

Big Data Management What, Why and How



Contents

What is Big Data?



Big data refers to large and diverse sets of data that have high volume, high velocity and originate from a variety of sources, and are present in different data formats. This increasing volume can extend to petabytes of data and can't be handled by traditional databases of the organisation.

Why the need for big data management ?



With enterprises growing in size continually, the amount of data has grown exponentially. Now managing this data, big data, is becoming increasingly challenging. The problems thrown up with this massive growth of data are:

- Inability to access desired data in real time
- Loss of tracking of outdated of data
- Increase in the number of servers, backup, and storage infrastructure quadrupling the costs to the company
- Inability to drive in-depth insights and inferences from data worth decision making for the company
- Inability to manage the security of data sources

These are only a few problems associated with big data



With coupling of IoT of devices, and various upcoming hardware and software infrastructure, the challenge is even more pertinent. The end result is the loss of trust of customers and users in data systems of the organisation, a decrease in productivity, and slowed down workflows and processes. It has also been found to demean the brand value over a period of time leading to an unmanageable elephant often increasing risks of non-compliance with regulatory policies and lack of transparency.

As enterprise data stores have continued to grow exponentially, managing that big data has become increasingly challenging. Organizations often find that the data they have is outdated, that it conflicts with other data in their systems, or that it is just plain inaccurate.

The trust factor dwindling with exponential growth of data was aptly highlighted in the 2017 Global Data Management Benchmark Report by Experian data quality. In the report, they stated, "While most organizations around the globe say that data supports their business objectives, less than half of organizations globally (44%) trust their data to make important business decisions. The report also added, "Nearly one in two organizations globally (52%) say that a lack of confidence in data contributes to an increased threat of non-compliance and regulatory penalties, and consequently, a downturn in customer loyalty (51%)."

So, the need for the management of big data acquired the utmost urgency. It started the research for a set of data policies, technologies, and tools that could help organizations manage big data with ease.

Big Data Management Challenges

For managing big data, the major challenges to the handling of big data were to be mapped out. After researching out various problems, organisations were going through in handling of big data, the major challenges converged out to be:

- **Siloed data**

Organisations use multiple applications in their work processes. The information and the data are generally stored in different databases. The same information sometimes may be stored in different databases and even the same piece of information can be stored in different locations conveniently eating up the storage space of the company. Now, if there is even a little bit of inconsistency of data stored in different databases, it may lead to multiple problems like conflicting data, loss of customer trust, inaccurate analytics leading to erroneous decision making, and multiple others.

- **Massive increase in data infrastructure**

When a company has millions of customers to cater to leading to petabytes of data, the company has to invest massively on data storage, access, retrieval and analysis, may be involving creation of multiple new data centres or data comms, more servers etc. The TCO increases massively often reducing the profit margins of the organisations radically.

- **Complex Data Architecture**

With petabytes of structure and unstructured data percolating from different data sources and present in different data formats like videos, gif, text, codes, sound files, image, presentations, cataloguing of this data becomes a nightmare. Your data management personnel may be having nightmares and sweating day in and day out simplifying the complexity of data arising every day.

- **Poor Data Quality**

The more the sources of data, the more the applications, the more the siloed databases, synchronisation between them becomes a hellish task. The chances of human errors are quadrupled and at the end inconsistent, poor data quality leads to faulty decision making occurs.

How to Manage Big Data?

To manage big data, it is important to define what big data management actually refers to. Big data management is the process of organising, governing, and administering the massive amount of data percolating from different processes, workflows, channels, platforms, IoT devices, and others. These terabytes of structured and unstructured data, logs, ETL, originating from different sources and present in different formats are synchronised, categorised, stored, updated, different access routes identified, refined to weed out redundant data to make data of high quality that can be readily used for business intelligence purposes for Agencies, corporations, and other organisations.

Now, how to manage big data?

