How Big Data is Changing the Way We Plan our IT Infrastructure

Dan Rosanova

Senior Architect

West Monroe Partners
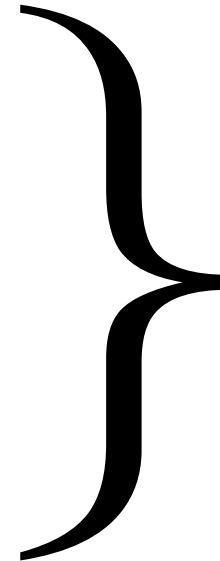
Big Data in Utilities

- Only prediction of use so far is "Big"

- We are only starting to understand how this new supply of data will affect our organizations

- Much of the innovation will be done in the future – by utilities and researchers

Big Data is a Big Challenge

- IT Infrastructure is new and very different

- Future demand and storage capabilities are not clear

- Tools we will use may not have been created yet

Current IT Infrastructure

- Virtualization

- Storage Area Networks

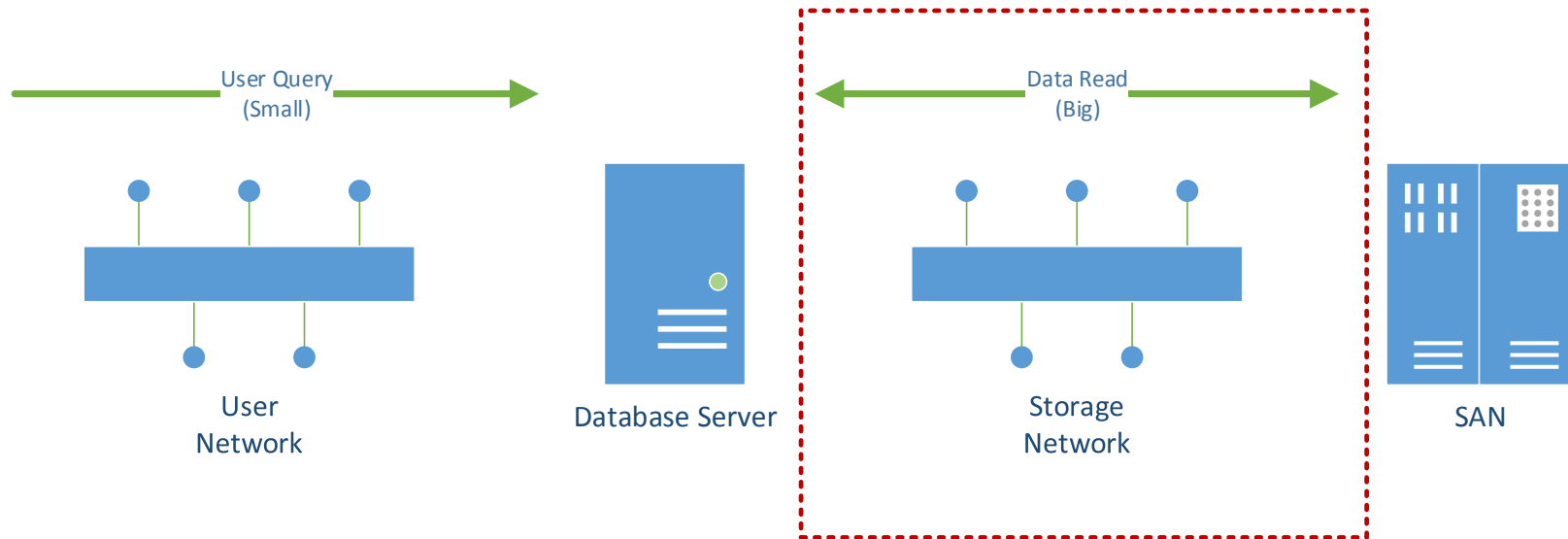- Relational Databases

- High Speed Networks

} Centralized IT Infrastructure Architecture

Centralized IT Architectures

- Designed to solve specific problems
- Are costly
- Require
  - Predictable growth
  - Well modelled consumption
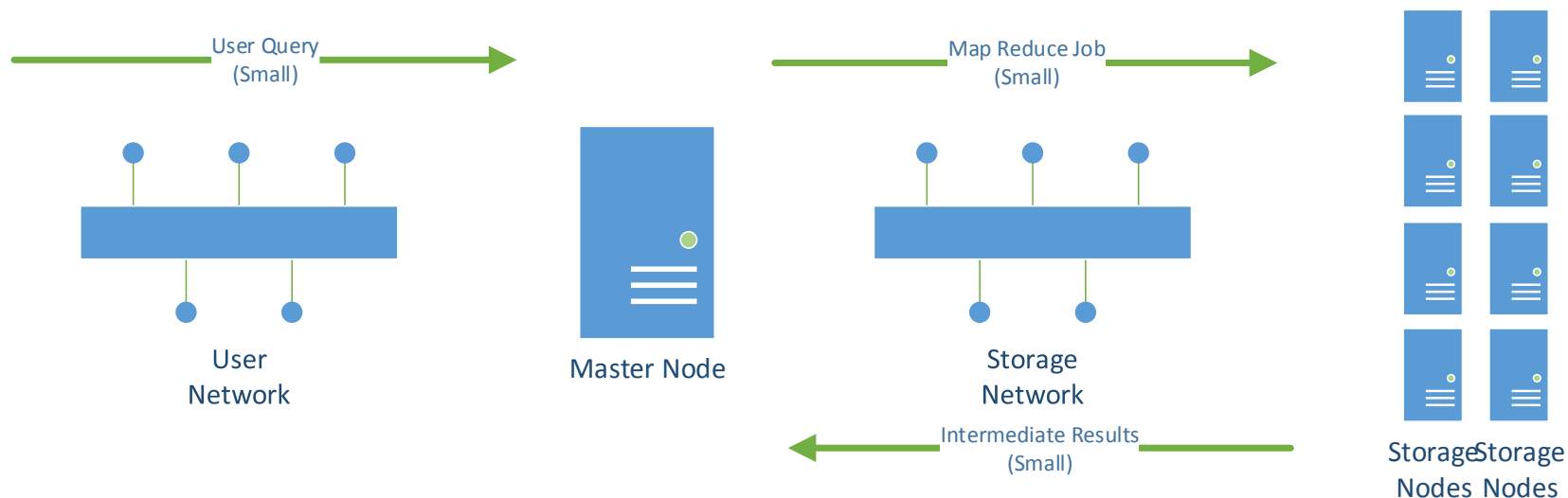
Centralized Query Model

