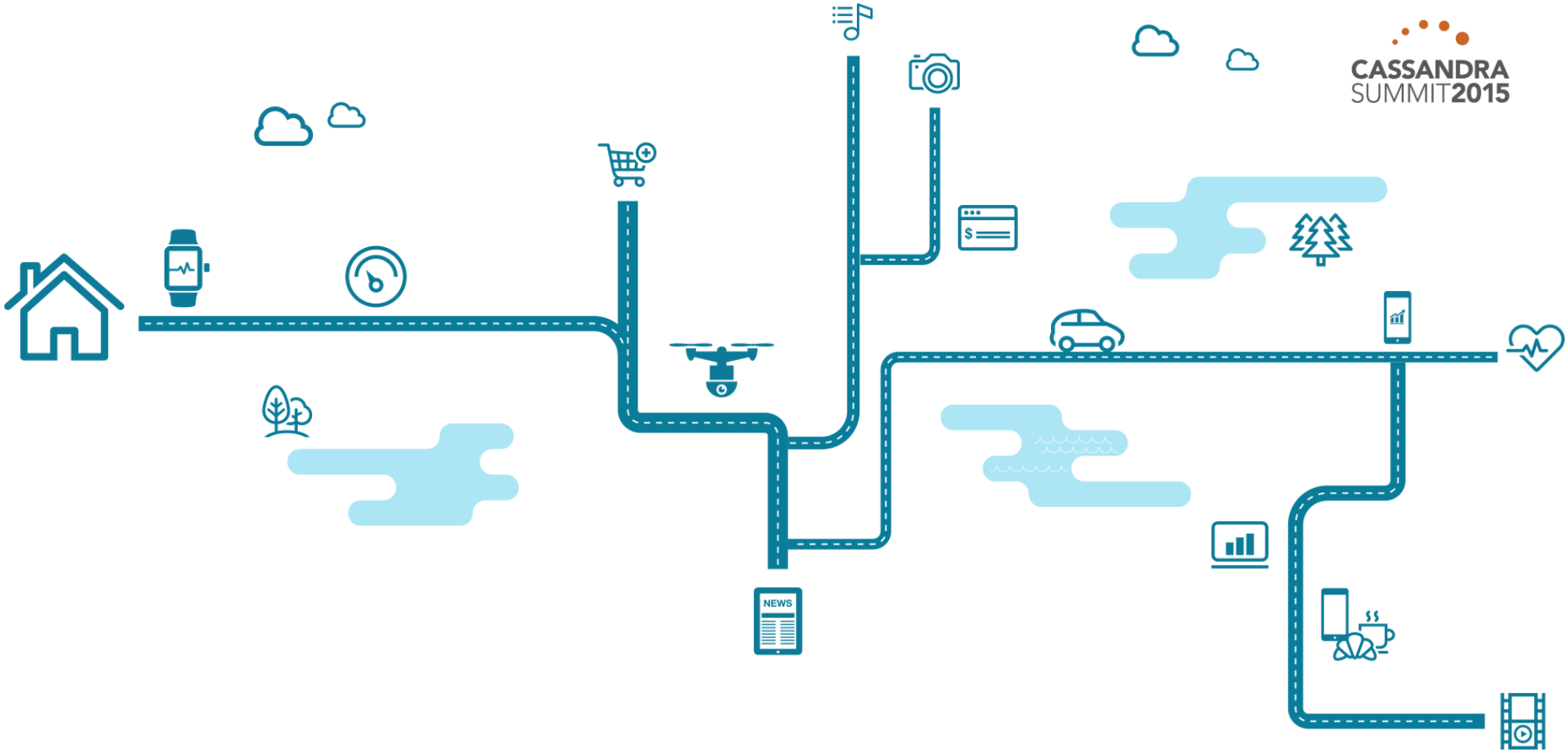


A deep look at the CQL WHERE clause

CQL WHERE clause

CASSANDRA SUMMIT2015

The WHERE clause restrictions are dependent on:

- The type of statement: SELECT, UPDATE or DELETE
- The type of column: partition key, clustering or regular column
- If a secondary index is used or not

SELECT statements

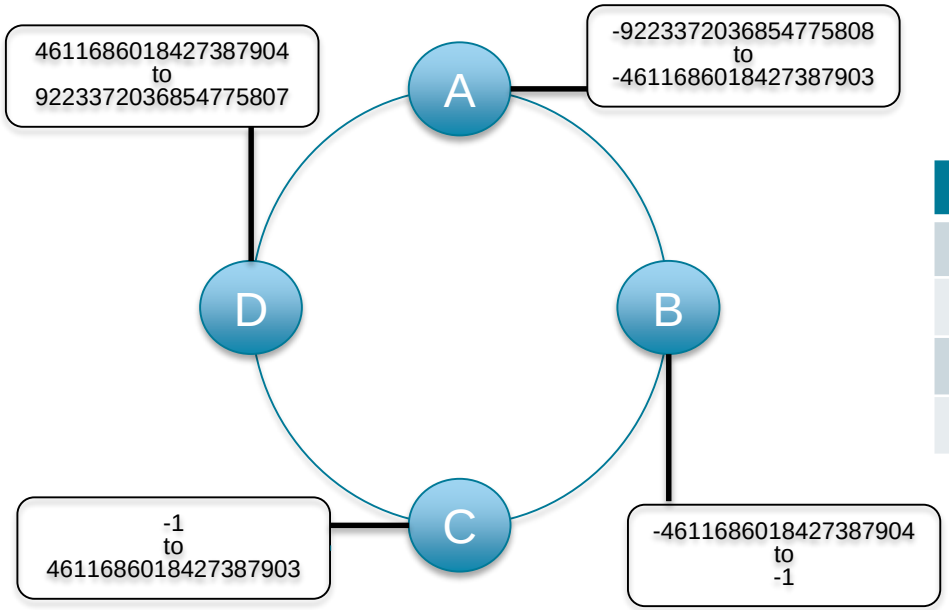
Partition key restrictions

```
CREATE TABLE numberOfRequests (
    cluster text,
    date text,
    time text,
    count int,
    PRIMARY KEY ((cluster, date))
)
```

