Practicing DS201 in Astra

# Exercise 1 – Install and Start Apache Cassandra™

- Using the astra database as a service

<https://astra.datastax.com/org/78d4705b-02c2-4a19-9381-9de9dc42ea39/database/0a53fadd-c405-4349-8c58-cb7979b0da96/connect>

# Exercise 2 – CQL

In this exercise, you will:

• Create a table to store video metadata

• Load the data for the video table

Create a single table called videos. video\_id is the primary key.

*CREATE TABLE videos (*

*video\_id TIMEUUID,*

*added\_date TIMESTAMP,*

*title TEXT,*

*PRIMARY KEY (video\_id)*

*);*

Inserting data into the table videos

*INSERT INTO killrvideo.videos (video\_id, added\_date, title)*

*VALUES (1645ea59-14bd-11e5-a993-8138354b7e31, '2014-01-29', 'Cassandra History');*

*INSERT INTO killrvideo.videos (video\_id, added\_date, title)*

*VALUES (245e8024-14bd-11e5-9743-8238356b7e32, '2012-04-03', 'Cassandra & SSDs');*

*INSERT INTO killrvideo.videos (video\_id, added\_date, title)*

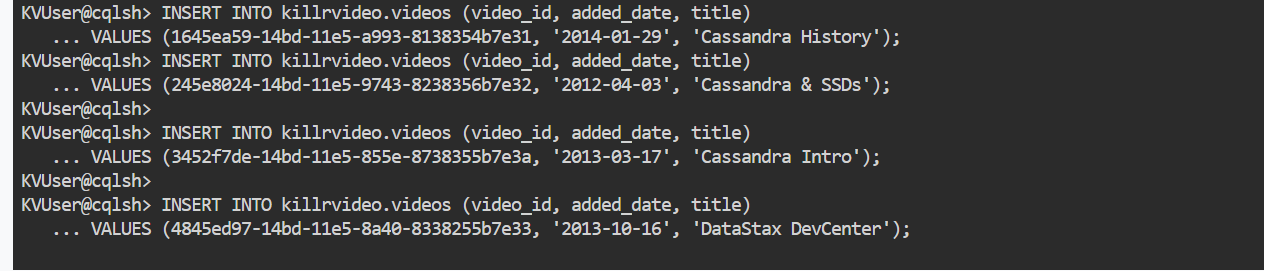
*VALUES (3452f7de-14bd-11e5-855e-8738355b7e3a, '2013-03-17', 'Cassandra Intro');*

*INSERT INTO killrvideo.videos (video\_id, added\_date, title)*

*VALUES (4845ed97-14bd-11e5-8a40-8338255b7e33, '2013-10-16', 'DataStax DevCenter');*

*INSERT INTO killrvideo.videos (video\_id, added\_date, title)*

*VALUES (5645f8bd-14bd-11e5-af1a-8638355b8e3a, '2013-04-16', 'What is DataStax Enterprise?');*



Check for data getting loaded into table

video\_id | added\_date | title

--------------------------------------+---------------------------------+------------------------------

245e8024-14bd-11e5-9743-8238356b7e32 | 2012-04-03 00:00:00.000000+0000 | Cassandra & SSDs

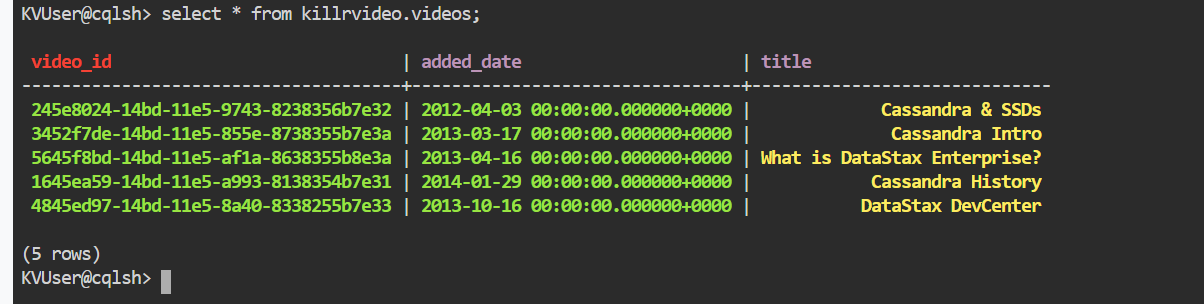
3452f7de-14bd-11e5-855e-8738355b7e3a | 2013-03-17 00:00:00.000000+0000 | Cassandra Intro

5645f8bd-14bd-11e5-af1a-8638355b8e3a | 2013-04-16 00:00:00.000000+0000 | What is DataStax Enterprise?

