



Facultad de Ingeniería de Sistemas y Electrónica
Carrera profesional de Ingeniería de Sistemas

Informe del proyecto final:

Solución de inteligencia de negocios
Caso: Admisión Universidad Nacional de San Agustín

Curso:

Inteligencia de negocios

Autores:

Canaza Tito, Eddy Wilmer
Yana Espinoza, Luis Fernando

Asesor(a):

Ing. Dennis Larry Quispe Chávez

Arequipa, marzo de 2017

Índice general

Presentación	3
1 Marco teórico	4
1.1 Inteligencia de negocios	4
1.2 Inteligencia de negocios en diferentes niveles	4
1.3 Actividades	5
1.4 Enfoque de desarrollo	6
1.5 Etapas de ingeniería y pasos de desarrollo	9
1.5.1 Justificación	9
1.5.2 Planificación	9
1.5.3 Análisis de negocio	10
1.5.4 Diseño	11
1.5.5 Construcción	12
1.5.6 Despliegue	13
1.6 Arquitectura y componentes	14
1.6.1 Data warehouse	14
1.6.2 Extract Transform Load (ETL)	14
1.6.3 Modelo de datos (BISM)	15
1.6.4 Visualización de datos	15
1.6.5 Gestión de datos maestros (MDM)	15
1.6.6 Servicios de calidad de datos	15
2 Aplicación	16
2.1 Análisis funcional	16
2.1.1 Descripción del problema	16
2.1.2 Definición de requerimientos de negocio	16
2.1.3 Análisis de información	16
2.2 Análisis técnico	17
2.2.1 Diseño de base de datos	17
2.2.2 Diseño de ETLs	33
2.2.3 Diseño de cubos	43
2.2.4 Diseño de reportes	48
Conclusiones	50
Recomendaciones	51

Índice de figuras

1.1	Metas específicas en cada nivel de la organización	4
1.2	Medidas concretas en cada nivel de la organización	5
1.3	Timing en cada nivel de la organización	5
1.4	Etapas de la ingeniería	7
1.5	El concepto de release de sistemas de BI	8
1.6	Dependencias de los pasos de desarrollo	13
1.7	Arquitectura y componentes	14
2.1	Modelo Entidad Relación Data Warehouse	26
2.2	Data Source View Cubo Ingresantes	43
2.3	Data Source View Cubo Postulantes	46

Índice de cuadros

2.1	Estructura de la hoja Areas	17
2.2	Estructura de la hoja Facultades	17
2.3	Estructura de la hoja Carreras	17
2.4	Estructura de la hoja TiposProcesosAdmision	18
2.5	Estructura de la hoja ProcesosAdmision	18
2.6	Estructura de la hoja Sedes	18
2.7	Estructura de la hoja EspecialidadesEducacion	19
2.8	Estructura de la hoja TiposExtraordinarios	19
2.9	Estructura de la hoja Ingresantes	20
2.10	Estructura de la hoja Estadisticas	20

Presentación

El presente documento corresponde al análisis y diseño de la solución de inteligencia de negocios propuesta para el área de Admisión de la Universidad Nacional de San Agustín en Arequipa.

En el primer capítulo, se define el concepto de inteligencia de negocios, la importancia, las actividades y la arquitectura de este tipo de software.

En el segundo capítulo, se realiza el análisis funcional y técnico de la solución de inteligencia de negocios.

La construcción de la herramienta, se realizó utilizando los componentes de inteligencia de negocios de Microsoft SQL Server 2016 y SQL Server Data Tools (SSDT) para Visual Studio.

El documento detalla cada uno de los componentes propuestos, entre ellos; la extracción, transformación y carga de las fuentes de datos con Integration Services; el diseño de los cubos con Analysis Services y la generación de reportes con Reporting Services.

Capítulo 1. Marco teórico

1.1. Inteligencia de negocios

Business Intelligence (BI) no es ni un producto ni un sistema. Se trata de una arquitectura y una colección de aplicaciones y bases de datos operativas integradas, así como de apoyo a las decisiones, que proporcionan a la comunidad empresarial un fácil acceso a los datos empresariales [1].

Hoy en día, BI es uno de los temas candentes en la mayoría de los mercados de trabajo en todo el mundo. La mayoría de las compañías están estableciendo o planeando establecer un sistema de Inteligencia de Negocios y un almacén de datos (DW). El conocimiento relacionado con el BI y el almacén de datos están en gran demanda en el mercado de trabajo [2].

1.2. Inteligencia de negocios en diferentes niveles

La inteligencia de negocios debe ser utilizados en todos los niveles de una organización para promover la toma de decisiones eficaces [3]. Si bien es cierto que la inteligencia de negocios es útil en toda la organización, no se necesita el mismo tipo de información en cada nivel. Diferentes niveles dentro de la organización requieren diferentes tipos de inteligencia empresarial para la toma de decisiones eficaces, como se nota en las figuras 1.1, 2.1 y 1.3.



Figura 1.1: Metas específicas en cada nivel de la organización [3]

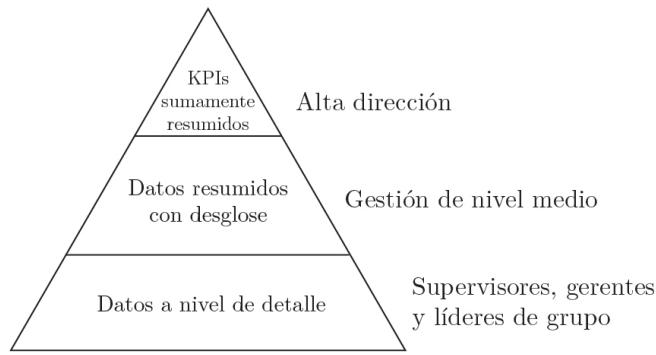


Figura 1.2: Medidas concretas en cada nivel de la organización [3]



Figura 1.3: Timing en cada nivel de la organización [3]

1.3. Actividades

Los sistemas de apoyo a la toma de decisiones de BI facilitan muchas actividades [1] como las que se listan a continuación:

- El análisis multidimensional
- Análisis de flujo de clics
- Minería de datos
- Predicción
- Análisis de negocio
- Preparación del cuadro de mando integral (CMI)
- Visualización

- Consultas, informes y gráficos
- Análisis geoespacial
- Gestión del conocimiento
- Implementación del portal empresarial
- Minería de texto, contenido y voz
- Acceso al tablero digital
- Otras actividades multifuncionales

Ejemplos de bases de datos de soporte para la toma de decisiones de BI, como las siguientes:

- Data warehouses para toda la empresa
- Data marts (funcionales y departamentales)
- Data warehouse de exploración (estadística)
- Bases de datos para minería de datos
- Data warehouses web (para datos de flujo de clics)
- Data warehouses operativos (ODS)

1.4. Enfoque de desarrollo

Al igual que el proceso de ingeniería estructural e ingeniería de sistema, un sistema de BI, pasa por seis etapas, desde el inicio hasta la implementación [1], como se ilustra en la figura 1.4.

- 1. Justificación:** Evaluar la necesidad empresarial que da lugar al nuevo proyecto de ingeniería
- 2. Planificación:** Desarrollar planes estratégicos y tácticos, que establezcan cómo se llevará a cabo y se desplegará el proyecto de ingeniería.

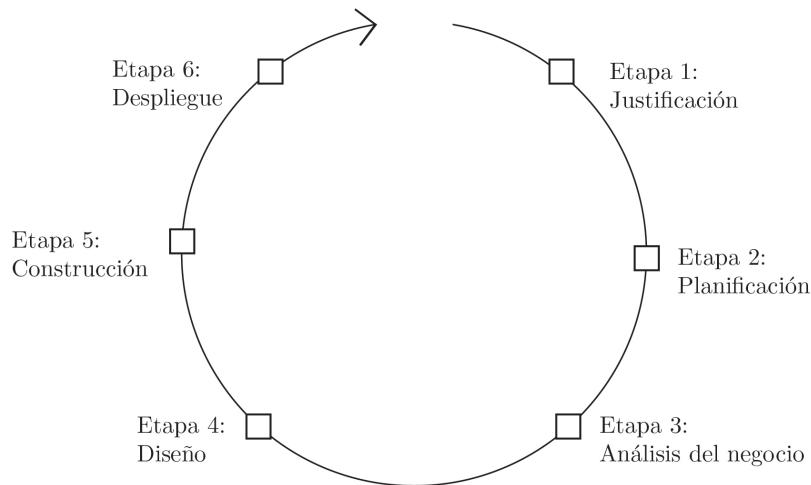


Figura 1.4: Etapas de la ingeniería [1]

- 3. Análisis de negocio:** Realizar un análisis detallado del problema empresarial o la oportunidad de negocio para obtener una comprensión sólida de los requisitos de negocio para una solución potencial (producto).
- 4. Diseño:** Concebir un producto que resuelva el problema del negocio o que permita la oportunidad de negocio.
- 5. Construcción:** Construir el producto, que debe proporcionar un retorno de la inversión dentro de un marco de tiempo predefinido.
- 6. Despliegue:** Implementar o vender el producto terminado, luego medir su efectividad para determinar si la solución cumple, excede o no cumple con el retorno esperado de la inversión.

Como indica la flecha de la figura 1.4, los procesos de ingeniería son iterativos. Una vez desplegado, un producto se mejora y mejora continuamente en función de la retroalimentación de la comunidad empresarial que utiliza el producto. Cada iteración produce una nueva versión del producto, a medida que el producto evoluciona y madura.

Las metodología de desarrollo de software tradicional no cubren los aspectos organizacionales y estratégicos fundamentales para BI, L. Moss [1] propone un modelo para solventar estas falencias, véase en la figura 1.5.

