Chapter 4: Data and Databases

- DataType describes what a sequence of bits represents
- Data = Bits (1/0) that represent Quantitative or Qualitative items
  - 1-bit Boolean 1/0 =True/False =On/Off =Yes/No =Checked/Unchecked
  - 8-bit Unsigned (0 to 255): $0000,0000_2 = 0_{10}$ 1 $111,1111_2 = 255_{10}$
  - 8-bit Signed (-127 to 128): $1000,0000_2 = -128_{10}$ $1111,1111_2 = -1_{10}$
  - 8-bit Text ASCII: $0100,0001_2$ ='A'$_{ASCII}$ $0111,1010_2$ ='z'$_{ASCII}$
  - 16-bit Unsigned (0 to 65,535): $1000,0000,0000,0000_2 = 32,768_{10}$
  - 16-bit Signed (32,767 to -32,768): $1000,0000,0000,0000_2 = -32,768_{10}$
  - 16-bit Unicode: 0011,0000,0100,0010 ='あ' 0101,1100,0111,0001 ='山'
  - 24-bit 3-Character String: 0100,0010,0110,1111,0110,0010 ="Bob"
  - 24-bit Unsigned (0 to 16,777,216): 0100,0010,0110,1111,0110,0010 = 4,353,890
  - 24-bit Color of one pixel: 0100,0010,0110,1111,0110,0010 = (Red, Green, Blue)
  - 32-bit Decimal (Floating Point)
    http://www.exploringbinary.com/floating-point-converter/
  - $0100,0000,0100,1000,1111,0101,1100,0011_2 = 3.14$

Processing Data into Information

- DataType provide a context for data as either
  - Quantitative data:
    - Integer
    - Floating Point
  - Qualitative data:
    - Characters
    - Text Strings
    - Boolean (True/False)
    - Images
    - Audio

- Data processing creates information from data
  - Custom Programs writen in Java, Python, Ruby, R
  - Database Management Systems

Database Management System

- *Database*
  - An organized collection of related data
  - All Data is described and associated with other data

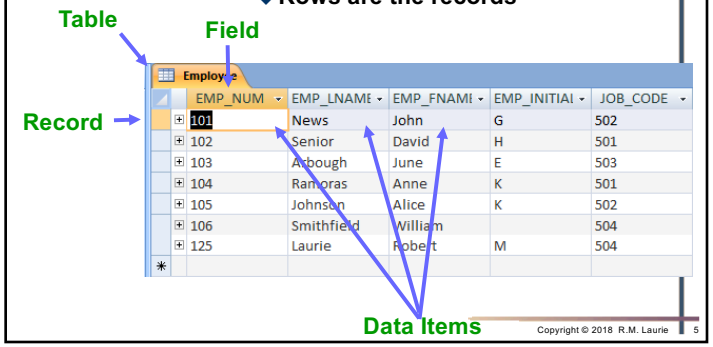
- *Database Management System* (*DBMS*)
  - Software that organizes data for fast & easy access
  - Desktop DBMS: Microsoft Access, LibreOffice Base
  - Enterprise DBMS: Oracle, Microsoft SQL Server, SAP
  - OpenSource WebServer DBMS: MySQL, PostgreSQL
- Phone books, file cabinets, and rolodex card files are non-computer versions of a database

Database Provides Information

- Information created from data
  - Timely relevant information key to decision making
  - Good decision making key to organization survival
- Database Management System (DBMS)
  - Manages database structure -- tables and relationships
  - Controls access to data – Security
  - Contains query language – SQL
  - All data in Database should be related
- Relational DBMS advantages
  - Integrated data (All items accessible)
  - Integrity (Accurate, up to date, no duplication)
  - Security Level Access
  - Easy Data Archive

A Database Table

- All data is organized into tables
  - Columns are the fields
  - Rows are the records



 5

Relational DB Model Terminology

- *Data Value* or *Data Items*
  - Contents of a field contained in a record
  - "Raw Facts" that can be recognized
- *Field* or *Attribute* or *Property* (Table Column)
  - Group of characters representing something with same data format
- *Record* or *Tuple* (Table Row)
  - Collection of related fields
- *Table* or *Entity*
  - Collection of related records and fields
  - Ordering of Columns and Rows is immaterial

