
DATASTAX

DATASTAX

DATASTAX ACADEMY

DataStax Enterprise

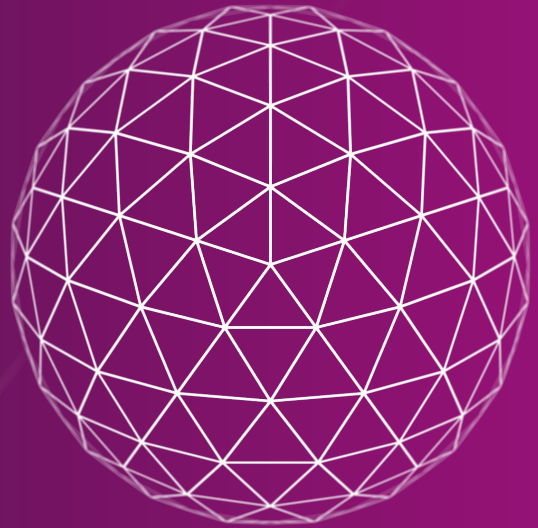
Foundations of Apache Cassandra™

DS201 - sections 6 -

July 27th, 2023

DATASTAX ACADEMY

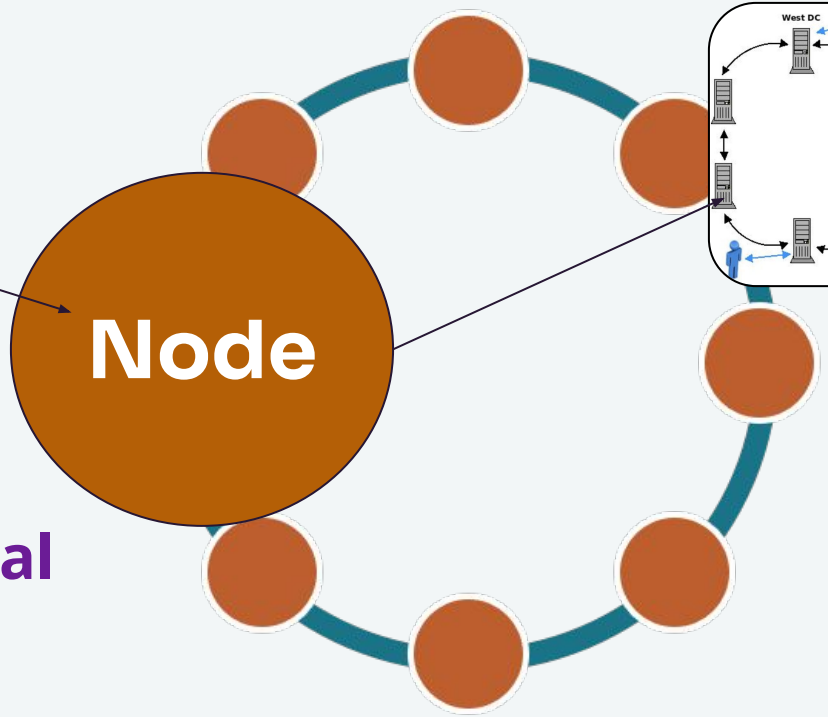
#6 - Nodes

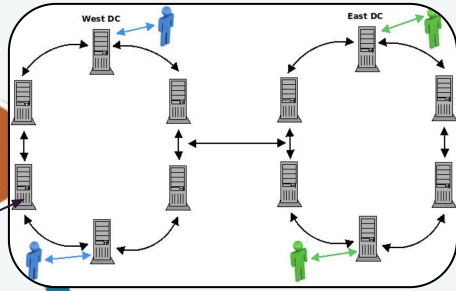


Apache Cassandra™ Node

DATASTAX ACADEMY

What is a Node?





Can be either physical or logical.

DATASTAX ACADEMY

Tips

› From the Trenches

Be careful when selecting disk for your nodes!

Always use SSDs.

Monitor IOPS.

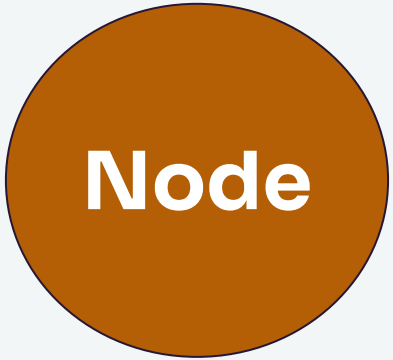
Understand your disk arrays.

