MIS2502:
Data Analytics
The Things You Can Do With Data
The Information Architecture of an Organization

Jaclyn Hansberry
jaclyn.hansberry@temple.edu

Why Data Analytics?

40%

- Of decisions by managers are made by using their “gut”

61%

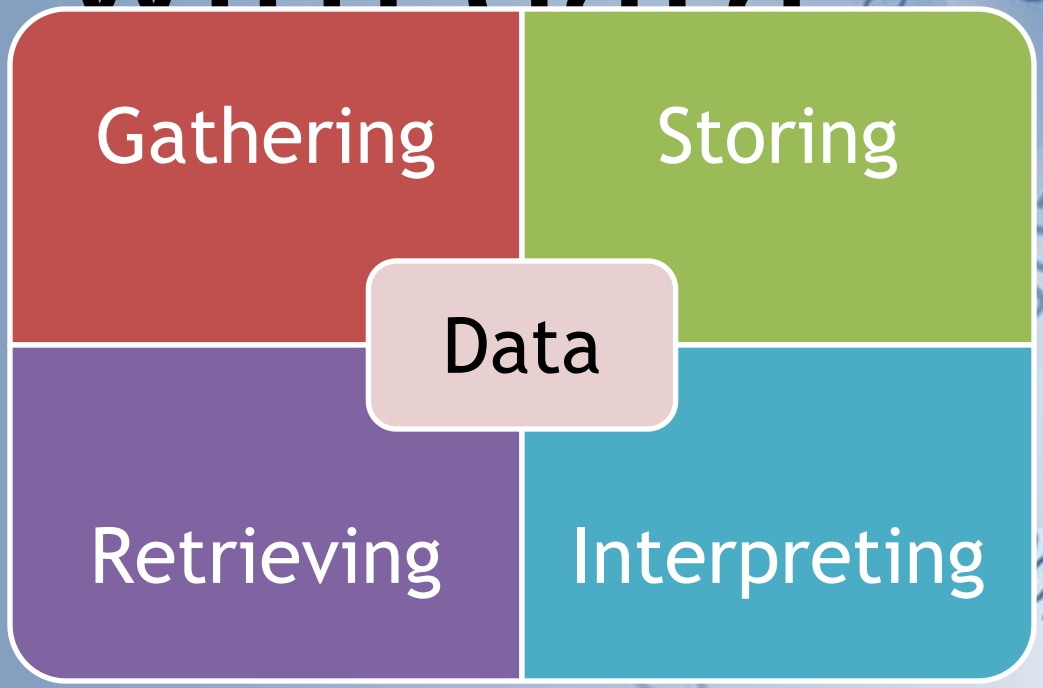
- Say this is because there is “no good data”

72%

- Want to increase their organization’s use of business intelligence

Source: http://advice.cio.com/thomas_wailgum/to_hell_with_business_intelligence_40_percent_of_execs_trust_gut

It all starts with data



Almost every business action requires at least one of these!

Data versus information

Data

Discrete, unorganized, raw facts

Information

The transformation of those facts into meaning

Examples of Data

Data	Information
• Quantity sold • Course enrollment • Customer name • Discount • Star rating	?

So then how do you turn data ***into*** information?

Example: Memphis Police

- Used historical crime data
- Place police where and when
  crime was likely to occur

http://www-03.ibm.com/press/us/en/pressrelease/32169.wss

Example: New York Mets

- Look at fan ticket purchase, social media, and mobile data
- Personalize communications and promotions
- Growth in corporate sales; ticket base now 14000+