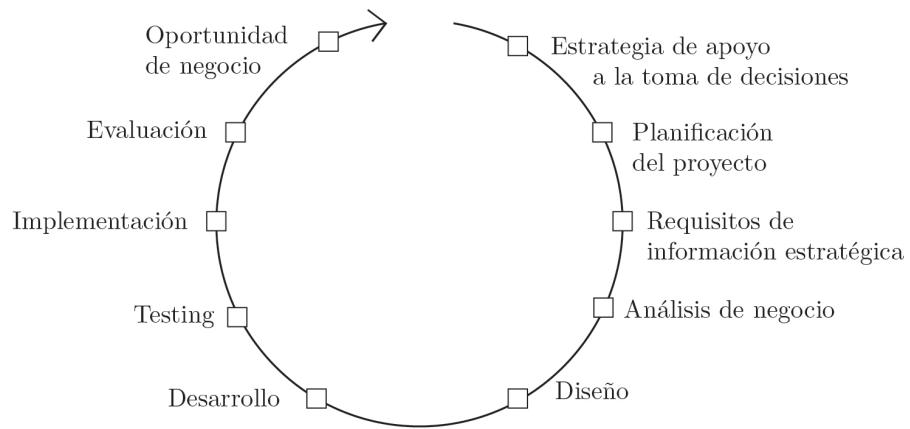


Figura 1.5: El concepto de release de sistemas de BI [1]

Los sistemas de BI son distintas a los sistemas stand-alone, las mayores diferencias son:

- Las aplicaciones de BI son en su mayoría impulsadas por la oportunidad de negocio en lugar de las necesidades del negocio.
- Las aplicaciones de BI implementan una estrategia de apoyo a la toma de decisiones entre organizaciones, en lugar de los silos de apoyo a las decisiones departamentales.
- Los requisitos de soporte de toma de decisiones de BI son en su mayoría requerimientos de información estratégica en lugar de requerimientos funcionales operacionales.
- El análisis de los proyectos de BI enfatiza el análisis de negocios en lugar de análisis de sistemas, y el análisis es la actividad más importante al desarrollar un entorno de soporte de decisiones de BI.
- Las evaluaciones de lanzamiento de aplicaciones de BI en curso promueven el desarrollo iterativo y el concepto de lanzamiento de software en lugar del desarrollo de big-bang.

1.5. Etapas de ingeniería y pasos de desarrollo

Los proyectos de BI se organizan de acuerdo con las mismas seis etapas comunes a cada proyecto de ingeniería. Dentro de cada etapa de ingeniería, se llevan a cabo ciertos pasos para ver el proyecto de ingeniería hasta su finalización. Moss [1], describe 16 pasos de desarrollo dentro de estas etapas, como se describe a continuación.

1.5.1. Justificación

Paso 1: Evaluación de casos de negocio

El problema empresarial o la oportunidad de negocio es definido y una solución de BI es propuesta. Cada lanzamiento de la aplicación de BI debe estar justificado por costos y debe definir claramente los beneficios de resolver un problema de negocio o aprovechar una oportunidad de negocio.

1.5.2. Planificación

Paso 2: Evaluación de la infraestructura de la empresa

Dado que las aplicaciones de BI son iniciativas interorganizativas, debe crearse una infraestructura empresarial para darles soporte. Una infraestructura empresarial tiene dos componentes:

- **Infraestructura técnica**, que incluye hardware, software, middleware, sistemas de administración de bases de datos, sistemas operativos, componentes de red, repositorios de metadatos, utilidades, etc.
- **Infraestructura no técnica**, que incluye estándares de metadatos, estándares de datanaming, modelo de datos lógicos empresariales (evolucionando), metodologías, directrices, procedimientos de prueba, procesos de control de cambios, procedimientos para la gestión de problemas y resolución de disputas, etc.

Paso 3: Planificación del proyecto

Los proyectos de apoyo a la decisión de BI son extremadamente dinámicos. Los cambios en el alcance, el personal, el presupuesto, la tecnología, los representantes

empresariales y los patrocinadores pueden afectar gravemente el éxito de un proyecto. Por lo tanto, la planificación del proyecto debe ser detallada, y el progreso real debe ser vigilado y reportado de cerca.

1.5.3. Análisis de negocio

Paso 4: Definición de requerimientos del proyecto

La gestión del alcance del proyecto es una de las tareas más difíciles en los proyectos de apoyo a la decisión de BI. El deseo de tener todo al instante es difícil de reducir, pero restringir ese deseo es uno de los aspectos más importantes de la negociación de los requisitos para cada entregable. Los equipos de proyecto deben esperar que estos requisitos cambien a lo largo del ciclo de desarrollo mientras los empresarios aprenden más sobre las posibilidades y las limitaciones de la tecnología BI durante el proyecto.

Paso 5: Análisis de datos

El mayor reto para todos los proyectos de soporte de decisiones de BI es la calidad de los datos de origen. Los malos hábitos desarrollados durante décadas son difíciles de romper, y los daños resultantes de los malos hábitos son muy costosos, consumen mucho tiempo y son tediosos de encontrar y corregir. Además, el análisis de datos en el pasado se limitó a la visión de una línea de negocio y nunca se consolidó o reconcilió con otros puntos de vista en la organización. Este paso toma un porcentaje significativo del tiempo asignado a toda la programación del proyecto.

Paso 6: Prototipo de la aplicación

El análisis de los entregables funcionales, que solía llamarse análisis de sistemas, se realiza mejor a través de la creación de prototipos para que pueda combinarse con el diseño de la aplicación. Nuevas herramientas y lenguajes de programación permiten a los desarrolladores probar o refutar relativamente rápidamente un concepto o una idea. El prototipado también permite a los empresarios ver el potencial y los límites de la tecnología, lo que les da la oportunidad de ajustar los requisitos de sus proyectos y sus expectativas.

Paso 7: Análisis del repositorio de meta-data

Tener más herramientas significa tener más metadatos técnicos además de los

metadatos del negocio, que usualmente se capturan en una herramienta de modelado de ingeniería de software asistida por computadora (CASE). Los metadatos técnicos deben asignarse a los metadatos del negocio y todos los metadatos deben almacenarse en un repositorio de metadatos. Los repositorios de metadatos pueden ser licenciados (comprados) o construidos. En cualquier caso, los requisitos para qué tipo de metadatos para capturar y almacenar deben documentarse en un modelo meta lógico. Al conceder licencias a un producto de repositorio de metadatos, los requisitos documentados en este modelo meta lógico deben compararse con el modelo meta del proveedor, si se proporciona uno. Además, deben analizarse los requisitos para entregar metadatos a la comunidad empresarial (por ejemplo, la función de ayuda en línea).

1.5.4. Diseño

Paso 8: Diseño de la base de datos

Una o más bases de datos de BI almacenarán los datos empresariales en forma detallada o agregada, dependiendo de los requisitos de informes de la comunidad empresarial. No todos los requisitos de reporte son estratégicos, y no todos son multidimensionales. Los esquemas de diseño de la base de datos deben coincidir con los requisitos de acceso a la información de la comunidad empresarial.

Paso 9: Diseño de ETL

El proceso de ETL es el proceso más complicado de todo el proyecto de apoyo a la decisión de BI. También es el menos glamoroso. Ventanas de procesamiento ETL (ventanas de proceso por lotes) son típicamente pequeñas, pero la mala calidad de los datos de origen suele requerir mucho tiempo para ejecutar los programas de transformación y limpieza. Finalizar el proceso ETL dentro de la ventana de lotes disponible es un desafío para la mayoría de las organizaciones.

Paso 10: Diseño de repositorio de meta-data

Si un repositorio de metadatos está licenciado, lo más probable es que tenga que ser mejorado con características que fueron documentadas en el metanálisis lógico pero que no son proporcionadas por el producto. Si se está construyendo un repositorio de metadatos, se debe tomar la decisión de si el diseño de la base de

datos del repositorio de metadatos estará basado en la entidad o en el objeto. En cualquier caso, el diseño tiene que cumplir con los requisitos del modelo meta lógico.

1.5.5. Construcción

Paso 11: Desarrollo de ETL

Muchas herramientas están disponibles para el proceso de ETL, algunas sofisticadas y otras sencillas. Dependiendo de los requisitos para la limpieza de datos y la transformación de datos desarrollados durante el Paso 5, Análisis de Datos, y el Paso 9, ETL Design, una herramienta ETL puede o no ser la mejor solución. En cualquier caso, con frecuencia se requiere preprocesar las extensiones de datos y escritura para complementar las capacidades de la herramienta ETL.

Paso 12: Desarrollo de la aplicación

Una vez que el esfuerzo de prototipado ha confirmado los requisitos funcionales, el verdadero desarrollo de la aplicación de acceso y análisis puede comenzar. Desarrollar la aplicación puede ser una simple cuestión de finalizar un prototipo operacional, o puede ser un esfuerzo de desarrollo más involucrado usando diferentes y más robustas herramientas de acceso y análisis. En cualquier caso, las actividades de desarrollo de aplicaciones front-end se llevan a cabo normalmente en paralelo con las actividades de desarrollo ETL de back-end y desarrollo de repositorios de metadatos.

Paso 13: Minería de datos

Muchas organizaciones no utilizan su entorno de soporte de decisiones de BI en la mayor medida posible. Las aplicaciones de BI a menudo se limitan a informes preescritos, algunos de los cuales no son ni siquiera nuevos tipos de informes sino reemplazos de informes antiguos. La verdadera recuperación proviene de la información oculta en los datos de la organización, que sólo se puede descubrir con herramientas de minería de datos.

