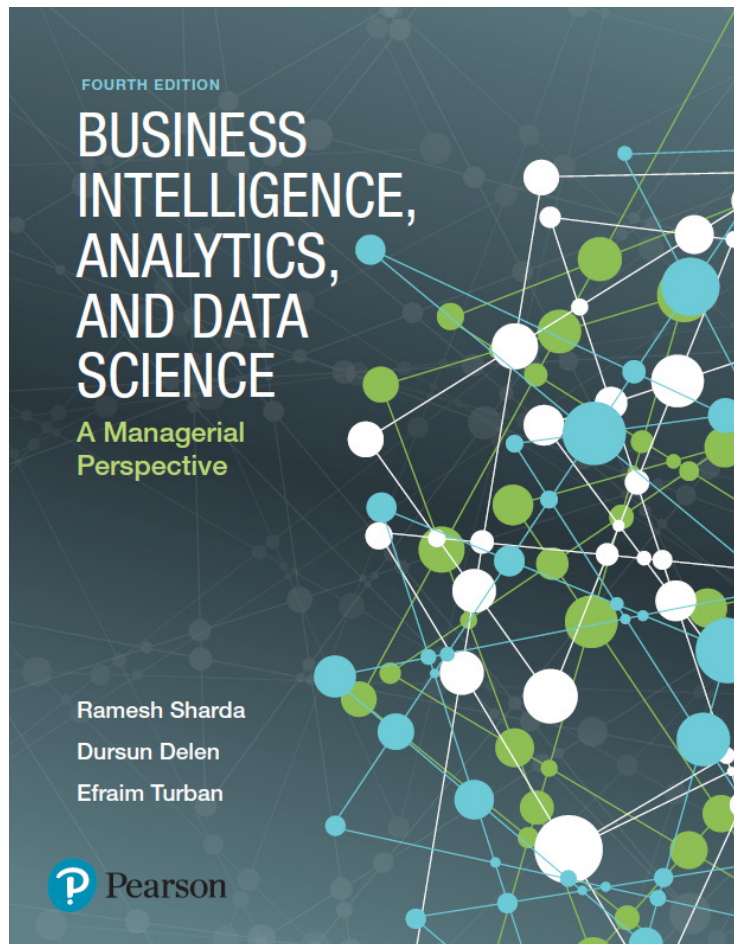


# Business Intelligence, Analytics, and Data Science: A Managerial Perspective

Fourth Edition



## Chapter 7 – Part B

Big Data Concepts  
and Tools

# Big Data and Data Warehousing

- What is the impact of Big Data on DW?
  - Big Data and RDBMS do not go nicely together
  - Will Hadoop replace data warehousing/RDBMS?
- Use Cases for Hadoop
  - Hadoop as the repository and refinery
  - Hadoop as the active archive
- Use Cases for Data Warehousing
  - Data warehouse performance
  - Integrating data that provides business value
  - Interactive BI tools