Apache Cassandra™ Node

DATASTAX ACADEMY

Stores data that it is responsible for.

```
{'-4142968581484834081',
'-74364765163325014 28',
'2565879255204039505',
'32285419931567747 22',
'4692907045319667757',
'5441372649615272288',
'5512830464417747154',
'64228252448376138 86'}
```



TX

-259495160448489 8973

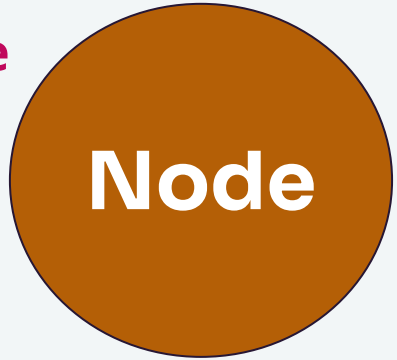
Max token range: -2^{63} to $2^{63}-1$

Apache Cassandra™ Node

DATASTAX ACADEMY

What can a single node handle?

- 6000 - 12000 ops/sec/core
- 2 - 4 TB

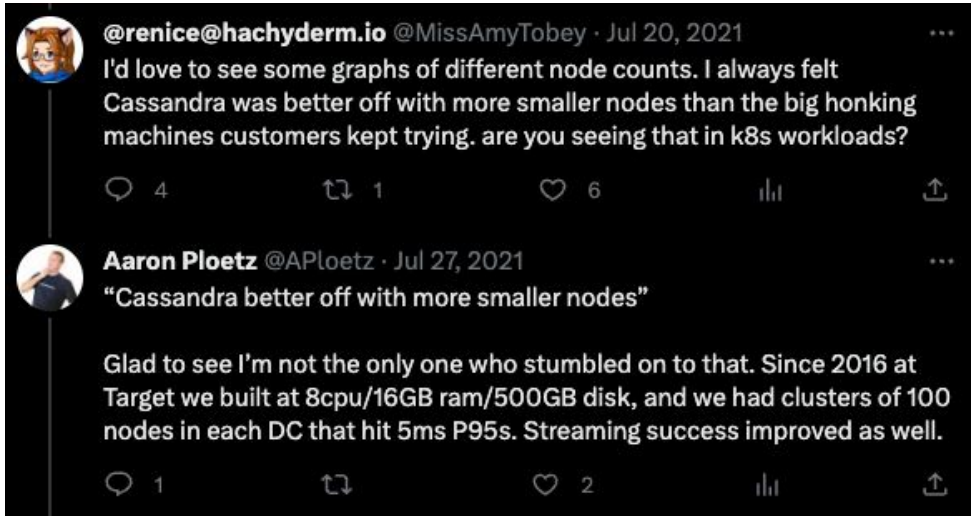


DATASTAX ACADEMY

Tips

› From the Trenches

More, smaller nodes will perform better than fewer, dense nodes.



Nodetool

DATASTAX ACADEMY

Management and health of a single Cassandra node

```
$ bin/nodetool <command>
```

Command	Description
help	Lists subcommands.
info	Information for current node.
status	Basic node health information.
tablestats	Metrics for a particular table on this node.

dsetool

DATASTAX ACADEMY

Management and health of a single DSE node

```
$ bin/dsetool <command>
```

Command	Description
help	Lists subcommands.
status	Basic node health information.

