



Getting started with Apache Cassandra and Apache Zeppelin

DuyHai DOAN – Apache Cassandra evangelist

- | | |
|---|---|
| 1 | Apache Zeppelin (quick) introduction |
| 2 | Cassandra interpreter |
| 3 | Advanced usage |
| 4 | Dynamic forms |
| 5 | Zeppelin with Spark/Cassandra connector |



Apache Zeppelin (quick) intro

What is Apache Zeppelin ?

Apache Zeppelin

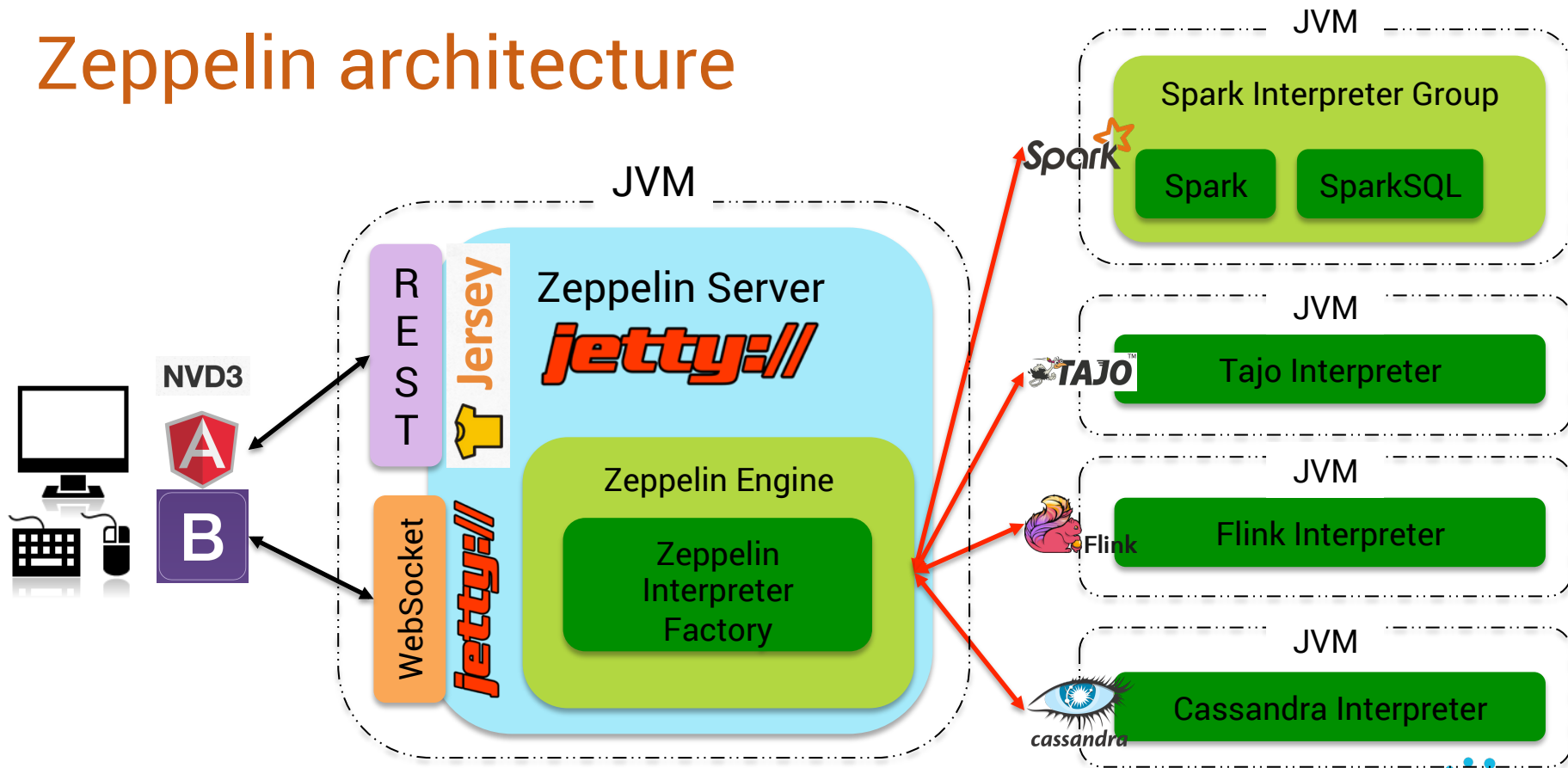
A web-based notebook that enables interactive data analytics.

