**Traditional Database vs Hadoop**

1. **Type of data**
   * **Traditional database** is good for **Structured Data:** data that resides within the fixed confines of record or file. Traditional database requires user to understand the data well enough to model it first (e.g. manage data with its relationship). The data resides in tables having relational structure defined by a schema.
   * **Hadoop** is better used with **Unstructured Data/Semi-structured Data:** data that comes from various sources e.g. emails text documents. Hadoop has an ability to join, aggregate and analyze vast stores of multi-source data without having to structure it first. Hadoop does not require any modeling as it uses key/value pairs as its basic data unit (flexible).
2. **The scalable analytics infrastructure**
   * **Traditional database** is suitable for companies whose **data workloads are constant and predictable.**
   * **Hadoop** better serves companies that have **increasing data demands** (growing/fluctuating workloads)
3. **Cost-effective**
   * **Traditional database** is designed more friendly to scaling up which can be expensive.
   * **Hadoop** is designed to be scale-out architecture operating on a cluster of commodity PC machines.

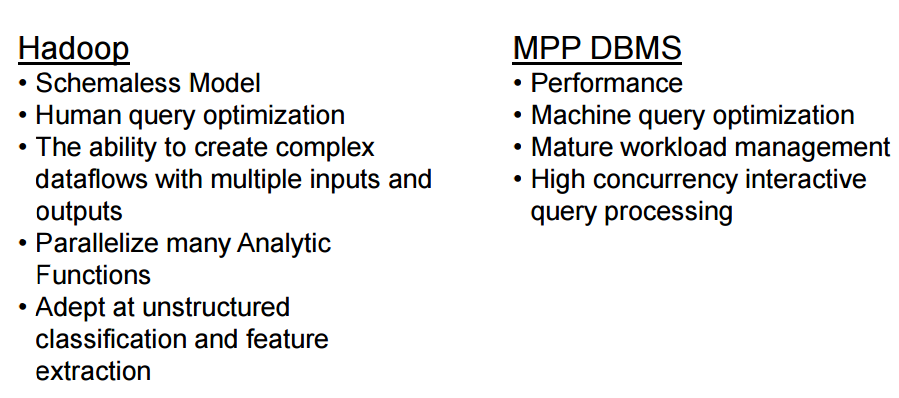
**Note:** Hadoop can be cost-effective by combining open source software with commodity servers. Cloud-based Hadoop platforms reduce costs by eliminating the expense of physical servers and warehouse space. Hadoop offers massive scale in processing power and storage at a very low comparable cost to the traditional database.

1. **Objective of the task**
   * **Traditional database** is better fit with organizations relying on **time-sensitive data analysis.** Traditional database is well equipped to analyze smaller data sets in real or near-real time.
   * **Hadoop** was designed for large distributed data processing that addresses every file in the database, which takes time. Hadoop is ideal for tasks where **a slower time-to-insight is acceptable** such as running end-of-day reports to review daily transactions, scanning historical data.

**Note:** Hive SQL is always a magnitude of times slower than SQL run in traditional databases.

1. **Properties of the system**
   * **Traditional database** has **ACID (Atomicity, Consistency, Isolation, Durability) properties** that guarantee that database transaction are processed reliably.
   * **Hadoop** does not have any of these properties.
2. **Downtime**
   * **Traditional database: a downtime is needed** for any maintenance on storage or data files or adding processing power such as more CPU, physical memory in non-virtualized environment.
   * **Hadoop** systems are individual nodes that can be added in as needed ensuring high-availability.

# Comparing the Strengths between Hadoop and massively parallel processing (MPP) database

****

**References:**

<http://www.qubole.com/blog/big-data/hadoop-vs-traditional/>

<http://www.quora.com/What-is-the-difference-between-a-Hadoop-database-and-a-traditional-Relational-Database>

<https://en.wikipedia.org/wiki/ACID>

<http://www.snia.org/sites/default/files2/ABDS2012/MainStage/GregBattas_Hadoop_Relational_Database.pdf>

https://sites.google.com/site/hadoopintroduction/home/comparing-sql-databases-and-hadoop

http://www.wikidifference.com/difference-between-hadoop-and-rdbms/