



# 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:** Consultant, I just don't see the point. My Excel macros function, my SQL queries are clear, and my team is aware of the precise location of each data row. When something isn't broken, why fix it?

**Consultant:** It's constricted rather than broken. In a suit and tie, you're attempting to complete a marathon. You're moving, of course, but the Three V's Volume, Velocity, and variety are going to cause you to overheat.

**Manager:** My SQL server handles my volume just fine.

**Consultant:** For now. But your Traditional BI relies on Vertical Scaling buying a bigger, more expensive server every time you grow. Eventually, you hit a ceiling. Big Data uses Horizontal Scaling, adding hundreds of cheap nodes to a cluster instead.

**Manager:** Okay, but what about the Variety part? Data is data.

**Consultant:** Not in the modern world. Your SQL setup is built for Structured Data neat little tables. But 80% of your customer insights are in Unstructured Data: voice notes, images, and sensor logs. Traditional BI tools literally can't see that content.

**Manager:** So I just need a better import tool?

**Consultant:** No, you need a different philosophy. You're currently using Schema-on-Write, where you must define exactly what the data is before you save it. Big Data uses Schema-on-Read we dump everything into a Data Lake and figure out the structure only when we need to ask a question.

**Manager:** Dumping everything in a lake sounds like a mess. How do you even find anything?

**Consultant:** That's where Hadoop comes in. It uses a Distributed File System (HDFS). It breaks your data into chunks and replicates them across the cluster. It's Fault Tolerant, meaning if one computer dies, the system doesn't even skip a beat.

**Manager:** My SQL database has relationships for a reason. Everything is connected.



# 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)

**Consultant:** And that's exactly why it's slow. For massive scale, we use NoSQL databases. They abandon the rigid table structure for high-speed, flexible storage that can handle millions of hits per second without locking up.

**Manager:** Is that why my real-time dashboard usually has a 24-hour delay?

**Consultant:** You're stuck in Batch Processing (ETL). By the time you see a trend, it's already historical. We need Stream Processing tools like Apache Spark to analyze data the second it's created.

**Manager:** So what's the Executive Summary? Why should I care?

**Consultant:** Because right now you're doing Descriptive Analytics . Big Data gives you Predictive Analytics. It uses Machine Learning to tell you what will happen predicting Customer Churn before the customer even knows they're unhappy.

**Manager:** Alright. You've convinced me that my spreadsheets are a safety net I've outgrown. Where do we start?

**Consultant:** We start by identifying your most valuable dark data the stuff you're collecting but not using. Once we unlock that with a modern stack, you'll stop looking at the past and start engineering the future. Let's grab a coffee and map out your first Hadoop cluster.