Cassandra

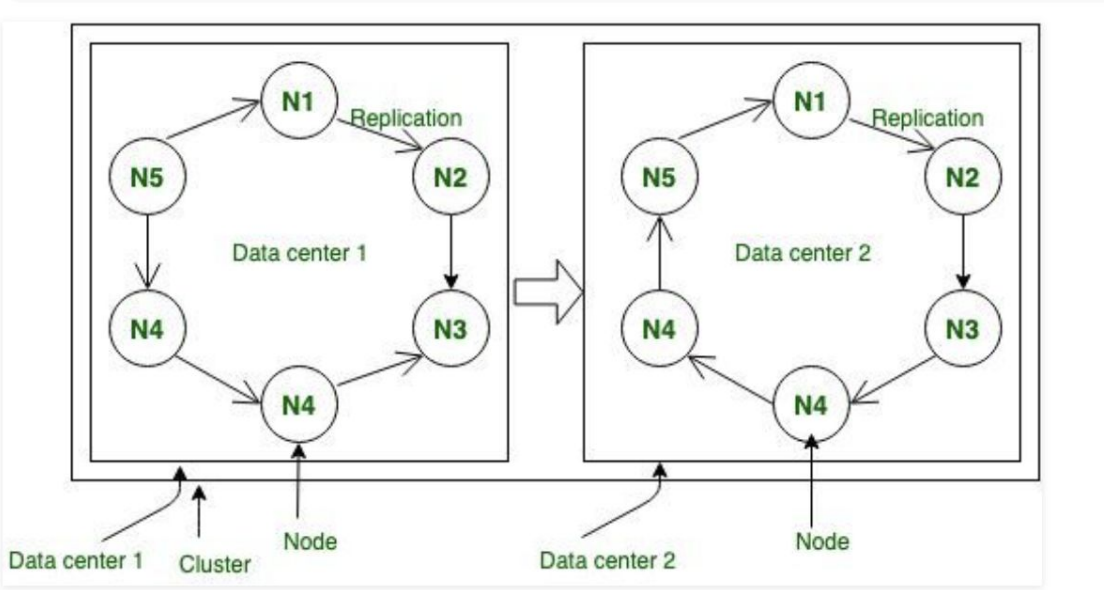
Cassandra features

- Open-source
- Written in Java
- NoSQL database
- Distributed Database
- Scalability
- Big data technology
- High availability (replication factor, multiple data center)
- Based on DynamoDB and Google Big Table

Use cases

- Product Catalog
- Recommendation
- Fraud detection
- IOT
- etc..

Distributed Database



Snitch

- Simple snitch
  - be used on one data center
- PropertyFileSnitch
  - Be used for multiple data center
  - Have to add/adjust all files to all nodes
- Gossip
  - Nodes communicate with other nodes to get info
-
-

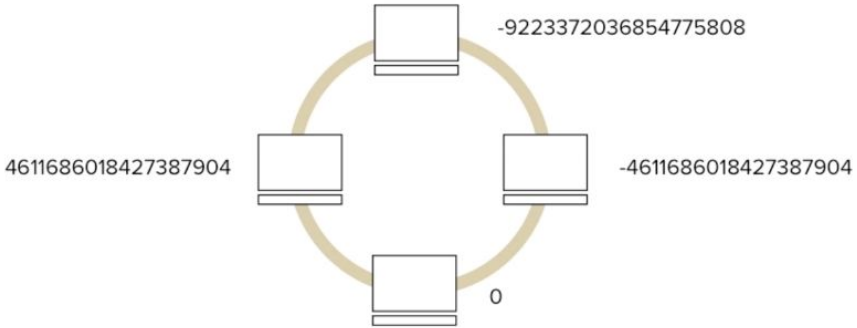
Cassandra Node IP=Data Center:Rack

```
# datacenter One

175.56.12.105=DC1:RAC1
175.50.13.200=DC1:RAC1
175.54.35.197=DC1:RAC1

120.53.24.101=DC1:RAC2
120.55.16.200=DC1:RAC2
120.57.102.103=DC1:RAC2

# datacenter Two

110.56.12.120=DC2:RAC1
110.50.13.201=DC2:RAC1
110.54.35.184=DC2:RAC1
```

Data Distribution

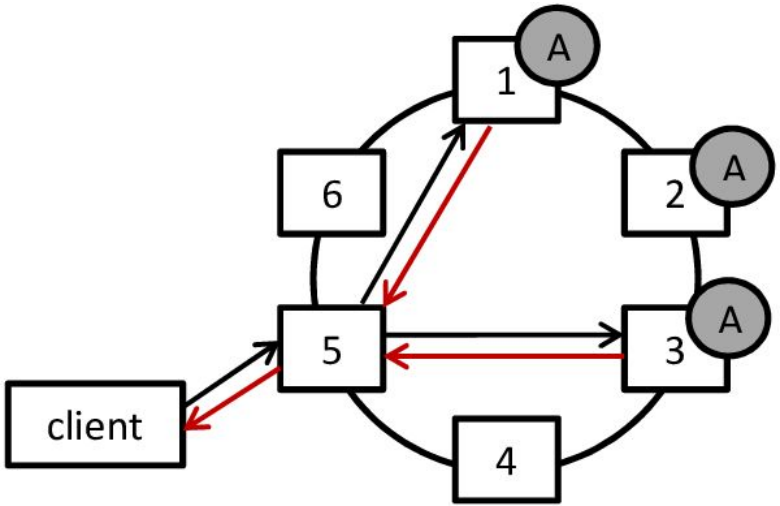
home_id	datetime	event	code_used
H01033638	2014-05-21 07:55:58	alarm set	2121
H01545551	2014-05-21 08:30:14	alarm set	8889
H00999943	2014-05-21 09:05:54	alarm set	1245

- Murmur3 is used for a partitioning algorithm
- Murmur3 creates a unique number for the first column
- Each node responsible for token range



Replication factor

- To store same data in different nodes
- Commonly specified 2,3
  - Depends on the criticality of the data



End