VANTAGEPOINT
I.T. TRAINING & CONSULTING
Technology Innovation Centre (UTECH), 237 Old Hope Road, Kingston 6, Jamaica W.I.
Tel: 876-970-0197, Fax: 927-1925, Email: info@vantagepointitc.com Website: www.vantagepointitc.com

Exam 70-475: Designing and Implementing Big Data Analytics Solutions

Skills measured

This exam measures your ability to accomplish the technical tasks listed below. The percentages indicate the relative weight of each major topic area on the exam. The higher the percentage, the more questions you are likely to see on that content area on the exam.

Please note that the questions may test on, but will not be limited to, the topics described in the bulleted text.

Design big data batch processing and interactive solutions (25–30%)

- Ingest data for batch and interactive processing
    - Ingest from cloud-born or on-premises data, store data in Microsoft Azure Data Lake, store data in Azure BLOB Storage, perform a one-time bulk data transfer, perform routine small writes on a continuous basis
- Design and provision compute clusters
    - Select compute cluster type, estimate cluster size based on workload
- Design for data security
    - Protect personally identifiable information (PII) data in Azure, encrypt and mask data, implement role-based security
- Design for batch processing
    - Select appropriate language and tool, identify formats, define metadata, configure output
- Design interactive queries for big data
    - Provision Spark cluster, set the right resources in Spark cluster, execute queries using Spark SQL, select the right data format (Parquet), cache data in memory (make sure cluster is of the right size), visualize using business intelligence (BI) tools (for example, Power BI, Tableau), select the right tool for business analysis

Design big data real-time processing solutions (25–30%)

- Ingest data for real-time processing
    - Select data ingestion technology, design partitioning scheme, design row key of event tables in HBase

Technology Innovation Centre (UTECH), 237 Old Hope Road, Kingston 6, Jamaica W.I.
Tel: 876-970-0197, Fax: 927-1925, Email: info@vantagepointitc.com Website: www.vantagepointitc.com

- Design and provision compute resources
    - Select streaming technology in Azure, select real-time event processing technology, select real-time event storage technology, select streaming units, configure cluster size, assign appropriate resources for Spark clusters, assign appropriate resources for HBase clusters, utilize Visual Studio to write and debug Storm topologies
- Design for Lambda architecture
    - Identify application of Lambda architecture, utilize streaming data to draw business insights in real time, utilize streaming data to show trends in data in real time, utilize streaming data and convert into batch data to get historical view, design such that batch data doesn't introduce latency, utilize batch data for deeper data analysis
- Design for real-time processing
    - Design for latency and throughput, design reference data streams, design business logic, design visualization output

Design Machine Learning solutions (20–25%)

- Create and manage experiments
    - Create, manage, and share workspaces; create training experiment; select template experiment from Machine Learning gallery
- Determine when to pre-process or train inside Machine Learning Studio
    - Select model type based on desired algorithm, select technique based on data size
- Select input/output types
    - Select appropriate SQL parameters, select BLOB storage parameters, identify data sources, select HiveQL queries
- Apply custom processing steps with R and Python
    - Visualize custom graphs, estimate custom algorithms, select custom parameters, interact with datasets through notebooks (Jupyter Notebook)
- Publish web services
    - Operationalize Azure Machine Learning models, operationalize Spark models using Azure Machine Learning, operationalize custom models

Operationalize end-to-end cloud analytics solutions (25–30%)

- Create a data factory

Technology Innovation Centre (UTECH), 237 Old Hope Road, Kingston 6, Jamaica W.I.
Tel: 876-970-0197, Fax: 927-1925, Email: info@vantagepointitc.com Website: www.vantagepointitc.com

    - Identify data sources, identify and provision data processing infrastructure, utilize Visual Studio to design and deploy pipelines
- Orchestrate data processing activities in a data-driven workflow
    - Leverage data-slicing concepts, identify data dependencies and chaining multiple activities, model complex schedules based on data dependencies, provision and run data pipelines
- Monitor and manage the data factory
    - Identify failures and root causes, create alerts for specified conditions, perform a restatement
- Move, transform, and analyze data
    - Leverage Pig, Hive, MapReduce for data processing; copy data between on-premises and cloud; copy data between cloud data sources; leverage stored procedures; leverage Machine Learning batch execution for scoring, retraining, and update resource; extend the data factory with custom processing steps; load data into a relational store, visualize using Power BI
- Design a deployment strategy for an end-to-end solution
    - Leverage PowerShell for deployment, automate deployment programmatically