

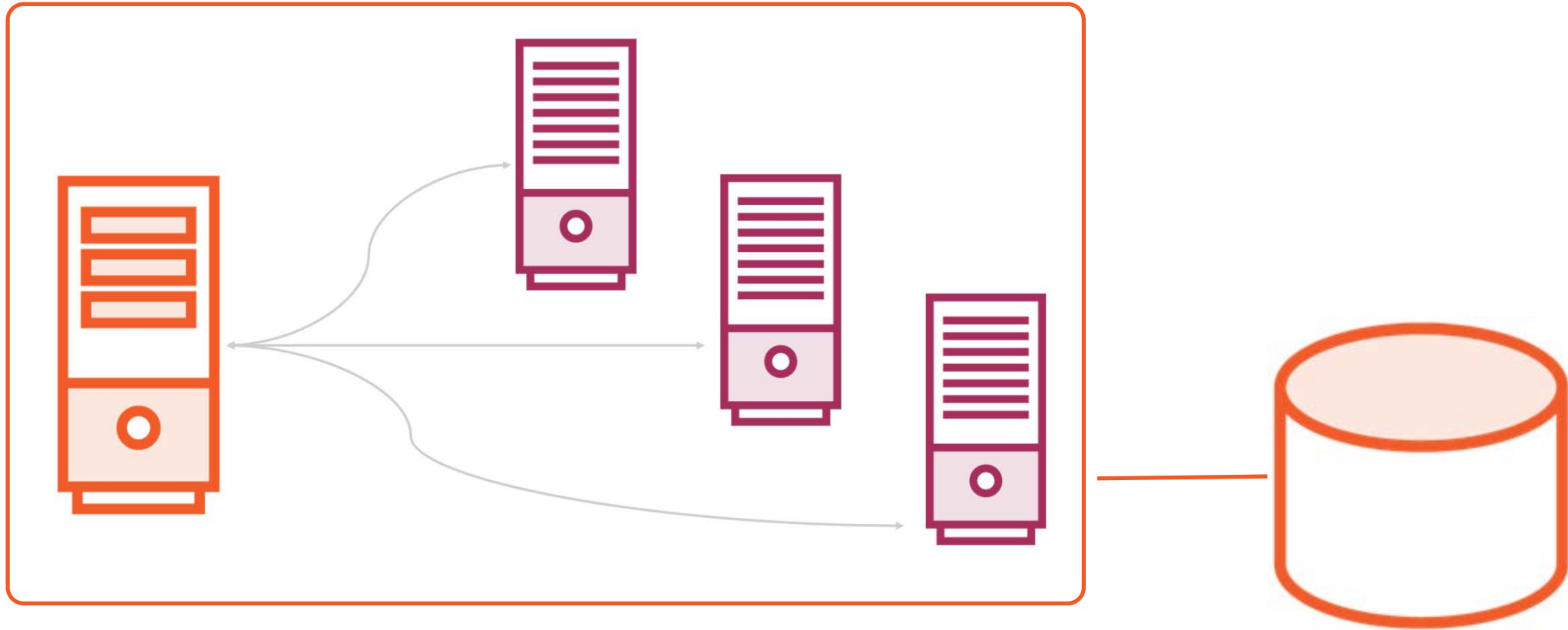
Navigating the Hadoop Ecosystem



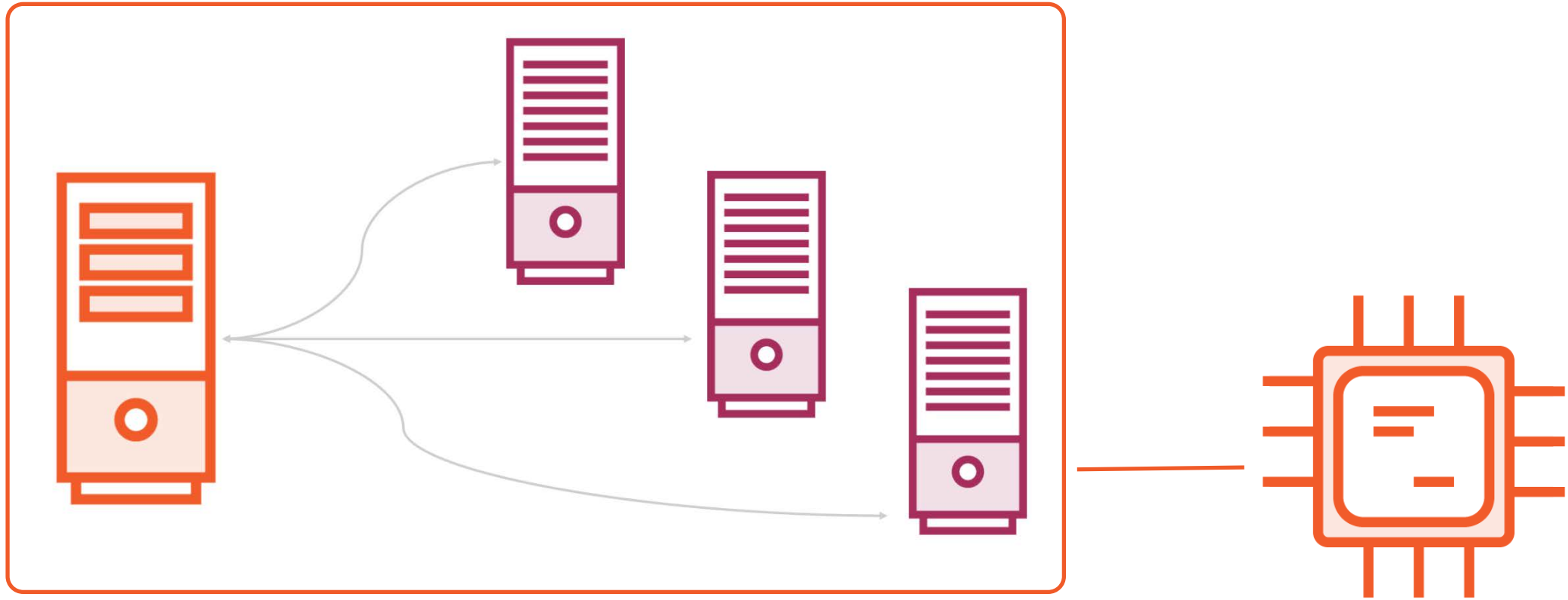
Elton Stoneman

SOFTWARE ARCHITECT

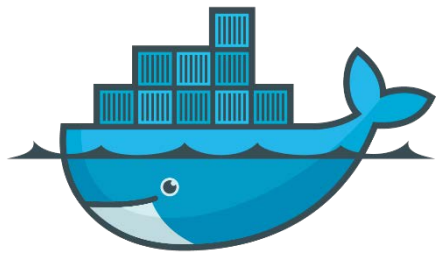
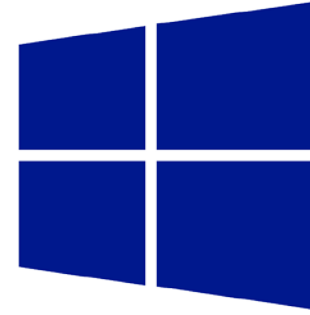
@EltonStoneman blog.sixeyed.com



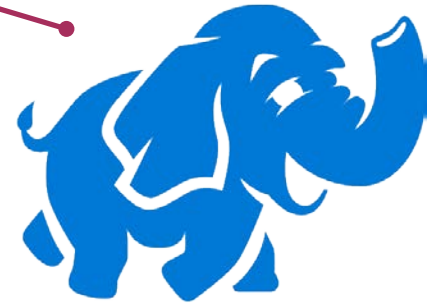
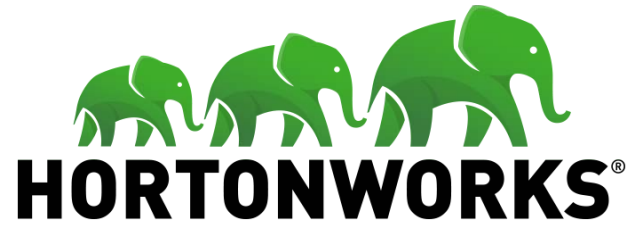
Hadoop Distributed File System

