BLOCKCHAINS

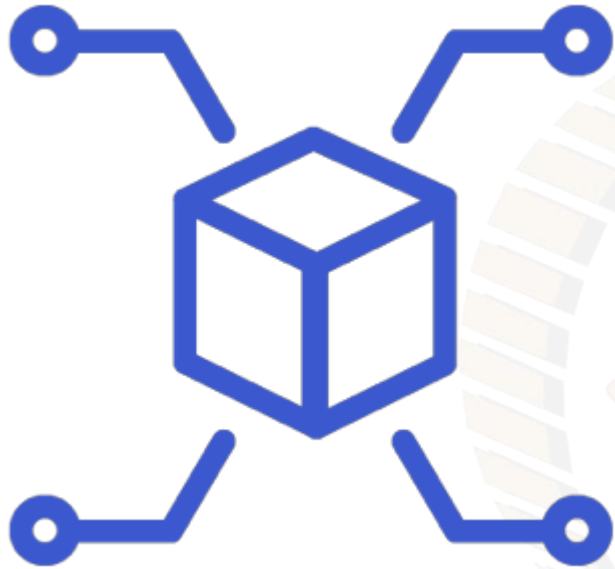## ARCHITECTURE, DESIGN AND USE CASES

SANDIP CHAKRABORTY
COMPUTER SCIENCE AND ENGINEERING,
IIT KHARAGPUR

PRAVEEN JAYACHANDRAN
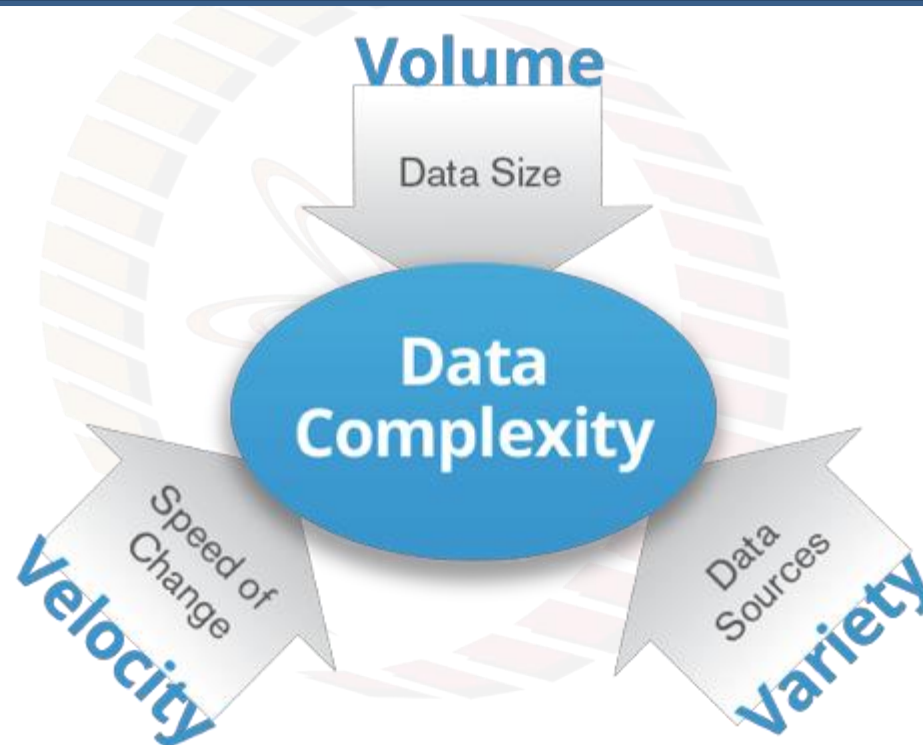IBM RESEARCH,
INDIA

IIT KHARAGPUR

1

Blockchain for Data Analytics I Blockchain and Big Data

Big Data

- **Volume**

- **Velocity**

- **Variety**

IIT KHARAGPUR

Big Data Processing

- Early and mid 2000s
  - ZooKeeper at Yahoo
  - BigTable and MapReduce at Google
  - Cassandra at Facebook

- Open Source Projects
  - Hadoop Distributed File System (HDFS)
  - Hadoop MapReduce

- **Who will control the infrastructure when there are multiple actors involved?**

    – Geospatial data - shared among multiple government and private agencies - ISRO, DRDO, Meteorological department, land registry department, and so on - Who will take charge of the data?

    – If you have multiple copies of the data at different locations, how will you know which one is the most updates?

IIT KHARAGPUR

- **How well can you trust the data?**
  - You have generated the data yourself, how will you proof that you are the originator?
  - How will you handle crash and malicious behaviors during data transfer from the source?

