8 + 8 + 8 Rule for 24hr day

```
8 hrs - Hardwork
8 hrs - Good Sleep
8 hrs - 3F + 3H + 2S
    3F - Family, Friends, Faith
    3H - Health, Hygine, Hobby
    2S - Soul & Smile
```

Road Map to DS

01. Programming Languages :

01. Python Lang :

- DataTypes, DS,
- Loops, CtrlStmnts
- Functions, Modules
- Viz(MatplotLib & Seaborn)
- OOP&Classes, Exceptions,
- Data Wrangling & Cleaning
- Image & Audio Files(Tkinter)
- PVM, Logging

02. R Lang :

- reading & Exporting Data,
- DataTypes, DS
- Manipulation & Processing
- CtrlStmnts [see if R Lang has loops]
- Functions
- Objects Within Objects
- Packages(Tidyverse, Dplyr, Tidyr, GGPlot)
- QueueingTheory

02. DataBases :

01. DataBases : SQL

01. Databases (SQL) :

1. Basic Sql Syntax
2. Data Types
   - String Types
   - Number Types
   - DateTime Types
   - Conversion Function - CAST as INT/FLOAT/STRING
3. DDL
   - CREATE, ALTER, DROP, TRUNCATE
4. DML
   - SELECT
   - FROM, WHERE, ORDER BY clauses
   - GROUP BY, HAVING
   - JOIN

- INSERT, UPDATE, DELETE

5. Aggregation Functions
    - SUM, AVG, COUNT, MIN, MAX
    - GROUP BY, HAVING clauses
6. CONSTRAINTS
    - PRIMARY KEY
    - FOREIGN KEY
    - UNIQUE
    - NOT NULL
    - CHECK
7. Joins
    - INNER, LEFT, RIGHT
    - OUTER / FULL OUTER
    - SELF
    - CROSS
8. Sub-Queries
    - Types
        - Scalar, Column, Row, Table
    - Nested Sub-Queries
    - Correlated Sub-Queries
9. String Functions
    - CONCAT
    - LENGTH
    - SUBSTRING, REPLACE,
    - UPPER, LOWER
10. Date Time Functions
    - DATE, TIME, TIMESTAMP, DATEPART, DATEADD, EXTRACT
    - Adding, Subtracting, Extracting Year/Month/Day
    - Formatting Date Time
11. Numeric Functions
    - ROUND, TRUNCATE
    - CEILING, FLOOR
    - ABS, MOD
12. Conditional Functions
    - CASE, COALESCE, NULLIF
13. Views
    - Create, Modify, Drop views
14. Indexes
    - CREATE INDEX
    - Optimize Query Using Index
15. Transactions
    - ACID properties
    - BEGIN, COMMIT
    - ROLLBACK, SAVEPOINT
    - Isolation Levels
16. Data Integrity & Security
    - Data Integrity Constraints
        - Referential Integrity
        - Entity Integrity
    - Permissions
        - GRANT
        - REVOKE
    - DB Security Best Practices

17. Stored Procedures
    - CREATE PROCEDURE
    - EXEC
18. Stored Functions
    - Create Functions
    - Using Functions In Queries
19. TRIGGERS
    - Creating & Using Triggers
20. PARTITION
    - Creating & Using Partitions
21. Regular Expressions
    - REGEXP '.', '*', '+', '?', '^', '$'
22. Schema Management
    - CREATE, ALTER, DROP schema
23. Error Handling
    - Try-CatchBlocks
    - Raising Custom Exceptions
24. Performance Optimization
    - Query Optimization Techniques
        - Indexes
        - Optimizing Joins
        - Reducing Sub-Queries
    - PerformaceTuningBestPractices
25. Advanced Sql Concepts
    - Window Functions
        - Numbering Functions
            - ROW_NUMBER()
            - RANK(), DENSE_RANK()
        - LEAD(), LAG()
    - CTEs (Common Table Expressions)
    - Recursive Queries
    - Pivot & Unpivot Operations
    - DynamicSQL
26. Merge Statement
    - for upsert
27. Qualify Statement
    - used in Teradata
28. Sequences And Identity Columns
    - Creating & Using Sequences/IdentityColumns
29. Advanced Sql Types
    - BLOB, CLOB, ENUM, SET, etc.
30. Temporal Tables
    - Creating & UsingTemporalTables
31. Cursors
    - Understanding Cursors
    - Using Cursors - DECLARE, OPEN, CLOSE, DEALLOCATE

