Las fuentes de datos a considerar son dos.

La primaria, o fuente de RAW DATA, es la herramienta de gestión del negocio que utiliza cada empresa. Puede ser un CRM, ERP, MRP u otro. La información de cada herramienta se volcará, mediante un proceso ETL, en un DATA WAREHOUSE (fuente de datos Core de la solución) que, a su vez, se dividirá en uno o más DATA MARTs de interés específico para cada una de las compañías del Holding.

Inicialmente se consideró la idea de implementar un DATA LAKE como fuente de datos general del DW, pero consideramos que toda la información a explotar está estructurada (del contenido multimedia de las empresas de medios sólo se explotará la METADATA), por lo que no sería necesario para el producto diseñado.

Tanto el proceso ETL, como la alimentación de los diferentes DM, se realizará mediante un proceso BATCH que correrá durante la noche (todos los días). Esto nos permitirá implementar una solución Centralizada Cloud y ahorrar costos en las horas que el DW no está en uso (se estima una inactividad de entre 18 y 20 hs diarias).

Para el Delivery de la información, la solución elegida es MicroStrategy. Gracias a su capa semántica, garantiza que los usuarios y los controles de acceso a los datos se apliquen de manera uniforme en todas las aplicaciones y análisis derivados. De esta manera, se apunta a homogeneizar la explotación de la información y estandarizar las habilidades del personal ejecutivo en cuanto al consumo de los datos. La capa semántica permite realizar la explotación de los datos en el “idioma” del negocio.

A poner en la presentación del 04/07:

* 1 DataWarehouse + 1 o más DataMarts para cada empresa del Holding
* Todo cloud (simplifica la gestión centralizada de los DMs y permite ahorrar costos de uso del DW)
* ETL a utilizar: Pentaho
* Para el DW y los DMs la solución elegida es Snowflake
* La explotación de la información (delivery) se haría con Microstrategy
* Adicionalmente, integraremos una solución de IA desarrollada en Python que permita la generación de Insights adicionales (**Prescriptivos**)

Ver Hadoop, Spark (labura en memoria; no tiene sentido aplicarlo ya que nuestros indicadores se actualizan durante la noche), Hive y **Snowflake**

Nota: El rol de **Arquitecto de Datos** es Clave en el diseño de solución (modelo de datos) de un Enterprise DataWarehouse

Metodologia

Produccion de contenidos y supermercados: perfilamiento de usuarios a través de metodos de clustering.

Seguros: modelos de scoring de usuarios. Clustering o regresion.

Fabricacion: datos de capacidad utilizada, paradas de planta, tiempos de procesos.