Database application

- Big data application
  - Connecting Dots — From small to big
  - Discovering specifics — From big to small — How do we find outliers, predict trends, provide summaries, and give explanations from a dataset.
  - Inferencing — Knowing unknown — Everything is related to everything else
  - Key values — The smallest unit of big data
    - Every key-value is unique with timestamp
    - Key values are associated to each other
    - All key values can be drawn as a
- Network science
  - The curse of dimensionality — Adding extra dimension to a data space will exponentially increase the volume of data space
  - Scale-free Network — the characteristic of network are independent of the size of network
  - Different type of networks
  - Computing issues of complex networks
- Explaining outliers in aggregate queries
- Effective storage of big data
  - Column-based VS Row-based data database stroages
  - Applicability of row and column storages
  - Compare and contrast of row and column storages
    - Row storage
      - Efficient when many columns of a single row are required at the same time
      - Well-suited for OLTP-liked workloads which are more heavily loaded with interactive transactions
    - Column storage
      - Efficient on aggregation operation over many row but only for a smaller subset of all columns of data
      - Efficient when inserting new values of a column for all rows at once
      - Well-suited for OLAP-like workloads
  - NoSQL(not only SQL)
    - Technology
      - Cloud Platform — A viable alternative to relational databases operating on cluster servers
      - No schema — Different types of data is collected, stored, accessed without a schema
      - Data fusion — A data integration technique for multi-source data
      - Flexible access — A query model accessing data without using traditional SQL
    - Property
      - To add/delete/query massive arrays and still allow for persistence and fault tolerance
      - To store objects using key-values
      - To implement large data query on MapReduce framework
    - System
      - CouchDB — A document-oriented database that can be queried and indexed in a MapReduce fashion using JavaScript.
      - MongoDB — A scalable, high-performance, open source, document-oriented database system
      - Hadoop — Hadoop develops open-source software for reliable, scalable, distributed computing
      - HBase — Hadoop database system supports random, real-time, read/write access to big data.