01. Databases (SQL : MySQL)

02. Databases (SQL : PostgreSQL)

03. Databases (SQL : T-SQL / Transact-SQL / MS-SQL)

02. DataBases : NoSQL

- Data Models
  - XML
  - JSON

01. Databases (No-SQL : MongoDB)

- MongoDB Architecture
- MongoDB vs. Cassandra

02. Databases (No-SQL : Cassandra)

- Cassandra Architecture
- MongoDB vs. Cassandra

03. DataBases : Concepts

- Data Warehousing
- Connecting DBs with Python
- Data Driven Decisions
- Enterprise Data Management
- Data Preparation
- Data Cleaning

03. Big Data :

- 5 V's of Big Data
- Background of Big Data

01. Data Warehouse

- Data Warehouse Fundamentals
  - OLAP vs. OLTP
  - Dimension Table vs. Fact Table
  - ER Modelling / Dimension Modelling
  - ETL vs ELT
    - Data Warehouseing & Data Lakes
    - Data Warehouse vs. Data Lake vs. Data Mart
- Data Warehouse Tools
  - SnowFlake, BigQuery, Amazon RedShift

02. Batch Processing

- Apache Hadoop
- Apache Spark
- Apache Flink

01. Apache Hadoop

- Hadoop Ecosystem
- HDFS
  - HDFS Architecture
  - HDFS cmdline & web interface
- Hadoop event stream processing
- Complex Event Processing
- MapReduce

  - MapReduce Architecture [flow / architecture]
  - MapReduce Features
  - Shuffle & Sort
  - Job Scheduling, Task Execution
  - MapReduce Types
- Hadoop Cluster, Specification, Configuration
- Hadoop Administration- Security, Monitoring, Maintainence
- Hive SQL
  - Hive Architecture
  - Hive Programming
- H-Base
  - H-Base Basics
  - H-Base Architecture
  - Java Client API
  - CRUD operations & Security

02. Apache Spark

- Apache Spark Fundamentals
  - Spark Architecture
  - initilializing Spark
  - RDDs
    - RDD operations
    - Key-Value Pairs
    - Shuffle Operations
    - RDD Persistence
    - Removing Data
    - RDD Operations
      - Transformations
        - map, flatMap, filter, etc.
      - Actions
        - count, reduce, reduceByKey, etc.
    - Shared Variables
      - Broadcast
      - Aggregators
  - Spark DataFrames / Spark SQL
    - Operations on DataFrames
      - count, printSchema, show, select, filter, groupBy, registerTempTable
  - Datasets
  - Job Optimization
  - EDA using PySpark
  - Spark MLib
  - Deployement to cluster Spark Streaming
- Apache Spark languages
  - PySpark
  - Spark Java
  - Scala
  - R
- Apache Spark Tools
  - Databricks
  - Amazon EMR
  - GCP Cloud Dataproc
  - AWS Glue

03. Stream/Real-Time Processing

- Apache Kafka

01. Apache Kafka

- Spark Streaming using Kafka
    - Spark-Kafka Integration
    - Setting Up Kafka Producer & Consumer
    - Kafka Connect API
    - MapReduce
- Connecting DBs with Spark

04. ETL Data Pipelines

- Data Warehousing for ETL Data Pipeline
- Data Lake for ETL Data Pipeline
- Apache Hive, Apache Airflow, Apache Beam

