

EC2 in Big Data

- On demand, Spot & Reserved instances:
 - Spot: can tolerate loss, low cost => checkpointing feature (ML, etc)
 - Reserved: long running clusters, databases (over a year)
 - On demand: remaining workloads
- Auto Scaling:
 - Leverage for EMR, etc
 - Automated for DynamoDB, Auto Scaling Groups, etc...
- EC2 is behind EMR
 - Master Nodes
 - Compute Nodes (contain data) + Tasks Nodes (do not contain data)