

Chapter 14 Big Data Analytics and NoSQL

NoSQL

- Name given to non-relational database technologies developed to address Big Data challenges
- **Key-value (KV) databases** store data as a collection of key-value pairs organized as **buckets** which are the equivalent of tables
- Document databases store data in key-value pairs in which the value components are tag-encoded documents grouped into logical groups called collections

NoSQL

- Column-oriented databases refers to 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
- Graph databases store data on relationship-rich data as a collection of nodes and edges
 - Properties are the attributes of a node or edge of interest to a user
 - Traversal is a query in a graph database

TABLE 14.2

NoSQL DATABASES

NoSQL CATEGORY	EXAMPLE DATABASES
Key-value database	Dynamo Riak Redis Voldemort
Document databases	MongoDB CouchDB OrientDB RavenDB
Column-oriented databases	HBase Cassandra Hypertable
Graph databases	Neo4J ArangoDB GraphBase

©2017 Cengage Learning®. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website or school-approved learning management system for classroom use.

Figure 14.7- Key-Value Database Storage

FIGURE 14.7 KEY-VALUE DATABASE STORAGE

Bucket:	= Customer
Key	Value
10010	"LName Ramas 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"

Figure 14.8- Document Database Tagged Format

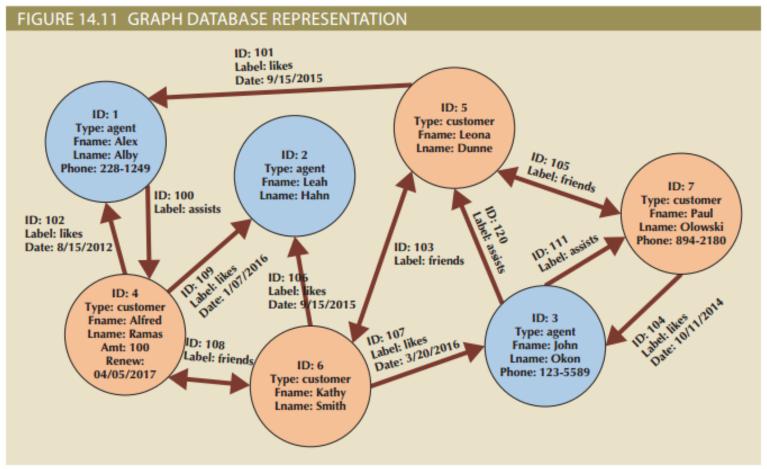
FIGURE 14.8 DOCUMENT DATABASE TAGGED FORMAT

Collecti	on = Customer
Key	Document
10010	{LName: "Ramas", 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"}

Figure 14.9- Comparison of Row-Centric and Column-Centric Storage

	cus	TOMER	relational tab	ole					
	Cu	Cus_Code Cus_LName Cus_ 10010 Ramas Alfr		Cus_F	Name	Cus_City	Cus_State		
				Alfred Leona		Nashville	TN		
						Miami	FL		
		10012	012 Smith		y	Boston	MA		
		10013	Olowski Orlando	Paul		Nashville	TN		
		10014		Myro	on				
		10015	O'Brian	Amy		Miami	FL		
		10016 Brown		James	s				
		10017	Williams	Geor	ge	Mobile	AL		
		10018	Farriss	Anne	!	Орр	AL		
	L	10019	Smith	Olett	e	Nashville	TN		
Row-cer	ntric stora	ge		_	_		Colum	n-cen	ntric storage
Block 1	Block 4	Block 4				1			Block 4
10010,Ramas,Alfred,Nashville,TN 10011,Dunne,Leona,Miami,FL		10016,Brown,James,NULL,NULL 10017,Williams,George,Mobile,AL			10010,10011,10012,10013,10014 10015,10016,10017,10018,10019				Nashville,Miami,Boston,Nashville,NULL Miami,NULL,Mobile,Opp,Nashville
Block 2	Block 5	Block 5			Block	Block 2			Block 5
10012,Smith,Kathy,Boston,MA 10013,Olowski,Paul,Nashville,TN	10018,Farriss,Anne,OPP,AL 10019,Smith,Olette,Nashville,TN				Ramas, Dunne, Smith, Olowski, Orlando O'Brian, Brown, Williams, Farriss, Smith				TN,FL,MA,TN,NULL, FL,NULL,AL,AL,TN
Block 3					Block	3			
10014,Orlando,Myron,NULL,NULL 10015,O'Brian,Amy,Miami,FL					Alfred,Leona,Kathy,Paul,Myron Amy,James,George,Anne,Olette				

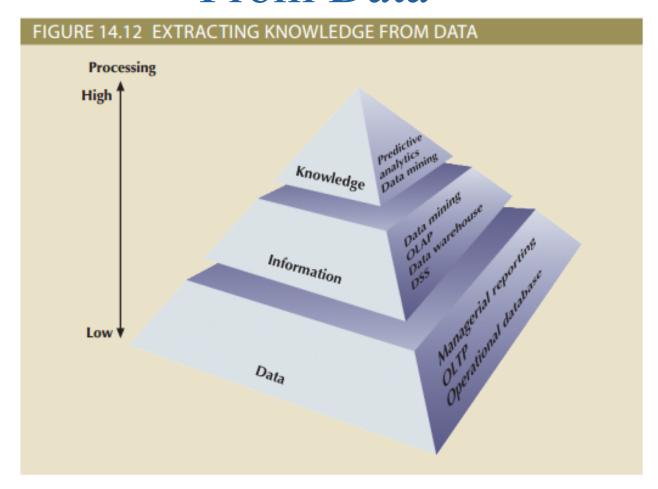
Figure 14.10- Graph Database Representation



