



Microsoft Azure



amazon  
web services™

# AWS Data Engineer Syllabus (Full Stack)

**BIGDATA + AWS + SNOWFLAKE**



snowflake®



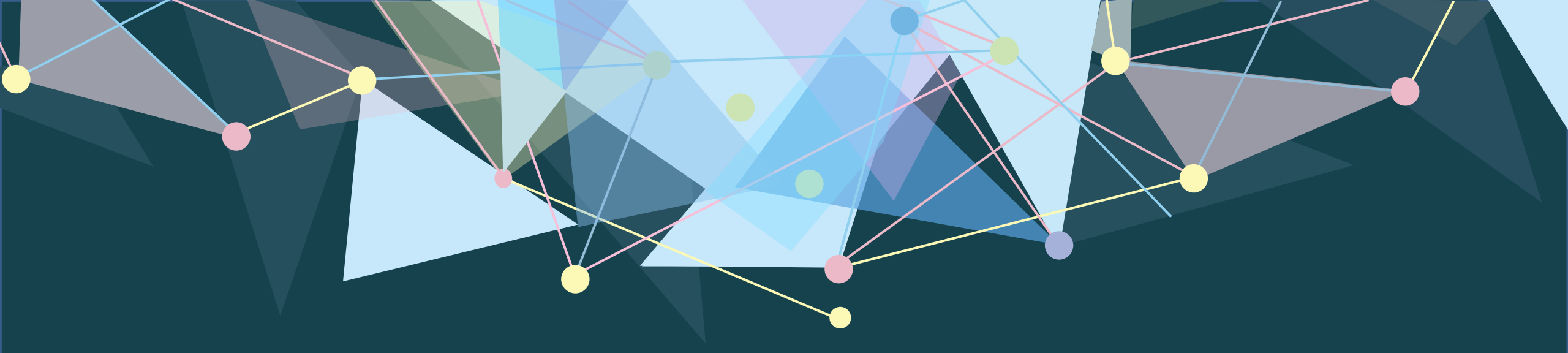
docker



APACHE  
Spark™



databricks™



# Experience with Multi Node Cluster

**3-5 Multi Node Cluster**

The screenshot shows a JupyterLab interface with a chat window on the left and a code editor on the right. The chat window, titled 'bigdatapedia', contains a message asking 'what is the use of this line' followed by the code 'set hive.exec.dynamic.partition=true;'. Below the message is a 'Copy To Clipboard' button. The code editor on the right shows the same code snippet: '[ ]: set hive.exec.dynamic.partition=true;'. The interface includes a top navigation bar with 'File', 'Edit', 'View', 'Run', 'Kernel', 'Tabs', 'Settings', and 'Help'. The bottom status bar shows 'Simple' mode, 'Hive | Idle', and 'Mode: Edit Ln 1, Col 38 HiveDay6Bucketing.ipynb 1'.

# First AI Supporting Training Program



# All Applications

- Cluster
- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler
- Tools

### Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources
3	0	1	2	2	<memory:1792, vCores:2>

### Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unlabeled Nodes
5	0	0	0	0

### Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum
Capacity Scheduler	[<name=memory-mb default-unit=Mi type=COUNTABLE>, <name=vcores default-unit= type=COUNTABLE>]	<memory:256, vC

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB
<a href="#">application_1638167425795_0003</a>	root	insert into newbatch values (1,'a','1234') (Stage-1)	MAPREDUCE	dev	0	Mon Nov 29 12:13:30 +0550 2021	Mon Nov 29 12:13:30 +0550 2021	N/A	RUNNING	UNDEFINED	2	2	1792
<a href="#">application_1638167425795_0002</a>	root	PySparkShell	SPARK	prod	0	Mon Nov 29 12:04:46 +0550 2021	Mon Nov 29 12:04:47 +0550 2021	Mon Nov 29 12:06:00 +0550 2021	FINISHED	SUCCEEDED	N/A	N/A	N/A
<a href="#">application_1638167425795_0001</a>	root	PySparkShell	SPARK	prod	0	Mon Nov 29 12:03:32 +0550 2021	Mon Nov 29 12:03:33 +0550 2021	Mon Nov 29 12:04:13 +0550 2021	FINISHED	SUCCEEDED	N/A	N/A	N/A