Partition Key

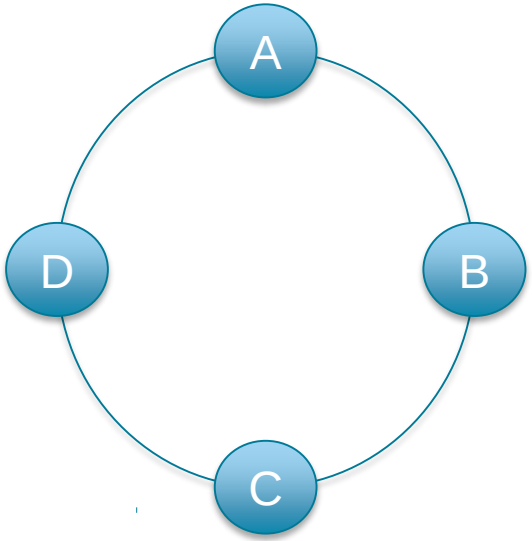
Cluster	Date	Time	Count
‘cluster 1’	‘2015-09-21’	‘12:00’	251
‘cluster 1’	‘2015-09-22’	‘12:00’	342
‘cluster 2’	‘2015-09-21’	‘12:00’	403
‘cluster 2’	‘2015-09-22’	‘12:00’	451

Partition key restrictions



Cluster	Date	Murmur3 hash
‘cluster 1’	‘2015-09-21’	-4782752162231423249
‘cluster 1’	‘2015-09-22’	4936127188075462704
‘cluster 2’	‘2015-09-21’	5822105674898716412
‘cluster 2’	‘2015-09-22’	2698159220916609751

Partition key restrictions

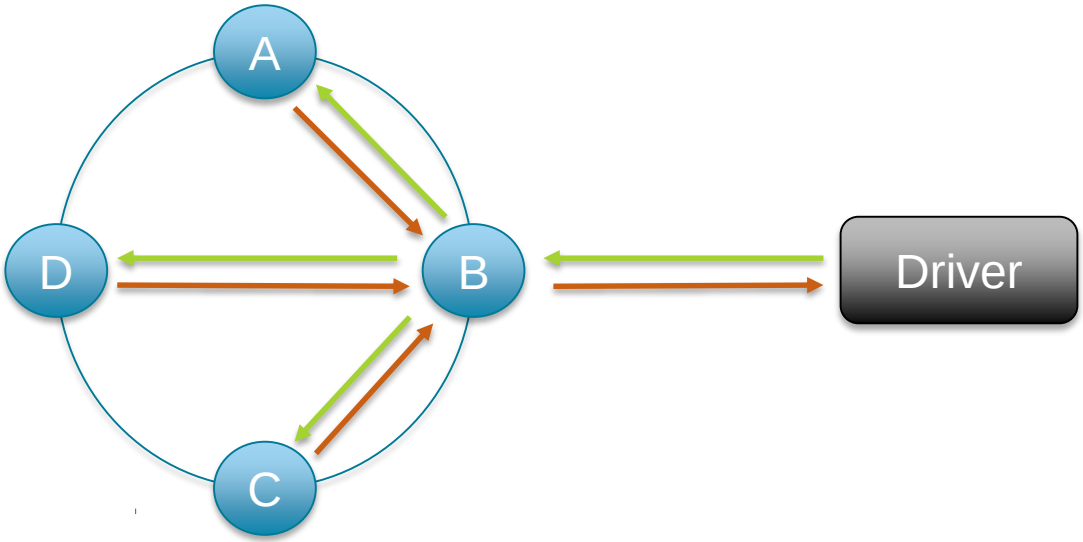


Cluster	Date	Node
‘cluster 1’	‘2015-09-21’	A
‘cluster 1’	‘2015-09-22’	D
‘cluster 2’	‘2015-09-21’	D
‘cluster 2’	‘2015-09-22’	C

Partition key restrictions

CASSANDRA SUMMIT2015

```
SELECT * FROM numberOfRequests;
```



Partition key restrictions

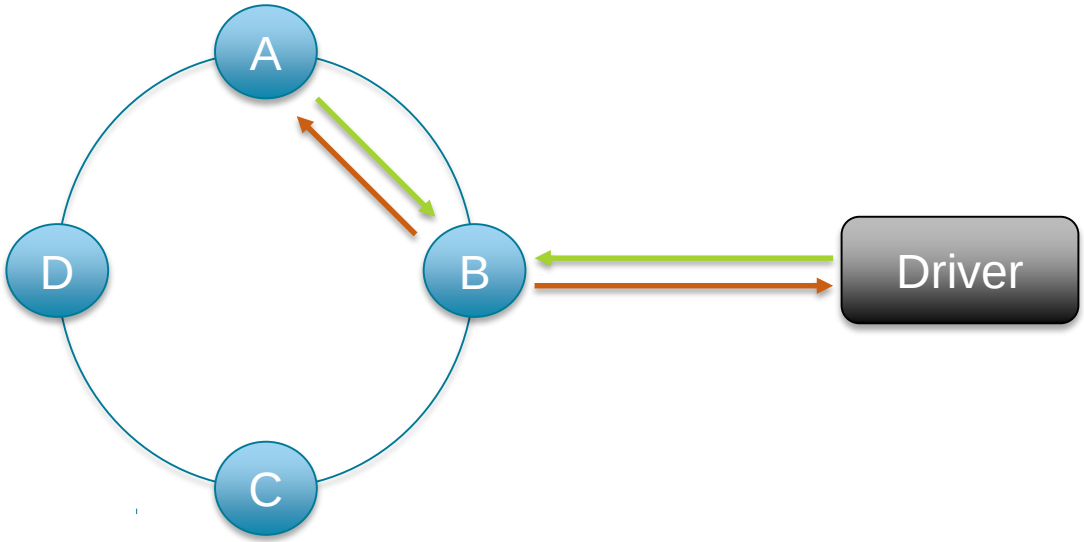
CASSANDRA SUMMIT2015

```
SELECT * FROM numberOfRequests WHERE cluster= ‘cluster 1’;
```

```
InvalidRequest: code=2200 [Invalid query]
message="Partition key parts: date must be restricted as other parts are"
```

