

## **Module – 4**

1. What is Apache Spark? With suitable diagram illustrate the architecture of a Spark Application. 10
2. Write the diagram of the relationship between the Spark Session and Spark's Language API. List and explain the languages used for Spark's Language APIs. 10
3. Define Data Frame. Discuss the difference between a single machine and dataframe data distributed over a cluster. 5
4. What is Lazy Evaluation? Write the difference between TRANSFORMATIONS and ACTIONS. 5
5. Describe Structured API logical planning process and physical planning process. 10

**OR**

**Overview of Structured API execution.10**

6. Write a note on Spark's Toolkit. 5

**OR**

**Discuss key components of Spark's Toolkit. 5**

7. Write a note on partitions of Spark. 5
8. Discuss datasets of Spark.5
9. Elaborate on structured streaming. 10
10. Spark's MLlib. 5
11. Present history of Spark. 5