





Kremlin Stwart Huaman Santos.
Senior Data Engineer

- Se desempeña como Big Data Software Engineer. Neo4J Certified Professional. Cuenta con certificaciones y especializaciones en Azure, Spark, Python. Cuento con +5 años de experiencia en TI y +3 años de experiencia en Data Engineering.
- Soy un apasionado estratega en tecnologías de la información, con una trayectoria de más de una 7 años dedicada a la excelencia y la innovación en el sector TI.



<https://www.linkedin.com/in/khuamans/>

Reglas de clases

- Mantener el micrófono apagado en caso no vayan a hablar.
- Preguntar en caso que tengan dudas
- Mantenerse atento a la clase.

Modo de evaluación y Producto de la sesión

Evaluación continua

Ejercicios, challenges y/o test.

Producto de la sesión

*Pautas claras para la obtención de la Certificación de Business
Intelligence en Certiprof de manera asincrónica*

Módulo 1

Data Analytics Fundamentals

Sesión 1

**Metodología para
construir un proyecto
en data**

Fundamentos en el análisis de datos

Ingresa al siguiente enlace y participa

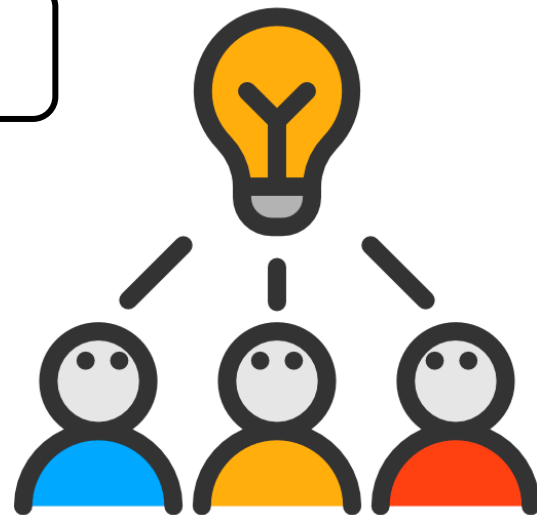


<https://jamboard.google.com/d/1DSTj0bir5oCg0lo0XsOvjSRG11s-vEF0u1GXkFI--Cg/viewer?f=0>

1. ¿Qué crees que hace un Data Engineer en el día a día?

2. ¿Cuál crees que es la diferencia entre un data engineer, un data scientist y un data analyst?

3. ¿Cuales son tus expectativas sobre el programa?



