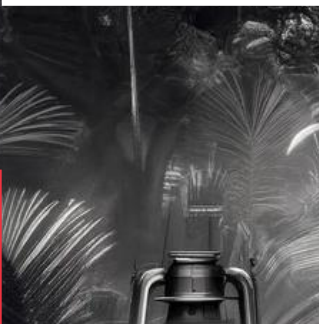
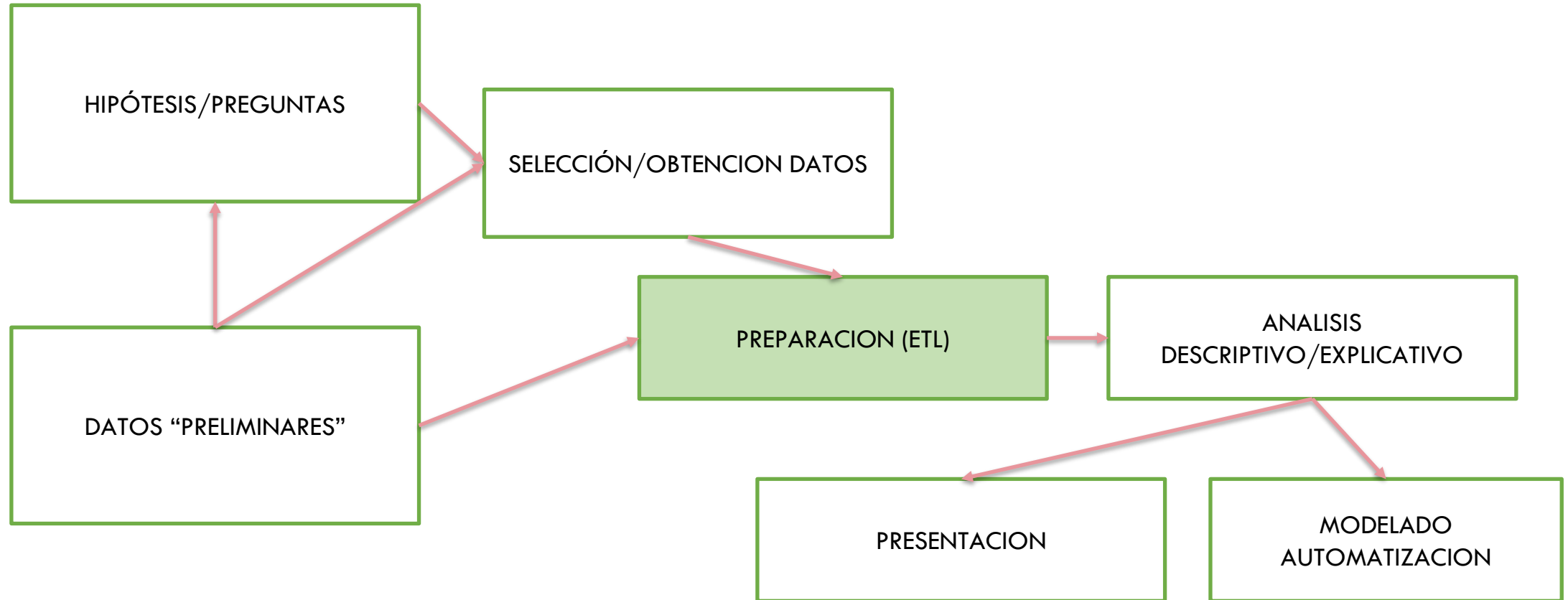




Análisis de Datos Proceso ETL



La importancia de empezar bien...



PROCESANDO DATOS (ETL)

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos
ETL?

- (E)xtract
- (T)ransform
- (L)oad



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

Hacer llegar los datos desde sus orígenes hasta un lugar propietario dentro de la Organización, llamado área de *staging*



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

Hacer llegar los datos desde sus orígenes hasta un lugar propietario dentro de la Organización, llamado área de *staging*

Puede ser un conjunto de ficheros o un gestor de bases de datos



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

Hacer llegar los datos desde sus orígenes hasta un lugar propietario dentro de la Organización, llamado área de *staging*

Puede ser un conjunto de ficheros o un gestor de bases de datos

Los datos llegan en bruto, crudo o *raw*, no hay procesamiento intermedio (eso es la T)



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad



Application.exe



Document.docx



Song.mp3



Spreadsheet.xlsx

ARCHIVOS

Se almacenan según su:

- Funcionalidad: Ejecutables o Archivos de Datos
- El formato del archivo y por tanto la forma de acceso depende del tipo de dato que almacene:
 - Texto
 - Hojas de cálculo
 - Bases de datos
 - Vídeo
 - Audio
 - Imágenes
 - ...



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad



Application.exe



Document.docx



Song.mp3



Spreadsheet.xlsx

ARCHIVOS

Se almacenan según su:

- Funcionalidad: Ejecutables o Archivos de Datos
- El formato del archivo y por tanto la forma de acceso depende del tipo de dato que almacene:
 - Texto plano
 - Texto enriquecido (pdf, word, etc)
 - Hojas de cálculo
 - Bases de datos
 - Vídeo
 - Audio
 - Imágenes
 - ...



PROCESANDO DATOS (ETL) EXTRACT

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

ARCHIVOS

Texto plano:

- Ficheros con información: txt, log, bat, sh, cfg, csv, xml, json...
- Ficheros con código de programación: java, r, py, php, jsp...



PROCESANDO DATOS (ETL) TRANSFORM

¿Qué son los procesos
ETL?

- (E)xtract
- (T)ransform
- (L)oad



PROCESANDO DATOS (ETL) TRANSFORM

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

- Unir fuentes de datos
- Analizar y tratar duplicados
- Limpiar datos y unificar formatos
- Generar nuevas columnas/datos a partir de los existentes
- Simplificar datos existentes
- Tratar los datos faltantes



PROCESANDO DATOS (ETL) LOAD

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

- Sistema de Archivos local o en la nube
- Bases de datos



PROCESANDO DATOS (ETL) LOAD

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

BASES DE DATOS



- Almacenes: Datos de forma organizada
- Gestor de Bases de Datos (SGBD)
- Gestores "famosos": Oracle, MySQL, PostgreSQL, SQL Server
- Para acceder: Lenguajes especializados (SQL, el más conocido)
- Tipos: Relacionales, Multidimensionales, NoSQL



PROCESANDO DATOS (ETL) LOAD

¿Qué son los procesos ETL?

- (E)xtract
- (T)ransform
- (L)oad

NO ES UNA BASE DE DATOS

- Excel no es una base de datos

Reino Unido olvidó registrar casi 16.000 positivos porque su Excel no admitía más filas



