



Modelado de datos y ETL

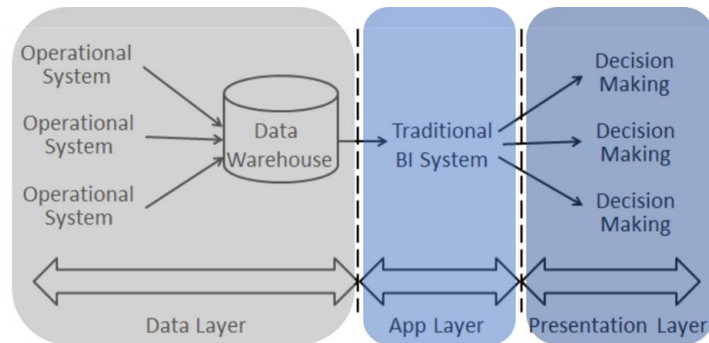
Arquitectura BI

Edinson Fernández

Medellín, octubre

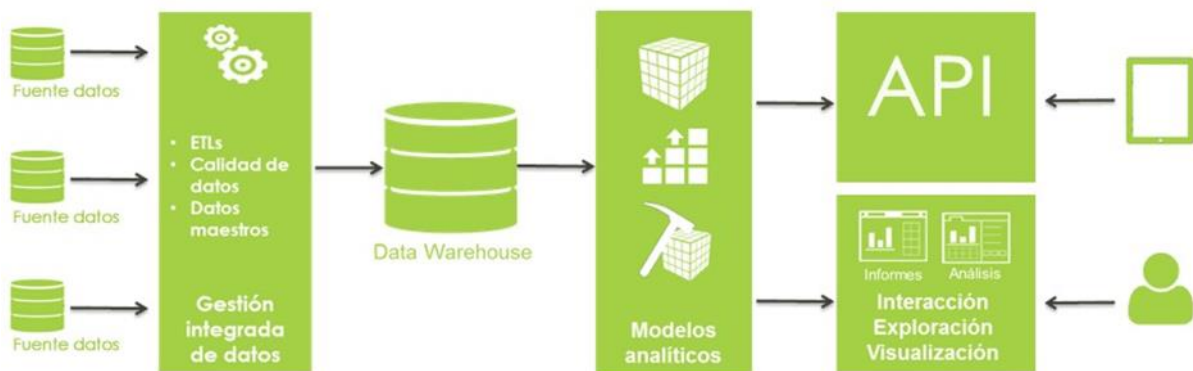
2021

Arquitectura soluciones BI Tradicional



Es frecuente que, con la arquitectura de tres niveles, sea difícil cumplir objetivos y exigencias de los sistemas modernos de nivel de servicio actuales, tales como el tiempo de respuesta máximo y las tasas de rendimiento mínimas. Esto se debe a las dificultades para predecir tiempos de ejecución en los que la capa de aplicación no conoce la gestión del almacenamiento de datos en las capas de bajo nivel.

En general, aunque el sistema de inteligencia de negocio estándar tradicional puede ofrecer una visión futura de negocio, es bien sabido que estos sistemas son lentos para muchos escenarios de datos actuales, rígidos, y el mantenimiento requiere un conocimiento profundo.



Arquitectura de inteligencia de negocio

En un sistema de inteligencia de negocio empresarial, una de las fuentes que siempre estará presente es el

sistema transaccional que soporta y mantiene las operaciones de la compañía. Los datos de transacciones se generan cuando se procesan las transacciones y se almacenan en el servidor de procesamiento de transacciones en línea (OLTP), también denominado fuentes de datos operacionales.

Desde los servidores OLTP, los datos se extraen, transforman y almacenan en un almacén de datos o data warehouse que finalmente es un repositorio de datos estructurado.

Se pueden aplicar diferentes técnicas de optimización de consultas en el almacén de datos para acelerar el

análisis de datos y la consulta de análisis puede ejecutarse en el almacén de datos. La aceleración adicional se puede lograr mediante la creación de submodelos de datos o Data Marts, que son subconjuntos del almacén de datos.

