

FIT 5137 Week 6

Part a

Breakout Room 1 - Q1

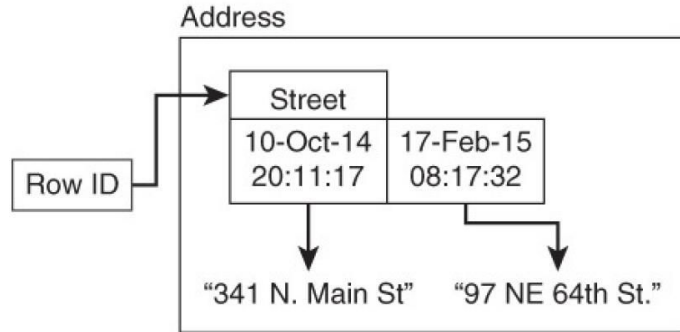
What is Column-Oriented DBMS?

- A database which use column to store the data rather than row
- Appropriate to distribute and high read/write requirement
- Peer-to-peer architecture
- Each node own the same states, which means that each node can read/write rather than only the master can read in the master-slave architecture
- Primary key have two part: partition key and cluster key
- Transparency to each data: we can trace their position via the hash ID
- It Should avoids the hotspotting in partition key

Breakout Room 1 - Q2

What are the three components of data values in Column-Oriented Databases?

- A column name
- A time stamp or version stamp
- Row id



Breakout Room 1 - Q3

Why are timestamps used in Column-Oriented Databases?

- Timestamps are used each time the column value is updated, it is mainly used for version control for the database
- Control the version of value
- Assist to trace previous data which may store in the log
- Satisfy the eventual consistency

Breakout Room 1 - Q4

Identify one similarity between column-oriented databases and document databases.

- Both of them are NoSQL
- Both of them are schemaless
- Both of them are identified by a unique identifier
- Both of them are self-contained

Breakout Room 2 - Q5

Identify one similarity between column-oriented databases and relational databases.

Both columnar and relational database use traditional query language (SQL- like language)

They store same data but in different ways (store by row/ store by column)

Breakout Room 2 - Q6

Describe the essential characteristics of a peer-to-peer architecture.

- Internode communication uses gossip protocol, all nodes can receive actions like update without master coordinator.
- All nodes in a cluster can do reads and writes.
- According to add servers or remove serves can simply release scaling up and scaling down.
- One of nodes breakdown, won't result in system breakdown.

Breakout Room 2 - Q7

Why does Cassandra use a gossip protocol to exchange server status information?

It allows all nodes to quickly receive updates without the need for a master coordinator.

- Gossip protocol is peer to peer protocol that efficient, lightweight, reliable-nodal broadcast protocol for diffusing data.
- Gossip protocol is decentralized, fault tolerant communication protocol.
- All characteristics are suitable for Cassandra work mechanism. So that Cassandra use Gossip protocol to do peer discovery and metadata propagation.

Breakout Room 2 - Q8

When would you use a column-oriented database instead of another type of NoSQL database?

purpose of a columnar database is to efficiently write and read data to and from hard disk storage in order to speed up the time it takes to return a query

When there are huge of data and update frequently, like database of social media (Facebook, Ins), billions of people post Anytime and anywhere.

Meanwhile column-oriented database can be used in data warehouse where businesses send massive amounts of data from multiple sources for BI analysis.

Breakout Room 3 - Q9

How do columns in column-oriented databases differ from columns in relational databases?

Rows are composed of several column families. Each family consists of a set of related columns. A data value is indexed by a row, a column name, and a time stamp.

Relational databases can be both row and column oriented.

In a column-oriented database, all values of a column are placed together on the disk.

Breakout Room 3 - Q10

When should columns be grouped together in a column family?

When Data values are indexed by row identifier, column name, and time stamp.

A group of columns that are logically related are called a super column. All of the super columns are grouped together to create a column family. A column family is conceptually similar to a table in the relational model.

Breakout Room 3 - Q8

When would you use a column-oriented database instead of another type of NoSQL database?

When we deal with Column with many null values, known as sparse data, can be dealt with more efficiently with column-oriented database, without wasting storage capacity for the empty cells.

Breakout Room 3 - Q7

Why does Cassandra use a gossip protocol to exchange server status information?

It allows all nodes to quickly receive updates without the need for a master coordinator.

Ensures high availability and fault tolerance of the data

- Sharing information about the state of servers in the cluster
- Ensuring nodes have the latest version of data
- Ensuring write data is stored when the server that should receive the write is unavailable

Breakout Room 4 - Q6

Describe the essential characteristics of a peer-to-peer architecture.

- Peer to peer networks are usually formed by groups of a dozen or less computers. These computers all store their data using individual security but also share data with all the other nodes.
- The nodes in peer to peer networks both use resources and provide resources. So, if the nodes increase, then the resource sharing capacity of the peer to peer network increases. This is different than client server networks where the server gets overwhelmed if the nodes increase.
- Since nodes in peer to peer networks act as both clients and servers, it is difficult to provide adequate security for the nodes. This can lead to denial of service attacks.
- Most modern operating systems such as Windows and Mac OS contain software to implement peer to peer networks.

Breakout Room 4 - Q5

Identify one similarity between column-oriented databases and relational databases.

- Column-oriented databases store data in columns, whereas relational databases can be both row and column oriented
- Both have a form of indexing?

Breakout Room 4 - Q4

Identify one similarity between column-oriented databases and document databases.

- Both use key-value pairs at its core
- Higher scalability
- Use distributed computing

Breakout Room 4 - Q3

Why are timestamps used in Column-Oriented Databases?

- Identifies when the data was last modified?
- Differentiate the valid content from stale ones
- Provides a history of updates