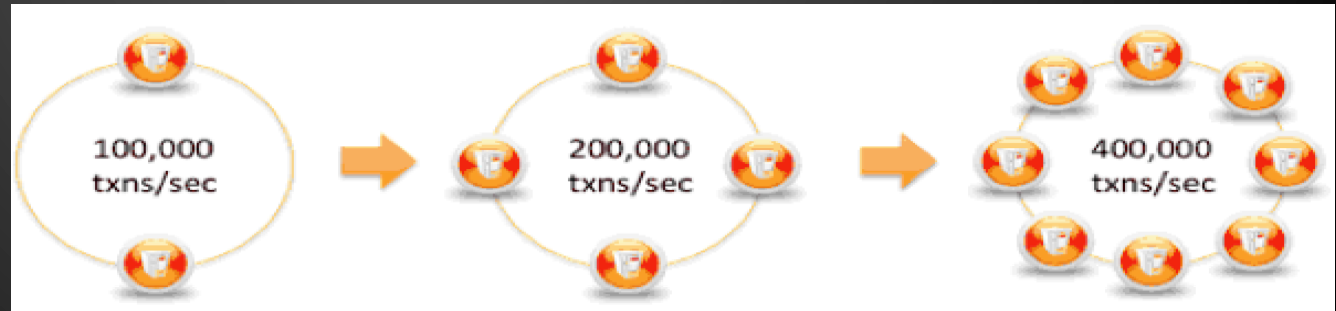# Cassandra

By: Sheefali Tewari

What is Cassandra?

- Scaled NoSQL database

- Good for managing large amounts of structured, semi-structured, and unstructured data

- No single point of failure

Features

- Elastic scalability

- Always on architecture

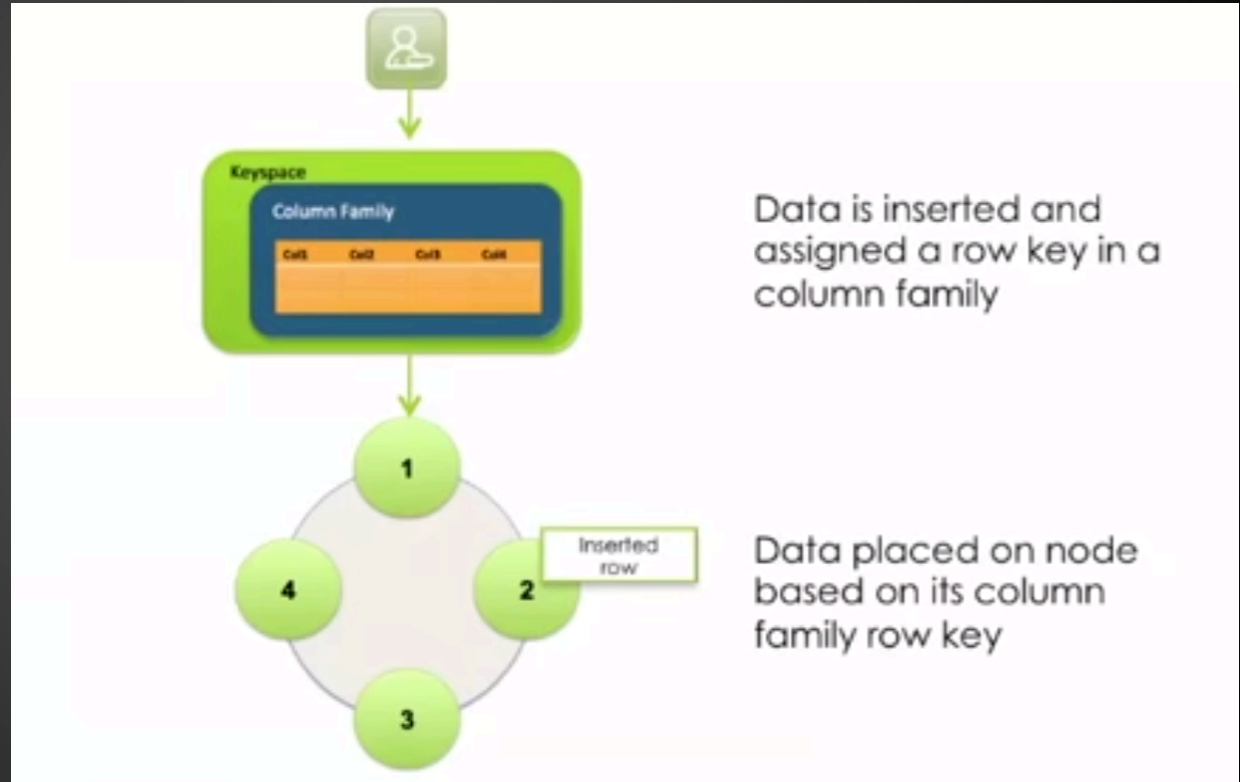- Fast linear-scale performance
  - Example:

More Features

- Flexible data storage

- Easy data distribution

- Operational simplicity

Even More Features

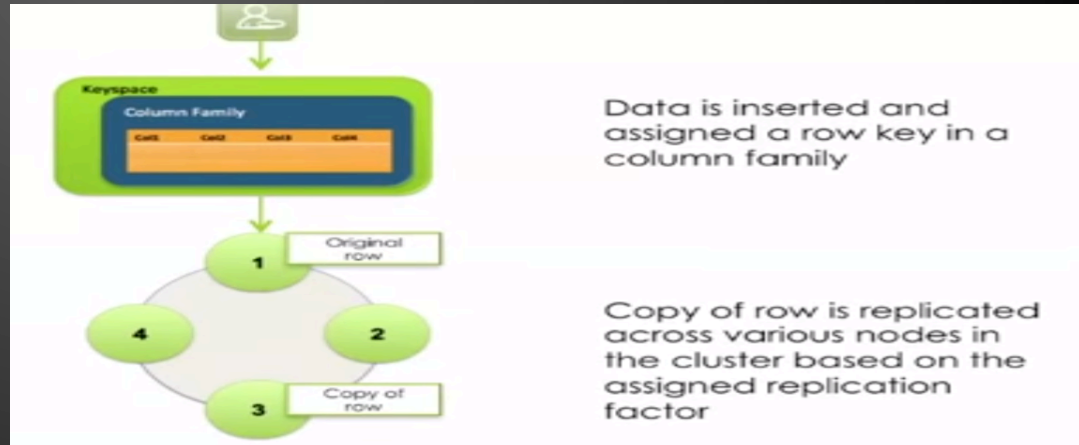- Transaction support

- Availability

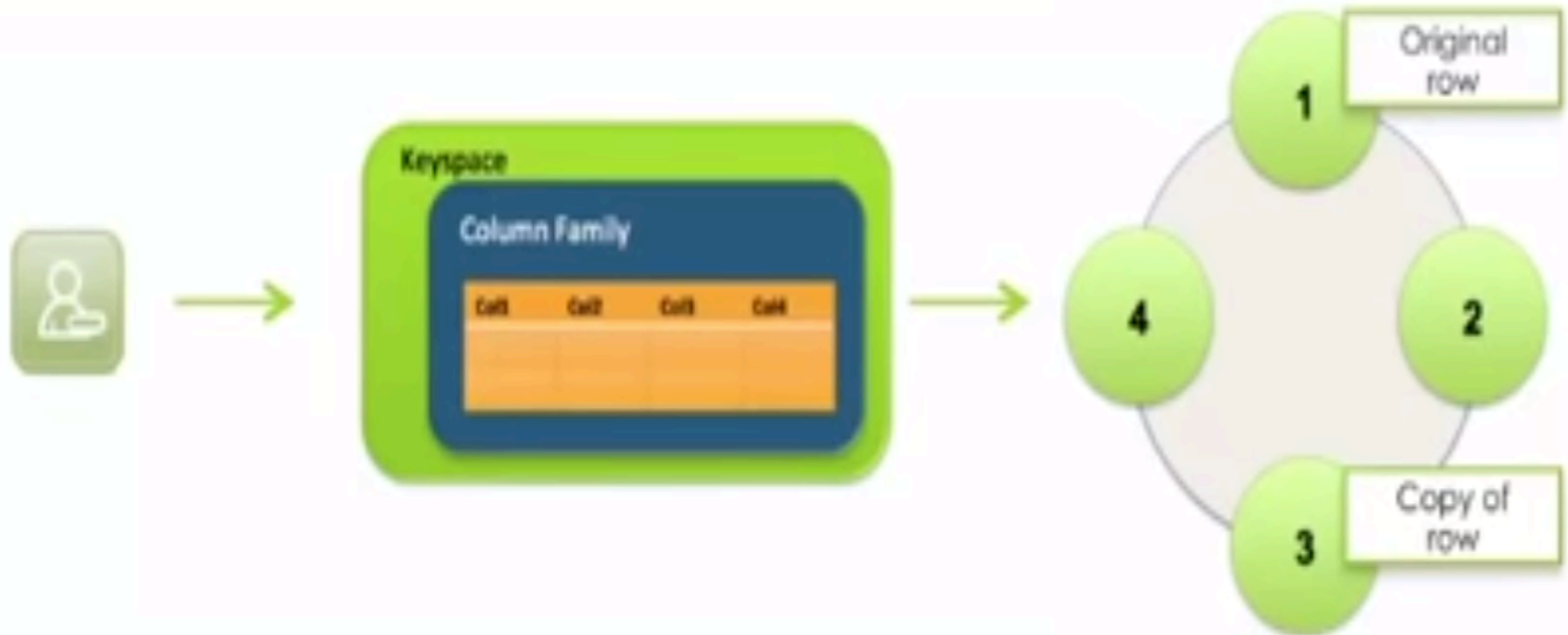# Data Partitioning in Cassandra

- Random

- Ordered

Replication in Cassandra

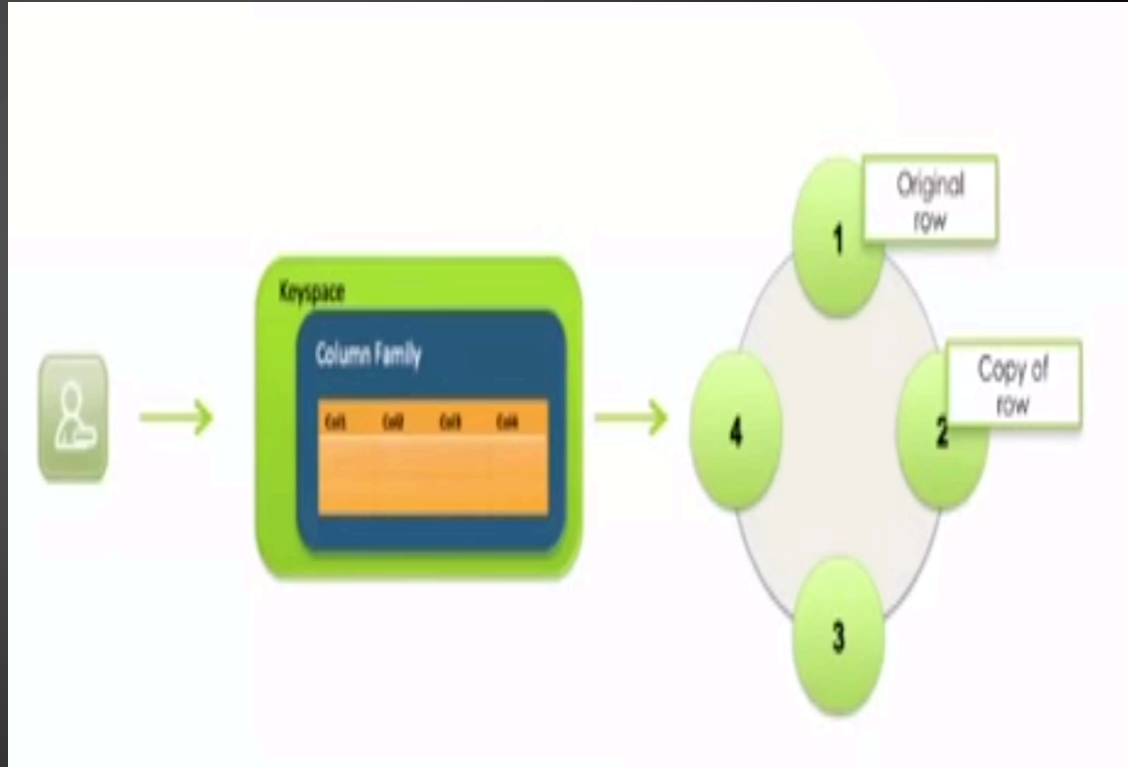- To ensure fault tolerance

