

The background is a dark grey collage featuring various logos and mascots associated with big data technologies. At the top left is the Hadoop logo (a red shield with a white 'H'). Next to it is the 'g1c' logo in a stylized brown font. To the right is the 'APACHE Spark' logo in black. Further right is a colorful geometric logo with a red vertical bar. Below these are several mascots: a yellow cartoon bee on the left, a yellow cartoon elephant in the center, a blue cartoon pig on the right, and a blue cartoon antelope at the bottom left. At the bottom center is a green logo consisting of a series of connected circles and squares.

# Big Data Processing Level-II

# Agenda

2

- ▶ Multinode cluster installation on AWS
- ▶ UDF
- ▶ Various file formats and serde
- ▶ Mini project on Spark
- ▶ Twitter streaming on Spark
- ▶ End-To-End job scheduling
- ▶ Kafka