05. Advanced Data Engineering Tools

- Tools
    - Learn some tools from Modern Data Stack, focus on core use case of that tool
        - DBT - Data Build Tool
- Deployement
    - Security, Networking & deployment
    - Kubernetes, Docker
- Use Cases
    - look at customer success story/case study of cloud platforms such as AWS, Azure, GCP, and observe how they solved some particular problem using some tool/service

04. Statistics :

01. Statistics (Basic) :

- Basic Terms
- Variables, Random Variables
- Sample & Population
- Sampling & Sampling Techniques
- Population Distribution & Sample Distribution
- Mean, Median, Mode, Range
- Variance, S.D.
- Probability / Data Distributions
    - Gaussian / Normal Distribution
    - Non-Gaussian / Non-Normal Distributions

02. Statistics (Descriptive) :

- Proportions
- Mean, Median, Mode, Range
- Variance, S.D.
- Skewness, Kurtosis
- IQR, BoxPlot, 5 Summary Statistics

03. Statistics (Probability)

- Population, Sample
- Odds
- Probability Theory
- Probability / Data Distributions

    - Gaussian / Normal Distribution
    - Non-Gaussian / Non-Normal Distributions
        - Binomial Distribution
        - Poisson Distribution
- CLT (Central Limit Theorem)
- Conditional Probability, Bayes Theorem
- Estimation
    - OLS

04. Statistics (Inferential) :

- Sampling & Sampling Techniques
- Sample & Population
- 3-SigmaRule
- HypothesisTesting
    - A/B Testing
- One-Tail/Way Test
    - Left-Tail/Way
    - Right-Tail/Way
- Two-Tail/Way Test
- Continuous Data Tests
    - Z-Score, Z-Test, t-Test
        - P-Value, LoS, Confidence Interval
    - ANOVA Test
    - Variance Tests
- Discrete Data Tests
    - χ² Test
    - Proportion Tests
    - Poisson Rate Tests

05. Statistics (Predictive) :

01. Supervised Learning

- (Data with Labels)

01. Regression Models

- (For predicting Continuous values)

1. Linear Regression Model
    1. Simple Linear Regression Model
    2. Multiple Linear Regression Model
    3. Categorical Regression Model
2. Non-Linear Regression Model
3. Logistic Regression Model (*is a Regression as well as Classification model*)
    1. Binary Logistic Regression Model
    2. Nominal Logistic Regression Model
    3. Ordinal Logistic Regression Model
4. Counts Regression Model
    1. Poisson Regression Model
    2. Negative Binomial Regression Model

02. Classification Models

- (for classification of categorical data)

1. Logistic Regression Model (*is a Regression as well as Classification model*)
    1. Binary Logistic Regression Model

        2. Nominal Logistic Regression Model
        3. Ordinal Logistic Regression Model
    2. Sequential / Probabilistic Classifier Algorithms
        1. Gaussian Naive Bayes Classifier
        2. Decision Tree Classifier
        3. Random Forest Classifier
    3. KNN (K-Nearest Neighbor) Classifier
    4. Support Vector Machine Classifier

03. Estimation Models

1. Ordinary-Least-Squares Regression (Similar to Linear Regression but not Linear Regression)

04. Regularization Models

1. Ridge Regularization
2. Lasso Regularization
3. ElasticNet Regularization

02. Unsupervised Learning

- (Data Without Labels)

1. Clustering (Document Analysis, Fake News Identification)
    1. K-Means Clustering
    2. Hierarchical Clustering
2. Association (Market Basket Analysis)
    1. APRIORI
3. Dimensionality Reduction (DNA & text analysis)
    1. Feature Extraction
        1. Principal Component Analysis
    2. Feature Selection
        1. Wrapper
        2. Filter
        3. Embedded Method

03. Reinforcement Learning

- (State and Action)

1. Model-Free
    1. Q-Learning
2. Model-Based
    1. Learn the Model
    2. Given the Model

06. Statistics (Advanced Statistics) :

- Q-Q Plot
- Chebyshev's Inequality
- Discrete Data Distributions
- Continuous Data Distributions
- Bernouli Distribution & Binomial Distribution
- Log Normal Distribution
- Power Law Distribution
- Box Cox Dististribution
- Poisson Distribution
- Applications Of Non-Gaussian Distributions
- Z-Test, t-Test
- Chi-Square Test