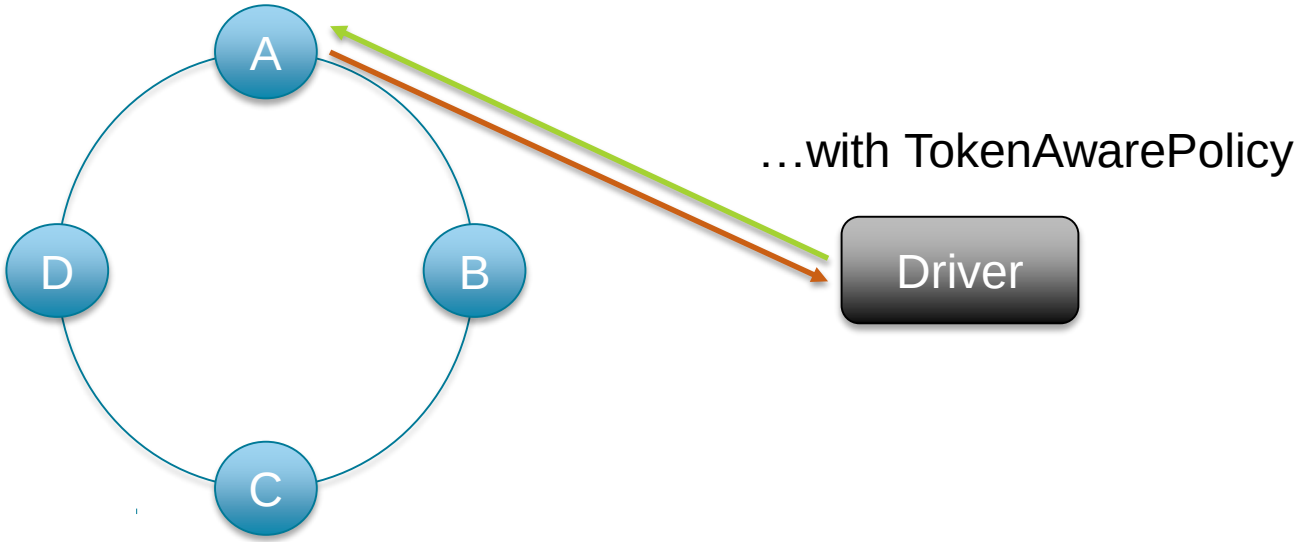
Partition key restrictions

SELECT * FROM numberOfRequests WHERE cluster= ‘cluster 1’
AND date = ‘2015-09-21’;



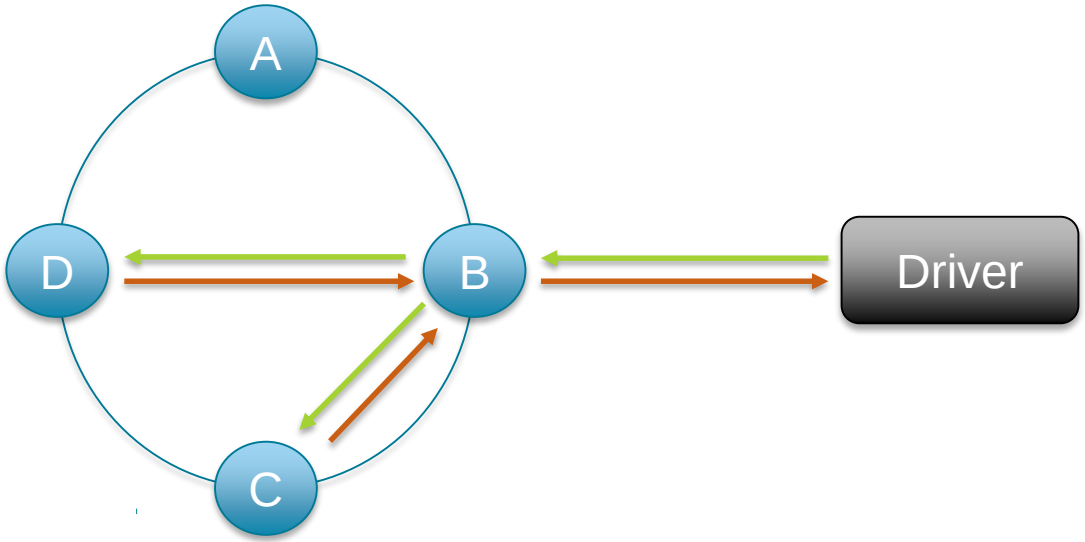
Partition key restrictions

SELECT * FROM numberOfRequests WHERE cluster= ‘cluster 1’
AND date = ‘2015-09-21’;



Partition key restrictions

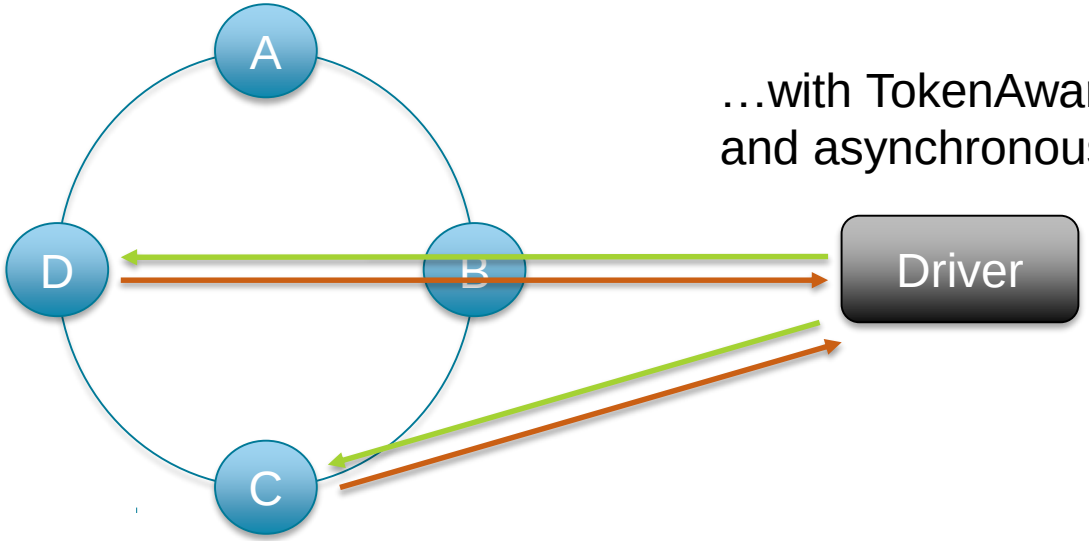
SELECT * FROM numberOfRequests WHERE cluster= ‘cluster 2’
AND date IN (‘2015-09-21’, ‘2015-09-22’);



Partition key restrictions

SELECT * FROM numberOfRequests WHERE cluster = ‘cluster 2’ AND date = ‘2015-09-21’;

