ASSIGNMENT REPORT
ON
MODULE 1

Subject Name: Big Data Analytics
Subject Code: BAD601

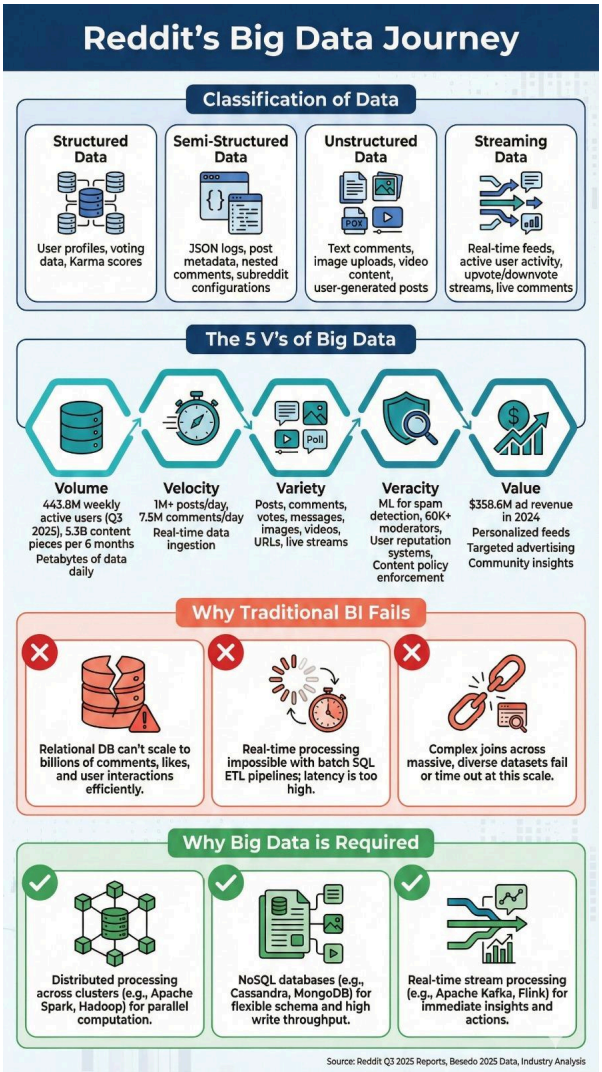
Submitted By:
Vishal Raj T (1AY23CD070)