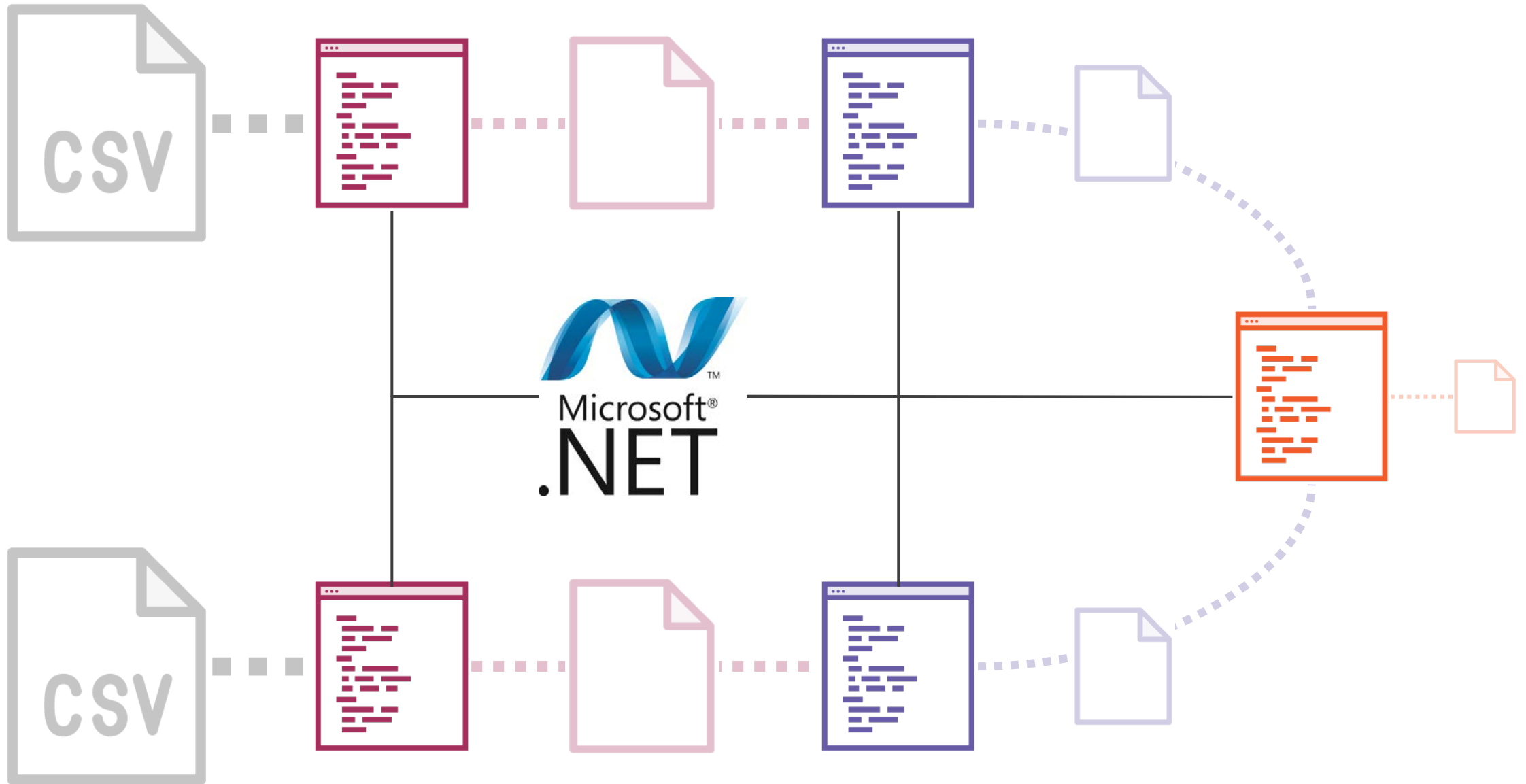


Yet Another Resource Negotiator



docker







Apache Ranger



Apache Ambari



Hive
SQL Over Hadoop

Impala
Alternative SQL Engine

Spark
In-Memory Access

Pig
Scripted Query

Storm
Event Streaming

Oozie
Workflow

Mahout
Machine Learning

Solr
Search

HBase
Real-time No SQL

YARN

Yet Another Resource Negotiator

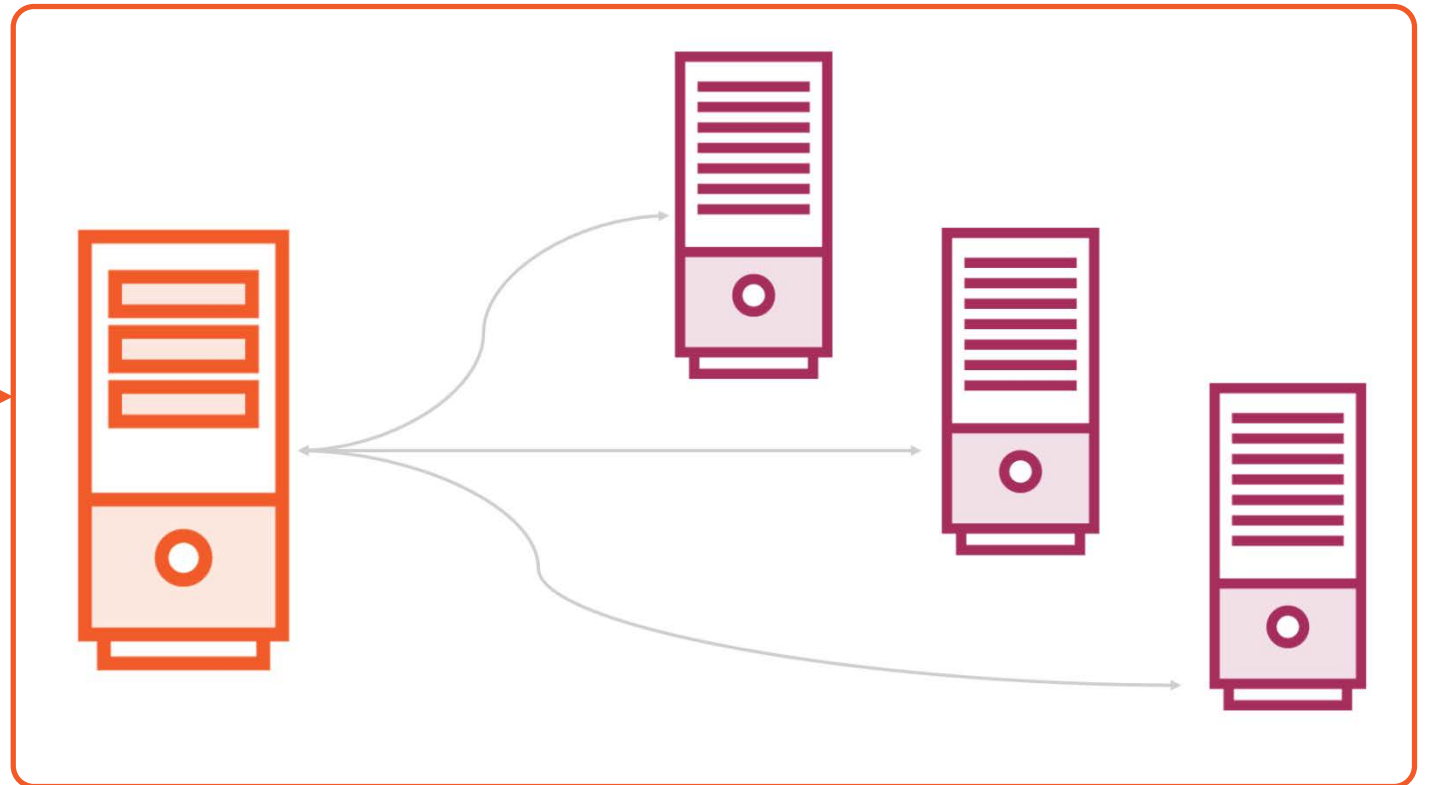
HDFS

Hadoop Distributed File System

Hive

SQL Over Hadoop

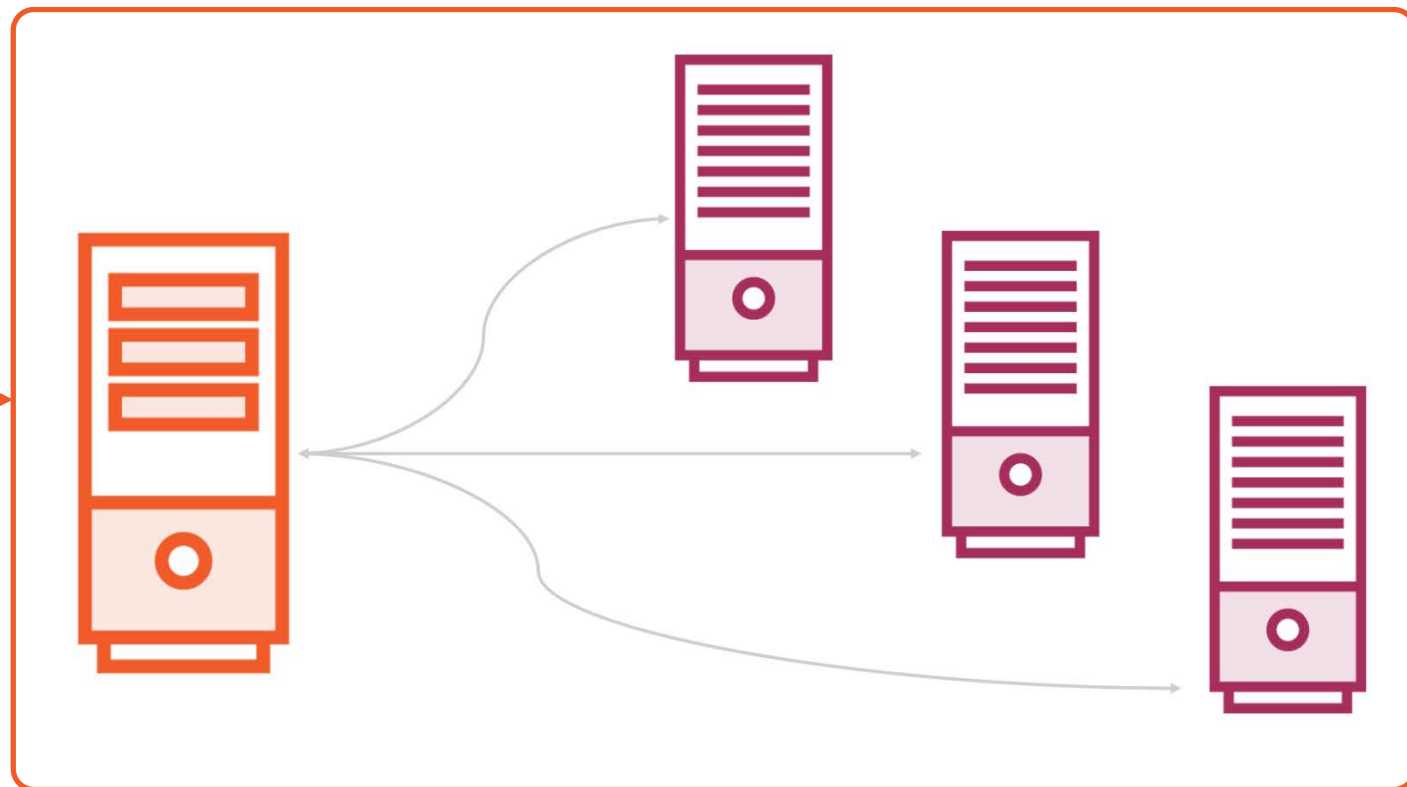
```
SELECT *  
FROM Postcodes
```



HBase

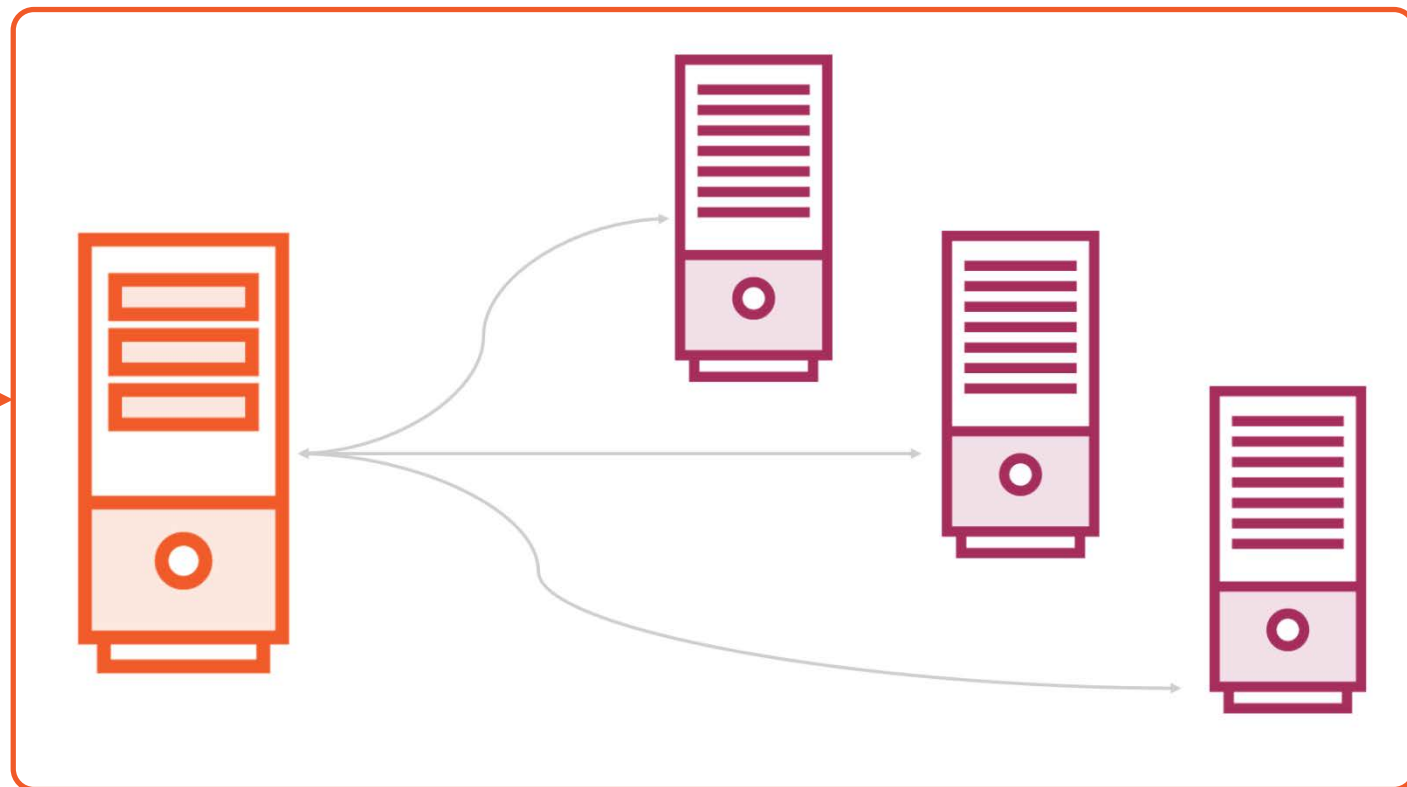
Real-time No SQL

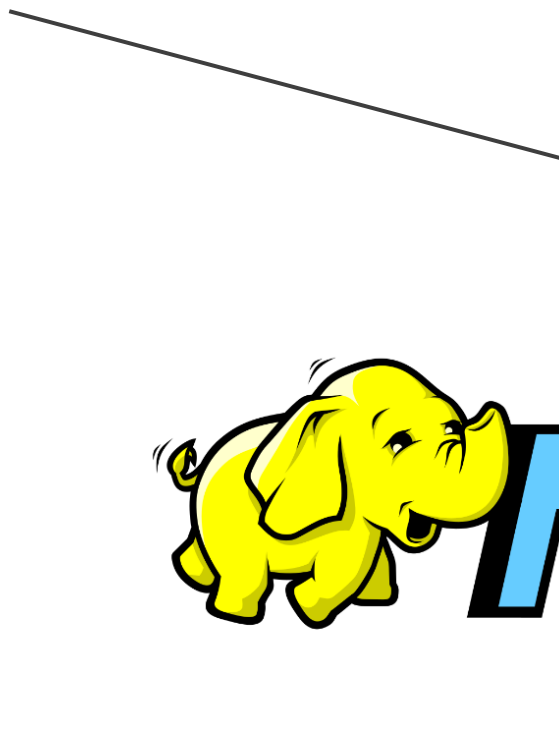
```
get 'PostcodesRT',  
    'ML6 6PY', 'ac:code'
```

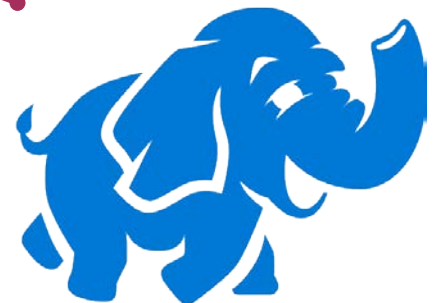
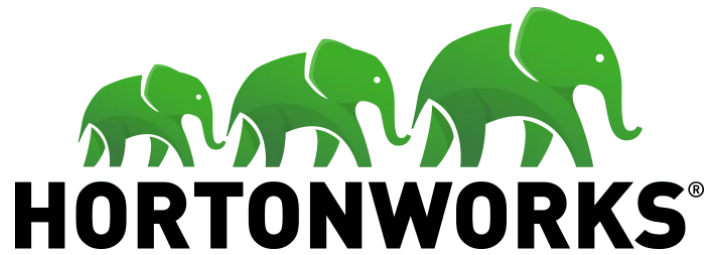


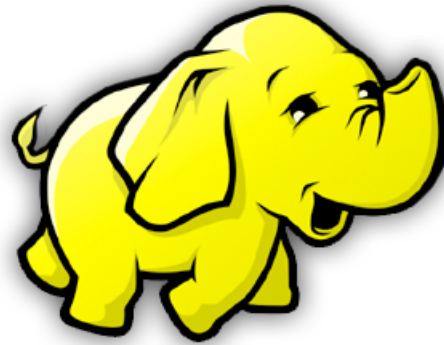
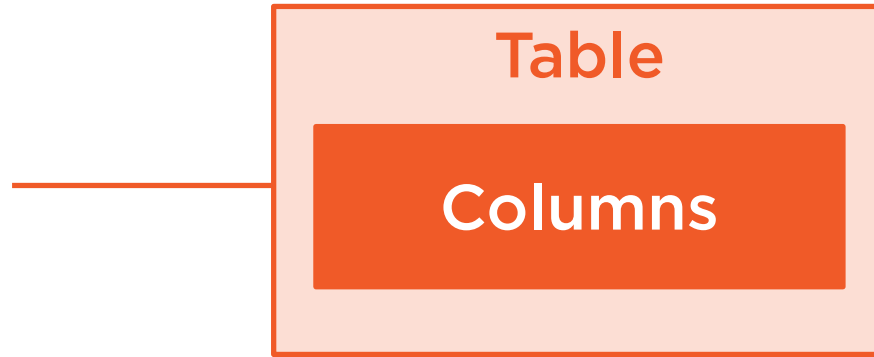
Spark In-Memory Access

```
acs.cache()
```







