1. Migrating existing data workloads to an environment (Ex - Hadoop, Spark Jobs) where one can analyze the data.
2. Analyzing large datasets at scale.
3. Building scalable pipelines that can handle streaming data.
4. Applying machine learning to data(building models).