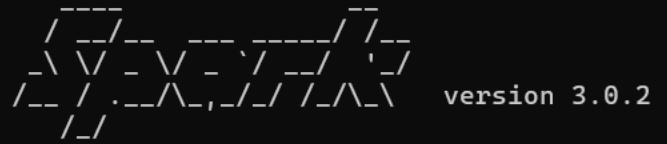
Showing 1 to 3 of 3 entries

Multi Node  
Realtime Practice

\*\*\*\*\*  
\* Welcome to BigDatapedia \*  
\*\*\*\*\*

Note\* If you find any issues on this images please raise it in GitHub page or reach us via WhatsApp +91 9715 010 010

```
root@f7503dd58916:/home/bigdatapedia#  
root@f7503dd58916:/home/bigdatapedia#  
root@f7503dd58916:/home/bigdatapedia#  
root@f7503dd58916:/home/bigdatapedia#  
root@f7503dd58916:/home/bigdatapedia# pyspark --master yarn --deploy-mode client  
Python 3.9.2 (default, Feb 28 2021, 17:03:44)  
[GCC 10.2.1 20210110] on linux  
Type "help", "copyright", "credits" or "license" for more information.  
Setting default log level to "WARN".  
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).  
21/11/29 06:34:44 WARN conf.HiveConf: HiveConf of name hive.metastore.event.db.notification.api.auth does not exist  
Welcome to
```



```
Using Python version 3.9.2 (default, Feb 28 2021 17:03:44)  
SparkSession available as 'spark'.
```

```
>>> sc  
<SparkContext master=yarn appName=PySparkShell>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>  
>>>
```

# Spark YARN

# Realtime Deployment

# Summary

Security is off.

Safemode is off.

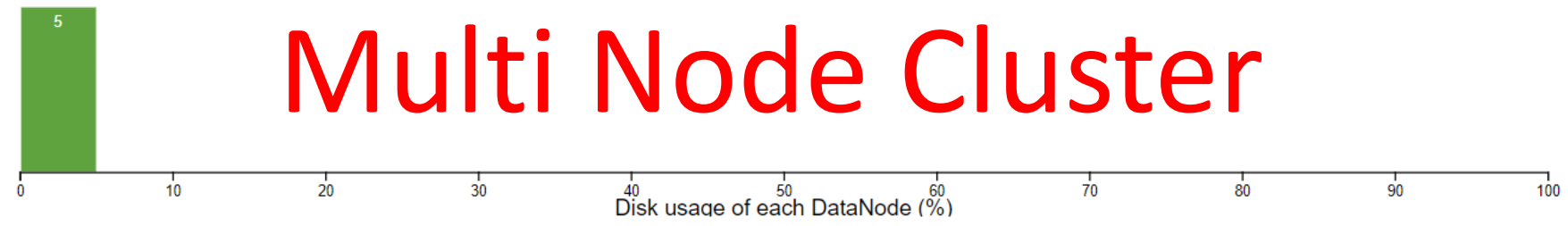
258 files and directories, 246 blocks = 504 total filesystem object(s).

Heap Memory used 102.64 MB of 213.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 53.78 MB of 54.78 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

<b>Configured Capacity:</b>	1.23 TB
<b>DFS Used:</b>	636.58 MB (0.05%)
<b>Non DFS Used:</b>	164.78 GB
<b>DFS Remaining:</b>	1 TB (81.71%)
<b>Block Pool Used:</b>	636.58 MB (0.05%)
<b>DataNodes usages% (Min/Median/Max/stdDev):</b>	0.04% / 0.05% / 0.06% / 0.01%
<b>Live Nodes</b>	5 (Decommissioned: 0, In Maintenance: 0)
<b>Dead Nodes</b>	0 (Decommissioned: 0, In Maintenance: 0)
<b>Decommissioning Nodes</b>	0
<b>Entering Maintenance Nodes</b>	0
<b>Total Datanode Volume Failures</b>	0 (0 B)
<b>Number of Under-Replicated Blocks</b>	0
<b>Number of Blocks Pending Deletion</b>	0