 6

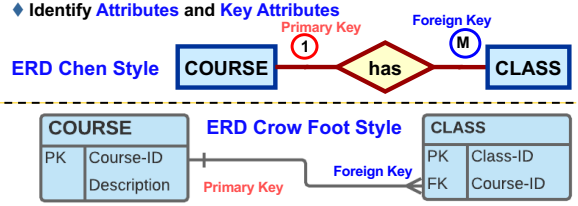
Field Name and Data Type

- Each Field must have a unique name in a table
  CustID LastName FirstName Address
- Fields may contain one of several data types:
  - **Character** = descriptive data (text).
  - **Numeric** = numbers used for calculation
  - **Date** = Month Day Year and/or time
  - **Logic** = T/F, Y/N, Checked/Unchecked
  - **Blob** = Images, Audio, Video
- Field width
  - maximum number of characters
  - Significant digits contained in the field

 7

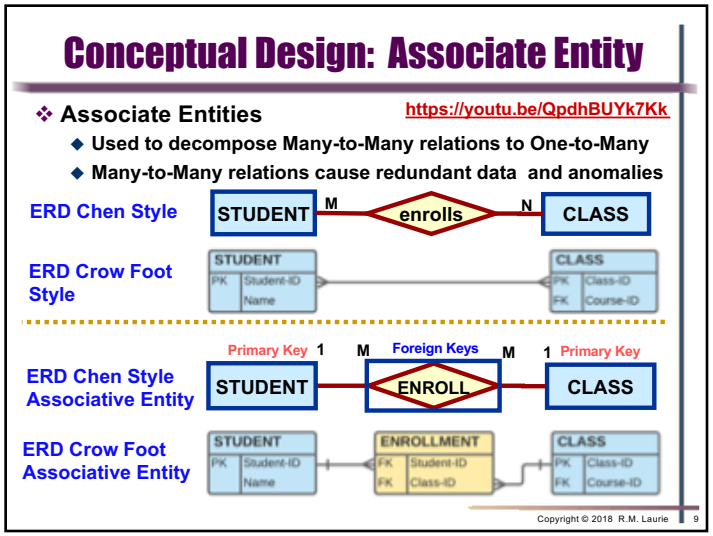
Conceptual Design

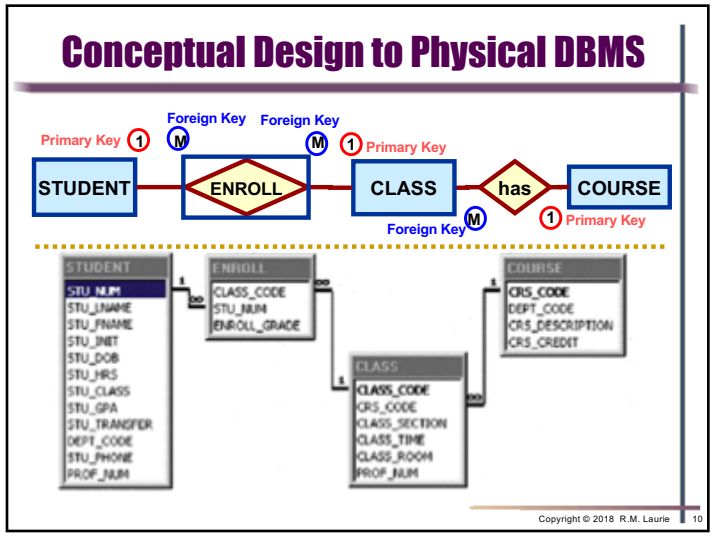
- Conceptual design https://youtu.be/FpJXQG7EIcE
  - Abstract model from Business Perspective
  - Entity Relationship Diagram (ERD) modeling
  - Define Relationships between Entities
    - Eliminate redundant database Entities
    - Identify Attributes and Key Attributes