SELECT * FROM numberOfRequests WHERE cluster = ‘cluster 2’ AND date = ‘2015-09-22’;



…with TokenAwarePolicy and asynchronous calls

Partition key restrictions

CASSANDRA SUMMIT2015

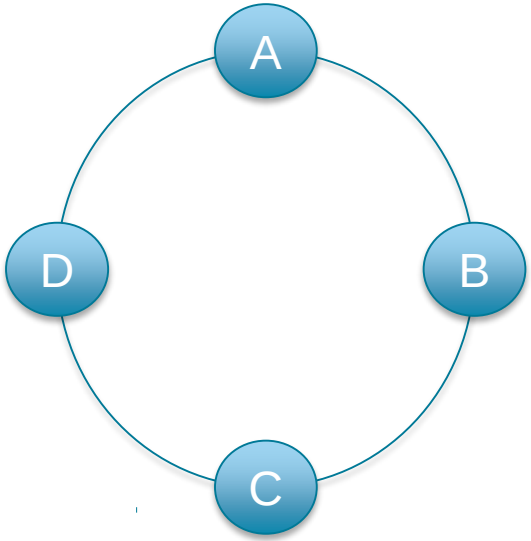
```
SELECT * FROM numberOfRequests WHERE cluster= ‘cluster 1’
AND date >= ‘2015-09-21’;
```

```
InvalidRequest: code=2200 [Invalid query]
message="Only EQ and IN relation are supported on the partition key (unless
you use the token() function)"
```

Partition key restrictions

CASSANDRA SUMMIT2015

```
SELECT * FROM numberOfRequests WHERE cluster= ‘cluster 1’
AND date >= ‘2015-09-21’;
```



Cluster	Date	Node
‘cluster 1’	‘2015-09-21’	A
‘cluster 1’	‘2015-09-22’	D
‘cluster 2’	‘2015-09-21’	D
‘cluster 2’	‘2015-09-22’	C

Partition key restrictions

CASSANDRA SUMMIT2015

- Murmur3Partitioner (default): uniformly distributes data across the cluster based on MurmurHash hash values.
- RandomPartitioner: uniformly distributes data across the cluster based on MD5 hash values.
- ByteOrderedPartitioner: keeps an ordered distribution of data lexically by key bytes

Partition key restrictions

```
SELECT * FROM numberOfRequests
WHERE token(cluster, date) > token('cluster 1', '2015-09-21')
AND token(cluster, date) < token('cluster 1', '2015-09-23');
```

Partition key restrictions (SELECT)

CASSANDRA SUMMIT2015

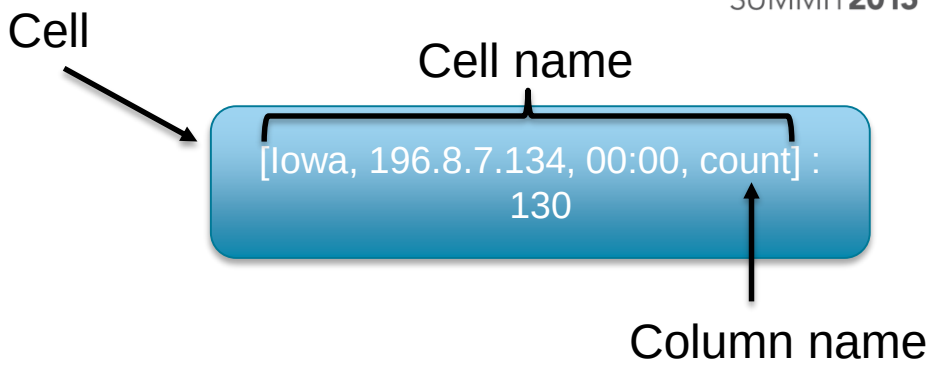
- Without secondary index, either all partition key components must be restricted or none of them
- = restrictions are allowed on any partition key component
- IN restrictions are allowed on any partition key component since 2.2
- Prior to 2.2, IN restrictions were only allowed on the last partition key component
- =, >, >=, <= and < restrictions are allowed with the token function

Clustering column restrictions

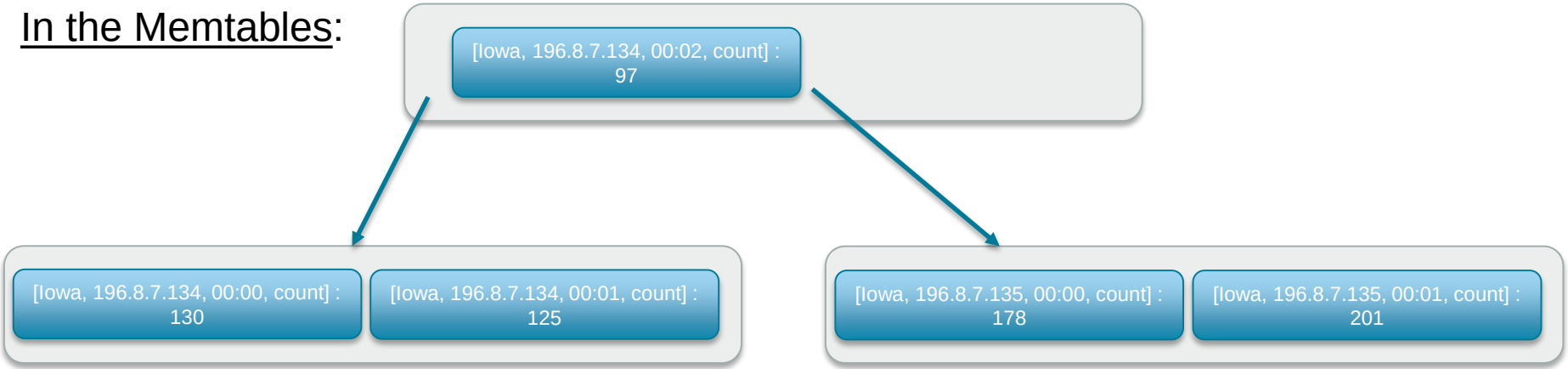
```
CREATE TABLE numberOfRequests (
    cluster text,
    date text,
    datacenter text,
    server inet,
    time text,
    count int,
    PRIMARY KEY((cluster, date), datacenter, server, time))
```

Clustering column restrictions

Datacenter	Server	Time	Count
Iowa	196.8.7.134	00:00	130
Iowa	196.8.7.134	00:01	125
Iowa	196.8.7.134	00:02	97
Iowa	196.8.7.135	00:00	178
Iowa	196.8.7.135	00:01	201