1645ea59-14bd-11e5-a993-8138354b7e31 | 2014-01-29 00:00:00.000000+0000 | Cassandra History

4845ed97-14bd-11e5-8a40-8338255b7e33 | 2013-10-16 00:00:00.000000+0000 | DataStax DevCenter

(5 rows)



# Exercise 3 – Partitions

In this exercise, you will:

• Experiment with partitions

Execute the following command to view the metadata for the videos table you created earlier.

DESCRIBE TABLE videos;

*KVUser@cqlsh:killrvideo> DESCRIBE TABLE videos;*

*CREATE TABLE killrvideo.videos (*

*video\_id timeuuid PRIMARY KEY,*

*added\_date timestamp,*

*title text*

*) WITH additional\_write\_policy = 'NONE'*

*AND bloom\_filter\_fp\_chance = 0.01*

*AND caching = {'keys': 'ALL', 'rows\_per\_partition': 'NONE'}*

*AND comment = ''*

*AND compaction = {'class': 'org.apache.cassandra.db.compaction.SizeTieredCompactionStrategy', 'max\_threshold': '32', 'min\_threshold': '4'}*

*AND compression = {'enabled': 'false'}*

*AND crc\_check\_chance = 1.0*

*AND default\_time\_to\_live = 0*

*AND gc\_grace\_seconds = 864000*

*AND max\_index\_interval = 2048*

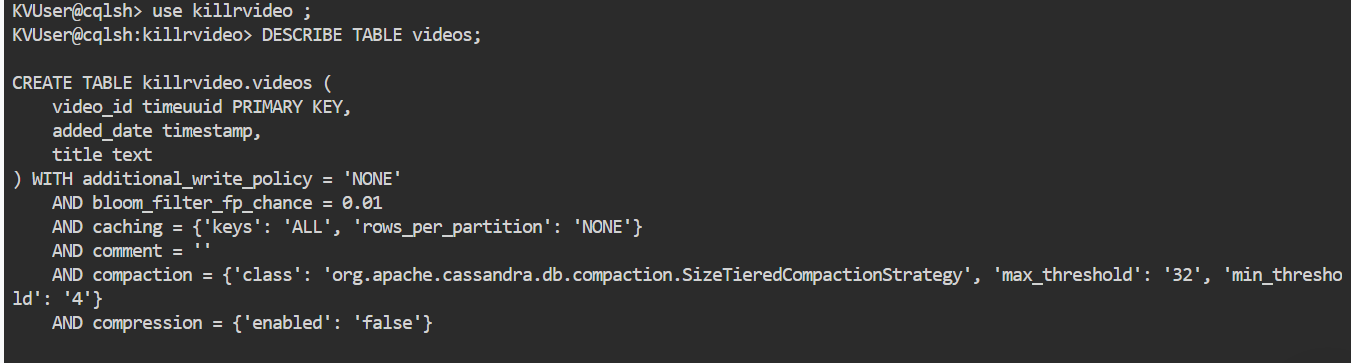
*AND memtable\_flush\_period\_in\_ms = 0*

*AND min\_index\_interval = 128*

*AND nodesync = {'enabled': 'true', 'incremental': 'true'}*

*AND read\_repair = 'BLOCKING'*

*AND speculative\_retry = 'NONE';*



• What is the partition key?

video\_id

• How many partitions are in this table?

One for each unique primary key value.

Execute the following query to view the partitioner token value for each video id.