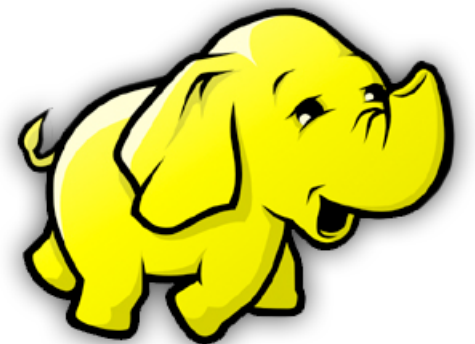


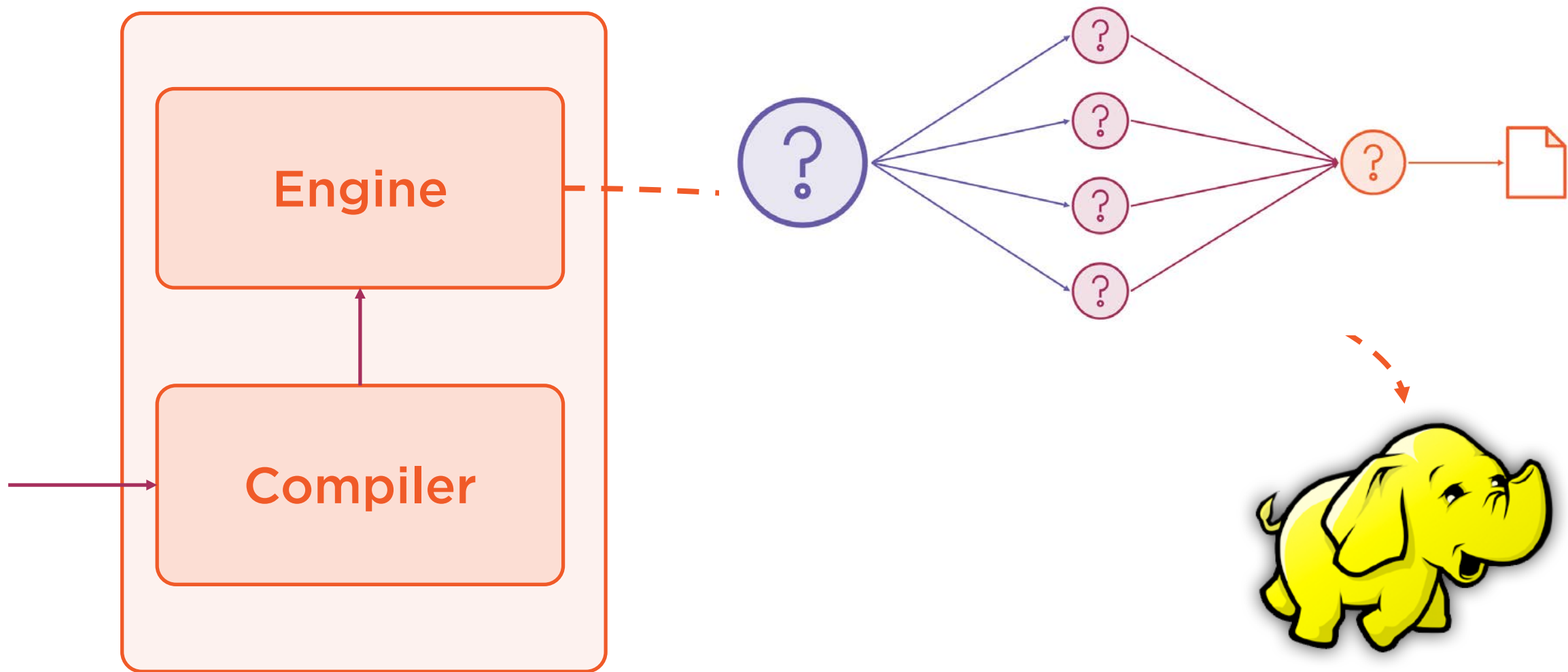
```
SELECT [Columns]  
FROM [Table]
```



Engine

Compiler





```
CREATE EXTERNAL TABLE Postcodes  
( Postcode STRING,  
  Classification STRING...
```

```
SELECT Classification  
FROM Postcodes  
WHERE Postcode = 'ML6 6PY'
```



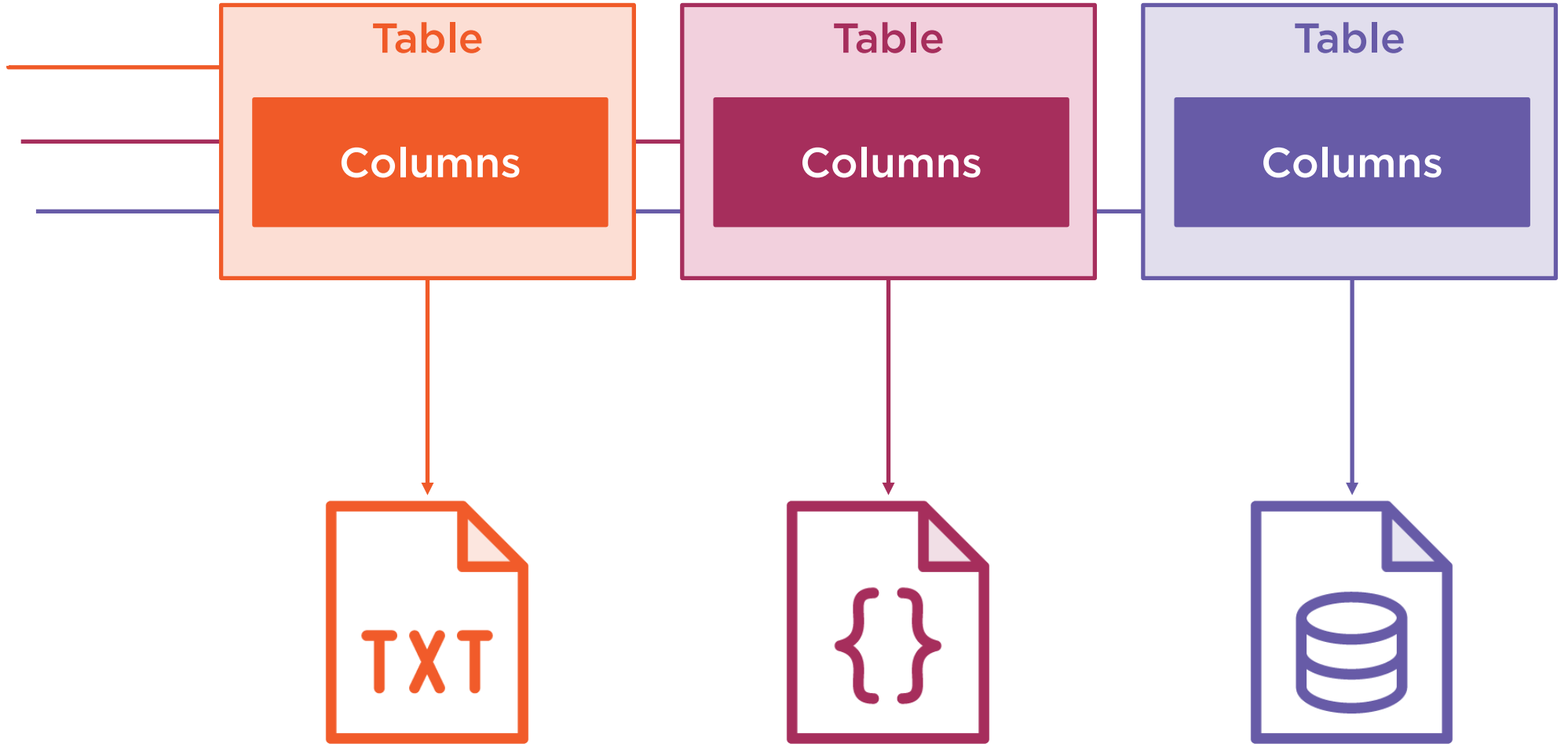

```
CREATE EXTERNAL TABLE Postcodes  
( Postcode STRING,  
  Classification STRING...
```



```
SELECT Classification  
FROM Postcodes  
WHERE Postcode = 'ML6 6PY'
```

Schema Mappings

Technical role

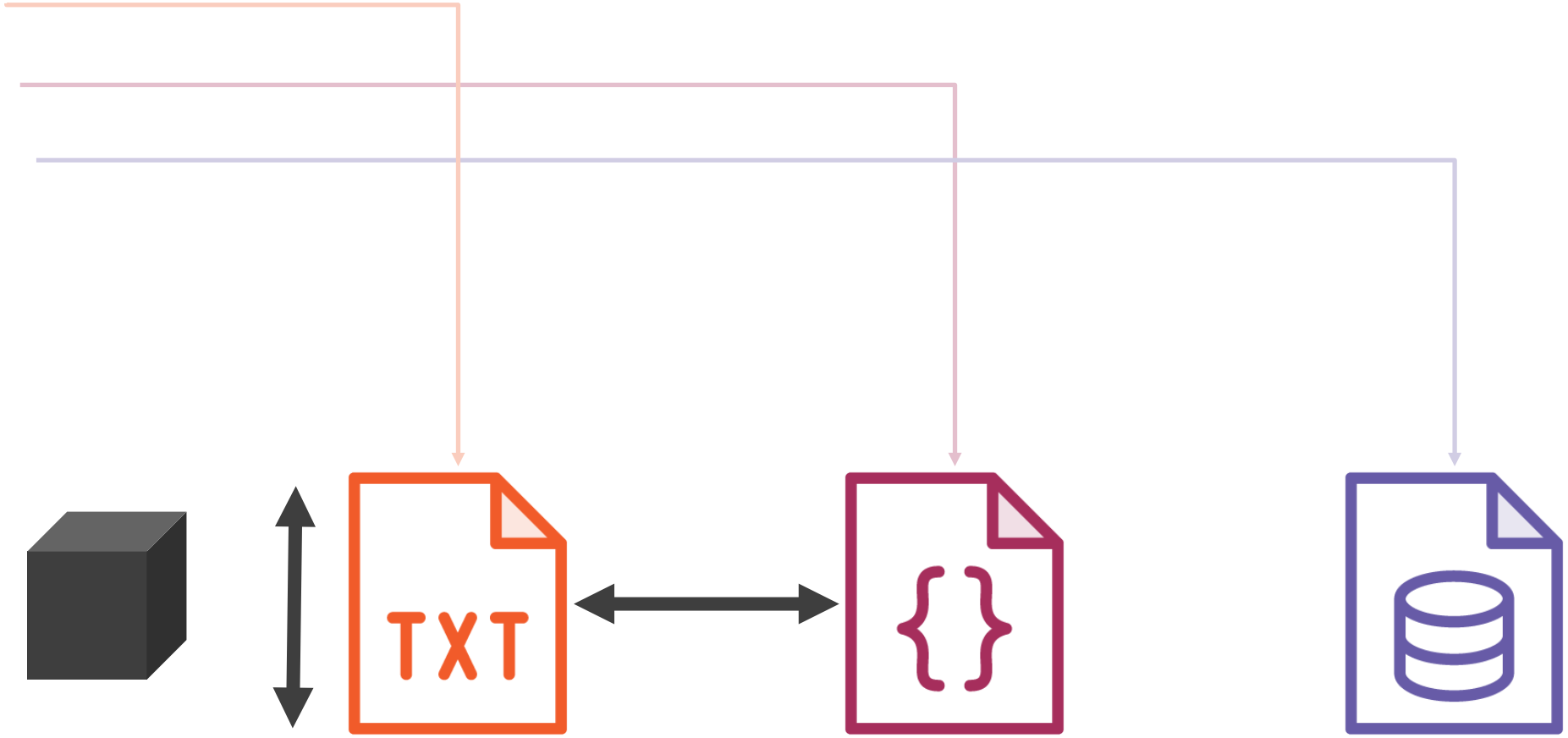


```
CREATE EXTERNAL TABLE Postcodes  
( Postcode STRING,  
  Classification STRING...
```



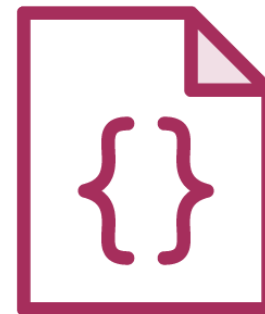
```
SELECT Classification  
FROM Postcodes  
WHERE Postcode = 'ML6 6PY'
```



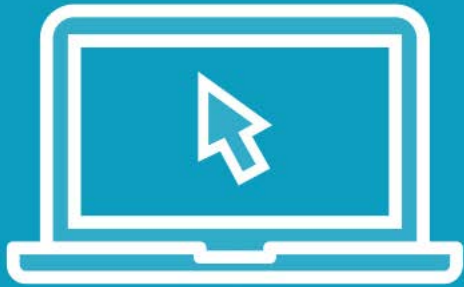




TRANSFORM...
USING...



