
DATA ENGINEERING PROGRAM

PYTHON

Python. Conditionals and Loops. Functions and Parameters. Classes and Inheritance. Modules and Packages. PIP. Pandas. Matplotlib. NumPy. Seaborn. GitHub. Scikit-Learn, SpaCy, Pytorch/Keras/Tensorflow, Spark NLP.

DATA MANAGEMENT & SQL

Structured Query Language (SQL) and NoSQL. Joins and Set Operators. Subqueries. Duplicate Records and Cleansing Data. Keys, Normalization, Tables, and Views. OLTP/OLAP, Data Warehousing Concepts, Star Schema/Snowflake Schema. DML and DDL. PostgreSQL/MS SQL Server/MySQL. Cloud Database with Amazon RDS, DynamoDB.

HADOOP & SPARK

Big Data. ETL Concepts. HDFS Architecture. Single Node and Clustering. MapReduce. YARN. HBase. Hive. HiveQL. Pig. Grunt Shell and Pig Latin. Sqoop. Oozie/Airflow. Avro. Spark. Scala.

AWS CLOUD

Cloud Computing Concepts. Amazon Web Services. EC2. S3. EMR. Glacier. DynamoDB. RedShift and Data Warehousing. Data Visualization with Quicksight. AWS Glue and Athena. AWS Kinesis and Firehose. Elasticsearch Cluster. Intro to CloudFormation, IAM, CloudWatch, Cognito, and Lambda. Intro to Databricks, Kafka, ZooKeeper, Docker, and Kubernetes.

CERTIFICATIONS

AWS Cloud Practitioner (Required)

AWS Certified Big Data – Specialty (Recommended)

PROJECTS AND EVALUATIONS

Weekly Technical Exam

The information transmitted by this document is intended only for the person or entity to which it is addressed. This document may contain proprietary, business-confidential and/or privileged material. If you are not the intended recipient of this document, be aware that any use, review, retransmission, distribution, reproduction or any action taken in reliance upon this document is strictly prohibited. If you received this in error, please contact the sender and delete the material from all computers. Thank you for your cooperation.