



ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.

Approved by AICTE, New Delhi

Department of Computer Science & Engineering (Data Science)



PROJECT REPORT

ON

From BI to Big Data: Explain, Design & Defend

Subject Name: Big Data Analytics

Subject Code: BAD601

Submitted By:
(Rakshith S– 1AY23CD401)

Submitted To:
Ms. Surbhi



ACHARYA INSTITUTE OF TECHNOLOGY
Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.
Approved by AICTE, New Delhi
**Department of Computer Science & Engineering
(Data Science)**

TABLE OF CONTENTS

SI.No.	Content	Page No.
1.	Task 1	3
2.	Task 2	4-6
3.	Task 3	7-9
4.	Task 4	10
5.	Bonus Challenge	11



ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.

Approved by AICTE, New Delhi

Department of Computer Science & Engineering (Data Science)

TASK 1: Big Data in Daily Life

The one real world Application I choose : **Flipkart**

Flipkart isn't just a shopping app, it's a massive data factory that processes billions of interactions

Infographic of Flipkart

Fipkart
Big Data at Flipkart

Classification of Data

- Structured Data**
 - Customer details,
 - Customer profiles, inventory
- Semi-Structured Data**
 - Clickstream logs,
 - Search history
- Unstructured Data**
 - Product images, reviews,
 - Chat messages

The 5 V's of Big Data

Volume	Velocity
Huge user data	Real-time orders
Variety	Variate
Text, images, GPS	Remove fake reviews
Value	Veracity
Personalized offers	Save mofies

Why Traditional BI Fails

- Small data capacity
- Slow processing
- Hard to scale up
- Can't handle unstructured data

Why Big Data is Required

- ✓ Live recommendations
- ✓ Personalized offers
- ✓ Predict delivery time
- ✓ Optimize warehouses

Without **Big Data**, Flipkart would be slow and outdated.
With **Big Data**, it's fast and super personalized.

Fig:1.1



ACHARYA INSTITUTE OF TECHNOLOGY
Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.
Approved by AICTE, New Delhi
**Department of Computer Science & Engineering
(Data Science)**

TASK 2: BI vs Big Data – Role Play

Manager: I don't understand why we need Big Data systems. Our SQL databases and Excel reports are working fine for Flipkart.

Consultant: They work for small-scale reporting, but Flipkart handles millions of users and products daily. The volume of data is far beyond traditional BI capacity.

Manager: But our SQL server stores all transaction records properly.

Consultant: Yes, it handles structured data like orders and payments. But Flipkart also collects product images, customer reviews, browsing patterns, and clickstream logs. That's unstructured and semi-structured data.

Manager: Can't we just upgrade to a more powerful server?

Consultant: That's vertical scaling buying a bigger machine. It's expensive and has limits. Big Data uses horizontal scaling, where many machines work together as a cluster.

Manager: We already generate daily and weekly reports. Isn't that enough?

Consultant: Those are batch reports. Flipkart needs real-time insights. When a customer searches for a product, recommendations must appear instantly.

Manager: Why is real-time processing so critical?

Consultant: Because customer behavior changes every second. During Big Billion Days, thousands of transactions happen per second. Traditional systems may slow down or even fail.



ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.

Approved by AICTE, New Delhi

Department of Computer Science & Engineering (Data Science)

Manager: Our database tables are well organized. What's the issue there?

Consultant: Traditional BI uses Schema-on-Write, meaning structure must be defined before storing data. Big Data uses Schema-on-Read, where raw data is stored first and structured later when needed.

Manager: Storing raw data sounds risky. How do we manage it?

Consultant: Hadoop's Distributed File System (HDFS) splits data into blocks and stores them across multiple nodes. It's fault-tolerant even if one system fails, the data remains safe.

Manager: What about relationships between tables? SQL manages that well.: True, but complex joins become slow with massive data. NoSQL databases are designed for high-speed access and can handle millions of user requests per second.

Manager: So are we moving from descriptive reports to something more advanced?

Consultant: Exactly. Traditional BI gives descriptive analytics what happened yesterday. Big Data enables predictive analytics what will happen next.

Manager: Can you give a Flipkart example?

Consultant: Big Data can predict which products will trend next week, which customers may cancel orders, and how much stock each warehouse needs in advance.

Manager: That would definitely improve efficiency.

Consultant: Yes, it also helps in personalized recommendations, dynamic pricing, fraud detection, and delivery route optimization.



ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.

Approved by AICTE, New Delhi

Department of Computer Science & Engineering (Data Science)

Manager: So we shouldn't completely remove our BI system?

Consultant: No. We can use a hybrid approach. Keep traditional BI for structured reporting and compliance. Use Big Data for real-time processing and advanced analytics.

Manager: I understand now. BI tells us where we've been, but Big Data helps us move ahead faster.