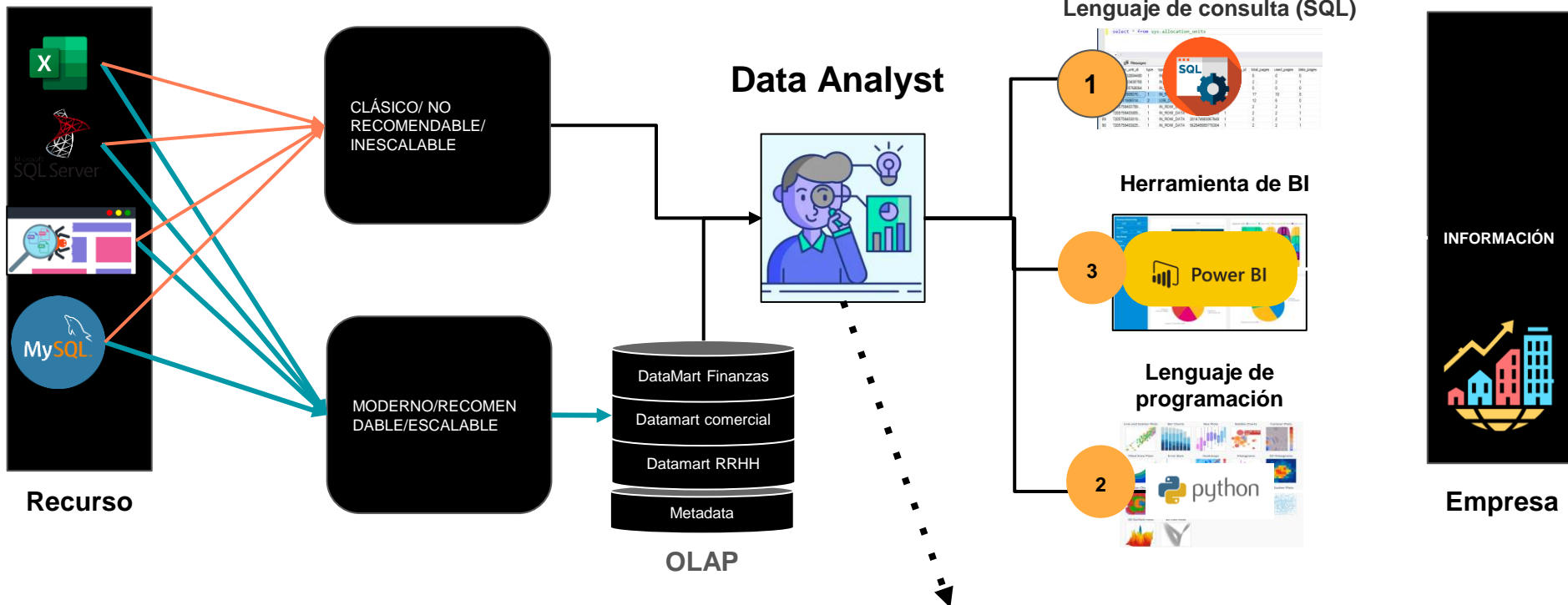
Overview del análisis de datos

Origen de datos

Obtención de datos

Análisis de datos

Información



Ayudo a mi empresa a tomar **decisiones informadas** convirtiendo sus **datos** en **información**



