

Big Data Analytics with Mongo DB

Prof. Shreeya D. Palkar¹, Prof. Swapnali M. Ponshe²

Department of Computer Science
Dapoli Urban Bank Senior Science College, Dapoli

Abstract- To build an intelligent platform for your data we need to find difference between the outperforming or lagging behind your competitors. Big Data means new opportunities for organizations to create business value and extract it. The MongoDB NoSQL database can underpin many Big Data systems, not only as a real-time, operational data store but in offline capacities as well. With MongoDB, organizations are serving more data, more users, more insight with greater ease and creating more value worldwide

Keywords- big data, Mongo DB.

I. INTRODUCTION

Big Data is the next generation of data warehousing and business analytics and is poised to deliver top line revenues cost efficiently for enterprises. Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse. This definition is intentionally subjective. We assume that, as technology advances over time, the size of datasets that qualify as big data will also increase.

Specifically, Big Data relates to data creation, storage, retrieval and analysis that is remarkable in terms of volume, velocity, and variety:

1. Volume: this is refer as amount of the data.
2. Velocity: this is refer as the speed of data creation
3. Variety: any organization contains variety of the data Structure of Big Data

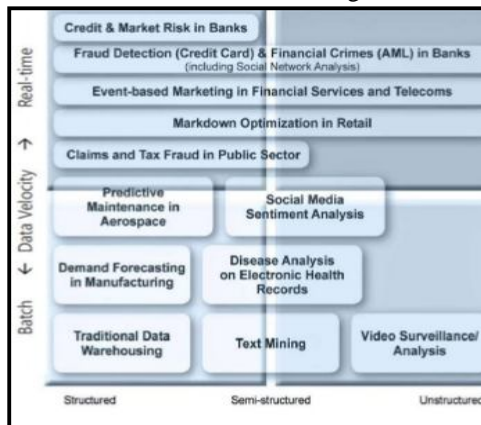


Fig 1: Structure of Big Data

Overview of Operational vs. Analytical Systems		
	Operational	Analytical
Latency	1 ms - 100 ms	1 min - 100 min
Concurrency	1000 - 100,000	1 - 10
Access Pattern	Writes and Reads	Reads
Queries	Selective	Unselective
Data Scope	Operational	Retrospective
End User	Customer	Data Scientist
Technology	NoSQL	MapReduce, MPP Database

Fig. 2: Overview of Operational vs. Analytical Systems

On the storage layer, traditional relational databases were not designed to take advantage of horizontal scaling. A class of new database architectures, dubbed NoSQL databases, are designed to take advantage of the cloud computing environment. NoSQL databases are natively able to handle load by spreading data among many servers, making them a natural fit for the cloud computing environment. Part of the

reason NoSQL databases can do this is that related data is always stored together, instead of in separate tables. This document data model, used in MongoDB and other NoSQL databases, makes them a natural fit for the cloud computing environment.

II. WHY USE MONGO DB FOR BIG DATA

MongoDB is built for the cloud. Its native scale-out architecture, enabled by 'sharding,' aligns well with the horizontal scaling and agility afforded by cloud computing. Sharding automatically distributes data evenly across multi-node clusters and balances queries across them. In addition, MongoDB automatically manages sets of redundant servers, called 'replica sets,' to maintain availability and data integrity even if individual cloud instances are taken offline.

To ensure high availability, for instance, users can spin up multiple members of a replica set as individual cloud instances across different availability zones and/or data centres. With MongoDB Atlas, both the infrastructure and the

storage layer are delivered as a service. Rather than managing the deployment of replica sets or sharded clusters, MongoDB Atlas automates these operational tasks for the end user.

III. INSTALLATION AND WORKING WITH MONGO DB

Following are the steps for the installation and running the mongo DB:

1. Open an Administrator command prompt.

2. Install MongoDB for Windows:

```
(msiexec.exe /q /i mongodb-win32-x86_64-2008plus-ssl-3.2.21-signed.msi ^
INSTALLLOCATION="C:\Program
Files\MongoDB\Server\3.2.21\" ^ ADDLOCAL="all")
```

3. Set up the MongoDB environment.

MongoDB requires a data directory to store all data. MongoDB's default data directory path is \data\db. Create this folder using the following commands from a Command Prompt.

If your path includes spaces, enclose the entire path in double quotes, for example:
`"C:\ProgramFiles\MongoDB\Server\3.2\bin\mongod.exe" --dbpath "d:\test\mongo db data"`

4. Start MongoDB

To start MongoDB, run mongod.exe. For example, from the Command Prompt
`"C:\ProgramFiles\MongoDB\Server\3.2\bin\mongod.exe"`

5. Connect to MongoDB

To connect to MongoDB through the ~bin.mongo.exe shell, open another Command Prompt.
`"C:\ProgramFiles\MongoDB\Server\3.2\bin\mongo.exe"`

6. Begin using MongoDB

IV. APPLICATION OF BIG DATA

Following are some applications of big data:

1. Healthcare
2. Manufacturing
3. Media & Entertainment
4. Internet of Things (IoT)

5. Government
6. Cyber security & Intelligence

These are some real time applications of Big Data

V. CONCLUSION

Big data is becoming an important element in the way organizations are leveraging high-volume data at the right speed to solve specific data problems. However, big data does not live in isolation. To be effective, companies often need to be able to combine the results of big data analysis with the data that exists within the business.

REFERENCES

- [1] Judith Hurwitz, Alan Nugent, Dr. Fern Halper, and Marcia Kaufman, "Big data for Dummies", A Wiley Brand.
- [2] Jure Leskovec, Anand Rajaraman, Jeffrey D. Ullman, "Mining of Massive Datasets".