

FORUMDE

We rise by lifting others.
Website:- <https://forumde.in/>
Contact Us:- 7870970617

100 DAY SELF STUDY MASTER PLAN TO BECOME AZURE DATA ENGINEER(ARUN KUMAR)

Day	Python	SQL	Azure
Day1		Database, Datawarehouse and Data Lake	What is Cloud?
Day1		Type of databases	What are the different major Cloud Platforms in the market?
Day1		EDWMS	What are Low-PaaS and SaaS?
Day1		Logical Schema Designing	Pricing Model in Azure
Day1		Data modelling	Create Account in Azure Portal(You will need credit card. Rs 2 will be deducted for authentication purpose and that will be returned back.)
Day2		Star Schema	Azure Portal Overview
Day2		Fact & Dimension table	Overview of Different Services relevant to Data Engineering in Azure
Day2		Normalization	Understanding what is Big Data.
Day2		Normal Forms	Provision ADLS Gen2
Day3		Sequence Diagram	Creating containers, directory and uploading files in ADLS Gen2
Day3		E-R Diagram	Provision Azure Databricks and its overview
Day4		FluxCharts	Creating clusters in Databricks
Day4		Physical Database Design	
Day5		Business Constraints - Primary key, Foreign Key, Candidate Key, Composite Key, Surrogate Key, Unique Key, Check, Default, Not Null	Widgets Utility
Day5		Index- Clustered and Non-Clustered	Widgets Utility
Day6			File System Utility, Understanding Cluster Configuration in Databricks
Day7	Revision		
Day8		Download and Install SSMS, SQL Server	Databricks CLI overview
Day8		SQL commands- DDL, DML, DCL, TCL	Service Principal
Day8		SQL commands- DDL, DML, DCL, TCL	Mount Point
Day9		SQL commands- DDL, DML, DCL, TCL	How to create Mount Point using Azure Key Vault
Day10		Select, INSERT, Update, Delete, Merge, Create, Alter, Truncate, Rename, Grant, Revoke, Commit, Rollback, Save point, Drop	Connect ADLS Gen2 from databricks using OAuth 2.0 with an Azure service principal
Day10		Query Execution sequence/Order	Connect ADLS Gen2 from databricks using Shared access signatures (SAS)
Day11		Clause, Operators, Predicates in SQL	Connect ADLS Gen2 from databricks using Account keys
Day11		LIKE, IN, LIMIT, TOP, CASE, INSTRING, SUBSTRING, AVERAGE, COUNT* MAX, MIN, SUM	Read and write data in Azure Databricks
Day12		UPPER, LOWER, CAST, EXTRACT, Wild Card Operators Etc.	Data processing in Azure Databricks
Day12		SUBQUERIES - Nested Subqueries, Correlated Subqueries, Inline View	Working with Dataframes in Azure Databricks
Day13		JOINs - Self Join, Inner Join, Left Outer Join, Right Outer Join, Cross Join	Working with different types of file formats like CSV, Parquet, Avro, Orc
Day14	Revision		
Day15		SET Operations - Union, Union All, Minus, Intersect	
Day15		CTE (Common Table Expression)	More dbutils commands
Day16		Window Function - Rank, Dense Rank, Row number, Lead, Lag, First Value, Last Value, Percent Rank	Databricks Runtime
Day17		Pivot and Unpivot	Concept of Delta Table
Day18		Rollup, Cube, Grouping Sets	Deep Dive into Delta Tables
Day19		Views, Materialized View	Managed and Unmanaged Table in Databricks
Day20			Table batch read and writes
Day20			Table delete, update, and merge
Day21	Revision		Different types of cluster in databricks
Day22			Scheduling jobs in Databricks Notebook
Day22	Basics of Python programming, working with Python interpreter identifiers		Azure Key Vault
Day22	Keywords contains variables,		Provisioning of Azure Key Vault
Day22	types of operators, precedence of operators data types		How to securely store Keys and Password in Azure Key-Vault
Day23	mutable and immutable data types, statements		How to access Keys and Password for Azure Key-Vault
Day23	expressions		How to connect AKV from ADB(Azure Databricks) and read the secrets of AKV
Day24	evaluation and comments input and output statements data type conversion		Secret Scope and databricks utility of secret scope
Day25			How to connect AKV from ADB(Azure Data Factory) and read the secrets of AKV.
Day25	selection (decision) and repetition (iteration)		Azure Data Lake Gen2(ADLS Gen2)
Day26	Chapter 3: Functions		Blob Storage
Day26	Introduction to		Azure Data Lake Gen1
Day27			Azure Data Lake Gen2
Day27			Installation of Azure Storage Explorer and connection with ADLS Gen2
Day27			Difference between Blob Storage, ADLS Gen1 and ADLS Gen2
Day27			How do we store data in ADLS
Day27			How to read/write data in ADLS Gen2 from databricks
Day28	Revision		
Day29	need of functions User defined functions: passing arguments to a function, returning values from functions,		
Day29	scope of variables		
Day30	Standard library, using built-in functions, importing modules-math, random, statistics, creating and importing user defined module		ETL CONCEPT
Day31	Periods, Strings: Initializing strings and accessing strings string operations		Data Warehouse Concept