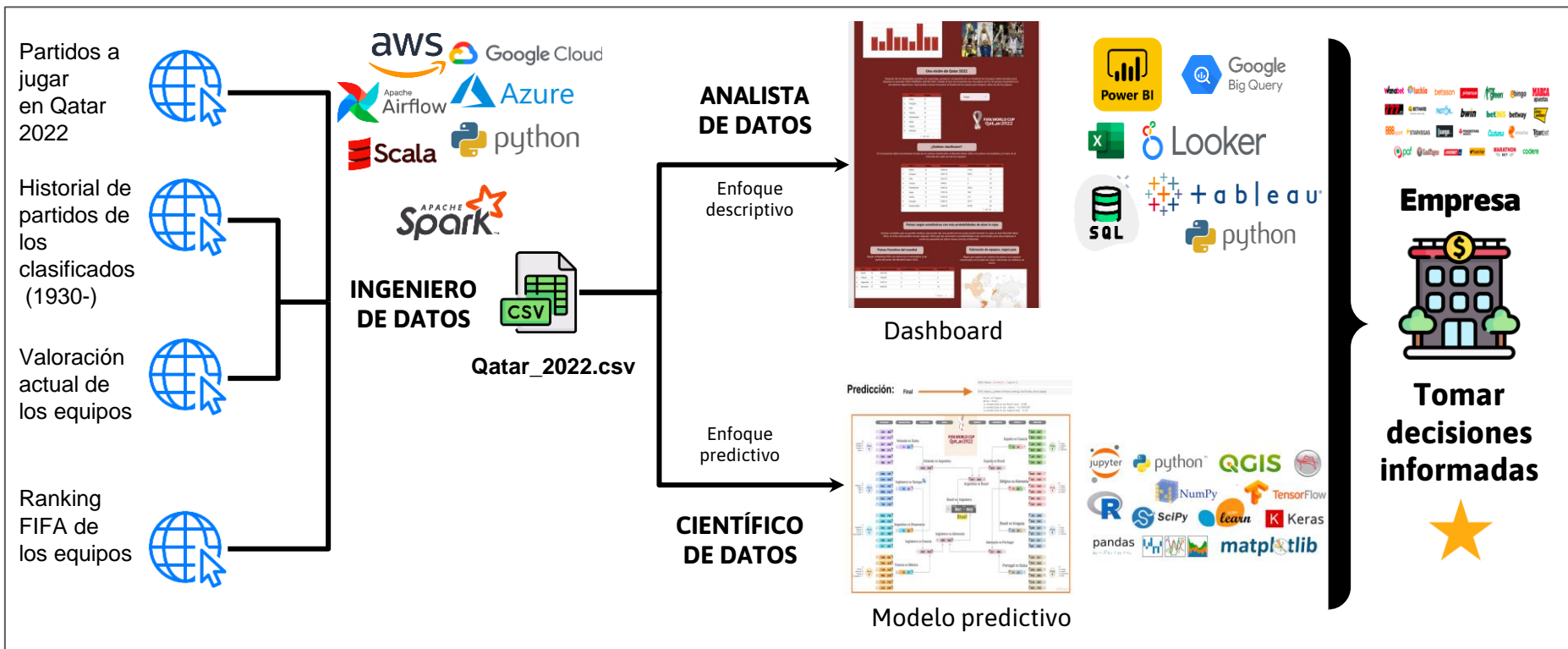
Metodología para construir un proyecto





Roles en un proyecto de análisis de datos

“Análisis y predicción del mundial Qatar 2022”





Roles de un Equipo de Datos



Chief Data Office

Gobierno de Datos

Advanced analytics Data engineering

Product Owner

Comunica las necesidades y contexto del negocio al equipo; y vela por cumplir los objetivos del proyecto.

Project Manager

Planifica y monitorea el proyecto. Vela por el cumplimiento de rituales y puntos de control de calidad. Gestiona riesgo y la relación con los sponsors. Impulsa la cultura de autogestión en el equipo.

Data Translator

Facilita la comunicación entre perfiles técnicos y de negocio. Fomenta un espíritu de experimentación desde la planificación hasta la ejecución del proyecto. Promueve la adopción de una cultura analítica y de procesos.

Data Visualization Specialist

Facilita la generación de insights con el negocio construyendo visualizaciones de datos que facilitan su análisis y entendimiento.

Data Quality Specialist

Participa activamente en las iniciativas de gestión de datos de la empresa y co-diseña soluciones eficientes (con foco en calidad) basadas en tecnologías avanzadas de data y analytics.

Data Scientist

Data Enabler

Diseña el modelado de datos para facilitar su consumo de forma eficiente y escalable asegurando la calidad de la información.

Machine Learning Engineer

Data Engineer

Construye el conjunto de variables transformadas que son la base del desarrollo de los modelos analíticos avanzados.

Data Analyst

Extrae, analiza datos y genera reportes para apoyar en la toma de decisiones.

Data Architect

Encargado del diseño e implementación de la Arquitectura de Datos. Lidera y propone nuevas estrategias del manejo de los datos. Diseña nuevas arquitecturas para cada tipo de caso de uso.

Data Governance Specialist

Participa activamente en las iniciativas de gestión de datos de la empresa y co-diseña soluciones eficientes (con foco en gobierno) basadas en tecnologías avanzadas de data y analytics.