There are simple sequential steps to do it

(a) Identify the big data management goals

Organisations can have various goals for implementing big data management projects. This can be:

- Managing the torrential amount of data being produced in a better manner
- Faster processing of data
- Have better synchronisation between IT processes, users
- Better communications with management
- More effective and accurate business decision making
- Lowering the cost of IT infrastructure and processes handling data of the company And many more .

(b) Prioritize Use cases and formulate big data roadmap

The roadmap would involve doing a gap analysis of the current state of data architecture and the desired state, outline the use cases to be executed and the capabilities to be developed through implementation of big data management.

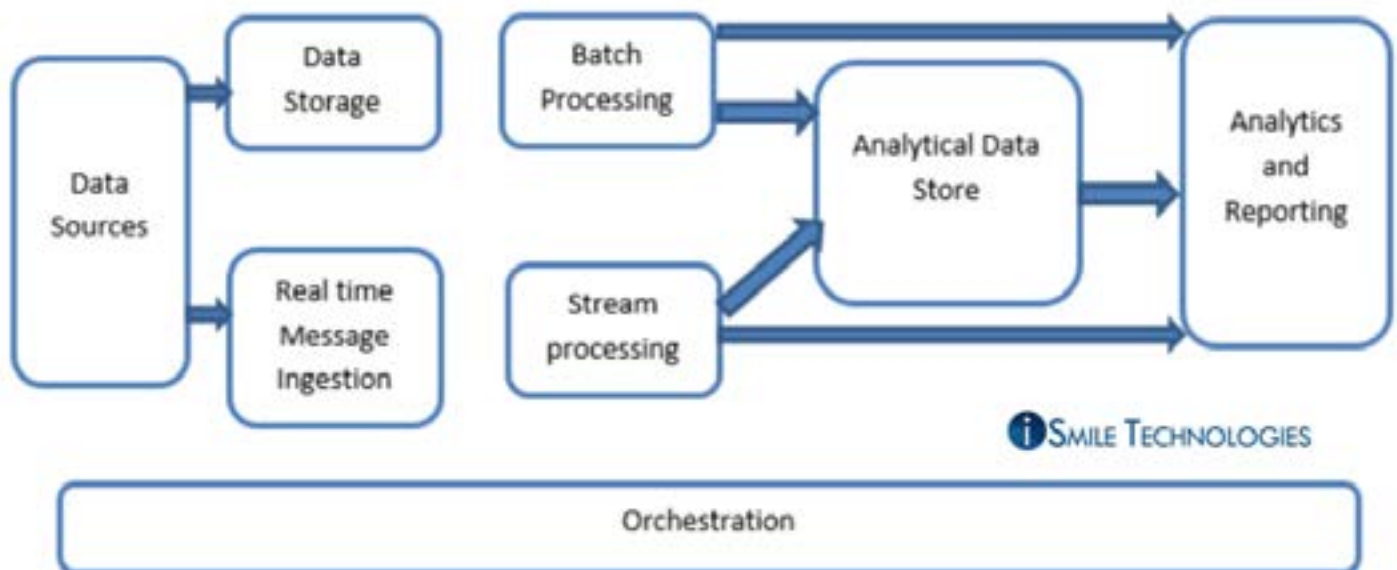
(c) Frame the architecture for big data

Framing big data architecture includes designing the IT processes, managing the IT infrastructure, creating a process framework, etc for handling the intake, processing, and analytics of big data workloads

Components of Big Data Architecture

The components which are universally present in any big data architecture involves:

- Data Sources including application stores, static files, data from IoT devices, etc
- Data Storage like distributed files storage having huge volumes of files in different formats, often called a data lake.
- Data Processing: Big data sets batch filtering, classifying, aggregating, and preparing data for analytics purposes.
- Real Time Data and Message Ingestion Options: The big data architecture includes options for capturing, storing, and stream processing of real time messages. For this, there are various options like Azure Event hubs, Kafka, etc.
- Analytical Data Stores: This includes creating big data solutions for processing data and presenting them in a structured manner so that it can be queried by analytics tools.