# Hadoop versus Data Warehouse

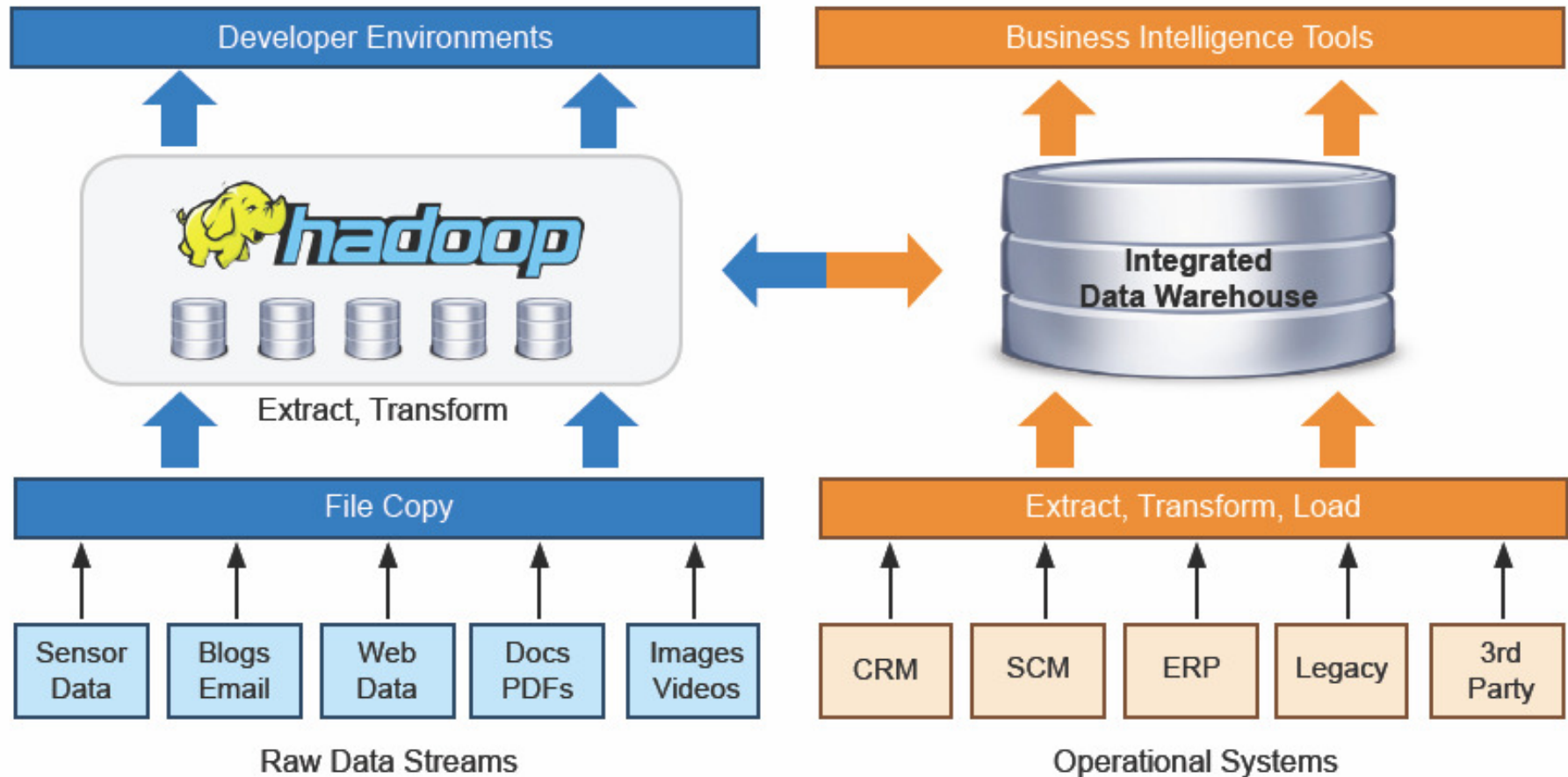
## When to Use Which Platform

TABLE 7.1 When to Use Which Platform—Hadoop versus DW		
Requirement	Data Warehouse	Hadoop
Low latency, interactive reports, and OLAP	<input checked="" type="checkbox"/>	
ANSI 2003 SQL compliance is required	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Preprocessing or exploration of raw unstructured data		<input checked="" type="checkbox"/>
Online archives alternative to tape		<input checked="" type="checkbox"/>
High-quality cleansed and consistent data	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
100s to 1,000s of concurrent users	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Discover unknown relationships in the data		<input checked="" type="checkbox"/>
Parallel complex process logic	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
CPU intense analysis	<input checked="" type="checkbox"/>	
System, users, and data governance		<input checked="" type="checkbox"/>
Many flexible programming languages running in parallel		<input checked="" type="checkbox"/>
Unrestricted, ungoverned sandbox explorations		<input checked="" type="checkbox"/>
Analysis of provisional data	<input checked="" type="checkbox"/>	
Extensive security and regulatory compliance	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

# Coexistence of Hadoop and DW

1. Use Hadoop for storing and archiving multi-structured data
2. Use Hadoop for filtering, transforming, and/or consolidating multi-structured data
3. Use Hadoop to analyze large volumes of multi-structured data and publish the analytical results
4. Use a relational DBMS that provides MapReduce capabilities as an investigative computing platform
5. Use a front-end query tool to access and analyze data

