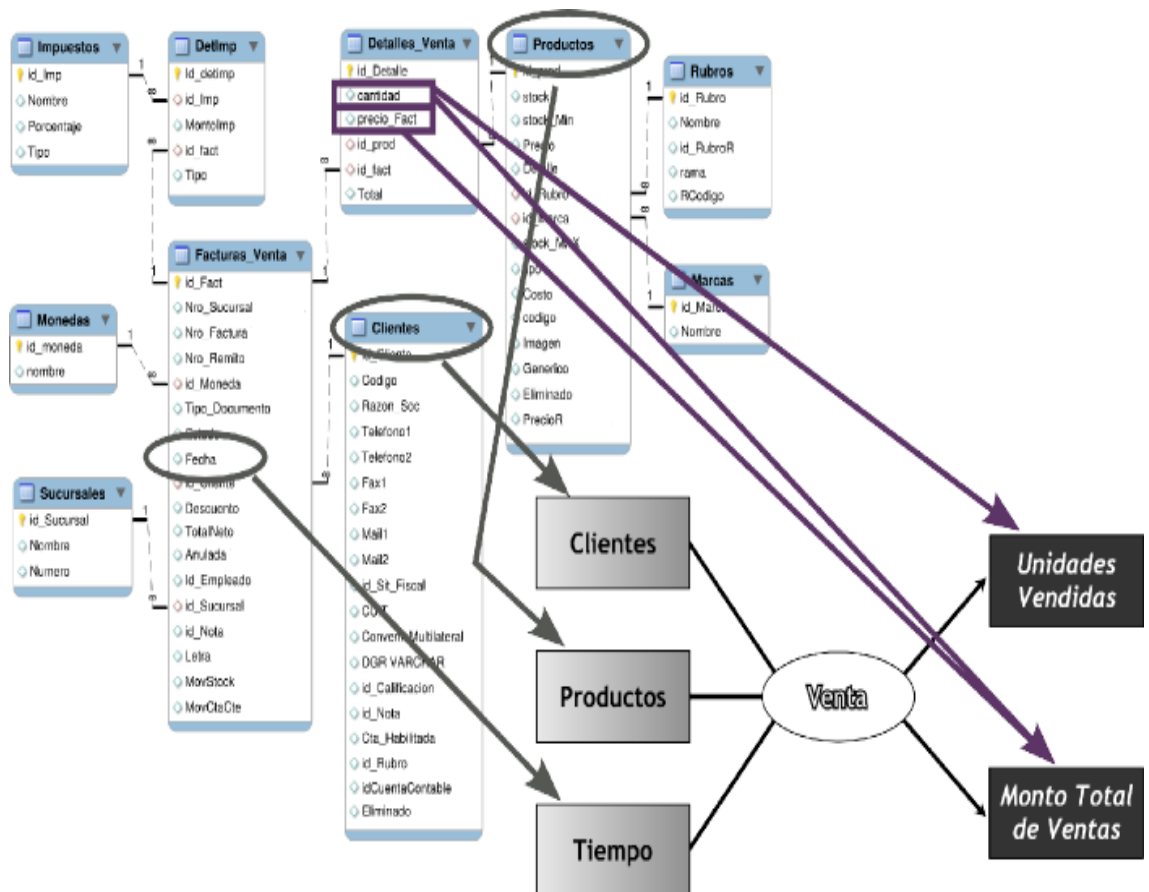


Figura 9. Ejemplo práctico, establecer correspondencia.

Las relaciones identificadas fueron las siguientes:

- tabla "Productos" se relaciona con la perspectiva "Productos".
- La tabla "Cientes" con la perspectiva "Cientes".
- El campo "fecha" de la tabla "Facturas_Venta" con la perspectiva "Tiempo" (debido a que es la fecha principal en el proceso de venta).
- El campo "cantidad" de la tabla "Detalles_Venta" con el indicador "Unidades Vendidas".
- El campo "cantidad" de la tabla "Detalles_Venta" multiplicado por el campo "precio_Fact" de la misma tabla, con el indicador "Monto Total de Ventas".

c. Nivel de granularidad

Una vez que se han establecido las conexiones con los OLTP, es necesario seleccionar los campos que contendrán cada perspectiva, ya que serán los campos utilizados para analizar y filtrar los indicadores.

Para lograr esto, basándose en las correspondencias identificadas en el paso anterior, es crucial presentar a los usuarios los datos de análisis disponibles para cada perspectiva. Es esencial tener una comprensión completa de cada campo y/o valor de los datos que se encuentran en los OLTP, por lo que es recomendable investigar su contexto, ya sea consultando diccionarios de datos, hablando con responsables del sistema o analizando los datos mismos.

Después de presentar a los usuarios los datos actuales y explicar su significado, valores potenciales y características, deben decidir cuáles de los indicadores consideran relevantes para consultar y cuáles no.

Es fundamental definir cómo se agruparán o sumarán los datos en relación con la perspectiva "Tiempo". El día de la semana, la quincena, el mes, el trimestre, el semestre y el año son algunos de sus posibles campos. Dado que esta acción determinará la granularidad de la información encontrada en el DW, se debe prestar especial atención al seleccionar los campos que integrarán cada perspectiva.[19][20]

Ejemplo práctico:

Los datos más comunes que se pueden usar con respecto a la perspectiva "Tiempo", que determinará la granularidad del depósito de datos, son los siguientes:

- Año.
- Semestre.
- Cuatrimestre.
- Trimestre.
- Número de mes.
- Nombre del mes.
- Quincena.
- Semana.
- Número de día.
- Nombre del día.

Después de recopilar toda la información relevante y consultar con los usuarios qué datos consideraban relevantes, para analizar los indicadores ya expuestos, los resultados fueron los siguientes:

- Perspectiva “Clientes”
 - “Razon_Soc” de la tabla “Clientes”. Ya que este hace referencia al nombre del cliente.
- Perspectiva “Productos”:
 - “detalle” de la tabla “Productos”. Ya que este hace referencia al nombre del producto.
 - “Nombre” de la tabla “Marcas”. Ya que esta hace referencia a la marca a la que pertenece el producto. Este campo es obtenido a través de la unión con la tabla “Productos”
- Perspectiva “Tiempo”:
 - “Mes”. Referido al nombre del mes.
 - “Trimestre”.
 - “Año”.

d. Modelo Conceptual ampliado

El modelo conceptual se extenderá en esta etapa para mostrar los resultados de los pasos previos. Se incluirán los campos seleccionados bajo cada perspectiva y las fórmulas de cálculo correspondientes bajo cada indicador. Esto se mostrará gráficamente:[19][20]. Ver figuras 11,12.

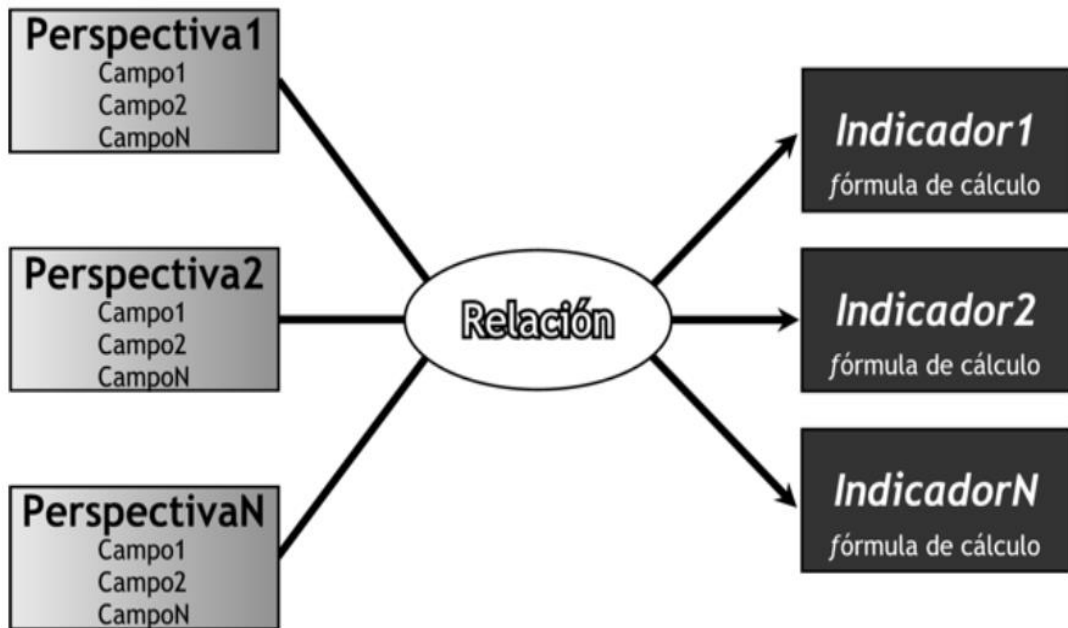


Figura 10. Modelo Conceptual ampliado.

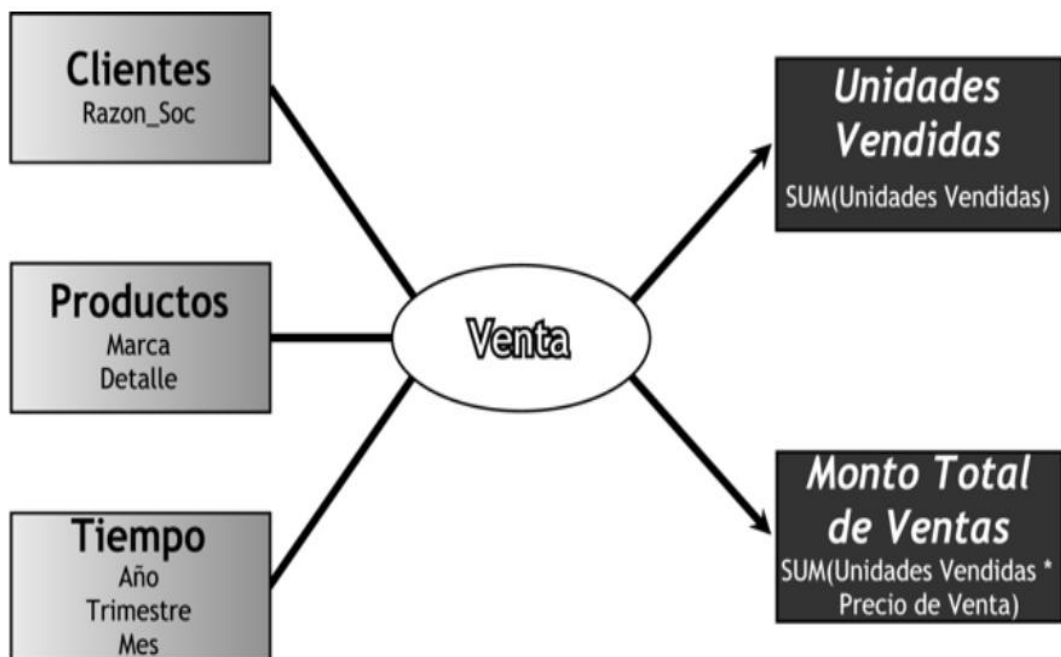


Figura 11. Caso práctico, Modelo Conceptual ampliado.

e. Modelo Lógico del DW

Se especificará el tipo de modelo a utilizar y se realizarán las tareas necesarias para diseñar las tablas de dimensiones y hechos, partiendo

del modelo conceptual creado anteriormente. Posteriormente, se crearán las relaciones necesarias entre estas tablas. El Modelo Lógico representa la estructura de datos que puede gestionar y almacenar un SGBD, que es un conjunto de programas para administrar bases de datos.[19][20]

a. **Tipo de Modelo Lógico del DW.**

Es fundamental seleccionar cuidadosamente el tipo de esquema que servirá como base para la estructura del almacén de datos, asegurándose de que se adapte a las necesidades y requisitos de los usuarios. Dado que la implementación de un esquema en una estrella, una constelación o un copo de nieve tendrá un impacto significativo en el desarrollo posterior del modelo lógico, es imperativo que se tome una decisión imparcial. [19][20]

Caso práctico:

Se empleará un esquema en estrella, debido a sus atributos distintivos, beneficios y diferencias en comparación con otros esquemas disponibles.

Tablas de dimensiones

En esta etapa, es necesario crear las tablas de dimensiones que formarán parte del Data Warehouse (DW).

Cada perspectiva definida en el modelo conceptual para los tres tipos de esquemas mencionados se convertirá en una tabla de dimensión. Para lograrlo, cada perspectiva y sus campos relacionados deben seguir los siguientes pasos:

- Se seleccionará un nombre que identifique claramente la tabla de dimensión.
- Se incluirá un campo que actúe como su clave principal.
- Se modificarán los nombres de los campos si no resultan suficientemente claros o intuitivos..[19][20]

Gráficamente. Ver Figuras 13,14,15,16.

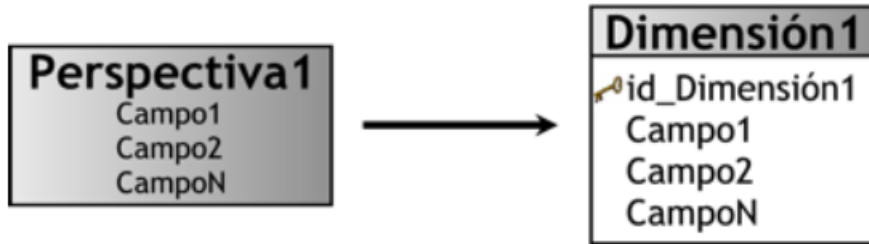


Figura 12. Diseño de tablas de dimensiones.

Caso Practico Perspectiva Productos se transforma en Tabla Dimensión Producto

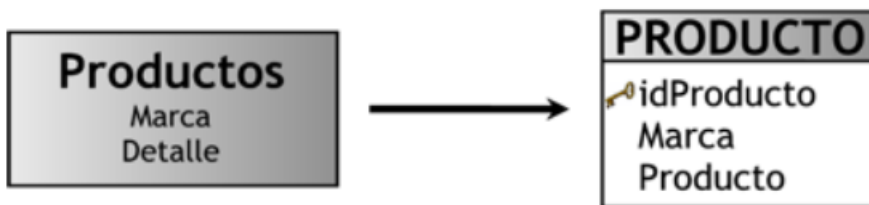


Figura 13. Caso práctico, tabla de dimensión" PRODUCTO".

Caso Practico Perspectiva Clientes se transforma en Tabla Dimensión Cliente

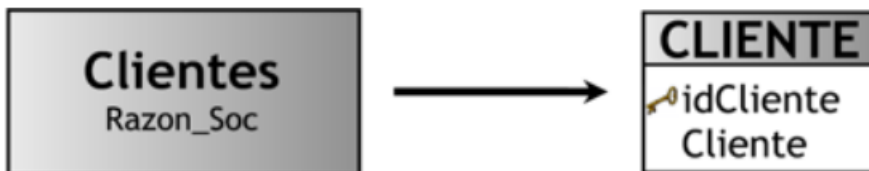


Figura 14. Caso práctico, tabla de dimensión" CLIENTE".

Caso Practico Perspectiva Tiempo se transforma en Tabla Dimensión Fecha

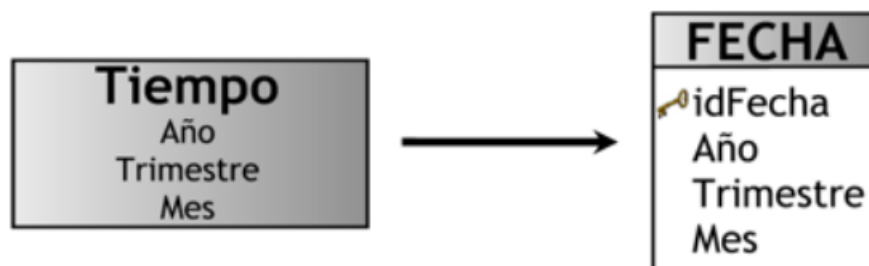


Figura 15. Caso práctico, tabla de dimensión" FECHA".

b. Tablas de hechos

Para construir los indicadores de análisis, en esta etapa se definirán las tablas de hechos.

Para los esquemas en estrella y copo de nieve, se llevarán a cabo los siguientes pasos:

- Se asignará un nombre a la tabla de hechos que describa la naturaleza de la información analizada, el área de estudio, el enfoque empresarial u otros aspectos pertinentes.
- Se establecerá su clave primaria, que será una combinación de las claves primarias de cada tabla de dimensión relacionada.
- Se crearán campos de hechos de acuerdo con los indicadores definidos en el modelo conceptual, utilizando los mismos nombres si se prefiere, o cualquier otro nombre deseado..[19][20]. Ver figura 17.

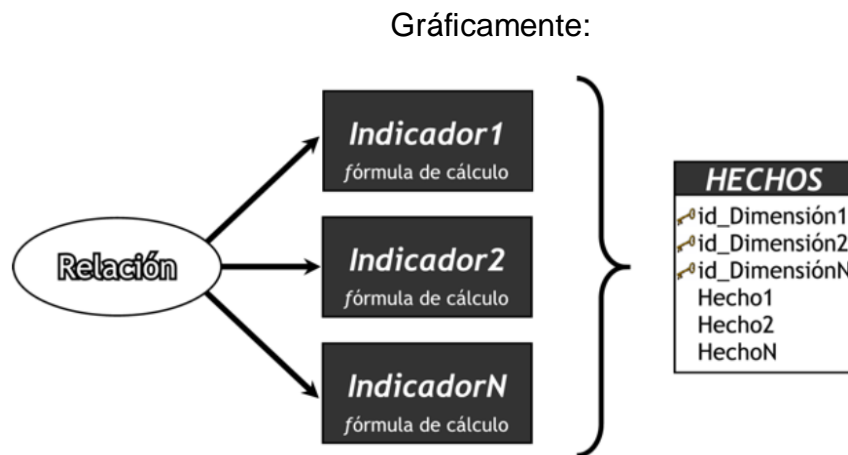


Figura 16. Tabla de hechos.

Caso práctico:

A continuación, se confeccionará la tabla de hechos:

- A la tabla de hechos se le asignará el nombre “VENTAS”.
- Su clave principal será la combinación de las claves principales de las tablas de dimensiones antes definidas: “idCliente”, “idProducto” e “idFecha”.

- Se crearán dos hechos, que se corresponden con los dos indicadores y serán renombrados, “Unidades Vendidas” por “Cantidad” y “Monto Total de Ventas” por “MontoTotal”.

En el gráfico siguiente se puede apreciar mejor este paso. Ver figura 18.

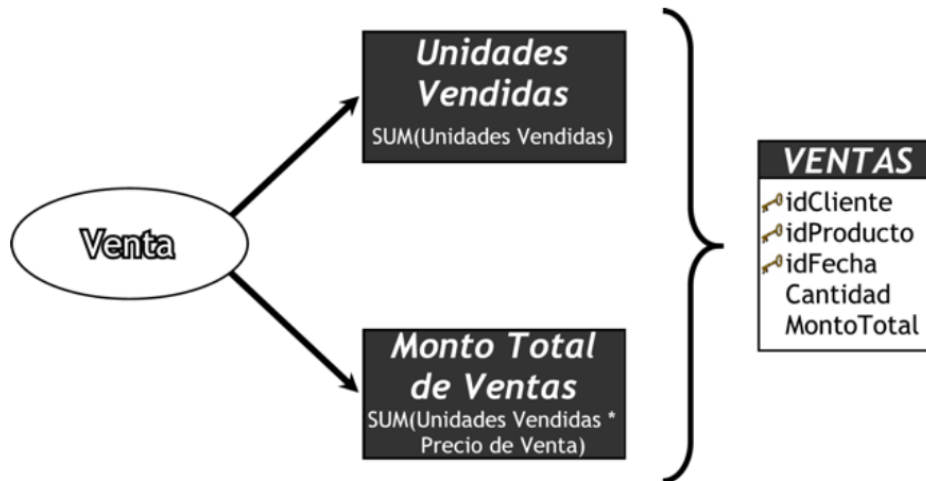


Figura 17.Caso práctico, diseño de la tabla de hechos.

c. Uniones

Para cada uno de los tres tipos de esquemas, se realizarán las conexiones adecuadas entre las tablas de dimensiones y las tablas de hechos.[19][20].

