



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: Kasinadhan M B

USN: 1AY23CD029

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

Task No.	Task	Page No
1	Big Data in Daily Life – Visual Storytelling	3
2	BI vs Big Data Play	4-5
3	Architecture Design Challenge	6
4	Analytics & Tool Match	7

TASK 1: Big Data in Daily Life – Visual Storytelling





Department of Computer Science & Engineering (Data Science)

TASK 2 : BI vs Big Data Play

Setting: Strategy meeting between Marketing Director and Data Consultant at Instagram HQ.

Marketing Director: We already use SQL dashboards and Excel reports. Engagement numbers look fine. Why complicate things with Big Data?

Consultant: Dashboards show last week's engagement. Instagram trends change every minute. By the time BI reports update, the trend is already dead.

Marketing Director: But we track likes, comments, and shares. That's enough, right?

Consultant: That's structured data. What about Reels videos, captions, hashtags, DMs, live streams? That's unstructured and semi-structured data. Traditional BI struggles with that.

Marketing Director: Can't we just scale our SQL servers?

Consultant: Scaling vertically gets expensive fast. Instagram processes billions of interactions daily. Big Data systems scale horizontally across hundreds of machines.

Marketing Director: Okay, but what real difference does that make to the user?

Consultant: Real-time personalization. When someone watches three fitness reels in a row, the Explore page updates instantly. BI systems use batch processing — meaning delayed insights.

Marketing Director: So Big Data is about speed?

Consultant: Speed, scale, and intelligence. Big Data platforms like Hadoop and Spark process massive datasets in parallel and support machine learning.

Marketing Director: How does machine learning fit into this?

Consultant: ML models analyze past behavior to predict what content a user will engage with next. That's predictive analytics. BI mostly gives descriptive analytics.

Marketing Director: And ads?



Department of Computer Science & Engineering (Data Science)

Consultant: Targeted ads rely on prescriptive analytics. Big Data systems evaluate user interests, behavior patterns, and engagement probability in real time before showing an ad.

Marketing Director: What about fake accounts and bots?

Consultant: Big Data tools detect anomalies using pattern recognition across millions of accounts. Traditional BI can't analyze behavioral signals at that scale.

Marketing Director: This sounds expensive and complex.

Consultant: Cloud-based Big Data platforms auto-scale. During viral trends or major events, infrastructure expands automatically. Traditional systems crash under sudden spikes.

Marketing Director: So BI tells us what happened. Big Data tells us what will happen and what to do next?

Consultant: Exactly. BI is reactive. Big Data is predictive and prescriptive.

Marketing Director: And the business benefit?

Consultant: Higher engagement, better ad targeting, improved user retention, and increased revenue.

Marketing Director: Alright. I'm convinced. It's not about replacing BI — it's about evolving beyond it.

Consultant: That's the shift. Hybrid intelligence — structured reporting plus real-time Big Data analytics.

