Big Data Science-Final Exam
Total Time: 3 hours
Total Marks: 40

Please read instructions carefully.

- Plagiarism Review - Turnitin will be on for the final exam. You will get zero marks in case you copy-paste answers from the internet.
- Your exam will be a take-home exam. You must submit the exam on Aug 7 between 12-3 pm. Please make sure you have uploaded your solutions sooner than later to avoid possible network and software problems.
- No extra time will be provided.
- You can use Microsoft Word/Pages etc. to write down all the answers and upload it back in pdf format in Quercus within given timeframe.
- No coding examples are needed for this exam but please make sure you address all the parts of the subjective questions.

Part A=> Multiple choices Qs/Ans. [15 Marks – 1 Mark/Qs]

Qs1: What are two characteristics of the public cloud? Each correct answer presents a complete solution.

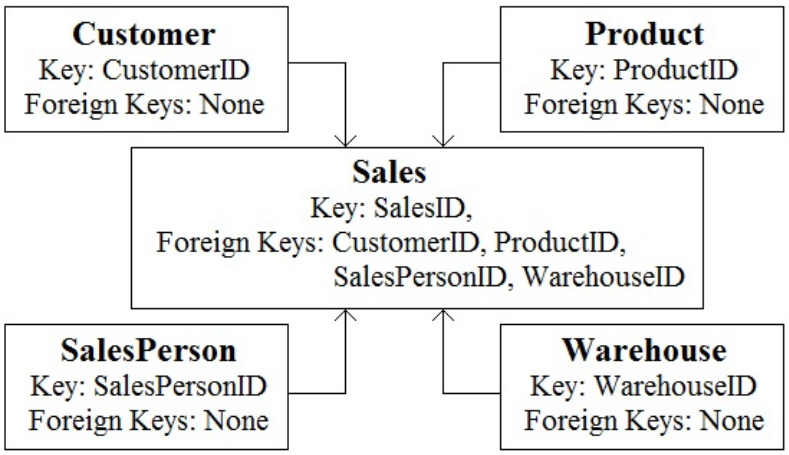
a. dedicated hardware
b. unsecured connections
c. limited storage
d. metered pricing
e. self-service management

Q2: The company's migration plan states that only Platform as a Service (PaaS) solutions must be used in Azure. You need to deploy an Azure environment that meets the company migration plan.

Solution: You create an Azure App Service and Azure virtual machines that have Microsoft SQL Server installed.
Does this meet the goal?
Options:

a. Yes
b. No

Q3: You are reviewing the data model shown below.



The data model is a **[answer choice]**.

- transactional model
- star schema
- snowflake schema

Customer is a **[answer choice]** table.

- fact
- dimension
- bridge

Qs4: To complete the sentence, select the appropriate option in the answer area.

A relational database is appropriate for scenarios that involve a high volume of

- changes to relationships between entities
- geographically distributed writes
- transactional writes
- writes that have varying data structures

Qs5: To complete the sentence, select the appropriate option in the answer area.

A visualization that shows a university's current student enrollment versus the maximum capacity is an example of

- cognitive
- descriptive
- predictive
- prescriptive

analytics.

Qs6: Choose Yes or No option for below statements:

	Yes	No
Platform as a service (PaaS) database offerings in Azure require less setup and configuration effort than infrastructure as a service (IaaS) database offerings.	○	○
Platform as a service (PaaS) database offerings in Azure provide administrators with the ability to control and update the operating system version.	○	○

Qs7: Match the types of data to the appropriate Azure data services.

Data Types

- Image files
- Key/value pairs
- Relationships between employees

Answer Area

Data type	Azure Blob storage
Data type	Azure Cosmos DB Gremlin API
Data type	Azure Table storage

Qs8: What are two characteristics of real-time data processing? Each correct answer presents a complete solution.

a. Data is processed periodically
b. Low latency is expected
c. High latency is acceptable
d. Data is processed as it is created

Qs9: Choose Yes or No option for below statements:

	Yes	No
a. IaaS database offerings in Azure require less setup and configuration effort than PaaS database offerings		
b. IaaS offerings in Azure provide administrators with ability to control and update the operating system version		

Qs10: Match the types of workloads to the appropriate scenarios.

Workload Types

- Batch
- Streaming

Answer Area

Workload type	Data for a product catalog will be loaded every 12 hours to a data warehouse.
Workload type	Data for online purchases will be loaded to a data warehouse as the purchases occur.
Workload type	Updates to inventory data will be loaded to a data warehouse every 1,000 transactions.

Qs11: To complete the sentence, select the appropriate option in the answer area.

Transparent Data Encryption (TDE) encrypts [▼]

- a column to protect data at rest and in transit.
- queries and their results in order to protect data in transit.
- the database to protect data at rest.
- the server to protect data at rest.

Qs12: To complete the sentence, select the appropriate option in the answer area.

Relational data uses [▼] to enforce relationships between different tables.

- collections
- columns
- keys
- partitions

Qs13: You have an e-commerce application that reads and writes data to an Azure SQL database. Which type of processing does the application use?

a. stream processing
b. batch processing
c. Online Analytical Processing (OLAP)
d. Online Transaction Processing (OLTP)

Qs14: For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statements	Yes	No
Relational database tables contain columns and rows	○	○
Indexes in a relational database describe the data types in a table	○	○
A database view is a virtual table whose content is defined by a query	○	○

Qs15: An application will use Microsoft Azure Cosmos DB as its data solution. The application will use the Cassandra API to support a column-based database type that uses containers to store items. You need to provision Azure Cosmos DB. Which container name and item name should you use? Each correct answer presents part of the solutions. Choose two options.

a. collection
b. rows
c. graph
d. entities
e. table

Part B: Explain below questions in 5 or more lines. [25 Marks – 5 Mark/Qs]

1. Explain Real-time Analytics and provide an appropriate technology available in Azure for real-time analytics and its main features.
2. Explain the difference between OLAP and OLTP. Explain one technology example for each of them.
3. Describe different consistency levels of Azure Cosmos DB.
4. What is the difference between ETL and ELT. Explain one of the Azure data integration offering briefly with its main components.
5. If you need to migrate 3 TB of data from On-Prem SQL Server to Azure cloud with easy lift and shift option with administrating operating system, what technology you will choose and why. Explain that technology and its capabilities.