*SELECT token(video\_id), video\_id*

*FROM killrvideo.videos;*

*system.token(video\_id) | video\_id*

*------------------------+--------------------------------------*

*-7805440677194688247 | 245e8024-14bd-11e5-9743-8238356b7e32*

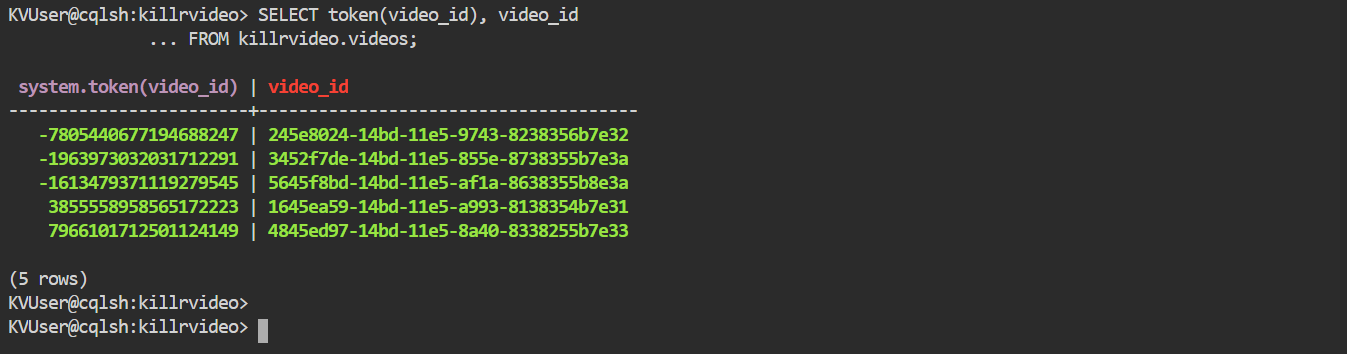
*-1963973032031712291 | 3452f7de-14bd-11e5-855e-8738355b7e3a*

*-1613479371119279545 | 5645f8bd-14bd-11e5-af1a-8638355b8e3a*

*3855558958565172223 | 1645ea59-14bd-11e5-a993-8138354b7e31*

*7966101712501124149 | 4845ed97-14bd-11e5-8a40-8338255b7e33*

*(5 rows)*



Create another table that will store this data partitioned by tags. With this given data set, there should be two partitions, one for each tag. Call your table videos\_by\_tag.

*CREATE TABLE videos\_by\_tag (*

*tag TEXT,*

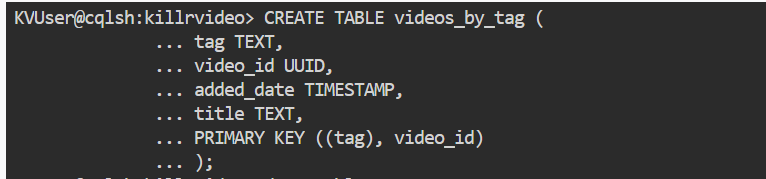
*video\_id UUID,*

*added\_date TIMESTAMP,*

*title TEXT,*

*PRIMARY KEY ((tag), video\_id)*

*);*



Insert data into videos\_by\_tag

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('cassandra',1645ea59-14bd-11e5-a993-8138354b7e31, '2014-01-29', 'Cassandra History');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('cassandra',245e8024-14bd-11e5-9743-8238356b7e32, '2012-04-03', 'Cassandra & SSDs');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

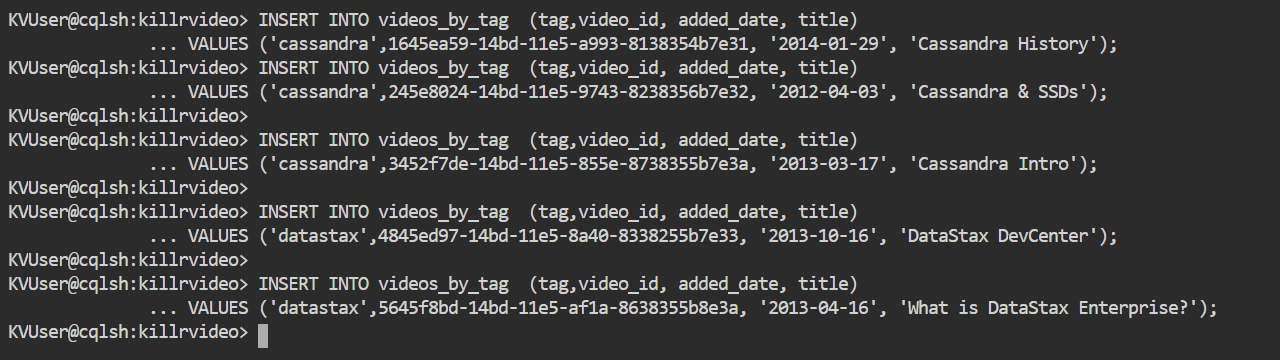
*VALUES ('cassandra',3452f7de-14bd-11e5-855e-8738355b7e3a, '2013-03-17', 'Cassandra Intro');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('datastax',4845ed97-14bd-11e5-8a40-8338255b7e33, '2013-10-16', 'DataStax DevCenter');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('datastax',5645f8bd-14bd-11e5-af1a-8638355b8e3a, '2013-04-16', 'What is DataStax Enterprise?');*



Verify data correctly by writing a SELECT \* command.

