

Name:	 UPES UNIVERSITY WITH A PURPOSE
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, July 2020

Course: Big Data Ingestion
Program: BTech – CS (BD)
Course Code: CSBD 2002

Semester: IV
Time 03 hrs.
Max. Marks: 60

Instructions: All questions are compulsory.

SECTION A

S. No.		Marks	CO
Q 1	_____ tool can list all the available database schemas a. sqoop-list-tables b. sqoop-list-databases c. sqoop-list-schem d. sqoop-list-columns	1.67	CO1
Q 2	Point out the correct statement. a. The sqoop command-line program is a wrapper which runs the bin/hadoop script shipped with Hadoop b. If \$HADOOP_HOME is set, Sqoop will use the default installation location for Cloudera’s Distribution for Hadoop c. The active Hadoop configuration is loaded from \$HADOOP_HOME/conf/, unless the \$HADOOP_CONF_DIR environment d. None of the mentioned	1.67	CO1
Q 3	Data can be imported in maximum _____ file formats. a. 1 b. 2 c. 3 d. 4	1.67	CO1
Q 4	_____ text is appropriate for most non-binary data types. a. Character b. Binary c. Delimiter d. None of the above	1.67	CO1
Q 5	Point out the wrong statement. a. Avro data files are a compact, efficient binary format that provides interoperability	1.67	CO1

	<p>with applications written in other programming languages</p> <p>b. By default, data is compressed while importing</p> <p>c. Delimited text also readily supports further manipulation by other tools, such as Hive</p> <p>d. None of the mentioned</p>		
Q 6	<p>_____ does not support the notion of enclosing characters that may include field delimiters in the enclosed string.</p> <p>a. Impala</p> <p>b. Oozie</p> <p>c. Sqoop</p> <p>d. Hive</p>	1.67	CO1
Q 7	<p>The _____ tool imports a set of tables from an RDBMS to HDFS</p> <p>a. export-all-tables</p> <p>b. import-all-tables</p> <p>c. import-tables</p> <p>d. none of the mentioned</p>	1.67	CO1
Q 8	<p>Which of the following argument is not supported by import-all-tables tool?</p> <p>a. -class-name</p> <p>b. -package-name</p> <p>c. -database-name</p> <p>d. -table-name</p>	1.67	CO1
Q 9	<p>For some databases sqoop can to faster data transefr by using the parameter</p> <p>a. bulkload</p> <p>b. fastload</p> <p>c. dump</p> <p>d. direct</p>	1.67	CO1
Q 10	<p>What is the disadvantage of storing password in the metastore as compared to storing in a password file?</p> <p>a. it is easily accessible</p> <p>b. it may get deleted accidentally</p> <p>c. It cannot be updated</p> <p>d. it is unencrypted</p>	1.67	CO1
Q 11	<p>The export and import of data between sqoop and relational system happens through which of the following programs?</p> <p>a. Sqoop client program</p> <p>b. Mapreduce job submitted by the sqoop command</p>	1.67	CO1

	<ul style="list-style-type: none"> c. Database stores procedure d. Hdfs file management program 		
Q 12	<p>The insert query used to insert exported data into tables is generate by</p> <ul style="list-style-type: none"> a. Sqoop command and processed as such b. Sqoop command and modified suitably by JDBC drive c. JDBC driver d. Database specific driver 	1.67	CO1
Q 13	<p>Sqoop can insert new rows and update existing changed rows into an already exported table by using the parameter</p> <ul style="list-style-type: none"> a. –update-insert b. update-else-insert c. –update-mode insert d. –update-mode allowinsert 	1.67	CO1
Q 14	<p>To sync a HDFS file with some deleted rows with a previously exported table for the same table the option is to</p> <ul style="list-style-type: none"> a. Using staging table b. Export the data again to a new database table and rename it c. Use a ETL tool d. Can not be done using sqoop 	1.67	CO1
Q 15	<p>For a column of data type numeric, the parameter used for substituting null values is</p> <ul style="list-style-type: none"> a. input-null-string b. input-null-non-string c. input-null-text d. input-null-varchar 	1.67	CO1
Q 16	<p>The –update-key parameter can take</p> <ul style="list-style-type: none"> a. Only one column name as key field b. Two column name as key filed c. Comma separated list of columns as keys d. Table name and column name as key 	1.67	CO2
Q 17	<p>The disadvantage of using a stored procedure to laod data is</p> <ul style="list-style-type: none"> a. Data gets loaded faster b. Parallel loads in the database table c. The store procedure cannot load multiple tables at a time d. It induces heavy load on database 	1.67	CO2
Q 18	<p>While importing data into Hadoop using sqoop the SQL SELCT clause is used. Similarly while exporting data form Hadoop the SQL clause used is</p>	1.67	CO2

