Topic 8
Introduction to NoSQL Database

SECD2523 Database

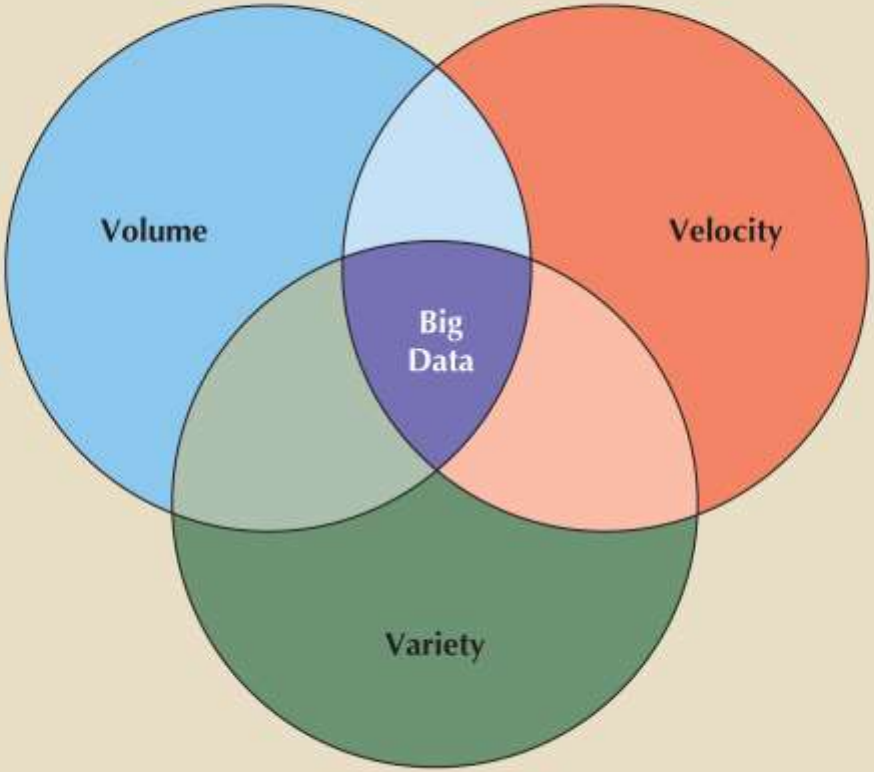
Semester 1 2022/2023

Learning Objective

- After completing this chapter, you will be able to:
  - Summarize the four major approaches of the NoSQL data model and how they differ from the relational model

Towards Big Data…

FIGURE 14.1 ORIGINAL VIEW OF BIG DATA



Towards Big Data…

- **Volume**: quantity of data to be stored
  - Scaling up: keeping the same number of systems but migrating each one to a larger system
  - Scaling out: when the workload exceeds server capacity, it is spread out across a number of servers
- **Velocity**: speed at which data is entered into system and must be processed
  - Stream processing: focuses on input processing and requires analysis of data stream as it enters the system
  - Feedback loop processing: analysis of data to produce actionable results
- **Variety**: variations in the structure of data to be stored
  - Structured data: fits into a predefined data model
  - Unstructured data: does not fit into a predefined model

Towards Big Data…

- Other characteristics
  - **Variability**: changes in meaning of data based on context
  - **Sentimental analysis**: attempts to determine if a statement conveys a positive, negative, or neutral attitude about a topic
  - **Veracity**: trustworthiness of data
  - **Value**: degree data can be analyzed for meaningful insight
  - **Visualization**: ability to graphically resent data to make it understandable
- Relational databases are not necessarily the best for storing and managing all organizational data
  - Polyglot persistence: coexistence of a variety of data storage and management technologies within an organization's infrastructure

NoSQL

- NoSQL: non-relational database technologies developed to address Big Data challenges. Four major approaches: Key-value DB, document DB, column-oriented DB, Graph DB.