Paso 14: Desarrollo de repositorio de meta-data

Si se toma la decisión de construir un repositorio de metadatos en lugar de uno de licencia, normalmente un equipo independiente se encarga del proceso de desarrollo. Esto se convierte en un subproyecto considerable en el proyecto general de BI.

1.5.6. Despliegue

Paso 15: Implementación

Una vez que el equipo ha probado exhaustivamente todos los componentes de la aplicación de BI, el equipo desarrolla las bases de datos y las aplicaciones. La capacitación está programada para el personal de negocios y otros interesados que usarán la aplicación de BI y el repositorio de metadatos. Comienzan las funciones de soporte, que incluyen el funcionamiento del servicio de asistencia, el mantenimiento de las bases de datos de destino de BI, la programación y ejecución de trabajos por lotes ETL, el monitoreo del rendimiento y el ajuste de bases de datos.

Paso 16: Evaluación del despliegue

Con un concepto de lanzamiento de aplicaciones, es muy importante beneficiarse de las lecciones aprendidas de los proyectos anteriores. Se deben examinar los plazos perdidos, los excesos de costos, las disputas y las resoluciones de disputas, y los ajustes del proceso deben hacerse antes de que comience el próximo lanzamiento. Todas las herramientas, técnicas, directrices y procesos que no fueron útiles deben ser reevaluados y ajustados, posiblemente descartados.

No es necesario realizar los pasos de desarrollo en secuencia; la mayoría de los equipos de proyectos probablemente los ejecutarán en paralelo. Sin embargo, debido a que existe un orden natural de progresión de una etapa de ingeniería a otra, existen ciertas dependencias entre algunos de los pasos de desarrollo, como se ilustra en la figura 1.6.

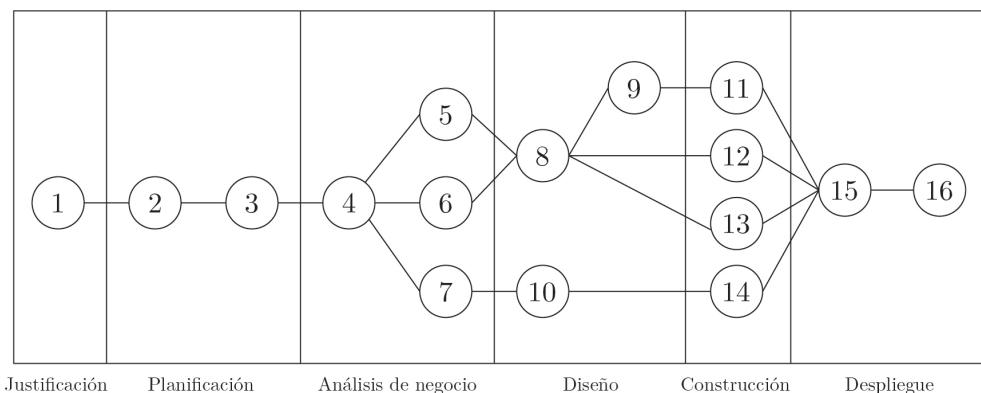


Figura 1.6: Dependencias de los pasos de desarrollo [4]

1.6. Arquitectura y componentes

La arquitectura y los componentes de BI difieren según las herramientas, el entorno, etc. La arquitectura mostrada en el diagrama anterior contiene componentes que son comunes en la mayoría de los sistemas de BI [4].

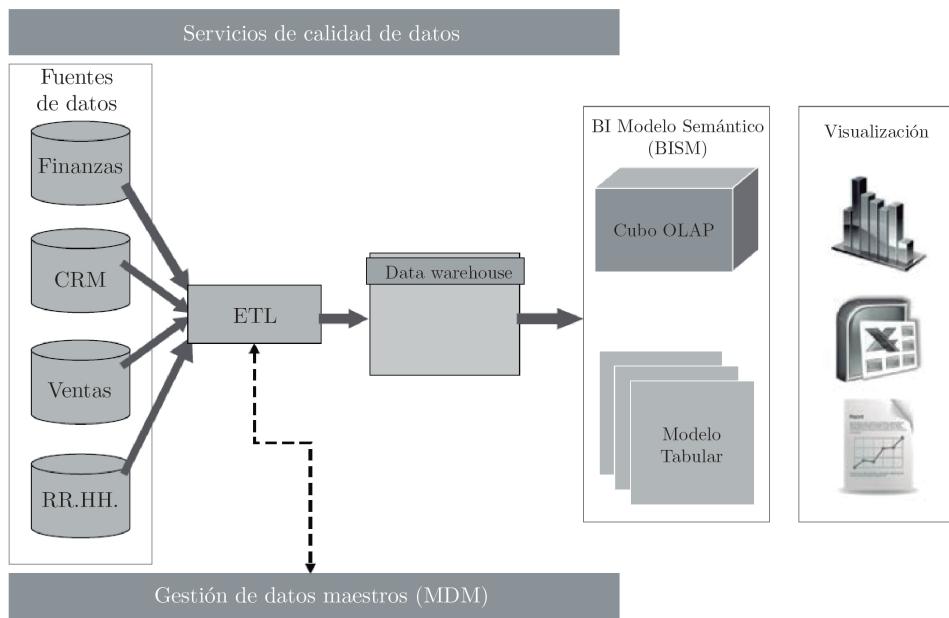


Figura 1.7: Arquitectura y componentes [4]

1.6.1. Data warehouse

El data warehouse es el núcleo del sistema BI. Un data warehouse es una base de datos construida con el propósito de análisis de datos e informes [4].

1.6.2. Extract Transform Load (ETL)

El proceso de Extract Transform Load (ETL) es la parte de la solución de BI que se encarga de la consolidación de datos. El proceso incluye la extracción de datos de diferentes fuentes, la transformación de forma que los datos se ajusten al data warehouse y la carga al data warehouse [4].

1.6.3. Modelo de datos (BISM)

El modelo de datos está basado en archivos o basado en memoria de los datos para producir respuestas muy rápidas a los reportes [4].

Microsoft divide esta capa en dos tecnologías: el cubo OLAP y el modelo tabular en memoria. El cubo OLAP es un almacenamiento de datos basado en archivos que carga datos de un data warehouse en un modelo en cubo. El cubo contiene información descriptiva como dimensiones y hechos [4].

1.6.4. Visualización de datos

La visualización de datos es una parte del sistema de BI que los usuarios pueden ver. Existen diferentes métodos para visualizar la información, como cuadros de mando estratégicos y tácticos, indicadores claves de rendimiento (KPI) e informes detallados o consolidados [4].

1.6.5. Gestión de datos maestros (MDM)

La gestión de datos maestros es el proceso de mantener la versión única de las entidades de datos maestros a través de múltiples sistemas [4].

1.6.6. Servicios de calidad de datos

La calidad de los datos es diferente en cada sistema operativo, especialmente cuando tratamos con sistemas heredados o sistemas que tienen una alta dependencia de las entradas de los usuarios. Dado que el sistema de BI se basa en datos, cuanto mejor sea la calidad de los datos, mejor será la salida de la solución de BI [4].

Capítulo 2. Aplicación

2.1. Análisis funcional

2.1.1. Descripción del problema

La Universidad Nacional de San Agustín, centro de educación superior ubicado en la ciudad de Arequipa, actualmente pone a disposición 4,000 vacantes para el ingreso y aproximadamente 25,000 personas rinden el examen de admisión para obtener uno alguno de estos lugares.

El área encargado de dirigir los procesos de admisión es el área de Admisión. Esta área actualmente no cuenta con una herramienta automatizada que le ayude a extraer conocimiento y le facilite la toma de decisiones dentro de su competencia.

Alguno de los problemas que se evidencia en el día a día de esta área son:

- Pérdida de tiempo, al realizar consultas a grandes volúmenes de información.
- Carencia de una plataforma que le permita revisar la información histórica.
- Carencia de una herramienta que le permita crear reportes interactivos.

2.1.2. Definición de requerimientos de negocio

Este proyecto tiene como objetivo mejorar la gestión del área de Admisión de la Universidad Nacional de San Agustín, incorporando elementos de inteligencia de negocios.

- Generar un Data Warehouse de los procesos de admisión desde el 2011 a 2017.
- Visualizar y analizar el Data Warehouse utilizando cubos OLAP.
- Generar reportes de los datos contenidos en el Data Warehouse.

2.1.3. Análisis de información

Las fuentes de datos que posee la universidad están almacenadas en hojas de cálculos. Incluye libros con información general y libros con los resultados de cada proceso de admisión.

2.2. Análisis técnico

2.2.1. Diseño de base de datos

A. Definición de fuentes de datos

a. **Libro Areas-Facultades-Carreras:** Este libro contiene la descripción de las áreas, facultades y carreras de la universidad.

i. **Areas:** Esta hoja contiene las áreas de la universidad, con la información que se muestra en el cuadro 2.1.

Cuadro 2.1: Estructura de la hoja Areas

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna

ii. **Facultades:** Esta hoja contiene la información de las facultades de la universidad, con la información que se muestra en el cuadro 2.2.

Cuadro 2.2: Estructura de la hoja Facultades

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna
Area	Texto	Asociado con hoja Areas

iii. **Carreras:** Esta hoja contiene las carreras profesionales que se dictan en la universidad, con la información que se muestra en el cuadro 2.3.

Cuadro 2.3: Estructura de la hoja Carreras

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna
Facultad	Texto	Asociado con hoja Facultades

b. **Libro Procesos-Admision:** Este libro contiene los tipos de admisión y la información de los procesos de admisión de la universidad.