The various layers in a big data architecture include:



a. Identify the right big data tools and software

There are various big data tools and software that are integral for big data management

Some of them are:

Apache Hadoop Software library (big data framework) - It allows distributed processing of big data sets across multiple computer clusters. It can be easily scaled to include hundreds of machines.

Atlas.ti (research software) - It allows you to manage projects containing thousands of project documents and data segments.

Other tools and software include:

- HPCC
 - Statwing
 - CouchDB
 - Cassandra
 - Storm
 - Qubole
 - Link
 - Pentaho
 - Cloudera
 - Knime
 - Rapidminer
 - Lumify
- And more .

b. Big data network management and monitoring

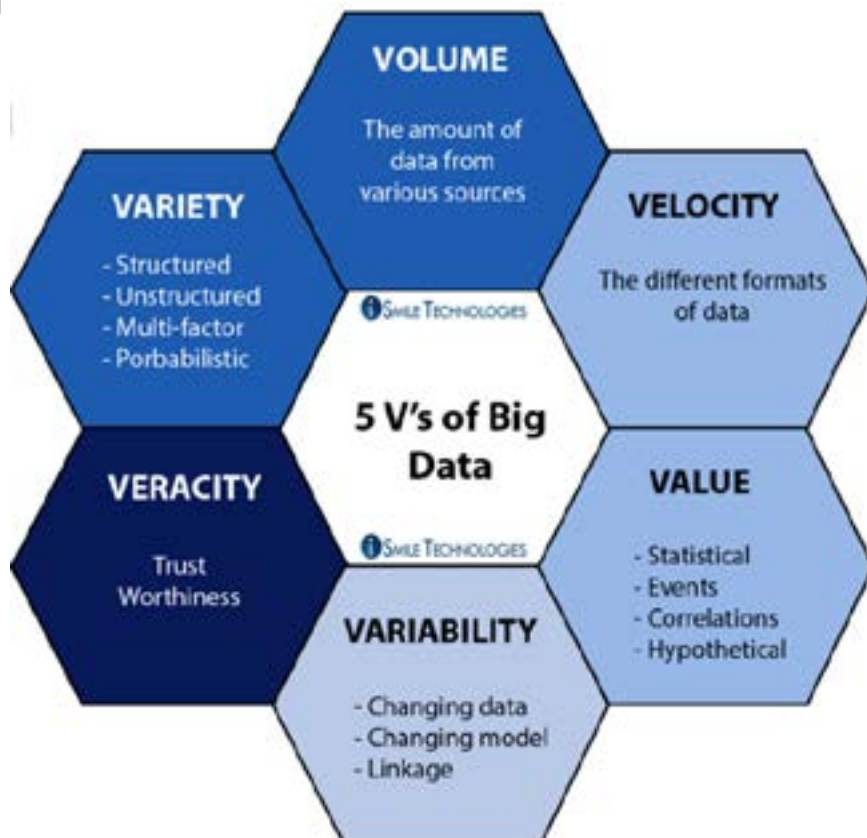
Big data network management and monitoring are required to decide the data flow for load balancing, handling real time data packets streams for higher through put, optimize bandwidth, and manage the virtual sources in a better manner.



c. Employing the big data processing programming models and tools

This takes into consideration the five Vs of big data

- Volume
- Velocity
- Variety
- Validity and
- Value





Efficient handling of large sources of company data means faster decision making, being able to better exploit the potentials of data driven marketing, have increased revenue, improved customer service, more accuracy in decision making through inferences drawn from analytics, and many more Hence, big data management carries utmost importance for organisations having an unusually high volume of data.



About Us

iSmile Technologies helps you devise and implement the best big data solutions for your business

WITH US, MORE DATA MEANS MORE PROFITS
Contact us at (732) 347-6245

Mail us at service@ismiletechnologies.com



 **SMILE TECHNOLOGIES**