Thank you for taking the Week 5: Assignment 5.

Week 5: Assignment 5

Your last recorded submission was on 2024-09-25, 22:15 IST Due date: 2024-09-25, 23:59 IST.

1) What distributed graph processing framework operates on top of Spark? ***1 point***

- ○ MLlib
- ◉ GraphX
- ○ Spark streaming
- ○ ALL

2) Which of the following frameworks is best suited for fast, in-memory data processing and supports advanced analytics such as machine learning and graph processing? ***1 point***

- ○ Apache Hadoop MapReduce
- ○ Apache Flink
- ○ Apache Storm
- ◉ Apache Spark

3) A financial institution needs to analyze historical stock market data to predict market trends and make investment decisions. Which Big Data processing framework is best suited for this scenario? ***1 point***

- ◉ Apache Spark
- ○ Apache Storm
- ○ Hadoop MapReduce

- Apache Flume

4) A telecommunications company needs to process real-time call logs from millions of subscribers to detect network anomalies. Which combination of Big Data tools would be appropriate for this use case? ***1 point***

- Apache Hadoop and Apache Pig
- Apache Kafka and Apache HBase
- Apache Spark and Apache Hive
- Apache Storm and Apache Pig

5) Do many people use Kafka as a substitute for which type of solution? ***1 point***

- log aggregation
- compaction
- collection
- all of the mentioned

6) Which of the following features of Resilient Distributed Datasets (RDDs) in Apache Spark contributes to their fault tolerance? ***1 point***

- DAG (Directed Acyclic Graph)
- In-memory computation
- Lazy-evaluation
- Lineage information

7) Point out the correct statement. ***1 point***

- Hadoop do need specialized hardware to process the data
- Hadoop allows live stream processing of real-time data
- In the Hadoop mapreduce programming framework output files are divided into lines or records
- None of the mentioned

8) Which of the following statements about Apache Pig is true? ***1 point***

- Pig Latin scripts are compiled into HiveQL for execution.
- Pig is primarily used for real-time stream processing.
- Pig Latin provides a procedural data flow language for ETL tasks.
- Pig uses a schema-on-write approach for data storage.

9) An educational institution wants to analyze student performance data stored in HDFS and generate personalized learning recommendations. ***1 point***

Assessment submitted.
X

Which Hadoop ecosystem components should be used?

- ○ Apache HBase for storing student data and Apache Pig for processing.
- ○ Apache Kafka for data streaming and Apache Storm for real-time analytics.
- ○ Hadoop MapReduce for batch processing and Apache Hive for querying.
- ◉ Apache Spark for data processing and Apache Hadoop for storage.

10) A company is analyzing customer behavior across multiple channels (web, mobile app, social media) to personalize marketing campaigns. Which technology is best suited to handle this type of data processing? ***1 point***

- ○ Hadoop MapReduce
- ○ Apache Kafka
- ◉ Apache Spark
- ○ Apache Hive

You may submit any number of times before the due date. The final submission will be considered for grading.

Submit Answers