Caso práctico:

Se realizarán las uniones pertinentes, de acuerdo corresponda:

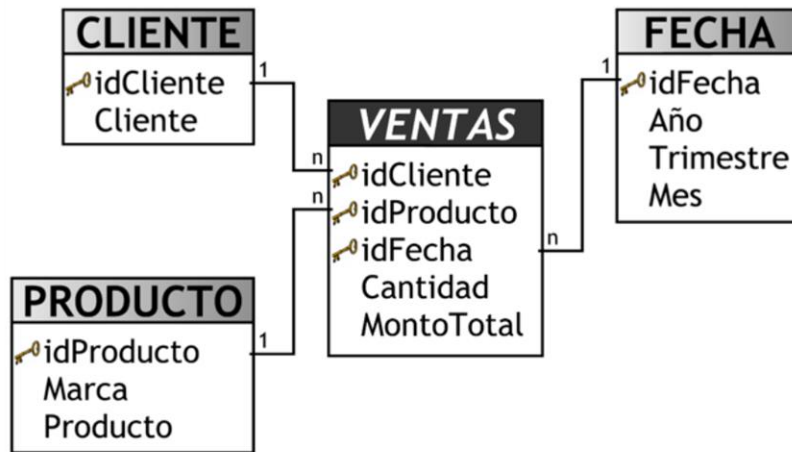


Figura 18.Caso práctico, uniones.[19][20]

d) Integración de Datos

Será esencial agregar información al modelo lógico una vez que se haya creado. Esto requiere la implementación de procedimientos para limpiar y garantizar la calidad de los datos, así como la implementación de procesos ETL (Extracción, Transformación y Carga) u otras técnicas apropiadas. Posteriormente, se establecerán normas y pautas para mantener los datos actualizados de manera constante, así como mecanismos para realizar esas actualizaciones de manera eficiente y precisa.[19][20].

a. Carga Inicial

Este paso implica comenzar cargando el Data Warehouse (DW) con el modelo de datos que previamente hemos construido. Entre otras cosas, esto implica una serie de tareas importantes, como la limpieza y garantía de la calidad de los datos y la ejecución de procesos ETL.

En algunos casos, estas actividades pueden tener una lógica bastante complicada. Afortunadamente, hay muchas herramientas de software disponibles en la actualidad que pueden simplificar significativamente este proceso.

Evitar la carga de valores innecesarios o extraños en el DW es crucial. Para lograrlo, es necesario establecer condiciones y restricciones que aseguren que los datos pertinentes sean utilizados exclusivamente.

Es importante tener en cuenta que diferentes tablas de hechos comparten varias tablas de dimensiones cuando trabajamos con un esquema en

estrella. Esto puede causar conflictos cuando las restricciones o condiciones de análisis de una tabla de dimensiones entran en conflicto con otras restricciones o condiciones de análisis de otras tablas de dimensiones.

El proceso de carga de datos debe comenzar con las dimensiones y luego seguir con las tablas de hechos, siempre asegurándose de que cada elemento tenga la correspondencia correcta. En caso de utilizarse un esquema de copo de nieve, las tablas de dimensiones se cargarán siguiendo las jerarquías establecidas, desde el nivel más general hasta el más detallado.

Es fundamental registrar minuciosamente las acciones llevadas a cabo con los diversos programas utilizados. Por ejemplo, es común que los sistemas ETL funcionen con "pasos" y "relaciones", cada uno de los cuales realiza una tarea específica dentro del proceso ETL y indica la dirección del flujo de datos. En este sentido, es crucial proporcionar una descripción detallada tanto del proceso general como de cada paso y/o relación, brindando una imagen completa y detallada del proceso en su totalidad.

Finalmente, es posible utilizar pre-agregaciones al cargar los datos en las tablas de hechos, ya sea a nivel de granularidad de la misma tabla o de cualquier otro nivel, según sea necesario.

La siguiente es la estructura del proceso ETL diseñado para la carga inicial:[19][20]

Las tareas que lleva a cabo este proceso son:

- Inicio: inicia la ejecución de los pasos en el momento en que se le indique.
- Establecer variables Fecha_Desde y Fecha_Hasta: establece dos variables globales que serán utilizadas posteriormente por algunos pasos.
 - Para la variable "Fecha_Desde" se obtiene el valor de la fecha en que se realizó la primera venta.

- Para la variable "Fecha_Hasta" se obtiene el valor de la fecha actual.
- Carga de Dimensión CLIENTE: ejecuta el contenedor de pasos que cargará la dimensión CLIENTE, más adelante se detallará el mismo.
- Carga de Dimensión PRODUCTO: ejecuta el contenedor de pasos que cargará la dimensión PRODUCTO, más adelante se detallará el mismo.
- Carga de Dimensión FECHA: ejecuta el contenedor de pasos que cargará la dimensión FECHA, más adelante se detallará el mismo.
- Carga de Tabla de Hechos VENTAS: ejecuta el contenedor de pasos que cargará la tabla de hechos VENTAS, más adelante se detallará el mismo.

Existe una variedad de herramientas y programas de software disponibles para crear esta tabla de dimensión, que es esencial en cualquier Data Warehouse (DW). Sin embargo, es posible completar el proceso manualmente o utilizando algún software, ingresando los datos en un archivo, tabla, hoja de cálculo, etc., y luego exportándolos según sea necesario.

En este caso particular, se ha implementado un procedimiento que realiza lo siguiente:

- Recibe como parámetros los valores de "Fecha_Desde" y "Fecha_Hasta".
- Itera a través de cada fecha comprendida dentro de este intervalo.
- Analiza cada fecha y lleva a cabo una serie de operaciones para generar los valores de los campos de la tabla de la dimensión FECHA.:

Carga de Tabla de Hechos VENTAS".

b. Actualización

Es esencial establecer normas y planes para la actualización de datos después de que se haya cargado por completo el Data Warehouse (DW). Después de este paso, se deben realizar las siguientes acciones:

- Definir las tareas específicas relacionadas con la extracción, transformación y carga (ETL) de los datos para mantener actualizados los datos del DW.
- Determinar de manera precisa los procedimientos de cada módulo que participa en este proceso de Business Intelligence (BI), manteniendo un nivel adecuado y coherente de especificidad. [19][20].

e) Visualización y evaluación

La visualización de la información sigue a la fase de integración de datos, lo que amplía el uso de la metodología HEFESTO. Esto ayuda a los usuarios a tomar decisiones al permitirles monitorear mejor los indicadores clave de rendimiento (KPIs) de la organización.[19][20]

f) System Usability Scale

El Sistema de Escalas de Usabilidad (SUS) es un cuestionario estandarizado que se utiliza para evaluar la calidad y la usabilidad de un proyecto o producto digital. El SUS se enfoca en tres elementos principales:

- Eficacia: ¿Los usuarios pueden lograr sus objetivos de manera exitosa utilizando el sistema?
- Eficiencia: ¿Cuánto esfuerzo se requiere para alcanzar estos objetivos?
- Satisfacción: ¿El uso del sistema resulta satisfactorio para los usuarios?

g) Cálculo SUS

El SUS consta de 10 enunciados predefinidos, divididos equitativamente entre positivos y negativos, numerados del 1 al 10. La elección de enunciados positivos y negativos es crucial para obtener un resultado final significativo. Los participantes evalúan estos enunciados utilizando la Escala de Likert, que ofrece 5 opciones graduadas:

- Totalmente en desacuerdo

- En desacuerdo
- Neutro
- De acuerdo
- Totalmente de acuerdo

Estas respuestas brindan una evaluación detallada de la experiencia del usuario con el sistema evaluado. La escala del método SUS se define a continuación:[21]

Tal como se visualiza en la Figura 20, la lectura de los resultados obtenidos en el SUS, básicamente califica cuan aceptable es nuestro aplicativo para el usuario, mismos resultados son comparados con el resultado promedio. Además, se considera como aceptable cuando los resultados están sobre el 70; mientras que será un inaceptable si esta menos de 50; por lo tanto, si los resultados están entre 50 y 70 se considerarían como solo aceptables.

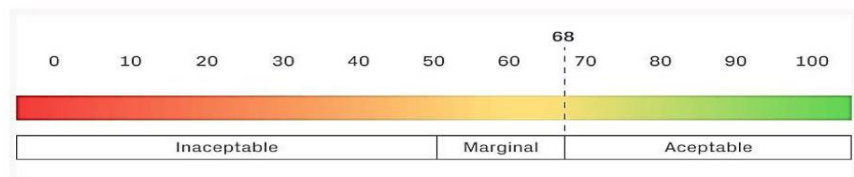


Figura 19. Escala SUS.

3. DESARROLLO E IMPLEMENTACIÓN.

Para el desarrollo e implementación de este proyecto, se decidió utilizar la metodología Hefesto. Esta decisión se basa en la intención de construir datamarts segmentados por áreas de la institución. Es relevante mencionar que inicialmente se enfocará en la automatización del área de producción y costos.

3.1. Implementación de Fases de HEFESTO.

Para implementar Business Intelligence, seguiremos los pasos establecidos por la metodología Hefesto. Se comenzará con el análisis de los requisitos del área de producción, donde se examinarán los indicadores y métricas, junto con sus respectivas perspectivas o tablas de datos, como productos y ventas. Después, se desarrollará el modelo conceptual y se procederá al análisis del OLTP.

Para este análisis, comenzaremos definiendo los indicadores y sus correspondientes fórmulas de cálculo. Luego, se establecerán las tablas de la base de datos desde las cuales se accederá a los datos, determinando su nivel de detalle según lo requerido. A continuación, ampliaremos el modelo conceptual, definiendo únicamente los campos necesarios para el desarrollo de BI.

Después de completar estos pasos, se utilizará el modelo estrella para definir el modelo lógico del almacén de datos. Para su uso en BI, en este modelo determinaremos las tablas de dimensiones como cliente y producto, junto con sus campos correspondientes. Además, crearemos tablas de hechos con identificadores de dimensiones y métricas e indicadores relevantes. Estableceremos relaciones de uno a varios entre las tablas de dimensiones y las tablas de hechos para realizar las uniones correspondientes.

Finalmente, se integrarán los datos. Esto incluirá la carga inicial de los datos, la definición de la periodicidad de actualización y el orden para mantener la integridad. Estas acciones son esenciales para crear una solución que satisfaga las necesidades de la institución para que sea evaluada adecuadamente. Después, crearemos el modelo de tablas con hechos y

dimensiones, usaremos procedimientos almacenados en la herramienta SQL Developer para cargar las tablas y luego crearemos el modelo de visualización en la herramienta Power BI.

3.2. Análisis de Requerimientos.

El Cuerpo de Ingenieros del Ejército (C.E.E.) es reconocido nacional e internacionalmente por su experiencia en la construcción civil, vial y servicios petroleros. Sin embargo, se han identificado varios desafíos en sus procesos gerenciales, como la repetición de información en informes de Excel extraídos manualmente, la falta de un sistema de consultas gerenciales en tiempo real, y la ausencia de predicciones basadas en datos históricos para futuros proyectos. Además, análisis importantes como rentabilidad y avances de proyectos se realizan manualmente, al igual que los prorrateos de gastos administrativos. Se lleva un seguimiento de pagos comprometidos en Excel, lo que sugiere la necesidad de una solución informática integral para mejorar la eficiencia y precisión en la gestión de proyectos. Se propone implementar un sistema de Business Intelligence (BI) para análisis gerencial, aprovechando la data existente en el Sistema Integrado del Cuerpo de Ingenieros del Ejército (SICEE). Este proyecto se enfocará en los procesos centrales del CEE, con énfasis en la construcción civil, vial y servicios petroleros. Se accederá al SICEE para recopilar datos relevantes y se utilizará Oracle 12c como base de datos. Los procesos clave serán Monitoreo Económico de proyectos, Monitoreo Administración de Proyectos, Monitoreo de Obras por Contrato, Monitoreo de Obras por Rubro

El objetivo es estandarizar la visualización de datos para facilitar la interpretación y análisis de información a través de dashboards clasificados según la lógica del negocio.

a. Identificar Preguntas

Para un levantamiento de requerimientos efectivo, es fundamental realizar la identificación de las preguntas clave del negocio. Esto permitirá entender de manera precisa las necesidades del proyecto de Business Intelligence (BI).

Tabla 3 identificar Preguntas

Número	Preguntas
1	Valores de producción de cada unidad ejecutora, centros de costos, estado de centros de costos en un tiempo determinada
2	Valores de planillaje técnico legalizado de cada unidad ejecutora, centros de costos, estado de centros de costos en un tiempo determinada
3	Valores de planillaje contabilizado de cada unidad ejecutora, centros de costos, estado de centros de costos en un tiempo determinada
4	Valores de costos de producción de cada unidad ejecutora, centros de costos, estado de centros de costos en un tiempo determinada
5	Valores de subcontratos no contabilizados de cada unidad ejecutora y centros de costos en un tiempo determinado
6	Valores de gastos matriz administrativos de cada unidad ejecutora y centros de costos en un tiempo determinado
7	Valores de margen operacional de cada unidad ejecutora y centros de costos en un tiempo determinado
8	Valores de superávit y/o déficit de cada unidad ejecutora y centros de costos en un tiempo determinado
9	Valores de producción de cada unidad ejecutora y centros de costos en un tiempo determinada, desglosado hasta el nivel de rubro
10	Valores de planillaje de cada unidad ejecutora y centros de costos en un tiempo determinada, desglosado hasta el nivel de rubro
11	Valores de cronograma de cada unidad ejecutora y centros de costos en un tiempo determinada, desglosado hasta el nivel de rubro

b. Identificar Indicadores y perspectivas

Basándonos en las interrogantes más relevantes para la construcción de BI, debemos ahora definir los indicadores y métricas. Los indicadores suelen ser valores numéricos que nos proporcionan información clave. Por otro lado, las perspectivas representan las diversas posibles fuentes de donde se extraerán estas métricas, e incluso, son la base desde la cual se pueden desglosar y agrupar los datos.

Tabla 4 Indicadores y perspectivas

Número	Indicadores	Perspectiva
1	Sumatoria total de Valores de producción	<u>unidad ejecutora, centros de costos, estado de centros de costos</u> en un <u>tiempo</u> determinada
2	Sumatorio del total Valores de planillaje técnico legalizado	<u>unidad ejecutora, centros de costos, estado de centros de costos</u> en un <u>tiempo</u> determinada
3	Sumatorio del total Valores de planillaje contabilizado	<u>unidad ejecutora, centros de costos, estado de centros de costos</u> en un <u>tiempo</u> determinada
4	Sumatorio del total Valores de costos de producción	<u>unidad ejecutora, centros de costos, estado de centros de costos</u> en un <u>tiempo</u> determinada
5	Sumatorio del total Valores de subcontratos no contabilizados	<u>unidad ejecutora y centros de costos estado de centros de costos</u> en un <u>tiempo</u> determinado

6	Sumatorio del total Valores de gastos matriz administrativos	<u>unidad ejecutora y centros de costos estado de centros de costos</u> en un <u>tiempo</u> determinado
7	Sumatorio del total Valores de margen operacional	<u>unidad ejecutora y centros de costos estado de centros de costos</u> en un <u>tiempo</u> determinado
8	Sumatorio del total Valores de superávit y/o déficit	<u>unidad ejecutora y centros de costos estado de centros de costos</u> en un <u>tiempo</u> determinado
9	Sumatorio del total Valores de producción	<u>Unidad ejecutora y centros de costos</u> en un <u>tiempo</u> determinada, desglosado hasta el nivel de <u>rubro</u>
10	Sumatorio del total Valores de planillaje Legalizado	<u>Unidad ejecutora y centros de costos</u> en un <u>tiempo</u> determinada, desglosado hasta el nivel de <u>rubro</u>
11	Sumatorio del total Valores de cronograma	<u>Unidad ejecutora y centros de costos</u> en un <u>tiempo</u> determinada, desglosado hasta el nivel de <u>rubro</u>

c. Modelo Conceptual

Dentro del modelo conceptual, se determinan los procesos o data marts que serán tratados en la solución de inteligencia empresarial (BI). También se indican los nombres de las fuentes potenciales de donde se extraerán los datos y se especifican las métricas que se visualizarán en cada proceso.

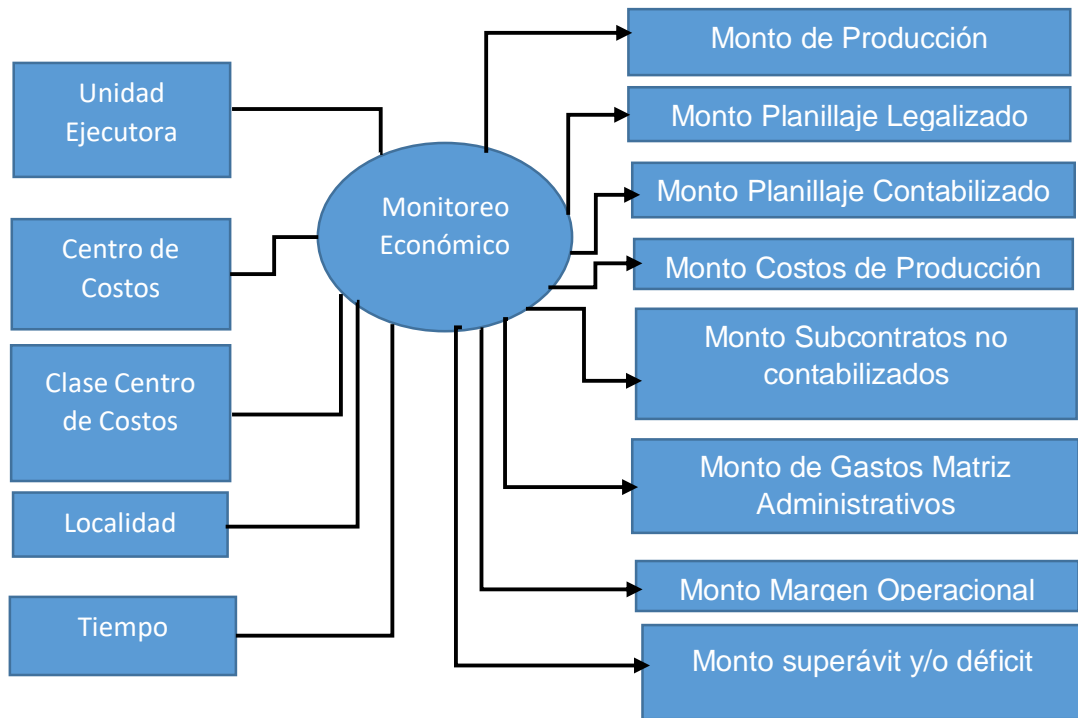


Figura 20 Modelo Conceptual Monitoreo Económico

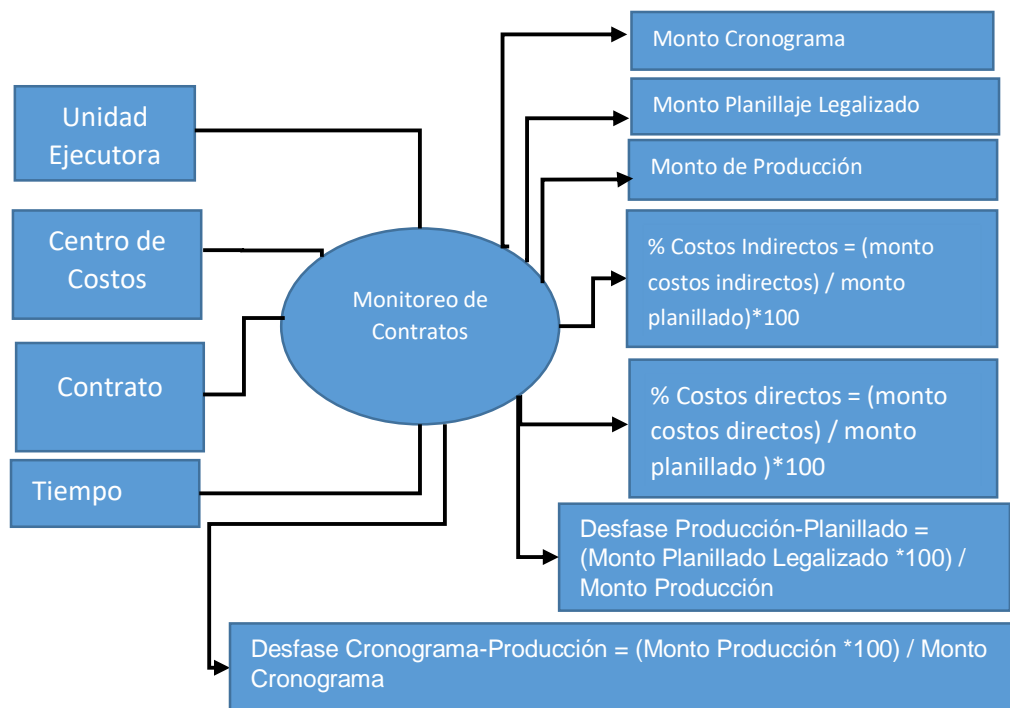


Figura 21 Modelo Conceptual Monitoreo de Contratos

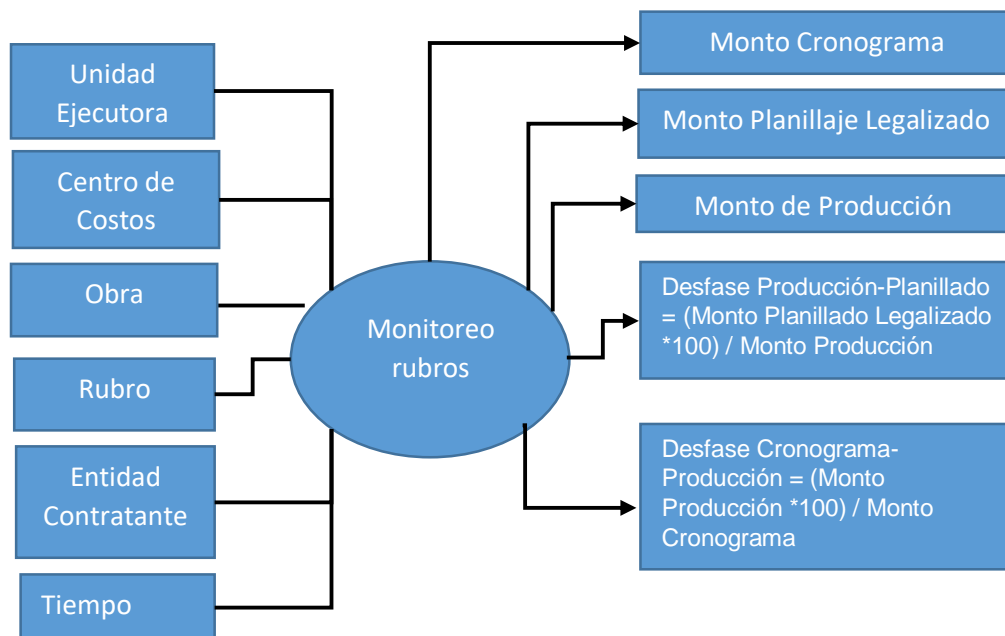


Figura 22 Modelo Conceptual Monitoreo de Rubro.

