

Data Modeling in Apache Cassandra

1 ALWAYS think queries first

One table per query is a great strategy

No JOINS

Denormalization

Must be done for fast reads

No concern about
slow writes of duplication

Apach Cassandra has been
optimized for fast writes

CQL

Cassandra Query Language

Similar to SQL

No JOINS, GROUP BY, subqueries

★ PRIMARY KEY

PARTITION KEY(must)

Determines the distribution of data

Its value will be hashed
to a number and will determine

Node

CLUSTERING COLUMNS(optional)

As many

Sort the data in ascending order

Must be unique

Except "ALLOW FILTERING"

Must for query

Fetch more specific data

Fast reads for big data

WHERE Clause

PARTITION KEY must be included

CLUSTERING COLUMNS follow the order