Demo



Querying Hadoop with Hive

- Mapping columns from CSV
- Running HiveQL queries
- Integrating .NET Core apps

```
CREATE EXTERNAL TABLE Postcodes  
( Postcode STRING,  
  Classification STRING...
```



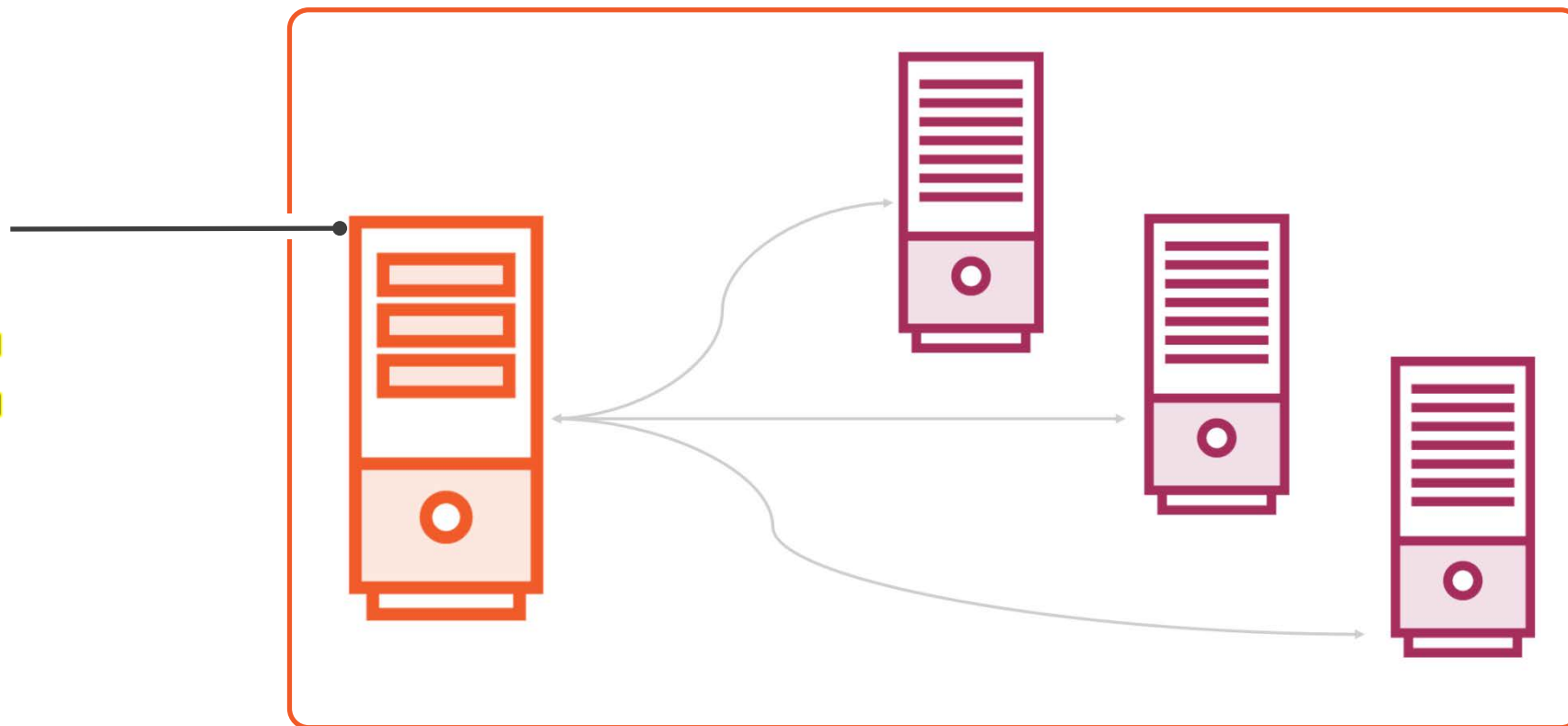
```
SELECT Classification  
FROM Postcodes  
WHERE Postcode = 'ML6 6PY'
```

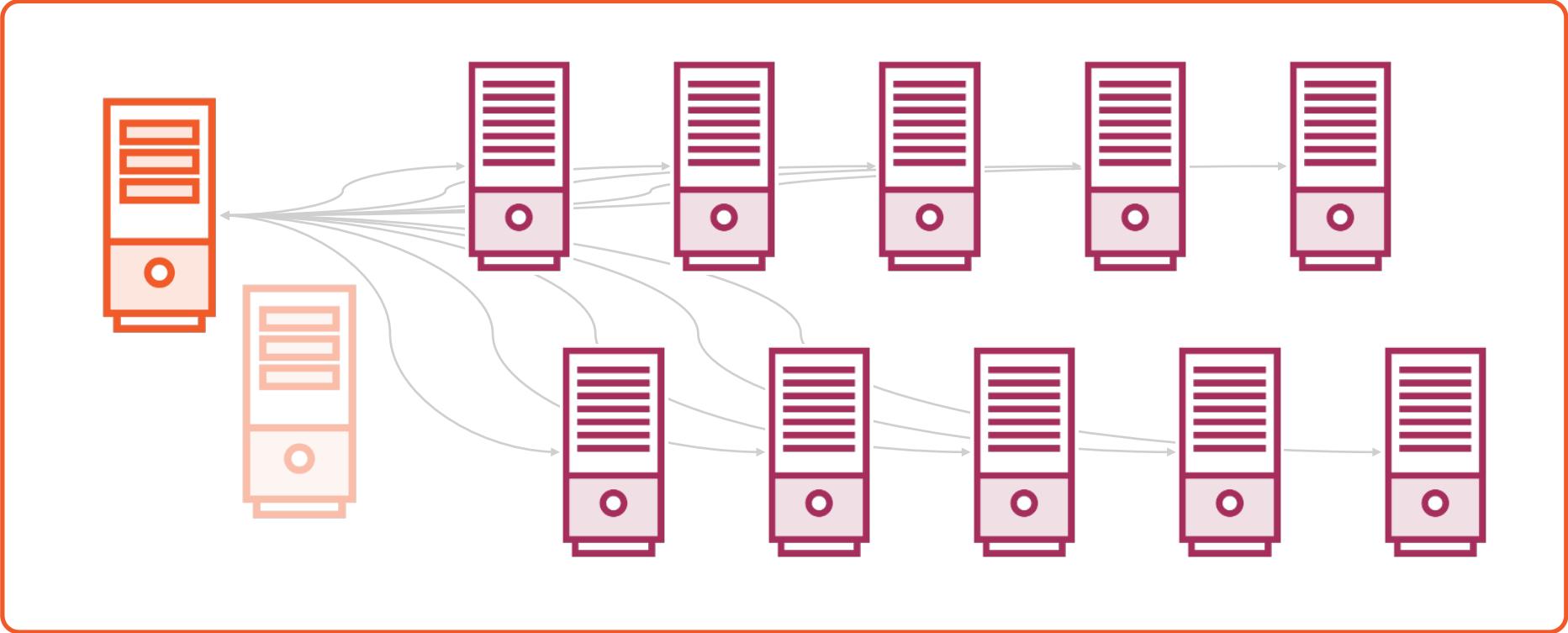


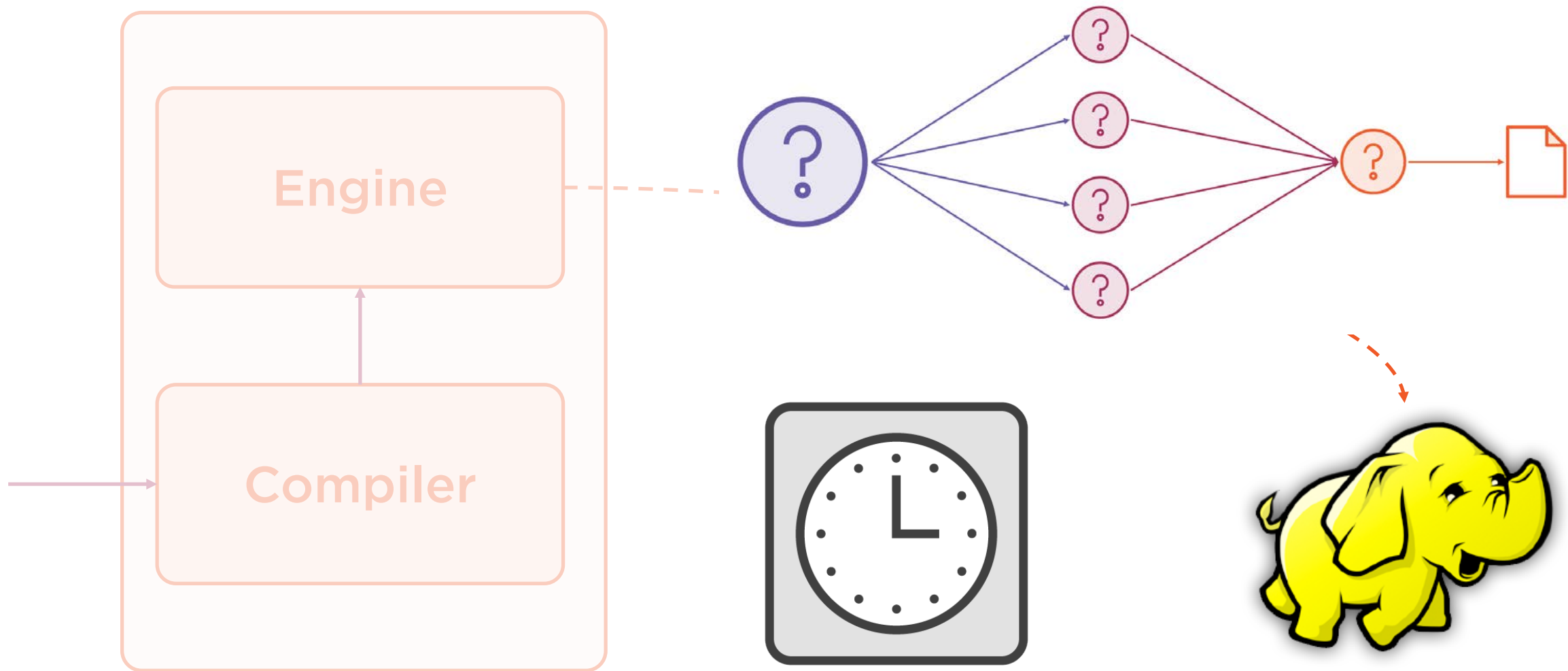


```
SELECT Postcode, Classification  
FROM Postcodes  
ORDER BY Postcode
```

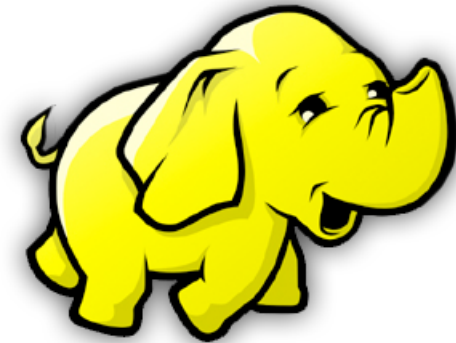
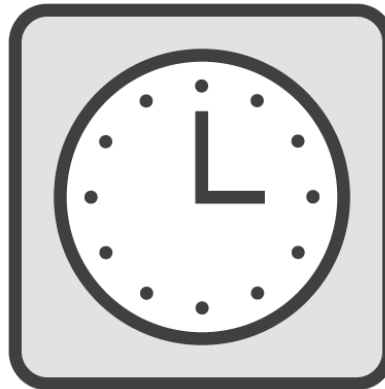




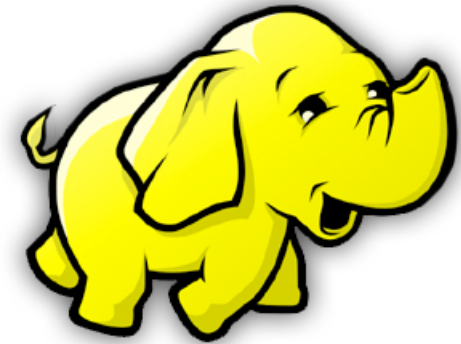




```
SELECT Classification  
FROM Postcodes  
WHERE Postcode = 'ML6 6PY'
```



```
get 'postcodes', 'ML6 6PY'
```