## Datanode usage histogram



## In operation

Show  entries

Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
✓ 73ff37c22de1:50010 (192.168.112.11:50010)	<a href="http://73ff37c22de1:50075">http://73ff37c22de1:50075</a>	0s	0m	250.98 GB <div style="display: inline-block; width: 100px; height: 10px; background-color: #ccc; position: relative;"><div style="width: 100%; background-color: #008000;"></div></div>	152	125.26 MB (0.05%)	2.10.1
✓ bc94daad9258:50010 (192.168.112.6:50010)	<a href="http://bc94daad9258:50075">http://bc94daad9258:50075</a>	0s	0m	250.98 GB <div style="display: inline-block; width: 100px; height: 10px; background-color: #ccc; position: relative;"><div style="width: 100%; background-color: #008000;"></div></div>	149	105.96 MB (0.04%)	2.10.1
✓ cc43bc7caef2:50010 (192.168.112.3:50010)	<a href="http://cc43bc7caef2:50075">http://cc43bc7caef2:50075</a>	0s	1m	250.98 GB <div style="display: inline-block; width: 100px; height: 10px; background-color: #ccc; position: relative;"><div style="width: 100%; background-color: #008000;"></div></div>	149	146.49 MB (0.06%)	2.10.1
✓ f8f9e903550d:50010 (192.168.112.4:50010)	<a href="http://f8f9e903550d:50075">http://f8f9e903550d:50075</a>	0s	0m	250.98 GB <div style="display: inline-block; width: 100px; height: 10px; background-color: #ccc; position: relative;"><div style="width: 100%; background-color: #008000;"></div></div>	141	126.88 MB (0.05%)	2.10.1
✓ fcb279b7c7ec:50010 (192.168.112.13:50010)	<a href="http://fcb279b7c7ec:50075">http://fcb279b7c7ec:50075</a>	0s	0m	250.98 GB <div style="display: inline-block; width: 100px; height: 10px; background-color: #ccc; position: relative;"><div style="width: 100%; background-color: #008000;"></div></div>	147	131.99 MB (0.05%)	2.10.1

```
*****  
*           Welcome to BigDatapedia           *  
*****
```

Note\* If you find any issues on this images please raise it in GitHub page or reach us via WhatsApp +91 9715 010 010

```
root@f7503dd58916:/home/bigdatapedia#
```

```
root@f7503dd58916:/home/bigdatapedia#
```

```
root@f7503dd58916:/home/bigdatapedia#
```

```
root@f7503dd58916:/home/bigdatapedia#
```

```
root@f7503dd58916:/home/bigdatapedia# hive
```

```
/home/bigdatapedia/hive/bin/hive: line 354: ps: command not found
```

```
SLF4J: Class path contains multiple SLF4J bindings.
```

```
SLF4J: Found binding in [jar:file:/home/bigdatapedia/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
```

```
SLF4J: Found binding in [jar:file:/home/bigdatapedia/hadoop/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
```

```
SLF4J: See http://www.slf4j.org/codes.html#multiple\_bindings for an explanation.
```

```
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
```

```
Logging initialized using configuration in jar:file:/home/bigdatapedia/hive/lib/hive-common-2.3.7.jar!/hive-log4j2.properties Async: true
```

```
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
```

```
hive>
```

```
> show databases;
```

```
OK
```

```
default
```

```
demo
```

```
Time taken: 0.641 seconds, Fetched: 2 row(s)
```

```
hive>
```

```
> use demo;
```

```
OK
```

```
Time taken: 0.024 seconds
```

```
hive>
```

```
> create external table newbatch (id int, name string, phno string)
```

```
> row format delimited fields terminated by ','
```

```
> stored as textfile
```

```
> location '/user/bigdatapedia/newbatch';
```

```
OK
```

```
Time taken: 0.194 seconds
```