KVUser@cqlsh:killrvideo> SELECT \*

... FROM videos\_by\_tag;

tag | video\_id | added\_date | title

-----------+--------------------------------------+---------------------------------+------------------------------

datastax | 4845ed97-14bd-11e5-8a40-8338255b7e33 | 2013-10-16 00:00:00.000000+0000 | DataStax DevCenter

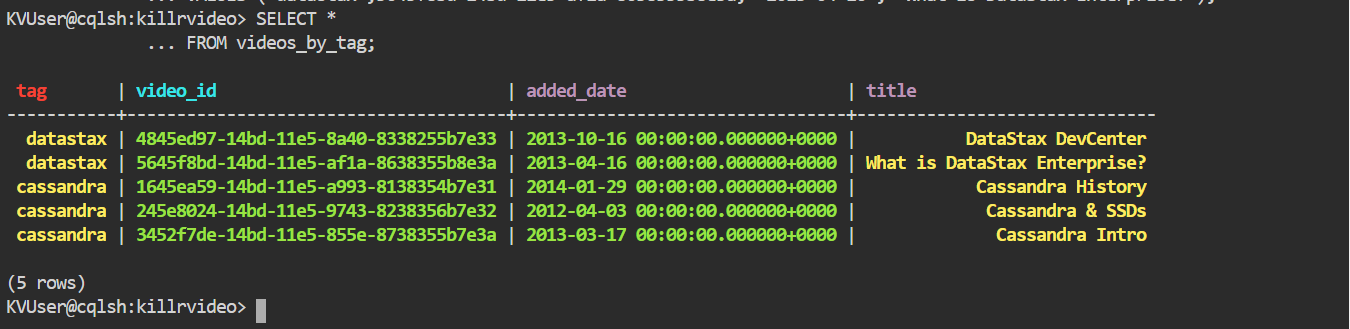
datastax | 5645f8bd-14bd-11e5-af1a-8638355b8e3a | 2013-04-16 00:00:00.000000+0000 | What is DataStax Enterprise?

cassandra | 1645ea59-14bd-11e5-a993-8138354b7e31 | 2014-01-29 00:00:00.000000+0000 | Cassandra History

cassandra | 245e8024-14bd-11e5-9743-8238356b7e32 | 2012-04-03 00:00:00.000000+0000 | Cassandra & SSDs

cassandra | 3452f7de-14bd-11e5-855e-8738355b7e3a | 2013-03-17 00:00:00.000000+0000 | Cassandra Intro

(5 rows)



Write a SELECT statement to retrieve all rows tagged with cassandra.

SELECT \*

FROM videos\_by\_tag

WHERE tag = 'cassandra';

tag | video\_id | added\_date | title

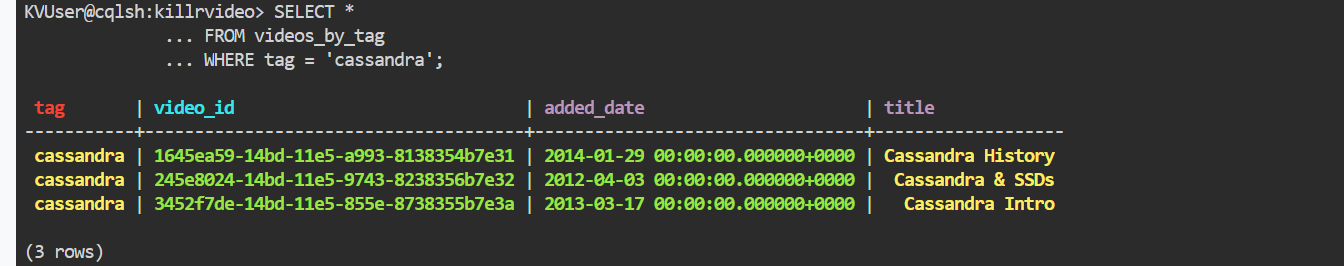
-----------+--------------------------------------+---------------------------------+-------------------

cassandra | 1645ea59-14bd-11e5-a993-8138354b7e31 | 2014-01-29 00:00:00.000000+0000 | Cassandra History

cassandra | 245e8024-14bd-11e5-9743-8238356b7e32 | 2012-04-03 00:00:00.000000+0000 | Cassandra & SSDs

cassandra | 3452f7de-14bd-11e5-855e-8738355b7e3a | 2013-03-17 00:00:00.000000+0000 | Cassandra Intro

(3 rows)



Now, find all videos tagged with datastax (similar to the previous query).