- ANOVA Test

05. Exploratory Data Analysis (EDA) :

1. Data CLeaning
2. Data Preparation
3. Feature ENcoding
4. Feature Scaling
5. Univariate Analysis
6. Bivariate Analysis
7. Correlation and Hypothesis Testing

01. Data Cleaning

-

02. Data Preparation

-

03. Feature Encoding

-

04. Feature Scaling

-

05. Univariate Analysis

-

06. Bi-Variate Analysis

-

07. Correlation and Hypothesis Testing

-

06. Mathematics (Other) :

01. Linear Algebra :

- Vectors & Matrics
- Transpose of Matrix, Inverse of Matrix, Determinent of Matrix, Trace of Matrix
- Dot Product
- Eigen Values, Eigen Vectors
- Single Value Decomposition

02. Differential Calculus :

- Chain Rule Of Derivatives
- Partial Derivatives
- Integrations
- Beta & Gamma Functions
- Functions Of Multiple Variable, Functions Of Limit, Functions Of Continuity, Functions Of Partial Derivatives
- Variants of Optimizers
- Loss Functions
- Back Propagation

- Minima & Maxima

07. Data Viz :

- Business Intelligence
    - Data Analytics Life Cycle
    - Analytics Processes & Tools
    - Analysis vs. Reporting
    - Visualization Techniques
- Advanced Excel
    - Functions, Formula, Charts, Pivots, Lookups
    - Data Analysis Tool Pack
        - Descriptive Summaries, Correlation, Regression
- Visualization Libraries
    - MatplotLib, Seaborn
- Modern Analytics Tools
    - PowerBI
        - Power BI project HR Analytics (HR Attrition Analysis!) by Rishabh Mishra
        - Power Bi Project - Sales dashboard (Super Store Sales Dashboard) by Rishabh Mishra
        - Power BI tutorial + project (Madhav Ecommerce Sales Dashboard) by RIshabh Mishra
        - Diwali Sales
    - Tableau
- Advanced Visualization

08. Machine Learning :

01. Machine Learning (Types):

1. Supervised Learning (Data with Labels)
    1. Regression Models (For predicting Continuous values)
        1. Linear Regression Model
            1. Simple Linear Regression Model
            2. Multiple Linear Regression Model
            3. Categorical Regression Model
        2. Non-Linear Regression Model
        3. Logistic Regression Model (*is a Regression as well as Classification model*)
            1. Binary Logistic Regression Model
            2. Nominal Logistic Regression Model
            3. Ordinal Logistic Regression Model
        4. Counts Regression Model
            1. Poisson Regression Model
            2. Negative Binomial Regression Model
    2. Classification Models (for classification of categorical data)
        1. Logistic Regression Model (*is a Regression as well as Classification model*)
            1. Binary Logistic Regression Model
            2. Nominal Logistic Regression Model
            3. Ordinal Logistic Regression Model
        2. Sequential / Probabilistic Classifier Algorithms
            1. Gaussian Naive Bayes Classifier
            2. Decision Tree Classifier
            3. Random Forest Classifier
        3. KNN (K-Nearest Neighbor) Classifier
        4. Support Vector Machine Classifier
    3. Ensemble Techniques / Gradient Boosting

