INTRODUCTION TO HADOOP:

Hadoop is an Apache open source framework written in java that allows

distributed processing of large datasets across clusters of computers using

simple programming models.

Hadoop Architecture

At its core, Hadoop has two major layers namely:

(a) Processing/Computation layer (MapReduce), and

(b) Storage layer (Hadoop Distributed File System).

OPEN SOURCE TECHNOLOGIES:

Open source software is computer software that is available in source code form under an open-source license that permits users to study, change and improve and at times also to distribute the software.

Top emerging technologies that are helping users cope with and handle Big Data in a cost-effective manner-

1. **Column-oriented databases**

Traditional, row-oriented databases are excellent for online transaction processing with high update speeds, but they fall short on query performance as the data volumes grow and as data becomes more unstructured. Column-oriented databases store data with a focus on columns, instead of rows, allowing for huge data compression and very fast query times. The downside to these databases is that they will generally only allow batch updates, having a much slower update time than traditional models.

2. **MapReduce**

This is a programming paradigm that allows for massive job execution scalability against thousands of servers or clusters of servers.

Any MapReduce implementation consists of two tasks:

- The "Map" task, where an input dataset is converted into a different set of key/value pairs, or tuples;
- The "Reduce" task, where several of the outputs of the "Map" task are combined to form a reduced set of tuples (hence the name).

3. **Hadoop**

Hadoop is by far the most popular implementation of MapReduce, being an entirely open source platform for handling Big Data.

4. **Hive**

Hive is a "SQL-like" bridge that allows conventional BI applications to run queries against a Hadoop cluster. It was developed originally by Facebook, but has been made open source for some time now, and it's a higher-level abstraction of the Hadoop framework that allows anyone to make queries against data stored in a Hadoop cluster just as if they were manipulating a conventional data store. It amplifies the reach of Hadoop, making it more familiar for BI users.

5. **PIG**

PIG is another bridge that tries to bring Hadoop closer to the realities of developers and business users, similar to Hive. Unlike Hive, however, PIG consists of a "Perl-like" language that allows for query execution over data stored on a Hadoop cluster, instead of a "SQL-like" language. PIG was developed by Yahoo!, and, just like Hive, has also been made fully open source.

6. **PLATFORA**

Perhaps the greatest limitation of Hadoop is that it is a very low-level implementation of MapReduce, requiring extensive developer knowledge to operate. Between preparing, testing and running jobs, a full cycle can take hours, eliminating the interactivity that users enjoyed with conventional databases. PLATFORA is a platform that turns user's queries into Hadoop jobs automatically, thus creating an abstraction layer that anyone can exploit to simplify and organize datasets stored in Hadoop.

7. **SkyTree**

SkyTree is a high-performance machine learning and data analytics platform focused specifically on handling Big Data.

CLOUD AND BIG DATA:

Big Data and cloud computing go hand-in-hand. Cloud computing enables companies of all sizes to get more value from their data than ever before, by enabling blazing(shining intensely) fast analytics at a fraction of previous costs. This, in turn drives companies to acquire and store even more data, creating more need for processing power and driving a virtuous(morally excellent) circle.

MOBILE BUSINESS INTELLIGENCE:

Mobile business intelligence is software that extends desktop business intelligence (BI) applications so they can be used on a mobile device.

MBI applications optimizes traditional BI reports so they can be viewed easily on a small screen and is ideal for displaying key performance indicators (KPIs) and alerts on small screens with simple charts, graphs and sparklines.

An additional benefit of MBI is that it allows data that's captured by the mobile device to be integrated on-the-fly so that reports are currents and mobile workers can make informed decisions in real time.

CROWD SOURCING ANALYTICS:

Crowdsourcing is the process of getting work or funding, usually online, from a crowd of people. The word is a combination of the words 'crowd' and 'outsourcing'. The idea is to take work and outsource it to a crowd of workers.