Day32	built-in functions for string manipulation string traversal		Difference between Database, Data Warehouse and Data Lake
Day34	Periods, Lists: list operations - creating, initializing, traversing and manipulating lists list methods and built-in functions		OLAP vs OLTP systems Concept
Day33	nested lists		Fact Table and Dimension Table
Day34			
Day36	list as argument to a function		Data Modelling Star Schema vs Snowflake Schema Concept
Day37	Tuples: Creating, initializing, accessing elements, tuple assignment, operations on tuples, tuple methods and built-in functions, nested tuples.		
Day38	Dictionary: concept of key-value pair, mutability, creating, initializing, traversing, updating and deleting elements, dictionary methods and built-in functions.		
Day39	Exception handling		SPARK
Day40	need of exception handling user-defined exceptions raising exceptions		Architecture of Spark
Day41	try, except, else clause try-finally clause		Job-Drigon and Task in Spark
Day42	Revision		
Day43	File Handling: text file, file types open and close files		SparkSQL concepts, PySpark concepts
Day44	reading and writing text files		Transformation and Action
Day45, Day46, Day47	Modules		Lazy Evaluation
Day48	Packages		rdd vs dataframe vs dataset
Day49	Revision		Immutability Concept
Day50, Day51, Day52	Regular Expressions		operation on RDD
Day53, Day54, Day55	Multi Threading		operation on dataframe
Day56	Revision/Start Thinking on Resume Preparation		Difference file formats like parquet, avro, orc and their differences
Day57, Day58	Connecting Databases with Python		Understanding Shuffling in Spark
Day59			Narrow Transformation
Day60			Wide Transformation
Day61			Accumulator
Day62			Broadcast variable
Day63	Revision/Prepare first round of resume		Partition By
Day64			
Day64			partition pruning, repartition and coalesce
Day64			cachet vs persist
Day65			Azure SQL Server
Day66			Provisioning of Azure SQL Server
Day67			Installation of SSMS and connection with Azure SQL Server
Day68			Installation of Azure Data Studio and connection with Azure SQL Server
Day69			Connect Azure SQL server from databricks
Day70			Connect Azure SQL server from Azure Data Factory
Day71, Day72, Day73			Azure Data Factory(ADF)
Day72			Pipelines
Day73			Activities
Day74			Datasets
Day75			Linked Services
Day76			
Day77	Revision/Make changes in the resume if required		Integration Runtimes
Day78			AutoResolve Infrastructure Runtime
Day79			Self-hosted Infrastructure Runtime(SelfIR)
Day80			SQL Infrastructure Runtime
Day81			Pipeline Run
Day82			Scheduling and Trigger of pipeline
Day83			Parameters
Day84			Variables
Day85			Orchestration of Pipeline in ADF
Day86			Triggers
Day87			Schedule Trigger and its System Variables
Day88			Event Based Trigger and its System Variables
Day89			Tumbling Window Trigger and its System Variables
Day90			
Day91	Revision/Make changes in the resume if required		Data Flow Pipeline
Day92, Day93			Pipeline Level Parameters
Day94, Day95, Day96, Day97	Creating data pipelines using all major components		Global Parameters
Day98	Revision/Make changes in the resume if required		Creating Dynamic Pipeline creation
Day99	Final Resume preparation/Mock interview		
Day100	Final Resume Preparation/Mock Interview		
			How to monitor ADF data pipelines?
			ADF Template
			Copy one table from My SQL(on-premise) to Azure SQL Server.
			Copy filtered data of one table from My SQL(on-premise) to Azure SQL Server.
			Copy All tables from My SQL Server(on-premise) to Azure SQL Server
			Copy selected tables from My SQL Server(on-premise) to Azure SQL Server
			Pipeline for Incremental Load
			Copy all tables from Azure SQL Database to ADLS Gen2
			Copy selected tables from Azure SQL Database to ADLS Gen2
			Error Logging of ADF pipelines in Azure SQL Server
			Azure Logic App
			How to send mail alert for pipeline failure and success using logic app.
			Deployment
			Introduction to Deployment
			DEV, UAT and PROD environment
			Azure DevOps
			Using CI/CD pipelines for deployment of Azure Databricks and Azure Data Factory

GET SET READY TO GIVE INTERVIEW!! ALL THE BEST!!!			

Tips:-

- 1)Resume preparation is a continuous process.It will take multiple iteration of resume building to build a robust resume.
- 2)Consider each failed interview as the mock interview and prepare the set of question you were not able to answer before sitting in the next interview.
- 3)Keep on practicing different sets of question in Python and SQL during the entire process of learning from any portal.
- 4)Continuous practice and revision is the key to learn technology and crack interviews.
- 5)That day will never come e