      1. Bagging
      2. Boosting
         1. XG Boost
         2. CAT Boost
      3. Complete-it
   4. Estimation Models
      1. Ordinary-Least-Squares Regression (Similar to Linear Regression but not Linear Regression)
   5. Regularization Models
      1. Ridge Regularization
      2. Lasso Regularization
      3. ElasticNet Regularization
2. Unsupervised Learning(Data Without Labels)
   1. Clustering (Document Analysis, Fake News Identification)
      1. K-Means Clustering
      2. Hierarchical Clustering
      3. K-Median Clustering
      4. Expectation Maximization
   2. Association (Market Basket Analysis)
      1. APRIORI
      2. Eclat
      3. FP-Growth
   3. Dimensionality Reduction (DNA & text analysis)
      1. Feature Extraction
         1. Principal Component Analysis
      2. Feature Selection
         1. Wrapper
         2. Filter
         3. Embedded Method
         4. Stepwise Regression
3. Semi-Supervised Learning
4. Reinforcement Learning (State and Action)
   1. Model-Free
      1. Q-Learning
      2. Hybrid
      3. Policy Optimization
   2. Model-Based
      1. Learn the Model
      2. Given the Model

02. Machine Learning (Algos):

- Predictive ALgos
- Clustering Algos
- //etc.// algorithms

09. Deep Learning :

01. Deep Learning (Algos):

- Basics of Neural Networks
- ANN
- RNN
- CNN
- Types Of Deep Learning Algorithms
  - Discriminative DL Algorithms
    - CV

    - Generative DL ALgorithms
        - NLP
- Note: [1 week for theory & Practical]

02. Deep Learning (Applications):

- CV
- NLP
- Note : [1 week for theory & Practical]

10. Libraries & Packages

01. Python Libraries :

- MatplotLibs, Seaborn
- Tkinter, Pillow
- scikit-learn
- TensorFlow & Keras
- PyTorch
- nltk
- OpenCV

02. R Packages :

- Tidyverse
- Dplyr
- Tidyr
- GGPlot

11. Platforms :

- Git
- Kaggle
- Medium
- Analytics Vidhya
- Medium

12. Cloud :

- Getting into cloud certifications
    - Self Taught framework to get into cloud by Gwyneth Peña-Siguenza YT channel
        - 2 certifications, 2 projects, 1 year entry level experience
        - Must Learn : CLI, Programming Language, Cloud Platform, and DevOps
            1. Associate level Cloud Certification
            2. Project that implements associate certificate knowledge
            3. Complementary Certification (Linux, Networking, Security, DevOps, etc.)
            4. Project that implements complementary certificate knowledge and/or builds on previous project
        - Cloud Exposure (Stepping Stone)

          ```
          Stop Caring about role titles, care about descriptions
          ```

        - Sales Based

          ```
          Junior Solutions Architect | Cloud Sales Representative
          ```

        - Content Based

```
Cloud Technical Writer | Cloud Training Specialist
```

- Support Based

```
Cloud Migration Specialist | Cloud Customer Success Manager
Cloud Support Specialist
```

- Others

```
Junior Cloud Developer | Junior DevOps Engineer
Junior Site Reliability Engineer (SRE)
Junior Data Engineer | Junior Cloud Administrator
Junior Cloud Security Analyst | Junior Cloud Solutions Architect
```

- Cloud Job (Cloud in role titles)

```
Cloud Engineer | Cloud Developer | DevOps Engineer
```