http://www.sas.com/en_us/insights/articles/analytics/New-York-Mets-sign-

Two types of data

Transactional

- Captures data describing and event
- An exchange between actors
- Real-time

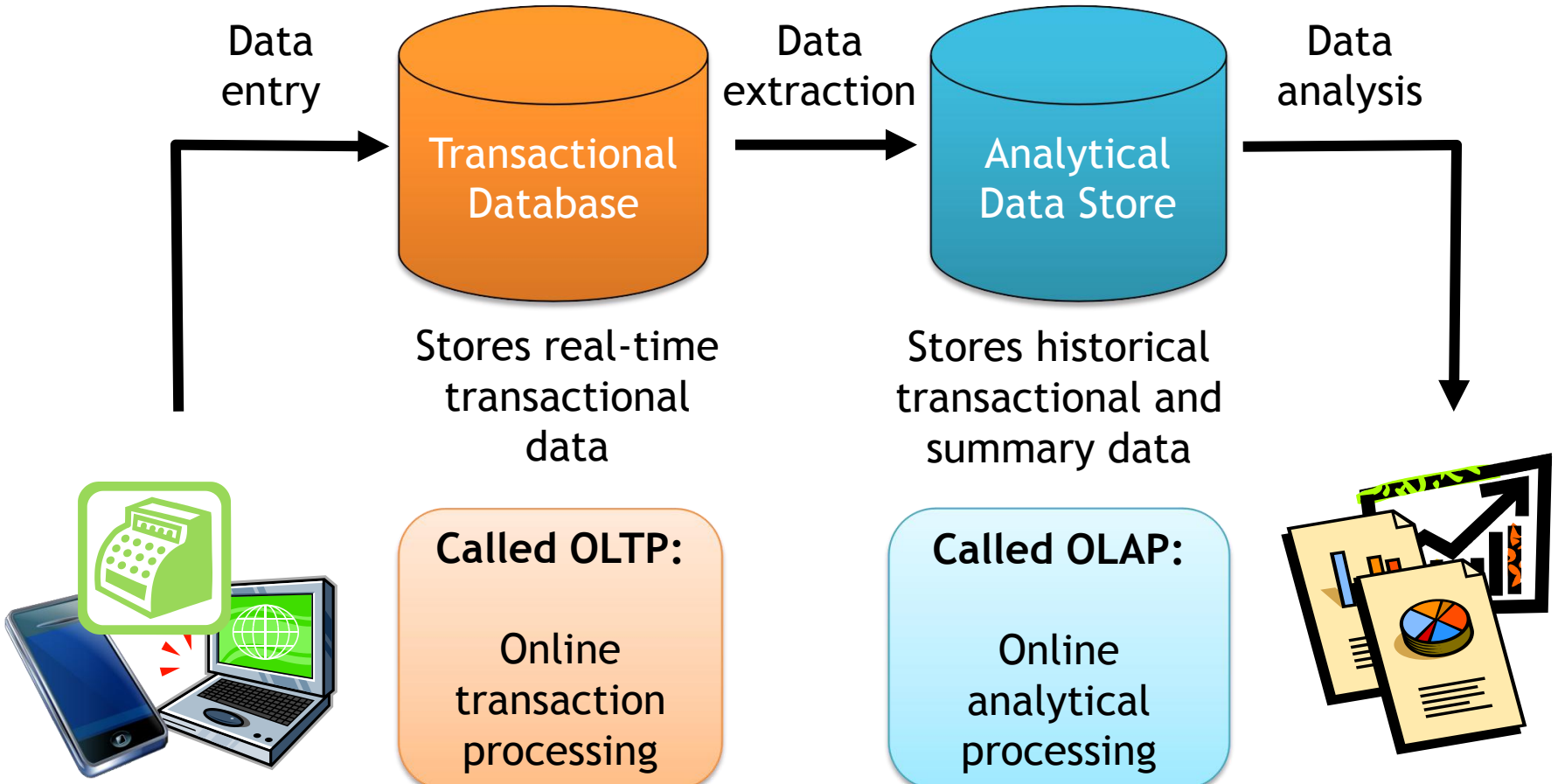
Analytical

- Captures data to support analysis and reporting
- An aggregated view of the business
- Historical

!