NewSQL Databases

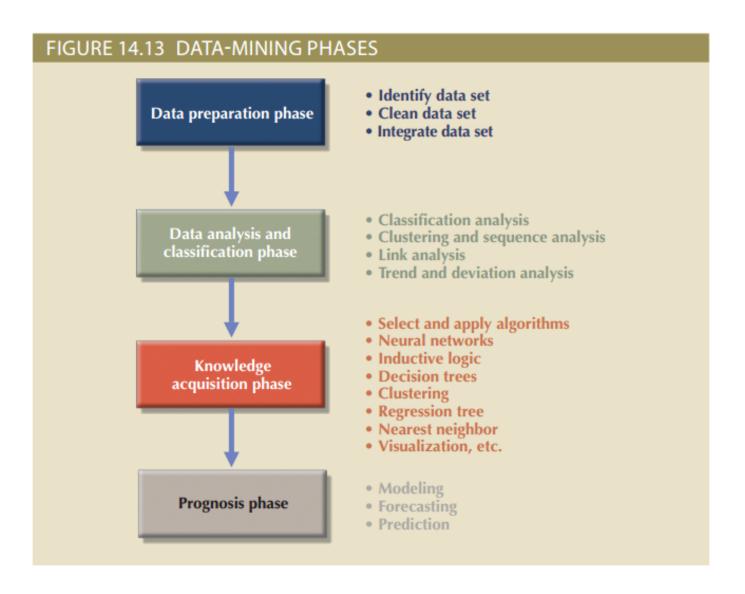
- Database model that attempts to provide ACIDcompliant transactions across a highly distributed infrastructure
 - Latest technologies to appear in the data management area to address Big Data problems
 - No proven track record
 - Have been adopted by relatively few organizations

Figure 14.12- Extracting Knowledge From Data



©2017 Cengage Learning®. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part, except for use as permitted in a license distributed with a certain product or service or otherwise on a password-protected website or school-approved learning management system for classroom use.

Figure 14.13DataMining Phases



Data Analytics

- Subset of business intelligence (BI) functionality that encompasses mathematical, statistical, and modeling techniques used to extract knowledge from data
 - Continuous spectrum of knowledge acquisition that goes from discovery to explanation to prediction
- **Explanatory analytics** focuses on discovering and explaining data characteristics based on existing data
- Predictive analytics focuses on predicting future data outcomes with a high degree of accuracy

Data Mining

- Focuses on the discovery and explanation stages of knowledge acquisition by:
 - Analyzing massive amounts of data to uncover hidden trends, patterns, and relationships
 - Forming computer models to simulate and explain findings and using them to support decision making
- Can be run in two modes:
 - Guided End-user decides techniques to apply to data
 - Automated End-user sets up the tool to run automatically and the data-mining tool applies multiple techniques to find significant relationships

Predictive Analytics

- Refers to the use of advanced mathematical, statistical, and modeling tools to predict future business outcomes with a high degree of accuracy
 - Focuses on creating actionable models to predict future behaviors and events
 - Most BI vendors are dropping the term data mining and replacing it with predictive analytics
- Models used in customer service, fraud detection, targeted marketing and optimized pricing
 - Can add value in many different ways but needs to be monitored and evaluated to determine return on investment

- <?xml version="1.0"?>
- <ancient_wonders>
- </ancient wonders>

- <?xml version="1.0"?>
- <ancient wonders>
- <wonder>
- <name>Colosus of Rhodes</name>
- </wonder>
- </ancient_wonders>

- <?xml version="1.0"?>
- <ancient wonders>
- <wonder>
- <name language="English">Colosus of Rhodes</name>
- <location> Rhodes, Greece</location>
- <height units="feet"> 107</height>
- </wonder>
- </ancient wonders>

- <?xml version="1.0"?>
- <ancient wonders>
- <wonder>
- <name > Colosus of Rhodes < /name >
- <location> Rhodes, Greece</location>
- <height units="feet">107</height>
- <main image file="colossus.jpg" w="528" h="349"/>
- </wonder>
- </ancient wonders>

```
<?xml version="1.0"?>
```

- <ancient wonders>
- <wonder>
- <name language="English">Colosus of Rhodes</name>
- <location> Rhodes, Greece</location>
- <height units="feet">107</height>
- <main image file="colossus.jpg" w="528" h="349"/>
- <source sectionid="101" newspaperid="21"></source>
- </wonder>
- </ancient wonders>

```
<?xml version="1.0"?>
<ancient_wonders>
```

<wonder>

<name language="English">Colosus of Rhodes</name>

<location> Rhodes, Greece/location>

<height units="feet">107</height>

<main_image file="colossus.jpg" w="528" h="349"/>

<!—this example comes from

Visual Quickstart Guide XML -->

<source sectionid="101" newspaperid="21"></source>

</wonder>

</ancient wonders>

```
<?xml version="1.0"?>
```

- <ancient_wonders>
- <wonder>
- <name language="English">Colosus of Rhodes</name>
- <location> Rhodes, Greece</location>
- <height units="feet">< 107</height>
- <main image file="colossus.jpg" w="528" h="349"/>
- <!—this example comes from
- Visual Quickstart Guide XML -->
- <source sectionid="101" newspaperid="21"></source>
- </wonder>
- </ancient wonders>