

The background is a dark gray collage featuring various Big Data logos and mascots. At the top, there is a red logo with a white 'i' and a dot, the 'git' logo in brown, the 'APACHE Spark' logo in black, and a colorful geometric logo with a white '1'. In the center, there is a yellow cartoon elephant (Hadoop mascot) and a yellow cartoon bee (Hive mascot) with the word 'HIVE' in black letters below it. To the right, there is a cartoon pig in blue overalls (Piggy Bank mascot). At the bottom, there is a blue silhouette of a reindeer head and a green outline of a piggy bank.

Big Data Processing Level-II

Agenda

2

- ▶ Multinode cluster installation on AWS
- ▶ UDF
- ▶ Various file formats and serde
- ▶ Mini project on Scala and advanced Spark
- ▶ Twitter streaming on Spark
- ▶ End-To-End job scheduling
- ▶ Kafka and Hbase
- ▶ Apache Ambari/Cloudera Manager for cluster management

Fun to learn

- ▶ How to check if the cluster is healthy
- ▶ Disk Usage
- ▶ Quotas
- ▶ Storage Policies
- ▶ How to stop/start/restart the cluster
- ▶ Log analysis and Troubleshooting
- ▶ Namenode stuck in safemode
- ▶ Memory issues
- ▶ Kerberos authentication