IIT KHARAGPUR

- **How do you monetize the data?**
  - How do you transfer the rights of the data?
  - Can we develop a universal data marketplace? - Look data like electricity or Internet

Image Source: http://www.narolainfotech.com

Blockchain Use Case: Shared Control of Big Data Infrastructure

- **Blockchain database** - control of the database infrastructure is shared across entities - within an enterprise, within a consortium, or across the planet

  - **BigchainDB** - a blockchain powered database (https://www.bigchaindb.com/)

- Infrastructure can be spread across different locations, still the properties of a database (integrity, consistency, etc.) are ensured through a blockchain

Blockchain Use Case: Shared Control of Big Data Infrastructure

- **A Use Case:**
    - Each regional office with its own *sysadmin* controls one node of the overall database
    - The entire database is controlled collectively
    - The data is still protected even if one or two sysadmin goes rogue, or a regional office is hacked

- Consider a data pipeline:

  - IoT Sensors $\rightarrow$ Kinesis/Event Hub + Stream analysis $\rightarrow$ HDFS storage $\rightarrow$ Spark data cleaning $\rightarrow$ Spark normalization $\rightarrow$ MongoDB storage $\rightarrow$ Tableau analytics

Blockchain Use Case: Audit Trails on Data

Before each data pipeline step starts, *time-stamp* the input data as follows:

- Create a transaction (a hash of the data and the meta-data)
- Cryptographically sign the transaction
- Write the transaction to the blockchain database - this will automatically time-stamp the transaction
  - You have immutable evidence that you had access to that data at that point of time

Blockchain Use Case: Audit Trails on Data

- **How will you proof that you are the originator of the data?** - The data is cryptographically signed and stored in an immutable database (blockchain)

- **What about the crashes and malicious behaviors?** - Rehash the data and check with the information from blockchain - if there is a mismatch, something is wrong !
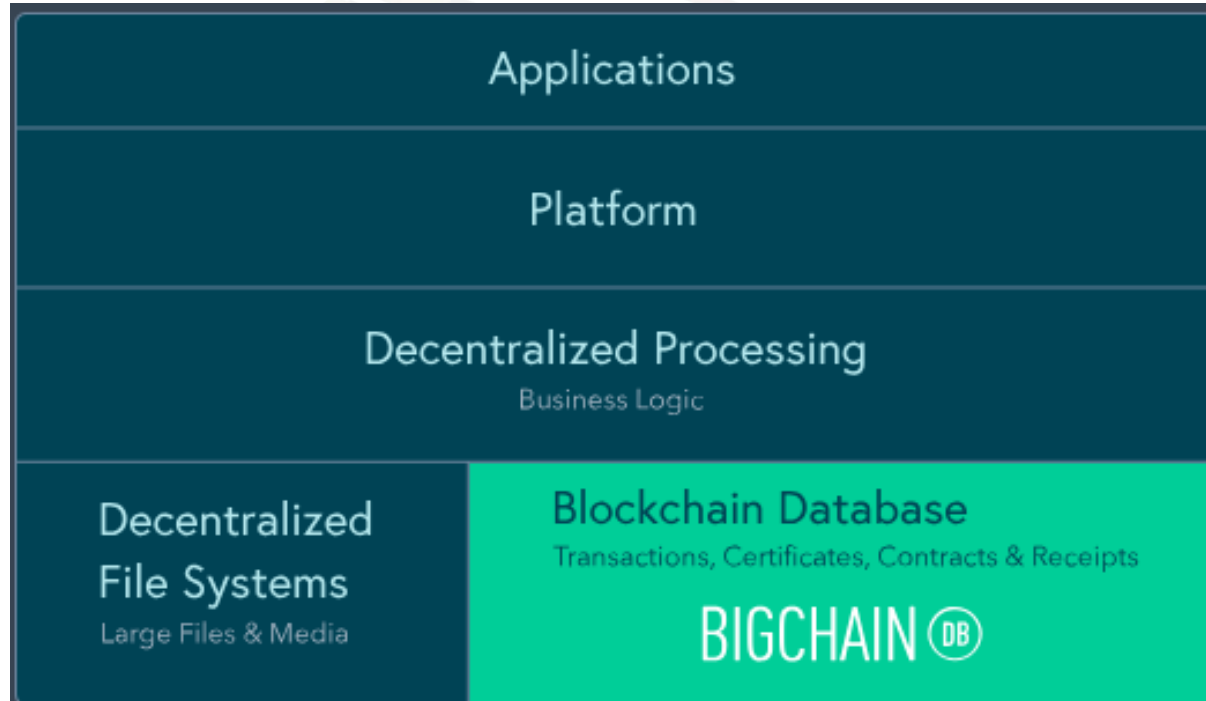
BigchainDB

BigchainDB - Features

- **Decentralization** - no single point of control, no single point of failure
- **Query** - You can write and run any MongoDB query over the database
- **Immutability** - Once stored, data can't be changed or deleted
- **Byzantine Fault Tolerant** - Up to one third of the nodes in the network can experience arbitrary faults
- **Low Latency** - Transaction finality happens first
- **Customizable** - You can design your own private network
- **Rich Permissioning** - You can set permissions at the transaction level

Decentralized Ecosystem

Blockchain vs Distributed Database vs BigchainDB

	Typical Blockchain	Typical Distributed Database	BigchainDB
Decentralization	✓		✓
Byzantine Fault Tolerance	✓		✓
Immutability	✓		✓
Owner-Controlled Assets	✓		✓
High Transaction Rate		✓	✓
Low Latency		✓	✓
Indexing & Querying of Structured Data		✓	✓

Further Readings

- BigchainDB whitepaper:
  https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf

- Try BigchainDB: https://www.bigchaindb.com/developers/getstarted/