Consultant: Exactly. For Flipkart, adopting Big Data means better customer experience, faster decisions, and stronger market competition. Let's start planning the implementation

TASK 3: Architecture Design Challenge

1 .Traditional Data Warehouse:

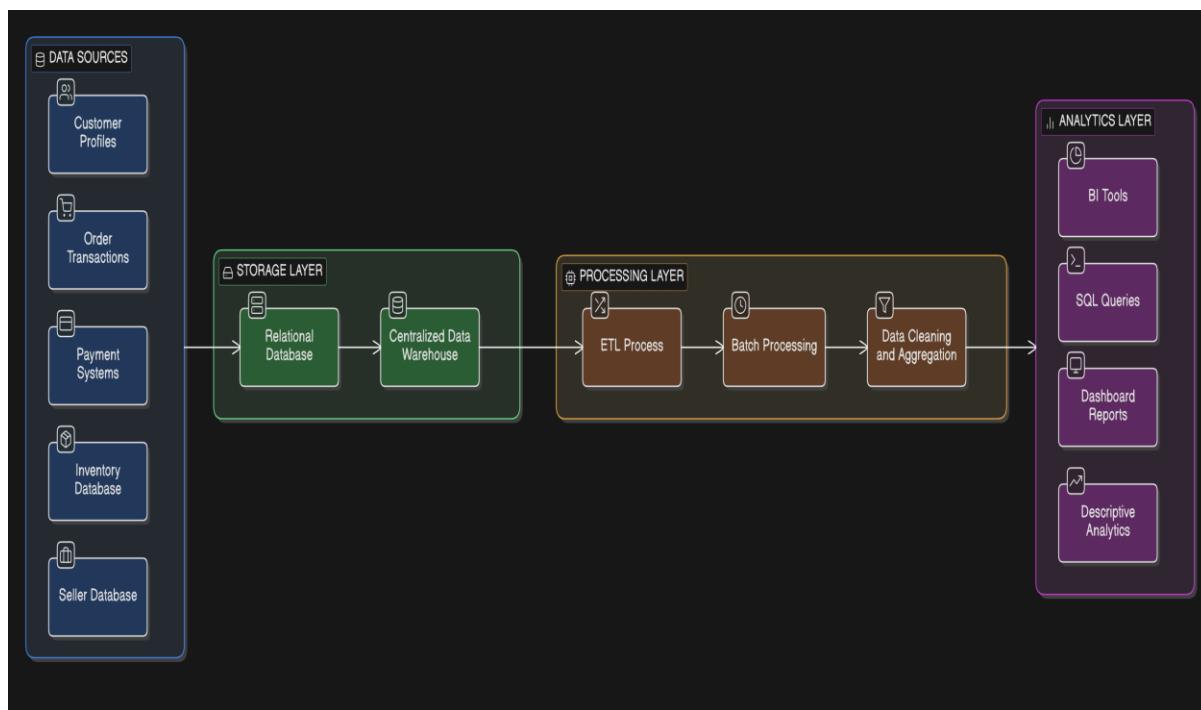


Fig 1.1

- Data passes through ETL (Extract, Transform, Load) process.
- Stored in centralized relational database (SQL).
- Processing is done in batch mode (daily/weekly).
- BI tools generate reports and dashboards