Obviamente, en un sistema de inteligencia de negocio, además de las fuentes de datos tradicionales, es decir,

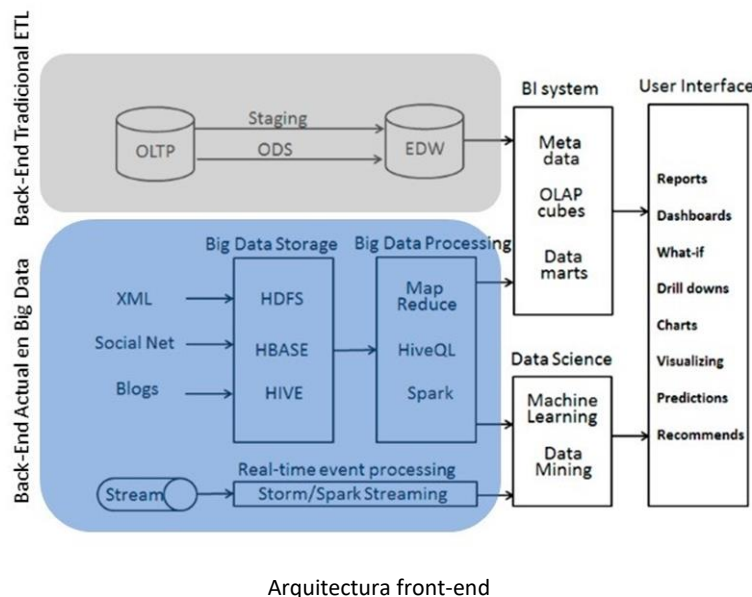
datos de transacción, las fuentes de datos de BI están evolucionando para incluir hasta los mensajes de redes

sociales, intranet, dispositivos móviles, sensores, etc.

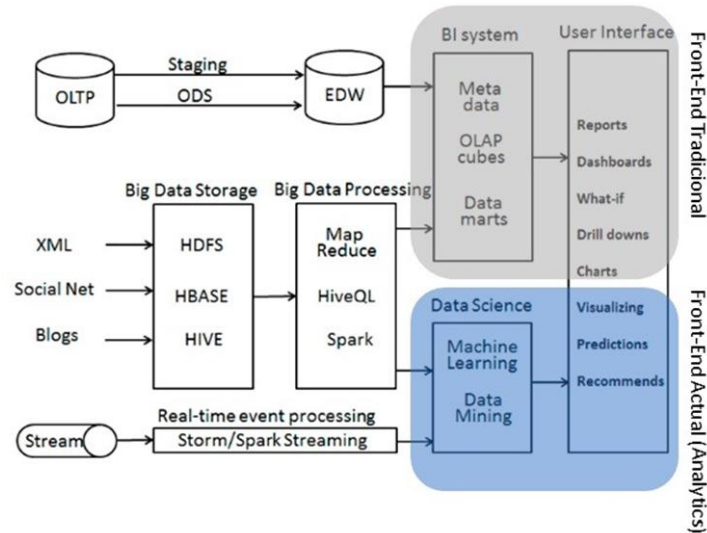
Esta información almacenada en el datawarehouse y/o submodelos debe ser accesible para el usuario analista mediante herramientas diseñadas para este propósito, de una forma ágil, eficiente e intuitiva.

Teniendo en cuenta el flujo de información, la mejor manera de hacer uso de las arquitecturas de inteligencia de negocios empresarias es dividir la arquitectura global en: Arquitectura del backend y arquitectura del frontend.

Back-end: Asociada con la recopilación de los datos y la organización de los mismos (ETL y almacén de datos en los sistemas tradicionales).



Front-end: Aquí es donde los motores de análisis desempeñan un papel fundamental, se prestará atención sobre todo a diferentes tendencias de aprendizaje automático (machine learning) que están permitiendo la evolución del BI, desde la herramienta de análisis histórico tradicional.