	<ul style="list-style-type: none"> a. APPEND b. MERGE c. UPDATE d. INSERT 		
Q 19	<p>What are the two different incremental modes of importing data into sqoop?</p> <ul style="list-style-type: none"> a. merge and add b. append and modified c. merge and lastmodified d. append and lastmodified 	1.67	CO2
Q 20	<p>While inserting data into Relational system from Hadoop using sqoop, the various table constraints present in the relational table must be</p> <ul style="list-style-type: none"> a. Disabled temporarily b. Dropped and re created c. Renamed d. Not violated 	1.67	CO2
Q 21	<p>To import data to HBase using sqoop the parameter(s) required is/are</p> <ul style="list-style-type: none"> a. hbase-table b. hbase-columns c. hbase-table and --column-family d. habse-table and --hbase-column-family 	1.67	CO2
Q 22	<p>What is the disadvantage of using the –columns parameter to insert a subset of columns to the relational table?</p> <ul style="list-style-type: none"> a. The relational table may have not null columns not covered in the –columns parameter. b. The relational table may store the data from HDFS in wrong columns. c. It may not load all the required data d. It will not be able to populate primary key values 	1.67	CO2
Q 23	<p>The argument in a saved sqoop job can be altered at run time by using the option</p> <ul style="list-style-type: none"> a. alter b. newval c. exec d. changeparam 	1.67	CO2
Q 24	<p>What is the advantage of storing password in a metastore as compared to storing in password in a file?</p> <ul style="list-style-type: none"> a. User intervention is required in password file but not in metastore. b. The password file can be encrypted while the metastore cannot be encrypted 	1.67	CO2

	<p>c. The password in metastore can be updated while that in password file cannot be</p> <p>d. It can be run by any user with valid access to sqoop environment</p>		
Q 25	<p>What does the --last-value parameter in sqoop incremental import signify?</p> <p>a. What is the number of rows successfully imported in append type import</p> <p>b. what is the date value to be used to select the rows for import in the last_update_date type import</p> <p>c. Both of the above</p> <p>d. The count of the number of rows that were successful in the current import.</p>	1.67	CO2
Q 26	<p>Sqoop supports</p> <p>a. full import of tables</p> <p>b. partial import of data from tables</p> <p>c. Both full and partial data import</p> <p>d. Import both the table and its partitions</p>	1.67	CO2
Q 27	<p>what option can be used to import only some of the table from a database while using the --import-all-tables parameter?</p> <p>a. skip-tables</p> <p>b. without-tables</p> <p>c. forgo-tables</p> <p>d. exclude-tables</p>	1.67	CO2
Q 28	<p>What option can be used to import the entire database from a relational system using sqoop?</p> <p>a. import-all-db</p> <p>b. import-all-tables</p> <p>c. import-all</p> <p>d. import-db</p>	1.67	CO2
Q 29	<p>What is the default value used by sqoop when it encounters a missing value while importing from CSV file.</p> <p>a. NULL</p> <p>b. null</p> <p>c. NaN</p> <p>d. no value</p>	1.67	CO2
Q 30	<p>What does the num-mappers parameter serve?</p> <p>a. force sqoop to use only one map task</p> <p>b. set the number of map tasks sqoop can use</p> <p>c. store the data imported by each map tasks in a separate file</p> <p>d. Fetch each row from the table using a new map task</p>	1.67	CO2

Q 31	<p>schemas type that are separate from the message payloads</p> <ul style="list-style-type: none"> a. JSON b. XML c. AVRO d. CSV 	1.67	C03
Q 32	<p>Which of the following does not support stream processing</p> <ul style="list-style-type: none"> a. Apache NIFI b. Apache Kafka c. Apache Storm d. Google Dataproc 	1.67	C03
Q 33	<p>What Is The Benefits Of Apache Kafka Over The Traditional Technique?</p> <ul style="list-style-type: none"> a. Scalability b. Sharding c. aggregation d. bulk data transfer 	1.67	C03
Q 34	<p>What Is The Meaning Of Broker In Kafka?</p> <ul style="list-style-type: none"> a. A server b. An object c. A program d. A datacenter 	1.67	C03
Q 35	<p>What is offset in kafka?</p> <ul style="list-style-type: none"> a. A bit of metadata b. A floating point number c. Address of a broker d. address of a consumer 	1.67	C03
Q 36	<p>Where does the consumer store the offset of a message?</p> <ul style="list-style-type: none"> a. in Consumer b. In Partition c. In Topic d. In Zookeeper 	1.67	C03
Q 37	<p>Who assign the offset to a message?</p> <ul style="list-style-type: none"> a. Producer b. Consumer c. Brocker d. Partition 	1.67	C03