SELECT \*

FROM videos\_by\_tag

WHERE tag = 'datastax';

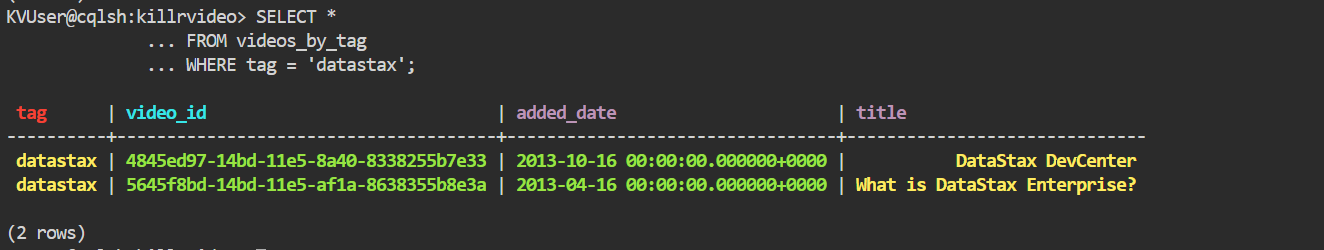
tag | video\_id | added\_date | title

----------+--------------------------------------+---------------------------------+------------------------------

datastax | 4845ed97-14bd-11e5-8a40-8338255b7e33 | 2013-10-16 00:00:00.000000+0000 | DataStax DevCenter

datastax | 5645f8bd-14bd-11e5-af1a-8638355b8e3a | 2013-04-16 00:00:00.000000+0000 | What is DataStax Enterprise?

(2 rows)



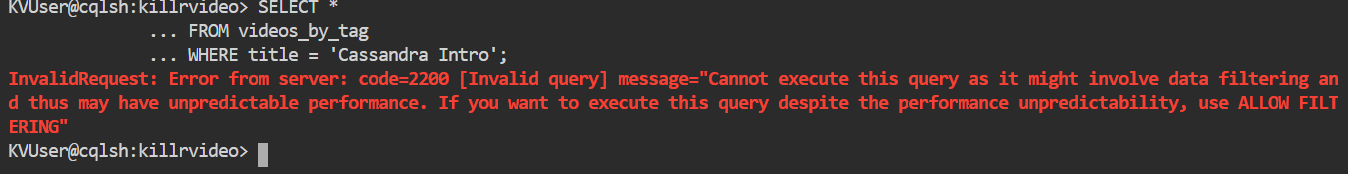
Finally, write a query to retrieve the video having a title of Cassandra Intro.

SELECT \*

FROM videos\_by\_tag

WHERE title = 'Cassandra Intro';

InvalidRequest: Error from server: code=2200 [Invalid query] message="Cannot execute this query as it might involve data filtering and thus may have unpredictable performance. If you want to execute this query despite the performance unpredictability, use ALLOW FILTERING"



# Exercise 4 – Clustering Columns

In this exercise, you will:

•Understand how clustering columns affect the underlying storage mechanism.

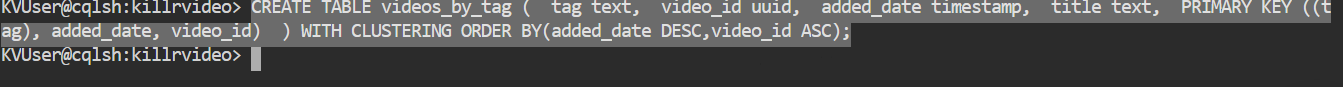
•Understand how clustering columns affect queries.