Sum 1 Trillion Data Points

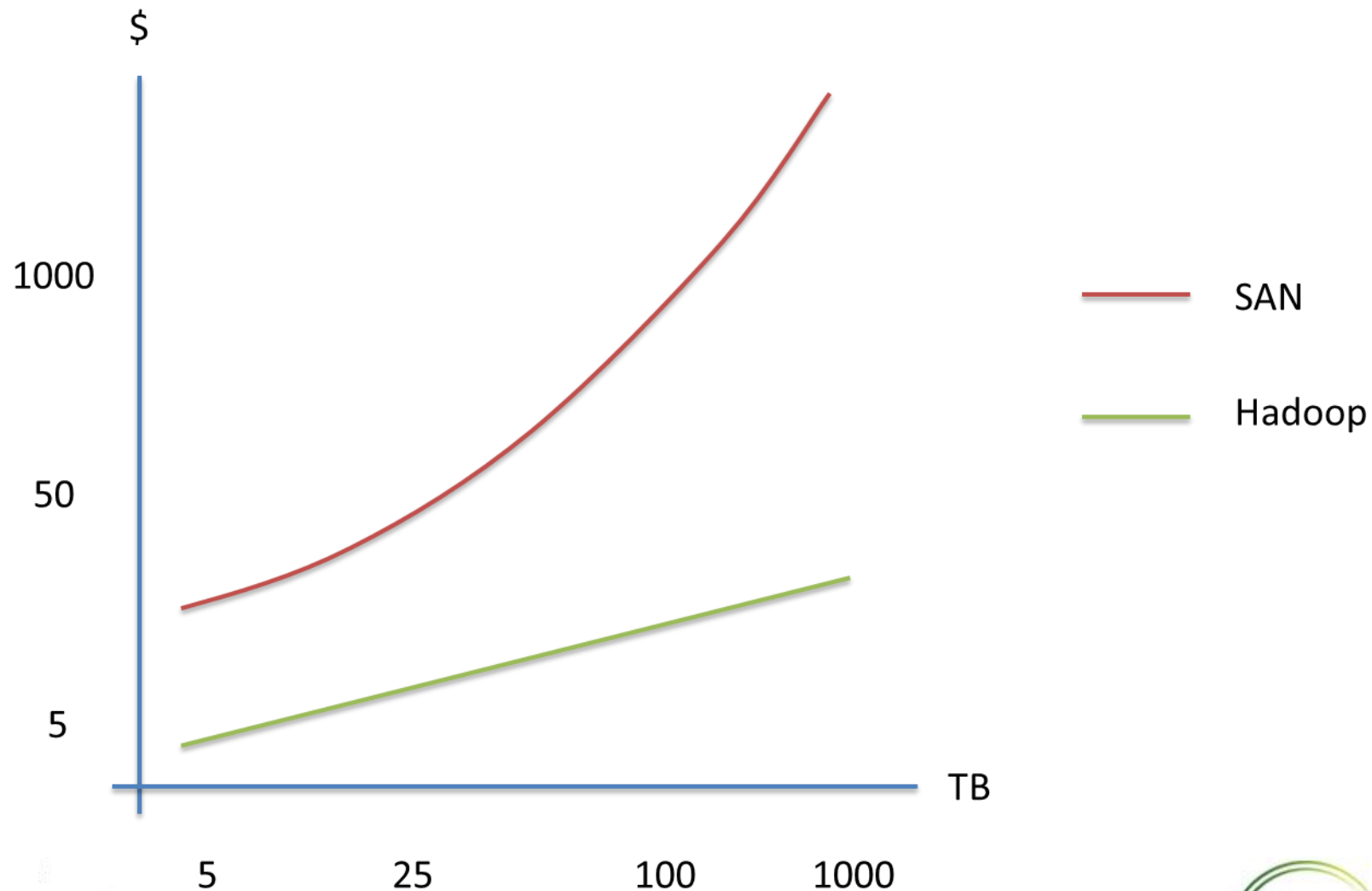
Copy 1 trillion items across network

Big Data (Hadoop) Model

Sum 1 Trillion Data Points

User Query
(Small)

Map Reduce Job
(Small)

User
Network

Master Node

Storage
Network

Intermediate Results
(Small)

Storage Storage
Nodes Nodes

Copy 1 query and N
results across network

Storage Costs

Summary

- Your current IT Infrastructure is good
- Use it for what it is good for
- Look to a different IT Infrastructure for Big Data
- This other infrastructure is
  - Commodity (cheap) hardware
  - JBOD (just a bunch of disks)