Q 38	<p>_____ assign the partition to the broker.</p> <ul style="list-style-type: none"> a. Controller b. Zookeeper c. Producer d. Consumer 	1.67	C03
Q 39	<p>What is the roll of the MirrorMaker?</p> <ul style="list-style-type: none"> a. Monitoring the failure of kafka cluster b. Monitoring the failure of kafka Broker c. Aggregate message from local clusters to a cluster d. Assign partitions to a broker 	1.67	C03
Q 40	<p>What is the optional field in the ProducerRecord?</p> <ul style="list-style-type: none"> a. Topic b. Value c. Partition d. Serializer 	1.67	C03
Q 41	<p>Suppose you want to transport critical financial tranction data from RDBMS to hadoop using Kafka. Which send method would you use?</p> <ul style="list-style-type: none"> a. Fire and forget b. Synchronous send c. Asynchronous send d. Can not transport data into hadoop 	1.67	C03
Q 42	<p>Which send method is the fastest?</p> <ul style="list-style-type: none"> a. Fire and forget b. Synchronous send c. Asynchronous send d. None 	1.67	C03
Q 43	<p>Which send method is the robust?</p> <ul style="list-style-type: none"> a. Fire and forget b. Synchronous send c. Asynchronous send d. None 	1.67	C03
Q 44	<p>Which error a producer may occure after sending a message?</p> <ul style="list-style-type: none"> a. TimeoutException b. SerializationException c. InterruptException d. Retriable 	1.67	C03

Q 45	Which parameter control the amount of time to wait for additional messages before sending the current batch? a. client.id b. linger.ms c. timeout.ms d. request.timeout.ms	1.67	CO3
Q 46	Which of the following provides the high availability and scalability to a consumer group? a. Partitioning b. replication c. rebalancing d. increasing brokers in the kafka cluster	1.67	CO4
Q 47	who helps consumers to maintain membership in a consumer group and ownership of partitions? a. A group Leader b. group coordinator c. Zookeeper d. Controller	1.67	CO4
Q 48	Suppose your applications that take longer to process records.what action should you take ? a. tune max.poll.interval.ms b. tune fetch.min.bytes c. tune fetch.max.wait.ms d. tune fetch.min.bytes	1.67	CO4
Q 49	If more than _____ passes without the consumer sending a heartbeat to the group coordinator, it is considered dead. a. session.timeout.ms b. max.partition.fetch.bytes c. max.poll.records d. fetch.min.bytes	1.67	CO4
Q 50	consumer commit an offset by producing a message to a special topic which is called_____. a. Topic_offsets b. consumer_offsets c. partition_offsets d. message_offsets	1.67	CO4

Q 51	Which send method uses callbacks? a. Fire and forget b. Synchronous send c. Asynchronous send d. None	1.67	CO4
Q 52	The controller is responsible for _____. a. creating partition b. electing leaders among the partitions c. Maintaining group membership d. creating a consumer group	1.67	CO4
Q 53	Which command is used in Kafka to add messages to a topic? a. kafka-create-message.sh b. kafka-add-message.sh c. kafka-console-producer.sh d. kafka-producer.sh	1.67	CO4
Q 54	A Kafka topic is setup with a replication factor of 5. Out of these, 2 nodes in the cluster have failed. Business users are concerned that they may lose messages. What do you tell them? a. They need to stop sending messages till you bring up the 2 servers b. They need to stop sending messages till you bring up at least one server c. They can continue to send messages as there is fault tolerance of 4 server failures. d. They can continue to send messages as you are keeping a tape back up of all the messages	1.67	CO4
Q 55	A kafka cluster has 20 nodes. There are 5 topics created, each with 6 partitions. How many total number of broker processes will be running? a. 20 processes, one on each node b. 100 processes, one process for each topic on each node. c. 30 processes, one process for each topic and partition. d. 120 processes, one process for each partition on each node	1.67	CO4
Q 56	You have tested that a Kafka cluster with five nodes is able to handle ten million messages per minute. Your input is likely to increase to twenty five million messages per minute. How many more nodes should be added to the cluster? a. 15 b. 13 c. 8 d. 5	1.67	CO4

Q 57	<p>Which of the following best describes the relationship between ZooKeeper and partial failures?</p> <ul style="list-style-type: none"> a. ZooKeeper eliminates partial failures b. ZooKeeper causes partial failures c. ZooKeeper detects partial failures d. ZooKeeper provides a mechanism for handling partial failures 	1.67	CO4
Q 58	<p>Replication of data can result in improved fault tolerance. Which of the following is a disadvantage of replication?</p> <ul style="list-style-type: none"> a. Inconsistent states b. Loss of data c. Deadlocks d. Partial failures 	1.67	CO4
Q 59	<p>Which server should be started before starting Kafka server?</p> <ul style="list-style-type: none"> a. Kafka producer b. Kafka consumer c. Kafka topic d. ZooKeeper server 	1.67	CO4
Q 60	<p>_____ is the subset of the replicas list that is currently alive and caught-up to the leader.</p> <ul style="list-style-type: none"> a. Replicas b. Leader c. Followers d. ISR 	1.67	CO4