Submitted To:
Ms. Surbhi

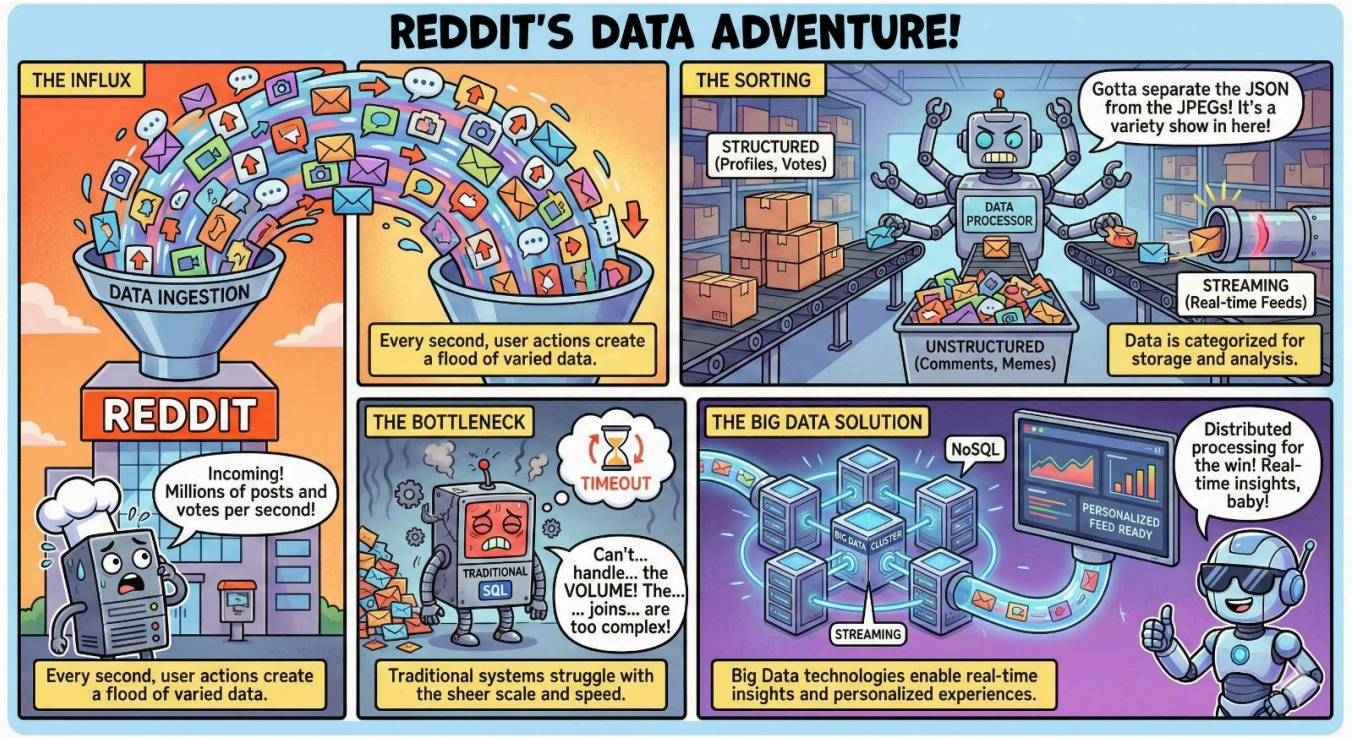
Table of Contents

TASK 1: Big Data in Daily Life – Visual Storytelling

Chosen Application: Reddit
Infographics Title: Reddit Infographics



Chosen Application: Reddit
Comic Strip Title: Journey of Reddit's Data



TASK 2: BI vs Big Data – Role Play

Dialogue: Data Consultant vs Business Manager

Data Consultant - Vishal

Business Manager - Arjun

Arjun: We already use Excel and SQL dashboards, so why are you pushing for Big Data?

Vishal: Excel and SQL are good for structured reports, but Reddit handles millions of posts and comments every hour.

Arjun: Our database stores all that information, doesn't it?

Vishal: It stores it, but processing massive unstructured text and media in real time is another challenge.

Arjun: What do you mean by unstructured data?

Vishal: Comments, memes, GIFs, videos, and constantly edited posts don't fit neatly into fixed tables.

Arjun: But we generate daily engagement reports just fine.

Vishal: Those are batch reports that summarize the past, not systems that react instantly.

Arjun: Why do we need instant reactions?

Vishal: Because harmful content, spam, or viral trends happen in seconds, not at the end of the day.

Arjun: Couldn't we just upgrade our SQL servers?

Vishal: Scaling one big server is expensive and still limited compared to distributed systems.

Arjun: So Big Data scales differently?

Vishal: Yes, it scales horizontally by adding more machines to share the workload.

Arjun: And that's where Hadoop comes in?

Vishal: Exactly, Hadoop stores and processes huge volumes of data across clusters.

Arjun: What about NoSQL, why is that necessary?

Vishal: NoSQL handles flexible data formats without redesigning schemas every time Reddit adds a new feature.

Arjun: Traditional BI tools seem simpler though.

Vishal: They are simpler, but they mainly handle structured historical data, not streaming and predictive analytics.

Arjun: How does Big Data improve user experience?

Vishal: It powers personalized feeds, detects toxic comments instantly, and identifies trending topics in real time.

Arjun: Can Excel not do predictive analysis?

Vishal: Not at Reddit's scale, especially when machine learning models analyze millions of interactions continuously.

Arjun: So Big Data helps us predict what users will engage with next?

Vishal: Yes, it helps forecast viral posts and recommend content tailored to each user.

Arjun: And that increases engagement and ad revenue?

Vishal: Exactly, better personalization leads to longer sessions and higher monetization.

Arjun: So BI looks backward, while Big Data helps us act and predict forward?

Vishal: That's right, BI explained yesterday, but Big Data shapes tomorrow.