- No single point of failure

Data is inserted and assigned a row key in a column family

Copy of row is replicated across various nodes in the cluster based on the assigned replication factor
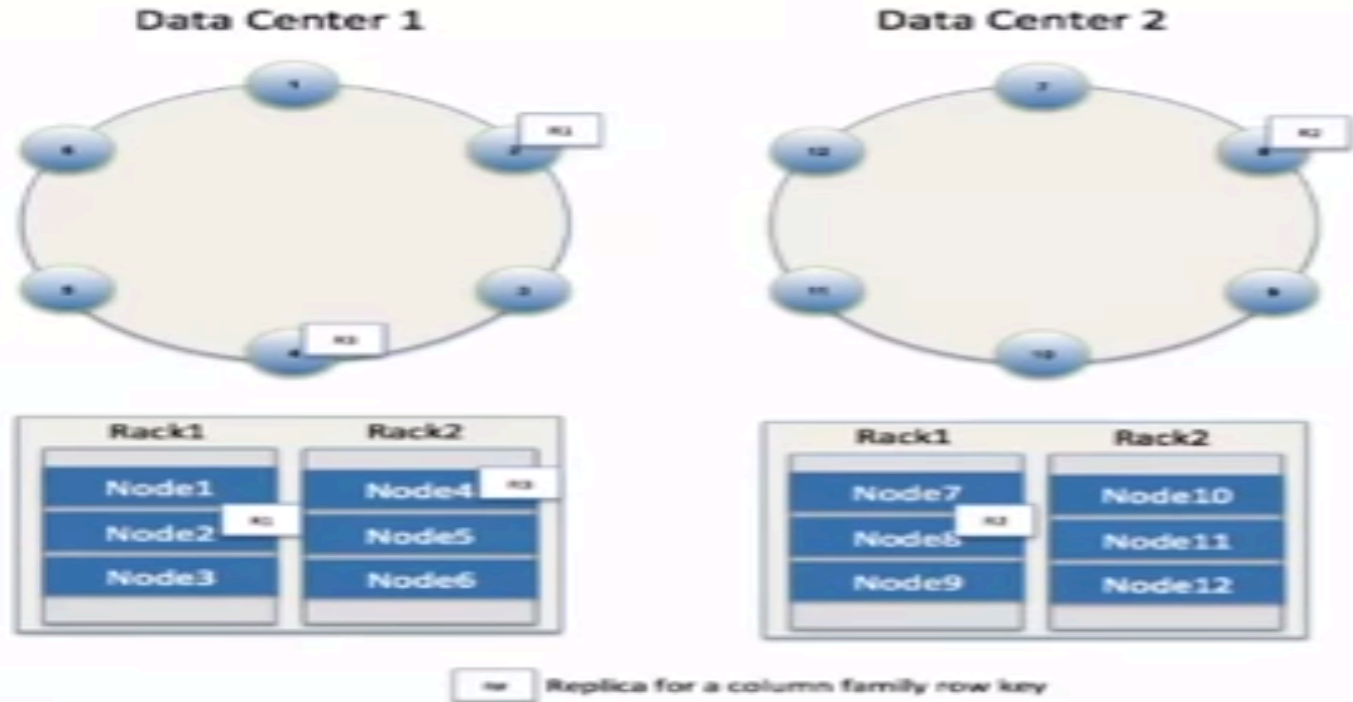
Replication in Cassandra

Replication Strategies: Simple

● Place original row on a node determined by the partitioner

Replication Strategies: Network Topology Strategy

Comparisons

- Apache Cassandra: scalable, high performance distributed database

- Apache HBase: open source, non-relational, distributed database

- MongoDB: cross-platform document-oriented database system

Who Uses Cassandra?

- Netflix
- Ebay
- Twitter
- Reddit
- Etc

Source

http://planetcassandra.org/what-is-apache-cassandra/