	INTERNAL ASSIGNMENT - 1
SURESH GYAN VIHAR UNIVERSITY Accredited by NAAC with 'A' Grade	

Course	**MCA Plus**	
Semester	**4**	**Spark on Azure HD**
Total Marks:	**15**	

Q.1. Write answers for any two questions from below. (5 Marks each – Word limit – 500)

a) What is Spark? Describe the various components of it.

b) What is Data Science. Explain the applications of Data science.

c) what are the offerings of Apache Spark on HDInsight?

Q.2. Write short notes on all the following topics (1 Mark each - Word limit - 100)

a) Explain the various Apache Spark Terminologies.

b) Describe the Building blocks of Data Preparation.

c) Explain Data Preparation Methodology.

d) What is Collaborative filtering?

e) What is Clustering in MLlib?

SURESH GYAN VIHAR UNIVERSITY Accredited by NAAC with 'A' Grade		**INTERNAL ASSIGNMENT - 2**
Course	**MCA Plus**	
Semester	**4**	**Spark on Azure HD**
Total Marks:	**15**	

Q.1. Write answers for any two questions from below. (5 marks each – Word limit – 500)

 a) Explain Feature extraction and transformation.

 b) Explain Optimization.

 c) Explain the features of R Server on HDInsight.

Q.2 Write short notes on all the following topics. (1 Mark each - Word limit - 100)

 a) Explain Architecture of Spark SQL.

 b) What are the components of Spark SQL?

 c) Describe the various Spark Streaming operations.

 d) What is Dimensionality reduction in MLlib ?

 e) Explain Random data generation.