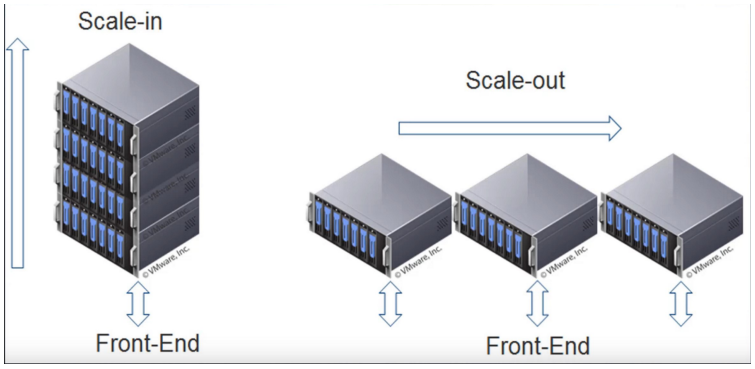
DAT229x Introduction to Big Data

Friday, 8 June 2018 1:14 PM

Lab 3. Introduction to NoSQL Databases

Lab 3.1 Overview of NoSQL Data Stores

- Relational databases are great for business applications that need to force relational constraints and manage data and highly structured formats, however many applications need more flexibility.
- A non-relational database (no tables), flexible used for big data & real-time web apps, multiple types of NoSQL databases to handle unstructured data
- Scaling-scale out / horizontal scaling

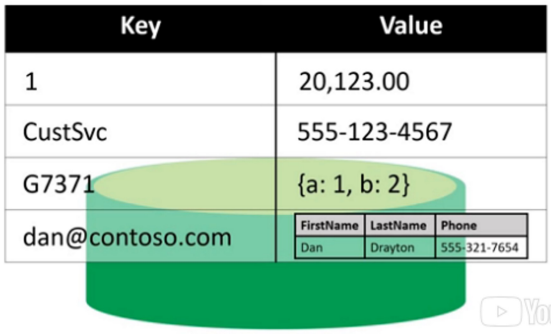


Horizontal scaling (scale-out) ~ NoSQL database

Scale-in~ Relational database, needs add components/storage/memory

Lab 3.2 Key-value Data Stores

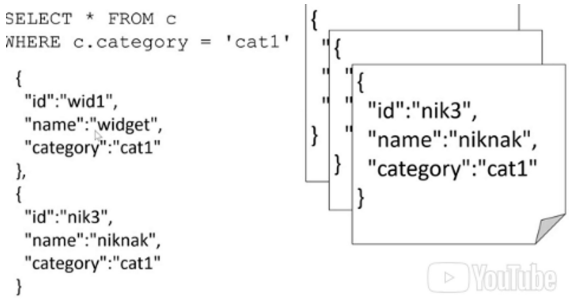
- One of the simplest and flexible ways to store disparate items of data
- Each data item is represented by a unique key and an associated value (any kind of)



- Approaches e.g., Cassandra, Apache Hbase, Azure table storage

Lab 3.3 Document Data Stores

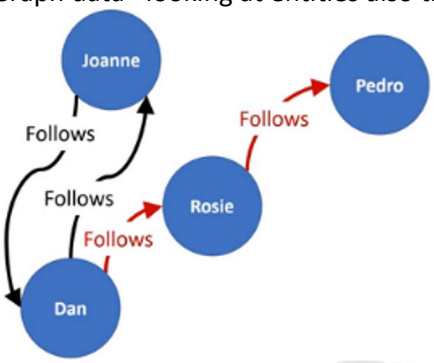
- JSON documents ~ query attributes ~ return results as JSON that can be desrialized into objects the application code can operate on efficiently



- Approaches e.g., MongoDB, Azure Document DB

Lab 3.4 Graph Data Stores

- Graph data~ looking at entities also the relationships between them



Traverse the graph by using edges to navigate between vertices and understand the relationships between the entities.

- Approaches e.g., Azure Cosmos DB, Neo4j
- NoSQL storage with Cosmos DB in Azure, see the Cosmos DB documentation at https://docs.microsoft.com/en-us/azure/cosmos-db/.