# Coexistence of Hadoop and DW



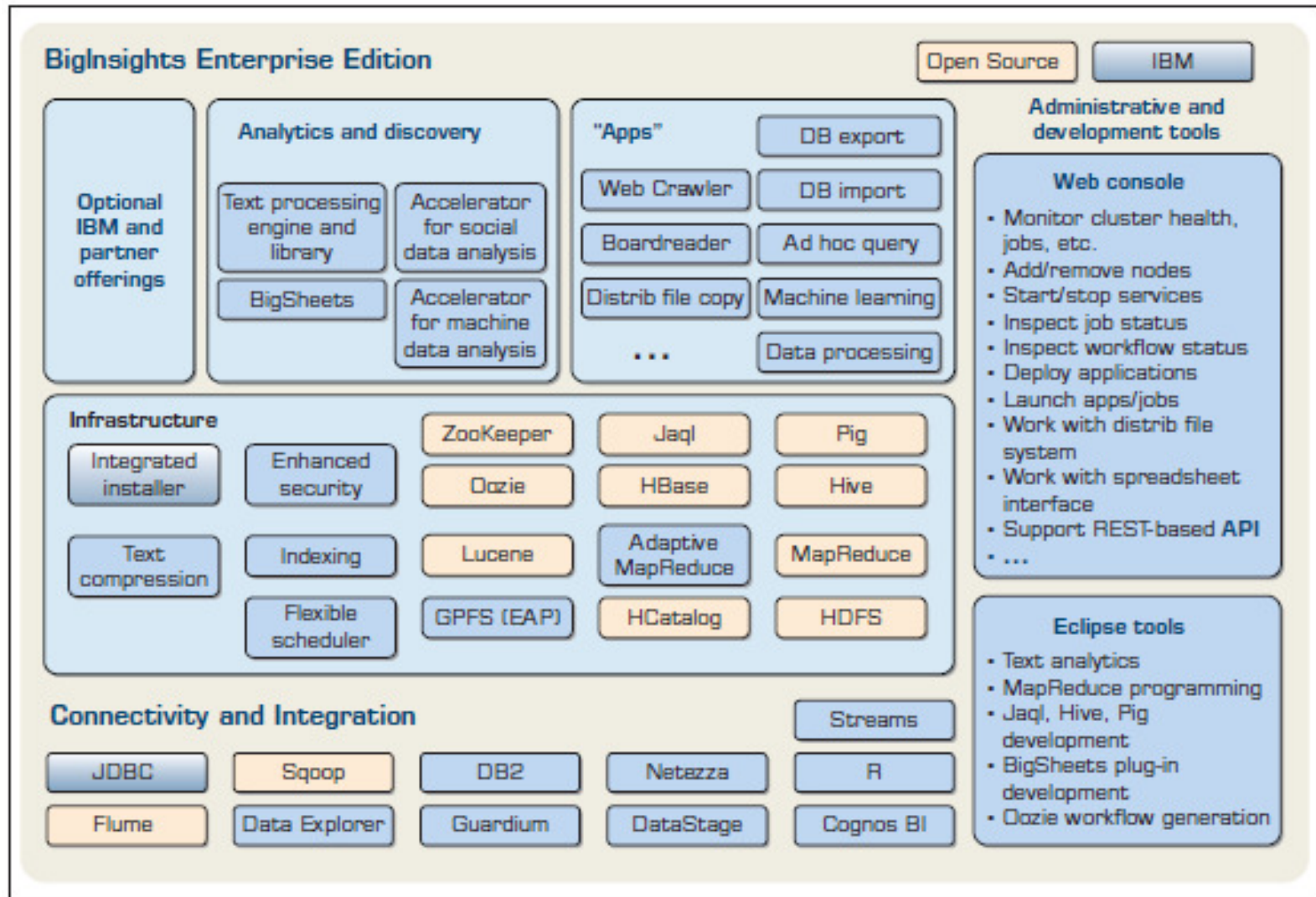
*Source: Teradata*

# Big Data Vendors

- Big Data vendor landscape is developing very rapidly
- A representative list would include
  - Cloudera - cloudera.com
  - MapR – mapr.com
  - Hortonworks - hortonworks.com
  - Also, IBM (Netezza, InfoSphere), Oracle (Exadata, Exalogic), Microsoft, Amazon, Google, ...

Software,  
Hardware,  
Service, ...