You can make beautiful data-driven, interactive and collaborative documents with SQL, Scala and more.



Demo

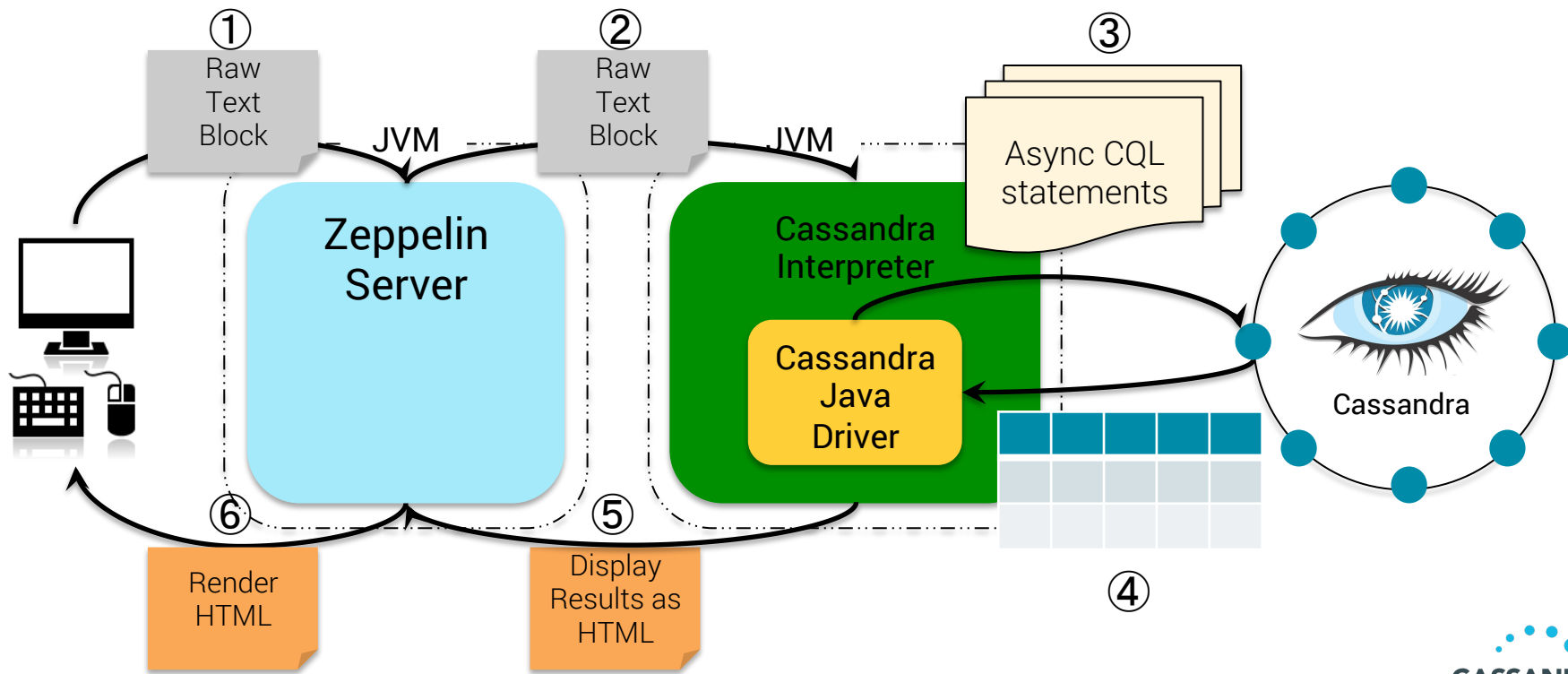
Zeppelin architecture





Cassandra interpreter

Cassandra interpreter architecture





Demo



CASSANDRA
SUMMIT **2016**

Advanced Usage



Demo



CASSANDRA
SUMMIT **2016**

Dynamic Forms



Demo



Zeppelin with Spark/Cassandra connector

Configuration matrix

- Vanilla Zeppelin + OSS Spark
- Vanilla Zeppelin + Datastax Enterprise (DSE)
- Pre-built Zeppelin + OSS Spark
- Pre-built Zeppelin + Datastax Enterprise (DSE)

Vanilla Zeppelin + OSS Spark

First add Spark/Cassandra connector as dependency for Spark interpreter in Zeppelin

