



Azure Data Engineering

 www.semicoretech.com  9700779739,8688018839

.....
1st floor, sri sai Arcade, Near Aditya Trade Center,
Ameerpet, hyderabad, Telangana, 500016

Module1: Introduction to Big Data&Cloud Computing

- What is Data?
- How data is generated
- Big Data and it's types: Structured,unstructured and semi-structured
- Information & DataBase
- What is ETL Processing?
- What is On-premise and Cloud
- Challenges of On-premise
- Features Of Cloud
- What is Cloud Computing?
- Architecture of Cloud Computing
- Types of Cloud Services Core Python and Adv.Python
- Python Basics

Module2: SQL

- Introduction and SSMS Installation
- Types of Databases
- DDL,DML(Create tables,insert data and update,delete operations)
- DCL,DQL,TCL(select and grant,revoke commands)
- SQLQueryexecutionOrderandhands-onproblems
- HandlingNulls
- Aliases
- Aggregations-sum,min,max,count,avg
- Group by and having clause & businesss cenarios
- Joins-Inner and left join along with hands-on
- Joins-right and full join with examples
- Joins-self and cross join with examples
- Sub queries-single row,multiple row & cor related
- Nested Sub queries

- Hands-on and assignments
- CTE(Common Table Expressions)with examples
- CTE Real time problems with complex businesss cenarios
- Set Operations-Union,Unionall,Except & intersect
- Case when
- Case when with groupby
- Introduction to window function sand syntax
- Ranking Functions
- Row number
- Rank
- Dense Rank
- Understanding the frame clauses-unbounded,preceding &following
- Aggregations over window functions
- Basicsto Advanced interview questions
- Basics to Advanced interview questions (covering over all of 50 in depth questions along with datasets)

Module3: Python

- Intro and basics
- Installation of Anaconda
- Working with Jupiter notebook
- Python conditional statements-ifal if&else
- Loops-for,while,range
- Data Structures-lists & tuples
- Data Structures-Sets & Dictionaries,list comprehensions
- Functions-definitions,lamba functions
- Static and global scope
- Interview Questions

Module4: Pandas

- Intro and library
- Data frames
- Loading the data set with different file formats
- Data frame operations
- Data transformations-cleaning,handling nulls,removing duplicates and other function
- Aggregations on data frame
- Joining different data frames
- Business Use cases
- Single worker node concept and it's disadvantages
- Interview Questions & Assignment

Module5 : Microsoft Azure

- Introduction to Microsoft Azure
- Create an account and signup for free trial
- Walk through of UI and services
- Resource group and creation
- Blob Storage
- Storage Tiers
- Azure storage Explorer
- Data lake Gen2
- Data lake vs Blob
- Redundancy-LRS, ZRSGRS & Read only GRS
- Performance-Standard and Premium with cost optimization
- Data factory and it's walk through
- Azure Relational database (Azure SQL)
- Azure Data bricks
- Data bricks community Edition

Module6: Pyspark

- Working with Clusters
- Cluster creation
- Types of cluster-single node,shared & isolation
- Worker and driver node
- Cluster Run time<S
- Cluster auto termination
- Data bricks Architecture
- Working with data frames
- Medalli on architecture-bronze , silver & gold
- Data frame operations
- Aggregations
- Derived columns
- In built functions-lit,with Column,With Column Re named
- Working with group by
- Schema Evaluation-Struct Field &Struct Type
- Write modes of dataframe-append & over write
- Working with DBFS
- With Column with when
- Window functions over data frame columns
- Joining multiple data frames
- Masking sensitive data
- Working with UDF(User Defined Functions)
- Explode-convert list of elements to rows
- Working with Json Parsing
- Select and aliases
- Walk through of Pyspark documentation with other functions necessary along with business scenarios
- Interview questions and assignment