- i. **TiposProcesosAdmision:** Esta hoja contiene los tipos de admisión que ofrece la universidad, con la información que se muestra en el cuadro 2.4.

Cuadro 2.4: Estructura de la hoja TiposProcesosAdmision

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna

- ii. **ProcesosAdmision:** Esta hoja contiene los procesos de admisión que lanzó la universidad, con la información que se muestra en el cuadro 2.5.

Cuadro 2.5: Estructura de la hoja ProcesosAdmision

Campo	Tipo	Observaciones
Año	Texto	Ninguna
Tipo	Texto	Asociado con hoja TiposProcesosAdmision
Fase	Texto	Expresado en números
Fecha	Fecha	Ninguna

- c. **Libro Sedes-Especialidades-Extraordinarios:** Este libro contiene la descripción de las sedes, especialidades de la carrera de educación y los tipos de exámenes extraordinarios.

- i. **Sedes:** Esta hoja contiene las sedes de la universidad, con la información que se muestra en el cuadro 2.6.

Cuadro 2.6: Estructura de la hoja Sedes

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna

- ii. **EspecialidadesEducacion:** Esta hoja contiene la información de las especialidades de la carrera de educación de la universidad, con la información que se muestra en el cuadro 2.7.

Cuadro 2.7: Estructura de la hoja EspecialidadesEducacion

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna

iii. **TiposExtraordinarios:** Esta hoja contiene los tipos de exámenes extraordinarios de la universidad, con la información que se muestra en el cuadro 2.8.

Cuadro 2.8: Estructura de la hoja TiposExtraordinarios

Campo	Tipo	Observaciones
Nombre	Texto	Ninguna

d. **Carpeta Resultados:** Esta carpeta contiene libros con los resultados de cada uno de los procesos de admisión de la universidad.

i. **Ingresantes:** Esta hoja contiene la información de los resultados de los exámenes de admisión de la universidad, con la información que se muestra en el cuadro 2.9.

Observaciones: Los libros de los procesos de admisión extraordinarios contienen el campo adicional Tipo. Los libros de los procesos de admisión sedes contiene el campo adicional Sede.

ii. **Estadísticas:** Esta hoja contiene la información de las vacantes y postulantes que se presentaron para los exámenes de admisión, con la información que se muestra en el cuadro 2.10.

Observaciones: Los libros de los procesos de admisión extraordinarios y sedes no contienen la información del número de vacantes e ingresantes.

Cuadro 2.9: Estructura de la hoja Ingresantes

Campo	Tipo	Observaciones
Carrera	Texto	Asociado con la hoja Carreras
Apellidos_Nombres	Texto	Expresada con la sintaxis “AP/AM, N”
Tipo	Texto	Exclusivo para extraordinarios
Puntaje	General	Expresado en decimal con una precisión de seis dígitos
Sede	Texto	Exclusivo para sedes
Especialidad	Texto	Exclusivo para la carrera de Educación

Cuadro 2.10: Estructura de la hoja Estadísticas

Campo	Tipo	Observaciones
Carrera	Texto	Asociado con hoja Carreras
Vacantes	General	Expresado en números
Postulantes	General	Expresado en números

B. Definición de base de datos Stage

- a. Nombre de base de datos: Admision_UNSA.Stage
- b. Ruta de los archivos:
 - i. Data: D:\Admision_UNSA\SQL\Admision_UNSA.Stage_Data.mdf
 - ii. Log: D:\Admision_UNSA\SQL\Admision_UNSA.Stage_Log.ldf
- c. Diccionario de datos

Nombre de tabla: Areas

Descripción: Tabla que contiene la información de las áreas.

Campo	Tipo y tamaño	Descripción
Nombre	VARCHAR(50)	Nombre del área

Nombre de tabla: Facultades

Descripción: Tabla que contiene la información de las facultades.

Campo	Tipo y tamaño	Descripción
Nombre	VARCHAR(100)	Nombre de la facultad
Area	VARCHAR(50)	Nombre del área

Nombre de tabla: Carreras

Descripción: Tabla que contiene la información de las carreras.

Campo	Tipo y tamaño	Descripción
Nombre	VARCHAR(100)	Nombre de la carrera
Facultad	VARCHAR(100)	Nombre de la facultad

Nombre de tabla: Modalidades

Descripción: Tabla que contiene la información de las modalidades de admisión.

Campo	Tipo y tamaño	Descripción
Nombre	VARCHAR(50)	Nombre del tipo de admisión

Nombre de tabla: ProcesosAdmision

Descripción: Tabla que contiene la información de los procesos de admisión.

Campo	Tipo y tamaño	Descripción
Anio	INT	Anio del proceso de admisión
Modalidad	VARCHAR(50)	Tipo del proceso de admisión
Fase	VARCHAR(50)	Fase del proceso de admisión
Fecha	DATE	Fecha del proceso de admisión

Nombre de tabla: Especialidades

Descripción: Tabla que contiene la información de las especialidades.

Campo	Tipo y tamaño	Descripción
Carrera	VARCHAR(100)	Carrera de la especialidad
Nombre	VARCHAR(100)	Nombre de la especialidad

Nombre de tabla: ModalidadesExtraordinarios

Descripción: Tabla que contiene la información de los tipos extraordinarios.

Campo	Tipo y tamaño	Descripción
Nombre	VARCHAR(50)	Nombre del tipo de extraordinario

Nombre de tabla: Sedes

Descripción: Tabla que contiene la información de las sedes.

Campo	Tipo y tamaño	Descripción
Nombre	VARCHAR(50)	Nombre de la sede

Nombre de tabla: Ingresantes

Descripción: Tabla que contiene la información de los ingresantes.

Campo	Tipo y tamaño	Descripción
Anio	INT	Anio del proceso de admisión
Modalidad	VARCHAR(50)	Tipo del proceso de admisión
Fase	VARCHAR(50)	Fase del proceso de admisión
OrdenMerito	INT	Puesto en el que ingresa
Carrera	VARCHAR(100)	Nombre de la carrera
ApellidosNombres	VARCHAR(100)	Nombre del postulante
Puntaje	DECIMAL(10,6)	Puntaje obtenido
Especialidad	VARCHAR(100)	Nombre de la especialidad
ModalidadExtraordinario	VARCHAR(50)	Modalidad de extraordinario
Sede	VARCHAR(50)	Nombre de la sede

Nombre de tabla: Estadisticas

Descripción: Tabla que contiene la información de las estadísticas.

Campo	Tipo y tamaño	Descripción
Anio	INT	Anio del proceso de admisión
Modalidad	VARCHAR(50)	Tipo del proceso de admisión
Fase	VARCHAR(50)	Fase del proceso de admisión
Carrera	VARCHAR(100)	Nombre de la carrera
NroVacantes	INT	Número de vacantes
NroPostulantes	INT	Número de postulantes

d. Mapeo de datos

Nombre de mapeo: Areas

Fuente origen:

Fuente destino:

Areas-Facultades-Carreras.xlsx

dbo.Areas

NOMBRE

Nombre

Nombre de mapeo: Facultades

Fuente origen:

Fuente destino:

Areas-Facultades-Carreras.xlsx

dbo.Facultades

NOMBRE

Nombre

AREA

Area

Nombre de mapeo: Carreras

Fuente origen:

Fuente destino:

Areas-Facultades-Carreras.xlsx

dbo.Carreras

NOMBRE

Nombre

FACULTAD

Facultad

Nombre de mapeo: Modalidades

Fuente origen:	Fuente destino:
Procesos-Admision.xlsx	dbo.Modalidades
NOMBRE	Nombre

Nombre de mapeo: Procesos de Admisión

Fuente origen:	Fuente destino:
Procesos-Admision.xlsx	dbo.ProcesosAdmision
ANIO	Anio
TIPO	Modalidad
FASE	Fase
FECHA	Fecha

Nombre de mapeo: Especialidades

Fuente origen:	Fuente destino:
Sedes-Especialidades-Extraordinarios.xlsx	dbo.Especialidades
CARRERA	Carrera
NOMBRE	Nombre

Nombre de mapeo: Modalidad extraordinarios

Fuente origen:	Fuente destino:
Sedes-Especialidades-Extraordinarios.xlsx	dbo.ModalidadesExtraordinarios
NOMBRE	Nombre

Nombre de mapeo: Sedes

Fuente origen:	Fuente destino:
Sedes-Especialidades-Extraordinarios.xlsx	dbo.Sedes
NOMBRE	Nombre

Nombre de mapeo: Ingresantes

Fuente origen:	Fuente destino:
Modalidad-Anio-Fase.xlsx	dbo.Ingresantes
CARRERA (Campo calculado)	Anio
NOMBRE (Campo calculado)	Modalidad
FASE (Campo calculado)	Fase
PUESTO	OrdenMerito
CARRERA	Carrera
APELLIDOS Y NOMBRE	ApellidosNombres
PUNTAJE	Puntaje
ESPECIALIDAD	Especialidad
TIPO	ModalidadExtraordinario
SEDE	Sede

Nombre de mapeo: Estadísticas

Fuente origen:	Fuente destino:
Modalidad-Anio-Fase.xlsx	dbo.Estadisticas
ANIO (Campo calculado)	Anio
MODALIDAD (Campo calculado)	Modalidad
FASE (Campo calculado)	Fase
CARRERA	Carrera
VACANTES	NroVacantes
POSTULANTES	NroPostulantes

