ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE
III SEMESTER EXAMINATION, OCTOBER 2019
M.SC IN BIG DATA ANALYTICS
BDA 3218: ENABLING TECHNOLOGIES FOR DATA SCIENCE

TIME 2 ½ HRS   **MAX MARKS 70**
This Question Paper Contains Two Parts

PART I

Answer Any Six Questions   (6*5=30)

1) What are the exception handling operators in Pig script?

2) Mention the date data type in Hive. Name the Hive data type collection

3) What is zookeeper? how does it work?

4) What is the difference between HDFS block and Input Split?

5) How to improve HIVE query performance?

6) Write all steps of dynamic partition table creation?

7) State the usage of 'filters', 'group', 'order by', 'distinct' keywords in pig scripts.

8) Write syntax for incremental updates using Sqoop?

9) What is the difference between gen1 (1.x) and gen2 (2.x) Hadoop with regards to the name node?

PART II

Answer Any Four Questions   (4*10=40)

1) With a neat diagram explain the complete architecture of HIVE and explain each component in detail?

2) Explain the complete map reduce life cycle with example?

3) What are the different phases of Map Reduce program and explain it with Word Count?

4) What are the components of YARN? Explain each component functionality with neat diagram?

5) Draw 2.x Hadoop architecture and explain the read and write operation of it?