# IBM InfoSphere BigInsights



# Application Case 7.5

## Using Social Media for Nowcasting the Flu Activity

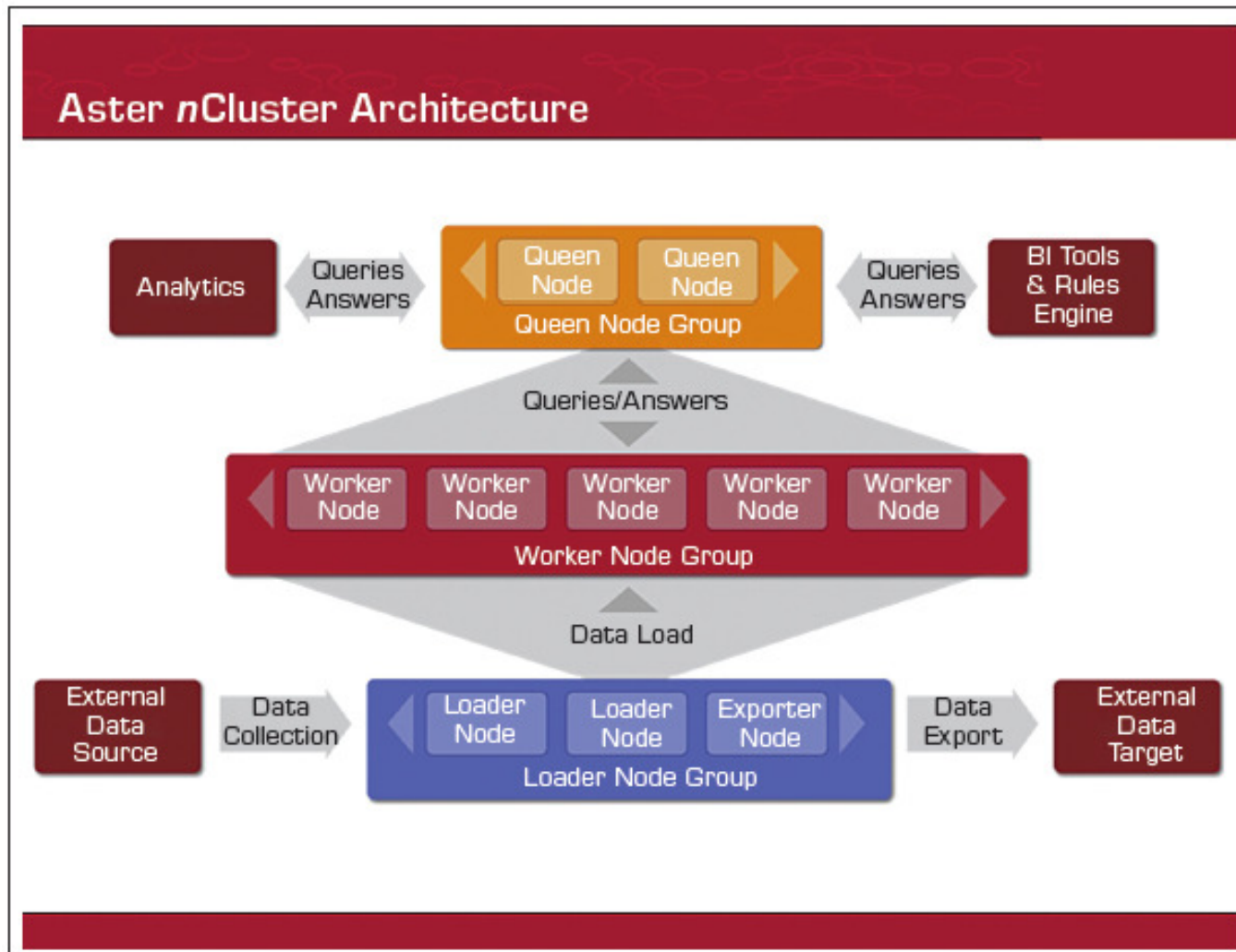
### Questions for Discussion

1. Why would social media be able to serve as an early predictor of flu outbreaks?
2. What other variables might help in predicting such outbreaks?
3. Why would this problem be a good problem to solve using Big Data technologies mentioned in this chapter?