Arquitectura back-end

Arquitectura soluciones BI Modernas

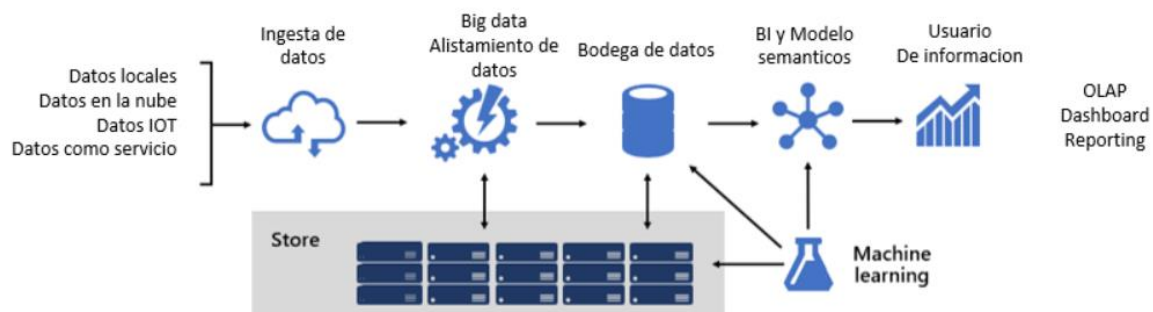
A continuación, se presenta una propuesta de arquitectura de solución para un servicio de analítica descriptiva.

Esta arquitectura tiene diferentes componentes que le permitirán no solamente soportar los requerimientos de negocio sino también los requerimientos de TI, siendo escalable y flexible.

La arquitectura planteada actúa como un pipeline entre los datos de origen transformados y enriquecidos con los consumidores de datos.

El desarrollo de esta arquitectura está orientada a componentes dentro de capas funcionales con responsabilidades clara mente identificadas:

- Orígenes de datos
- Ingesta de datos
- Preparación de datos o macrodatos
- Almacenamiento de datos
- Modelos semánticos de BI
- Informes y tableros de control



Arquitectura BI

La plataforma debe admitir demandas específicas y ahí es donde radica su flexibilidad, no solo para entender requerimientos de información preestablecidos sino requerimientos al vuelo, de este modo atenderá las expectativas de negocio de forma eficiente y eficaz.

Esta arquitectura además permitirá poner en línea nuevos datos extendiendo la capacidad de conexión con otras fuentes, dentro y fuera de la organización.

Marco de trabajo

Se propone un enfoque de tipo sistémico, invirtiendo de esta manera en un marco de trabajo (framework) tanto en procesos técnicos, empresariales y funcionales, permitiendo la reutilización de los diseños y la lógica, la flexibilidad en el escalamiento, en arquitecturas de TI y de negocio reduciendo además la carga de TI.

Con la implementación de un marco de trabajo arquitectónico se optimiza:

- Linaje de datos
- Análisis de impacto
- Mantenibilidad de la lógica de negocio
- Administración de la taxonomía
- Optimización de la gobernanza
- Implementación ágil
- Ambiente colaborativo

Modelo de datos

Los modelos de datos toman el control sobre cómo se estructuran los datos y su acceso. En caso de servicios de negocio y consumidores de datos, los modelos de datos son la interfaz de la plataforma de BI.

Una plataforma de BI ofrece tres tipos de modelos de datos:

Modelo empresarial

Modelo semántico BI

Modelo de Machine Learning (ML)

Modelo empresarial

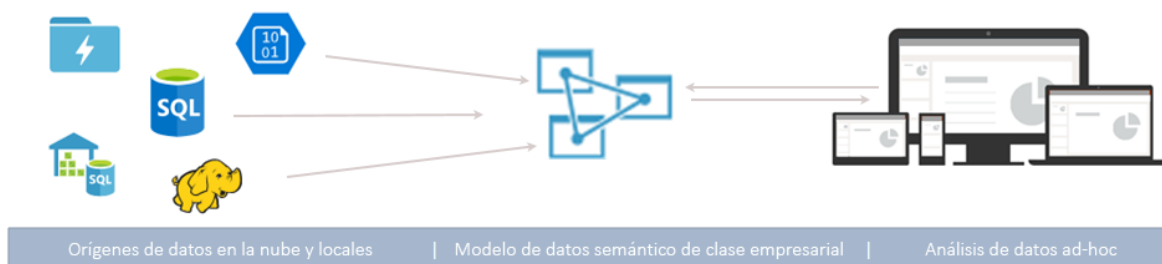
Esto se refiere a modelos dimensionales o *datamart*. Aquí se almacenarán los datos en formato relacional, como tablas dimensionales y de hechos. Estas tablas almacenan datos limpios y enriquecidos, consolidados de muchos sistemas y/o fuentes y representan un origen autoritativo para informes y análisis.

