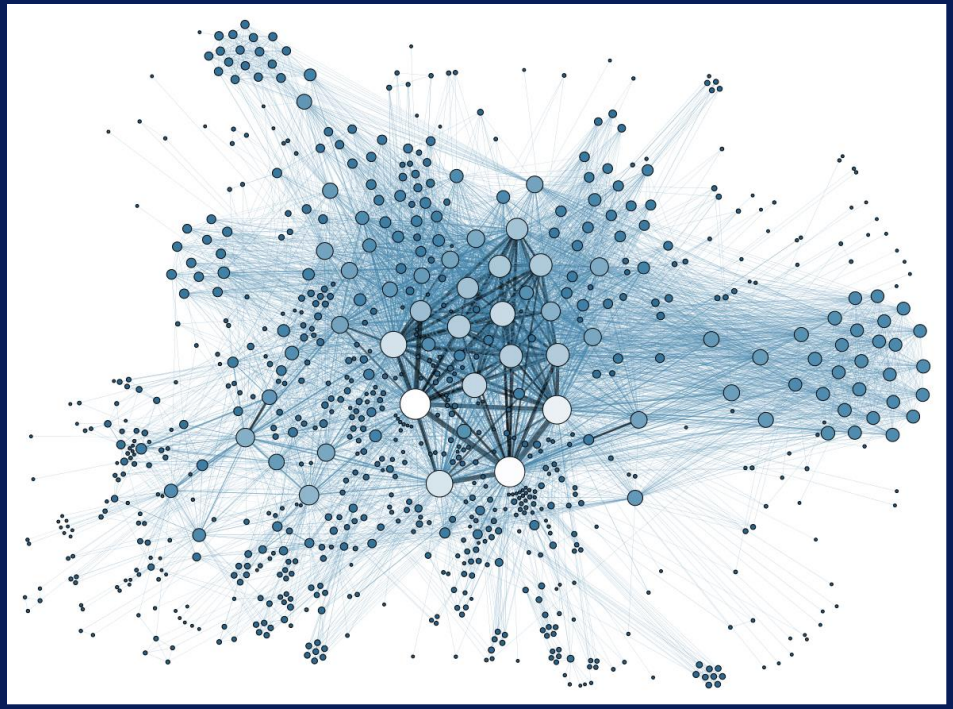
Summary of Big Data Modeling and Management



After this video you will be able to..

- Recall why big data modeling and management is essential in preparing to gain insights from your data
- Summarize different kinds of data models
- Describe streaming data and the different challenges it presents
- Explain the differences between a DBMS and a BDMS

Big Data Modeling and Management

- Data modeling tells you
  - How your data is structured
  - What operations can be done on the data
  - What constraints apply to the data
- Database Management Systems
  - Typically handle many low-level details of data storage, manipulation, retrieval, transactional updates, failure and security
  - Relieves a user to focus on higher level operations like querying and analysis

Different Data Models

- Relational Data
  - Where data look like tables
- Semi-structured Data
  - Document data, XML and JSON
- Graph Data
  - Social Networks, email networks
- Text Data
  - Articles, reports

Streaming Data

- An infinite flow of data coming from a data source
  - Sensor data from instruments
  - Stock price data
- Data rates vary – can be too fast and too large to store
- Often processed in memory
- May need to be processed immediately
  - Inform whenever 3 tech stocks go up by 3% within a 30 second span
  - Used for event detection and prediction

DBMS and BDMS

- BDMS
  - Designed for parallel and distributed processing
    - Data-partitioned parallelism
  - May not always guarantee consistency for every update
    - More likely to guarantee eventual consistency
  - Often built-on Hadoop
    - Offer Map-reduce style computation
    - Utilizes replication natively offered by HDFS