M/T/W/T/F/S/SU/ Date

Big data Modeling and Management

Week 1 :- 1st part :- Summary of Introduction to Big data,
2nd part :- Big data management.

→ Describe what "data management" means
→ Identify the primary issues involved in the management of big data.

* What is Data Management ?

→ How do we ingest the data ?
→ Where and how do we store it ?
→ How can we ensure data quality
→ What operations do we perform on the data ?
→ How can these operations be efficient
→ How to scale up data volⁿ, variety, velocity and access ?
→ How to keep the data secure ?

Ingestion Infrastruxe -

① How many data source ?
② How large are data items
③ Will the no. of data sources grow
④ Rate of data ingestion ?
⑤ What to do with bad data ?
⑥ What to do when data is too little or too much ?

Storage Infrastructure - where hardware meets Data Manage

→ How much data to store
→ How fast do we need to read/write

* Memory Hierarchy

Pric

Data Quality :- • Better quality means better analy
decision making.
M T W T F S SU • Quality assurance means needed for Date
regulatory compilance.
• Quality leads to better engagement and interaction
with external entities

* Operations of Data :-
→ Operation on single data items that produces a sub-item
→ Operation of collections of data items.
→ Operation that select a part of a collection.
→ Operations that combine two collections.
→ Operations that compute a function on a collection.

* Efficiency of Data Operations
• Measured by time and space
• should use parallelism

Achieving Scalability

① Vertical Scaling : (scale up)
Adding more processors and RAM, buying a more expensive and robust server.
② many operations perform better with more memory, more cores.
Maintenance can be difficult, expensive.

Scaling up and Scaling Out

① Horizontal scaling (scale-out)
Adding more, possibly less powerful machines that interconnect over a network
② Parallel operations will possibles be slower.
③ Easier in practice to add more machine.

* The server industry has many sol^n's for scale up / scale out

Keeping Data secure :-
Data security :- a must for sensitive data
Increasing the no of machines leads to more security
Data in transit must be secure.
Encryption and decryption ↑se security but make data operations expensive.