Key-value (KV) DB	Document DB	Column-oriented DB	Graph DB
• conceptually the **simplest** of the NoSQL data models • Store data as a **collection of key-value pairs** organized as buckets which are the equivalent of tables	• like key-value databases and can almost be considered a subtype of KV databases • Store data in **key-value pairs** in which the value components are **encoded documents** grouped into large groups called **collections**	• two technologies: • **Column-centric storage:** data stored in blocks which hold data from a single column across many rows • **Row-centric storage:** data stored in block which hold data from all columns of a given set of rows	• store data on **relationship-rich data** as a collection of **nodes** and **edges** • **Properties**: like attributes; they are the data that we need to store about the node • **Traversal**: query in a graph database

Example: Key-value (KV) DB

FIGURE 14.7 KEY-VALUE DATABASE STORAGE

Bucket = Customer

Key	Value
10010	"LName Collin FName Alfred Initial A Areacode 615 Phone 844-2573 Balance 0"
10011	"LName Dunne FName Leona Initial K Areacode 713 Phone 894-1238 Balance 0"
10014	"LName Orlando FName Myron Areacode 615 Phone 222-1672 Balance 0"

Example: Document DB

FIGURE 14.8 DOCUMENT DATABASE TAGGED FORMAT

Collection = Customer

Key	Document
10010	{LName: "Collin", FName: "Alfred", Initial: "A", Areacode: "615", Phone: "844-2573", Balance: "0"}
10011	{LName: "Dunne", FName: "Leona", Initial: "K", Areacode: "713", Phone: "894-1238", Balance: "0"}
10014	{LName: "Orlando", FName: "Myron", Areacode: "615", Phone: "222-1672", Balance: "0"}

Example: Column-oriented DB

FIGURE 14.9 COMPARISON OF ROW-CENTRIC AND COLUMN-CENTRIC STORAGE

CUSTOMER relational table

Cus_Code	Cus_LName	Cus_FName	Cus_City	Cus_State
10010	Collin	Alfred	Nashville	TN
10011	Dunne	Leona	Miami	FL
10012	Smith	Kathy	Boston	MA
10013	Olowski	Paul	Nashville	TN
10014	Orlando	Myron		
10015	O'Brian	Amy	Miami	FL
10016	Brown	James		
10017	Williams	George	Mobile	AL
10018	Farriss	Anne	Opp	AL
10019	Smith	Olette	Nashville	TN

