Lesson 3 Design Layers in Data Processing Architecture

2019

"Big Data Analytics ", Ch.01 L03: Introduction To ... Big Data Analytics Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

1

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

Layer 5 Data consumption	Export of datasets to cloud, web etc.		Datasets usages: BPs, BIs, knowledge discovery			Analytics (real-time, near real-time, scheduled batches), reporting, visualization		
Layer 4 Data processing	Processing techn- ology: MapReduce, Hive, Pig, Spark		Processing in real- time, scheduled batches or hybrid		uled	Synchronous or asynchronous processing		
Layer 3 Data storage	Considerations of types (historical or incremental), formats, compression, frequency of incoming data, patterns of querying and data consumption			file se	Hadoop distributed file system (scaling, self-managing and self-healing), Spark, Mesos or S3		NoSQL data stores – Hbase, MongoDB, Cassandra, Graph database	
Layer 2 Data ingestion and acquisition	Ingestion using Extract Load and Transform (ELT)	Data semantics (such as replace, append, aggregate, compact, fuse)			(v trans tra	-processing validation, formation or anscoding) quirement		Ingestion of data from sources in batches or real time
Layer 1 Identification of internal and external sources of data	Sources for ingestion of data	of	Push or pull data from t sources for ingestion	he	data	Data types for database, files, web or service		Data formats: structured, semi- or unstructured for ingestion

2019

Big Data Analytics ¹⁷, Cn.01 L03: Introduction 10 ... Big Data Analytics Raj Kamal and Preeti Saxena, [©] McGraw-Hill Higher Edu. India

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

"Big Data Analytics ", Ch.01 L03: Introduction To ... Big Data Analytics Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

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.

"Big Data Analytics ", Ch.01 L03: Introduction To ... Big Data Analytics Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

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

2019

"Big Data Analytics ", Ch.01 L03: Introduction To ... Big Data Analytics Raj Kamal and Preeti Saxena, © McGraw-Hill Higher Edu. India

Summary

We learnt

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



We learnt:

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

2019

End of Lesson 3 on Design Layers in Data Processing Architecture