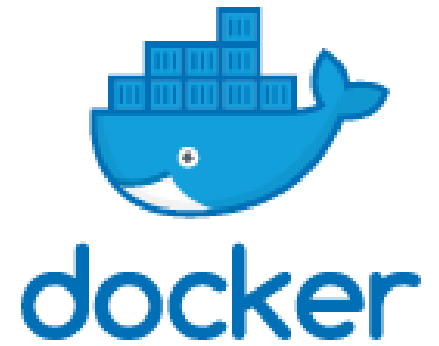
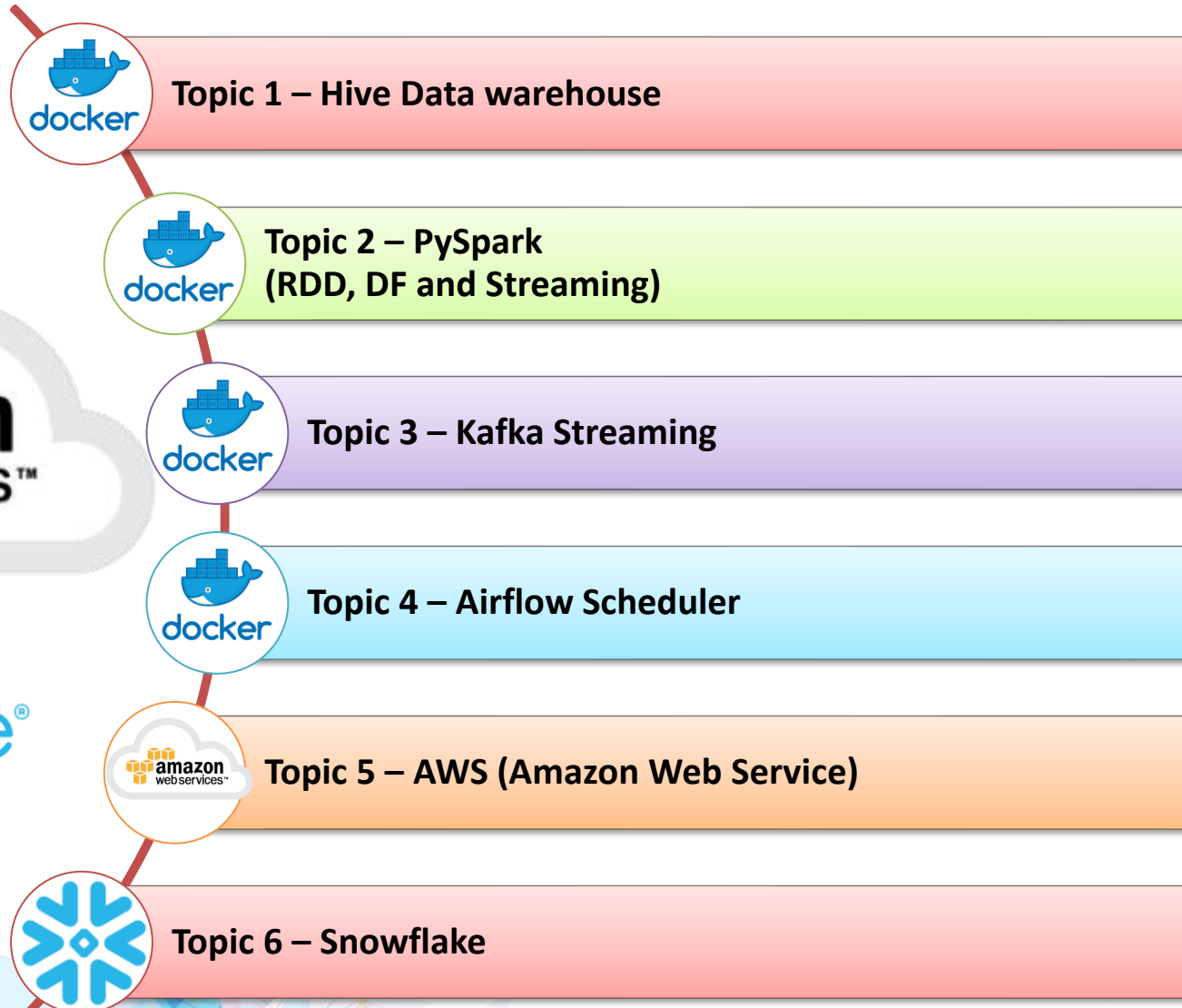
```
hive> show tables;
```

```
OK
```

```
newbatch
```



# Course Overview





1 0 1 0 1 0 1 0 1 0 1  
0 1 0 0 0 1 0 0 0 1 0  
0 0 1 1 0 0 1 1 0 0 1

**BIGDATAPEDIA**

Learning will lead the future...

# Our Policy

- No Fast Track (25% Theory + 75% Hands-on)
- 100% Refund if you are not satisfied
- Interactive and Concept basis
- Latest Version upgrades (Spark 3x)
- Private WhatsApp Group for your queries



10101010101  
01000100010  
00110011001

**BIGDATAPEDIA**

Learning will lead the future...

# Course Materials

- Practice **Hadoop Multi Node Cluster** (with Latest Version)
- Recorded Videos
- Practice Materials
- Online eLibraries
- Sample Interview Questions with Answer