APACHE
HBASE



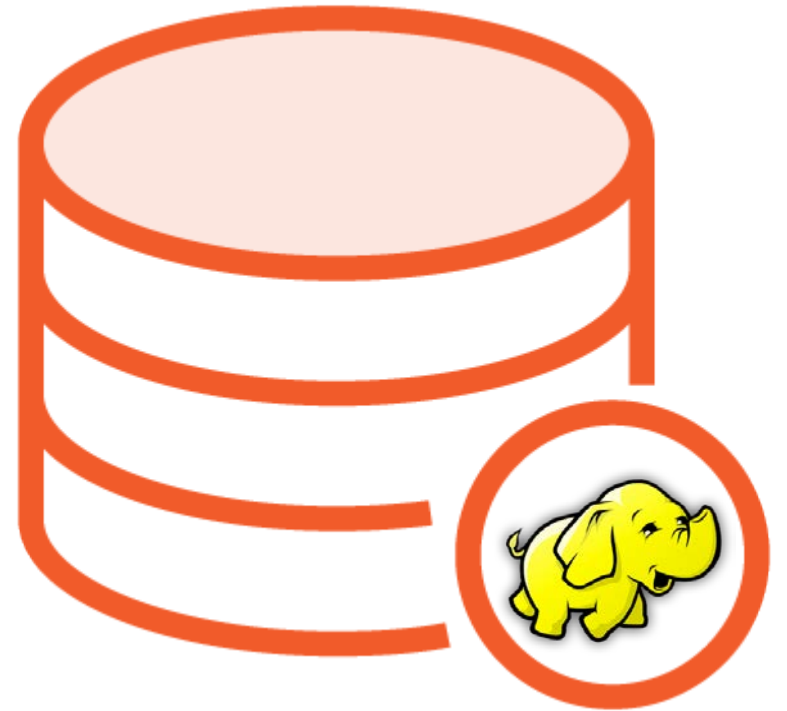
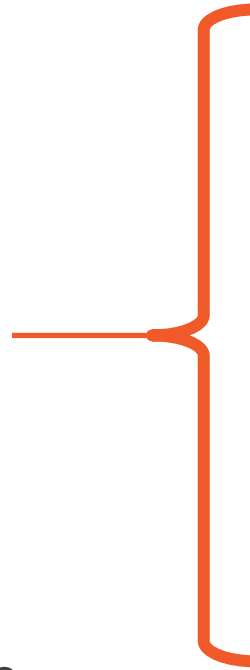
Master
Server



Zookeeper



Region
Server



postcodes	ac
------------------	-----------

RowKey	ac:code	ac:desc
ML6 6PY	6B4	Older workers and retirement
RowKey	ac:code	
ML7 1AB	6B4	
RowKey		
ML8 0AY		

postcodes	ac	
RowKey	ac:code	ac:desc
ML6 6PY	6B4	Older workers and retirement
RowKey	ac:code	
ML7 1AB	6B4	
RowKey		
ML8 0AY		

Column Family

Zero+ columns

postcode s	ac	
RowKey	ac:code	ac:desc
ML6 6PY	6B4	Older workers and retirement
RowKey	ac:code	
ML7 1AB	6B4	
RowKey		
ML8 0AY		

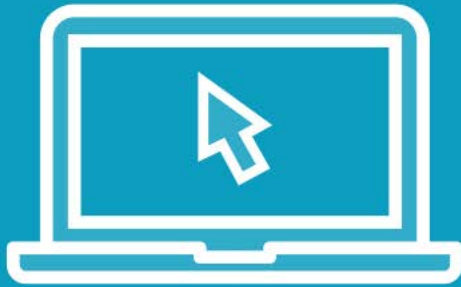
Row Key
Unique ID

postcode s	ac	
RowKey	ac:code	ac:desc
ML6 6PY	6B4	Older workers and retirement
RowKey	ac:code	
ML7 1AB	6B4	
RowKey		
ML8 0AY		

<1s

get 'postcodes', 'ML6 6PY'

Demo



Real-time Big Data with HBase

- Using the HBase command line
- Mapping HBase tables in Hive
- Loading into HBase from Hadoop

APACHE
HBASE



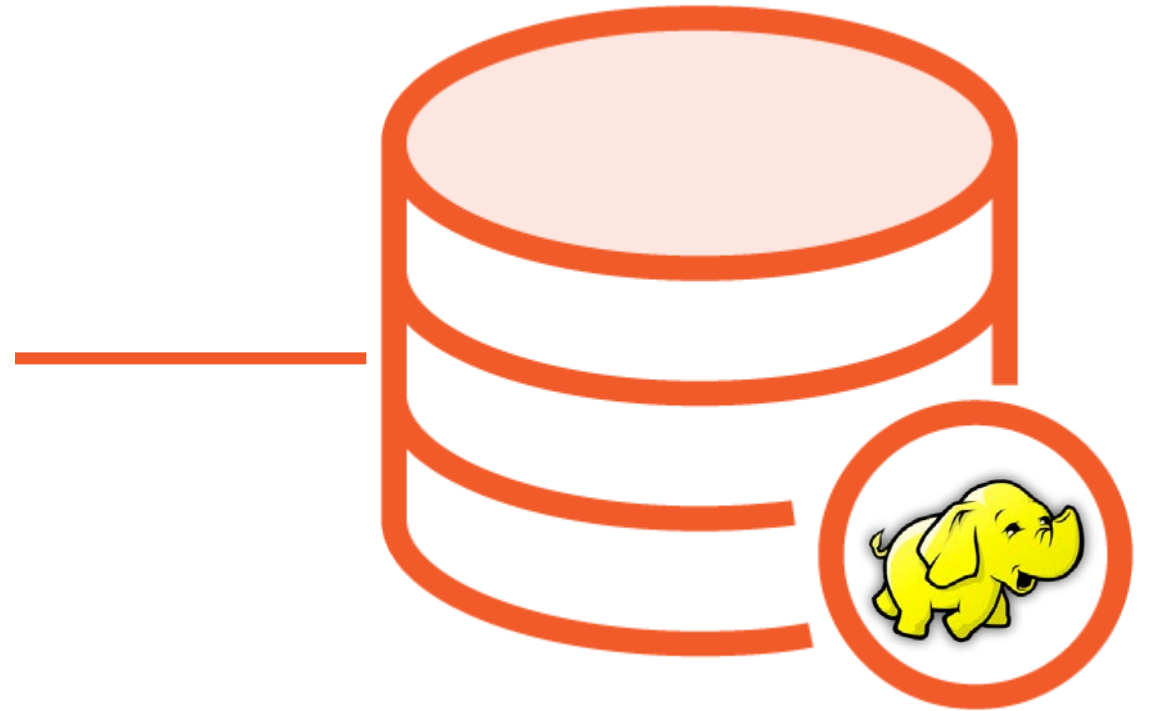
Master
Server



Zookeeper



Region
Server

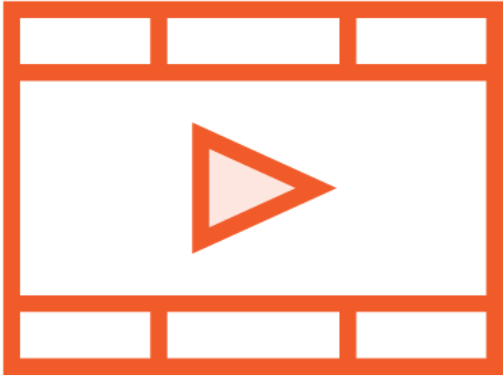


postcode s	ac
-----------------------	-----------

RowKey	ac:code	ac:desc
ML6 6PY	6B4	Older workers and retirement
RowKey	ac:code	
ML7 1AB	6B4	

x BILLIONS

x THOUSANDS

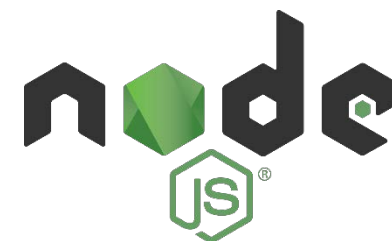
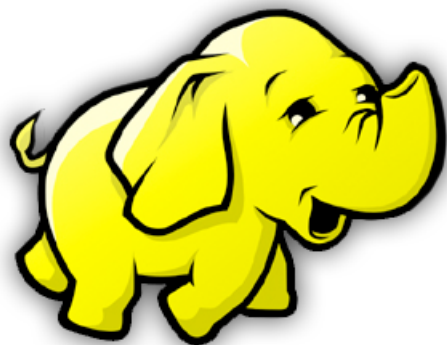


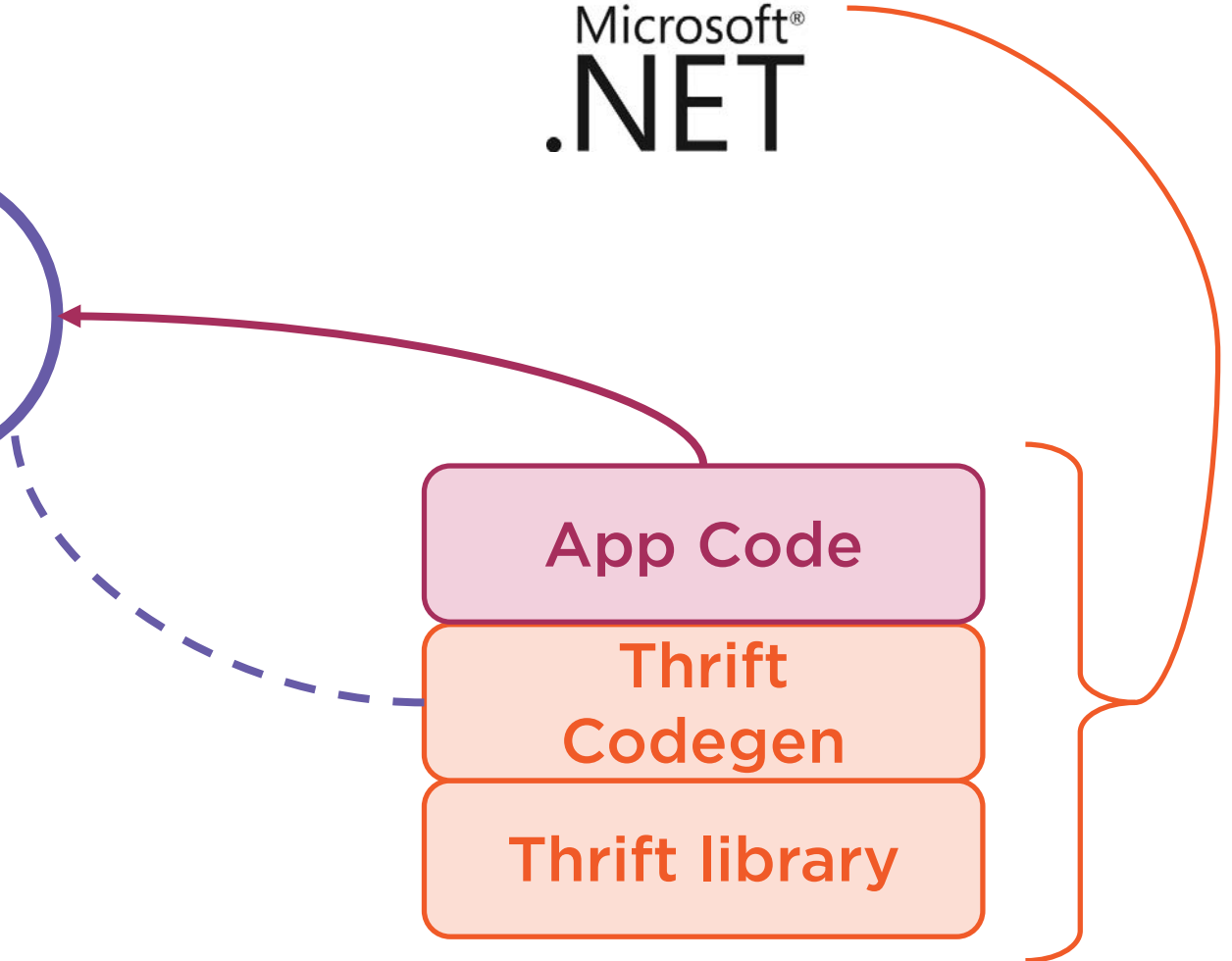
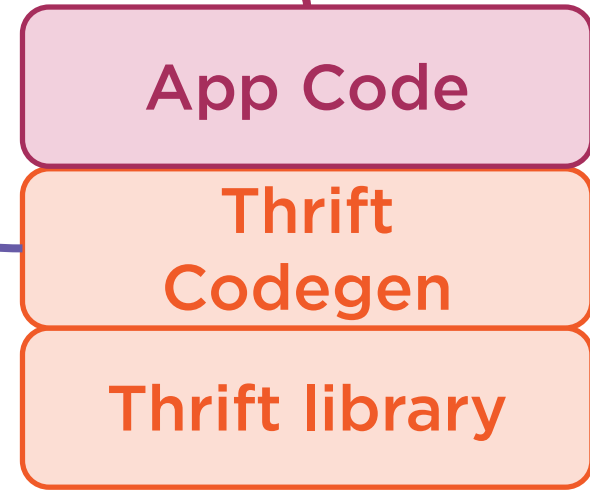
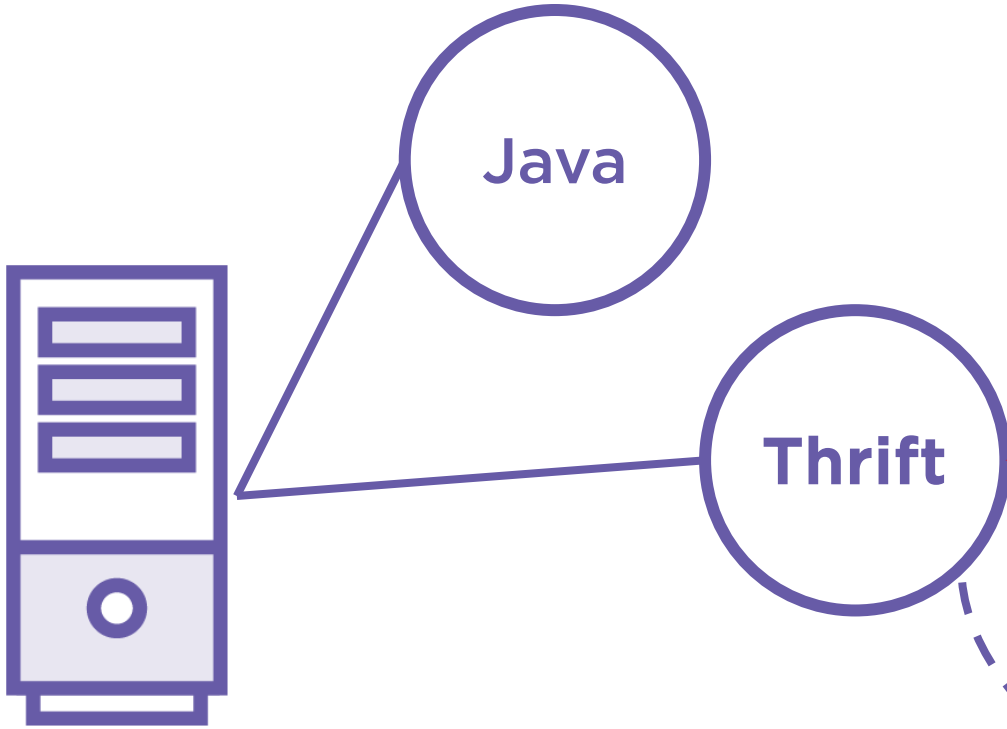
HDInsight Deep Dive: Storm, HBase and Hive



