**Explain the big data architecture with neat diagram.**

Big Data Architecture –

Big data architecture must fill all the foundational requirements such as -

* Capturing data from different sources.(Collecting data)
* Cleaning and integrating data of different formats.
* Sorting and organizing data.
* Analyzing data.
* Identifiying relationships and patterns.
* Deriving conclusions based on the data analysis results.

The big data architecture can be divided into the number of layer as follow –

* Data sources layer.(1) – relational database, NoSQL database (Text,video,audio,image)
* Ingestion layer.(2) – identify, transfromation, validation, testing… etc
* Storage layer.(3) – Hadoop storage layer – big data(NoSQL Databse) – HDFS. 2 – Data Warehouse is use to store the data.
* Physical infrastructure layer.(4) – Bare metal clustered worksstations, virtuallized cloud services.
* Platform management layer.(5) – plg, sqoop, hive, mapreduce, zoopkeeper.
* Security layer.(6) – provide security.
* Monitoring layer.(7) – monitor the task.
* Analytics layer.(8) – Statical Analytics, Test Analytics, search engine, real time engine.
* Visualization layer.(9) – Hadoop Adminstration, Data Analyst IDE / SDK, visuallization.
* Big data application.(10) – raak.

**Data Source Layer -**

* To absorb and integrate the data coming form various sources at varying velocity and in different formats.
* Data has to be cleaned before use.
* Differebtiate noise and infromation.

**Ingestion Layer -**

* To absorb the huge inflow of data and sort it out in different catogeries.
* Validates, cleanses, transforms, reduces and integrates the structure data for further processing.
* Stages in the ingestion layer are
* Identification
* Filtration
* Validation
* Noise reduction
* Transfromation
* Compression
* Integration

**Storage Layer –**

* HDFS used to store huge volumes of data.
* Fault tolerant and parallelization.
* Stores data in the from of blocks.
* Follows write once read many model.
* All types of data storage requirements can be addressed by NoSQL.

**Physical Layer –**

* Takes care of the H/W and network requirements.
* Provide either a cloud environment or distributed grid of commodity.

**Platform Management Layer –**

* To provide tools and query language for accessing NoSQL databases.

Uses

Security Layer –

* Handles basic security principles.

Security Check –

* It must authenticate nodes by using protocols
* It must enable file layer encrptions
* It must suscribe a key management services for trusted keys and certificates
* Use tools validating

Monitoring Layer –

* Consist of a number of monitoring systems
* Aware of all the configurations and functions of different O/S and Hardware.
* Facility of machine communication
* Provides tools for data storage and visualization

Analytics Engine –

Visualization Layer –

* Handles the task of intresting and visualizing big data
* Done by data analyst and scientist
* Describe as viewing a piece of information form different perspectives
* Interpreting it in differnet

Big Data Application –

Applications can be either horizontal or vertical

Horizontal Application

* used