Modify the following command to create a new videos\_by\_tag table partitioned based on the tag. The table should also store the rows of each partition so that the newest videos are listed first within the partition

*DROP TABLE videos\_by\_tag;*



*CREATE TABLE videos\_by\_tag ( tag text, video\_id uuid, added\_date timestamp, title text, PRIMARY KEY ((tag), added\_date, video\_id) ) WITH CLUSTERING ORDER BY(added\_date DESC,video\_id ASC);*



Insert data into videos\_by\_tag

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('cassandra',1645ea59-14bd-11e5-a993-8138354b7e31, '2014-01-29', 'Cassandra History');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('cassandra',245e8024-14bd-11e5-9743-8238356b7e32, '2012-04-03', 'Cassandra & SSDs');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

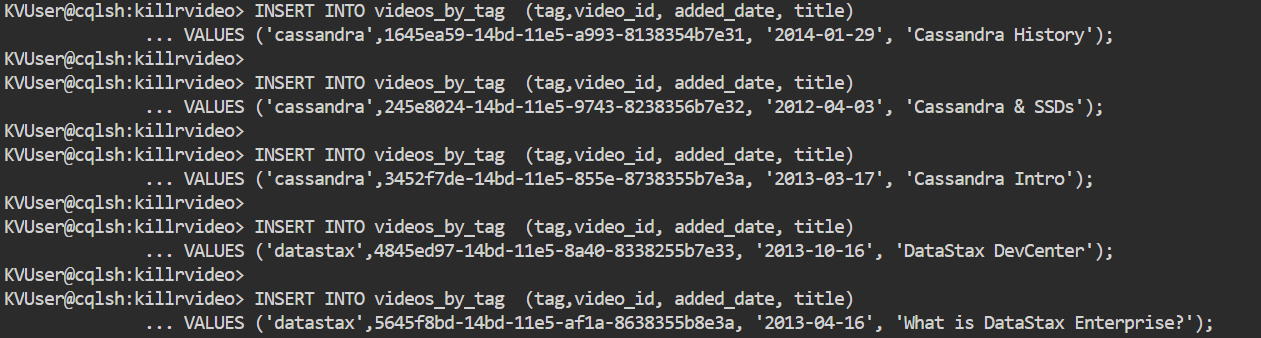
*VALUES ('cassandra',3452f7de-14bd-11e5-855e-8738355b7e3a, '2013-03-17', 'Cassandra Intro');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('datastax',4845ed97-14bd-11e5-8a40-8338255b7e33, '2013-10-16', 'DataStax DevCenter');*

*INSERT INTO videos\_by\_tag (tag,video\_id, added\_date, title)*

*VALUES ('datastax',5645f8bd-14bd-11e5-af1a-8638355b8e3a, '2013-04-16', 'What is DataStax Enterprise?');*



Perform a SELECT \* query on videos\_by\_tag.

SELECT \*

FROM videos\_by\_tag;

tag | added\_date | video\_id | title

-----------+---------------------------------+--------------------------------------+------------------------------

datastax | 2013-10-16 00:00:00.000000+0000 | 4845ed97-14bd-11e5-8a40-8338255b7e33 | DataStax DevCenter

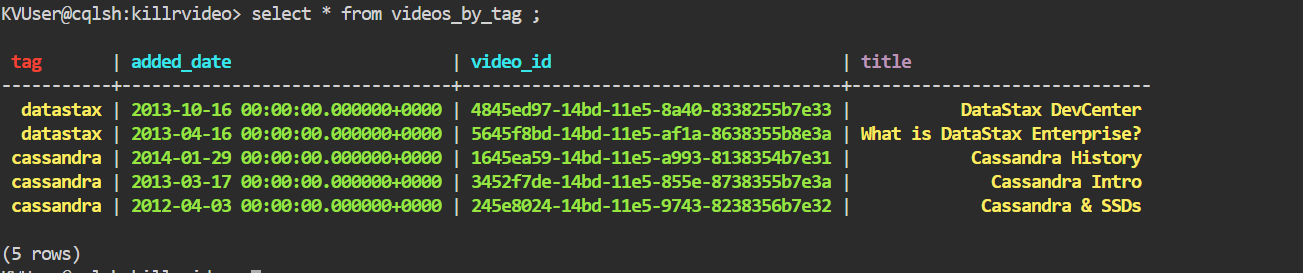
datastax | 2013-04-16 00:00:00.000000+0000 | 5645f8bd-14bd-11e5-af1a-8638355b8e3a | What is DataStax Enterprise?

