Spark includes all these in one software:

1. Batch data: historical data last 1 year, Hadoop, Hive, pig, sqoop, oozie
2. Streaming data: live data
3. Machine learning: recommendation, predicting future, sklearn
4. Graphical data: neo4j

Advantages of spark:

1. Very fast as compare to other technologies.
2. Spark is general purpose, open source, puttable (can easily integrate with any technologies) framework.
3. Api used in spark:
   1. Rdd api (10%)
   2. Dataframe (70%)
   3. Dataset Api (15%)
   4. Dstream Api (5%)
4. Databricks: took spark and modified it (unifying spark, cloud)
5. Cloudera: unifying all big data (Hadoop, hive, sqoop, oozie, kafka, spark)
6. Aws glue:aws etl service use spark.
7. Azure databricks: took spark+