Exercise #6

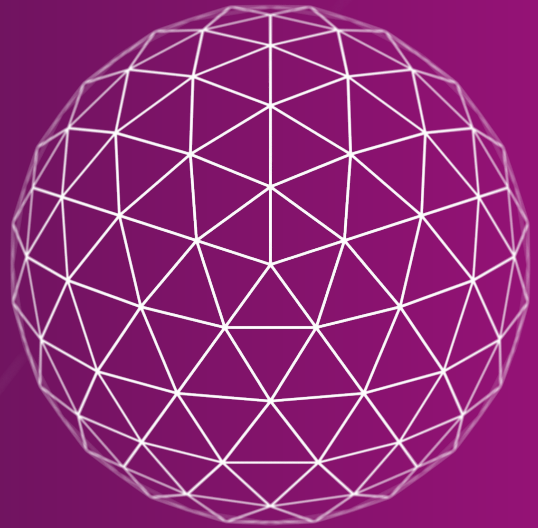
DATASTAX ACADEMY

Hands-on Exercise #6

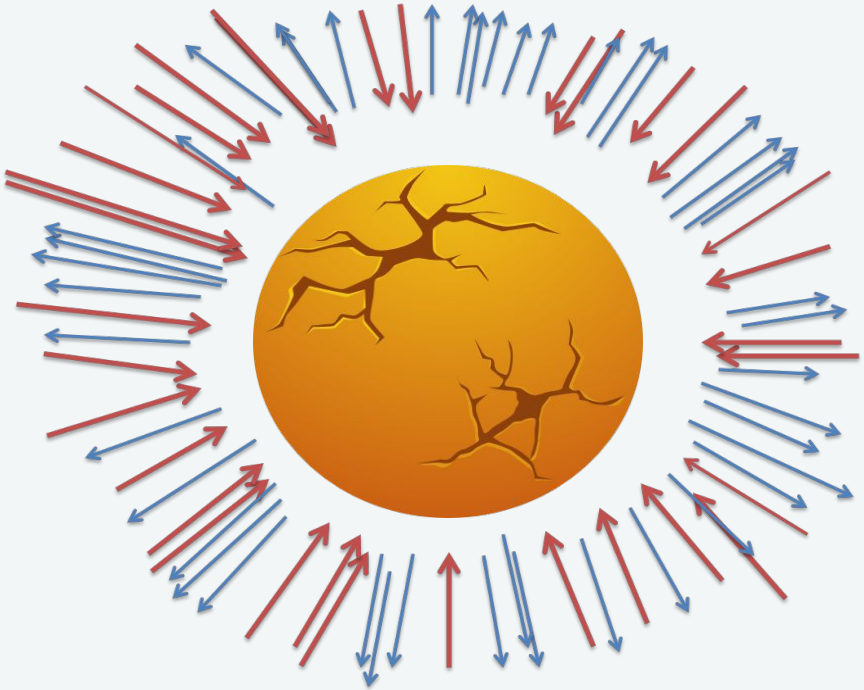
- Learn more about nodetool and dsetool
- Run several common commands

DATASTAX ACADEMY

#7 - The Ring



Pressures of scale

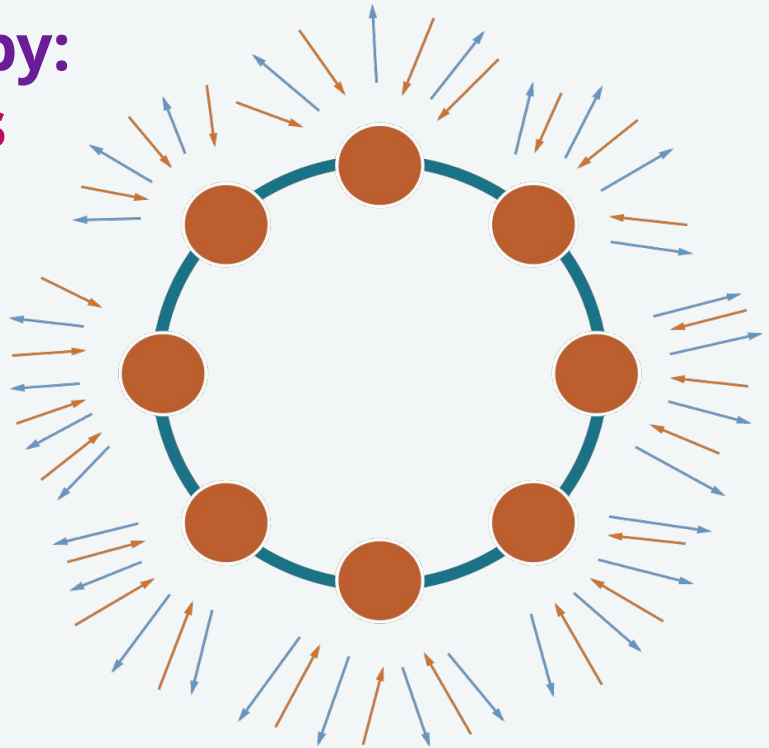


Pressures of scale

DATASTAX ACADEMY

Cassandra gets around this by:

- **Scaling w/ multiple nodes**
- **Evenly spreads:**
  - **Data**
  - **Traffic**



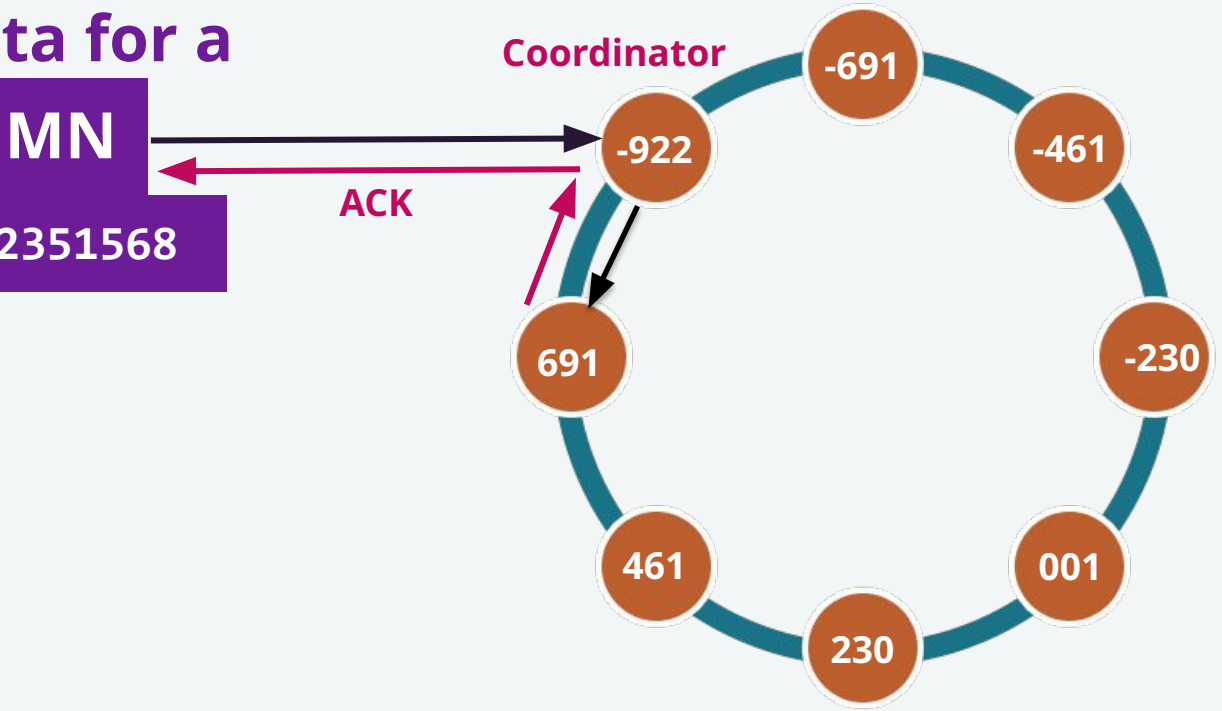
Cassandra Write Process

DATASTAX ACADEMY

Writing data for a partition

MN

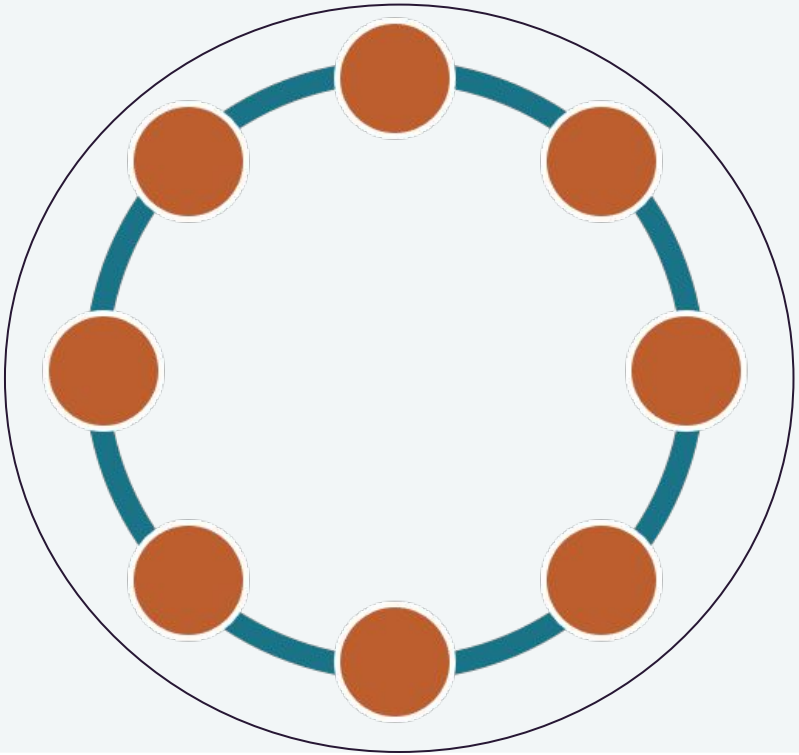
7688153959642351568



Actual Token Range

DATASTAX ACADEMY

Max token range:
-2^{63} to $2^{63}-1$



Actual Token Range

DATASTAX ACADEMY

sampleCode/tokenRanges.py

```
for counterj in range(numNodes):
    endRanges.append(str(int((2**64 / numNodes) * counterj) - 2**63))

 How many nodes are in your cluster? 8
node  start range              end range
0)    6917529027641081857 to -9223372036854775808
1)    -9223372036854775807 to -6917529027641081856
2)    -6917529027641081855 to -4611686018427387904
3)    -4611686018427387903 to -2305843009213693952
4)    -2305843009213693951 to 0
5)    1 to 2305843009213693952
6)    2305843009213693953 to 4611686018427387904
7)    4611686018427387905 to 6917529027641081856
```

DATASTAX ACADEMY

Tips

From the Trenches

For even data distribution, size your clusters as a factor of the replication factor.

If RF == 3 ***(which it should be)***

Then appropriate cluster sizes == [3, 6, 9, 12,15… 30… 60… 180 …204]

Drivers

DATASTAX ACADEMY

Drivers also understand the cluster topology

- Drivers intelligently chooses coordinators.
- Different coordinator chosen on each query.
- **`TokenAwarePolicy`** - driver chooses the node responsible for the data.
- **`RoundRobinPolicy`** - Driver “round robins” the ring.
- **`DCAwareRoundRobinPolicy`** - Driver only “round robins” a specific data center.

New Node Joins the Cluster

DATASTAX ACADEMY

What happens?

- Nodes join the cluster by communicating w/ any node.
- The new node “seeds” its discovery with its “seed list.”
- Listed seed nodes communicate cluster topology.
- Some token ranges are bisected, and assigned to the new node.
- Data is streamed to the new node.
- Once the new node joins, all nodes are peers.