# Topic - 1

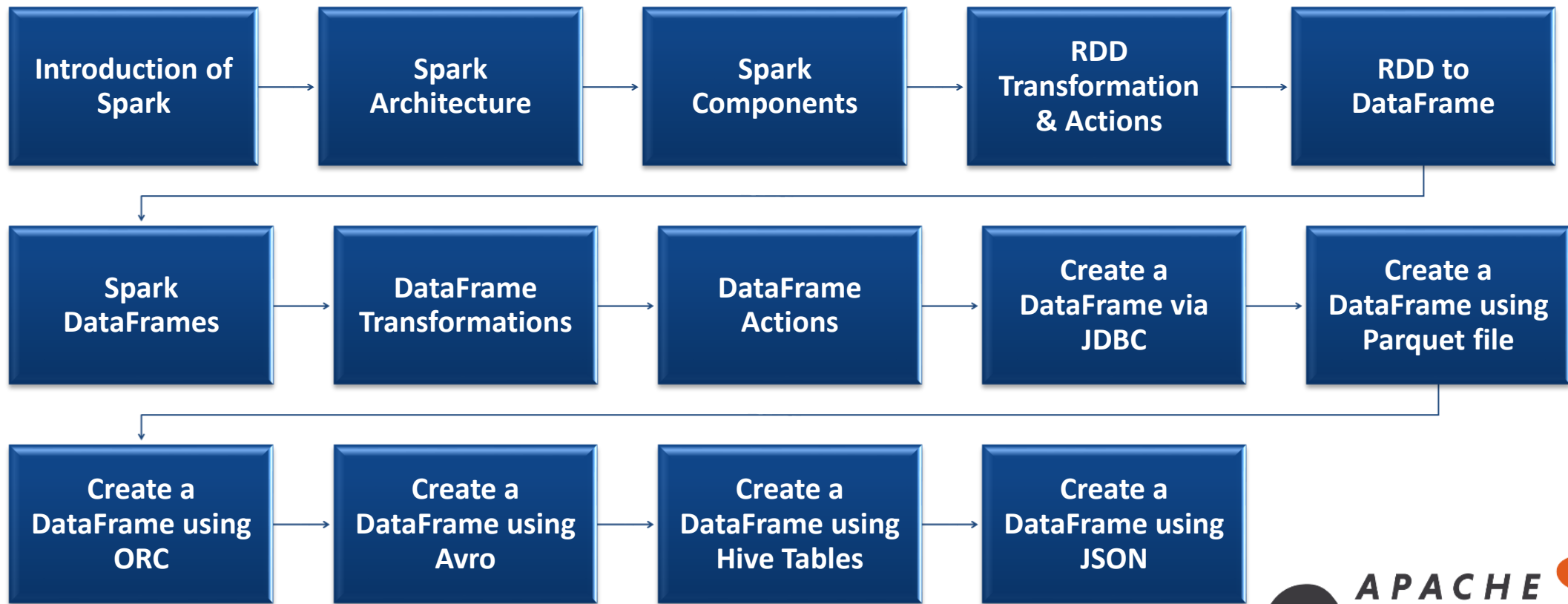
## Hive Introduction



# Topic - 3

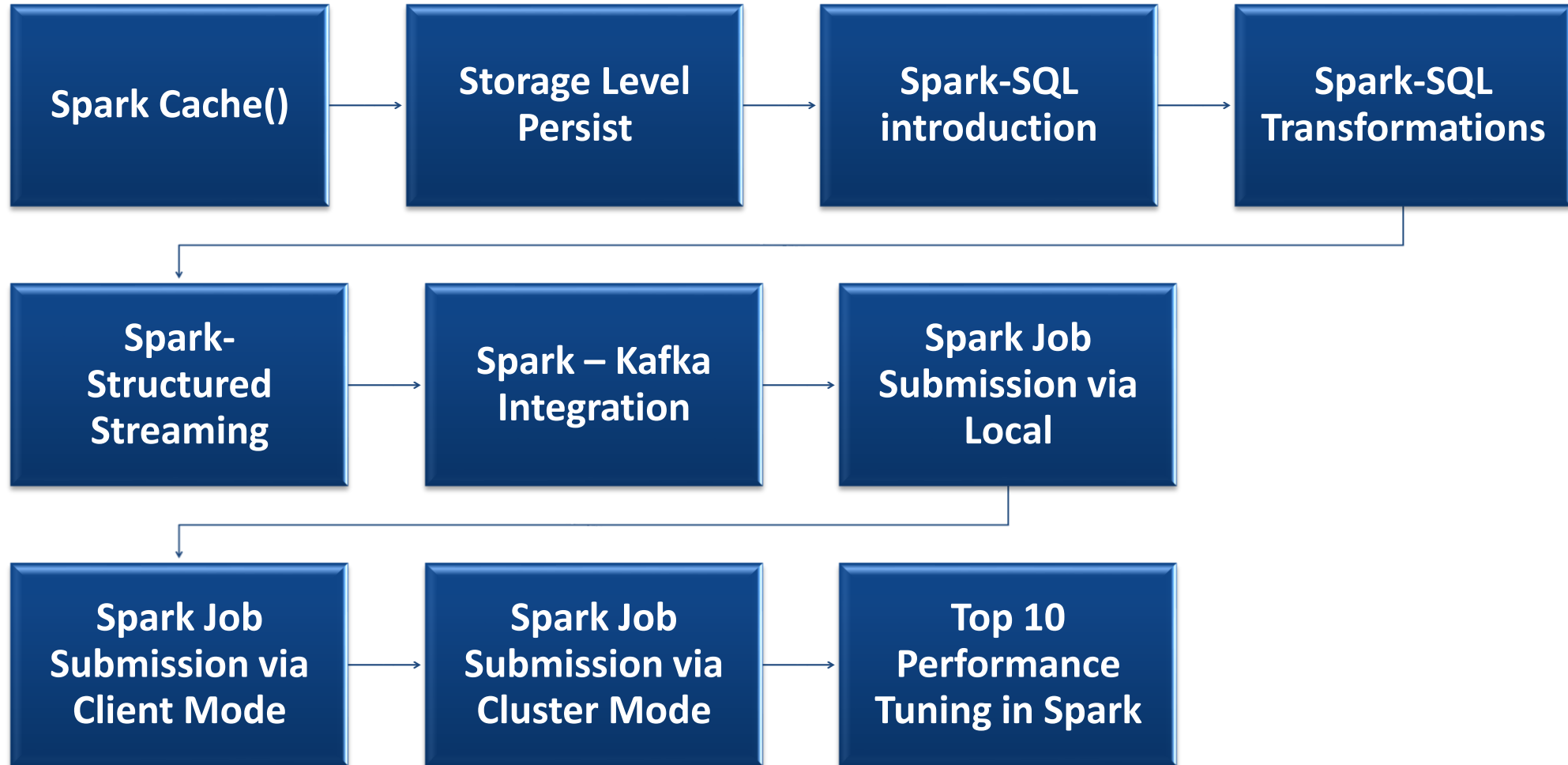