3.3. Análisis de los OLTP.

Esta fase, es crucial definir los indicadores y las funciones de cálculo correspondientes, como sumatoria, distintos, contar, entre otras, que se utilizarán para su agregación. Además, se establecerán las relaciones entre la base de datos y las perspectivas, indicadores y/o métricas obtenidas en el paso anterior. Seguidamente, definiremos el nivel de granularidad de cada perspectiva; por ejemplo, en la perspectiva de tiempo se determinarán los niveles de año, mes, etc. Con todas estas actividades completadas, generaremos el diagrama conceptual ampliado, en el cual se incluirán las perspectivas con sus respectivos campos y las fórmulas adecuadas para los indicadores/métricas.

a) Conformar Indicadores

Debemos proporcionar una descripción detallada de las fórmulas de cálculo y las funciones de sumarización, resta o contabilización de cada uno de los indicadores y métricas. Estas estarán incluidas dentro de la tabla de hechos, donde se almacenan las métricas.

Tabla 5. Indicadores de Monitoreo Económico.

Indicador / Métrica	Hechos	Función
Total Monto de producción por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto de Producción por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Monto de Producción por unidad ejecutora, centro costos, estado centro costos por año, mes)
Total Monto planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Montos Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos por año, mes)
Total Monto planillaje Contabilizado por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto planillaje Contabilizado por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (planillaje Contabilizado por unidad ejecutora, centro costos, estado centro costos por año, mes)
Total Monto de costos de producción por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto de costos de Producción por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Monto de costos de Producción por unidad ejecutora, centro costos, estado centro costos por año, mes)

Total Monto de Subcontratos no Contabilizados por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto de Subcontratos no Contabilizados por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Monto de Subcontratos no Contabilizados por unidad ejecutora, centro costos, estado centro costos por año, mes)
Total Monto de Gastos Matriz por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto de Gastos Matriz por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Monto de Gastos Matriz por unidad ejecutora, centro costos, estado centro costos por año, mes)
Total Monto Margen Operacional por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Margen Operacional por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Margen Operacional por unidad ejecutora, centro costos, estado centro costos por año, mes)
Total Monto Superávit y/o Déficit por unidad ejecutora, centro costos, estado centro costos en un tiempo determinado	Total Monto Superávit y/o Déficit por unidad ejecutora, centro costos, estado centro costos por año, mes	Sumatorio (Monto Superávit y/o Déficit por unidad ejecutora, centro costos, estado centro costos por año, mes)

Tabla 6. Monitoreo de Contratos

Indicador / Métrica	Hechos	Función
Total Monto de Cronograma por unidad ejecutora, centro costos, estado centro costos, contrato, en un tiempo determinado	Total Monto de Cronograma por unidad ejecutora, centro costos, estado centro costos , contrato por año, mes	Sumatorio (Monto de Cronograma por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes)
Total Monto de Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Total Monto de Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes	Sumatorio (Monto de Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes)
Total Monto de Producción por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Total Monto de Producción por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes	Sumatorio (Monto Producción por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes)
Total Monto de Costos Producción Directos e Indirectos por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Total Monto de Costos Producción Directos e Indirectos por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes	Sumatorio (Monto de Costos Producción Directos e Indirectos por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes)
Total Porcentaje Planillado Legalizado por unidad ejecutora, centro costos, estado	% Costos Indirectos = (monto costos indirectos) / monto planillado)*100	Sumatorio (Porcentaje Costos Indirectos por unidad ejecutora, centro costos, estado

centro costos, contrato en un tiempo determinado		centro costos, contrato por año, mes).	
Total Producción ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Porcentaje por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	% Costos directos = (monto costos directos) / monto planillado) * 100	Sumatorio (Porcentaje Costos directos por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes).
Total Cronograma por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Porcentaje por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Desfase Cronograma-Producción = (Monto Producción * 100) / Monto Cronograma	Sumatorio (Desfase Cronograma-Producción por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes).
Total Planillado por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Porcentaje por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Desfase Producción-Planillado = (Monto Planillado * 100) / Monto Producción	Sumatorio (Desfase Producción-Planillado por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes).

Tabla 7 Monitoreo de Rubro

Indicador / Métrica	Hechos	Función
Total Monto de Cronograma por unidad ejecutora, centro costos, estado centro costos, contrato, obra, capitulo, rubro en un tiempo determinado	Total Monto de Cronograma por unidad ejecutora, centro costos, estado centro costos , obra, capitulo, rubro por año, mes	Sumatorio (Monto de Cronograma por unidad ejecutora, centro costos, estado centro costos, obra, capitulo, rubro por año, mes)

Total Monto de Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos, contrato, obra, capitulo, rubro en un tiempo determinado	Total Monto de Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos, obra, capitulo, rubro por año, mes	Sumatorio (Monto de Planillaje Legalizado por unidad ejecutora, centro costos, estado centro costos, obra, capitulo, rubro por año, mes)
Total Porcentaje Cronograma por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Desfase Cronograma-Producción = (Monto Producción*100) / Monto Cronograma	Sumatorio (Desfase Cronograma-Producción por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes).
Total Porcentaje Planillado por unidad ejecutora, centro costos, estado centro costos, contrato en un tiempo determinado	Desfase Producción-Planillado = (Monto Planillado *100) / Monto Producción	Sumatorio (Desfase Producción-Planillado por unidad ejecutora, centro costos, estado centro costos, contrato por año, mes).

b) Establecer Correspondencias.

Esta etapa tiene como objetivo identificar las fuentes de datos de los sistemas transaccionales (OLTP) y establecer una conexión entre ellas y el modelo conceptual. Esto permitirá obtener los datos necesarios para el análisis.

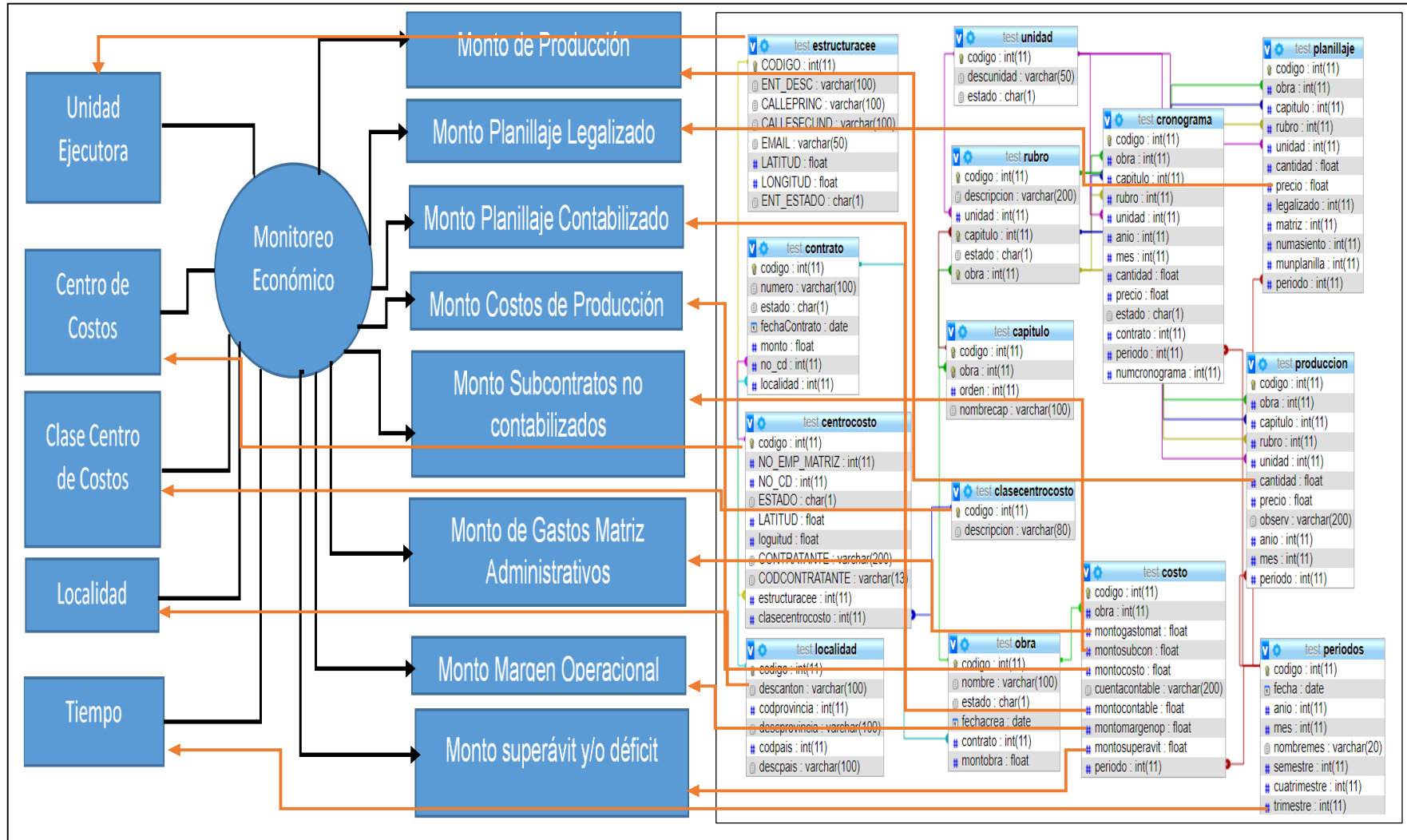


Figura 23 Correspondencia de Monitoreo Económico

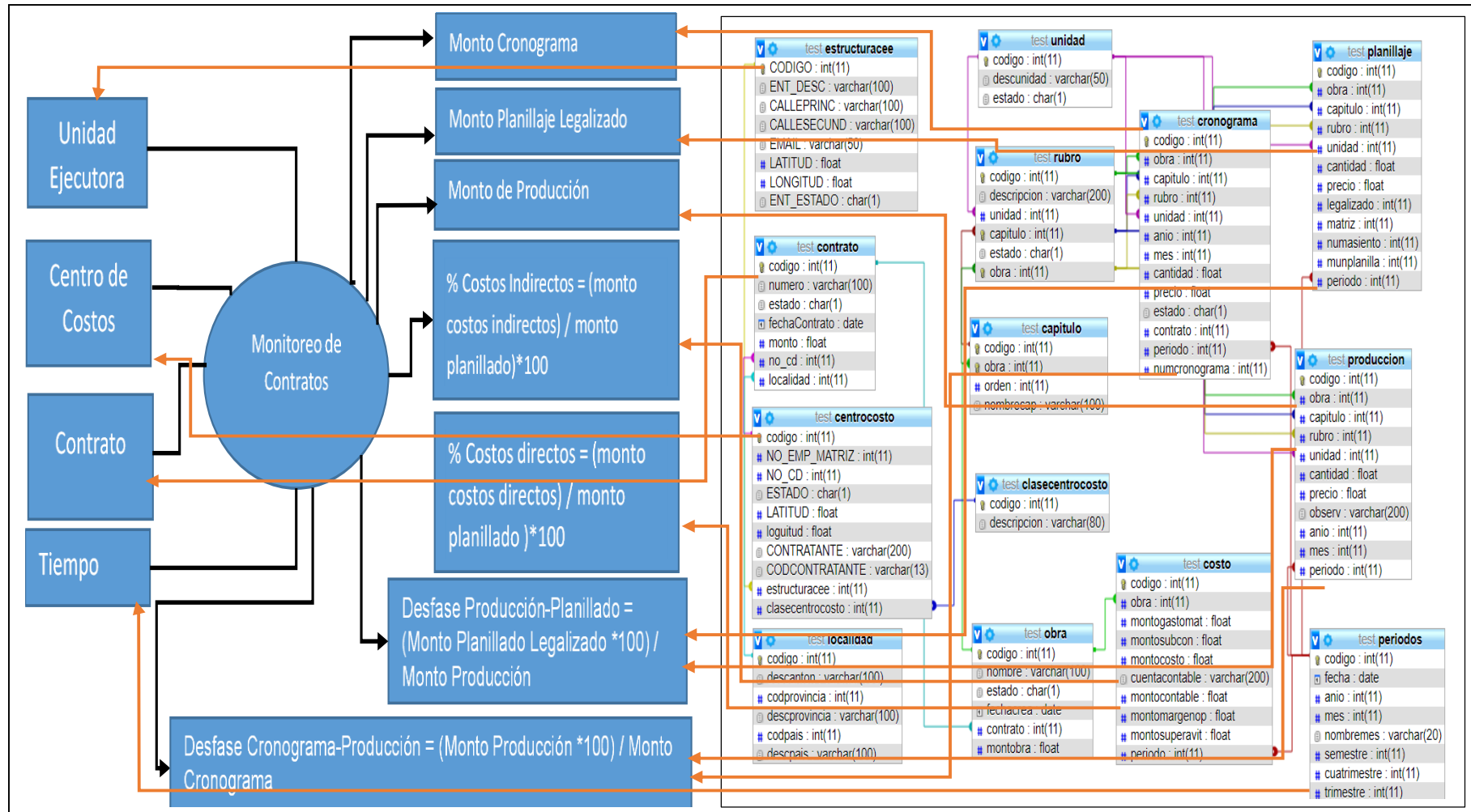


Figura 24 Correspondencia Monitoreo Proyectos Centro de Costos

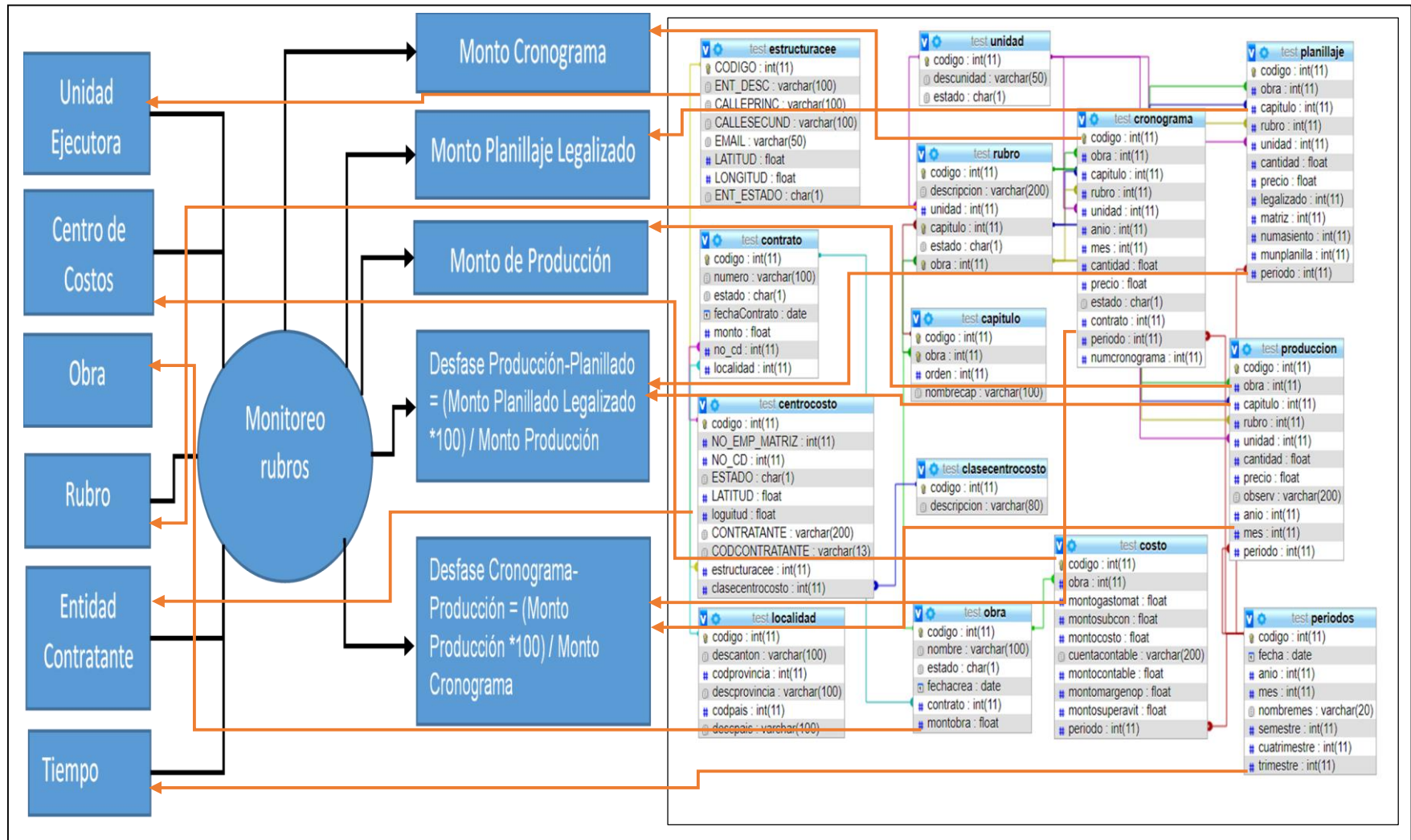


Figura 25 Correspondencia Monitoreo Rubros

c) Nivel de Granularidad.

En esta fase, se procederá a identificar los campos y a describir el contenido de cada uno, con el propósito de comprender claramente la naturaleza de los datos, los cuales corresponden a diferentes perspectivas o dimensiones. Estos campos serán esenciales para responder a las preguntas planteadas inicialmente, ya que permitirán aplicar los filtros necesarios para la búsqueda de información.