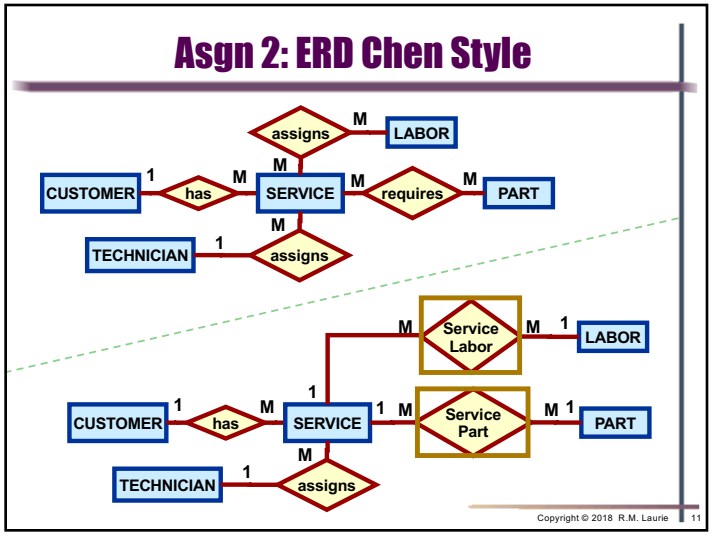


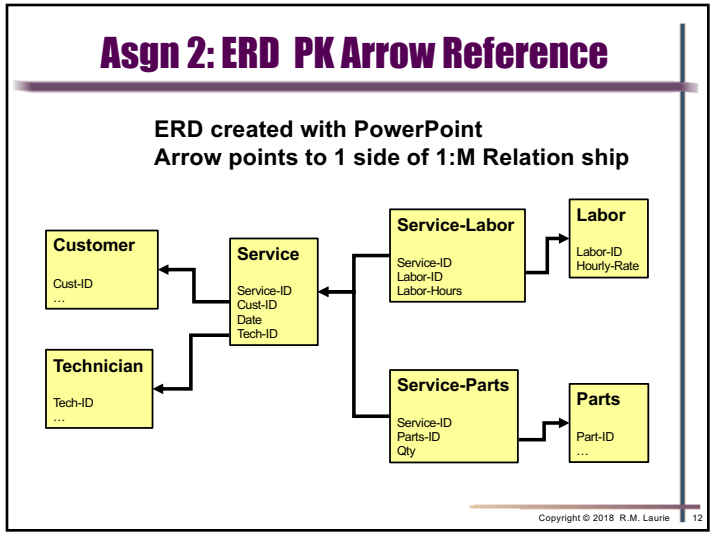
- Normalization
  - Minimize redundant data elements
  - Eliminate many-to-many relationships
  - This is CMIS320 with prerequisite <= CMIS102

 8



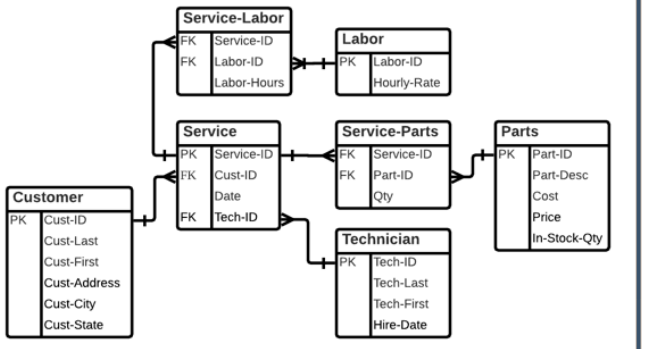






Asgn 2: ERD using Crow Foot Notation

Created using Blank ERD at LucidChart.com



https://youtu.be/-CuY5ADwn24

 13

MS Access Navigation

- ***Access Objects***
  Provides interface to database components
  - ***Tables***
    Containers for data
  - ***Forms***
    Input one record
  - ***Reports***
    Information output
  - ***Queries***
    Ask?