01. Cloud (AWS) :

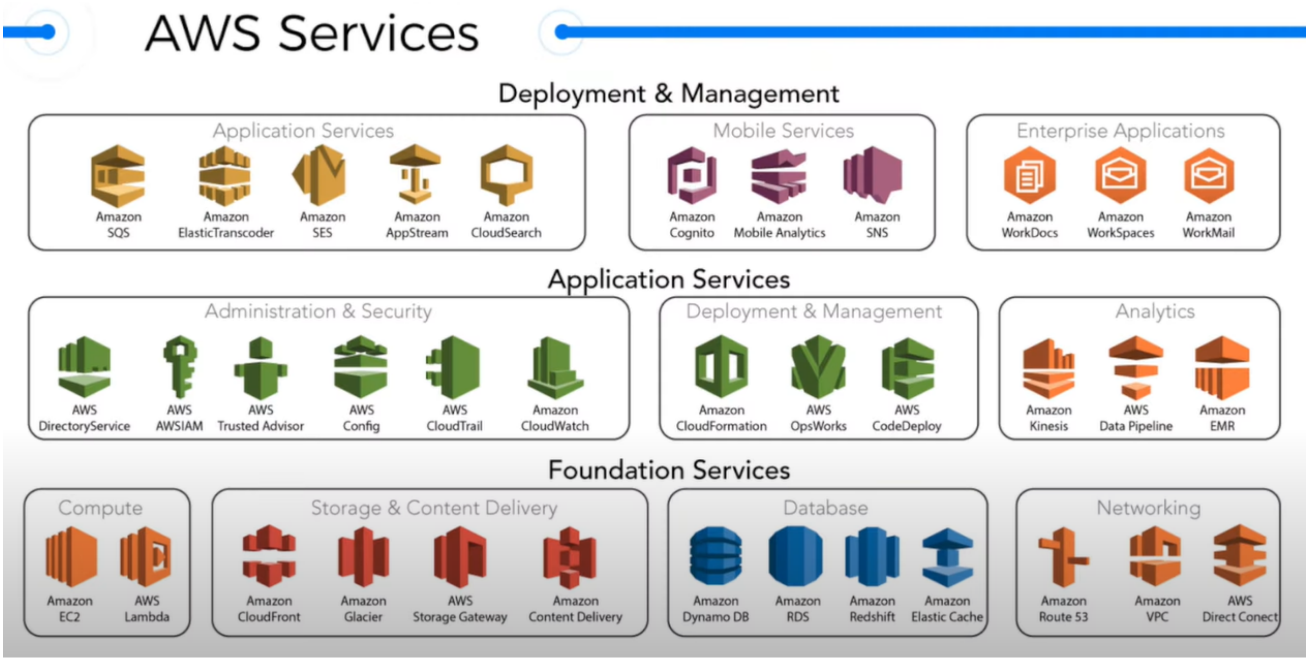
- `Amazon S3` : SImple Storage Service, Object Storage in a bucket in a hierarchical manner
- `AWS Glue` : serverless ETL using Python, PySpark, Spark scripts on it
- `Amazon EMR` : Elastic MapReduce, has managed servers/clusters, for batch processing
- `Amazon Redshift` : Datawarehouse service for querying data using SQL only after loading data into datawarehouse from S3, for batch processing
- `Amazon Athena` : for querying data using SQL (from S3) without loading data into datawarehouse , data is not modified in S3, for batch processing
- `AWS Lambda` : for transformation, extract data from API or basic operation like automation, serverless compute service, focus on writing scripts in multiple languages supported, not on managing clusters, clusters managed by AWS, can set triggers for automation for batch processing
- `AWS Kinesis` : for real-time data processing,
- `AWS DMS` : Data Migration Service, to migrate data from one source (like on-premise) datacentre to target (like cloud, oracle, etc.)
- `AWS Lake Formation` : to make data lakes
- `Amazon Managed Workflows` : for Apache airflow
- `Amazon QuickSight` : BI tool
- `Amazon RDS (Relational DataBase Service)` : Relational Database service offers PostgreSQL, MySQL, MariaDB, Amazon Aurora, etc.
- Fundamental Services
  - `Amazon VPC` : Virtual Private Cloud used for networking & Security
  - `Amazon EC2` : Elastic Compute Cloud used as an online computer
  - `Amazon SNS` : Simple Notification Service for notification purposes
  - `Amazon Cloudwatch` : to understand logs
  - `AWS IAM` : Identity Access Maangement to manage users and role
- Kafka
- SnowFlake
- Terraform

AWS Services

- Deployment & Management
  - Application Services
    - Amazon SQS (Simple Queue Service)
    - Amazon ElasticTranscoder
    - Amazon SES (Simple Email Service)
    - Amazon AppStream
    - Amazon CloudSearch
  - Mobile Services
    - Amazon Cognito
    - Amazon Mobile Analytics
    - Amazon SNS (Simple Notification Service)