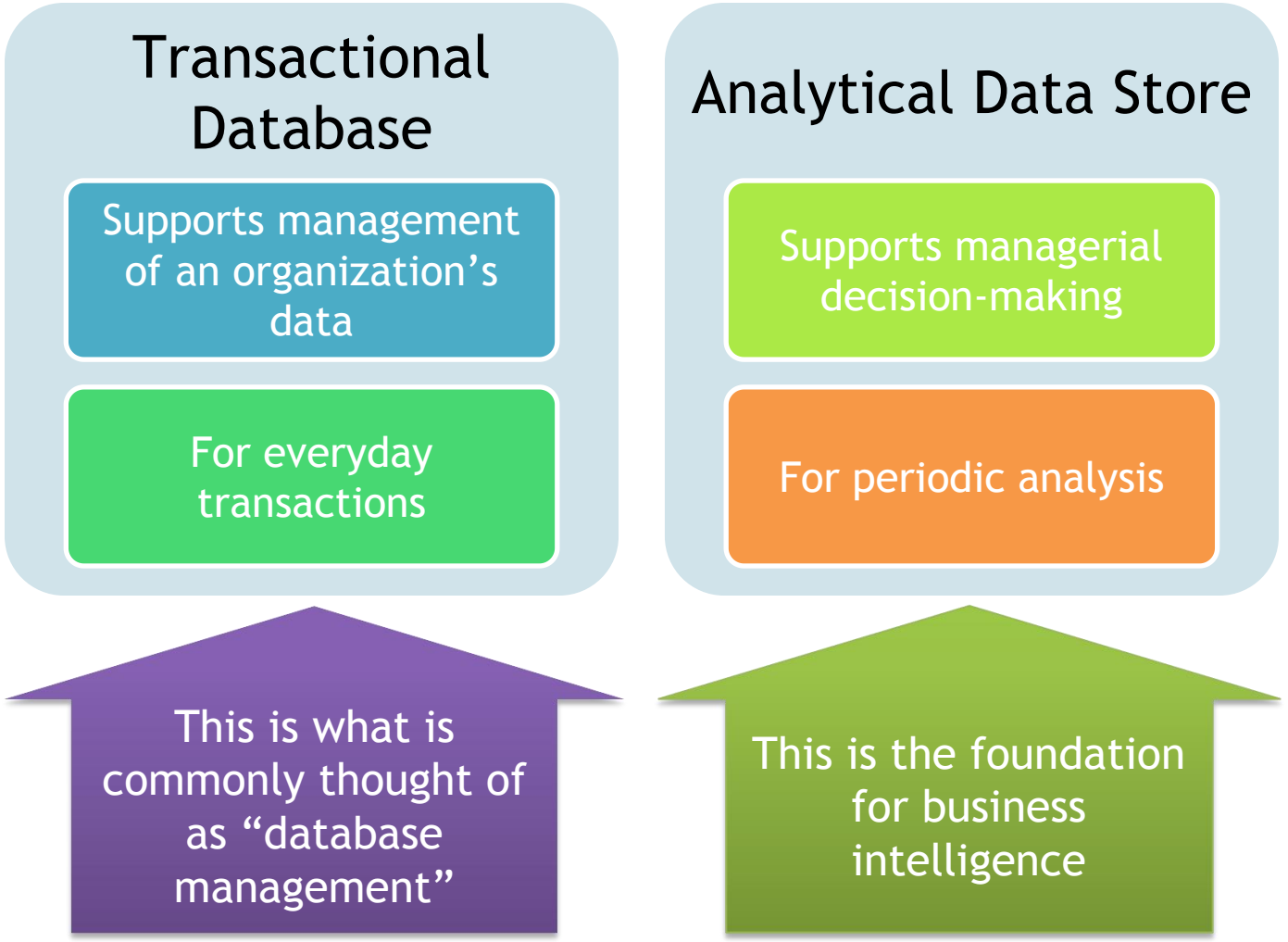
Explain the role of transactional and analytical data in the examples on the previous slides.

The Information Architecture of an Organization



! But this is changing rapidly….

Components of an information infrastructure



The Transactional Database

- Stores real-time, transactional data

In business, a transaction is the exchange of information, goods, or services.

For databases, a transaction is an action performed in a database management system.

Operational databases deal with both: they store information about business transactions using database transactions

- Examples of transactions
  - Purchase a product
  - Enroll in a course
  - Hire an employee
- Data is in real-time
  - Reflects current state
  - How things are "now"