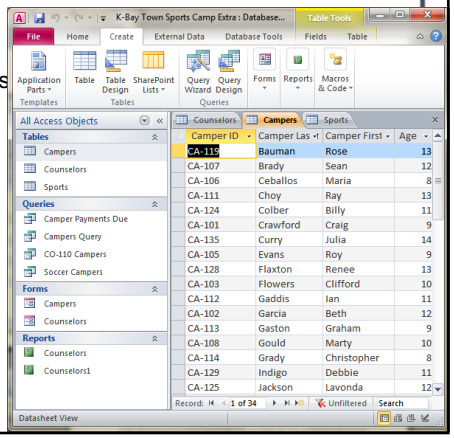


Table Design View

Table Name

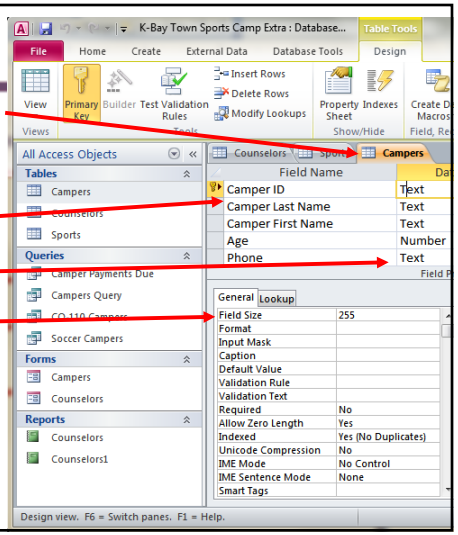
Design Field Structure

Field Name

Data Types

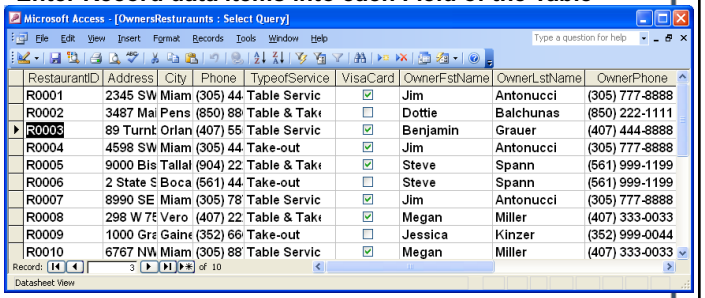
Field Width

Data Dictionary:
Contains data about each file in database and each field within those files



Database Table: Restaurants

Enter Record data items into each Field of the Table



- **Do you see any potential problems with this table?**
  - ***Data Redundancy leads to Data Inconsistencies***
  - ***Update Data Anomaly***
  - ***Deletion Data Anomaly***

 16

Normalization: Removing Data Redundancy

RestaurantID	Address	City	Phone	TypeofService	VisaCard	OwnerFstName	OwnerLstName	OwnerPhone
R0001	2345 SW	Miam	(305) 44	Table Servic	☑	Jim	Antonucci	(305) 777-8888
R0002	3487 Mai	Pens	(850) 88	Table & Take	☐	Dottie	Balchunas	(850) 222-1111
R0003	89 Turnb	Orlan	(407) 55	Table Servic	☑	Benjamin	Grauer	(407) 444-8888
R0004	4598 SW	Miam	(305) 44	Take-out	☑	Jim	Antonucci	(305) 777-8888
R0005	9000 Bis	Tallal	(904) 22	Table & Take	☑	Steve	Spann	(561) 999-1199
R0006	2 State S	Boca	(561) 44	Take-out	☐	Steve	Spann	(561) 999-1199
R0007	8990 SE	Miam	(305) 78	Table Servic	☑	Jim	Antonucci	(305) 777-8888
R0008	298 W 75	Vero	(407) 22	Table & Take	☑	Megan	Miller	(407) 333-0033
R0009	1000 Gra	Gaine	(352) 66	Take-out	☐	Jessica	Kinzer	(352) 999-0044
R0010	6767 NW	Miam	(305) 88	Table Servic	☑	Megan	Miller	(407) 333-0033

1. Remove any duplicate records
2. Determine Primary Key Fields: RestaurantID
3. *Normalize* to remove *non key data dependencies*

RestaurantID	Address	City	Phone	TypeofService	VisaCard	OwnerFstName	OwnerLstName	OwnerPhone

RestaurantID	Address	City	Phone	TypeofService	VisaCard	FranchiseeID

OwnerID	OwnerFstNa	OwnerLstNan	OwnerPhone

 17

Normalization: Making a Better Database

RestaurantID	Address	City	Phone	TypeofService	VisaCard	FranchiseeID
R0001	2345 SW 98 St	Miami	(305) 444-8787	Table Service	☑	F001
R0002	3487 Main High	Pensacola	(850) 886-5555	Table & Take	☐	F002
R0003	89 Turnberry Dri	Orlando	(407) 555-9999	Table Service	☑	F004
R0004	4598 SW 136 S	Miami	(305) 444-4444	Take-out	☑	F001
R0005	9000 Biscayne	Tallahassee	(904) 222-1111	Table & Take	☑	F003
R0006	2 State Street	Boca Raton	(561) 444-1100	Take-out	☐	F003
R0007	8990 SE 2 Aven	Miami	(305) 787-7889	Table Service	☑	F001
R0008	298 W 75 Terra	Vero Beach	(407) 222-9999	Table & Take	☑	F005
R0009	1000 Grand Ave	Gainesville	(352) 666-7788	Take-out	☐	F006
R0010	6767 NW 75 St	Miami	(305) 887-8877	Table Service	☑	F005