C. Definición de base de datos Data Warehouse

a. **Nombre de base de datos:** Admision_UNSA_DW

b. **Ruta de los archivos:**

i. **Data:** D:\Admision_UNSA\SQL\Admision_UNSA_DW_Data.mdf

ii. **Log:** D:\Admision_UNSA\SQL\Admision_UNSA_DW_Log.ldf

c. Definición de Modelo Entidad Relación

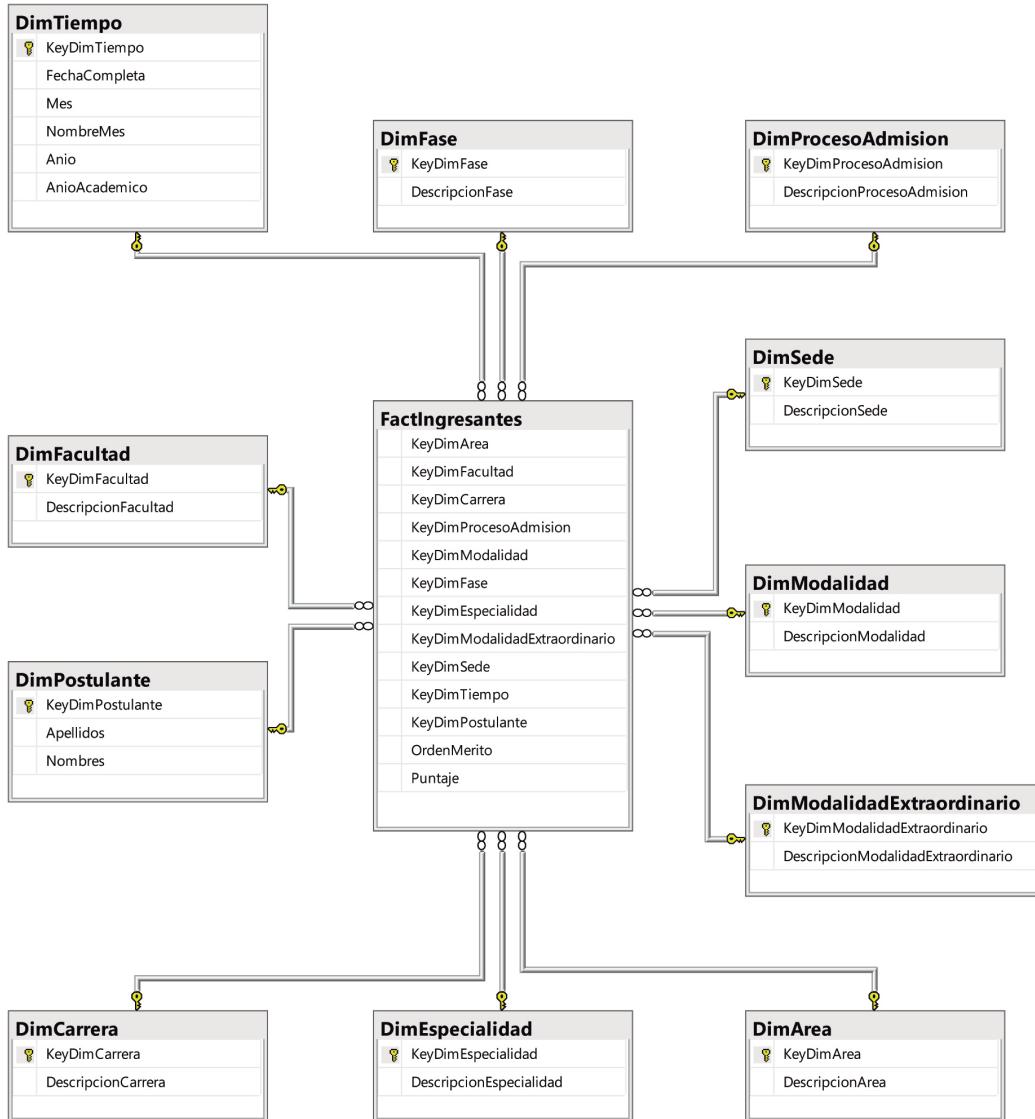


Figura 2.1: Modelo Entidad Relación Data Warehouse

d. Diccionario de datos

Nombre de tabla: DimArea

Descripción: Tabla que contiene la dimensión de las áreas

Campo	Tipo y tamaño	Descripción
KeyDimArea	INT (Autogenerada)	Identificador de la dimensión
DescripcionArea	VARCHAR(50)	Nombre del área

Nombre de tabla: DimFacultad

Descripción: Tabla que contiene la dimensión de las facultades.

Campo	Tipo y tamaño	Descripción
KeyDimFacultad	INT (Autogenerada)	(PK) Identificador de la dimensión
DescripcionFacultad	VARCHAR(100)	Nombre de la facultad

Nombre de tabla: DimCarrera

Descripción: Tabla que contiene la dimensión de las carreras.

Campo	Tipo y tamaño	Descripción
KeyDimCarrera	INT (Autogenerada)	(PK) Identificador de la dimensión
DescripcionCarrera	VARCHAR(100)	Nombre de la facultad

Nombre de tabla: DimModalidad

Descripción: Tabla que contiene la dimensión de las modalidades de admisión.

Campo	Tipo y tamaño	Descripción
KeyDimModalidad	INT (Autogenerado)	(PK) Identificador de la dimensión
DescripcionModalidad	VARCHAR(50)	Nombre del tipo de admisión

Nombre de tabla: DimFase

Descripción: Tabla que contiene la dimensión de las fases.

Campo	Tipo y tamaño	Descripción
KeyDimFase	INT (Autogenerada)	(PK) Identificador de la dimensión
DescripcionFase	VARCHAR(50)	Nombre de la fase

Nombre de tabla: DimProcesoAdmision

Descripción: Tabla que contiene la dimensión de los procesos de admisión.

Campo	Tipo y tamaño	Descripción
KeyDimProcesoAdmision	INT (Autogenerada)	(PK) Id de la dimensión
DescripcionProcesoAdmision	VARCHAR(100)	Nombre del proceso

Nombre de tabla: DimEspecialidad

Descripción: Tabla que contiene la dimensión de las especialidades.

Campo	Tipo y tamaño	Descripción
KeyDimEspecialidad	INT (Autogenerada)	(PK) Id de la dimensión
DescripcionEspecialidad	VARCHAR(100)	Nombre de la especialidad

Nombre de tabla: DimModalidadExtraordinario

Descripción: Tabla que contiene la dimensión de los tipos extraordinarios.

Campo	Tipo	Descripción
KeyDimModalidadExtraordinario	INT (Auto)	(PK) Id de la dimensión
DescripcionModalidadExtraordinario	VARCHAR(50)	Nombre del extraordinario

Nombre de tabla: DimSede

Descripción: Tabla que contiene la dimensión de las sedes.

Campo	Tipo y tamaño	Descripción
KeyDimSede	INT (Autogenerada)	(PK) Identificador de la dimensión
DescripcionSede	VARCHAR(50)	Nombre de la sede

Nombre de tabla: DimPostulante

Descripción: Tabla que contiene la dimensión de los postulantes.

Campo	Tipo y tamaño	Descripción
KeyDimPostulante	INT (Autogenerada)	(PK) Identificador de la dimensión
Nombres	VARCHAR(50)	Nombres del postulante
Apellidos	VARCHAR(50)	Apellidos del postulante

Nombre de tabla: DimTiempo

Descripción: Tabla que contiene la dimensión del tiempo.

Campo	Tipo y tamaño	Descripción
KeyDimTiempo	INT (Autogenerada)	(PK) Identificador de la dimensión
FechaCompleta	DATE	Fecha completa
Mes	INT	Mes calendario
NombreMes	VARCHAR(20)	Nombre del mes
Anio	INT	Anio calendario
AnioAcademico	INT	Anio académico

Nombre de tabla: FactIngresantes

Descripción: Tabla que contiene los hechos de los ingresantes.

Campo	Tipo	Descripción
KeyDimArea	INT	(FK) DimArea
KeyDimFacultad	INT	(FK) DimFacultad
KeyDimCarrera	INT	(FK) DimCarrera
KeyDimProcesoAdmision	INT	(FK) DimProcesoAdmision
KeyDimModalidad	INT	(FK) DimModalidad
KeyDimFase	INT	(FK) DimFase
KeyDimEspecialidad	INT	(FK) DimEspecialidad
KeyDimModalidadExtra*	INT	(FK) DimModalidadExtraordinario
KeyDimSede	INT	(FK) DimSede
KeyDimTiempo	INT	(FK) DimTiempo
KeyDimPostulante	INT	(FK) DimPostulante
OrdenMerito	INT	Puesto de ingreso
Puntaje	DECIMAL(10,6)	Puntaje obtenido

e. Mapeo de datos

i. **Base de dato origen:** Admision_UNSA_Stage

ii. **Base de dato destino:** Admision_UNSA_DW

Nombre de mapeo: Areas

Fuente origen:

dbo.Areas

(Autogenerado)

Nombre

Fuente destino:

dbo.DimArea

KeyDimArea

DescripcionArea

Nombre de mapeo: Facultades

Fuente origen:

dbo.Facultades

(Autogenerado)

Nombre

Fuente destino:

dbo.DimFacultad

KeyDimFacultad

DescripcionFacultad

Nombre de mapeo: Carreras

Fuente origen:

dbo.Carreras

(Autogenerado)

Nombre

Fuente destino:

dbo.DimCarrera

KeyDimCarrera

DescripcionCarrera

Nombre de mapeo: Modalidades

Fuente origen:

dbo.Modalidades

(Autogenerado)

Nombre

Fuente destino:

dbo.DimModalidad

KeyDimModalidad

DescripcionModalidad

Nombre de mapeo: Fases

Fuente origen:

Fuente destino:

dbo.ProcesosAdmision

dbo.DimFase

(Autogenerado)

KeyDimFase

Fase

DescripcionFase

Nombre de mapeo: Procesos de Admisión

Fuente origen:

Fuente destino:

dbo.ProcesosAdmision

dbo.DimProcesoAdmision

(Autogenerado)

KeyDimProcesoAdmision

Modalidad-Anio-Fase (Campo calculado)

DescripcionAdmision

Nombre de mapeo: Especialidades

Fuente origen:

Fuente destino:

dbo.Especialidades

dbo.DimEspecialidad

(Autogenerado)

KeyDimEspecialidad

Nombre

DescripcionEspecialidad

Nombre de mapeo: Modalidad extraordinarios

Fuente origen:

Fuente destino:

dbo.ModalidadesExtraordinarios

dbo.DimModalidadExtraordinario

(Autogenerado)

KeyDimModalidadExtraordinario

Nombre

DescripcionModalidadExtraordinario

Nombre de mapeo: Sedes

Fuente origen:

Fuente destino:

dbo.Sedes

dbo.DimSede

(Autogenerado)

KeyDimSede

Nombre

DescripcionSede

Nombre de mapeo: Postulantes

Fuente origen:	Fuente destino:
dbo.Ingresantes	dbo.DimPostulante
(Autogenerado)	KeyDimPostulante
NombresApellidos (Campo calculado)	Apellidos
NombresApellidos (Campo calculado)	Nombres

Nombre de mapeo: Tiempo

Fuente origen:	Fuente destino:
dbo.ProcesosAdmision	dbo.DimTiempo
(Autogenerado)	KeyDimTiempo
Fecha	FechaCompleta
Fecha (Campo calculado)	Mes
Fecha (Campo calculado)	NombreMes
Fecha (Campo calculado)	Anio
Fecha (Campo calculado)	AnioAcademico

Nombre de mapeo: Ingresantes

Fuente origen:	Fuente destino:
dbo.*	dbo.FactIngresantes
	KeyDimArea
Fecha	KeyDimFacultad
dbo.DimCarrera (PK)	KeyDimCarrera
dbo.DimProcesoAdmision (PK)	KeyDimProcesoAdmision
dbo.DimModalidad (PK)	KeyDimModalidad
dbo.DimFase (PK)	KeyDimFase
dbo.DimEspecialidad (PK)	KeyDimEspecialidad
dbo.DimModalidadExtraordinario (PK)	KeyDimModalidadExtraordinario
dbo.DimSede (PK)	KeyDimSede
dbo.DimTiempo (PK)	KeyDimTiempo
dbo.DimPostulante (PK)	KeyDimPostulante
dbo.Ingresantes OrdenMerito	OrdenMerito
dbo.Ingresantes Puntaje	Puntaje

2.2.2. Diseño de ETLs

A. Definición de unidades de programación Admision_UNSA_Stage

Elemento	Pack_ETL_Admission_UNSA_Stage
Tipo	Project Deployment Model
Descripción	Este paquete contiene todos los paquetes relacionados con la creación de las ETLs para poblar la base de datos Stage.
Objetivo	Realizar la limpieza y transformación de las fuentes de datos.
Paquetes	1. Pack_ETL_Areas.dtsx 2. Pack_ETL_Carreras.dtsx 3. Pack_ETL_Especialidades.dtsx 4. Pack_ETL_EstadisticasCeprunsas.dtsx 5. Pack_ETL_EstadisticasOrdinarios.dtsx 6. Pack_ETL_Facultades.dtsx 7. Pack_ETL_IngresantesCeprunsas.dtsx 8. Pack_ETL_IngresantesExtraordinarios.dtsx 9. Pack_ETL_IngresantesOrdinarios.dtsx 10. Pack_ETL_IngresantesSedes.dtsx 11. Pack_ETL_Modalidades.dtsx 12. Pack_ETL_ModalidadesExtraordinarios.dtsx 13. Pack_ETL_ProcesosAdmision.dtsx 14. Pack_ETL_Sedes.dtsx
Parámetros	Path_Carpeta_Resultados_Ceprunsas Path_Carpeta_Resultados_Extraordinarios Path_Carpeta_Resultados_Ordinarios Path_Carpeta_Resultados_Sedes Path_Excel_Areas_Facultades_Carreras Path_Excel_Procesos_Admission Path_Excel_Resultados_Ceprunsas Path_Excel_Resultados_Extraordinarios Path_Excel_Resultados_Ordinarios Path_Excel_Resultados_Sedes Path_Excel_Sedes_Especialidades_Extraordinarios Path_Log Path_Server_Name
Archivos	D:\Admision_UNSA\Log\Log_Pack_ETL_Admission_UNSA_Stage

Elemento	Pack_ETL_Areas
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Areas
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Con*: P::Con_Excel_Libro_Areas-Facultades-Carreras 1.2. Hoja: Areas 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Filtrar los registros duplicados 6. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 6.1. Conexión: P::Con_BD_Admision_UNSA_Stage 6.2. Tabla: Areas

Elemento	Pack_ETL_Carreras
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Carreras
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Con*: P::Con_Excel_Libro_Areas-Facultades-Carreras 1.2. Hoja: Carreras 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Filtrar los registros duplicados 6. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 6.1. Conexión: P::Con_BD_Admision_UNSA_Stage 6.2. Tabla: Carreras

Elemento	Pack_ETL_Especialidades
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Especialidades
Elemento	Tarea de flujo de datos
Proceso	<p>1. Obtener conexión con la fuente de origen</p> <p>1.1. Conexión: P::Con_Excel_Libro_Sedes-Especialidades-Extraordinarios</p> <p>1.2. Hoja: Especialidades</p> <p>2. Limpiar los registros nulos</p> <p>3. Convertir los campos a los tipos soportados en la BD</p> <p>4. Quitar los espacios antes y después de cada dato</p> <p>5. Filtrar los registros duplicados</p> <p>6. Obtener conexión con la fuente de destino</p> <p>6.1. Conexión: P::Con_BD_Admission_UNSA_Stage</p> <p>6.2. Tabla: Especialidades</p>

Elemento	Pack_ETL_EstadisticasCeprunsa
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Estadísticas
Elemento	Foreach Loop Container
Datos	<p>1. Directory: P::Path_Carpeta_Resultados_Ceprunsa</p> <p>2. Iterador: V::Aux_Iterator</p>
Elemento	Tarea de flujo de datos
Proceso	<p>1. Obtener conexión con la fuente de origen</p> <p>1.1. Conexión: P::Con_Excel_Resultados_Ceprunsa</p> <p>1.2. Hoja: Estadisticas</p> <p>2. Limpiar los registros nulos</p> <p>3. Obtener la ruta del nombre del archivo</p> <p>4. Obtener la modalidad, año y fase de los procesos</p> <p>5. Convertir los campos a los tipos soportados en la BD</p> <p>6. Quitar los espacios antes y después de cada dato</p> <p>7. Filtrar los registros duplicados</p> <p>8. Obtener conexión con la fuente de destino</p> <p>8.1. Conexión: P::Con_BD_Admision_UNSA_Stage</p> <p>8.2. Tabla: Estadisticas</p>

Elemento	Pack_ETL_EstadisticasOrdinarios
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Estadisticas
Elemento	Foreach Loop Container
Datos	<ol style="list-style-type: none"> 1. Directory: P::Path_Carpeta_Resultados_Oordinarios 2. Iterador: V::Aux_Iterator
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Resultados_Oordinarios 1.2. Hoja: Estadisticas 2. Limpiar los registros nulos 3. Obtener la ruta del nombre del archivo 4. Obtener la modalidad, año y fase de los procesos 5. Convertir los campos a los tipos soportados en la BD 6. Quitar los espacios antes y después de cada dato 7. Filtrar los registros duplicados 8. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 8.1. Conexión: P::Con_BD_Admision_UNSA_Stage 8.2. Tabla: Estadisticas

Elemento	Pack_ETL_Facultades
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Facultades
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Libro_Areas-Facultades-Carreras 1.2. Hoja: Facultades 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Filtrar los registros duplicados 6. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 6.1. Conexión: P::Con_BD_Admision_UNSA_Stage 6.2. Tabla: Facultades

Elemento	Pack_ETL_IngresantesCeprunsas
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Ingresantes
Elemento	Foreach Loop Container
Datos	<ol style="list-style-type: none"> 1. Directory: P::Path_Carpeta_Resultados_Ceprunsas 2. Iterador: V::Aux_Iterator
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Resultados_Ceprunsas 1.2. Hoja: Ingresantes 2. Limpiar los registros nulos 3. Obtener la ruta del nombre del archivo 4. Obtener la modalidad, año y fase de los procesos 5. Obtener la sede y modalidad de extraordinario 6. Convertir los campos a los tipos soportados en la BD 7. Quitar los espacios antes y después de cada dato 8. Verificar si no aplica los atributos 9. Filtrar los registros duplicados 10. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 10.1. Conexión: P::Con_BD_Admision_UNSA_Stage 10.2. Tabla: Ingresantes

Elemento	Pack_ETL_IngresantesExtraordinarios
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Ingresantes
Elemento	Foreach Loop Container
Datos	<ol style="list-style-type: none"> 1. Directory: P::Path_Carpeta_Resultados_Extraordinarios 2. Iterador: V::Aux_Iterator
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Resultados_Extraordinarios 1.2. Hoja: Ingresantes 2. Limpiar los registros nulos 3. Obtener la ruta del nombre del archivo 4. Obtener la modalidad, año y fase de los procesos 5. Obtener la sede y modalidad de extraordinario 6. Convertir los campos a los tipos soportados en la BD 7. Quitar los espacios antes y después de cada dato 8. Verificar si no aplica los atributos 9. Filtrar los registros duplicados 10. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 10.1. Conexión: P::Con_BD_Admision_UNSA_Stage 10.2. Tabla: Ingresantes