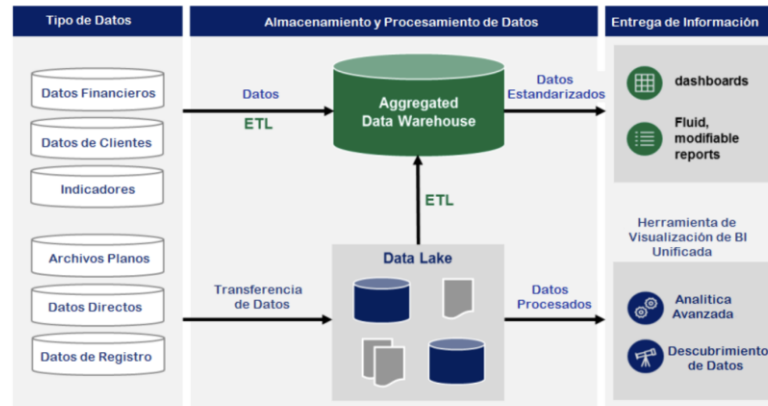
Skill	Data Analyst	Data Scientist	Data Engineer
Negocio	Avanzado (más de 1 proceso)	Avanzado (sobre un proceso)	Medio (sobre un proceso)
Análisis de requerimientos	Avanzado	Avanzado	Medio
Identificación y análisis de fuentes	Avanzado	Avanzado	Medio
Analítico y estadístico	Avanzado	Avanzado	Básico
Comunicación y visualización de datos	Avanzado	Medio	Medio
Limpieza y calidad de datos	Avanzado	Avanzado	Avanzado
SQL	Avanzado	Avanzado	Avanzado
Python	Medio	Avanzado	Avanzado
Matemática aplicada	Medio	Avanzado	Básico
Machine Learning	Básico	Avanzado	Básico
Programación	Básico	Avanzado	Avanzado
Tecnologías de BigData/Cloud	Básico	Medio	Avanzado

Fuente: César Aldana

Business Intelligence Fundamentals

“Business Intelligence es combinación de **tecnología**, **herramientas** y **procesos** que me permiten transformar los **datos** almacenados en **información**, esta información en conocimiento y este conocimiento dirigido a un **plan o una estrategia comercial**” (Certiprof).

“El conjunto de **estrategias**, **procesos**, **aplicaciones**, **datos**, **tecnologías** y **arquitecturas técnicas** que se utilizan para respaldar la **recopilación**, el **análisis de datos**, la **presentación** y la difusión de información comercial” (Wikipedia).



ETL (Extract, Transform, and Load)
Extraer, Transformar, Cargar Datos

Certificación | Tarea 01

Business Intelligence Foundation Professional Certification-BIFPC™



(1) <https://certiprof.com/pages/business-intelligence-foundation-professional-certification-bifpc>

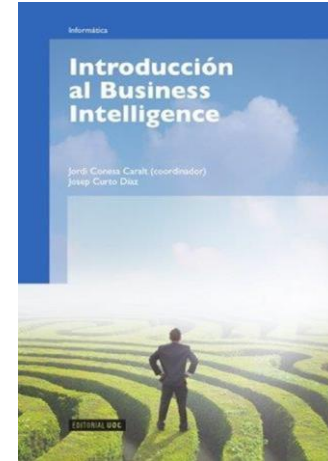
¿Qué es el Business Intelligence?

En 1989 Howard Dresden, analista de Gartner propone una definición formal del concepto:

Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basado en hechos.

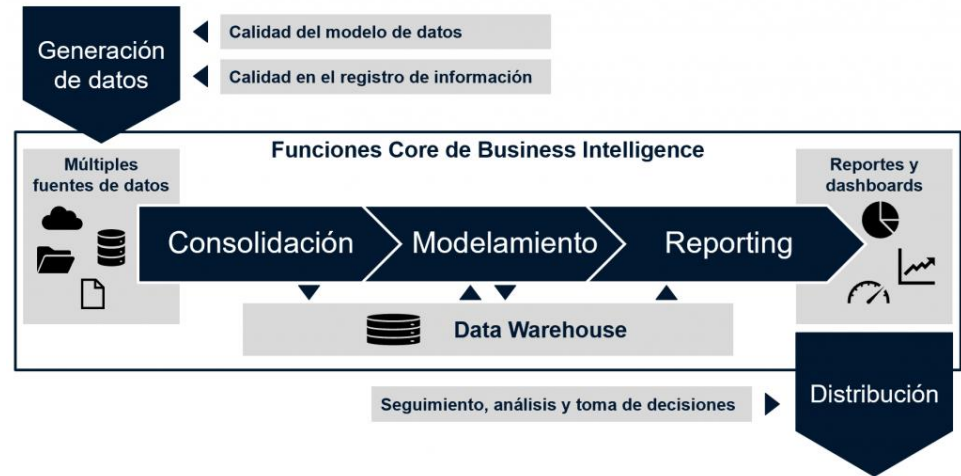


Conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información, que permite tomar mejores decisiones a los usuarios de una organización