2. Hadoop based Big data Architecture

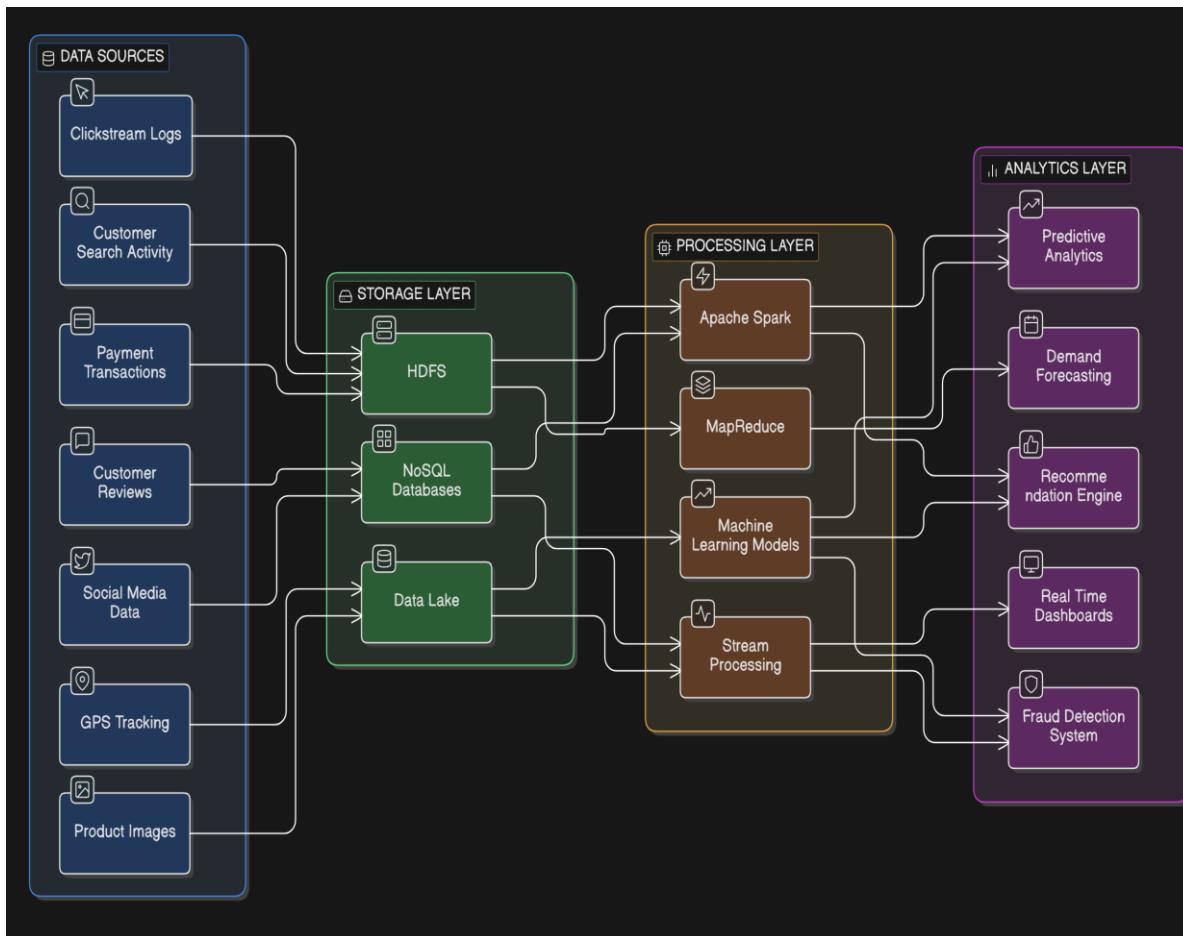


Fig 1.2

- Raw data stored in HDFS (Data Lake).
- Uses distributed storage across multiple machines.
- Processing done using Spark / MapReduce (real-time + parallel).



ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.

Approved by AICTE, New Delhi

Department of Computer Science & Engineering (Data Science)

3.1 Traditional Data Warehouse Architecture

In the traditional architecture, data from various sources such as sales systems, payment gateways, and inventory databases is collected and processed through an ETL process. The processed data is stored in a centralized relational data warehouse.

The storage layer uses SQL-based databases. Processing occurs in batch mode. The analytics layer consists of BI tools that generate reports and dashboards for business analysis.

While this architecture works for structured data and historical reporting, it struggles with scalability and real-time processing.

3.2 Hadoop-Based Big Data Architecture

In the Big Data architecture, data from Flipkart's website, mobile app, seller systems, and logistics platforms is ingested in real time using streaming tools. The data is stored in distributed storage systems such as HDFS (Hadoop Distributed File System).

The processing layer uses Apache Spark and MapReduce for parallel and real-time processing. The analytics layer applies machine learning models and BI dashboards to generate insights.

This architecture supports horizontal scaling, fault tolerance, and high-speed processing, making it suitable for large-scale operations.



**Department of Computer Science & Engineering
(Data Science)**

TASK 4: Analytics & Tool Match

Business Question	Analytics Type	Tool
What happened?	Descriptive Analytics	BI Tools (Power BI, Tableau)
Why did it happen?	Diagnostic Analytics	Hadoop + SQL
What will happen next?	Predictive Analytics	Spark (Machine Learning)
What action should be taken?	Prescriptive Analytics	Spark + NoSQL



ACHARYA INSTITUTE OF TECHNOLOGY

Affiliated to Visvesvaraya Technological University, Belagavi, Govt. of Karnataka.

Approved by AICTE, New Delhi

Department of Computer Science & Engineering (Data Science)

Bonus Challenge

Explain Big Data to a 10-year-old :

Big Data is a very large amount of information collected from many people every second. Imagine if millions of people are using their phones at the same time and clicking, searching, and buying things. All that information gets stored somewhere. When the data becomes too big for a normal computer to handle, it is called Big Data. For example, when you use Flipkart and search for shoes, the app remembers what you looked at. It also stores what other people searched and bought. This creates a huge collection of information. Special powerful computers are used to manage and study this data. These computers work together like a team to process everything quickly. Big Data helps companies understand what people like and dislike. It also helps them suggest products that you may want to buy. During big sales, it helps manage thousands of orders every second. It can even detect fake reviews or fraud. Big Data also helps companies deliver products faster by choosing the best routes.

Github Account : [rakshith70](#)