Primary Key

Foreign Key

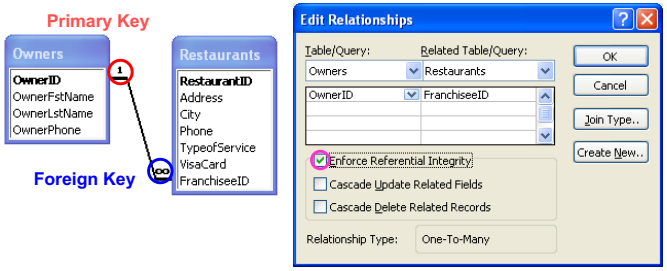
OwnerID	OwnerFstNa	OwnerLstNan	OwnerPhone
F001	Jim	Antonucci	(305) 777-8888
F002	Dottie	Balchunas	(850) 222-1111
F003	Steve	Spann	(561) 999-1199
F004	Benjamin	Grauer	(407) 444-8888
F005	Megan	Miller	(407) 333-0033
F006	Jessica	Kinzer	(352) 999-0044
F007	Carlos	Portu	(305) 787-8778

Data Redundancy Eliminated

 18

Enforcing Referential Integrity

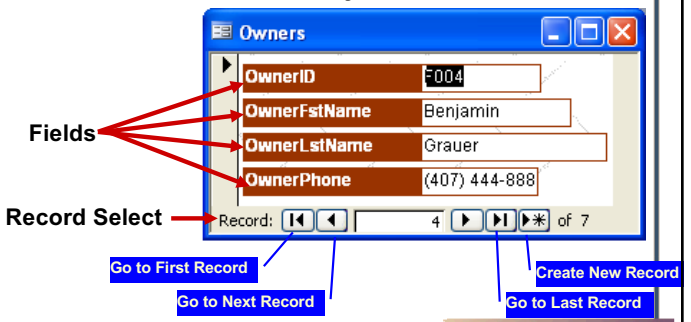
- Foreign key must match primary key values or be null value
- Impossible to delete row whose primary key has matching foreign key values in other table



 19

Database Form

- Forms allow the user to enter or view fields for one record at a time
- Forms can be attractively Formatted



 20

REPORTS: Information Output

You cannot enter data or edit data using reports.

Label Unbound Control

Textbox Bound Control

Owners

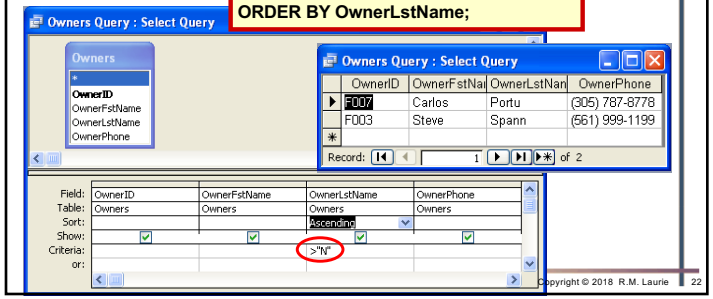
OwnerFst	OwnerLstN	OwnerP	Restau	Address	City	Phone	TypeofService	VisaC
Benjamin	Grauer	(407) 44						
			R0003	89 Turnberry	Orlando	(407) 555-9999	Table Service	☑
Dottie	Balchunas	(850) 22						
			R0002	3487 Main H	Pensacola	(850) 886-5555	Table & Take-o	☐
Jessica	Kinzer	(352) 99						
			R0009	1000 Grand	Gainesville	(352) 666-7788	Take-out	☐
Jim	Antonucci	(305) 77						
			R0001	2345 SW 98	Miami	(305) 444-8787	Table Service	☑
			R0004	4598 SW 13	Miami	(305) 444-4444	Take-out	☑
			R0007	8990 SE 2 A	Miami	(305) 787-7889	Table Service	☑
Megan	Miller	(407) 33						
			R0008	298 W 75 Te	Vero Beach	(407) 222-9999	Table & Take-o	☑
			R0010	6767 NW 75	Miami	(305) 887-8877	Table Service	☑
Steve	Spann	(561) 99						
			R0005	9000 Biscay	Tallahassee	(904) 222-1111	Table & Take-o	☑
			R0006	2 State Stre	Boca Raton	(561) 444-1100	Take-out	☐

 21

Query-by-Example Design View and SQL

- Select Fields of each Table that will be Queried
- Set Sort and Criteria to Query

```
SELECT *
FROM Owners
WHERE OwnerLstName>"N"
ORDER BY OwnerLstName;
```