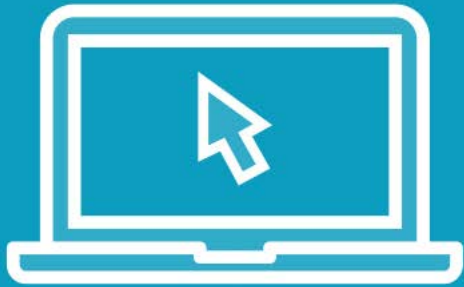
Elton Stoneman

@EltonStoneman blog.sixeyed.com



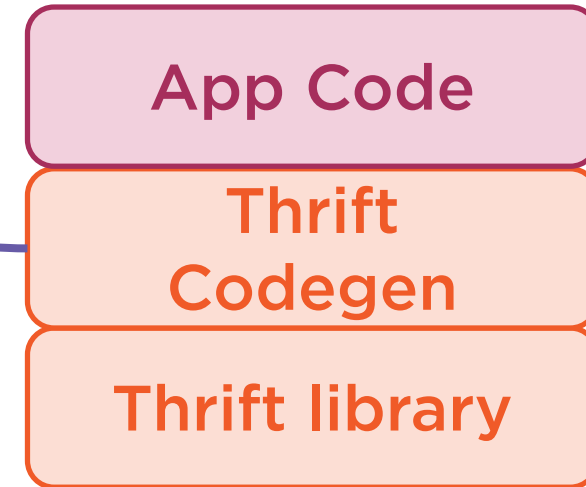
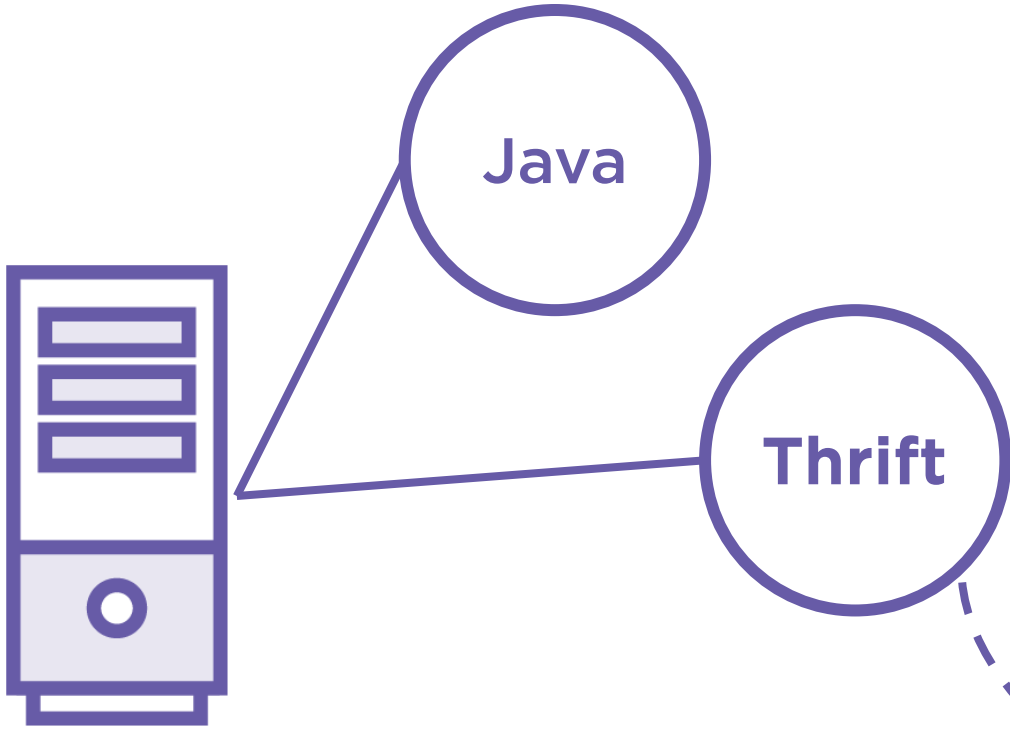


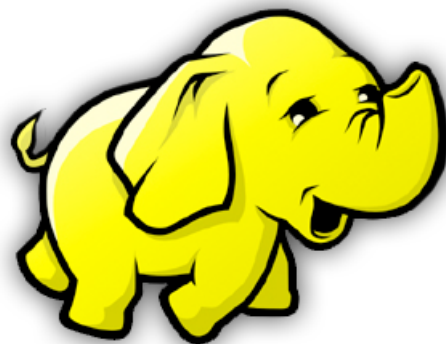
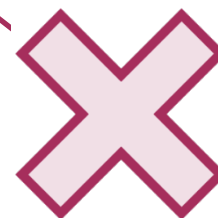
Demo

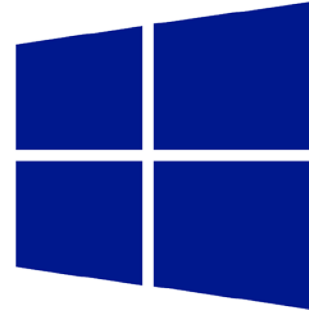


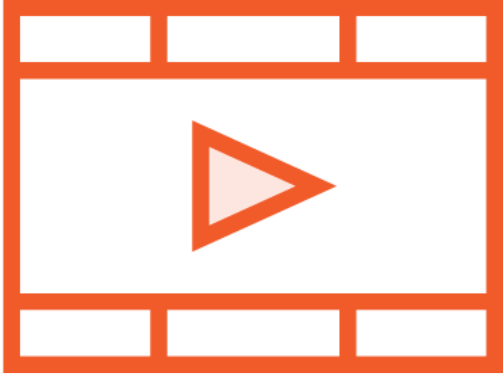
Using the HBase Thrift Server

- Generated HBase Thrift code
- Thrift NuGet package
- Connecting to HBase with .NET







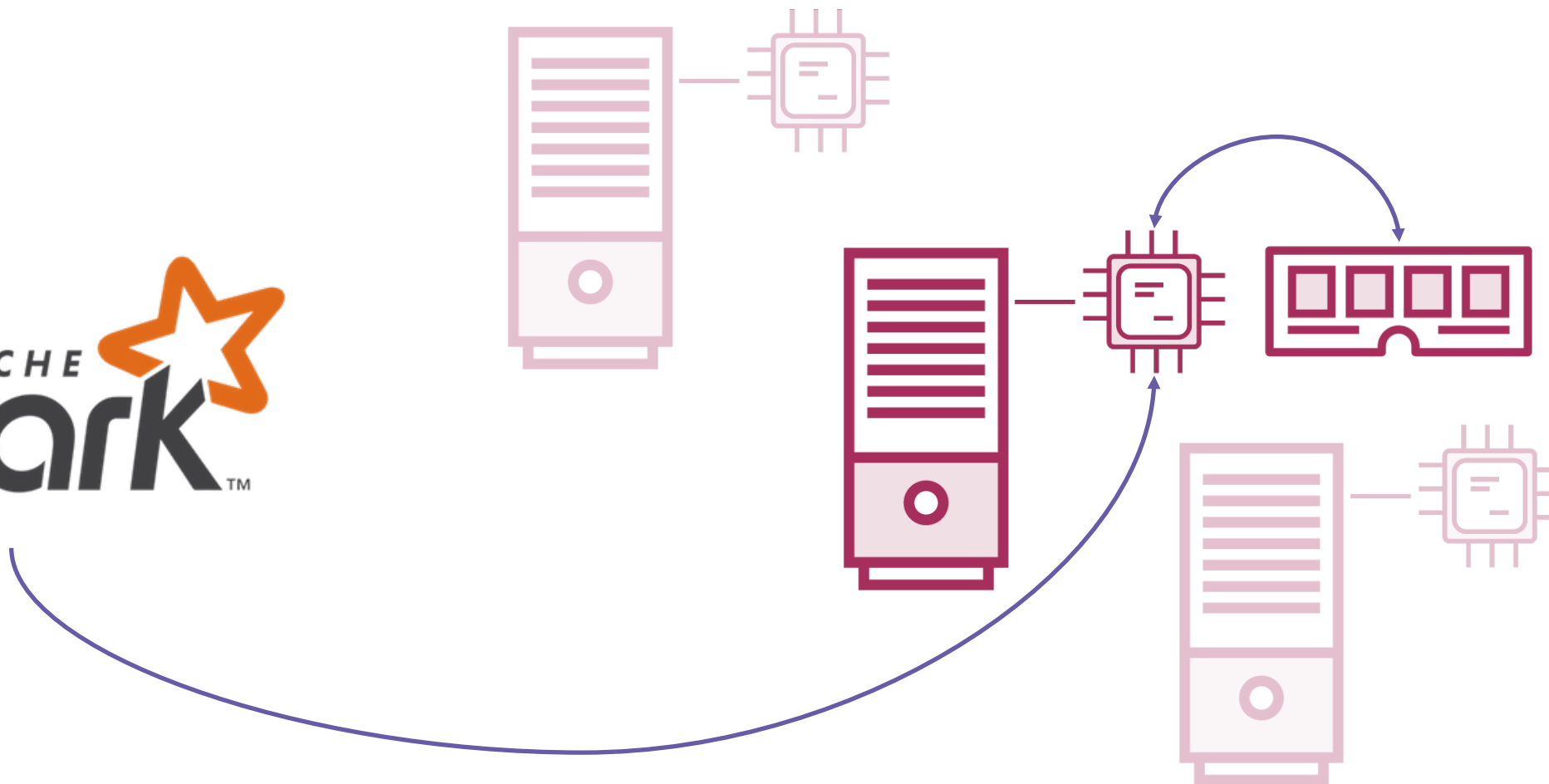


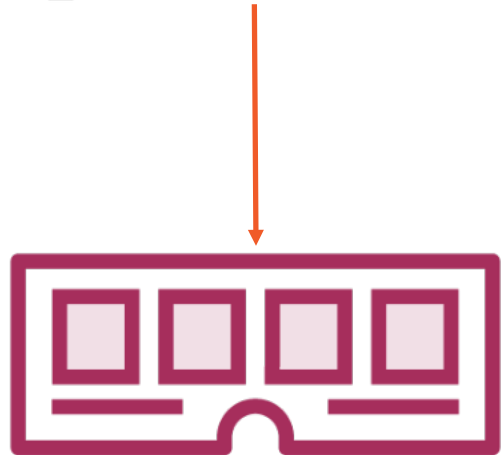
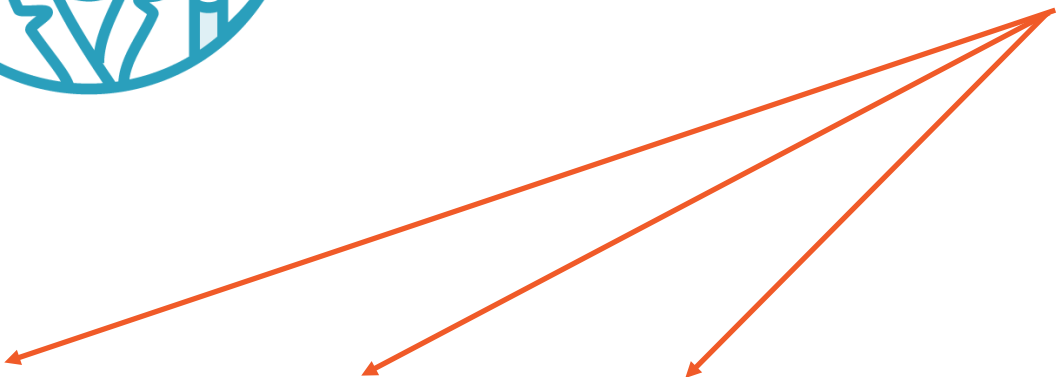
Apache Spark Fundamentals