Tabla 8. EstructuraCEE

EstructuraCEE	
CODIGO	clave primaria de la tabla.
ENT_DESC	Descripción de la Unidad ejecutora de construcción
CALLEPRINC	Calle Principal de ubicación correspondiente a la unidad ejecutora de construcción.
CALLESECUND	Calle Secundaria de ubicación correspondiente a la unidad ejecutora de construcción.
LATITUD	Dato de la latitud de ubicación de Unidad Ejecutora
LONGITUD	Dato de la longitud de ubicación de Unidad Ejecutora
ESTADO	Estado de la latitud de ubicación de Unidad Ejecutora (ejecución, cerrada, proceso cierre, liquidación)

Tabla 9. Clase Centro de Costos

ClaseCentroCosto	
CODIGO	Clave primaria de la tabla.
DESCRIPCION	Descripción de la clase del centro de costo

Tabla 10. Localidad

Localidad	
CODIGO	Clave primaria de la tabla.
DESCANTON	Descripción del cantón
CODPROVINCIA	Descripción de la Provincia
DESCPROVINCIA	Descripción de provincia
CODPAIS	Código del País
DESCPAIS	Descripción de País

Tabla 11. Periodos

Periodos (fechas de búsqueda Usuario)	
CODIGO	Clave primaria de la tabla.
FECHA	Fecha
ANIO	Descripción año
SEMESTRE	semestre
CUATRIMESTRE	cuatrimestre
TRIMESTRE	trimestre
MES	mes
NOMBRE_MES	Descripción del mes

Tabla 12. Fecha de Corte

Fecha de Corte (fechas de mes actual y mes actual BBD)	
PE_PK_CODIGO	Clave primaria de la tabla.
PE_PK_FECHA	Fecha
PE_ANIO	Descripción año
PE_SEMESTRE	semestre
PE_CUATRIMESTRE	cuatrimestre
PE_TRIMESTRE	trimestre
PE_MES	mes
PE_NOMBRE_MES	Descripción del mes

Tabla 13. Rubro

Rubro	
CODIGO	Clave primaria de la tabla.
DESCRIPCION	Descripción del Rubro
UNIDAD	Descripción Unidad de medida del rubro
CAPITULO	Clave foránea de capitulo
ESTADO	Estado de rubro (vigente, cancelado)
OBRA	Clave foránea de obra

Tabla 14. Capitulo

Capitulo	
CODIGO	Clave primaria de la tabla.
OBRA	Clave foránea de Obra
ORDEN	Orden de ejecución de capitulo
NOMBRECAP	Nombre de Capitulo

Tabla 15. Unidad

Unidad	
CODIGO	Clave primaria de la tabla.
DESCUNIDAD	Descripción de la unidad
ESTADO	Estado del registro (Vigente, Cancelado)

Tabla 16. Centro de Costos

CentroCosto	
CODIGO	Clave primaria de la tabla.
NO_EMP_MATRIZ	Código de la empresa matriz
NO_CD	Número del Centro de costos secuencial
ESTADO	Estado centro de costos (ejecución, cerrado, en proceso de liquidación, Mantenimiento)
LATITUD	Dato de Latitud ubicación del centro de costos
LONGITUD	Dato de Longitud ubicación del centro de costos
CONTRATANTE	Descripción de la entidad Contratante
CODCONTRATANTE	Código de la entidad Contratante
ESTRUCTURACEE	Clave Foránea de la Estructura del Grupo de trabajo
CLASECENTROCOSTO	Clave Foránea de la clase del centro de costo

Tabla 17. Contrato

Contrato	
CODIGO	Clave primaria de la tabla.
NUMERO	Descripción de la unidad
ESTADO	Estado del registro (Vigente, Cancelado)
FECHACONTRATO	Fecha de firma de contrato
MONTO	Monto del contrato
NO_CD	Clave foránea del centro de costos
LOCALIDAD	Clave foránea de la localidad
NOMBRE	Nombre del contrato

Tabla 18. Obra

Obra	
CODIGO	Clave primaria de la tabla.
NOMBRE	Descripción de la unidad
ESTADO	Estado del registro (Vigente, Cancelado)
FECHACREA	Fecha de firma de contrato
CONTRATO	Monto del contrato
MONTOBRA	Clave foránea del centro de costos

Tabla 19. Cronograma

Cronograma	
CODIGO	Clave primaria de la tabla.
OBRA	Clave foránea de la obra
CAPITULO	Clave foránea del capitulo
RUBRO	Clave foránea del rubro
UNIDAD	Clave foránea de unidad
ANIO	Descripción del año
MES	Descripción del Mes
CANTIDAD	Cantidad a ejecutarse
PPRECIO	Precio del rubro

ESTADO	Estado del cronograma (vigente, cancelado)
CONTRATO	Clave foránea del contrato
PERIODO	Clave foránea del periodo (año, mes)
NUMCRONOGRAMA	Numero de cronograma

Tabla 20. Planillaje

Planillaje	
CODIGO	Clave primaria de la tabla.
OBRA	Clave foránea de la obra
CAPITULO	Clave foránea del capitulo
RUBRO	Clave foránea del rubro
UNIDAD	Clave foránea de la unidad de medida del rubro
CANTIDAD	Cantidad del planillaje
PRECIO	Precio del rubro
LEGALIZADO	Estado de lesgalizada del planilla (Legalizado ='S')
MATRIZ	Enviado a la matriz la planilla legalizada ='SI'
NUMASIENTO	Número de asiento
NUMPLANILLA	Número de planilla
PERIODO	Clave foránea del periodo

Tabla 21. Producción

Producción	
CODIGO	Clave primaria de la tabla.
OBRA	Clave foránea de la obra
CAPITULO	Clave foránea del capitulo
RUBRO	Clave foránea del rubro
UNIDAD	Clave foránea de la unidad de medida del rubro

CANTIDAD	Cantidad de producción y/o avance de obra
PRECIO	Precio del rubro
OBSERVACION	Se ingresa la observación en caso de no ejecutar lo planificado
ANIO	El año de ingreso del avance (2023,2024.)
MES	El mes de ingreso del avance (01,02...)
PERIODO	Clave foránea del periodo

Tabla 22. Costos

Costos	
CODIGO	Clave primaria de la tabla.
OBRA	Clave foránea de la obra
MONTOGASTOMAT	Valores de gastos matriz
MONTOSUBCON	Valores de subcontratos
MONTOCOSTO	Valores de costos y/o gastos para la producción
CUENTACONTABLE	Código de la cuenta contable de costos
MONTOCONTABLE	Valor de planilla contabilizadas
MONTOMARGENOP	Valor de margen operacional
MONTOSUPERAVIT	Valor de superávit o déficit
PERIODO	Clave foránea del periodo

d) Modelo Conceptual Ampliado

En el siguiente paso, será necesario elaborar un gráfico que represente los campos de las perspectivas o dimensiones establecidos en el paso anterior para cada datamart. Además, deberemos incluir los indicadores o métricas correspondientes, junto con sus respectivas fórmulas de cálculo, que pueden ser sumas, conteos o cálculos de porcentajes, según corresponda.

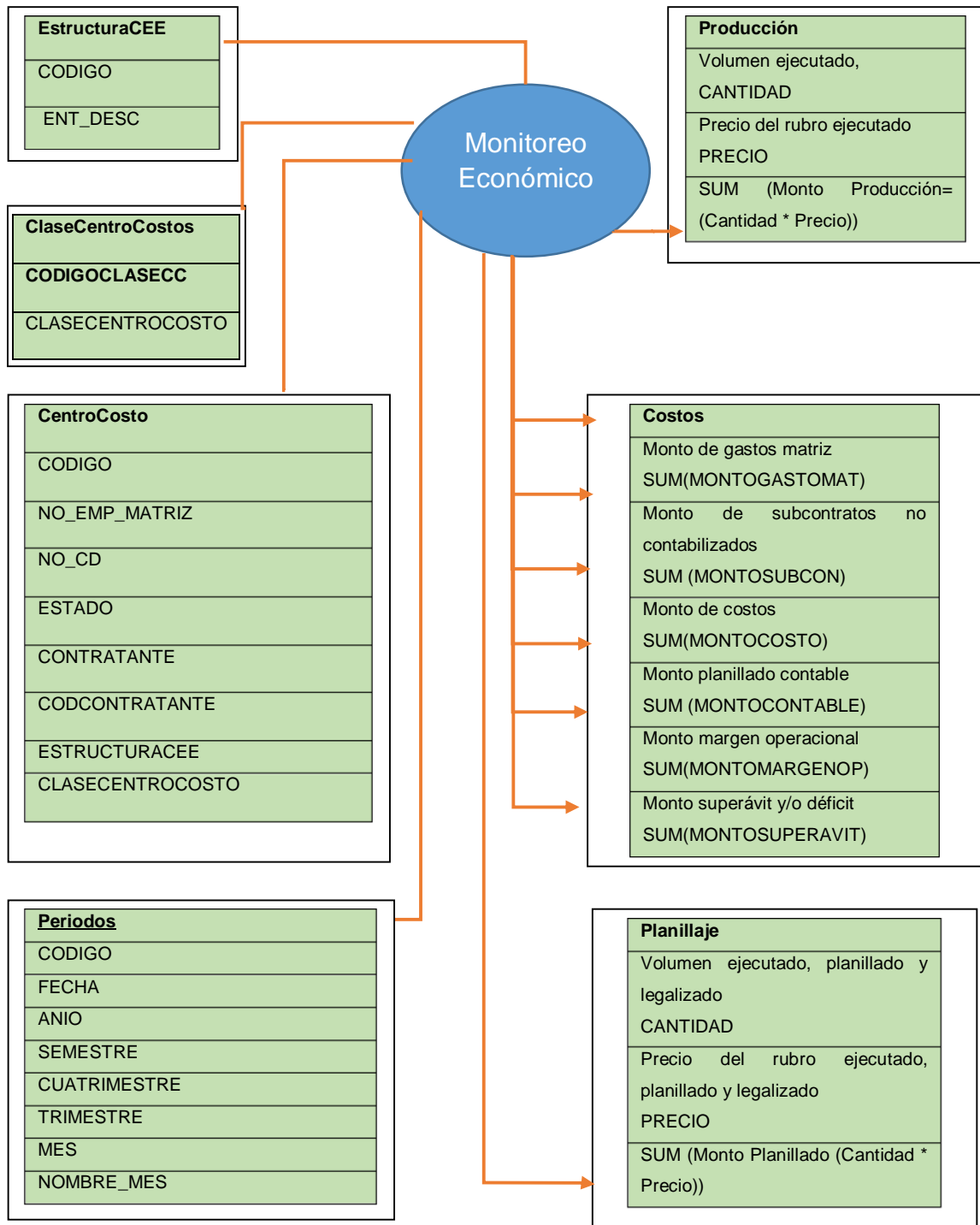


Figura 26 Monitoreo Económico

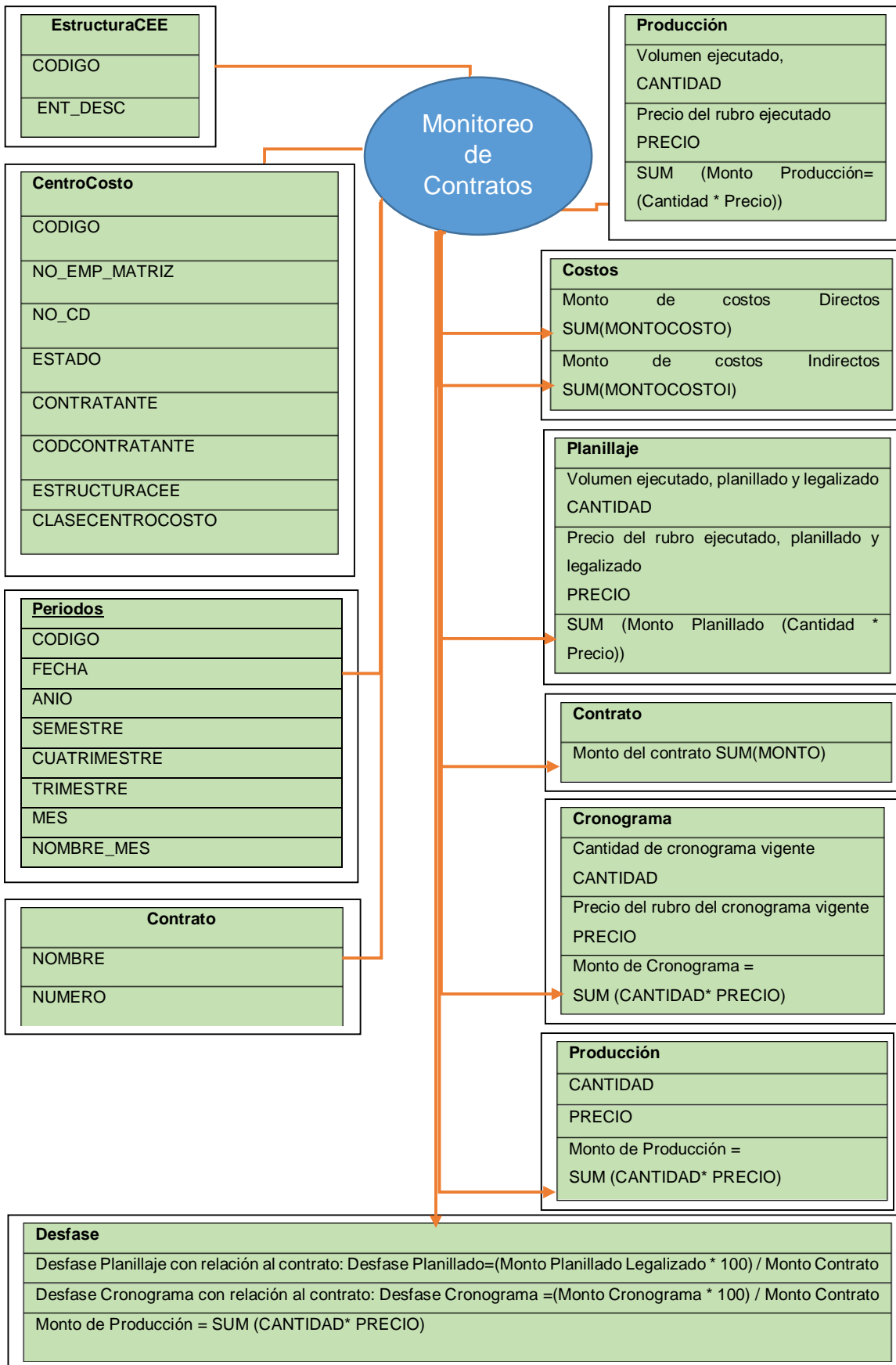


Figura 27 Monitoreo de Contratos

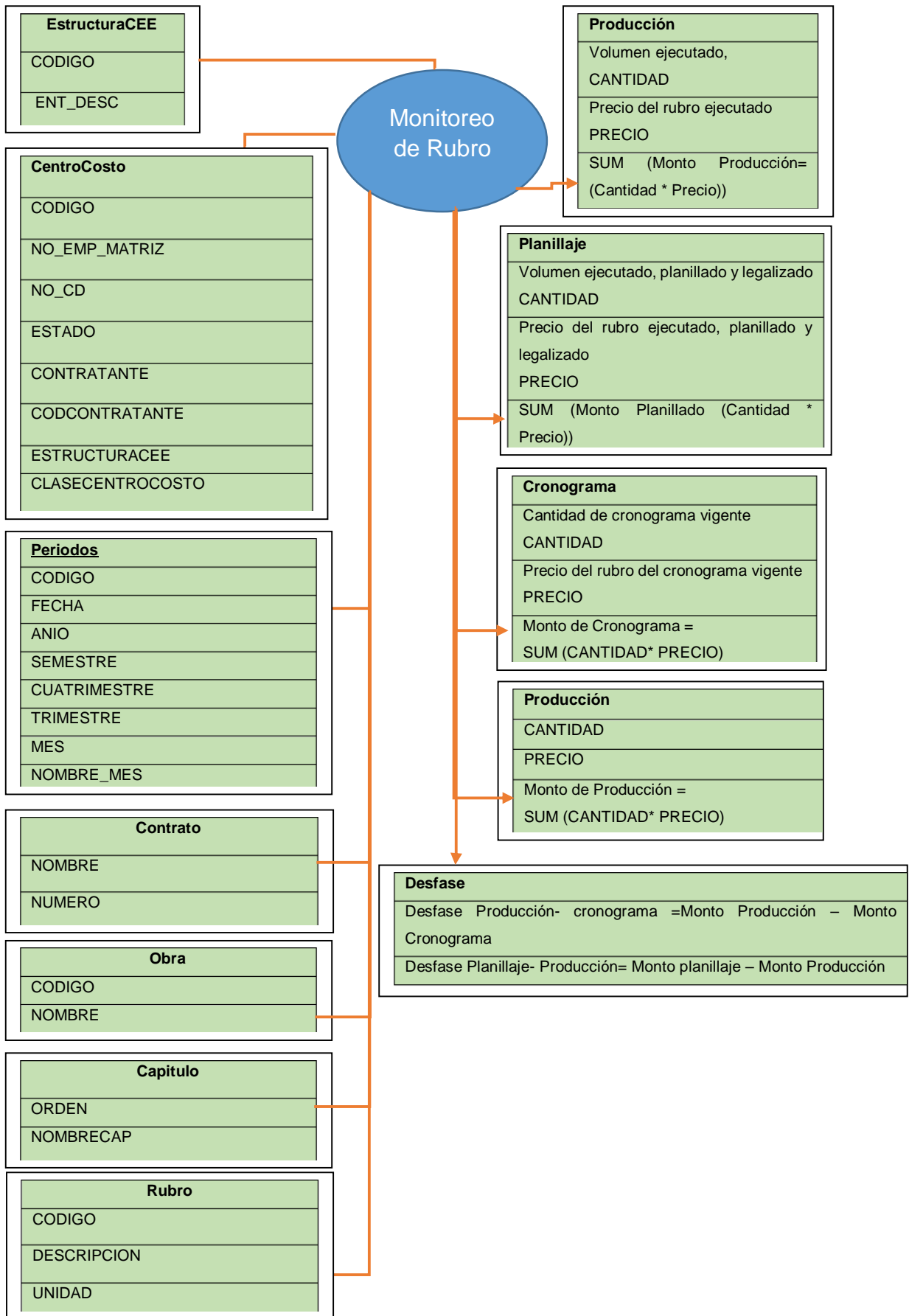


Figura 28 Monitoreo de Rubros

3.4. Modelo Lógico del DWH.

a) Tipo de Modelo Lógico del DW

Se implementará un modelo estrella, respaldado por la base de datos Oracle 12c y gestionado mediante el SQL Developer. En este entorno, se desarrollarán los paquetes y procedimientos necesarios para realizar el proceso de Extracción, Transformación y Carga (ETL). Las tablas de dimensiones y hechos estarán ubicadas en el mismo esquema de Oracle 12c.

El modelo estrella fue elegido por su facilidad para añadir nuevas tablas de hechos y dimensiones, permitiendo una fácil adaptación a las necesidades del C.E.E. Además, este modelo admite la integración de múltiples fuentes de datos con diversos niveles de granularidad.

b) Tablas de dimensiones

Las tablas de dimensiones serán incorporadas en el Datawarehouse correspondiente a los tres esquemas de desarrollo mencionados previamente, siguiendo la normativa establecida.

- El nombre que identifique el contenido que tendrá la tabla de dimensiones
- Necesariamente deberá tener la clave primaria de la tabla
- Se definirán el nombre de los campos los mismos que deberán ser intuitivos.

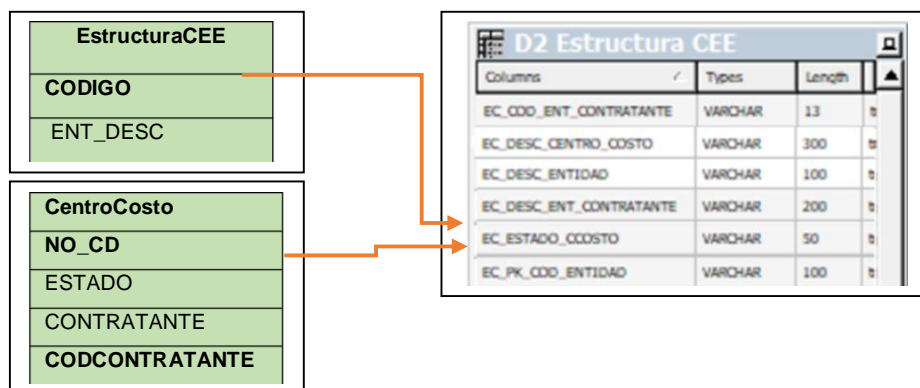


Figura 29. Estructura C.E.E.