Modelos semánticos de BI

Estos modelos representan una capa semántica de los modelos empresariales. Los desarrolladores BI y los usuarios empresariales los compilan y mantienen. El propósito finalmente es habilitar

modelo ya sean con vista empresarial o vista usuario de menor escala centrándose regularmente en una única área temática con amplio rango de compartición.

La implementación de modelos semánticos puede lograrse en Azure Analysis Services o en capacidades de Power BI Premium.

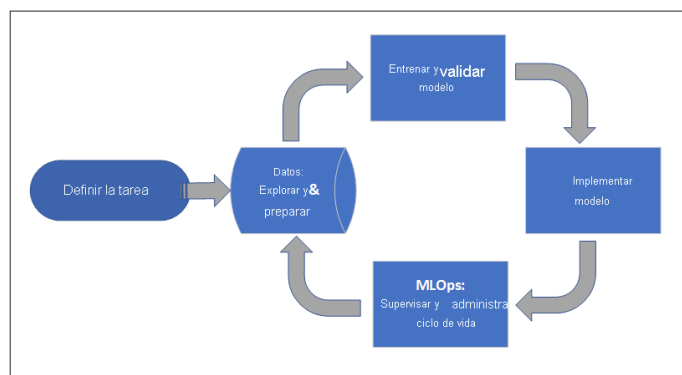


Modelo de Machine Learning

Los modelos de Machine Learning los compilan y mantienen los científicos de datos. Se desarrollan principalmente a partir de orígenes sin procesar en el lago de datos.

Con el entrenamiento de modelos se busca identificar patrones dentro de los datos. En algunas situaciones estos patrones pueden usarse para hacer pronósticos para enriquecer los propios datos.

En una plataforma BI en la nube pueden usarse Azure Machine Learning. A continuación, se describe el ciclo de vida de un proyecto de machine learning:



Almacenamiento de datos

El centro de una plataforma BI se centra en el almacenamiento de los datos, datos que hospedan los modelos empresariales.

Se trata de un origen de datos autorizados, como sistema de registro y como centro de conectividad, que presenta modelos empresariales para informes, BI y ciencia de datos.

Un lago de datos puede almacenar grandes volúmenes de datos sin procesar para su uso posterior junto con transformaciones de datos de almacenamiento provisional.

Los datos refinados se almacenan en una base de datos relacional para ofrecer un almacén de datos de alto rendimiento y altamente escalable para los modelos empresariales, con seguridad,

gobernanza y capacidad de administración. Los datamarts específicos de cada tema se almacenan en Azure Synapse Analytics, que luego se cargan mediante consultas de T-SQL de Polybase o de Azure Databricks.

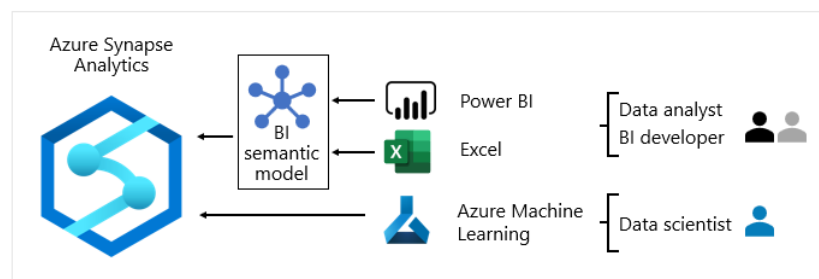
Marco de ingesta

Se propone un marco de ingesta como un conjunto de procedimientos y tablas de configuración. Admite un enfoque controlado por datos para adquirir grandes volúmenes de datos a alta velocidad y con código mínimo. En resumen, este marco simplifica el proceso de adquisición de datos para cargar el almacenamiento de datos.

Consumo de datos

En la capa de informes, los servicios de negocio consumen datos empresariales procedentes del almacenamiento de datos. También tienen acceso a los datos directamente en el lago de datos para realizar tareas de análisis ad hoc o de ciencia de datos.