Algunas de las tecnologías que forman parte del Business Intelligence son:

- Data warehouse
- Reporting
- Análisis OLAP (On-Line Analytics Processing)
- Análisis Visual
- Reglas de negocio
- Previsiones
- Dashboards
- Integración de datos (incluye ETL)



Un data warehouse es un repositorio de datos que proporciona una visión, global, común e integrada de los datos de la organización, con las siguientes propiedades: estable, coherente, fiable y con información histórica.

Orientado a un tema

Se realizan fotos de los datos basados en fechas

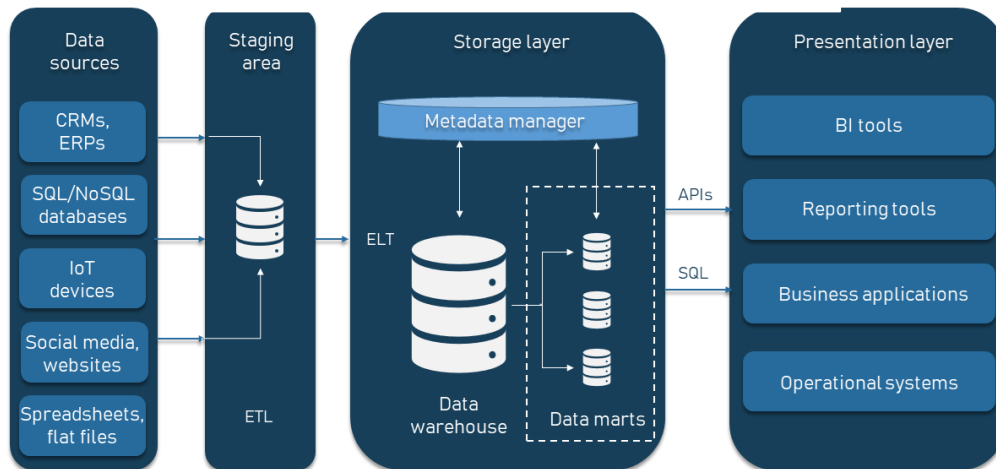
Solo de lectura para los usuarios finales

Organiza una colección de información alrededor de un tema central

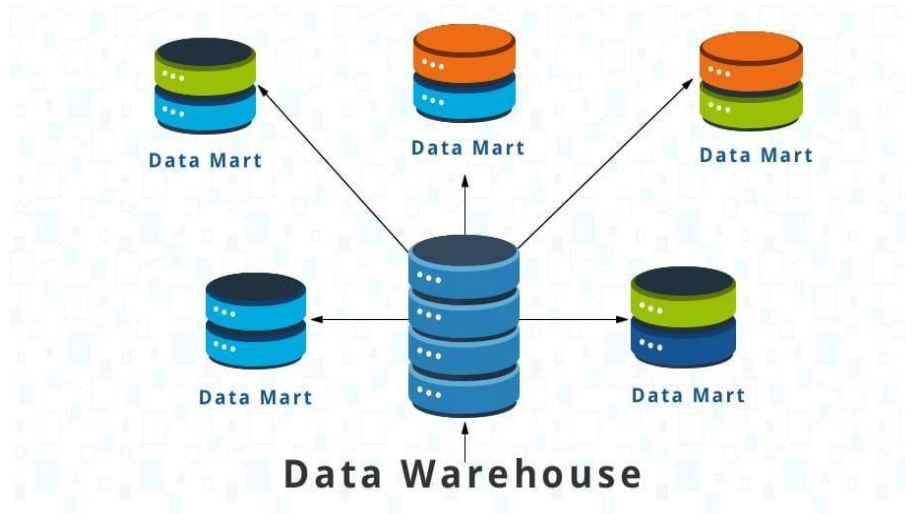
Integrado

Variable en el tiempo

No volátil



Es un subconjunto de los datos del data warehouse cuyo objetivo es responder a un determinado análisis, función o necesidad, con una población específica de usuarios.