Dependencies

These dependencies will be added to classpath when interpreter process starts.

artifact

`org.apache.commons:commons-csv:1.1`

`com.datastax.spark:spark-cassandra-connector_2.10:1.5.0`

Vanilla Zeppelin + OSS Spark

Then configure Spark to include the connector dependencies: `$SPARK_HOME/conf/spark-env.sh`

Either

```
export SPARK_CLASSPATH=/xxx/spark-cassandra-connector-assembly-1.5.0.jar
```

Or

```
export SPARK_SUBMIT_OPTIONS="--package com.datastax.spark:spark-cassandra-connector_2.10:1.5.1"
```

Vanilla Zeppelin + DSE

Only need to add Spark/Cassandra connector as dependency for Spark interpreter in Zeppelin

Dependencies

These dependencies will be added to classpath when interpreter process starts.

artifact

org.apache.commons:commons-csv:1.1

com.datastax.spark:spark-cassandra-connector_2.10:1.5.0

Pre-built Zeppelin + OSS Spark

Only need to configure Spark to include the connector dependencies:

`$SPARK_HOME/conf/spark-env.sh`

Either

```
export SPARK_CLASSPATH=/xxx/spark-cassandra-connector-assembly-1.5.0.jar
```

Or

```
export SPARK_SUBMIT_OPTIONS="--package com.datastax.spark:spark-cassandra-connector_2.10:1.5.1"
```

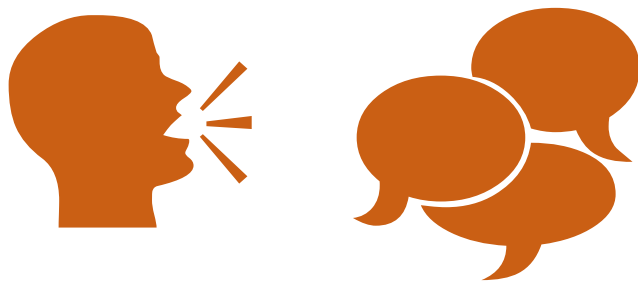
Pre-built Zeppelin + DSE

Nothing to do ...

Where to get pre-built Zeppelin ?

<http://goo.gl/8S4Vv6>

Zeppelin version	Spark version/DSE version	Spark-Cassandra connector version	Tarball
0.6.0	Spark 1.4.0	1.4.4	zeppelin-0.6.0-cassandra-spark-1.4.0.tar.gz
0.6.0	Spark 1.4.1	1.4.4	zeppelin-0.6.0-cassandra-spark-1.4.1.tar.gz
0.6.0	Spark 1.5.0	1.5.1	zeppelin-0.6.0-cassandra-spark-1.5.0.tar.gz
0.6.0	Spark 1.5.1	1.5.1	zeppelin-0.6.0-cassandra-spark-1.5.1.tar.gz
0.6.0	Spark 1.5.2	1.5.1	zeppelin-0.6.0-cassandra-spark-1.5.2.tar.gz
0.6.0	Spark 1.6.0	1.6.0	zeppelin-0.6.0-cassandra-spark-1.6.0.tar.gz
0.6.0	Spark 1.6.1	1.6.0	zeppelin-0.6.0-cassandra-spark-1.6.1.tar.gz
0.6.0	Spark 1.6.2	1.6.0	zeppelin-0.6.0-cassandra-spark-1.6.2.tar.gz
0.6.0	DSE 4.8.3, DSE 4.8.4 (Spark 1.4.1)	1.4.1	zeppelin-0.6.0-dse-4.8.3-4.8.4.tar.gz
0.6.0	DSE 4.8.5, DSE 4.8.6 (Spark 1.4.1)	1.4.2	zeppelin-0.6.0-dse-4.8.5-4.8.6.tar.gz
0.6.0	DSE 4.8.7 (Spark 1.4.1)	1.4.3	zeppelin-0.6.0-dse-4.8.7.tar.gz
0.6.0	DSE 4.8.8, DSE 4.8.9 (Spark 1.4.1)	1.4.4	zeppelin-0.6.0-dse-4.8.8-4.8.9.tar.gz
0.6.0	DSE 5.0.0, DSE 5.0.1 (Spark 1.6.1)	1.6.0	zeppelin-0.6.0-dse-5.0.0-5.0.1.tar.gz



Q & A

Thank You



@doanduyhai



duy_hai.doan@datastax.com

<https://academy.datastax.com/>