TASK 3: Architecture Design Challenge & Traditional Data Warehouse Architecture

Figure 1: Traditional Data Warehouse Architecture

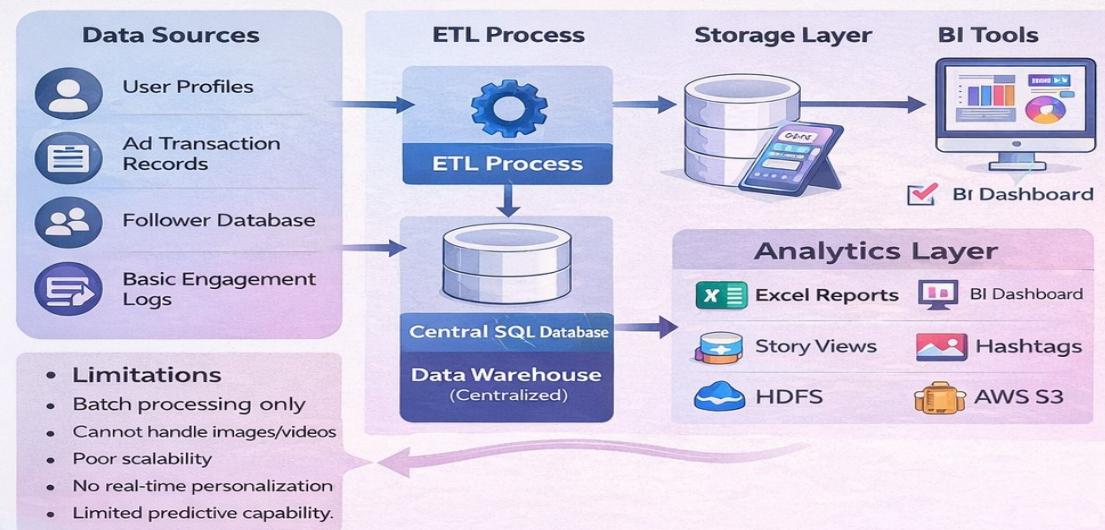
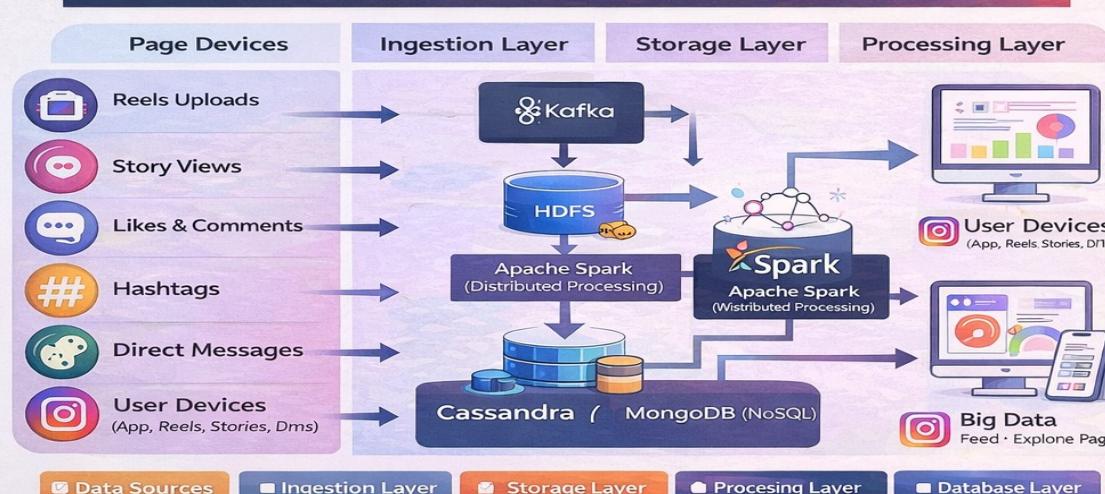


Figure 2: Big Data Architecture for Instagram



Traditional Data Warehouse systems are centralized and optimized for structured data and periodic reporting. In contrast, Big Data architecture is distributed, scalable, and capable of processing high-volume, high-velocity, and high-variety data in real-time. This enables Instagram to deliver personalized content, targeted advertisements, and predictive analytics efficiently.



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 4: Analytics & Tool Match

Business Question	Analytics Type	Big Data Tool	Instagram Example
What happened? (Historical data)	DESCRIPTIVE Analytics	Hadoop + Hive	"This reel received 2 million views and 150K likes."
Why did it happen? (Pattern discovery)	DIAGNOSTIC Analytics	Apache Spark SQL	"The reel went viral due to trending hashtags and high share rate."
What will happen next? (Future trends)	PREDICTIVE Analytics	Spark ML lib / Machine Learning Models	"This creator's next reel is likely to gain high engagement."
What action should be taken? (Recommendations)	PRESCRIPTIVE Analytics	ML Models + NoSQL (Cassandra/Mongo DB)	"Show this reel to users interested in fitness content."



TASK 5:Bonus Challenge

Okay, imagine Instagram is like a magic TV that knows exactly what you like.

When you watch funny dog videos, like superhero reels, or follow football pages, Instagram secretly takes notes. Not scary notes — just simple notes like:

“Ohhh, this kid likes dogs!”

“Ohhh, they watched that football video till the end!”

Now imagine not just you... but millions and millions of people using Instagram at the same time. Everyone is liking, scrolling, watching, and commenting.

That's a LOT of information.

So much that one computer would say,

“Whoa, this is too much for me!”

So Instagram uses many, many computers working together like a big team of helpers. One helper checks what videos you liked. Another helper checks what you skipped. Another helper looks at what kids who like the same things as you are watching.

Then a super smart brain (that's called Artificial Intelligence) looks at all the notes and decides:

“Hmm... this kid likes funny dogs and football. Let's show more of that!”

That's why when you open Instagram, it feels like it knows you.

Big Data is just a big way of saying:

“We have sooo much information, we need lots of computers working together to understand it fast.”

Without Big Data, Instagram would just show random stuff.

With Big Data, it feels like magic.