# Big Data Platforms

## Teradata Aster



# Application Case 7.6

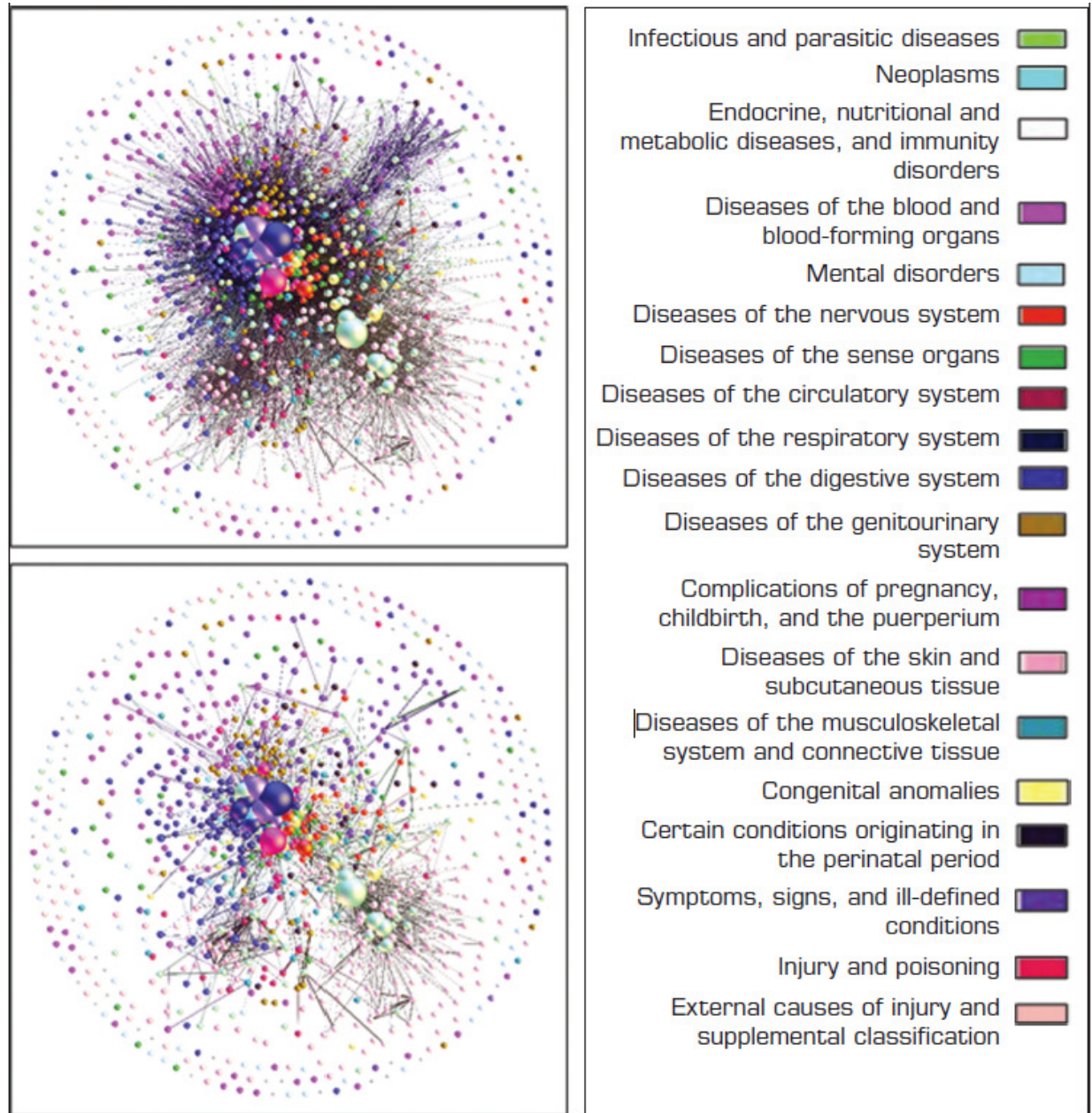
## Analyzing Disease Patterns from an Electronic Medical Records Data Warehouse

### Questions for Discussion

1. Why could comorbidity of diseases be different between rural and urban hospitals?
2. What is the issue about the huge difference between rural and urban patient encounters?
3. What are the main components of a network?
4. Where else can you apply the network approach?

