Lesson 3

Design Layers in Data Processing Architecture

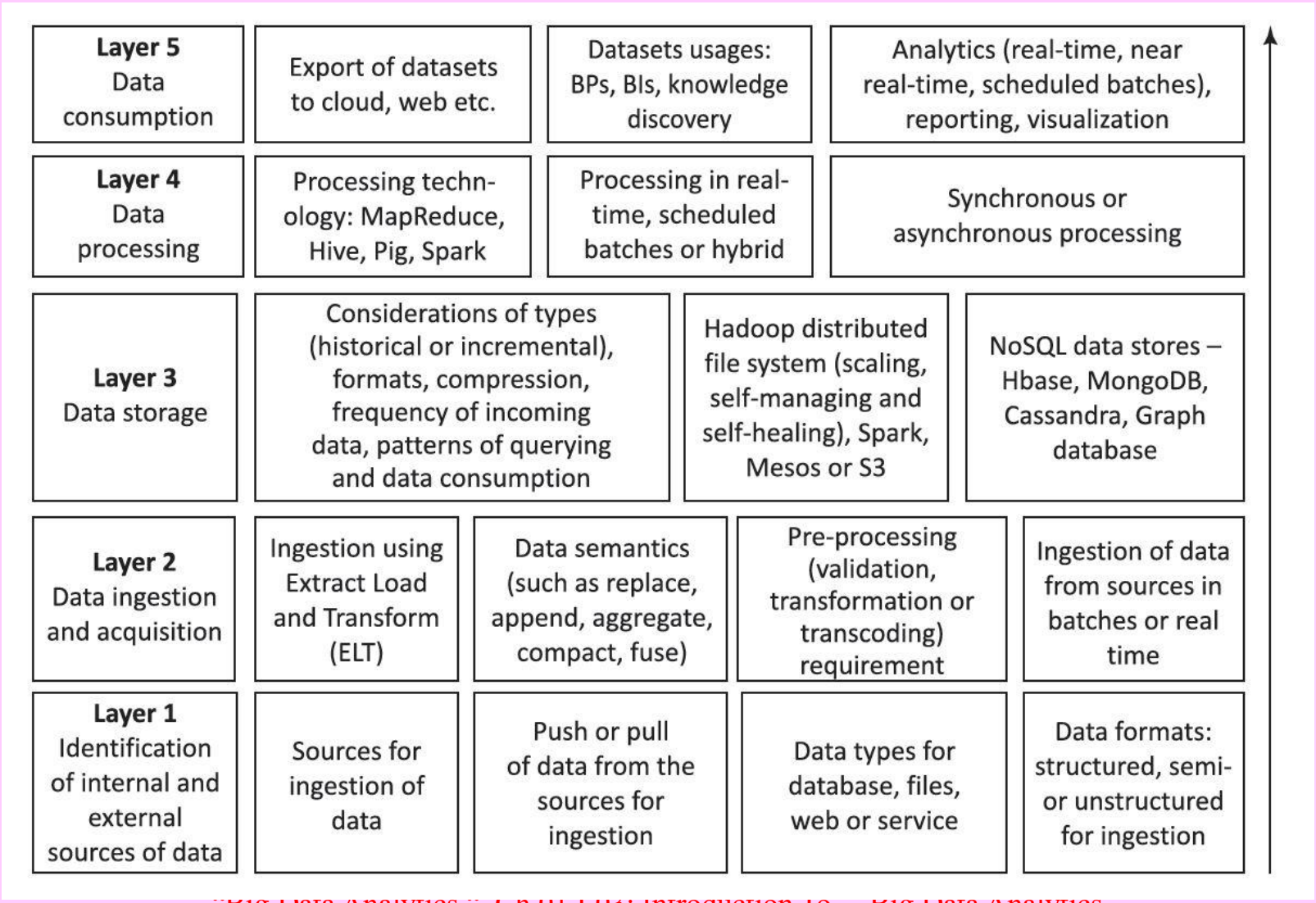
Big Data Architecture

- "Big Data architecture is the logical and/or physical layout/structure of how Big Data will be stored, accessed and managed within a Big Data or IT environment" Techopedia

Big Data Architecture

- Logically defines how Big Data solution will work, the core components (hardware, database, software, storage) used, flow of information, security and more

Figure 1.2 Design of logical layers in a data processing architecture



Lowest Layer L1

- Considers amount of data needed at ingestion layer 2 (L2) and either Push from L1 or pull by L2 as per the mechanism for the usages
- Source data-types: Database, files, web or service
- Source formats, i.e., semi-structured, unstructured or structured.

Data Ingestion and Acquisition Layer L2

- Considers Ingestion and ETL processes either in real time, which means store and use the data as generated, or in batches
- Batch processing is using discrete datasets at scheduled or periodic intervals of time.

Data Storage Layer L3

- Data storage type (historical or incremental), format, compression, incoming data frequency, querying patterns and consumption requirements for L4 or L5
- Data storage using Hadoop distributed file system or NoSQL data stores—HBase, Cassandra, MongoDB

Data Processing Layer L4

- Data processing software such as MapReduce, Hive, Pig, Spark, Spark Mahout, Spark Streaming
- Processing in scheduled batches or real time or hybrid
- Processing as per synchronous or asynchronous processing requirements at L5.

Data Consumption Layer L5

- Data integration
- Datasets usages for reporting and visualization, Analytics (real time, near real time, scheduled batches), BPs, BIs, knowledge discovery
- Export of datasets to cloud, web or other systems

Summary

We learnt

- Five Design Layers
- L1: Identification of Internal and External Sources of Data for ingestion and acquisition
- L2 Ingestion and Acquisition Layer

… Summary

We learnt:

- L3 Data Storage in Required formats for processing at L4
- L4 Data Processing Layer
- L5 data consumption (usage) layer

End of Lesson 3 on

Design Layers in Data Processing Architecture