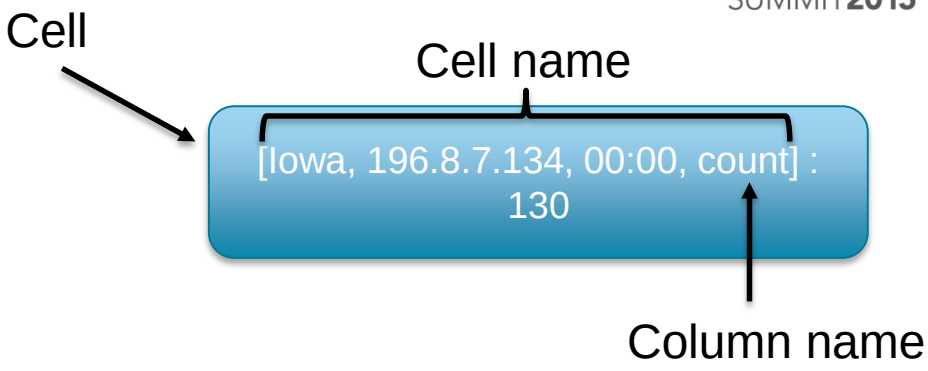


In the Memtables:



Clustering column restrictions

Datacenter	Server	Time	Count
Iowa	196.8.7.134	00:00	130
Iowa	196.8.7.134	00:01	125
Iowa	196.8.7.134	00:02	97
Iowa	196.8.7.135	00:00	178
Iowa	196.8.7.135	00:01	201



In the SSTables:

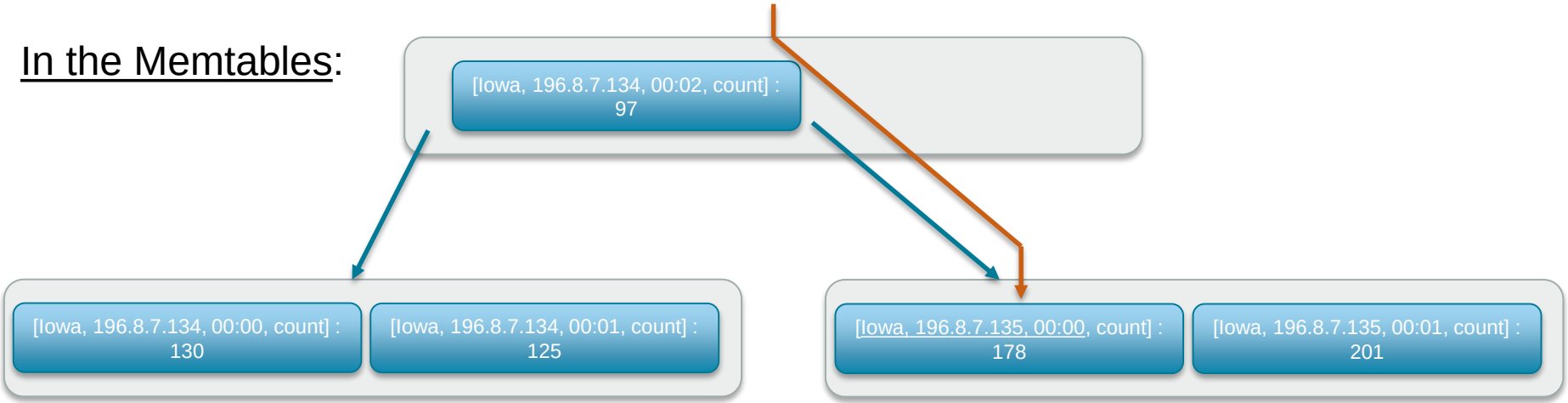
Clustering column restrictions

CASSANDRA SUMMIT2015

```
SELECT * FROM numberOfRequests
WHERE cluster = ‘cluster1’ AND date =‘2015-09-21’
AND datacenter = ‘Iowa’ AND server = ‘196.8.7.135’ AND  time = ‘00:00’;
```

[Iowa,196.8.7.135,00:00]

In the Memtables:



Clustering column restrictions

CASSANDRA SUMMIT2015

```
SELECT * FROM numberOfRequests
WHERE cluster = ‘cluster1’ AND date =‘2015-09-21’
AND datacenter = ‘Iowa’ AND server = ‘196.8.7.135’ AND  time = ‘00:00’;
```

[Iowa,196.8.7.135,00:00]

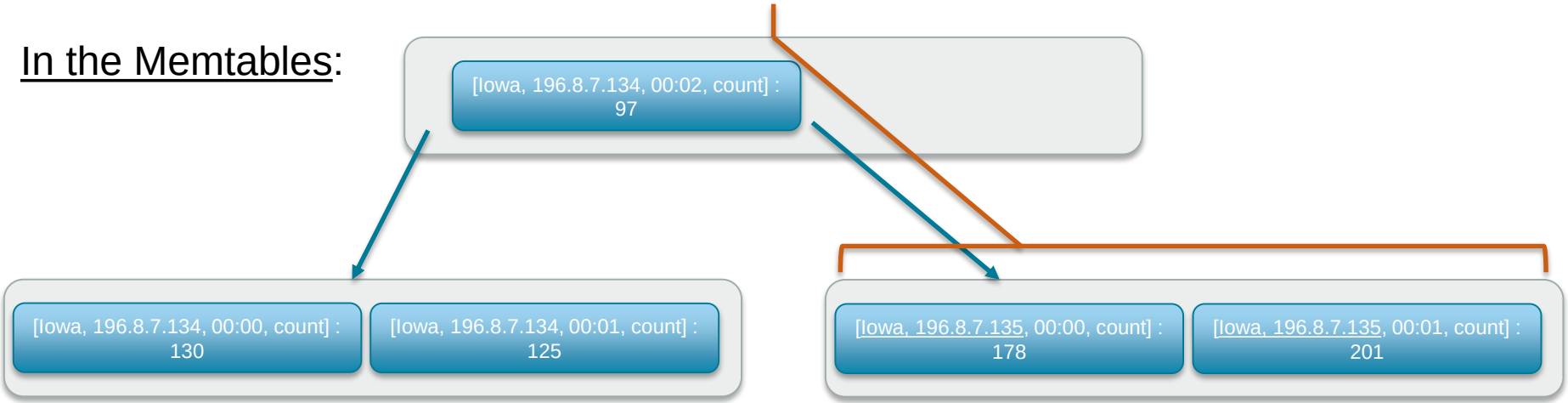
In the SSTables:

Clustering column restrictions

```
SELECT * FROM numberOfRequests
WHERE cluster = ‘cluster1’ AND date =‘2015-09-21’
AND datacenter = ‘Iowa’ AND server = ‘196.8.7.135’;
```

[Iowa,196.8.7.135]

In the Memtables:

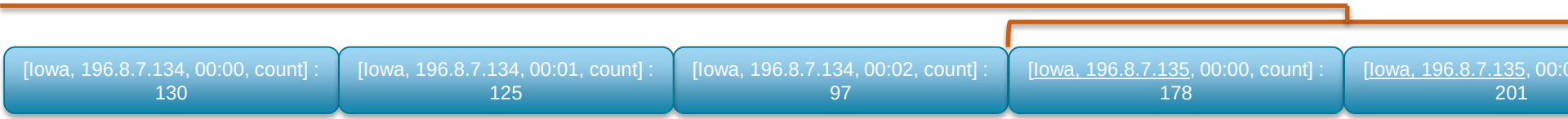


Clustering column restrictions

```
SELECT * FROM numberOfRequests
WHERE cluster = ‘cluster1’ AND date =‘2015-09-21’
AND datacenter = ‘Iowa’ AND server = ‘196.8.7.135’;
```

[Iowa,196.8.7.135]

In the SSTables:



Clustering column restrictions

CASSANDRA SUMMIT2015

```
SELECT * FROM numberOfRequests
WHERE cluster = ‘cluster1’ AND date =‘2015-09-21’
AND time = ‘00:00’;
```

[?,?,00:00]

InvalidRequest: code=2200 [Invalid query]
message="PRIMARY KEY column "time" cannot be restricted as preceding column "datacenter" is not restricted"

Clustering column restrictions

CASSANDRA SUMMIT2015

In 2.2:

```
…
AND datacenter = ‘Iowa’
AND server IN (‘196.8.7.134’, ‘196.8.7.135’)
AND time = ‘00:00’;
```

[Iowa,196.8.7.134,00:00]
[Iowa,196.8.7.135,00:00]

In the SSTables:

Clustering column restrictions

CASSANDRA SUMMIT2015

In 2.1:

```
…
AND datacenter = ‘Iowa’
AND server IN (‘196.8.7.134’, ‘196.8.7.135’)
AND time = ‘00:00’;
```

```
InvalidRequest: code=2200 [Invalid query]
message="Clustering column "server" cannot be restricted by an IN relation"
```

Clustering column restrictions

CASSANDRA SUMMIT2015

= multi-column restriction:

(clustering1, clustering2, clustering3) = (?, ?, ?)

IN multi-column restriction:

(clustering1, clustering2, clustering3) IN ((?, ?, ?), (?, ?, ?))

Slice multi-column restriction:

(clustering1, clustering2, clustering3) > (?, ?, ?)
(clustering1, clustering2, clustering3) >= (?, ?, ?)
(clustering1, clustering2, clustering3) <= (?, ?, ?)
(clustering1, clustering2, clustering3) < (?, ?, ?)

Clustering column restrictions

In 2.1:

```
…
AND datacenter = ‘Iowa’
AND (server, time) IN ((‘196.8.7.134’, ‘00:00’),
                       (‘196.8.7.135’, ‘00:00’));
```

[Iowa,196.8.7.134,00:00]
[Iowa,196.8.7.135,00:00]

In the SSTables:

Clustering column restrictions

CASSANDRA SUMMIT2015

```
…
AND datacenter = ‘Iowa’
AND server = ‘196.8.7.134’
AND time > ’00:00’;
```

from after [Iowa,196.8.7.134,00:00]
to end of [Iowa,196.8.7.134]

In the SSTables:

Clustering column restrictions (SELECT)

CASSANDRA SUMMIT2015

- Without secondary index, a clustering column cannot be restricted if one of the previous ones was not
- = restrictions (single and multi) are allowed on any clustering column
- IN restrictions (single and multi) are allowed on any clustering column since 2.2
- Prior to 2.2, IN restrictions (single and multi) were only allowed on the last clustering column or set of clustering columns
- >, >=, <=, < restrictions (single and multi) are only allowed on the last restricted clustering column or set of clustering columns
- CONTAINS and CONTAINS KEY restrictions are only allowed on indexed collections

Secondary index queries

```
CREATE TABLE numberOfRequests (
    cluster text,
    date text,
    datacenter text,
    server inet,
    time text,
    count int,
    PRIMARY KEY((cluster, date), datacenter, server, time));

CREATE INDEX ON numberOfRequests (time);
```

Secondary index queries

```
CREATE INDEX ON numberOfRequests (time);
```

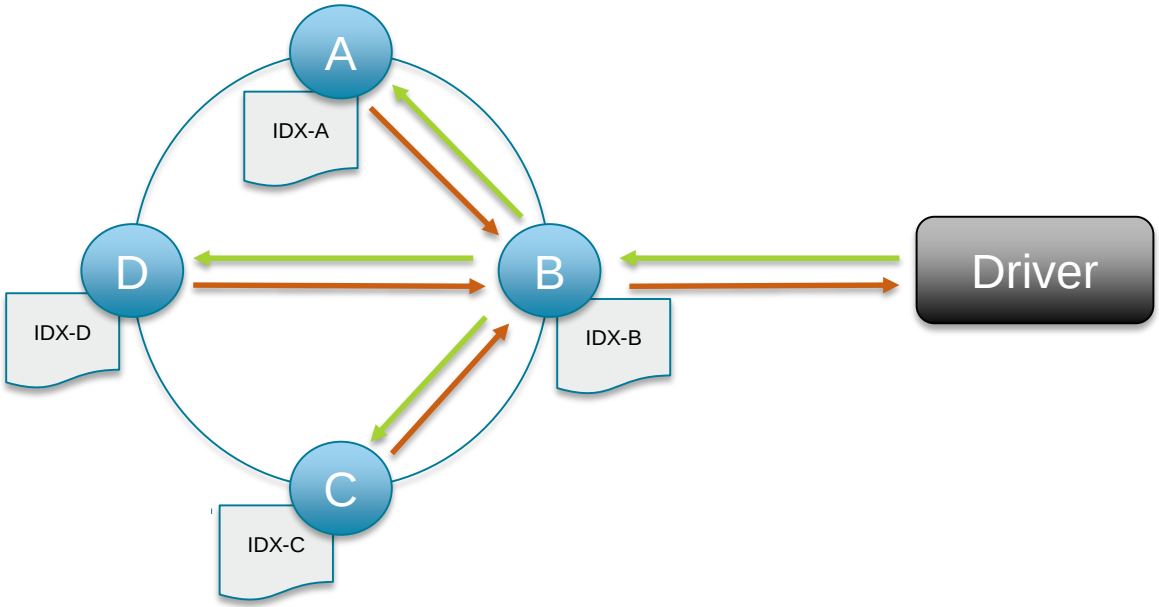
```
CREATE LOCAL TABLE numberOfRequests_time_idx (
    time text,
    cluster text,
    date text,
    datacenter text,
    server inet,
    PRIMARY KEY(time, cluster, date, datacenter, server);
```

