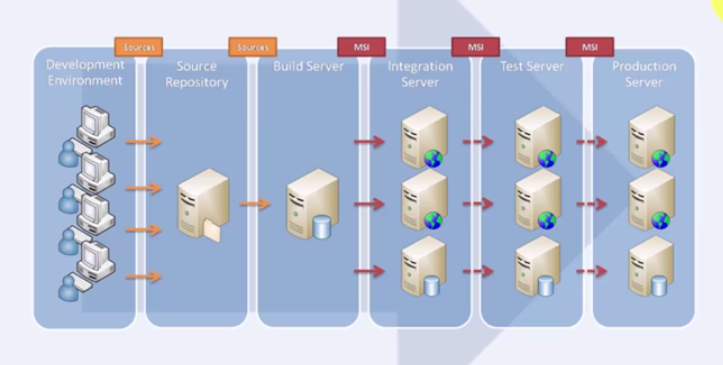
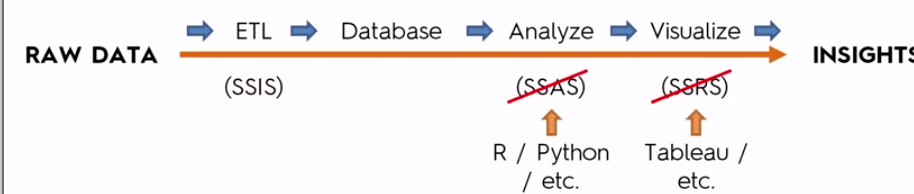
* Data must undergo journey from raw data to insights which can consist of 4 major steps: **ETL 🡪 Database 🡪 Analyze 🡪 Visualize**
* 70% of time is in the **prep phase** (ETL + database) and 30% is in the **creative phase** (Analyze + Visualize)

*Data Warehousing + ETL*

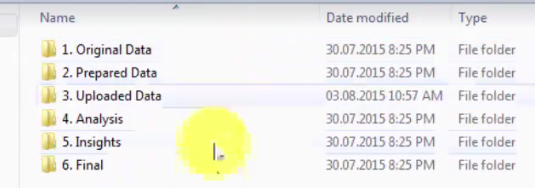
* **DW** = storing the data in a non-transactional RDB, but storing data in a way that it can be analyzed + visualized to derive insights
* Advantages: integrity, combination of datasets, easily scalable, processing power
* Common IT infrastructure:



* Data is stored in multiple sources, and analyzing it in the source could cause multiple issues, such as overwriting the data or severely impact a critical business process or crash a business system
* So we put it into a DW via an ETL process to remove anomalies, transform the data, and basically clean it up all into standard, agreed-upon formats



* File Structure Template



* Original Data 🡪 from source systems (don't modify any data in this folder)
* Prepared Data 🡪 Any modifications done to original/raw data, including cleaning
* Uploaded Data 🡪 a temporary stop for the data w/ subfolders for upload dates
* Analysis 🡪 Analystics code/scripts/investigaions, etc.
* Insights 🡪 Visualizations, comments, documents, etc.
* Final 🡪 Final reports + presentation