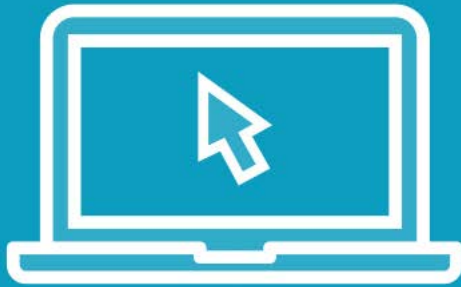
Justin Pihony

@justinpihony





Demo



Exploring Data with Apache Spark

- Running a Jupyter notebook
- Querying Hive with Spark
- Visualizing data

Hive
SQL Over Hadoop

Impala
Alternative SQL Engine

Spark
In-Memory Access

Pig
Scripted Query

Storm
Event Streaming

Oozie
Workflow

Mahout
Machine Learning

Solr
Search

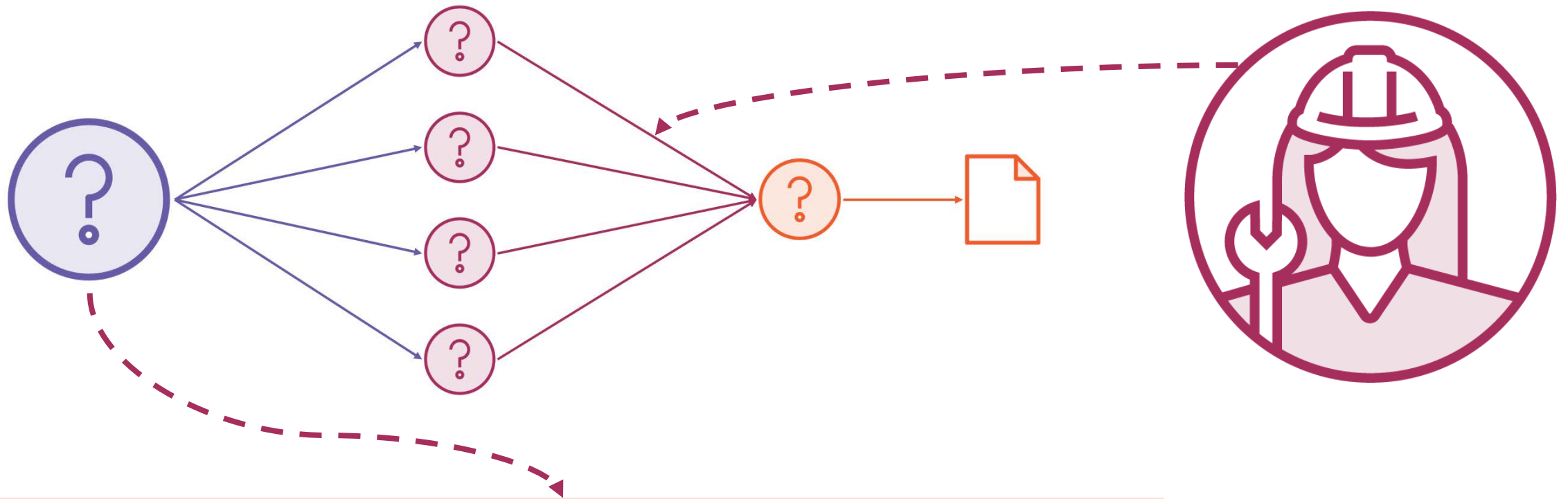
HBase
Real-time No SQL

YARN

Yet Another Resource Negotiator

HDFS

Hadoop Distributed File System



YARN
Yet Another Resource Negotiator

HDFS
Hadoop Distributed File System

Hive
SQL Over Hadoop

```
SELECT Classification, COUNT(1)  
FROM Postcodes  
GROUP BY Classification
```

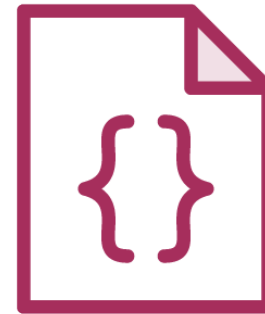


YARN
Yet Another Resource Negotiator

HDFS
Hadoop Distributed File System



TRANSFORM...
USING...





get 'postcodes', 'ML6 6PY'

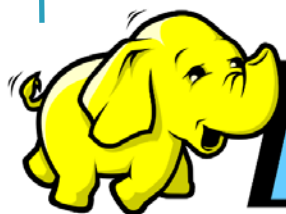
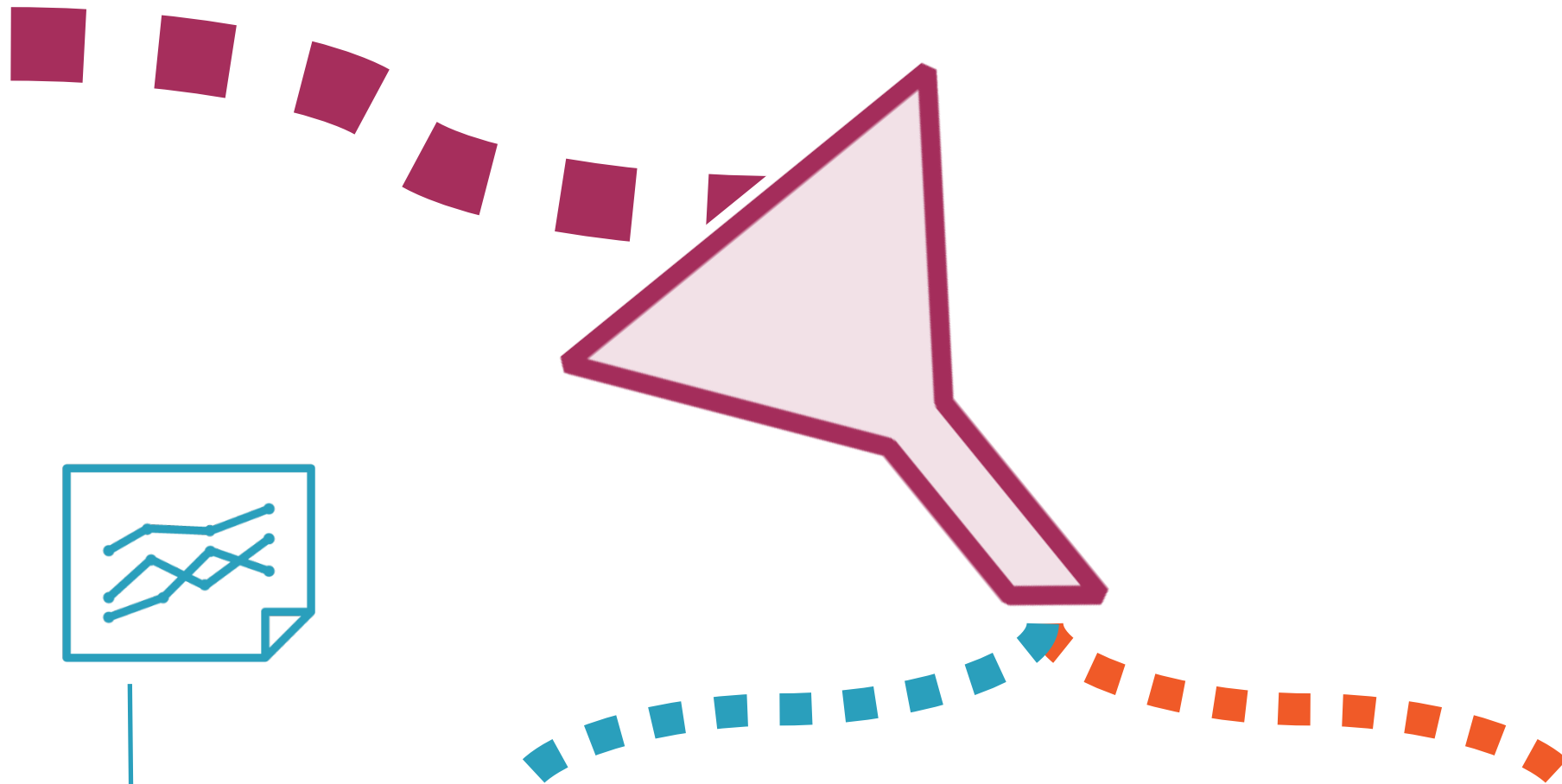
HBase
Real-time No SQL

YARN

Yet Another Resource Negotiator

HDFS

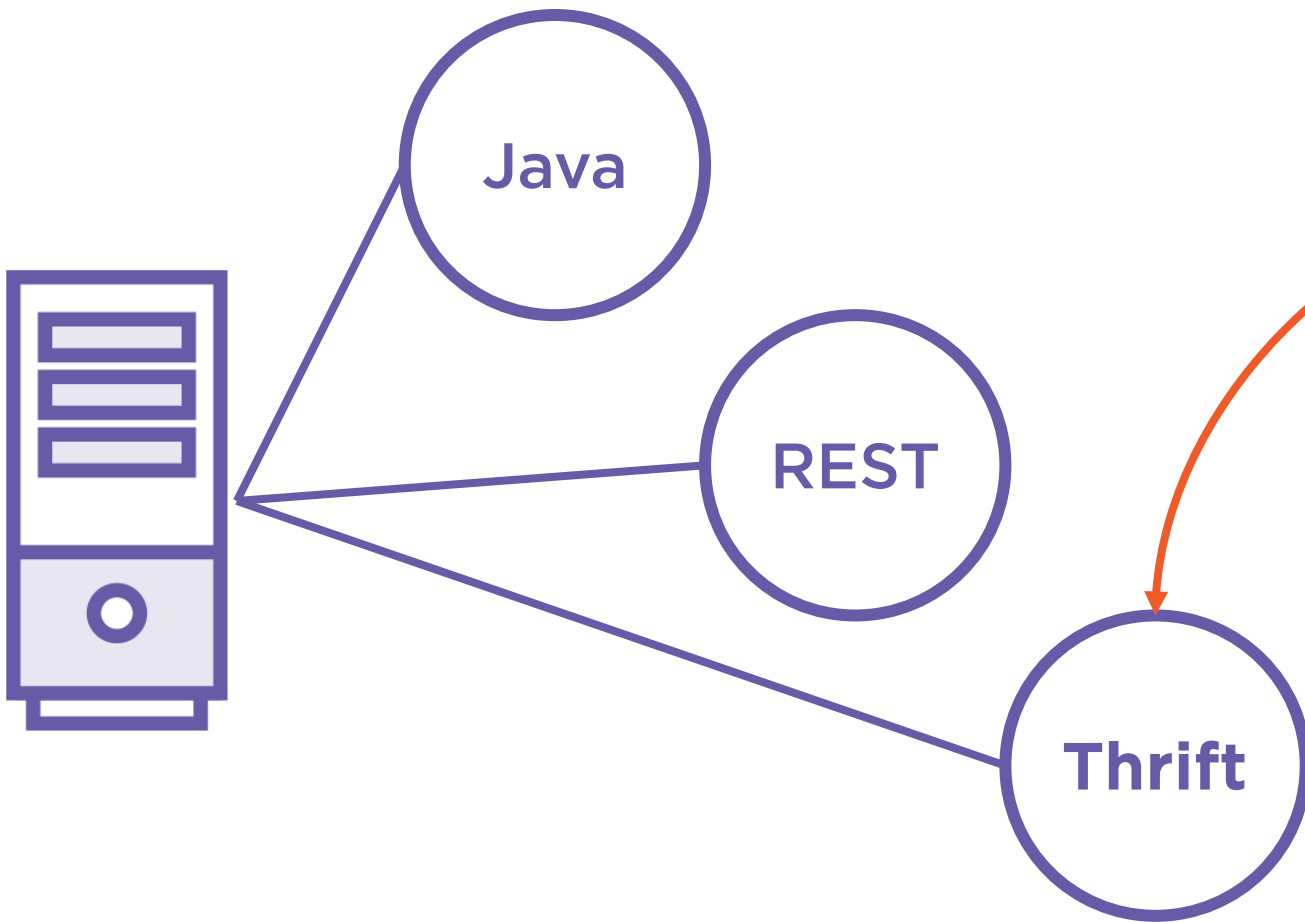
Hadoop Distributed File System



hadoop



APACHE
HBASE





Spark
In-Memory Access

```
p_acs.sort('PostcodeCount',  
          ascending=0)[:10]
```