Table Partition Key

Table remaining clustering columns

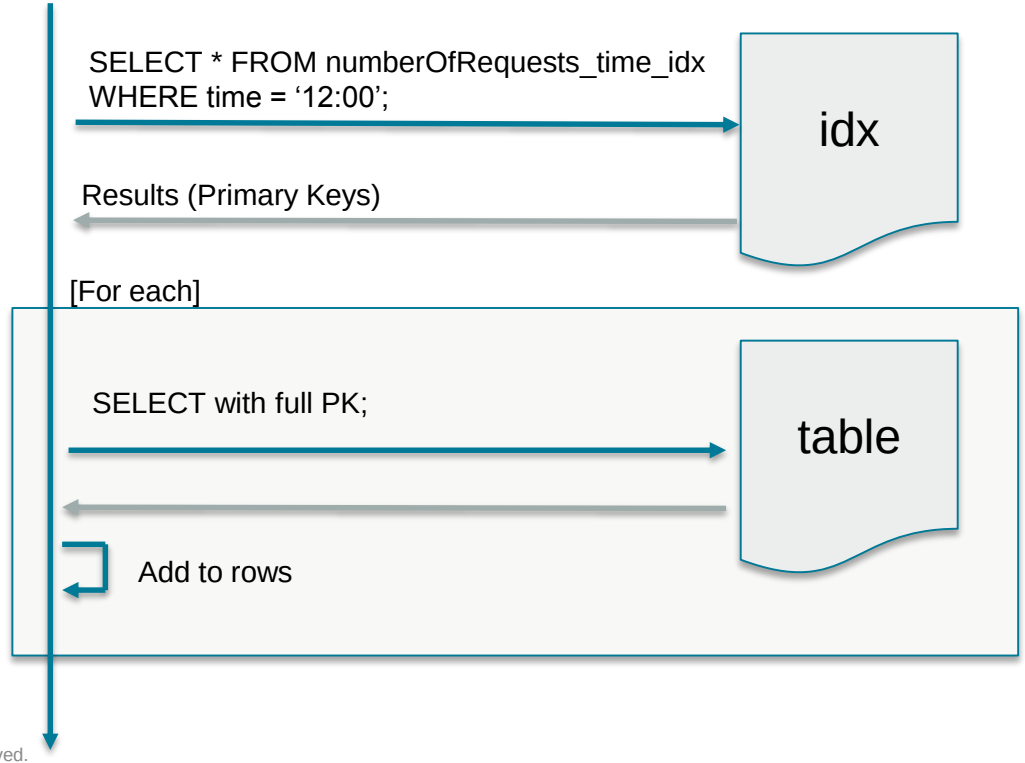
Secondary index queries

SELECT * FROM numberOfRequests WHERE time = ‘12:00’;



Secondary index queries

SELECT * FROM numberOfRequests WHERE time = ‘12:00’;



Secondary index queries

CASSANDRA SUMMIT2015

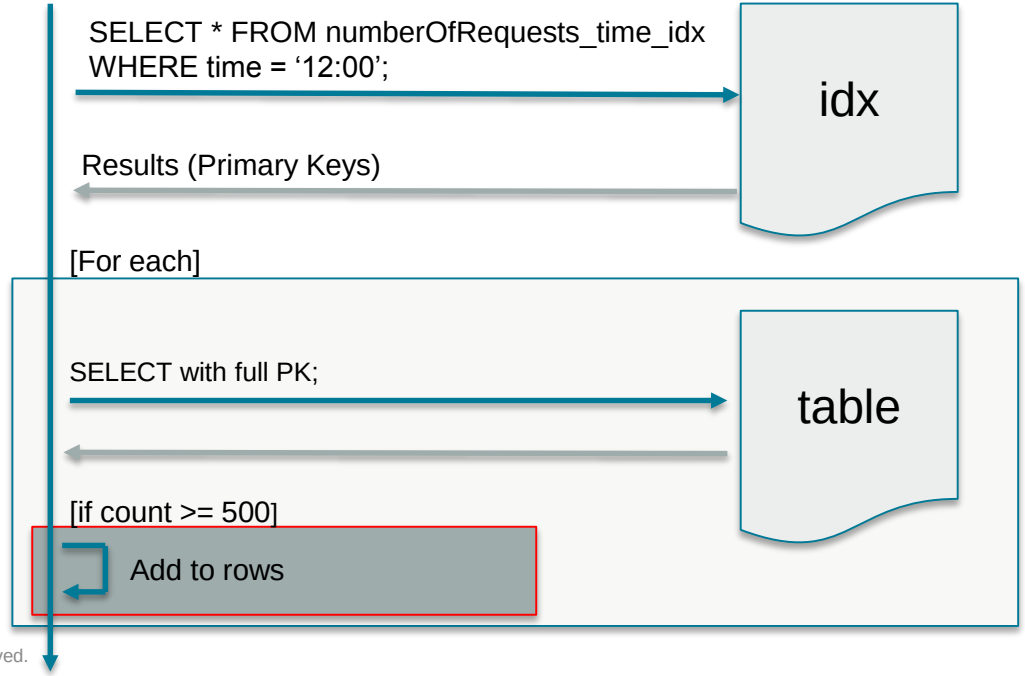
```
SELECT * FROM numberOfRequests WHERE time >= ‘12:00’;
```

```
InvalidRequest: code=2200 [Invalid query]
message="PRIMARY KEY column "time" cannot be restricted as preceding
column "datacenter" is not restricted"
```

Direct queries on secondary index support only =, CONTAINS or CONTAINS KEY restrictions.

