**Semana 11**

Video introducción: Explicación breve de lo que se verá en la semana. Comentar que el proceso de ETL es importante tanto para lo que vemos en el curso como para técnicas de analítica avanzada como Machine Learning ya que tener datos erróneos o “sucios” puede llevarnos a tomar la decisión incorrecta

Video 1: “Importancia del ETL en BI y en analítica en general”

Destacar la importancia de la limpieza de los datos ya que datos “sucios” pueden llevar a conclusiones erróneas y por lo tanto a tomar decisiones de forma equivocada. Así mismo pasa con el tratamiento de outliers, ya que un promedio se puede disparar si existe un dato anómalo que lo aumente o disminuya de forma considerable

Video 2: “ETL – Extracción”

Ejemplos de diferentes fuentes para la obtención de datos, como por ejemplo encuestas donde puede quedar mal escrito algo o faltar algún dato, comentar la existencia de web scrapping para obtención de datos (páginas de ofertas utilizan esta técnica como ejemplo). En el caso del curso se utilizarán consultas SQL para obtener los datos desde una base OLTP.

Video 3: “ETL - Transform”

Ejemplos de limpieza de datos o de estandarización como los que se mencionaron previamente. Dentro de la explicación incluir que se puede realizar la selección de datos relevante eliminando duplicados, columnas o datos anómalo, o también realizando perfilado de datos para llenar algunos valores perdidos o corregir datos inconsistentes. Mencionar el uso de la tabla resumen de atributos.

Video 4: “ETL - Load”

Ejemplo de carga de datos ya limpios, este proceso es posterior a la trasformación por lo que los datos deberían estar limpios en el Data Warehouse. En el caso de la carga inicial, se mueven grandes cantidades de volúmenes, por ejemplo, pensar en un retail que está iniciando el camino a BI. Por otro lado, los refresh, que están determinados por la granularidad, son volúmenes de datos mucho más pequeños, ya que pueden ser diarios o semanales, por ejemplo.

Video 5: “GROUP BY para construcción de OLAP a partir de OLTP”

Explicar cómo funciona GROUP BY para obtener sumas, promedios o cantidades. Luego mostrar cómo se aplicaría para crear los atributos de la Tabla de Hechos del modelo ROLAP explicando que el GROUP BY agruparía más de un valor.

Material de apoyo para el desarrollo de videos extraído desde las diapositivas del profesor Alex Arenas

