SARDAR VALLABHBHAI PATEL INSTITUTE OF TECHNOLOGY, VASAD.
Information Technology Department
Subject Name: **BIG DATA ANALYTICS (2171607)**

1   What is Apache Spark? What are the advantages of using Apache Spark over Hadoop? Explain in brief four major libraries of Apache Spark.
2   What is transformation and actions in Apache Spark? Discuss various commands available for these activities in Apache Spark?
3   What is Resilient Distributed Dataset in Apache Spark? Explain in detail. Make a note on why RDD is better than Map Reduce data storage?
4   What are the Components of Spark? Also state the features of Spark.
5   What is Spark? State the advantages of using Apache Spark over Hadoop MapReduce for big data processing with example.
6   What is RDD? Explain about transformations and actions in the context of RDDs. State and explain RDD operations in brief.
7   What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset? Explain RDDs in detail.
8   Explain Spark components in detail. Also list the features of spark.
9   Write a brief short note on: Spark Unified Stack