

Assignments 14

Column-Family Databases

Column Store for Relational Database

Given is our relational database lesson:

1. For what use case(s) would you store the content as rows, for what use case(s) would you store the content as columns?
2. Write down 2 queries for our lesson db that are efficiently executed within a row store.
3. Write down 2 queries for our lesson db that are efficiently executed within a column store.

Column Store for Relational Database

- Which column(s) of the table lesson in the lesson database would be suitable for compression in a column store?
- Take the first 6 rows of the table lesson. (Populate if not yet done.)
- Create bitmap and run-length encoding for the column and fill the index for the first 6 rows of your table.

Create the table (column family) for Q3

Q3: Return a teacher list by given, specific postcode
= Return the teachers that live at a given, specific postcode

- What is the primary key?
- Is it a composite or an atomic key?
- what is the partition key?
- Do you have a clustering key?
- Write the Cassandra create command for table for Q3

Populate / query the table

Populate the table – only in writing - with a teachers' list for two different postalcodes.

Are the following queries possible / allowed? How are they executed? Are they efficient or not?

- `select * from teachers_by_postalcode where postalcode = "xxxx"`
- `select * from teachers_by_postalcode where postalcode between "xxxx" and "zzzz"`
- `select * from teachers_by_postalcode where t_email = "xxxxx"`
- `select * from teachers_by_postalcode`
- `select * from teachers_by_postalcode where postalcode = "xxxx" and t_email in {"xxxx", "yyyy", "www", "zzzz"}`

summer@kiu.edu.ge

fname: sally: timestamp
lname:summer: timestamp
DoB: 1990-05-05: timestamp
...
...
subjects [EN, GE,DE, MA]:
timestamp

s_email: s1@kiu.edu.ge,
t_fname: sally: timestamp
t_lname:summer: timestamp
s_fname: giorgi: timestamp
...
s_subjects: [EN,MA] timestamp

s_email: s5@kiu.edu.ge,
t_fname: sally: timestamp
t_lname:summer: timestamp
s_fname: ana: timestamp
...
s_subjects: [EN,DE] : timestamp

What are the timestamps used for?

Partition key

Info

Students

summer@kiu.edu.ge

fname: sally,
lname:summer,
DoB: 1990-05-05,
...
...
subjects [EN, GE,DE, MA]

s_email: s1@kiu.edu.ge,
t_fname: sally,
t_lname:summer,
s_fname: giorgi,
s_lname: daravadze,
s_phone: 599-123-456,
s_subjects: [EN,MA]

doe@kiu.edu.ge

fname: john
lname:doe
DoB: 1999-12-02
...
...
subjects [CS, PH, MA]

s_email: s2@kiu.edu.ge,
t_fname: john,
t_lname: doe,
s_fname: mariam,
s_lname: chikhvadze
s_phone: 599-123-654,
s_subjects: [CS,MA]

A query does not need to load the whole row, that is all column families, if a rowkey (partition key) controls multiple column families. Give an example of a query that loads the whole row (all column families) and a query the does not need to load the whole row.

Use Case Examples

Use cases:

- Global Press Agencies storing their articles / press releases by date / topic
 - write a query for this database (keyspace)
 - write down a possible table (column-family) definition
- Social media: feeds by followed persons per user
 - write a query for this database (keyspace)
 - write down a possible table (column-family) definition