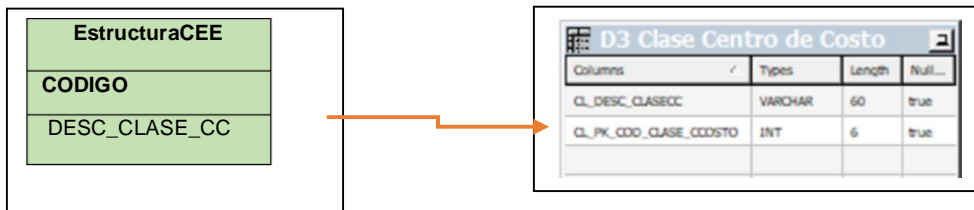


Figura 30. Clase Centro de Costos

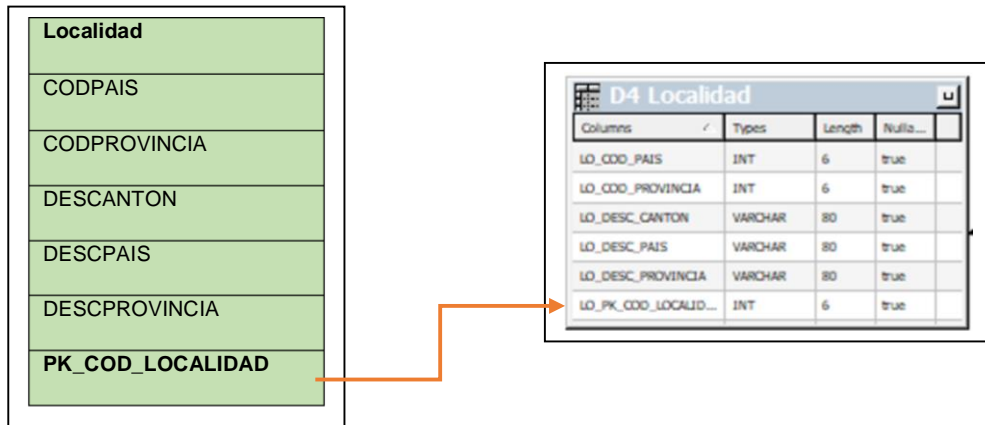


Figura 31. Localidad

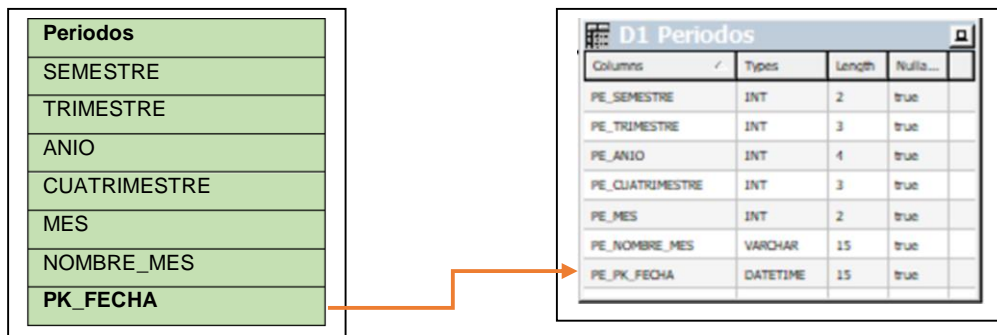


Figura 32. Periodos

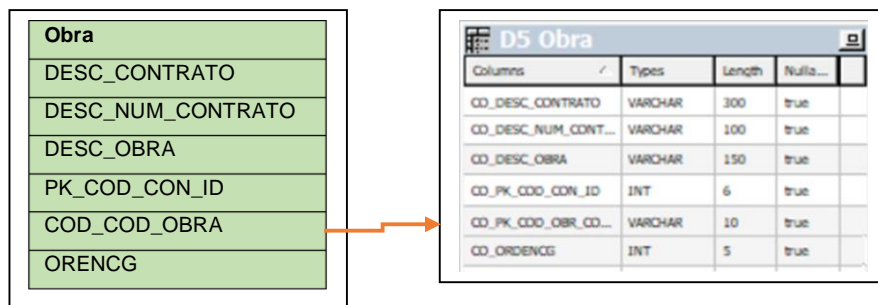


Figura 33. Obra

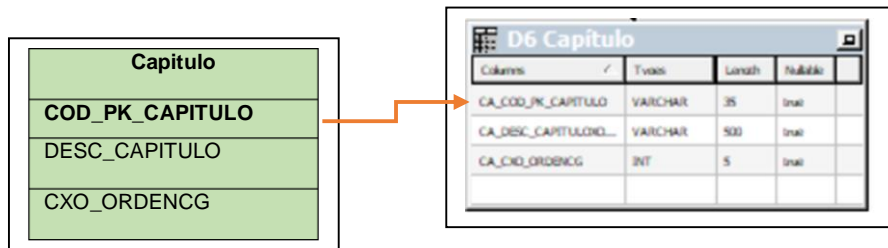


Figura 34. Capitulo

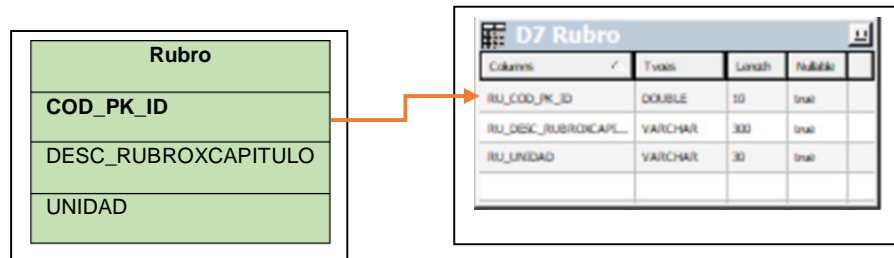


Figura 35. Rubro

c) Tablas de Hechos

Es necesario definir las tablas de hechos, donde se almacenarán las métricas e indicadores esenciales para la toma de decisiones en el C.E.E. Es importante señalar que las claves foráneas en estas tablas actuarán como filtros de búsqueda, permitiendo la filtración de información específica en las tablas de hechos. Para ello, se deben seguir estas pautas básicas:

- Asignar un nombre apropiado a cada tabla de hechos, que describa claramente el tipo de datos que contendrá.
- Asegurar que cada tabla de hechos tenga una clave primaria.
- Nombrar los campos de las tablas de hechos de manera descriptiva y clara, para que reflejen correctamente el contenido de los datos que representan.

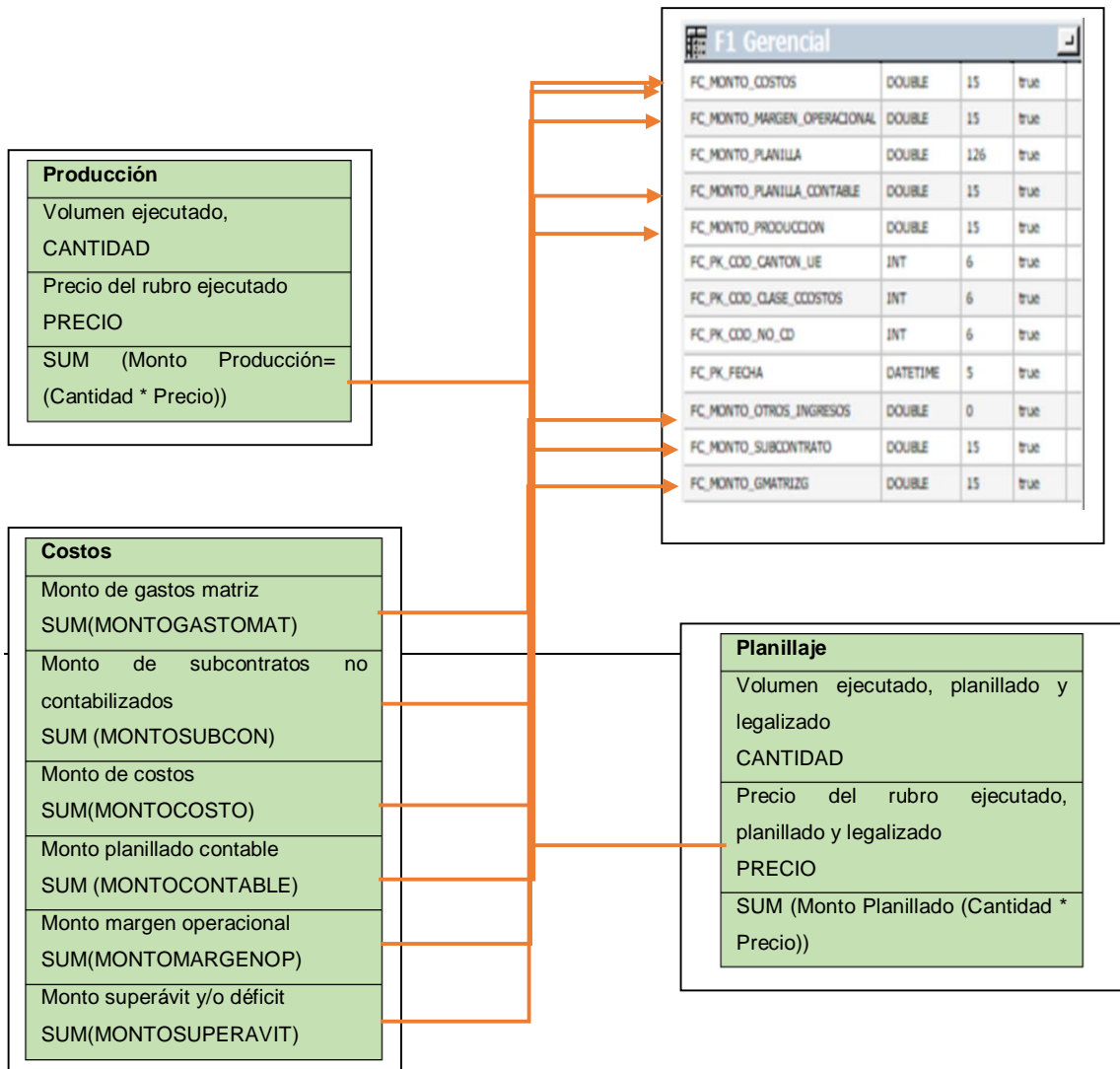


Figura 36. Monitoreo Económico

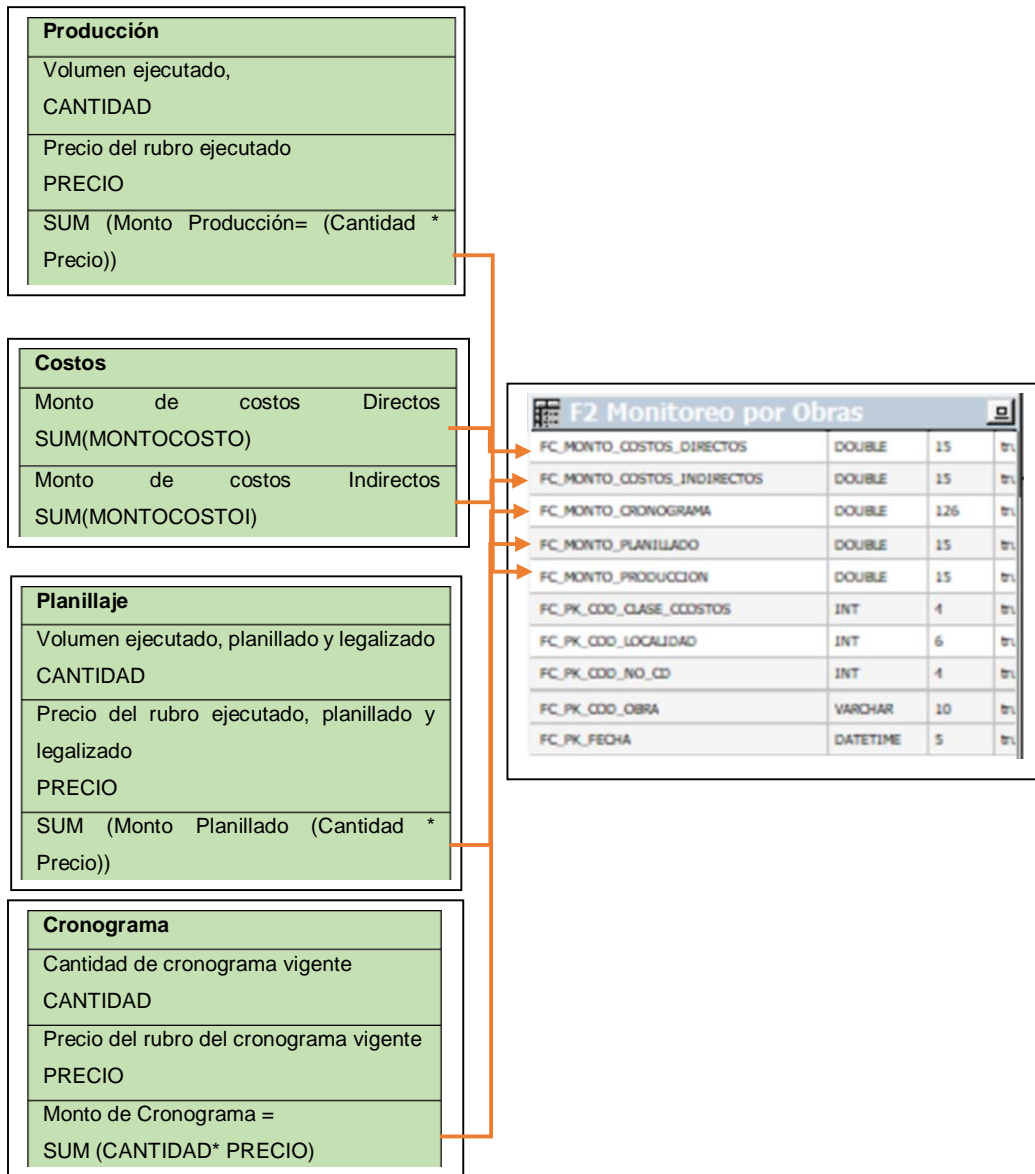


Figura 37. Monitoreo por Contrato

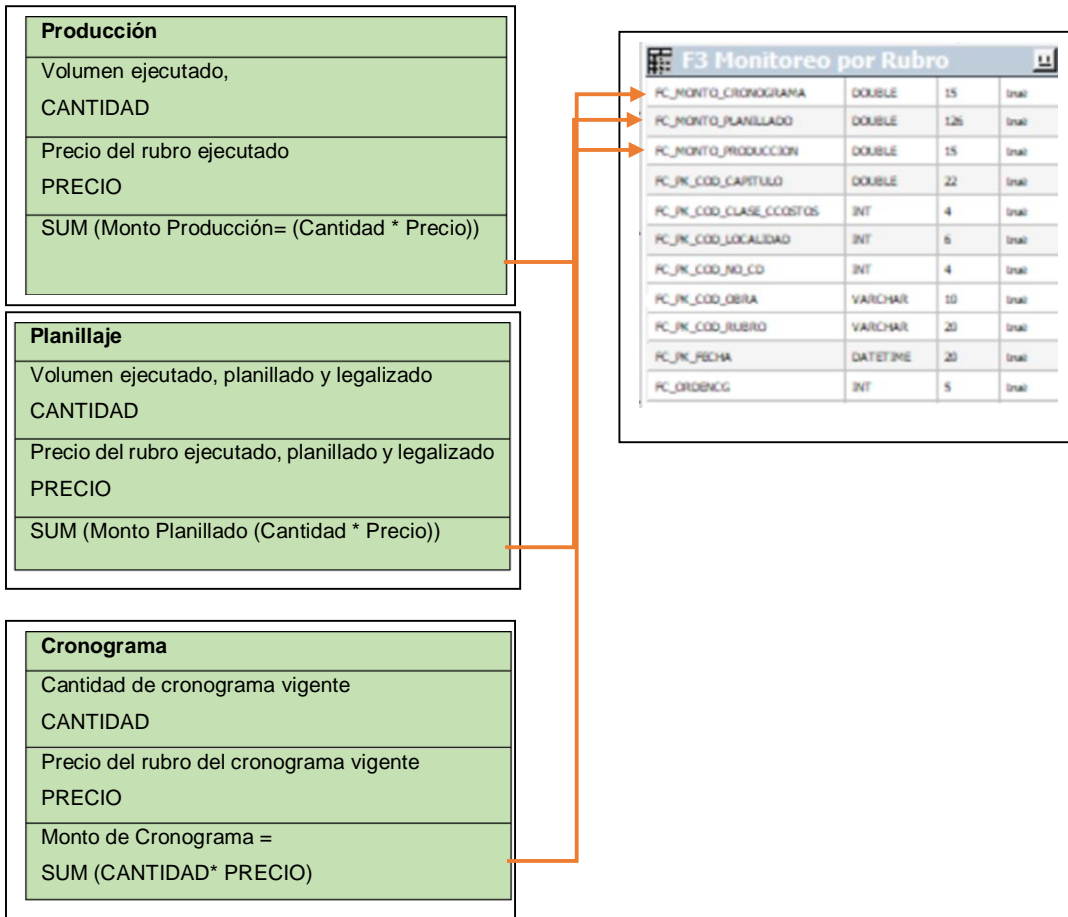


Figura 38. Monitoreo Por Rubro

e) Uniones

Se realizarán las relaciones respectivas entre las tablas de dimensiones y tablas de hechos por cada datamart establecido en este proyecto.

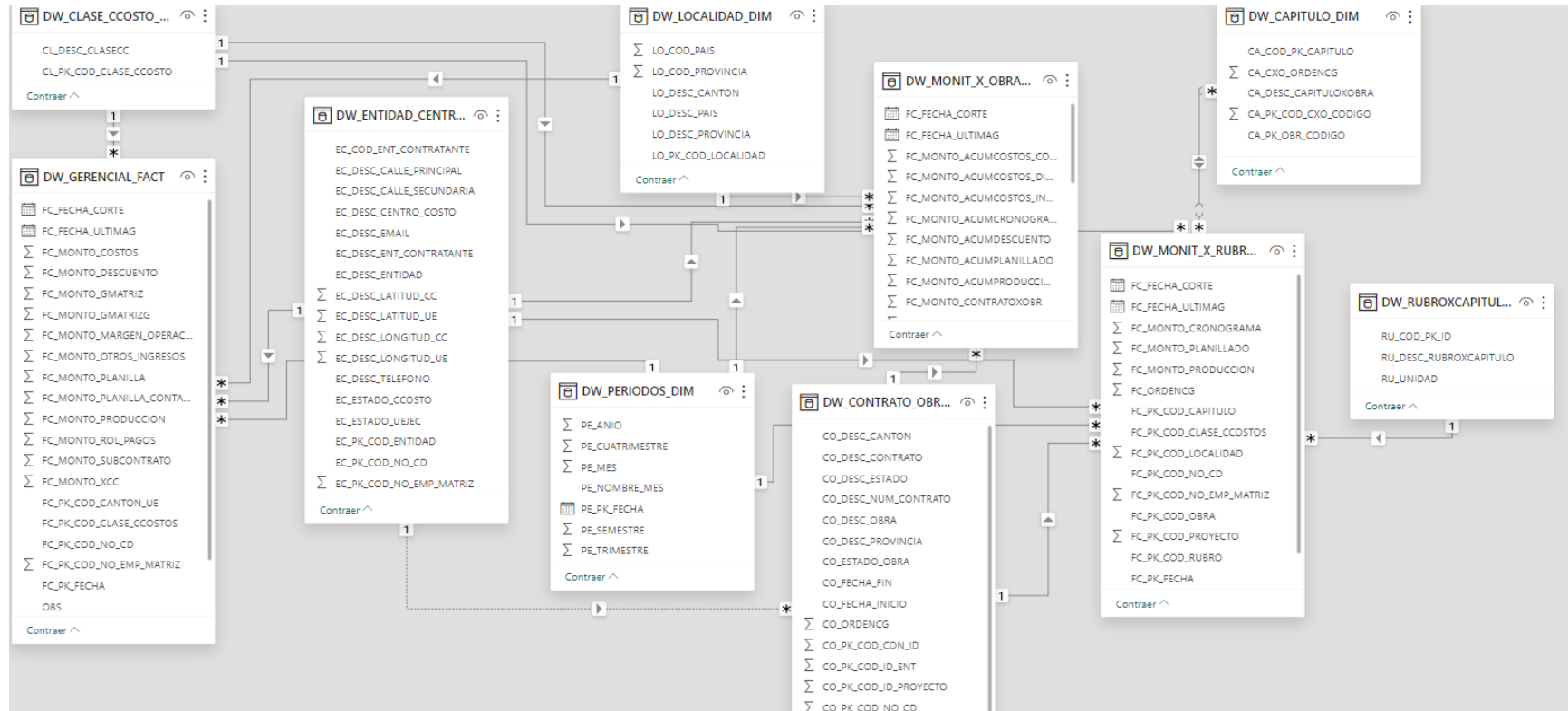


Figura 39. Modelo Completo BI

3.5. Integración de datos.

Después de establecer la estructura de las tablas, se llevará a cabo la carga de datos utilizando procesos de ETL, tal como se mencionó anteriormente. Estos procesos se ejecutarán mediante procedimientos almacenados agrupados en paquetes en Oracle 12c. Posteriormente, se programarán las cargas de datos según las necesidades del C.E.E.

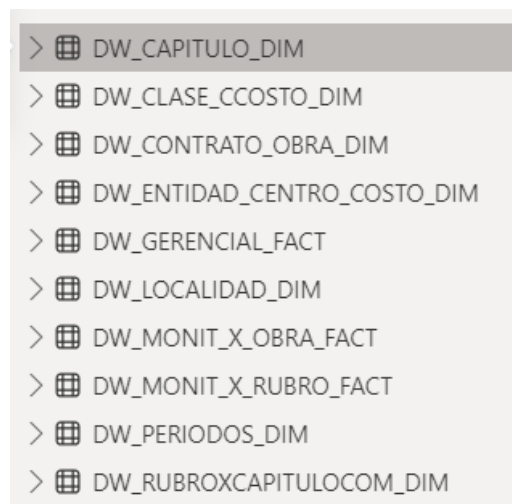


Figura 40. Listado de Datos Data Warehouse

a) Carga Inicial

Los datos serán obtenidos del sistema transaccional SICEE, y se enfocará en los procesos relacionados con la cadena de valor de la institución, específicamente en los procesos de construcción y evaluación de proyectos.