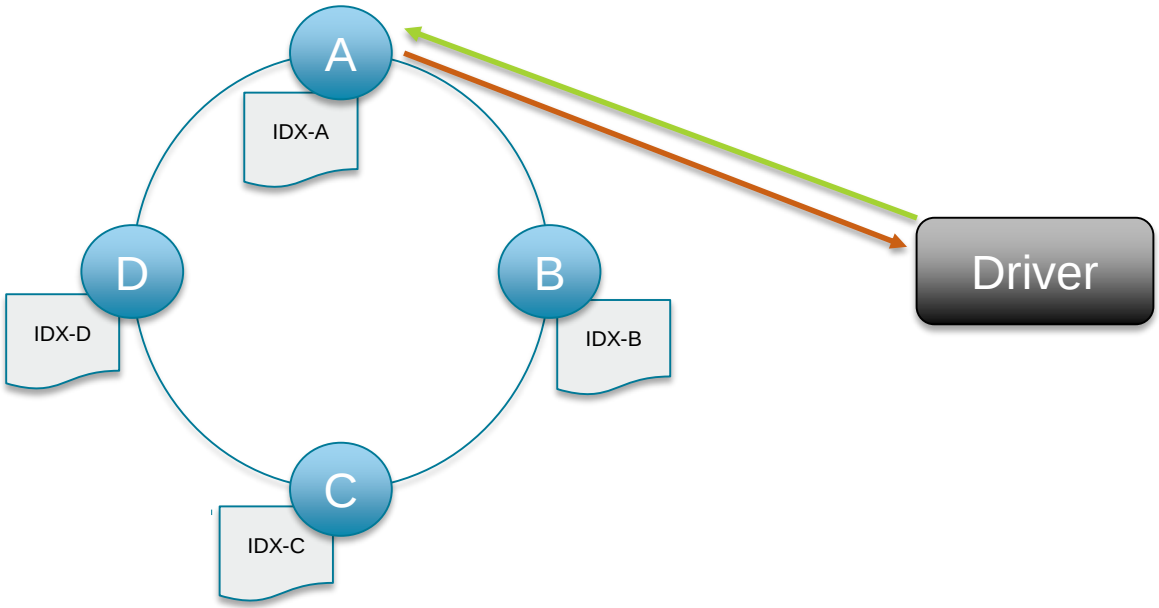
Secondary index queries

SELECT * FROM numberOfRequests WHERE time = ‘12:00’
AND count >= 500 ALLOW FILTERING;



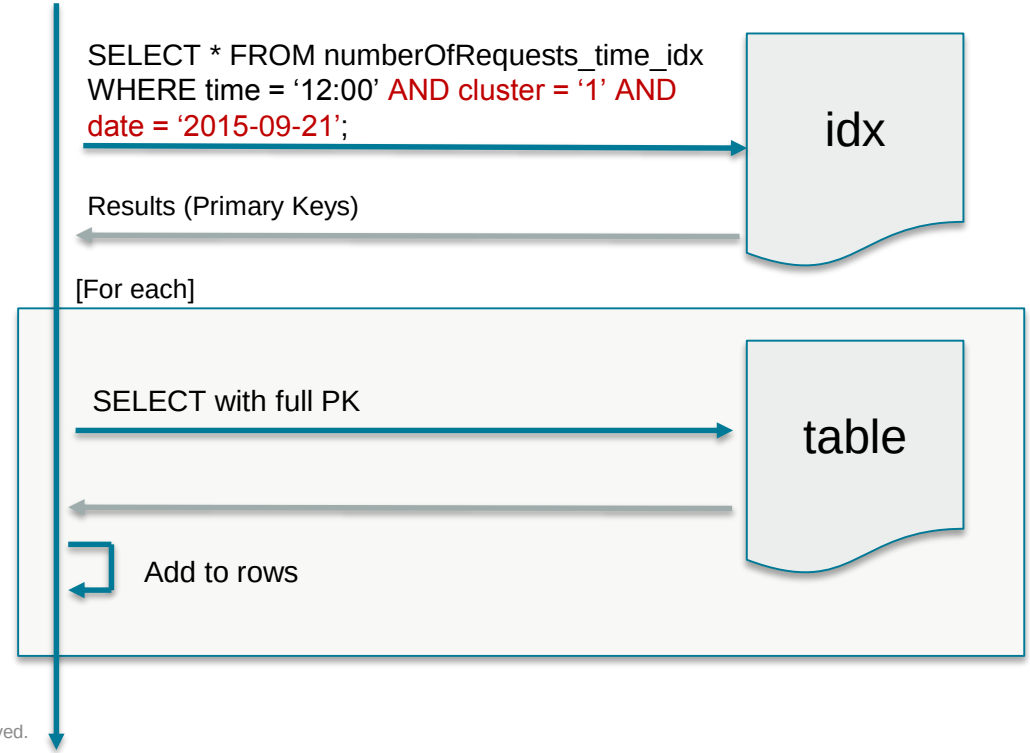
Secondary index queries

```
SELECT * FROM numberOfRequests
WHERE  cluster = ‘cluster 1’ AND date = ‘2015-09-21’AND time = ‘12:00’;
```



Secondary index queries

```
SELECT * FROM numberOfRequests
WHERE  cluster = ‘cluster 1’ AND date = ‘2015-09-21’ AND time = ‘12:00’;
```



UPDATE/DELETE statements

UPDATE statements

CASSANDRA SUMMIT2015

In the UPDATE statements all the primary key columns must be restricted and the only allowed restrictions are:

- Prior to 3.0:
  - Single column = restriction on any partition key or clustering column
  - Single column IN restriction on the last partition key column
- In 3.0:
  - = and IN single column restrictions on any partition key column
  - = and IN single or multi column restrictions on any clustering column

DELETE statements

Before 3.0, in the DELETE statements all the primary key columns must be restricted and the only allowed restrictions were:

- Single column = restriction on any partition key or clustering column
- Single column IN restriction on the last partition key column

DELETE statements

CASSANDRA SUMMIT2015

Since 3.0:

- The partition key columns must be restricted by = or IN restrictions
- A clustering column might not be restricted if none of the following is
- Clustering columns can be restricted by:
  - Single or multi column = restriction
  - Single or multi column IN restriction
  - Single or multi column >, >=, <=, < restriction

Design your tables for the queries you want to perform.

Thank you