  - Enterprise Services
    - Amazon WorkDocs
    - Amazon WorkSpace
    - Amazon WorkMail
- Application Services
  - Administration & Security
    - AWS DirectoryService
    - AWS AWSIAM
    - AWS Trusted Advisor
    - AWS Config
    - AWS CloudTrail
    - AWS CloudWatch
  - Deployment & Management
    - Amazon CloudFormation
    - Amazon OpsWorks
    - Amazon CodeDeploy
  - Analytics
    - Amazon Kinesis
    - AWS Data Pipeline
    - AWS EMR
- Foundation Services
  - Compute
    - Amazon EC2 (Elastic Compute Cloud)
    - Amazon Lambda
  - Storage & Content Delivery
    - Amazon S3 (Simple Storage Service)
    - Amazon CloudFront
    - Amazon Glacier
    - AWS Storage Gateway
    - AWS Content Delivery
  - Database
    - Amazon Dynamo DB
    - Amazon RDS (Relational Database Service)
    - Amazon Redshift
    - Amazon Elastic Cache
  - Networking
    - Amazon Route S3
    - Amazon VPC (Virtual Private Cloud)
    - Amazon Direct Conect

- Figure: AWS Services



-

AWS Exams

- AWS certified Developer Associate -
- AWS certified solution architect -

02. Cloud (MS-Azure) :

- Azure DataBricks
- Azure DataLake
- Azure DataFactory

MS-Azure Services

- Azure Storage
  - Types of azure storage
    - File, BLOB, Queue, Table
- Azure Data Lake Storage Gen2 - for building data lakes
- Azure Data Factory - to process data and pipelines
- Azure Synapse Analytics - for data analysis and building pipelines
- Azure DataBricks - for managed spark environements and to write transformation jobs
- Stream Analytics - to process real-time data,
- HDInsight - Open-source version of Spark, Hive & HBase
- Azure Cosmos DB - NoSQL solution by Azure

MS-Azure Exams

- DP-203
  - Certification Exam : Exam DP-203: Data Engineering on Microsoft Azure
  - Association Certification : Microsoft Certified Azure Data Engineer Associate
  - Rs. 4800 approx
  - prerequisite : Python, SQL, data engineering, Azure Services and is components
  - https://learn.microsoft.com/en-us/training/
  - Three sections
    1. Design and Implement data storage (15-20%)
    2. Develop Data Processing (40-45%)
    3. Secure, Monitor, and optimize data storage and data processing (30-35%)
  - Resources

        - Cloudacademy : DP-203 Exam Preparation: Data Engineering on Microsoft Azure https://cloudacademy.com/learning-paths/dp-203-exam-preparation-data-engineering-on-microsoft-azure-3191/
        - Coursera : Microsoft Azure Data Engineering Associate (DP-203) Professional Certificate https://www.coursera.org/professional-certificates/microsoft-azure-dp-203-data-engineering
        - Azure Data Engineer Associate Certification Guide - Newton Alex
        - For Practice Exam
            - Exam Topics
            - Search exam dumps on google

- DP-900
    - Certification Exam : Exam DP-900: Microsoft Azure Data Fundamentals
    - Association Certification : Microsoft Certified Azure Data Fundamentals
- PL-300
    - Certification Exam : Exam PL-300: Microsoft Power BI Data Analyst
    - Association Certification : Microsoft Certified Power BI Data Analyst Associate
- AZ-104
    - Certification Exam : Exam AZ-104: Microsoft Azure Administrator
    - Association Certification :
- AZ-204
    - Certification Exam : Exam AZ-204: Developing Solutions for Microsoft Azure
    - Association Certification :

03. Cloud (GCP) :

-

04. ETL Tools :

- Xplenty
- Stitch
- Alooma
- Talend
- Informatica
- Oracle
- Integrate.io
- AWS

13. Projects :

- End to End ML & DL projects, with deployment

14. Update Online Presence

- Linkedin
- GitHub
- HackerRank
- HackerEarth
- LeetCode
- Naukri
- Glassdoor

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.