cassandra | 2014-01-29 00:00:00.000000+0000 | 1645ea59-14bd-11e5-a993-8138354b7e31 | Cassandra History

cassandra | 2013-03-17 00:00:00.000000+0000 | 3452f7de-14bd-11e5-855e-8738355b7e3a | Cassandra Intro

cassandra | 2012-04-03 00:00:00.000000+0000 | 245e8024-14bd-11e5-9743-8238356b7e32 | Cassandra & SSDs

(5 rows)



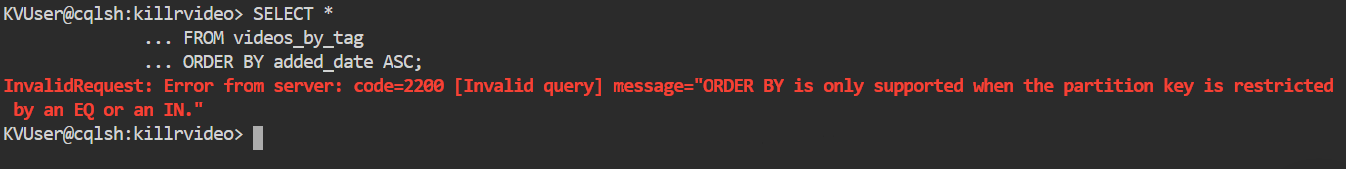
Execute your query again, but list the oldest videos first.

SELECT \*

FROM videos\_by\_tag

ORDER BY added\_date ASC;

InvalidRequest: Error from server: code=2200 [Invalid query] message="ORDER BY is only supported when the partition key is restricted by an EQ or an IN."



Change your query to restrict the partition key value to 'cassandra'.

SELECT \*

FROM videos\_by\_tag

WHERE tag = 'cassandra'

ORDER BY added\_date ASC;

tag | added\_date | video\_id | title

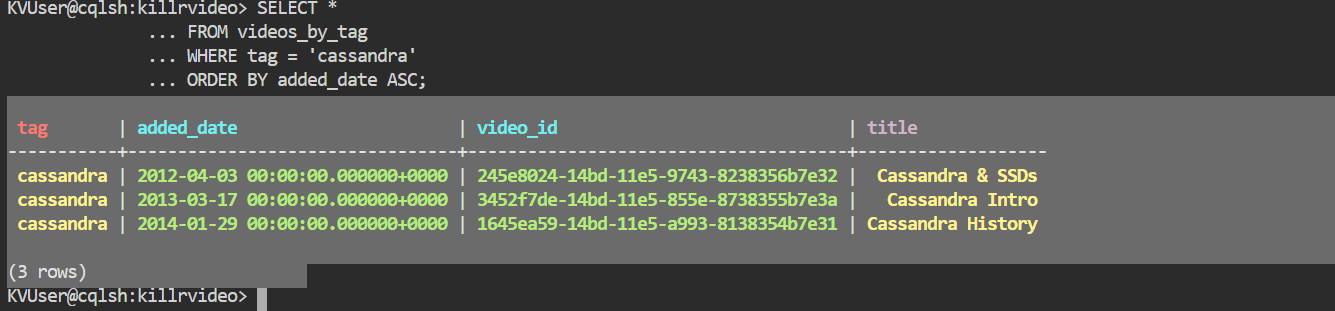
-----------+---------------------------------+--------------------------------------+-------------------

cassandra | 2012-04-03 00:00:00.000000+0000 | 245e8024-14bd-11e5-9743-8238356b7e32 | Cassandra & SSDs

cassandra | 2013-03-17 00:00:00.000000+0000 | 3452f7de-14bd-11e5-855e-8738355b7e3a | Cassandra Intro

cassandra | 2014-01-29 00:00:00.000000+0000 | 1645ea59-14bd-11e5-a993-8138354b7e31 | Cassandra History

(3 rows)



Change your query to retrieve videos made in 2013 or later.

tag | added\_date | video\_id | title

-----------+---------------------------------+--------------------------------------+-------------------

cassandra | 2013-03-17 00:00:00.000000+0000 | 3452f7de-14bd-11e5-855e-8738355b7e3a | Cassandra Intro

cassandra | 2014-01-29 00:00:00.000000+0000 | 1645ea59-14bd-11e5-a993-8138354b7e31 | Cassandra History

(2 rows)