# FIGURE 7.11

## Urban and Rural Comorbidity Networks



# Technology Insights 7.3

## How to Succeed with Big Data

1. Simplify
2. Coexist
3. Visualize
4. Empower
5. Integrate
6. Govern
7. Evangelize

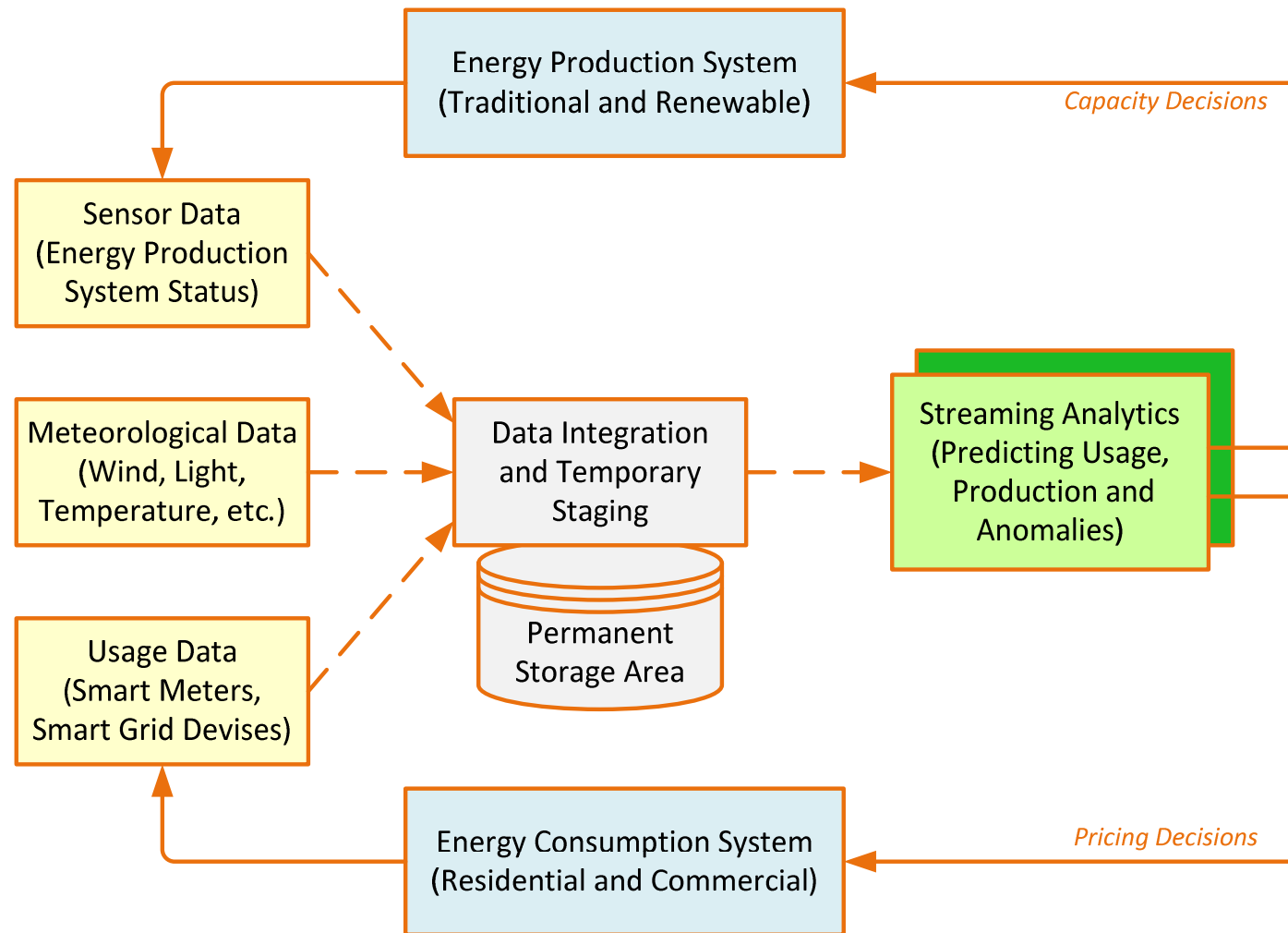


# Big Data And Stream Analytics

- Data-in-motion analytics and real-time data analytics
  - One of the Vs in Big Data = Velocity
- Analytic process of extracting actionable information from continuously flowing data
- Why Stream Analytics?
  - It may not be feasible to store the data, or lose its value
- Stream Analytics Versus Perpetual Analytics
- Critical Event Processing?

# Stream Analytics

## A Use Case in Energy Industry



# Stream Analytics Applications

- e-Commerce
- Telecommunication
- Law Enforcement and Cyber Security
- Power Industry
- Financial Services
- Health Services
- Government

# Application Case 7.7

## Salesforce Is Using Streaming Data to Enhance Customer Value

### Questions for Discussion

1. Are there areas in any industry where streaming data is irrelevant?
2. Besides customer retention, what are other benefits of using predictive analytics?