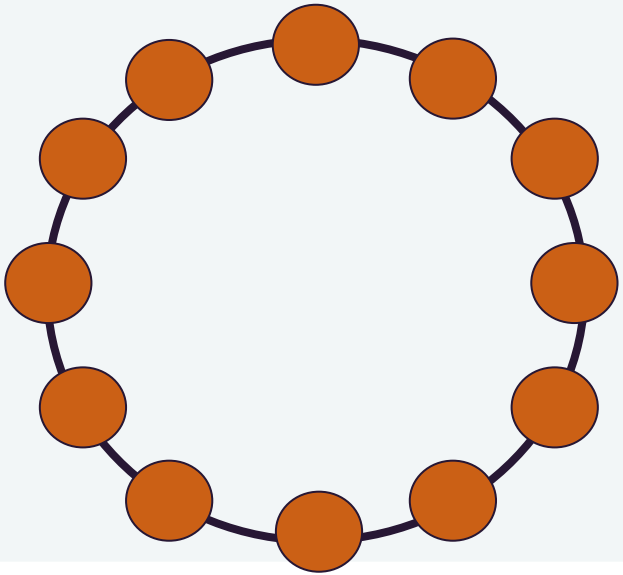
Scaling w/o Downtime

DATASTAX ACADEMY

Nodes can be added (or removed) to meet demand

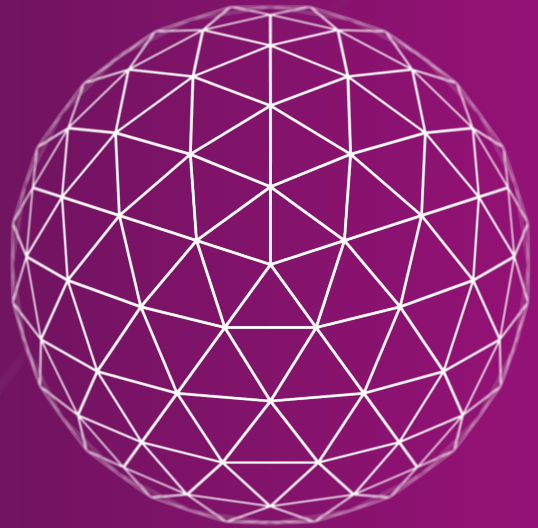
- Cassandra performs linearly based on horizontal scaling.