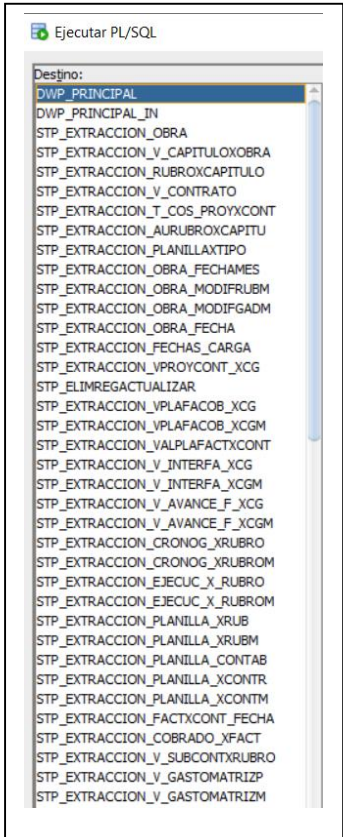
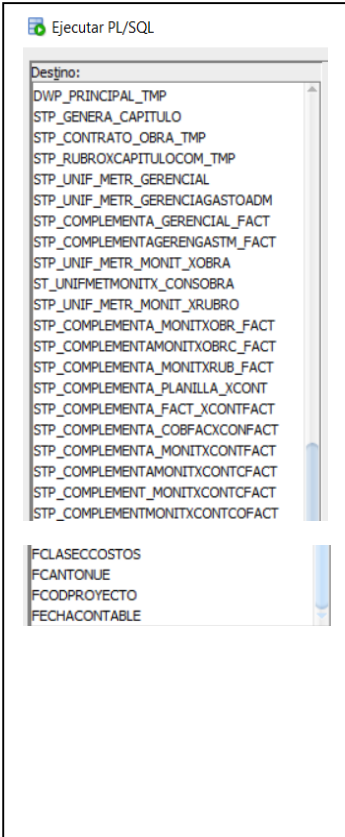
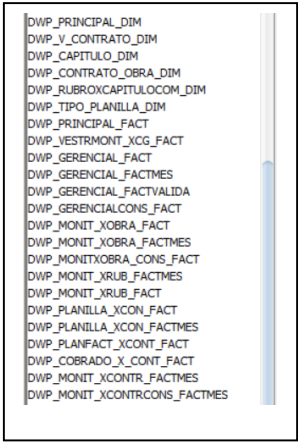
Extracción	Transformación	Carga
 <pre> Destino: DWP_PRINCIPAL DWP_PRINCIPAL_IN STP_EXTRACCION_OBRA STP_EXTRACCION_V_CAPITULO_XOBRA STP_EXTRACCION_RUBROXCAPITULO STP_EXTRACCION_V_CONTRATO STP_EXTRACCION_T_COS_PROYXCONT STP_EXTRACCION_AURUBROXCAPITU STP_EXTRACCION_PLANILLAXTIPO STP_EXTRACCION_OBRA_FECHAMES STP_EXTRACCION_OBRA_MODIFRUBM STP_EXTRACCION_OBRA_MODIFGADM STP_EXTRACCION_OBRA_FECHA STP_EXTRACCION_FECHAS_CARGA STP_EXTRACCION_VPROYXCONT_XCG STP_ELIMREGACTUALIZAR STP_EXTRACCION_VPLAFACOB_XCG STP_EXTRACCION_VPLAFACOB_XCGM STP_EXTRACCION_VALPLAFACXCONT STP_EXTRACCION_V_INTERFA_XCG STP_EXTRACCION_V_INTERFA_XCGM STP_EXTRACCION_V_AVANCE_F_XCG STP_EXTRACCION_V_AVANCE_F_XCGM STP_EXTRACCION_CRONOG_XRUBRO STP_EXTRACCION_CRONOG_XRUBROM STP_EXTRACCION_EJECUC_X_RUBRO STP_EXTRACCION_EJECUC_X_RUBROM STP_EXTRACCION_PLANILLA_XRUB STP_EXTRACCION_PLANILLA_XRUBM STP_EXTRACCION_PLANILLA_XCONTR STP_EXTRACCION_PLANILLA_XCONTRM STP_EXTRACCION_FACTXCONT_FECHA STP_EXTRACCION_COBRADO_XFACT STP_EXTRACCION_V_SUBCONTRRUBRO STP_EXTRACCION_V_GASTOMATRIZP STP_EXTRACCION_V_GASTOMATRIZM </pre>	 <pre> Destino: DWP_PRINCIPAL_TMP STP_GENERA_CAPITULO STP_CONTRATO_OBRA_TMP STP_RUBROXCAPITULOCOM_TMP STP_UNIF_METR_GERENCIAL STP_UNIF_METR_GERENCIAGASTOADM STP_COMPLEMENTA_GERENCIAL_FACT STP_COMPLEMENTA_GERENGASTM_FACT STP_UNIF_METR_MONIT_XOBRA ST_UNIFMETMONITX_CONSOBRA STP_UNIF_METR_MONIT_XRUBRO STP_COMPLEMENTA_MONITXOBR_FACT STP_COMPLEMENTA_MONITXRUB_FACT STP_COMPLEMENTA_MONITXCONTR_FACT STP_COMPLEMENTA_PLANILLA_XCONT STP_COMPLEMENTA_FACT_XCONTRFACT STP_COMPLEMENTA_COBFACXCONFACT STP_COMPLEMENTA_MONITXCONTRFACT STP_COMPLEMENTA_MONITXCONTRCFAC STP_COMPLEMENTA_MONITXCONTRCOFACT STP_COMPLEMENTMONITXCONTRCOFACT FCLASECOSTOS FCANTONUE FCODPROYECTO FECHACONTABLE </pre>	 <pre> DWP_PRINCIPAL_DIM DWP_V_CONTRATO_DIM DWP_CAPITULO_DIM DWP_CONTRATO_OBRA_DIM DWP_RUBROXCAPITULOCOM_DIM DWP_TIPO_PLANILLA_DIM DWP_PRINCIPAL_FACT DWP_VESTRMONIT_XCG_FACT DWP_GERENCIAL_FACT DWP_GERENCIAL_FACTMES DWP_GERENCIAL_FACTVALIDA DWP_GERENCIALCONS_FACT DWP_MONIT_XOBRA_FACT DWP_MONIT_XOBRA_FACTMES DWP_MONIT_XOBRA_CONS_FACT DWP_MONIT_XRUB_FACT DWP_MONIT_XRUB_FACTMES DWP_MONIT_XRUB_FACT DWP_PLANILLA_XCON_FACT DWP_PLANILLA_XCON_FACTMES DWP_PLANFACT_XCONT_FACT DWP_COBRADO_X_CONT_FACT DWP_MONIT_XCONTR_FACTMES DWP_MONIT_XCONTRCONS_FACTMES </pre>

Figura 41. Carga inicial Extracción, Transformación, Carga

Una vez finalizado el análisis, se procederá a programar los procesos almacenados en el gestor de bases de datos SQL Developer. Estos procesos facilitarán la ejecución del ETL del proyecto, con el objetivo de nutrir los tres datamarts propuestos.

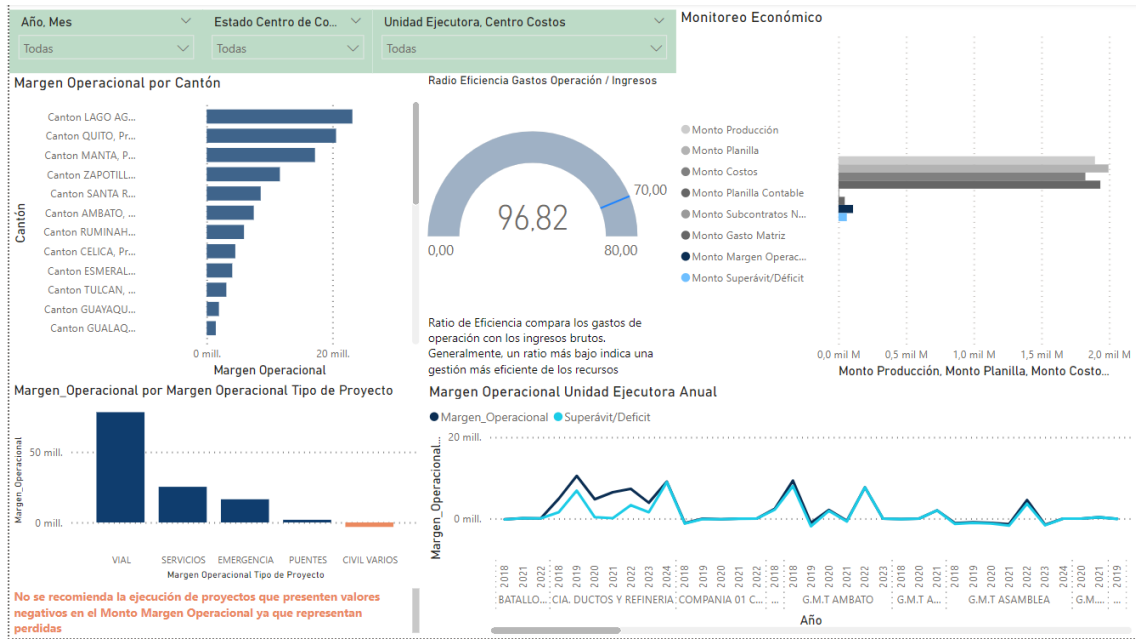


Figura 42. Dashboard Final Monitoreo Económico

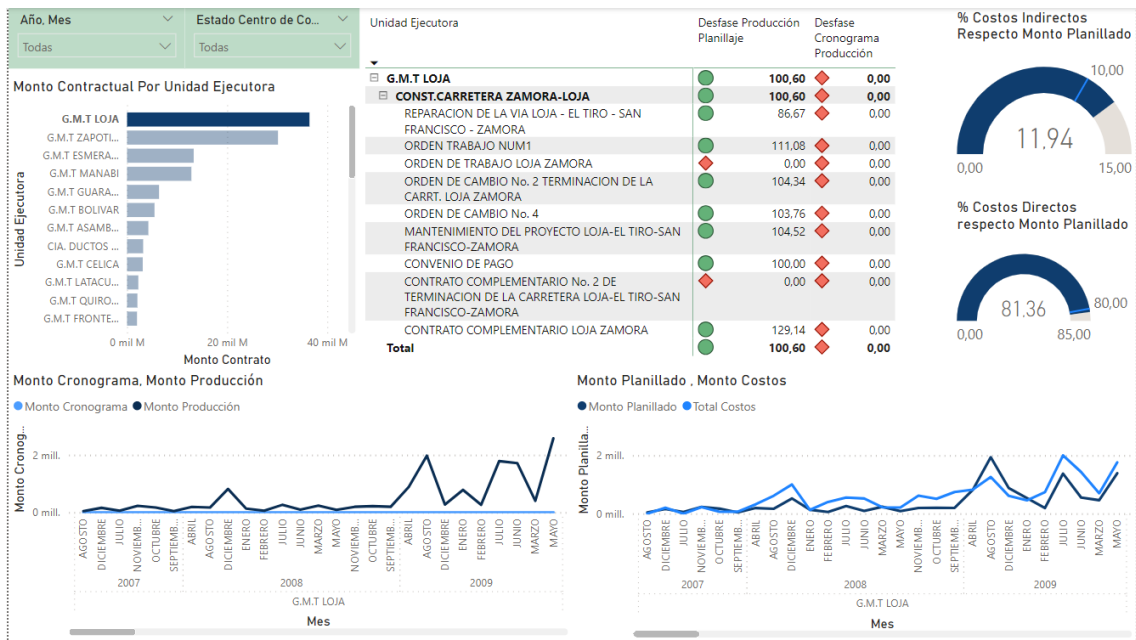


Figura 43. Dashboard Final Monitoreo Por Contrato

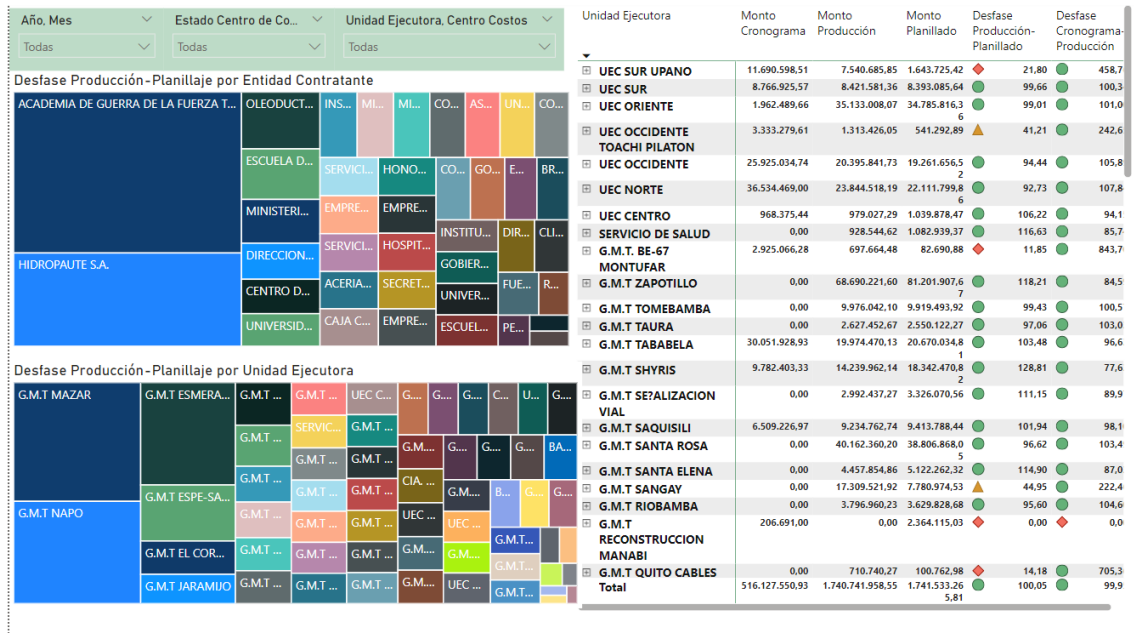


Figura 44. Dashboard Final Monitoreo por Rubro

b) Actualización

Se establece que la carga de datos en las tablas de dimensiones y hechos se realizará diariamente, debido a la importancia de estos análisis para el área de evaluación de proyectos y el cumplimiento de las políticas institucionales que requieren análisis diarios.

3.6. Resultados y despliegue.

Creando tableros de control y visualización para supervisar el avance y gastos de los proyectos de construcción desde el punto de administración técnica. Estos tableros van a incluir los datamart monitoreo económico, monitoreo de contratos, monitoreo de rubros. Cada uno de estos tableros van a tener métricas y/o indicadores específicos todo diseñado para adaptarse a lo que requiere el C.E.E., para tomar las mejores decisiones. Para llevar a cabo esta tarea, se introducirán los datos en Power BI, una plataforma donde se podrá crear y visualizar la información de manera efectiva.

3.7. Entrenamiento del modelo de regresión lineal y la respectiva predicción de datos.

El Cuerpo de Ingenieros del Ejército requiere mejorar su capacidad para predecir el margen operacional, utilizando los datos históricos de ingresos y

egresos. Para esto, se utilizará la información obtenida del sistema transaccional SICEE. Se han aplicado técnicas de regresión lineal, ya que es un método efectivo de predicción que ayudan a ajustar las estrategias y a mejorar la eficacia y eficiencia de los resultados que esperamos obtener, para lo cual es necesario realizar el modelo Predictivo para el Margen Operacional en Proyectos de Construcción.

Para lo cual seguimos los pasos de limpieza del dato:

- Al inspeccionar la base de datos a primera vista se observa que no hay categorías. Además, sus valores en su totalidad son numéricos, la variable que deseamos predecir es numérica lo cual nos indica que tenemos que aplicar un modelo de regresión lineal, adicional se puede ver que no tengo diferentes tipos de datos todos son numéricos, por lo tanto, no es necesario aplicar escalado (unificación) de tipos de datos, como se ve en la figura 46.

```

▶ dft.info()
↳ <class 'pandas.core.frame.DataFrame'>
  RangeIndex: 10376 entries, 0 to 10375
  Data columns (total 12 columns):
   #   Column                                     Non-Null Count  Dtype
  ---  ---
   0   FC_MONTO_PLANILLA                         10376 non-null  float64
   1   FC_MONTO_PLANILLA_CONTABLE                10376 non-null  float64
   2   FC_MONTO_COSTOS                           10376 non-null  float64
   3   FC_MONTO_MARGEN_OPERACIONAL              10376 non-null  float64
   4   FC_MONTO_OTROS_INGRESOS                   10376 non-null  float64
   5   FC_MONTO_PRODUCION                        10376 non-null  float64
   6   FC_MONTO_XCC                              8364 non-null   float64
   7   FC_MONTO_DESCUENTO                        10376 non-null  float64
   8   FC_MONTO_SUBCONTRATO                      10376 non-null  float64
   9   FC_MONTO_GMATRIZG                         10376 non-null  float64
  10  FC_MONTO_CRONOGRAMA                       10376 non-null  float64
  11  FC_SUPERAVIT                              10376 non-null  float64
dtypes: float64(12)
memory usage: 972.9 KB

```

Figura 45. Información del contenido del DataFrame.

Al realizar una descripción del DataFrame (dft), nos permite analizar los siguientes aspectos.

- Se verifica que el error absoluto resultante siempre debe estar muy por debajo de la media, señalada en la variable a predecir `fc_monto_margen_operacional`.
- Además, se puede analizar la desviación estándar, lo cual es el grado de dispersión o separación del grupo de datos de su valor promedio. Podemos establecer que la desviación estándar de la data está estable y no tienen una dispersión considerable.
- Valores máximos y mínimos de la data. Se observa que no existen demasiados valores mínimos y máximos. Sin embargo, una vez generado el análisis de Correlación de Pearson definiremos si son o no relevantes estas variables que al momento tienen datos mínimos y máximos.
- Visualizar los percentiles los cuales se dividen en 4 grupos 25%, 50%, 75%, 100% de los cuales los valores que se encuentran en el grupo de mínimos <1% y >99% son los valores atípicos.

De este análisis podemos fijarnos que nuestra data mayormente se encuentra en los percentiles 25% al 75% y su desviación estándar es estable, lo cual es bueno esto indica que la data es relativamente compacta. Como se ve en la figura 47.

```
dft.describe().transpose()
```

	count	mean	std	min	25%	50%	75%	max
FC_MONTO_PLANILLA	10376.0	1.935465e+05	5.767113e+05	-28383.07	0.0000	0.00	1.028916e+05	1.502551e+07
FC_MONTO_PLANILLA_CONTABLE	10376.0	1.867791e+05	6.699231e+05	-1771596.07	0.0000	0.00	1.928423e+04	1.627943e+07
FC_MONTO_COSTOS	10376.0	1.762256e+05	4.425930e+05	-1620567.32	1048.3200	24046.44	1.616116e+05	1.642300e+07
FC_MONTO_MARGEN_OPERACIONAL	10376.0	1.030049e+04	5.661483e+05	-9384317.37	-64226.6275	-6162.79	0.000000e+00	1.162410e+07
FC_MONTO_OTROS_INGRESOS	10376.0	4.366381e+03	8.050913e+04	-156957.97	0.0000	0.00	0.000000e+00	3.707612e+06
FC_MONTO_PRODUCCION	10376.0	1.836259e+05	1.108465e+06	-64273163.13	0.0000	0.00	8.417316e+04	6.589852e+07
FC_MONTO_XCC	8364.0	3.489723e+07	1.670870e+08	0.00	224102.3200	2547997.98	1.277634e+07	1.735480e+09
FC_MONTO_DESCUENTO	10376.0	-7.721698e+01	2.593307e+03	-214243.82	0.0000	0.00	0.000000e+00	0.000000e+00
FC_MONTO_SUBCONTRATO	10376.0	2.530014e+02	4.030374e+03	0.00	0.0000	0.00	0.000000e+00	2.122582e+05
FC_MONTO_GMATRIZG	10376.0	4.507348e+03	2.446959e+04	-97822.55	0.0000	0.00	0.000000e+00	7.270193e+05
FC_MONTO_CRONOGRAMA	10376.0	6.509832e+03	1.619334e+05	0.00	0.0000	0.00	0.000000e+00	1.166675e+07
FC_SUPERAVIT	10376.0	5.793143e+03	5.669170e+05	-9384317.37	-66699.8100	-6614.70	0.000000e+00	1.162410e+07

Figura 46. Muestra resumen estadístico de columnas del DataFrame.

Se realizará el análisis de Correlación de Pearson, para evaluar la relación existente entre las variables, enfocándose lógicamente en la variable principal del análisis la cual corresponde FC_MARGEN_OPERACIONAL.

En el análisis de Correlación de Pearson se realizó un análisis tanto horizontal como verticalmente, lo cual permitió establecer que las siguientes variables no aportan al modelo de análisis de forma significativa, ya que tienen una correlación baja entre las variables. Como se observa en la figura 48.