## Apache Spark Introduction

---



# Topic - 3

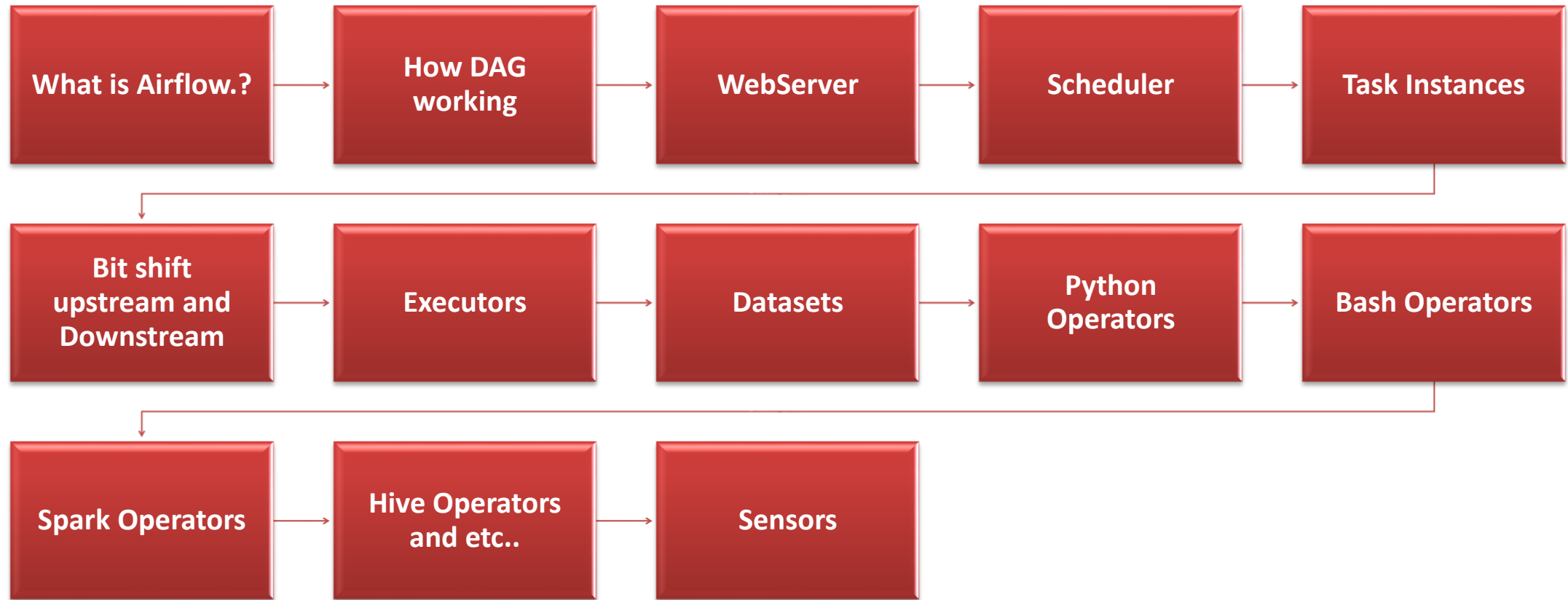
## Apache Spark Introduction (cont..)