Elemento	Pack_ETL_IngresantesOrdinarios
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Ingresantes
Elemento	Foreach Loop Container
Datos	<ol style="list-style-type: none"> 1. Directory: P::Path_Carpeta_Resultados_Oordinarios 2. Iterador: V::Aux_Iterator
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Resultados_Oordinarios 1.2. Hoja: Ingresantes 2. Limpiar los registros nulos 3. Obtener la ruta del nombre del archivo 4. Obtener la modalidad, año y fase de los procesos 5. Obtener la sede y modalidad de extraordinario 6. Convertir los campos a los tipos soportados en la BD 7. Quitar los espacios antes y después de cada dato 8. Verificar si no aplica los atributos 9. Filtrar los registros duplicados 10. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 10.1. Conexión: P::Con_BD_Admision_UNSA_Stage 10.2. Tabla: Ingresantes

Elemento	Pack_ETL_IngresantesSedes
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admision_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Ingresantes
Elemento	Foreach Loop Container
Datos	<ol style="list-style-type: none"> 1. Directory: P::Path_Carpeta_Resultados_Sedes 2. Iterador: V::Aux_Iterator
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Resultados_Sedes 1.2. Hoja: Ingresantes 2. Limpiar los registros nulos 3. Obtener la ruta del nombre del archivo 4. Obtener la modalidad, año y fase de los procesos 5. Obtener la sede y modalidad de extraordinario 6. Convertir los campos a los tipos soportados en la BD 7. Quitar los espacios antes y después de cada dato 8. Verificar si no aplica los atributos 9. Filtrar los registros duplicados 10. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 10.1. Conexión: P::Con_BD_Admision_UNSA_Stage 10.2. Tabla: Ingresantes

Elemento	Pack.ETL.Modalidades
Tipo	Paquete de Integration Services
Dependencia	Pack.ETL.Admision.UNSA.Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Modalidades
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con.Excel.Libro.Procesos-Admision 1.2. Hoja: Modalidades 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Filtrar los registros duplicados 6. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 6.1. Conexión: P::Con.BD_Admision.UNSA.Stage 6.2. Tabla: Modalidades

Elemento	Pack.ETL.ModalidadesExtraordinarios
Tipo	Paquete de Integration Services
Dependencia	Pack.ETL.Admision.UNSA.Stage
Descripción	Este paquete contiene el ETL de la carga a tabla ModalidadesExtraordinarios
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con.Excel.Libro.Sedes-Especialidades-Extraordinarios 1.2. Hoja: Facultades 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Filtrar los registros duplicados 6. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 6.1. Conexión: P::Con.BD_Admision.UNSA.Stage 6.2. Tabla: ModalidadesExtraordinarios

Elemento	Pack_ETL_ProcesosAdmision
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admission_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla ProcesosAdmision
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Libro_Procesos-Admision 1.2. Hoja: ProcesosAdmision 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Verificar si no aplica los atributos 6. Filtrar los registros duplicados 7. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 7.1. Conexión: P::Con_BD_Admission_UNSA_Stage 7.2. Tabla: ProcesosAdmision

Elemento	Pack_ETL_Sedes
Tipo	Paquete de Integration Services
Dependencia	Pack_ETL_Admission_UNSA_Stage
Descripción	Este paquete contiene el ETL de la carga a tabla Sedes
Elemento	Tarea de flujo de datos
Proceso	<ol style="list-style-type: none"> 1. Obtener conexión con la fuente de origen <ol style="list-style-type: none"> 1.1. Conexión: P::Con_Excel_Libro_Sedes-Especialidades-Extraordinarios 1.2. Hoja: Facultades 2. Limpiar los registros nulos 3. Convertir los campos a los tipos soportados en la BD 4. Quitar los espacios antes y después de cada dato 5. Filtrar los registros duplicados 6. Obtener conexión con la fuente de destino <ol style="list-style-type: none"> 6.1. Conexión: P::Con_BD_Admission_UNSA_Stage 6.2. Tabla: Sedes

2.2.3. Diseño de cubos

A. Cubo Ingresantes

a. Datos generales:

- i. Data Source: Admision_UNSA_DW
- ii. Data Source View: DataSource_Ingresantes

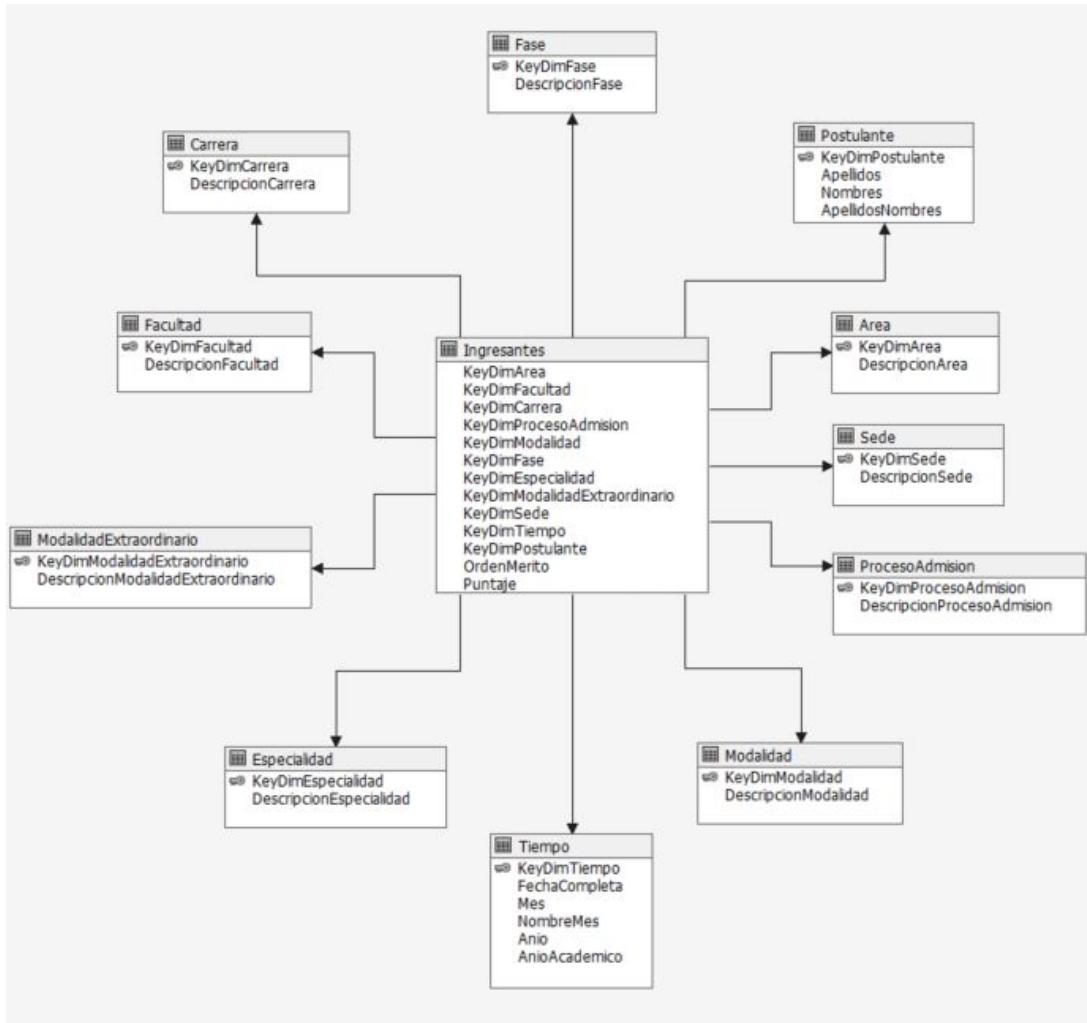


Figura 2.2: Data Source View Cubo Ingresantes

b. Dimensiones:

Nombre de dimensión: Area

Atributo	Usage	KeyColumn	NameColumn
Área	Key	KeyDimArea	DescripcionArea

Nombre de dimensión: Facultad

Atributo	Usage	KeyColumn	NameColumn
Facultad	Key	KeyDimFacultad	DescripcionArea

Nombre de dimensión: Carrera

Atributo	Usage	KeyColumn	NameColumn
Carrera	Key	KeyDimCarrera	DescripcionCarrera

Nombre de dimensión: Modalidad

Atributo	Usage	KeyColumn	NameColumn
Modalidad	Key	KeyDimModalidad	DescripcionModalidad

Nombre de dimensión: Fase

Atributo	Usage	KeyColumn	NameColumn
Fase	Key	KeyDimFase	DescripcionFase

Nombre de dimensión: ProcesoAdmision

Atributo	Usage	KeyColumn	NameColumn
Proceso	Key	KeyDimProcesoAdmision	DescripcionProcesoAdmision

Nombre de dimensión: Especialidad

Atributo	Usage	KeyColumn	NameColumn
Especialidad	Key	KeyDimEspecialidad	DescripcionEspecialidad

Nombre de dimensión: ModalidadExtraordinario

Atributo	Usage	KeyColumn	NameColumn
Modalidad Extraordinario	Key	KeyDimModalidadExt	DescripcionModalidadExt

Nombre de dimensión: Sede

Atributo	Usage	KeyColumn	NameColumn
Sede	Key	KeyDimSede	DescripcionSede

Nombre de dimensión: Postulante

Atributo	Usage	KeyColumn	NameColumn
Apellidos y Nombres	Key	KeyDimPostulante	ApellidosNombres