FC_MONTO_OTROS_INGRESOS', 'FC_SUPERAVIT', 'FC_MONTO_CRONOGRAMA', 'FC_MONTO_GMATRIZG', 'FC_MONTO_XCC', 'FC_MONTO_DESCUENTO',

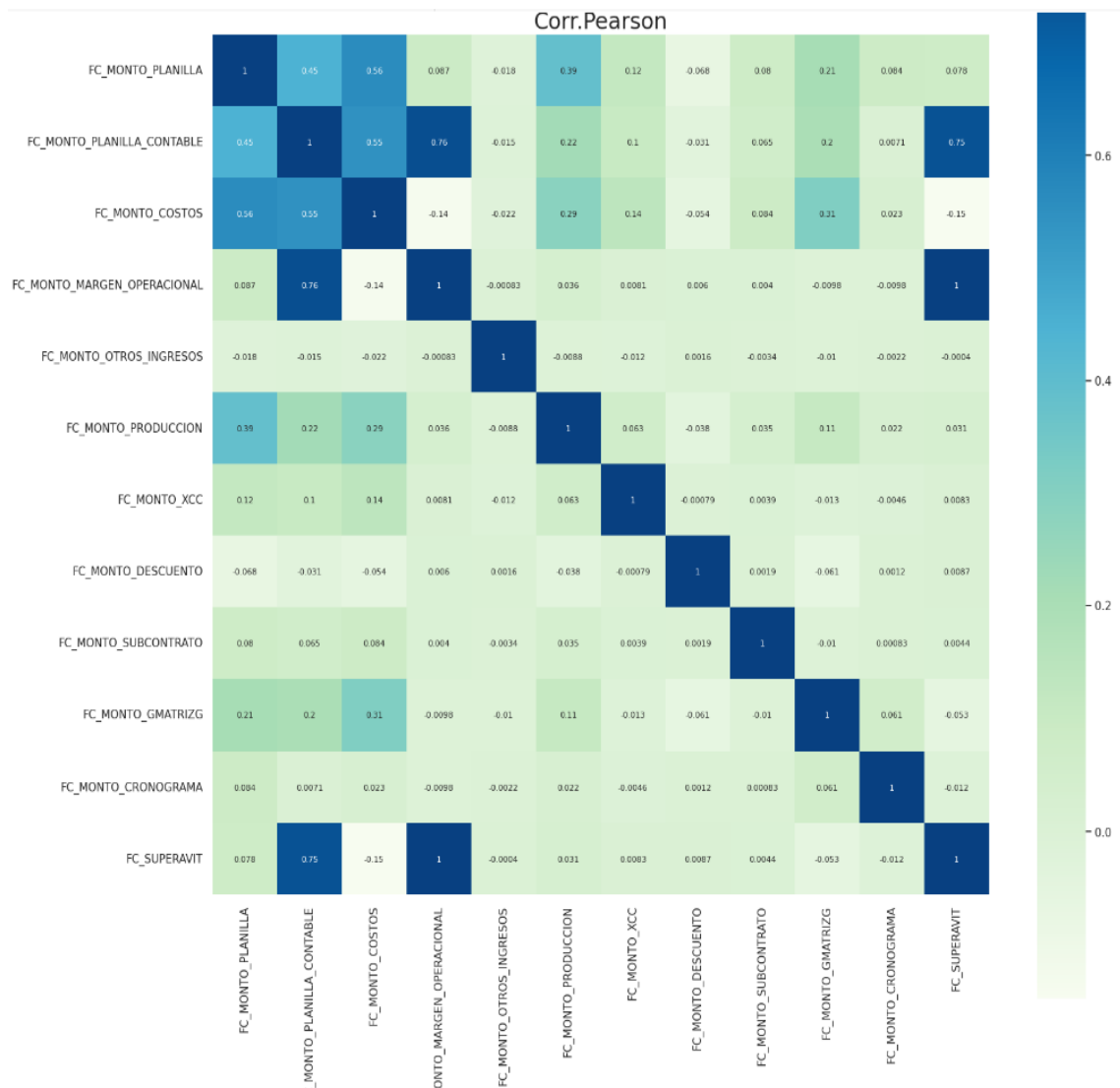


Figura 47 Correlación de Pearson Inicial

Se procedió a eliminar las variables que no aportan al modelo de forma significativa, lo cual generó que se trabajaran con 6 variables en este modelo. Ver Figura 49.

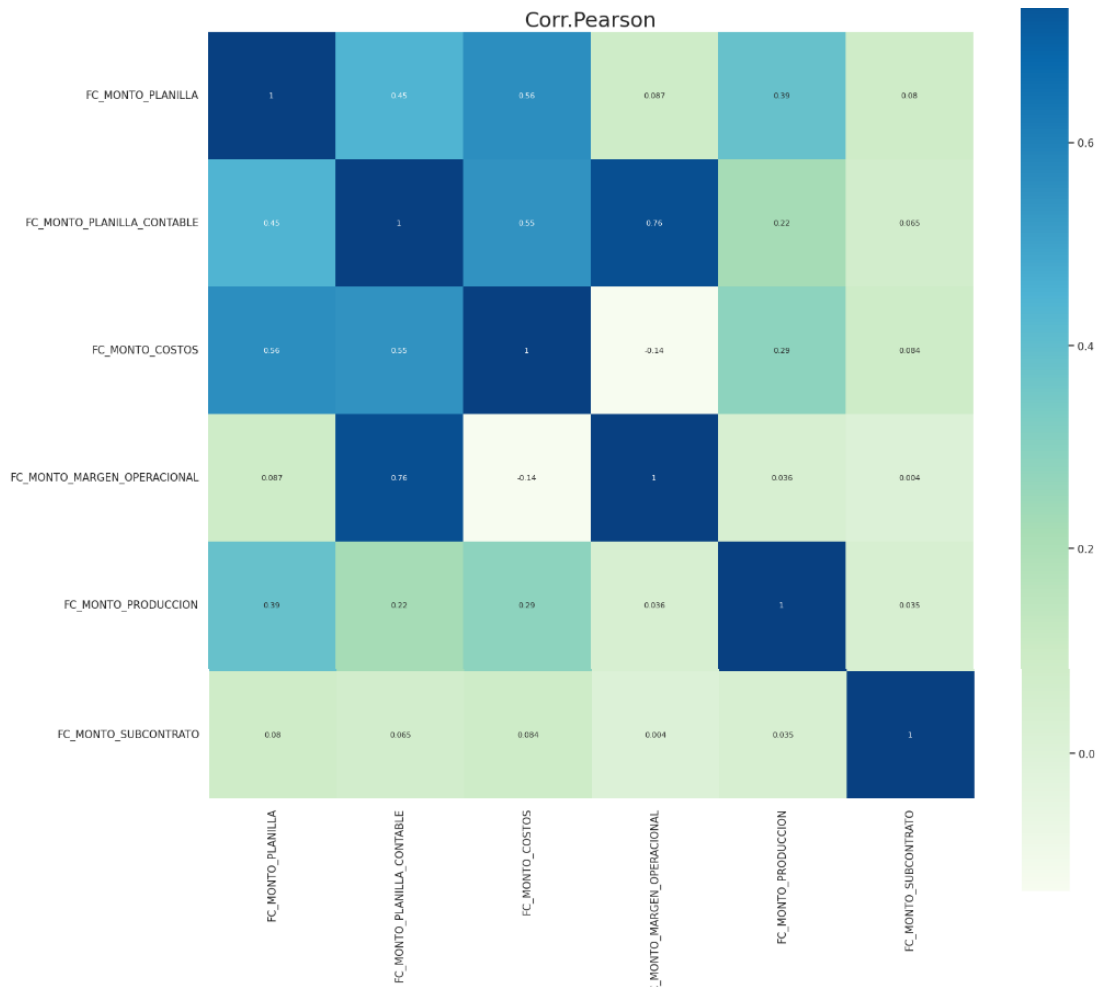


Figura 48. Correlación de Pearson Modificado.

Se verifica si existen valores nulos en la data, obteniendo como resultado que la data está completa sin valores nulos. Como se muestra en la figura 50.

```

dft.isnull().sum()
FC_MONTO_PLANILLA          0
FC_MONTO_PLANILLA_CONTABLE 0
FC_MONTO_COSTOS            0
FC_MONTO_MARGEN_OPERACIONAL 0
FC_MONTO_PRODUCCION        0
FC_MONTO_SUBCONTRATO       0
dtype: int64

```

Figura 49. Verificación de valores nulos.

Se grafica la distribución de la data mediante histogramas, en la cual se observa que la distribución de los datos que están siendo analizados tienen una distribución lineal, por lo tanto, estamos hablando de un modelo lineal. Para extraer las mejores características de las variables se deberá aplicar un rectificarte lineal denominado (RELU). Ver histogramas en la figura 51.

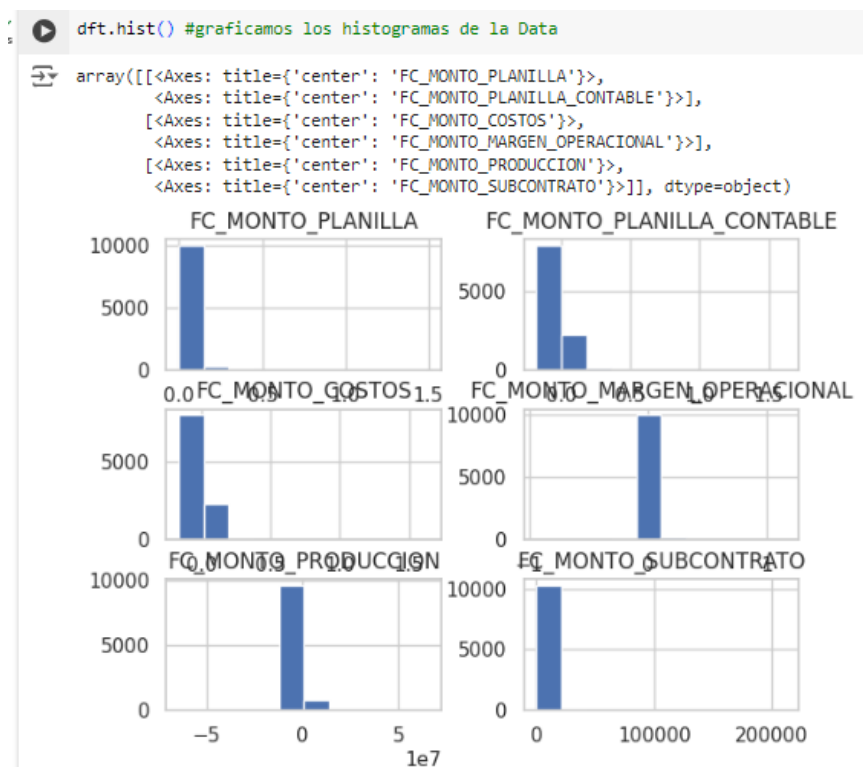


Figura 50. Histogramas por variables.

Aplicaremos el comando `Msno.bar`, el cual nos permite visualizar en un gráfico de barras la cantidad de datos faltantes por columnas del `dft`. Y como se puede evidenciar en la figura 52, no existen datos faltantes en ninguna columna.

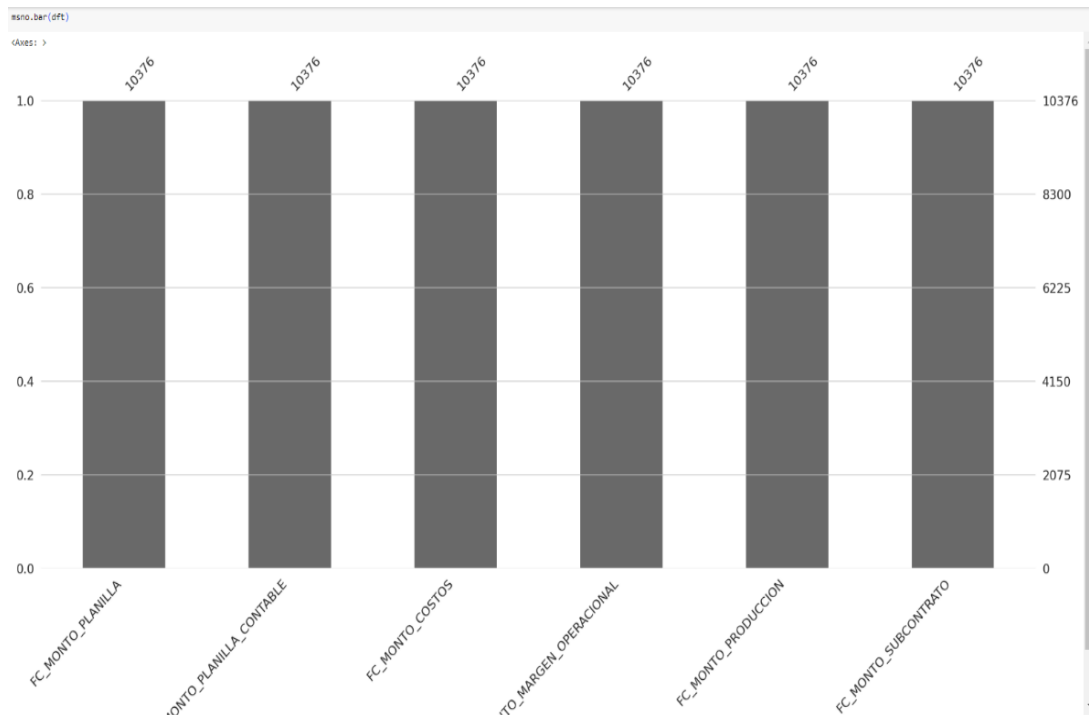


Figura 51 Verificación de Datos Faltantes en columnas del DataFrame.

Se aplicará el comando `msno.matriz(dft)`, el cual nos muestra la matriz de la gráfica de datos perdidos en las filas con datos vacíos o nulos. Por lo tanto, se confirma en la figura 53 que no existen datos vacíos en las filas del DataFrame.

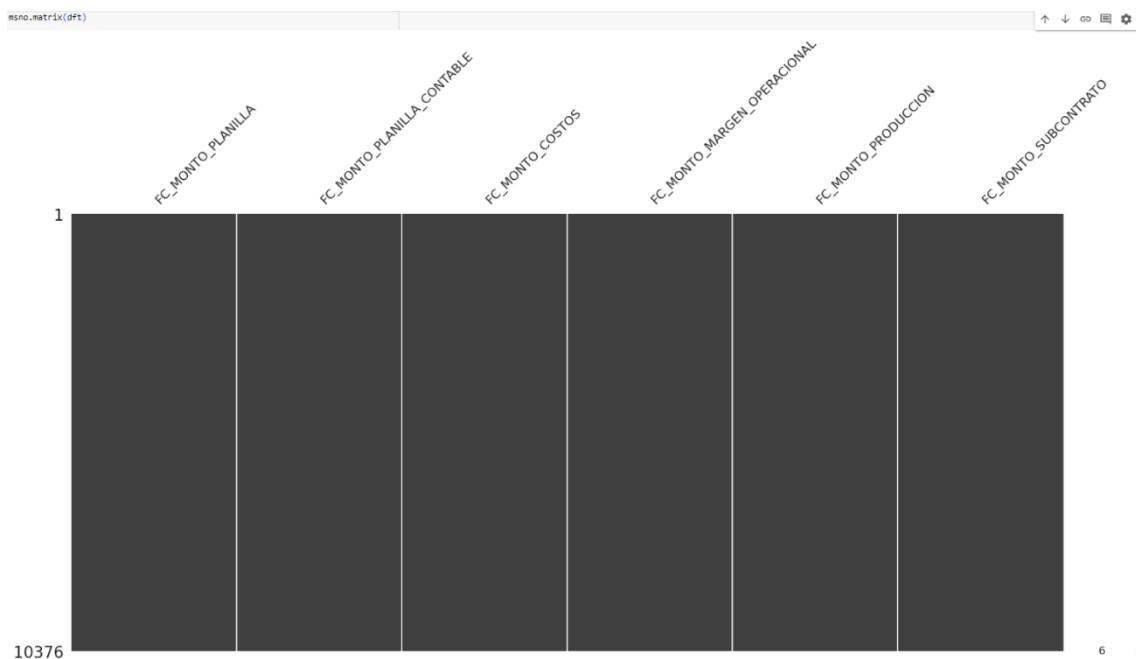


Figura 52. Verificación de Datos Faltantes en las filas del DataFrame.

Se genera la distribución de las variables. Necesitamos asegurarnos de que los datos estén agrupados y sigan una distribución, lo cual es crucial para el análisis. Confirmar una distribución lineal de los datos es esencial. Sin embargo, si están muy dispersos debido a una desviación estándar excesivamente grande, los datos no serán útiles. Por lo tanto se observa que, con las variables actuales, se podrán usar para predecir el margen operacional mediante un modelo de regresión lineal. Como se muestra en la figura 54.

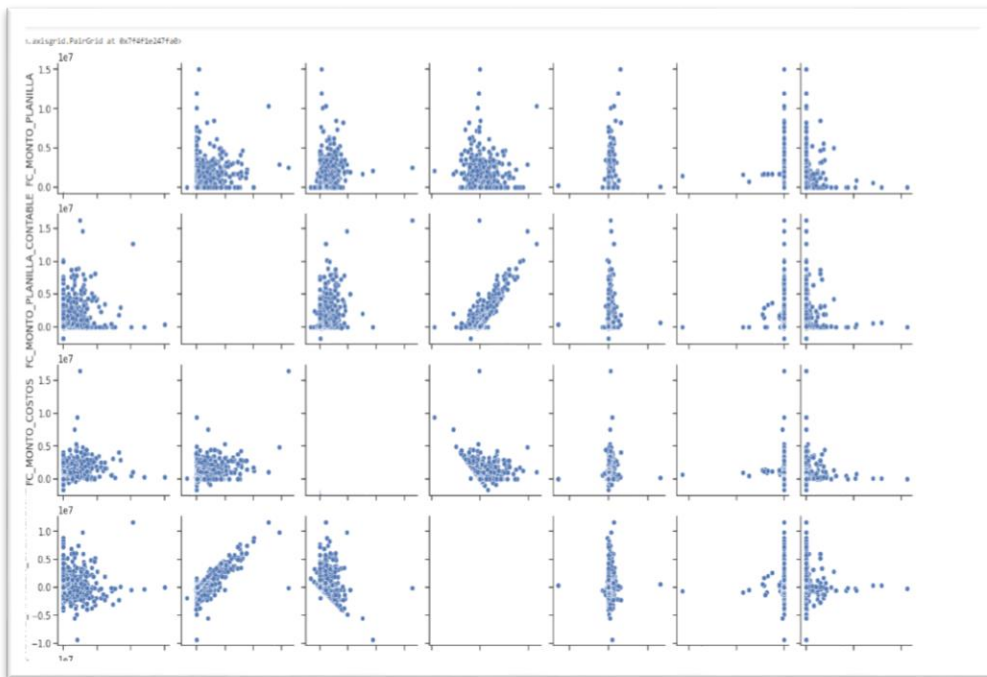


Figura 53. Gráfico de Dispersión de Variables.

Una vez limpia la data se establece la arquitectura, del modelo de regresión lineal, para cual separamos las características “X” de los valores a predecir denominados “Y”. Como se muestra en la figura 55.

```
#Separamos las características de los Y= valores a predecir
x=dft.drop('FC_MONTO_MARGEN_OPERACIONAL',axis=1)
y= dft[['FC_MONTO_MARGEN_OPERACIONAL']]
```

Figura 54. Separar Características del valor a predecir.

Se separan los datos en dos grupos, uno para entrenamiento y otro para testeo, usando una muestra de 100 individuos en cada barajada. En esta división, el conjunto `x_train` contiene el 70% de los datos, mientras que `x_test` contiene el 30%, una práctica recomendada. En resumen, se utilizaron 70 individuos para entrenar y 30 para test, logrando un total de 100 individuos en la muestra. Se sugiere que el tamaño de la muestra sea del 1 al 2% del total de los datos. Como se muestra en la figura 56.

```
[60] # separamos datos entre conjunto de datos de entrenamiento con una semilla de 100 individuos por cada barajada
#y testeo en este conjunto x_train tiene 70% de datos x_test 30% es lo recomendable,
#es decir 70 individuos para datos y 30 individuos para test hasta llegar a 100 individuos que es la semilla
#el porcentaje de la semilla es recomendable se 1 o 2% de total de datos
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state=100,test_size=0.3)
```

Figura 55. Separar los datos de entrenamiento y testeo.

