



AZURE DATA ENGINEER ASSOCIATE (DP-203)

- Learn at your own place
- Taught by industry professionals
- Interactive classes

Join us today!!



ZETLAN TECHNOLOGIES

www.zetlantech.com

Follow us on:



Course Modules



1. Implement a partition strategy

- Implement a partition strategy for files
- Implement a partition strategy for analytical workloads
- Implement a partition strategy for streaming workloads
- Implement a partition strategy for Azure Synapse Analytics
- When partitioning is need in Azure Data Lake Storage Gen2

2. Design and implement the data exploration layer

- Create and execute queries by using a compute solution
- Implement Azure Synapse Analytics database templates
- Push new or updated data lineage to Microsoft Purview
- Browse & search metadata in MS Purview Data Catalog

3. Ingest and transform data

- Design and implement incremental loads
- Transform data by using Apache Spark
- Transform data by usg Transact-SQL in Az Synapse Analytics
- Ingest and transform data by using Azure Synapse Pipelines
- Transform data by using Azure Stream Analytics
- Cleanse data
- Handle duplicate data
- Avoiding duplicate data by using Azure Stream Analytics
- Handle missing data
- Handle late-arriving data
- Split data



4. Develop a batch processing solution

- Develop batch processing solutions
- Use PolyBase to load data to a SQL pool
- Implement Azure Synapse Link and query the replicated data
- Create data pipelines
- Scale resources
- Configure the batch size
- Create tests for data pipelines
- Integrate Jupyter or Python notebooks into a data pipeline
- Upsert data
- Revert data to a previous state
- Configure exception handling



5. Develop a stream processing solution

- Stream process soln by using Stream Analytics & Az Event Hubs
- Process data by using Spark structured streaming
- Create windowed aggregates
- Handle schema drift
- Process time series data
- Process data across partitions
- Process within one partition
- Configure checkpoints and watermarking during processing
- Scale resources
- Create tests for data pipelines
- Optimize pipelines for analytical or transactional purposes
- Configure exception handling



6. Manage batches and pipelines

- Trigger batches
- Handle failed batch loads
- Validate batch loads
- Manage data pipelines in Azure Data Factory
- Schedule data pipelines in Data Factory
- Implement version control for pipeline artifacts
- Manage Spark jobs in a pipeline



7. Implement data security

- Implement data masking
- Encrypt data at rest and in motion
- Implement row-level and column-level security
- Implement Azure role-based access control (RBAC)
- Implement POSIX-like access control lists (ACLs)
- Implement a data retention policy
- Implement secure endpoints (private and public)
- Implement resource tokens in Azure Databricks
- Load a DataFrame with sensitive information
- Write encrypted data to tables or Parquet files
- Manage sensitive information

8. Monitor data storage and data processing

- Implement logging used by Azure Monitor
- Configure monitoring services
- Monitor stream processing
- Measure performance of data movement
- Monitor and update statistics about data across a system
- Monitor data pipeline performance
- Measure query performance
- Schedule and monitor pipeline tests
- Interpret Azure Monitor metrics and logs
- Implement a pipeline alert strategy



9. Optimize & troubleshoot data storage and data processing

- Compact small files
- Handle skew in data
- Handle data spill
- Optimize resource management
- Tune queries by using indexers
- Tune queries by using cache
- Troubleshoot a failed Spark job
- Troubleshoot a failed pipeline run,



**LEARN
REMOTELY!!**

The efficiency of online learning
in terms of time management,
flexibility, and the ability
to access resources anytime,
anywhere can be compelling.



ZETLAN TECHNOLOGIES
www.zetlantech.com

**For contact: +91 8680961847
+91 9600579474**