TASK 3: Architecture Design Challenge (Reddit)

Architecture A: Traditional Data Warehouse Architecture (Reddit)

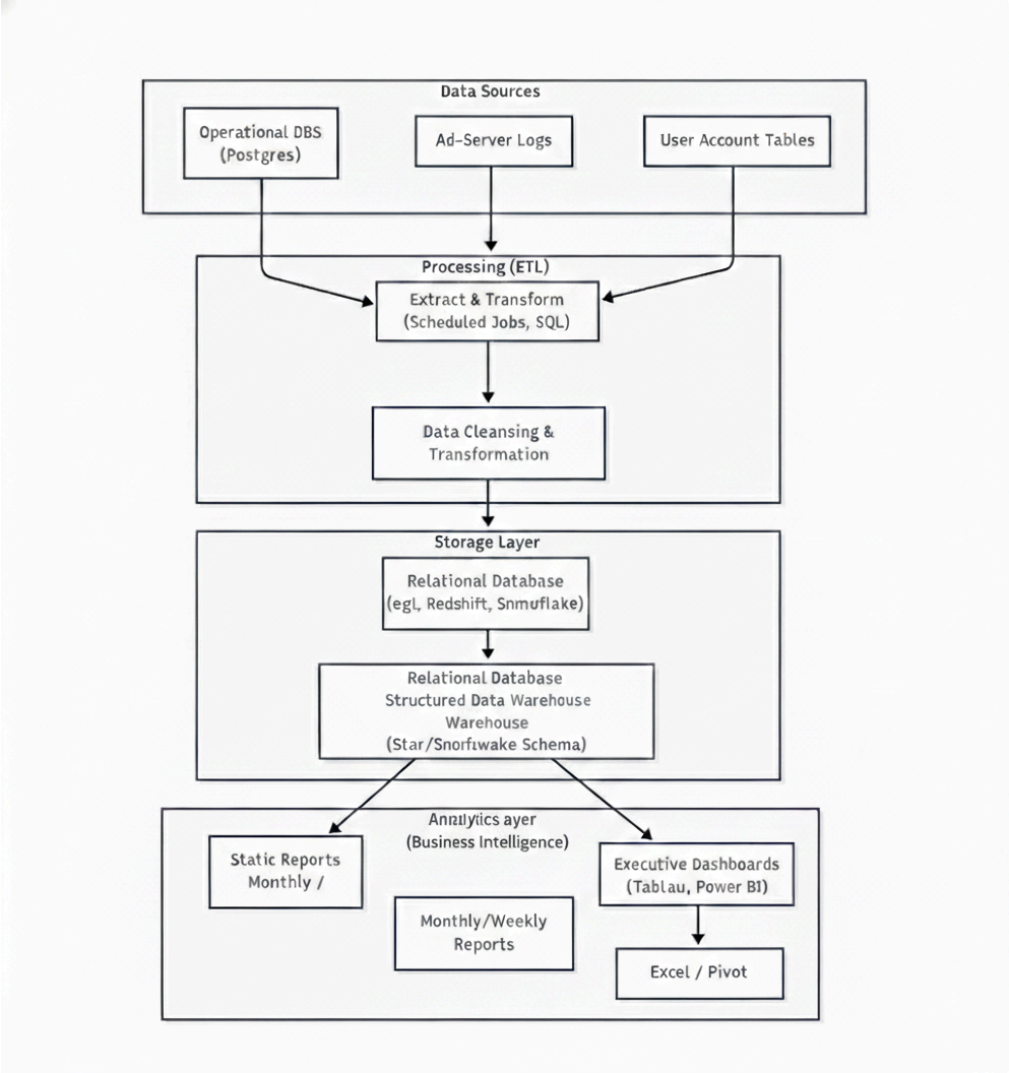


Fig 1. Traditional Data Warehouse Architecture for Reddit

Architecture B: Hadoop-Based Big Data Architecture (Reddit)