Se utilizará una red neuronal que se basa en la función RELU (Rectified Linear Unit), que es conocida por acceder a las mejores características de las variables y hacer predicciones de manera efectiva. En este caso, se realizará la red con 5 neuronas en la primera capa, 3 en la segunda y 1 en la tercera. Para mejorar el rendimiento del modelo, se utilizará el algoritmo 'adam' y como métrica para evaluar el rendimiento, se usará el error cuadrático medio (MSE). Durante el proceso de entrenamiento, cada iteración generará un error de aprendizaje que será evaluado utilizando MSE, que es adecuado para modelos de regresión. El parámetro `batch_size` determinará cuántas filas se procesarán en cada iteración, mientras que `epochs` indicará el número total de iteraciones. Se entrenará el modelo para predecir el margen operacional, utilizando `x_train` e `y_train` como datos de entrenamiento, y `x_test` e `y_test` para la validación del modelo. Ver figuras 57,58.

```
ann= Sequential()
ann.add(Dense(5,activation='relu')) #Pasamos 5 neuronas
ann.add(Dense(3,activation='relu')) #Pasamos 3 neuronas
ann.add(Dense(1)) #Pasamos 1 neuronas
ann.compile(optimizer='adam',loss='mse') #optimizamos con la metrica de optimizacion adam, y como loss metrica de
ann.fit(x=x_train,y=y_train,validation_data=(x_test,y_test),batch_size=150,epochs=330,verbose=1)
```

Figura 56. Código crear red neuronal

```

Epoch 277/330
49/49 [=====] - 0s 5ms/step - loss: 229513363456.0000 - val_loss: 238121418752.0000
Epoch 278/330
49/49 [=====] - 0s 4ms/step - loss: 229146935296.0000 - val_loss: 237856555008.0000
Epoch 279/330
49/49 [=====] - 0s 4ms/step - loss: 229020663808.0000 - val_loss: 237844054016.0000
Epoch 280/330
49/49 [=====] - 0s 4ms/step - loss: 229012291584.0000 - val_loss: 237813153792.0000
Epoch 281/330
49/49 [=====] - 0s 3ms/step - loss: 229003821056.0000 - val_loss: 237818183680.0000
Epoch 282/330
49/49 [=====] - 0s 4ms/step - loss: 228988829696.0000 - val_loss: 237805355008.0000
Epoch 283/330
49/49 [=====] - 0s 6ms/step - loss: 228998316032.0000 - val_loss: 237806518272.0000
Epoch 284/330
49/49 [=====] - 0s 5ms/step - loss: 228985847808.0000 - val_loss: 237802913792.0000
Epoch 285/330
49/49 [=====] - 0s 5ms/step - loss: 228982276096.0000 - val_loss: 237813350400.0000
Epoch 286/330
49/49 [=====] - 0s 5ms/step - loss: 228976443392.0000 - val_loss: 237801340928.0000
Epoch 287/330
49/49 [=====] - 0s 5ms/step - loss: 228965187584.0000 - val_loss: 237800816640.0000
Epoch 288/330
49/49 [=====] - 0s 5ms/step - loss: 228984094720.0000 - val_loss: 237799964672.0000
Epoch 289/330
49/49 [=====] - 0s 7ms/step - loss: 228973412352.0000 - val_loss: 237802143744.0000
Epoch 290/330
49/49 [=====] - 0s 8ms/step - loss: 228960698368.0000 - val_loss: 237797212160.0000
Epoch 291/330
49/49 [=====] - 0s 7ms/step - loss: 228965892096.0000 - val_loss: 237799686144.0000
Epoch 292/330
49/49 [=====] - 0s 5ms/step - loss: 228958453760.0000 - val_loss: 237796261888.0000
Epoch 293/330
49/49 [=====] - 0s 6ms/step - loss: 228964335616.0000 - val_loss: 237801455616.0000
Epoch 294/330
49/49 [=====] - 0s 6ms/step - loss: 228954914816.0000 - val_loss: 237800718336.0000
Epoch 295/330
49/49 [=====] - 0s 6ms/step - loss: 228960337920.0000 - val_loss: 237800374272.0000
Epoch 296/330
49/49 [=====] - 0s 5ms/step - loss: 229037457408.0000 - val_loss: 237807140864.0000

```

Figura 57. Ejecución Red neuronal

Se grafica la curva de aprendizaje la cual debe tener una caída suave, eso implica que el modelo está aprendiendo y no memorizando. Como resultado de esta ejecución se verifica que efectivamente el modelo está aprendiendo cumpliendo de esta manera el objetivo. Es decir, el aprendizaje es suave como lo esperado, por lo cual no aplicaremos ajuste de datos y el análisis lo realizaremos con la data real sin ajuste a los valores atípicos. Ver figura 59.

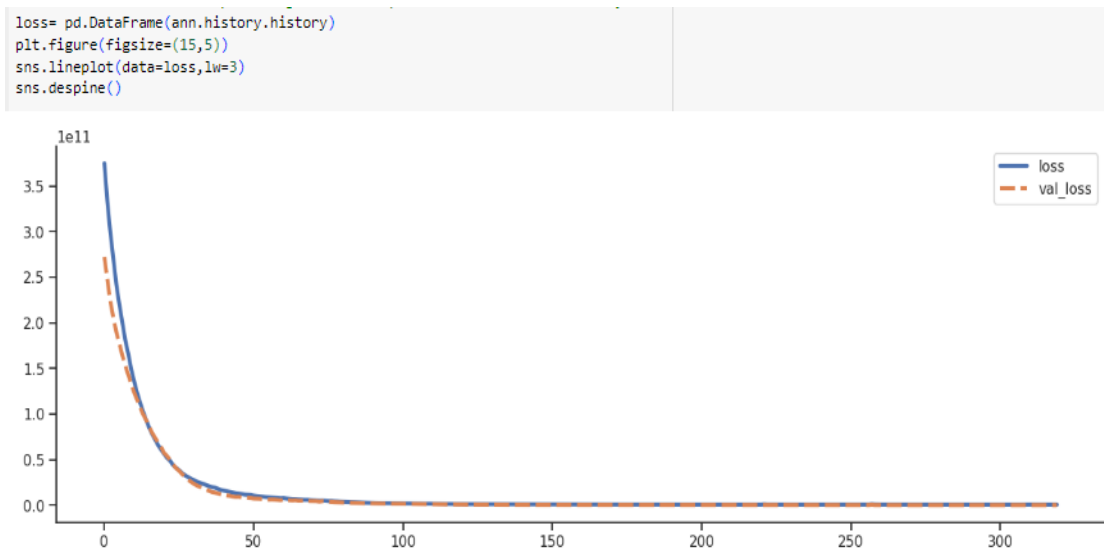


Figura 58. Grafica curva de aprendizaje.

Calcularemos el error absoluto medio, mismo que estará representado en dólares, es decir cuánto podría desviarse la predicción del margen operacional, respecto al valor real. Como se puede evidenciar el modelo está arrojando un error muy por debajo de la media de la variable margen operacional, que se analizó inicialmente. Ver figura 60.

```

y_predic= ann.predict(x_test)
errorAbsoluto= mean_absolute_error(y_test,y_predic)
print(errorAbsoluto)

```

98/98 [=====] - 0s 1ms/step
1235.5524606541944

Figura 59. Error medio absoluto.

Se eliminará la columna de margen operacional y se colocará en un h1. Ver figura 61.

```

[24] #eliminamos la columna margen operacional
      h1= dft.drop('FC_MONTO_MARGEN_OPERACIONAL',axis=1).iloc[0]

```

```

[38] h1
      FC_MONTO_PLANILLA          0.00
      FC_MONTO_PLANILLA_CONTABLE  0.00
      FC_MONTO_COSTOS           97357.66
      FC_MONTO_PRODUCCION        0.00
      FC_MONTO_SUBCONTRATO       0.00
      Name: 0, dtype: float64

```

Figura 60. Eliminar variable a predecir del modelo inicial.

Se genera el valor predicho de la fila número 1 y obtenemos como resultado el margen operacional predicho, en este caso cumpliendo con una proximidad de -69286.1655685. Esto es adecuado como resultado de este análisis. Ver figura 62.

```
[41] #escalamos los datos de dato de la fila num. 1
      scaler = StandardScaler()
      scaler.fit(h1.values.reshape(-1, 5))
      h1N= scaler.transform(h1.values.reshape(-1,5))

#Este es el margen operacional predicho
print("Marge Operacional Predicho",ann.predict(h1N)[0,0])

1/1 [=====] - 0s 65ms/step
Marge Operacional Predicho -69286.1655685
```

Figura 61. Margen operacional Predicho.

Por el momento no se pondrá en producción este análisis, sin embargo, se generará el archivo denominado ANNMargenOperacional.h5, el cual se podrá implementar mediante la ejecución de un JSON, para obtener el margen operacional y ponerlo en producción. Ver figura 63.

```
#Generamos el archivo modelANNMargenOperacional.h5 el cual se podría poner en producción
ann.save("/content/sample_data/modelANNMargenOperacional.h5")
```

Figura 62. Generación del archivo para posterior puesta en marcha.

3.8. Evaluación y Usabilidad.

Para asegurarnos de que la solución que hemos implementado sea fácil de usar y cumpla con las necesidades de los usuarios, se llevarán a cabo encuestas de usabilidad de la solución implementada. Para ello, utilizaremos el método SUS.

Resultados SUS Dashboard CMP

Después de analizar la accesibilidad y la utilidad de la solución de inteligencia empresarial (BI) que se ha puesto en marcha, los resultados resultaron alentadores. Con una calificación promedio del 85,43%, queda claro que esta solución es altamente funcional, adaptable y fiable en términos de información. Además, destaca por su capacidad para presentar los datos de manera visualmente atractiva y comprensible.

Usuarios	Te resultó fácil de usar el Dashboard	Consideras que requieres ser experto en el área para explorar el sistema	Fue fácil para ti moverte por el Dashboard y encontrar lo que necesitabas	Consideras que se seleccionó mal la herramienta de visualización	Encontraste rápidamente la información que estabas buscando	Piensas que llevaría mucho tiempo para dominar eficientemente la información	Cómo describirías la forma en que estaban organizados los datos en el Dashboard	Hubo algo que te confundiera o te resultara difícil de entender en el Dashboard	Qué tan contento estás en general con tu experiencia usando el Dashboard	Consideras que el sistema es muy complicado de utilizar	SUS	Ponderación Final
usuario1	5	1	5	1	5	1	5	2	5	1	39	97,5
usuario2	4	2	5	2	5	2	5	1	4	1	35	87,5
usuario3	5	2	5	1	5	2	4	1	4	2	35	87,5
usuario4	4	1	5	4	5	2	5	1	5	5	31	77,5
usuario5	5	2	5	2	5	2	5	2	5	1	36	90
usuario6	5	1	4	1	5	2	5	2	5	2	36	90
usuario7	5	1	5	2	4	2	4	2	5	5	31	77,5
usuario8	5	2	5	1	5	1	5	2	5	2	37	92,5
usuario9	3	2	4	5	5	2	5	1	5	2	30	75
usuario10	4	1	4	2	4	2	4	2	5	2	32	80
usuario11	5	2	5	2	5	4	5	2	4	2	32	80
usuario12	5	1	4	1	4	2	4	1	4	1	35	87,5
usuario13	5	2	4	2	4	4	4	1	5	2	31	77,5
usuario14	4	2	5	1	5	2	5	2	5	2	35	87,5
usuario15	5	1	5	5	5	2	4	2	4	2	31	77,5
usuario16	5	5	4	1	5	1	5	1	5	1	35	87,5
usuario17	4	2	4	1	5	1	4	2	5	1	35	87,5
usuario18	5	1	5	1	5	1	5	1	4	2	38	95
usuario19	5	4	5	2	5	1	5	1	5	2	35	87,5
usuario20	5	2	4	2	4	2	5	1	5	2	34	85
usuario21	5	5	5	1	5	2	5	2	5	1	34	85
usuario22	4	2	4	2	4	1	5	2	4	1	33	82,5
usuario23	5	1	5	2	5	2	5	2	5	2	36	90
Promedio	4,65	1,96	4,61	1,91	4,74	1,87	4,7	1,57	4,7	1,91	34,17	85,43

Figura 63. Resultados Ponderados de Usabilidad BI.

La fórmula empleada para la ponderación fue la siguiente: $=(B2-1)+(5-C2)+(D2-1)+(5-E2)+(F2-1)+(5-G2)+(H2-1)+(5-I2)+(J2-1)+(5-K2)$. En este proceso, se resta -1 a las respuestas de los ítems pares y 5 a las respuestas de los ítems impares. Luego, se suman los resultados de los ítems pares e impares para obtener el puntaje SUS por fila. Posteriormente, este puntaje se multiplica por 2.5 para obtener el puntaje SUS Final. El promedio del puntaje SUS Final fue de 85.43 sobre 100. Lo cual indica que el sistema BI implementado en el C.E.E. es fácil de usar y entender.

4. CONCLUSIONES.

- Las necesidades del C.E.E. se satisficieron por completo con la implementación de una solución de BI para la Evaluación, Seguimiento y Control de Proyectos. Para lograr los objetivos establecidos, fue necesario recopilar los requisitos para comprender la problemática de manera más precisa. Esto implicó trabajar con directores de áreas como construcción y planificación estratégica. Como resultado, se determinó que el enfoque sería en los procesos de la cadena de valor institucional. Esta premisa, que identificaba el objetivo principal del proyecto, fue fundamental para el desarrollo y la comprensión de la solución durante su desarrollo e implementación.
- Se implementó la solución integral, sólida, confiable y robusta de BI, con la finalidad que las áreas directivas administrativas de nivel superior y medio accedan a la información de forma más entendible, resumida y gráfica para el control y evaluación de los proyectos de CCVSP, mediante tableros de visualización (dashboards) estandarizados, optimizados y clasificados de acuerdo a las diferentes áreas de análisis (económicos, control y evaluación de proyectos), y de esta manera la gerencia pueda optimizar los recursos económicos y humanos, para la mejora continua de la Institución.
- Como parte de la funcionalidad, se llevó a cabo una encuesta de satisfacción del cliente con la finalidad de evaluar la conformidad del usuario en lo referente al uso y manejo de la aplicación gerencial de BI. Esta encuesta permitió analizar la facilidad de uso de la solución de BI, implementada para la supervisión y evaluación de proyectos., utilizando el método de Escala de Usabilidad del Sistema (SUS). En el conjunto de preguntas, se encuentran preguntas específicas sobre la facilidad de uso, flexibilidad y presentación visual del aplicativo. Estas preguntas fueron generadas de manera clara para evaluar estas variables, asignándoles un peso en porcentaje del total. Los resultados obtenidos fueron positivos, con un promedio de aceptación del 88%, lo que demuestra que el sistema es de mucha ayuda para el equipo gerencial dando eficiencia en el ahorro de tiempo y recursos para el análisis.
- Para la elaboración del proyecto de titulación, se emplearon datos provenientes de la base de datos de pruebas debido a consideraciones de

confidencialidad. A pesar de esto, nos hemos asegurado de que la calidad de los datos proporcionados sea óptima para el desarrollo del proyecto en cuestión.

5. RECOMENDACIONES.

- Es importante revisar con regularidad el funcionamiento de los paneles de visualización y la ubicación de los datos almacenados. De esta manera, podemos detectar problemas o bloqueos y tomar medidas para arreglarlos antes de que se conviertan en un gran problema.
- Es fundamental cuidar la integridad de los datos, especialmente aquellos provenientes del sistema transaccional SICEE. Esto implica implementar controles y validaciones rigurosas en fechas, formatos de texto y campos, asegurando que los datos estén limpios y actualizados. También es crucial eliminar cualquier información obsoleta o duplicada, y mantener nuestros procesos de recopilación y análisis de datos al día para garantizar que contemos siempre con la información más precisa y útil disponible
- Es de vital importancia realizar capacitaciones continuas a los usuarios. Es importante que todas las características y herramientas disponibles sean aprovechadas al máximo provecho de ellas y les permita tomar las mejores decisiones.
- Cada vez que se realicen cambios en la solución de inteligencia de negocios, es importante hacerlo de forma organizada y controlada. Esto nos permite evitar problemas como la redundancia en los gráficos o la repetición de procesos de análisis. En otras palabras, debemos seguir una secuencia organizativa que esté alineada con los procesos institucionales para garantizar una implementación efectiva y eficiente de los cambios.
- Evaluar la usabilidad de los usuarios para conocer su percepción sobre la solución y qué podríamos mejorar. Esto nos ayudará a identificar áreas donde podemos hacer cambios para seguir cumpliendo con sus necesidades y expectativas.

Biografía

- [1] J. García, “El Project Manager aplicado a la construcción,” *Partida Doble*, vol. 164, p. 26 a 33, 2005.
- [2] V. F. Villarreal, M. X. Sanchez, and S. C. A. Rincón, “Contabilización del cuarto elemento del costo,” *Entramado*, vol. 4, pp. 2–15, 2008.
- [3] L. E. Silva Solano, “Business Intelligence: un balance para su implementación,” *Innovag*, pp. 27–36, 2017, [Online]. Available: <https://revistas.pucp.edu.pe/index.php/innovag/article/view/19742/19808>
- [4] D. A. ARIAS ROMERO, “DEFINICIÓN E IMPLEMENTACIÓN DE DASHBOARD PARA EL SEGUIMINETO Y CONTROL DE LAS OBRAS CIVILES VIALES EN LA CIUDAD DE BOGOTÁ,” UNIVERSIDAD LA GRAN COLOMBIA, 2023. [Online]. Available: https://repository.ugc.edu.co/bitstream/handle/11396/7854/Arias_Diego_2023.pdf?sequence=1&isAllowed=y
- [5] L. F. LAPUERTA PAREDES, “IMPLEMENTACIÓN DE UN DASHBOARD COMO HERRAMIENTA PARA DEFINIR TERRITORIOS DE LOS VISITADORES MÉDICOS DE UN LABORATORIO FARMACÉUTICO EN ECUADOR,” *Fac. Ing. Maest. EN Sist. Inf. MENCIÓN Cienc. DATOS*, vol. 0, no. 0, p. 117, 2023, [Online]. Available: <https://repositorio.puce.edu.ec/server/api/core/bitstreams/8d7dc51f-331c-4e71-9141-cbbd2e5e2dbb/content>
- [6] A. K. Vivas Paspuel and D. A. Vivas Paspuel, “Estadísticas y análisis del Covid-19 en Ecuador utilizando Microsoft Power BI,” *Rev. INGENIO*, vol. 04, pp. 1–39, 2021, doi: <https://doi.org/10.29166/ingenio.v4i1.3068>.
- [7] L. R. Quisaguano Collaguazo, G. G. Paula, Esquivel, S. P. David, Darwin, and Á. J. Alajo, Lasluisa, “Desarrollo de inteligencia empresarial: Análisis comparativo entre Power BI y Qlik Sense en la implementación de tableros estratégicos de indicadores clave,” *Technol. Rain J. ISSN 2953-464X*, vol. 3, p. 15od, 2024, doi: <https://doi.org/10.55204/trj.v3i1.e30>.
- [8] IFD Technologies, “Ventajas y las desventajas del Power BI ~ IFD Technologies.” 2021. [Online]. Available: <https://ifd.com.mx/es/power-bi/>
- [9] J. Yala, J. Ortiz, C. Guevara, and E. Maya, “Herramientas de Business Intelligence (BI) modernas, basadas en memoria y con lógica asociativa,” *Rev. PUCE. ISSN 2528-8156. NÚM.106.*, vol. 0, no. 0, p. 376, 2018, [Online]. Available: <https://www.revistapuce.edu.ec/index.php/revpuce/article/view/144/246>
- [10] D. A. Vanegas, G. M. Tarazona Bermúdez, and L. A. Rodríguez Rojas, “Mejora de la toma de decisiones en ciclo de ventas del subsistema comercial de servicios en una empresa de IT,” *abril 2020*, Bogotá, Colombia, p. 183, Apr. 2020. doi: <https://doi.org/10.14483/23448350.15241>.
- [11] J. HAN, M. KAMBER, and J. PEI, “Data Mining: Concepts and

- Techniques,” *DATA Min. CONCEPTS Tech. 3RD Ed.*, vol. 44, no. 8, pp. 1689–1699, 2011, doi: 10.1088/1751-8113/44/8/085201.
- [12] Ministerio de Tecnologías de la Información y las Comunicaciones, “Herramientas de analítica para la explotación de datos,” *Herramientas analítica para la Explot. datos73*, vol. 1, p. 73, 2018, [Online]. Available: https://herramientas.datos.gov.co/sites/default/files/2020-11/Inventario_herramientas_analitica_0.pdf
- [13] Y. Xu, Y. Zhou, P. Sekula, and L. Ding, “Machine learning in construction: From shallow to deep learning,” *Dev. Built Environ.*, vol. 6, p. 100045, May 2021, doi: 10.1016/J.DIBE.2021.100045.
- [14] P. T. la economía de nuestros países entrenando a la próxima generación de profesionales en Tecnología, “Regresión lineal en machine learning.”
- [15] Oracle Corporation, “Oracle ® Database PL/SQL Language Reference,” *Oracle*, vol. 1, no. July, 2015.
- [16] C. y P. INESEM Bussiness School Escuela de lideres Master Online, “Data mart vs Data warehouse: diferencias.” [Online]. Available: <https://www.inesem.es/revistadigital/informatica-y-tics/diferencia-entre-data-mart-y-data-warehouse/>
- [17] H. Zerpa, R. García, and Izquierdo Henry, “DATAMART BASADO EN EL MODELO ESTRELLA PARA LA IMPLEMENTACIÓN DE INDICADORES CLAVE DE DESEMPEÑO COMO SALIDA DEL BIG DATA,” *UNIVERSIDAD, Cienc. y Tecnol.*, vol. 24, p. 54, 2020, [Online]. Available: <https://orcid.org/0000-0003-0499-2089>
- [18] Diego Armando Vanegas, Giovanni Mauricio Tarazona, and Luz Andrea Rodriguez Rojas, “Dialnet-ImplementacionDeUnDataWarehouseMedianteLaMetodolog-8229747 (4),” *Rev. Cient.*, vol. 38, pp. 174–183, 2019.
- [19] A. M. Jaramillo Tacuri and P. A. S. Leopoldo, “Diseño de un modelo físico de Data Warehouse para la gestión de incidencias para una empresa de telecomunicaciones, aplicando la metodología Hefesto Design of a physical Data Warehouse model for incident management for a telecommunications company, applyi,” *Ciencias de la computación y telecomunicaciones*, Cuenca, Ecuador, p. 118, 2019. doi: 10.23857/pc.v4i7.1026.
- [20] E. I. Leonard Brizuela and C. B. Yudi, “Metodologías para desarrollar Almacén de Datos.,” *Rev. Arquit. e Ing. E-ISSN 1990-8830*, vol. 3, p. 12, 2013, [Online]. Available: <http://www.redalyc.org/articulo.oa?id=193930080003>
- [21] Brandy Klug, “An Overview of the System Usability Scale in Library Website and System Usability Testing,” *J. Libr. User Exp.*, vol. 1, no. 6, p. 10, 2017, doi: <https://doi.org/10.3998/weave.12535642.0001.602>.