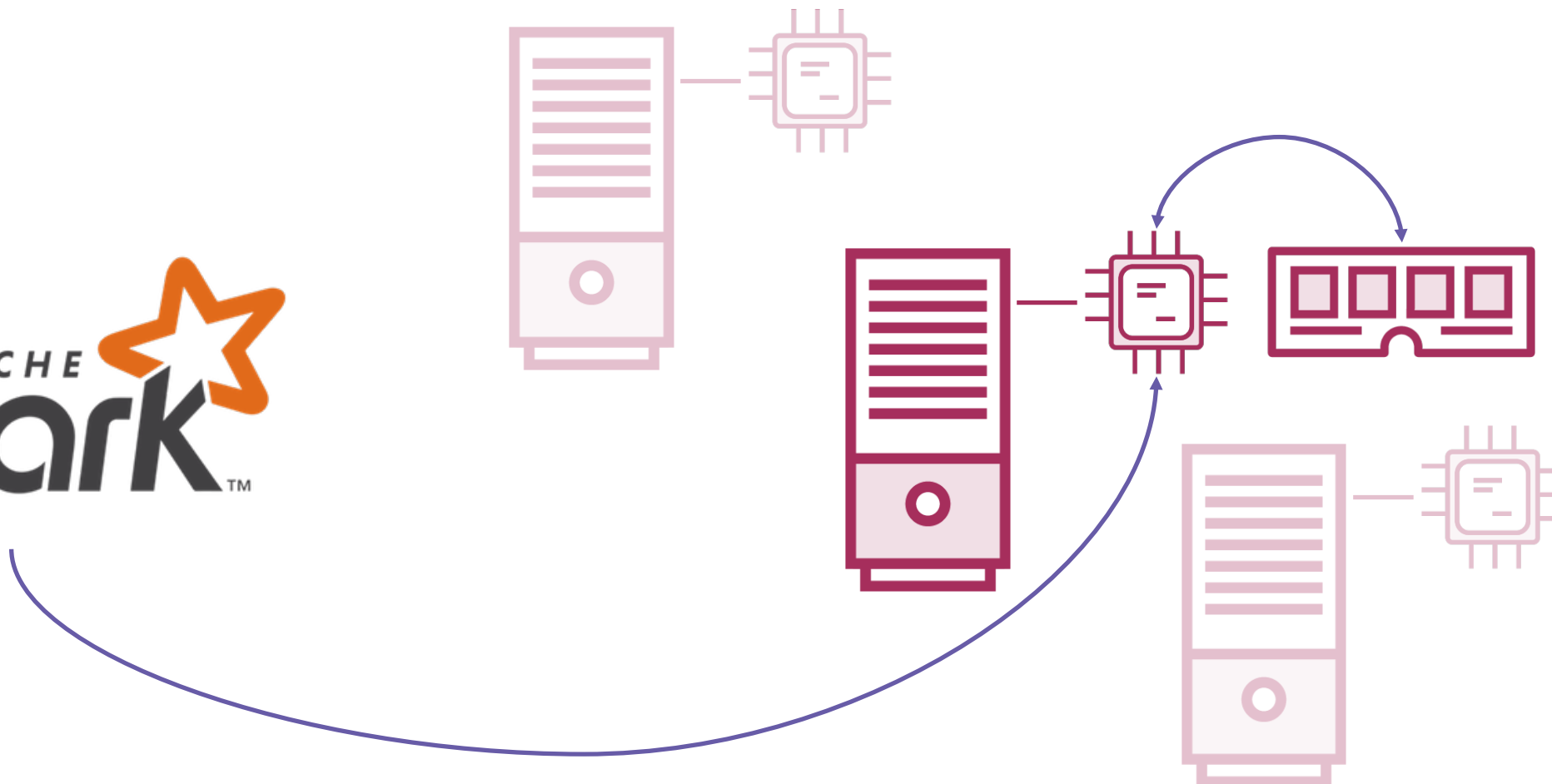


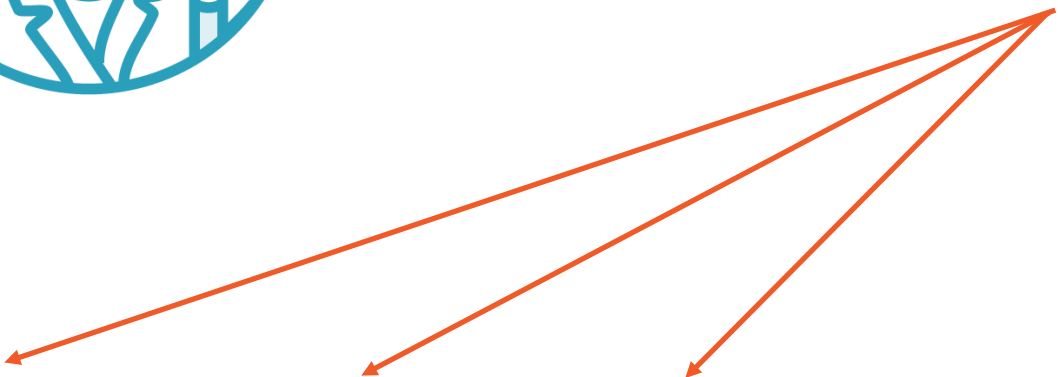
YARN

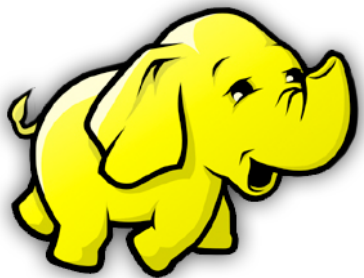
Yet Another Resource Negotiator

HDFS

Hadoop Distributed File System

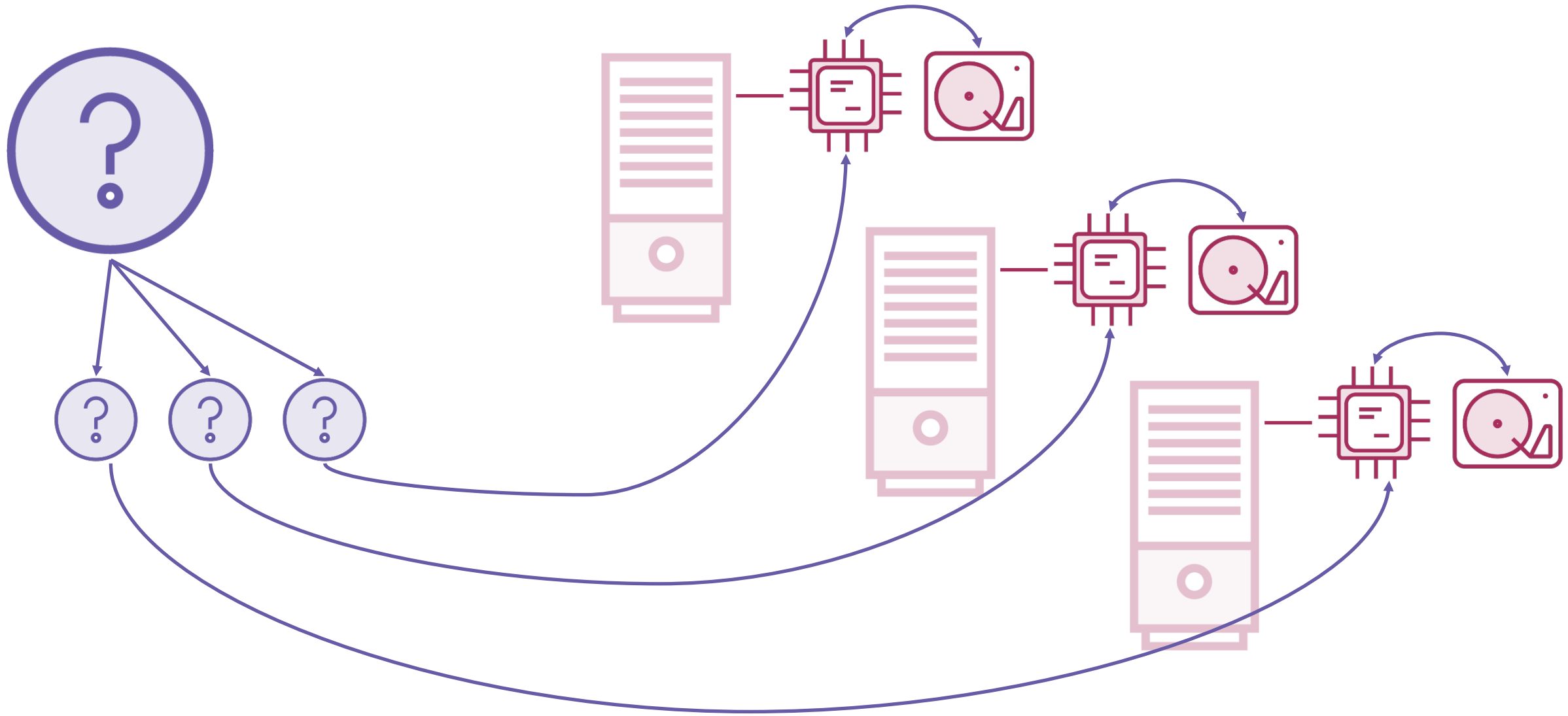


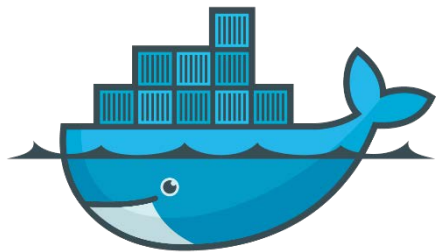
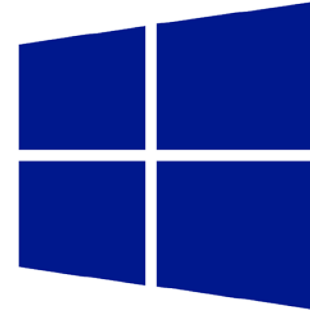




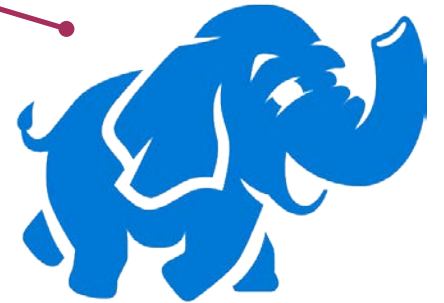
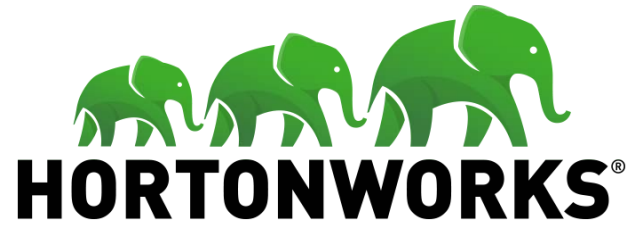
HDFS & YARN

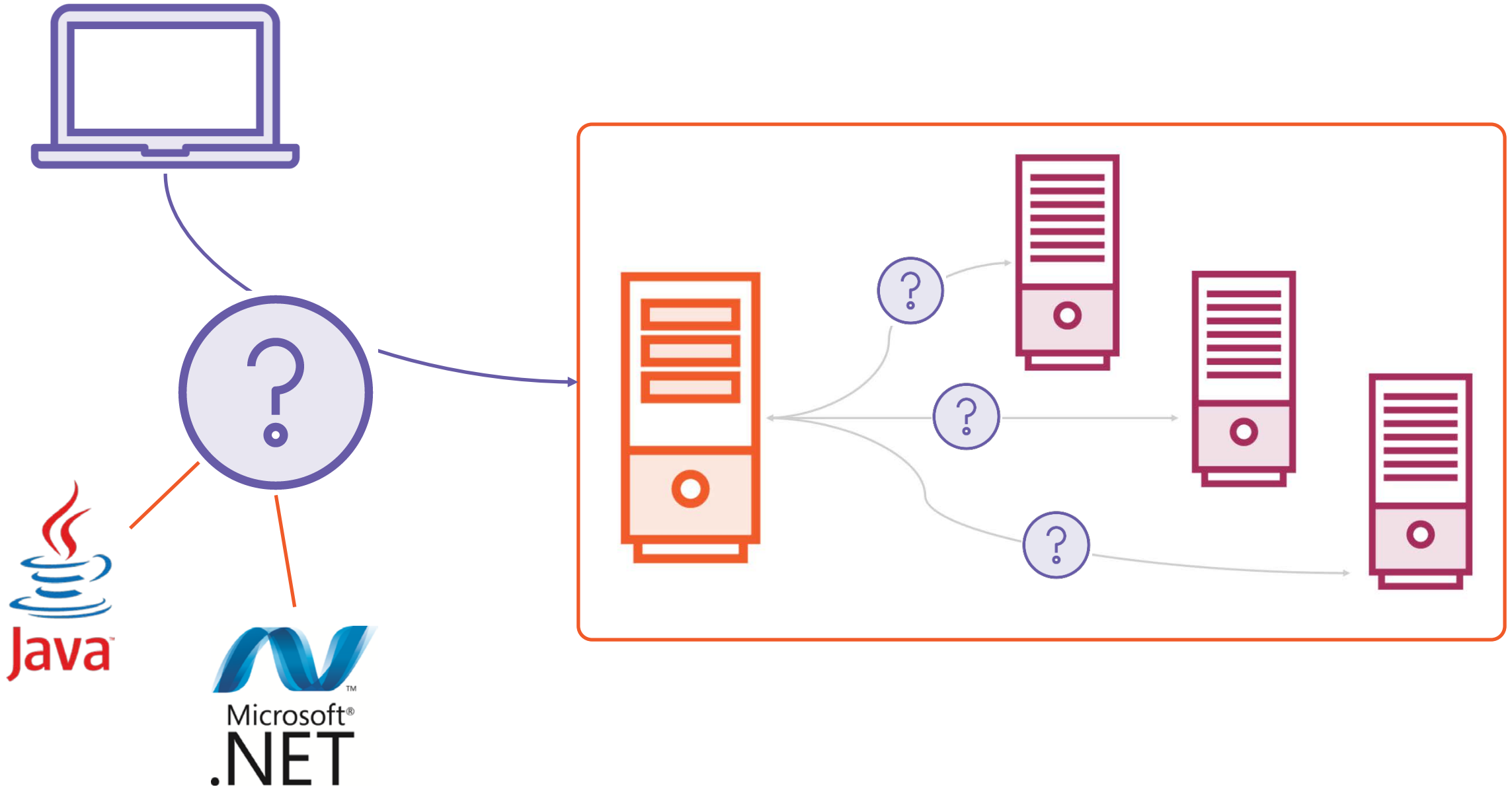







docker







HADOOP WITH .NET POC

- Does Hadoop run well on Windows?
- Can you write queries in .NET?
- What about **fast queries** in .NET?
- Does the Hadoop ecosystem  Microsoft?