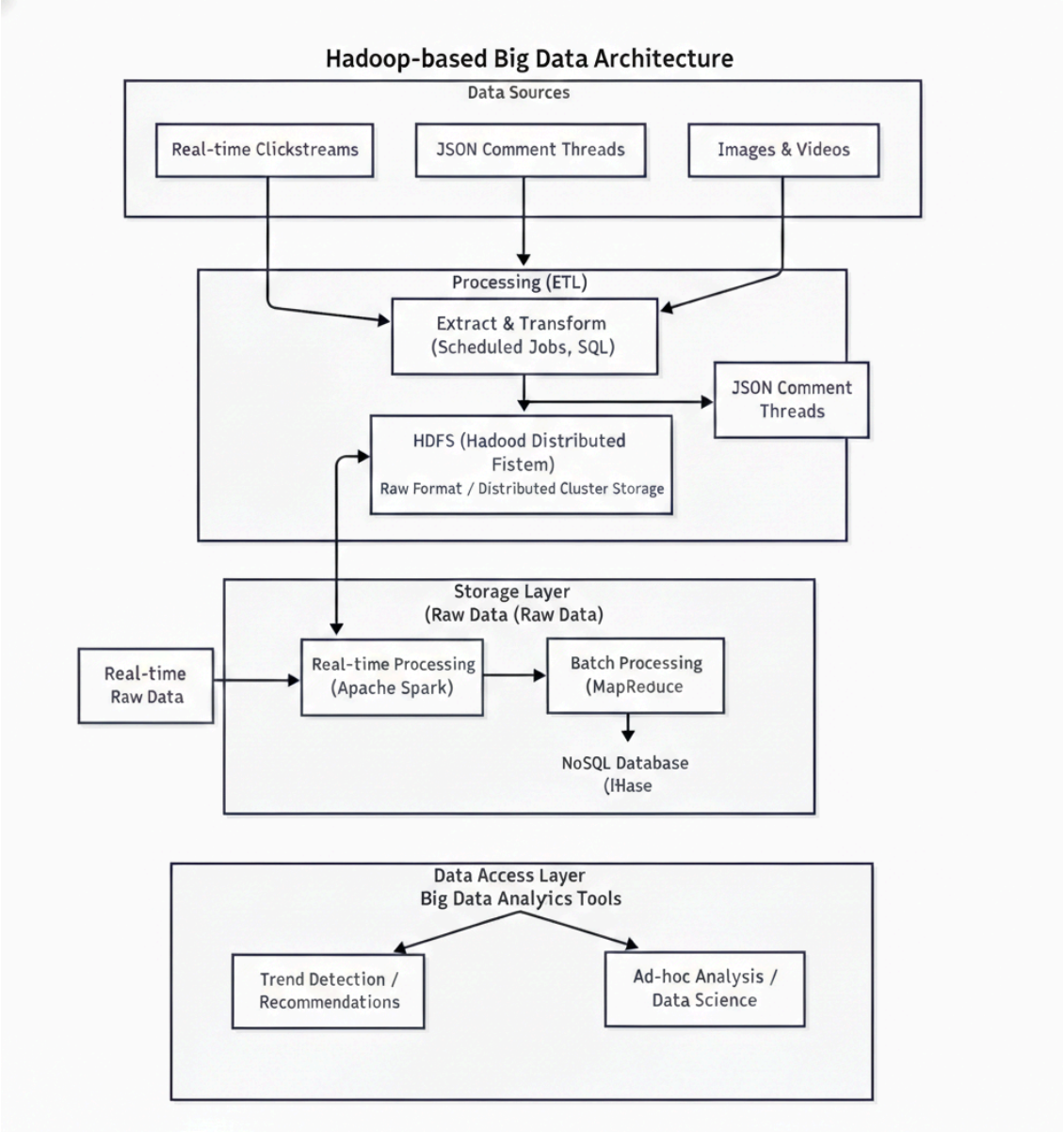


Fig 2. Hadoop-Based Big Data Architecture for Reddit

TASK 4: Analytics & Tool Match

Business Question	Analytics Type	Tool
What happened?	Descriptive Analytics	BI Tools (Tableau, Power BI, SQL, Excel)
Why did it happen?	Diagnostic Analytics	Python, R, SQL, Data Mining Tools
What will happen next?	Predictive Analytics	Hadoop, Spark ML, Machine Learning Models
What action should be taken?	Prescriptive Analytics	AI Systems, Optimization Tools, Advanced ML

BONUS: Explain Big Data to a 10-Year-Old

Imagine you have a small toy box in your room. You know exactly which toy you play with the most, which one is your favourite, and which one you don't like anymore. It's easy to remember because there are only a few toys.

Now imagine a giant toy store in a big city. Every day, thousands of children visit the store. They look at different toys, pick some up, put others back, and buy their favourites. Some toys suddenly become very popular, while others stay on the shelves. New toys arrive every week, and some old ones stop selling.

The store owner cannot sit with a notebook and track everything. There are simply too many toys and too many children making choices every second. This is where Big Data comes in.

Big Data is like a super smart robot working inside the toy store. This robot watches which toys children look at, which toys they buy together, and which toys become popular during holidays. It remembers millions of choices and learns patterns from them. After learning, it can even predict which toy will be the next big trend.

Because of Big Data, the toy store knows how many toys to order, where to place them, and which toys to recommend to different children. It helps the store make better decisions quickly and keeps the shelves full of toys that kids actually love.

In simple words, Big Data is when there is so much information that normal tools cannot handle it, so we use powerful and intelligent systems to understand it and make smart decisions.