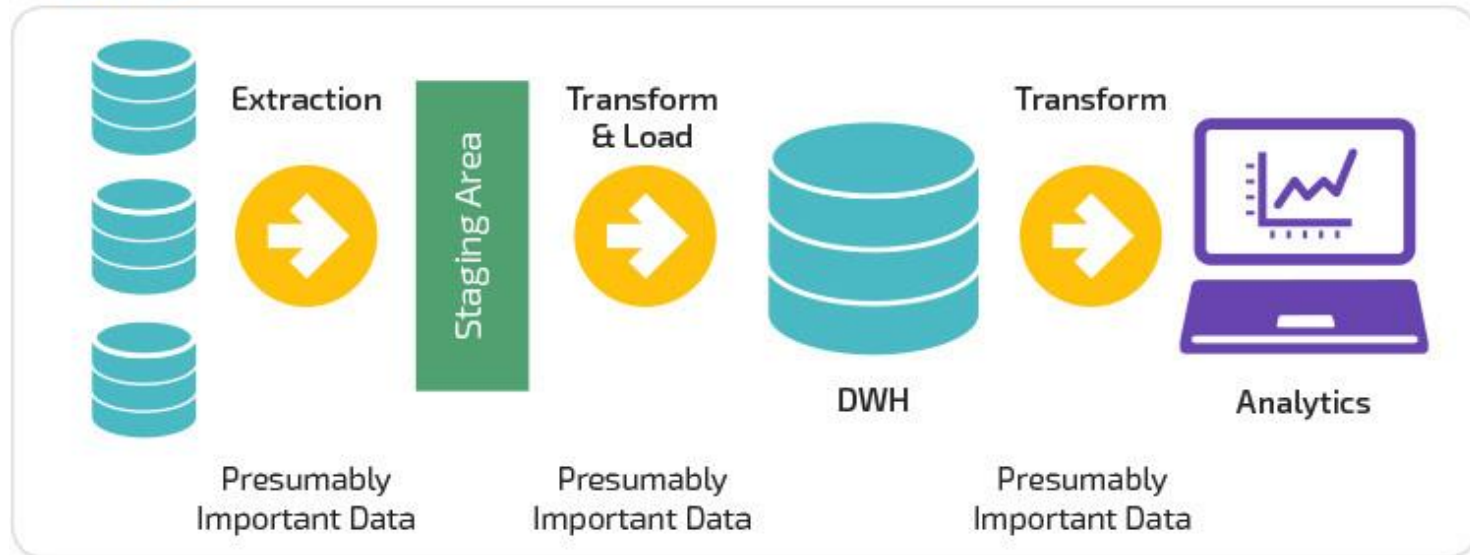


El datamart está pensando para cubrir las necesidades de un grupo de trabajo o de un determinado departamento dentro de la organización.

Característica	Data Warehouse	Data Marts
Alcance	Empresa completa	Departamento o área específica
Tamaño de Datos	Muy grande	Relativamente pequeño
Tiempo de Implementación	Largo	Corto
Costos	Alto	Menor
Estructura de Datos	Compleja (esquema estrella o copo de nieve)	Más simple
Flexibilidad	Menos flexible (por su tamaño y complejidad)	Más flexible y adaptable
Rendimiento	Optimizado para grandes volúmenes de datos y análisis complejos	Rápido para consultas específicas

¿Qué procesos
de ETL
conoces?

Tecnología de integración de datos basada en la consolidación de datos que se usa tradicionalmente para alimentar data warehouse, datamart y staging área.