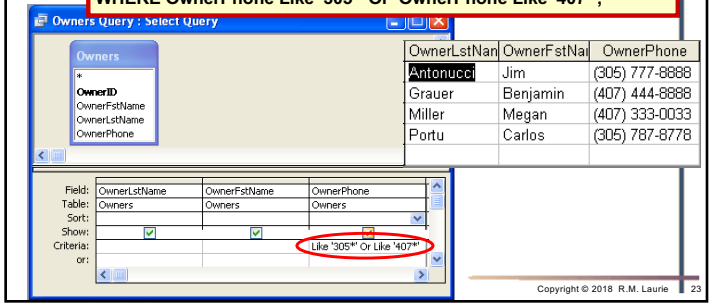


 22

Query-by-Example Design View and SQL

- Select Fields of each Table that will be Queried
- Set Sort and Criteria to Query

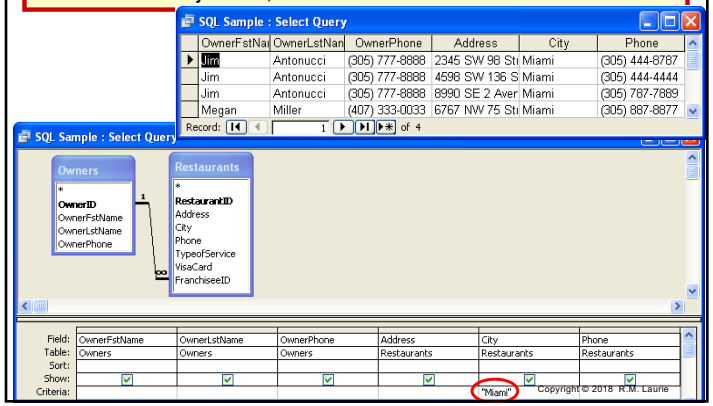
```
SELECT OwnerLstName, OwnerFstName, OwnerPhone
FROM Owners
WHERE OwnerPhone Like '305*' Or  OwnerPhone Like '407*';
```



 23

Query-by-Example Design View and SQL

```
SELECT Owners.OwnerFstName, Owners.OwnerLstName, Owners.OwnerPhone,
Restaurants.Address, Restaurants.City, Restaurants.Phone
FROM Owners INNER JOIN Restaurants ON Owners.OwnerID = Restaurants.FranchiseeID
WHERE Restaurants.City="Miami";
```

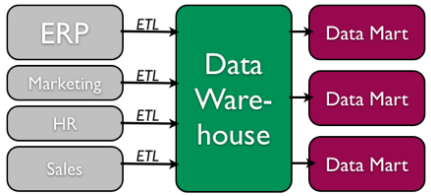


Enterprise Databases

- Large Databases
  - Commercial: Oracle, Microsoft, IBM, and Amazon
  - Open-source alternatives: MySQL and PostgreSQL
- NoSQL (not only SQL) is "Big Data" alternative
  - Relational database model does not scale well
  - NoSQL database works with data in a looser way
  - More easily scaled on multiple servers worldwide
    - Google now offers the App Engine Datastore
    - Amazon DynamoDB
- Big Data
  - Massively large data sets that are analysed
  - Amazon processes millions of customer transactions per hour

Data Warehouses

- Data Warehouses are huge databases that store and manage data required to analyze historical and current transactions
  - Data access but not alterable (Read Only)
  - Data mart is subset of data warehouse often associated with a firm's functional unit
  - ETL = Extraction-Transformation-Load data conversion



Data Mining

- Process of analyzing data to make decisions
  - Find trends, patterns, and associations
  - Automating using Big Data from Data Warehouse
  - Test hypothesis like automatic market trading
  - Privacy concerns especially Artificial Intelligence
- Business Intelligence and Analytics
  - Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make queries to support better business decisions
  - Analysis to obtain a competitive advantage
  - Text Mining use email and other documents
  - Web Mining use World Wide Web sources

Videos to View

1. ERD Systems Analysis
   https://youtu.be/FpJXQG7EIcE
2. LucidChart.com ERD Tutorial - Part 1
   https://youtu.be/QpdhBUYk7Kk
3. LucidChart.com ERD Tutorial - Part 2
   https://youtu.be/-CuY5ADwn24
4. Real Data Warehouse?
   https://youtu.be/y5-3Pjbk8Zk
5. Benefits of a Data Warehouse
   https://youtu.be/KGHbY_Sales