# Topic - 4

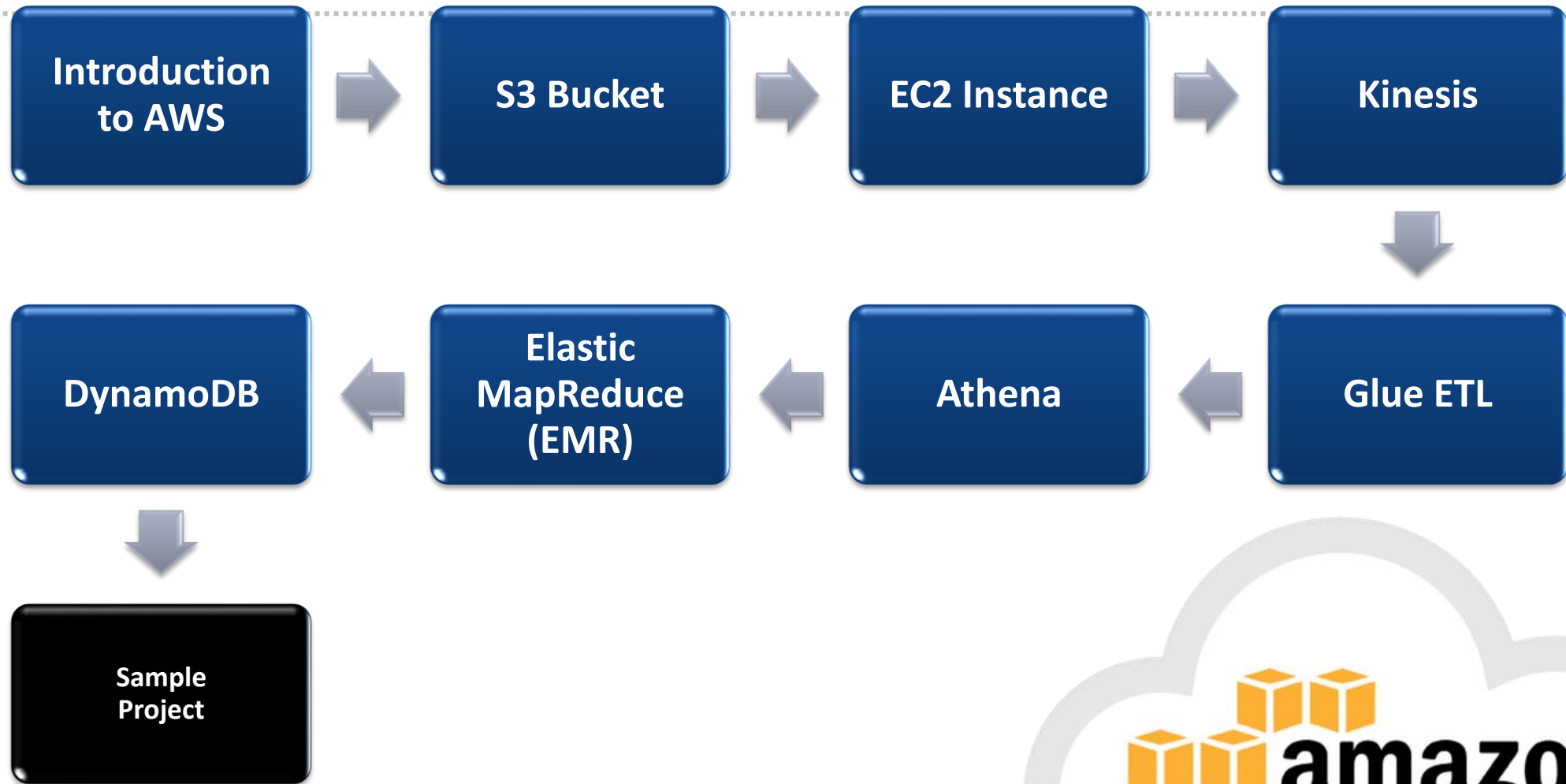
## Kafka Introduction



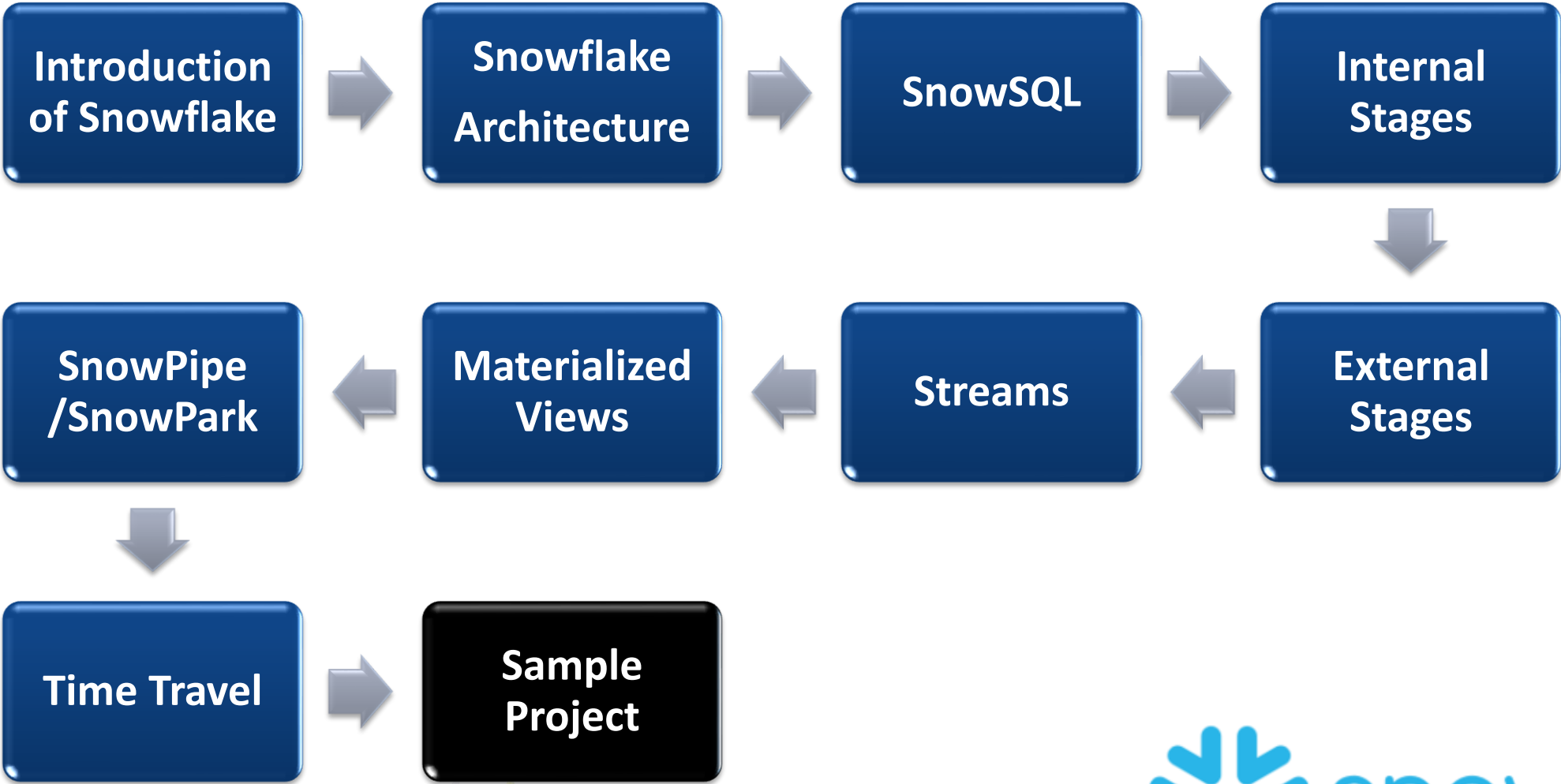




# Amazon Web Services



# Snowflake Cloud Data Warehouse





- Get Free Lifetime Practice with all recent upgraded VM.***
- Pay only for the new tools (Only for Old students).***
- Bigdata Interview related tips and tricks & Interview question with answers.***