The Relational Paradigm

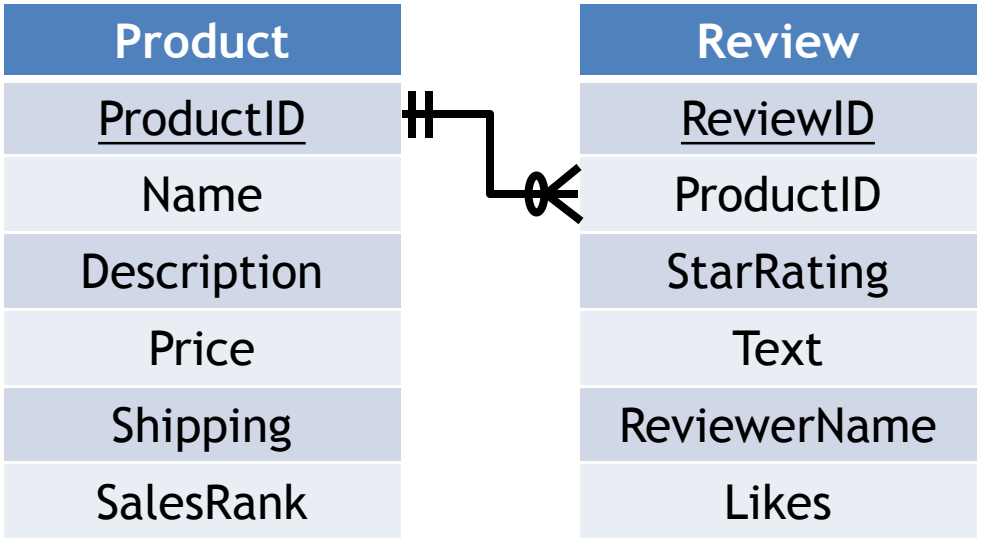
- How transactional data is collected and stored
- Primary Goal: Minimize redundancy
  - Reduce errors
  - Less space required

Which of these do you think is more important today ?

- Most database management systems are based on the relational paradigm
  - Oracle, Microsoft Access, SQL Server

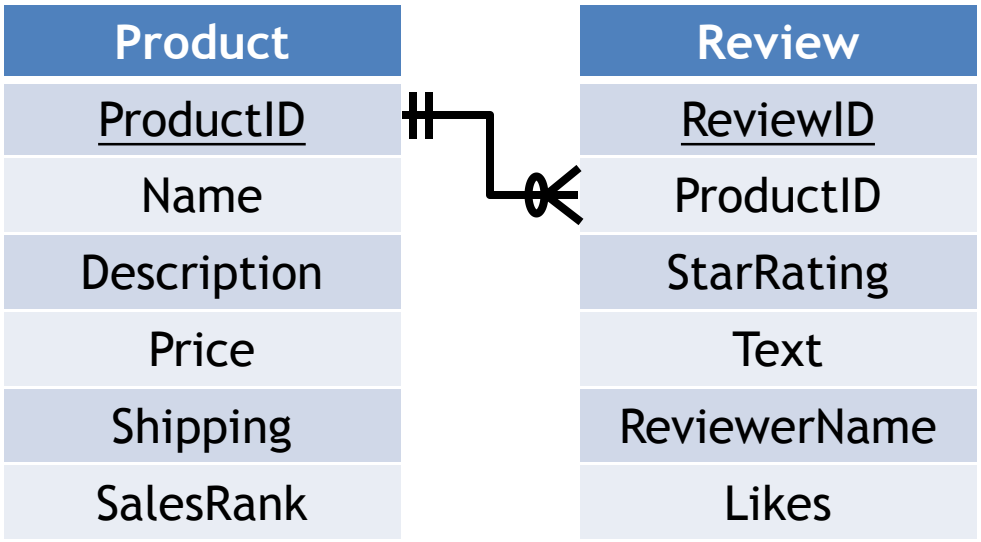
The Relational Database
Online Retailer Example

- A series of tables with logical associations between them
- The associations (relationships) allow the data to be combined



Why more than one table?

- Every review has an associated product
- Every product *can* have a review
- Products and reviews have a unique ID number
- Split the details off into separate tables



This is good because:

- Information is entered and stored once
- Minimizes redundancy

Analyzing transactional data

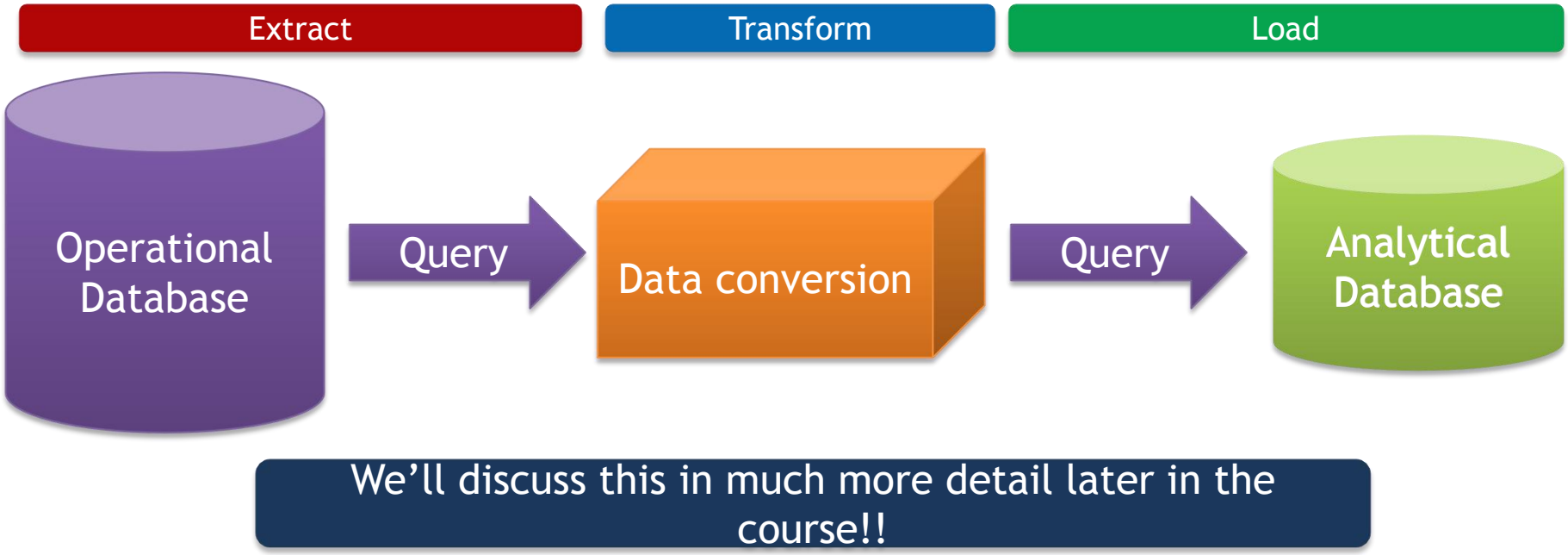
- Can be difficult to do from a relational database
- Having multiple tables is good for storage and data integrity, but bad for analysis
  - Tables must be “joined” together before analysis can be done
- The solution is the Analytical Data Store

Operational databases are optimized for storage efficiency, not retrieval

Analytical databases are optimized for retrieval and analysis, not storage efficiency and data integrity

The Analytical Data Store

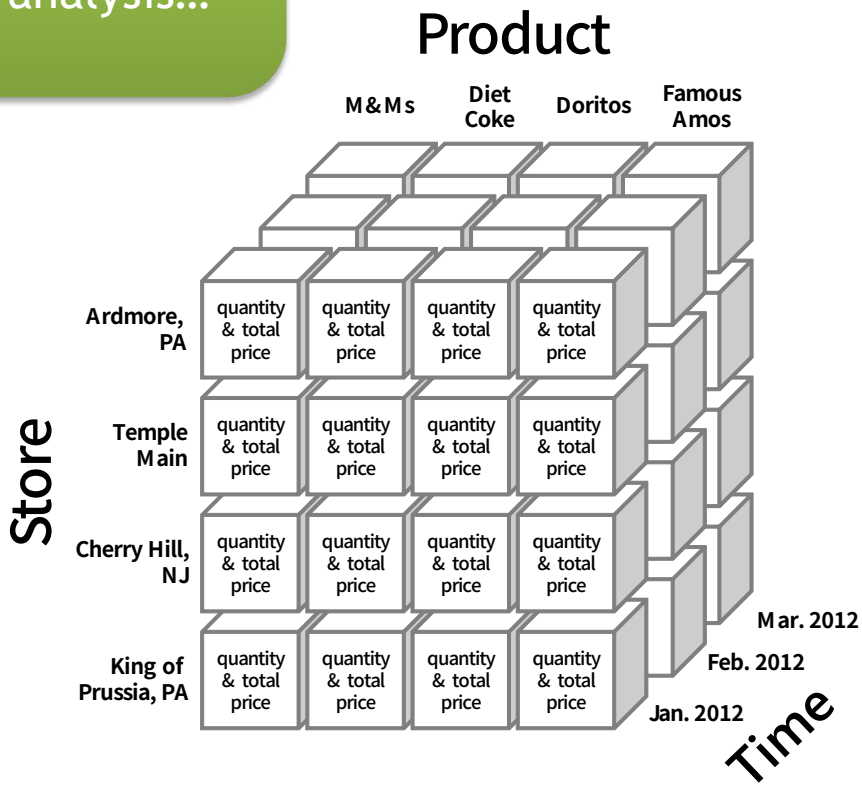
- Stores historical and summarized data
  - "Historical" means we keep **everything**
- Data is extracted from the operational database and reformatted for the analytical database



The Dimensional Paradigm

Data is stored like this around a business event…

…and can be summarized like this for analysis…



Dimensional Data and the Data Cube

…or it can be expanded in detail like this so that data mining (complex statistical analysis) can be done.

