# Types of Big Data Technologies (+ Management Tools)

## 1. Data storage

- Apache Hadoop
- Apache Spark
- Apache Hive
- Apache Flume
- ElasticSearch
- MongoDB

## 2. Data mining

- Rapidminer
- Presto

## 3. Data analytics

- Apache Spark
- Splunk
- KNIME

## 4. Data visualization

- Tableau
- Power BI

## 1. Data storage

### Apache Hadoop

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

### MongoDB

MongoDB is a source-available cross-platform document-oriented database program.

Classified as a NoSQL database program, MongoDB uses JSON-like documents with optional schemas

## 2. Data mining

### Rapidminer

pros:

1. Multiple deployment options based on our preference.
2. Strong visualization.
3. Accurate Preprocessing.
4. Multiple interfaces.
5. Java API available that can be used in programs.

cons:

1. It takes too much memory and so slows down your system.
2. Less forums for support.

### Presto

A single Presto query can process data from multiple sources like HDFS, MySQL, Cassandra, Hive and many more data sources. Presto is built in Java and easy to integrate with other data infrastructure components. Presto is powerful, and leading companies like Airbnb, DropBox, Groupon, Netflix are adopting it.

## Apache Spark

Apache Spark is an open-source unified analytics engine for large-scale data processing. Spark provides an interface for programming clusters with implicit data parallelism and fault tolerance

## Splunk

Splunk is a program that enables the search and analysis of computer data. It analyzes semi-structured data and logs generated by various processes with proper data modeling as per the need of the IT companies. The user produces the data by means of any device like- web apps, sensors, or computers. It has built-in functionality for defining data types, field separators, and search process optimization. For the searched result, it also provides visualization of data.

| | |
|---|---|
| Tableau can handle huge columns of data and still offer better performance. | Power BI is best for a limited volume of data. |
| Tableau has better data visualization. | Power BI offers many data points for data visualization. |
| Tableau works best with huge data. | Power BI is suboptimal with huge data. |
| Experts and experienced users use Tableau. | Power BI is used by beginners and experienced alike. |

1. Airflow

2. Delta Lake

3. Drill

4. Druid

5. Flink

6. Hadoop

7. Hive

8. HPCC Systems

9. Hudi

10. Iceberg

11. Kafka

12. Kylin

13. Pinot

14. Presto

15. Samza

16. Spark

17. Storm

18. Trino

MongoDB

# 1. Data storage

- Apache Hadoop
- Hortonworks
- Data lake
- Apache Spark
- MongoDB
- Apache Cassandra
- P r e s t o
- Cloudera
- Cloud storage
- Elastic search
- Hybrid storage

# Cloud Service Providers

Microsoft Azure

Google Cloud Platform

Amazon Web Service (AWS)

Oracle Cloud

IBM Cloud Services

Rackspace

Adobe Creative Cloud

Red Hat

SAP

Kamatera

Salesforce

Verizon Cloud

VMware