Row-centric storage

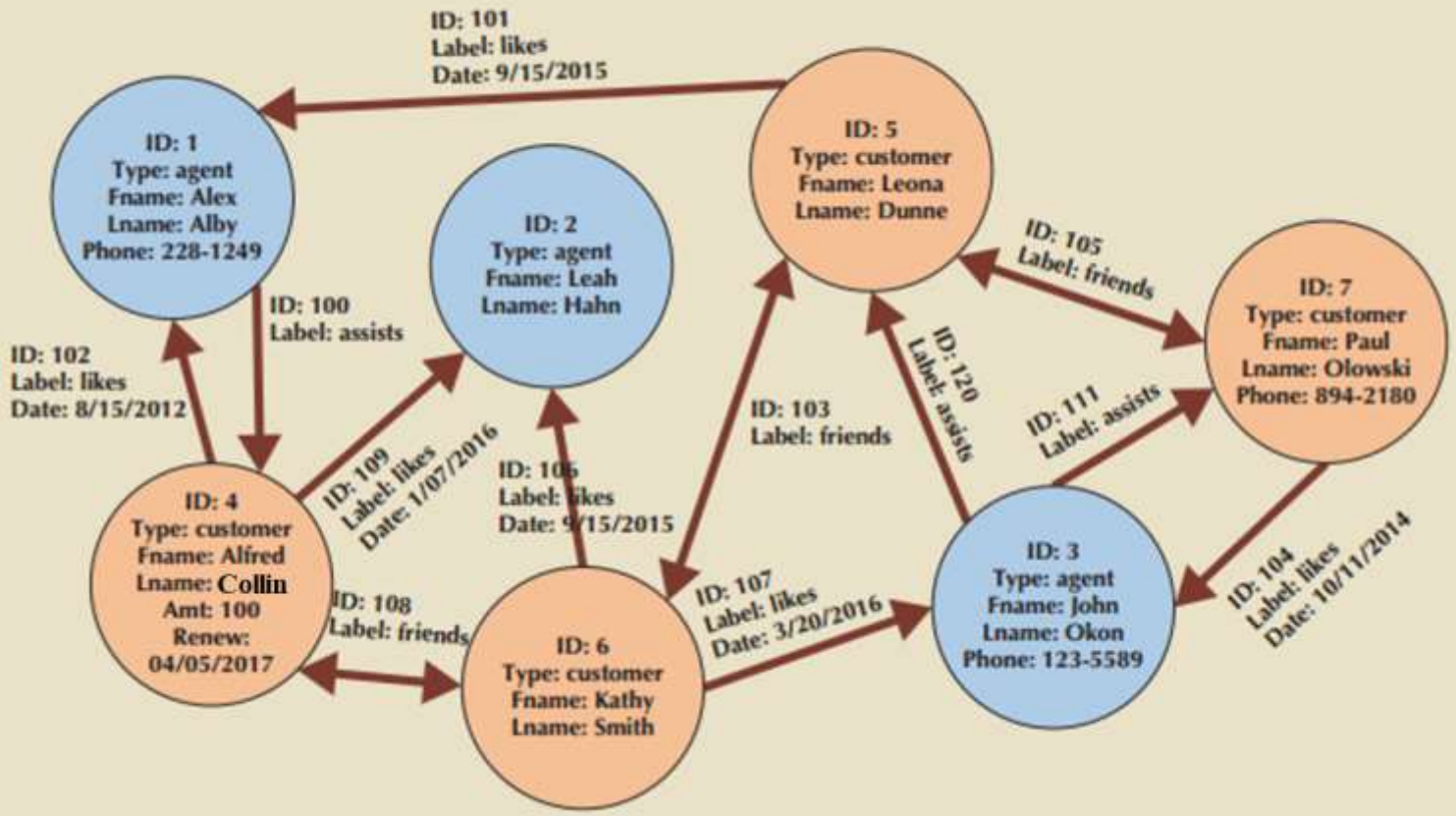
Block 1	Block 4
10010,Collin,Alfred,Nashville,TN 10011,Dunne,Leona,Miami,FL	10016,Brown,James,NULL,NULL 10017,Williams,George,Mobile,AL
Block 2	**Block 5**
10012,Smith,Kathy,Boston,MA 10013,Olowski,Paul,Nashville,TN	10018,Farriss,Anne,OPP,AL 10019,Smith,Olette,Nashville,TN
Block 3	
10014,Orlando,Myron,NULL,NULL 10015,O'Brian,Amy,Miami,FL	

Column-centric storage

Block 1	Block 4
10010,10011,10012,10013,10014 10015,10016,10017,10018,10019	Nashville,Miami,Boston,Nashville,NULL Miami,NULL,Mobile,Opp,Nashville
Block 2	**Block 5**
Collin,Dunne,Smith,Olowski,Orlando O'Brian,Brown,Williams,Farriss,Smith	TN,FL,MA,TN,NULL, FL,NULL,AL,AL,TN
Block 3	
Alfred,Leona,Kathy,Paul,Myron Amy,James,George,Anne,Olette	

Example: Graph DB

FIGURE 14.11 GRAPH DATABASE REPRESENTATION



Summary

- Big Data is characterized by data of such volume, velocity, and/or variety that the relational model struggles to adapt to it
- Volume, velocity, and variety are collectively referred to as the 3 Vs of Big Data
- NoSQL is a broad term to refer to any of several nonrelational database approaches to data management
- Key-value databases store data in key-value pairs
- Document databases also store data in key-value pairs, but the data in the value component is an encoded document
- Column-oriented databases, also called column family databases, organize data into key-value pairs in which the value component is composed of a series of columns, which are themselves key-value pairs
- Graph databases are based on graph theory and represent data through nodes, edges, and properties