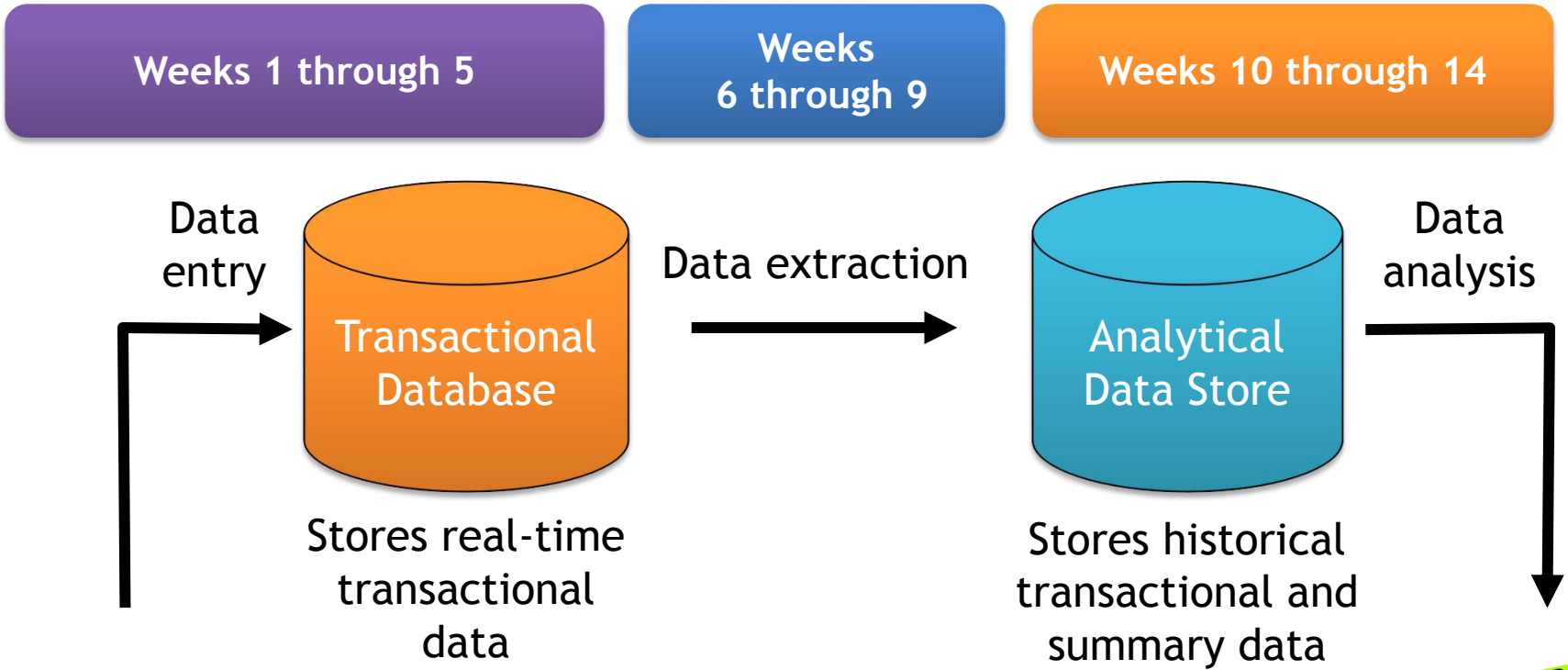
Sales ID	Qty. Sold	Total Price	Prod. ID	Prod. Name	Prod. Price	Prod. Weight	Store ID	Store Address	Store City	Store State	Store Type	Time ID	Day	Month	Year
1000															
1001															
1002															

Sales Fact | Product Dimension | Store Dimension | Time Dimension

Comparing Operational and Analytical Data Stores

Operational Data Store	Analytical Data Store
Based on Relational paradigm	Based on Dimensional paradigm
Storage of real-time transactional data	Storage of historical transactional data
Optimized for storage efficiency and data integrity	Optimized for data retrieval and summarization
Supports day-to-day operations	Supports periodic and on-demand analysis

The agenda for the course



Data interpretation, visualization, communication