100,000 ops/sec



No Exercises for section #7

#8 - Peer to Peer

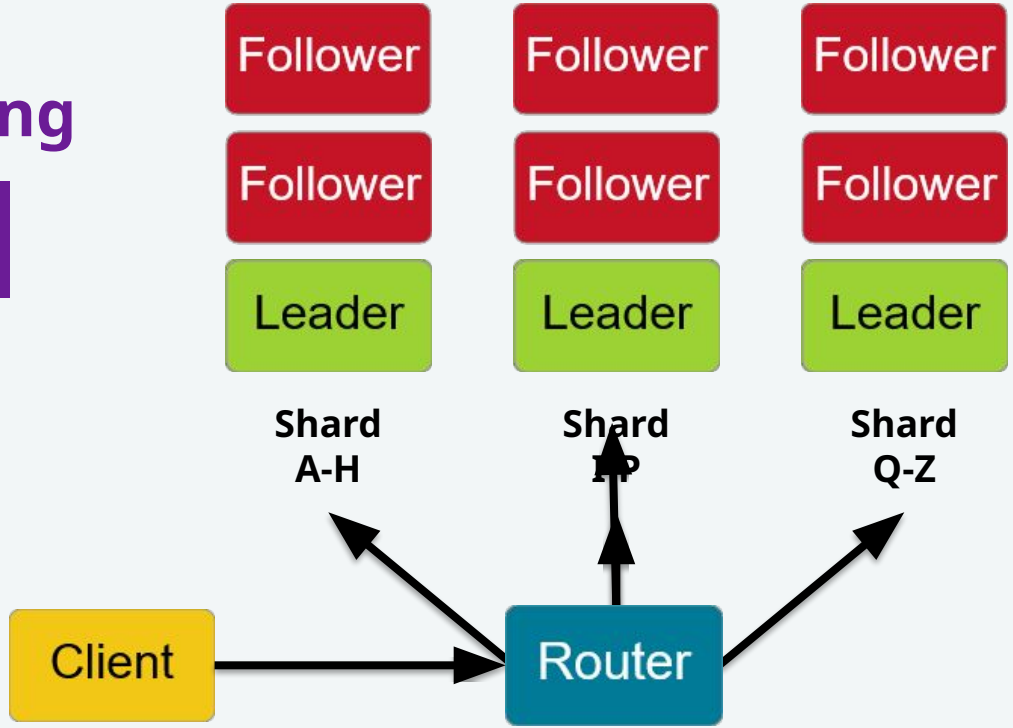


Leader – Follower Complexity

DATASTAX ACADEMY

Sharding

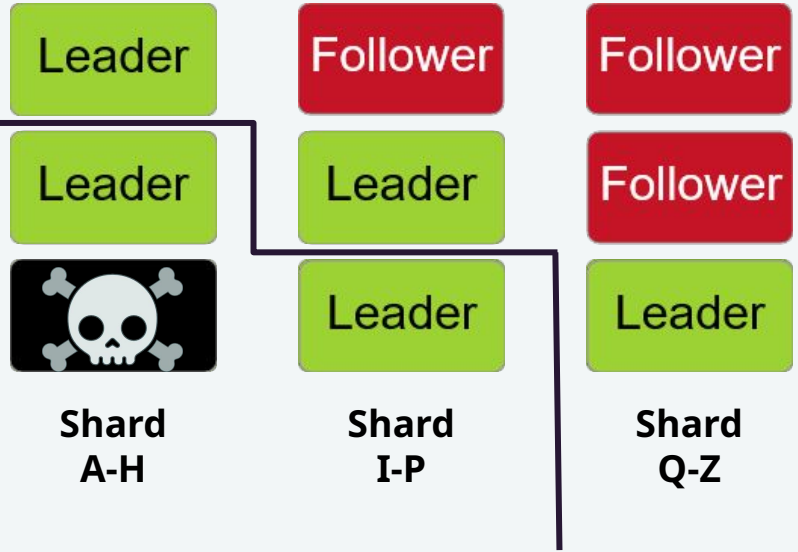
TX



Leader – Follower Failure Scenarios

DATASTAX ACADEMY

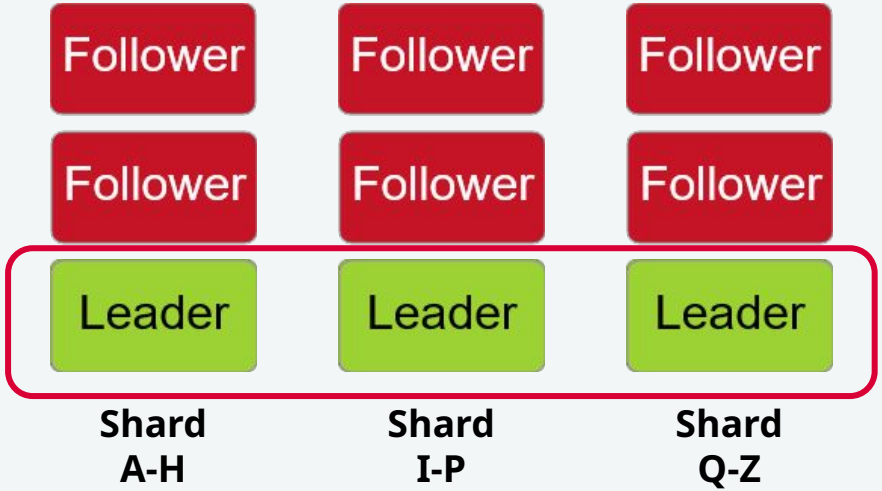