Herramientas ETL

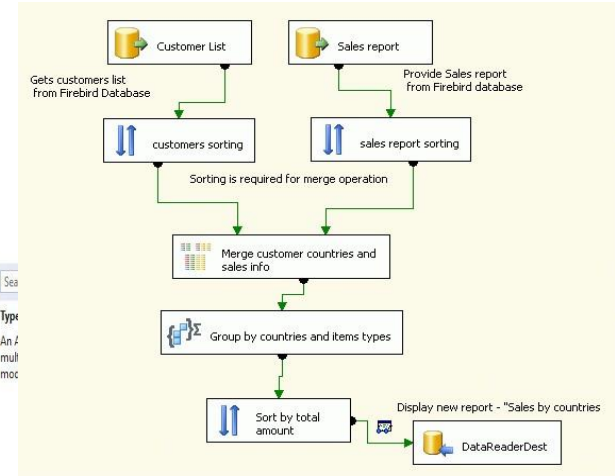
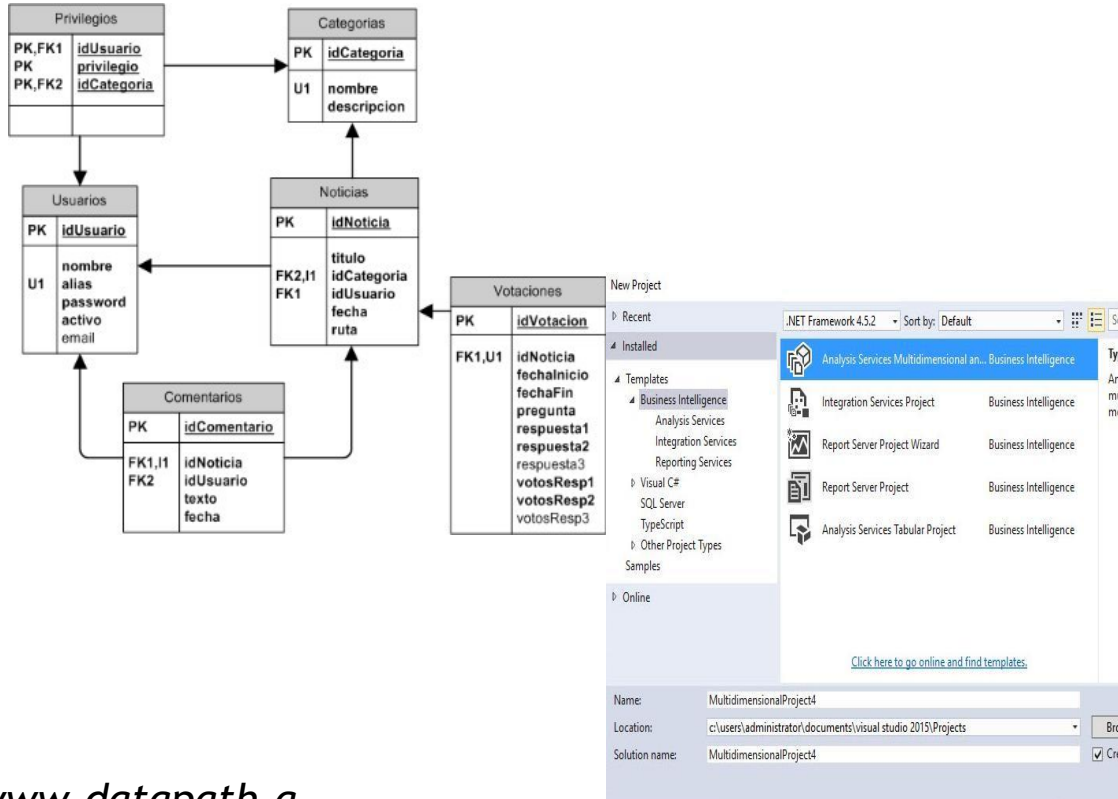
 datapath



www.datapath.a

i

Ambientes de desarrollo



Áreas de negocio



Reporte

Dashboard

“Presentación visual de la **información más importante** necesaria para alcanzar uno o más objetivos los cuales encajan en **una sola pantalla** de computadora de tal manera que pueda ser **monitoreada a una mirada**” (Few, 2013 p34)

- Pequeña, concisa, clara, intuitiva y a menudo interactiva presentación.
- Foco en monitoreo y medición
- Combinación de Charts(Gráficas) y tablas