Module7: Data Warehouse

- OLTP & OLAP
- Components of data warehouse-data mart,meta data,EtL
- Normalization & its types
- Denormalization
- Data Lake vs Data Warehouse
- Facts & dimensions
- Types of Dimensions
- Slowly changing dimensions(SCD)&its types
- Data modelling & its types
- DM Process & Techniques
- Star & Snow flake schema
- ER diagram
- Indexes-clustered & non clustered
- Performance optimization
- Dimensions of DQ
- Common data quality issues and handling them
- DQ tools
- Demo on DQ

Module8 : Spark Architecture

- Understanding the application manager
- Hadoop vs Spark
- Spark design
- Application master container with main() & py4J module
- Spark Deploy Modes
- Spark API's-transformation & actions
- Job sands tages

Module9: Azure Data Factory

- Intro and walk through of ADF
- Working with data set,linked service and integration run time.
- Creating pipelines
- Copy activity
- Look up
- Get meta data
- if condition
- For loop
- Wait
- Until
- Execute Pipeline
- Working with SQL script
- Stored Procedures
- Creating dynamic content
- Creating one data set for one storage account and using it across multiple pipe lines
- Triggers-schedule, tumbling and event based
- Types of integration runtime
- Attaching triggers to pipeline
- Monitor
- Data Factory linked services
- Realtime problem deck
- Interview questions

Module10 : E2E project (F1 Racing)

- Walk through on data sets and business problem
- Project design and approach
- Project requirements
- Understanding data bricks secrets and key vault
- Understanding the old fashion mounting
- Service principal and other mount point setup
- Unity Cata log and its advantages over mount
- Creating unity cata log enabled workspace
- Creating meta store and establishing the connection to ADLS
- Understanding meta store-catalog,data base and tables
- Data Ingestion to bronze
- Silver Layer transformations
- Gold layer table creation as per business requirements
- Optimizing the code to create utilities,calling one note book from another
- Create includes notebook
- Understanding the DBU til sand widgets
- Converting entire logic to handle increment alload □ Handling missing files to avoid failures
- Delta Tables-managed and external
- Batch load & upsert
- Orche strating the entire process
- Choosing the right trigger to load entire data and to handle incremental process

Module 11: Azure Repos

- Understand the Devops Portal
- Walk through on all sections-boards,work items,repos, project management
- Creating a project repo
- Understanding the branches-local,feature,dev,qaandprod
- Creating and integrating gitto workspace
- Creating local branch and pushing all the notebooks/code
- Raising pull request(PR)
- Understand the industry standards for PR
- Setup the CI/CD process

Module 12: Azure Stream Analytics

- Intro & basics
- Understand the read stream and micro batches
- Background query
- Sources and sinks
- Write stream
- Output modes
- Triggers & its types
- Auto Loader
- Schema enforcement,evaluation & evolution
- E2E project: Vehicle tracking management
 - Data understanding
 - Loading the data
 - Transform the data
 - Ingestion to gold layer
 - Create a workflow
 - Trigger sand monitoring the job

Module13: Azure Synapse Analytics

- Creation of service
- How to create workspace
- Creating data base and tables
- Using synapse note books
- Create synapse pipe lines

Module14 : Azure Fabric

- Introduction
- Process to sign up and account creation
- Work spaces
- One lake
- Role sand accessin workspace
- Lake house
- Connecting Azure Data lake to Fabric
- Creating shortcuts
- Git on fabric

Module15: Projects & Scenarios

- Healthcare domain,Retail domain
- Project challenge sand how to answer the min interview
- Scenario of tough situation and solution designed based questions for interview
- Realtime E2E project showcase to client-Single Stack Solution
 - APPTdesignflowtoshowrealtimedeckfollowedinrealtimeto design single stack solution presenting it to the end users

OUR CERTIFICATES





CERTIFICATE OF COMPLETION

THIS CERTIFICATE IS AWARDED TO

has completed course on _____

from _____ *to* _____ *successfully*

we wish this intern all the best for future endeavours

www.semicoretech.com Authorised Signature





CERTIFICATE OF INTERNSHIP

THIS CERTIFICATE IS AWARDED TO

has completed internship on _____

from _____ *to* _____ *successfully*

we wish this intern all the best for future endeavours

www.semicoretech.com Authorised Signature

Thank You
- Team Semicore