Network Failure



Leader – Follower Failure Scenarios

DATASTAX ACADEMY

Write bottleneck

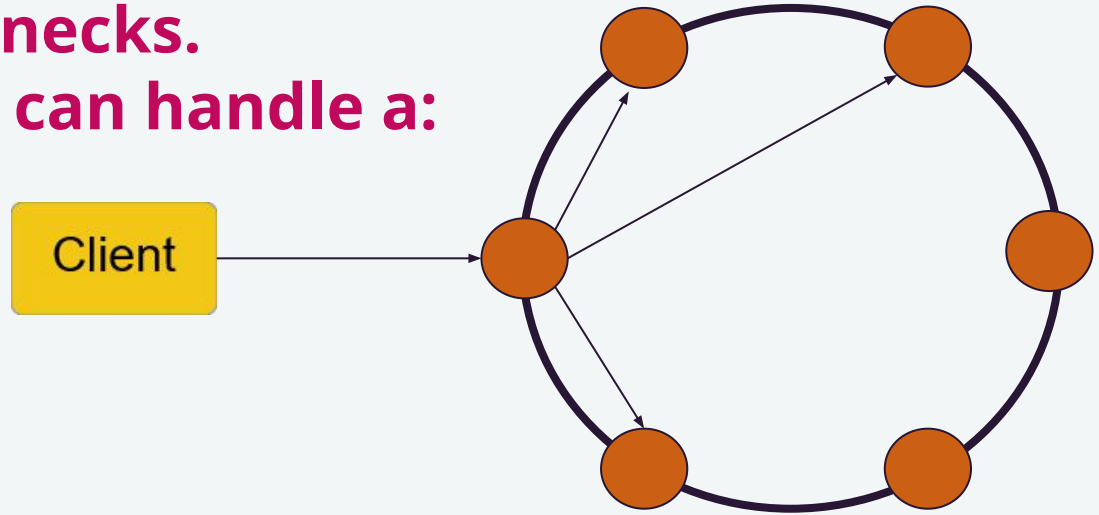


Cassandra Peer to Peer

DATASTAX ACADEMY

All nodes are created and treated equally

- **No need for shards.**
- **No bottlenecks.**
- **Any node can handle a:**
  - **read**
  - **write**