Nombre de dimensión: Tiempo

Atributo	Usage	KeyColumn	NameColumn
Key Dim Tiempo	Key	KeyDimTiempo	KeyDimTiempo
Año	Regular	Anio	Anio
Mes	Regular	Anio, Mes	NombreMes
Año Académico	Regular	AnioAcademico	AnioAcademico

c. Cubo:

i. Medidas:

Nombre	Campo	Usage	Tipo	Visible
Nro Ingresantes	*	Count	Integer	Visible
Sum Puntaje	Puntaje	Sum	Double	Hidden
Count Puntaje	Puntaje	Count	Integer	Hidden
Mínimo Puntaje	Puntaje	Min	Double	Visible
Máximo Puntaje	Puntaje	Max	Double	Visible
Orden Merito	OrdenMerito	No aggregations	Integer	Visible
Puntaje	Puntaje	No aggregations	Double	Visible

ii. Calculations

Nombre	Parent hierarchy	Expression
Puntaje Promedio	Measures	Sum Puntaje/Count Puntaje

B. Cubo Postulantes

a. Datos generales:

- i. Data Source: Admision_UNSA_DW
- ii. Data Source View: DataSource_Postulantes

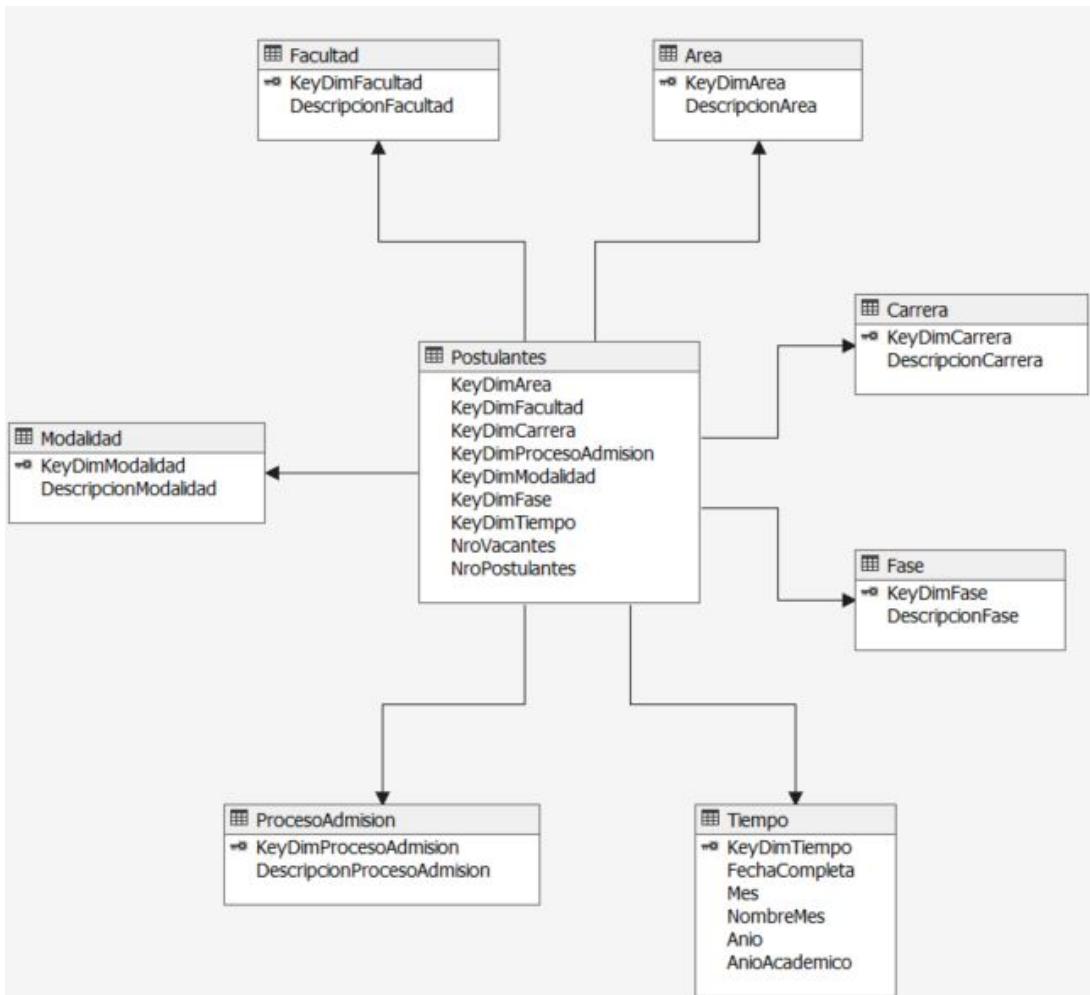


Figura 2.3: Data Source View Cubo Postulantes

b. Dimensiones:

Nombre de dimensión: Área

Atributo	Usage	KeyColumn	NameColumn
Área	Key	KeyDimArea	DescripcionArea

Nombre de dimensión: Facultad

Atributo	Usage	KeyColumn	NameColumn
Facultad	Key	KeyDimFacultad	DescripcionArea

Nombre de dimensión: Carrera

Atributo	Usage	KeyColumn	NameColumn
Carrera	Key	KeyDimCarrera	DescripcionCarrera

Nombre de dimensión: Modalidad

Atributo	Usage	KeyColumn	NameColumn
Modalidad	Key	KeyDimModalidad	DescripcionModalidad

Nombre de dimensión: Fase

Atributo	Usage	KeyColumn	NameColumn
Fase	Key	KeyDimFase	DescripcionFase

Nombre de dimensión: ProcesoAdmision

Atributo	Usage	KeyColumn	NameColumn
Proceso	Key	KeyDimProcesoAdmision	DescripcionProcesoAdmision

Nombre de dimensión: Tiempo

Atributo	Usage	KeyColumn	NameColumn
Key Dim Tiempo	Key	KeyDimTiempo	KeyDimTiempo
Año	Regular	Anio	Anio
Mes	Regular	Anio, Mes	NombreMes
Año Académico	Regular	AnioAcademico	AnioAcademico

c. Cubo:

i. Medidas:

Nombre	Campo	Usage	Tipo	Visible
Nro Vacantes	NroVacantes	Sum	Integer	Visible
Nro Postulantes	NroPostulantes	Sum	Integer	Visible

ii. Calculations

Nombre	Parent hierarchy	Expression
Profit	Measures	Nro Vacantes * 10

iii. KPIs

Nombre	Value	Goal	Status Indicator
KPI_Cupos	Nro Postulantes	Profit	Value/Goal > 1 : 1
			0.85 <= Value/Goal <= 1 : 0
			Otherwise : -1

2.2.4. Diseño de reportes

A. Data Sources:

- a. **Admision_UNSA_DW:** Datawarehouse de la solución.
- b. **Cube_Ingresantes:** Cubo Ingresantes, contiene la información de los ingresantes de proceso de admisión.
- c. **Cube_Postulantes:** Cubo Postulantes, contiene la información de las estadísticas de vacantes y postulante de cada proceso de admisión.

B. Reportes:

Nombre de reporte: Dashboard

Descripción: Contiene información presentada a través de cuadros.

Parámetros	Datasets	Dependencia
Ninguno	Ingresantes_Area	Child: KPI_Cupos

Nombre de reporte: IngresantesPorCarrera

Descripción: Contiene información de los ingresantes agrupados por carrera en un proceso de admisión específico.

Parámetros	Datasets	Dependencia
AnioAcademico	Ingresantes_AreaFacultadArea	C: IngresantesPorCarrera_Detalle
ProcesoAdmision	AniosAcademicos ProcesosAdmisionPorAnio	

Nombre de reporte: IngresantesPorCarrera_Detalle

Descripción: Contiene el detalle de ingresantes por carrera.

Parámetros	Datasets	Dependencia
ProcesoAdmision Carrera	Ingresantes_Postulante	Parent: IngresantesPorCarrera

Nombre de reporte: KPI_Cupos

Descripción: Contiene la relación entre número de vacantes y postulantes.

Parámetros	Datasets	Dependencia
AnioAcademico	KPI_Carreras	Parent: Dashboard
ProcesoAdmision	ProcesoAdmision	
Facultad	AnisoAcademicos ProcesosAdmisionPorAnio Facultades	

Conclusiones

1. La solución de inteligencia de negocios propuesta, procesará la información de los procesos de admisión y apoyará con la toma de decisiones en el área de admisión de la Universidad Nacional de San Agustín.
2. Los cubos procesados a través de Analysis Services, serán consumidos desde aplicaciones externas que facilitará la interacción con los usuarios.
3. Los reportes diseñados permitirán visualizar la información de forma sistematizada para mejorar el análisis de la data y tomar decisiones acertadas.

Recomendaciones

1. Realizar un limpieza exhaustiva de los fuentes de datos, para utilizar la información correcta y procesar mejor la data.
2. Configurar el servicio de Active Directory, en el servidor, para controlar el acceso de los usuarios a la información contenida en la solución de inteligencia de negocios.
3. Redactar un manual de usuario que facilite el aprendizaje de uso de las herramientas que se incorporan en la solución.

Referencias bibliográficas

- [1] L. T. Moss and S. Atre, *Business intelligence roadmap: The complete project lifecycle for decision-support applications.* Boston, U.S.A.: Addison Wesley, 2003.
- [2] R. Rad, *Microsoft SQL Server 2014 Business intelligence development beginner's guide.* Birmingham, U.K.: Packt Publishing, 2014.
- [3] B. Larson, *Delivering business intelligence with Microsoft SQL Server 2016.* McGraw-Hill, 2017.
- [4] C. Howson, *Successful business intelligence: Secrets to making BI a killer app.* McGraw-Hill, 2008.