Los permisos específicos se aplican en todas las capas: en el lago de datos, los modelos empresariales y los modelos semánticos de BI. Los permisos garantizan que los consumidores de datos solo pueden ver los datos para los que tienen derechos de acceso.

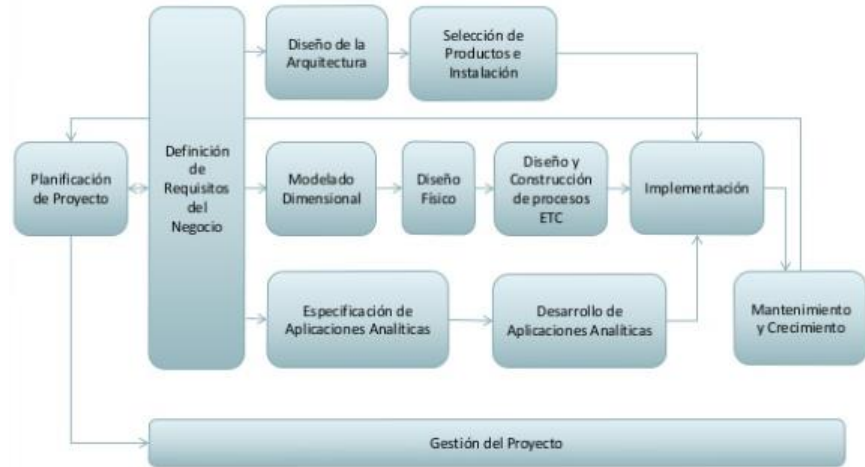


Metodología

La metodología propuesta para el desarrollo del proyecto se basa en lo que Kimball denomina Ciclo de Vida Dimensional del Negocio (Business Dimensional Lifecycle). Este ciclo de vida del proyecto de DW, está basado en cuatro principios básicos (Microsoft, 2019):

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

La construcción de una solución de DW/BI (Datawarehouse/Business Intelligence) es sumamente compleja, y Kimball propone una metodología que ayuda a simplificar esa complejidad en las siguientes fases.



Metodología BI Kimbal

Referencias

Anuario ThinkEPI 2018. (2018). Análisis de tendencias en información y documentación.

Recuperado de [https://ebookcentral-proquest](https://ebookcentral-proquest.com/loginbiblio.poligran.edu.co/lib/bibliopoligransp/detail.action?docID=5486106)

[com.loginbiblio.poligran.edu.co/lib/bibliopoligransp/detail.action?docID=5486106](https://ebookcentral-proquest.com/loginbiblio.poligran.edu.co/lib/bibliopoligransp/detail.action?docID=5486106)

Biasca, R. E. (2007). Los 12 pasos para construir: el tablero de comando. (P. E. Central, Trad.). El

Cid. Recuperado de [https://ebookcentral-proquest-](https://ebookcentral-proquest.com/loginbiblio.poligran.edu.co/lib/bibliopoligransp/detail.action?docID=3173428)

[com.loginbiblio.poligran.edu.co/lib/bibliopoligransp/detail.action?docID=3173428](https://ebookcentral-proquest.com/loginbiblio.poligran.edu.co/lib/bibliopoligransp/detail.action?docID=3173428)

De Carvalho, V. D., Poleto, T. y Seixas, A. P. (2018). Information technology Contracting out Correlation (Statistics) Surveys Business partnerships. Expert Systems, 1-12.

Gorostiza, E. I. y Fontao, A. (2018). Google Analytics: mide y vencerás. ProQuest Ebook Central.

Recuperado de

<http://ebookcentral.proquest.com/lib/bibliopoligransp/detail.action?docID=5756879>

Microsoft. (17 de 06 de 2019). Microsoft. Recuperado de The Microsoft Modern Data Warehouse:

<http://www.microsoft.com/en-us/server-cloud/solutions/data-warehouse-big-data.aspx>

Microsoft (17 de 10 de 2021). BI solution architecture in the Center of Excellence

<https://docs.microsoft.com/en-us/power-bi/guidance/center-of-excellence-business-intelligence-solution-architecture>