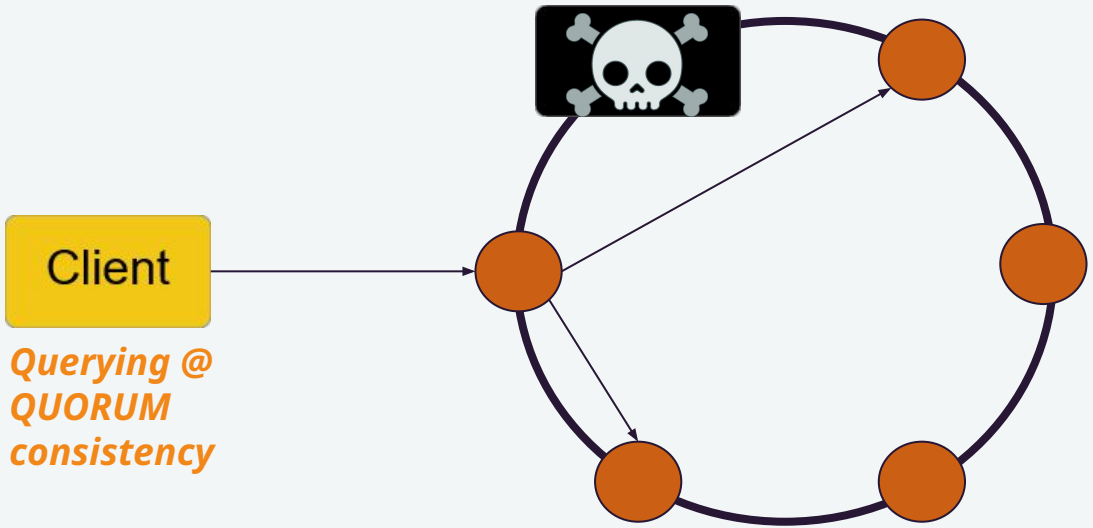


Cassandra Failure Scenarios

DATASTAX ACADEMY

Cassandra is designed to withstand a *tuneable amount* node/hardware failure.

RF == 3



Cassandra Failure Scenarios

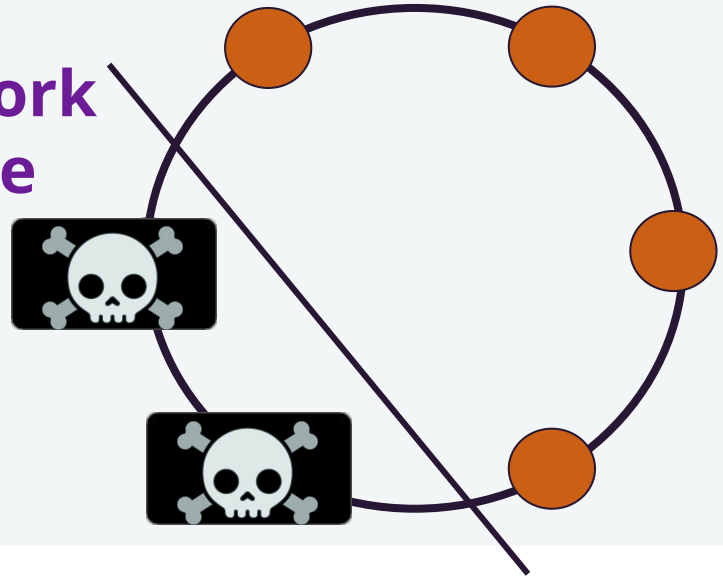
DATASTAX ACADEMY

Cassandra is designed to withstand a *tuneable amount* of node/hardware failure.

RF == 3

Network Failure

Still ok @ ONE consistency



DATASTAX ACADEMY

Pop › Quiz!

What is a QUORUM of 3?

$$(3 / 2) + 1$$

$$= 2$$

DATASTAX ACADEMY

Pop › Quiz!

What is a QUORUM of 2?

$$(2 / 2) + 1$$

$$= 2$$

No Exercises for section #8