



# Cloud Data Engineering

100% Placement Guarantee

Supporting Job seekers in career Change and skill enhancement program



# Why Cloud Data Engineering?



**Data engineering is an fascinating IT career path with tremendous opportunity in the market.**

**Being a data engineer we will be building pipelines for business and explore many interesting scenarios.**

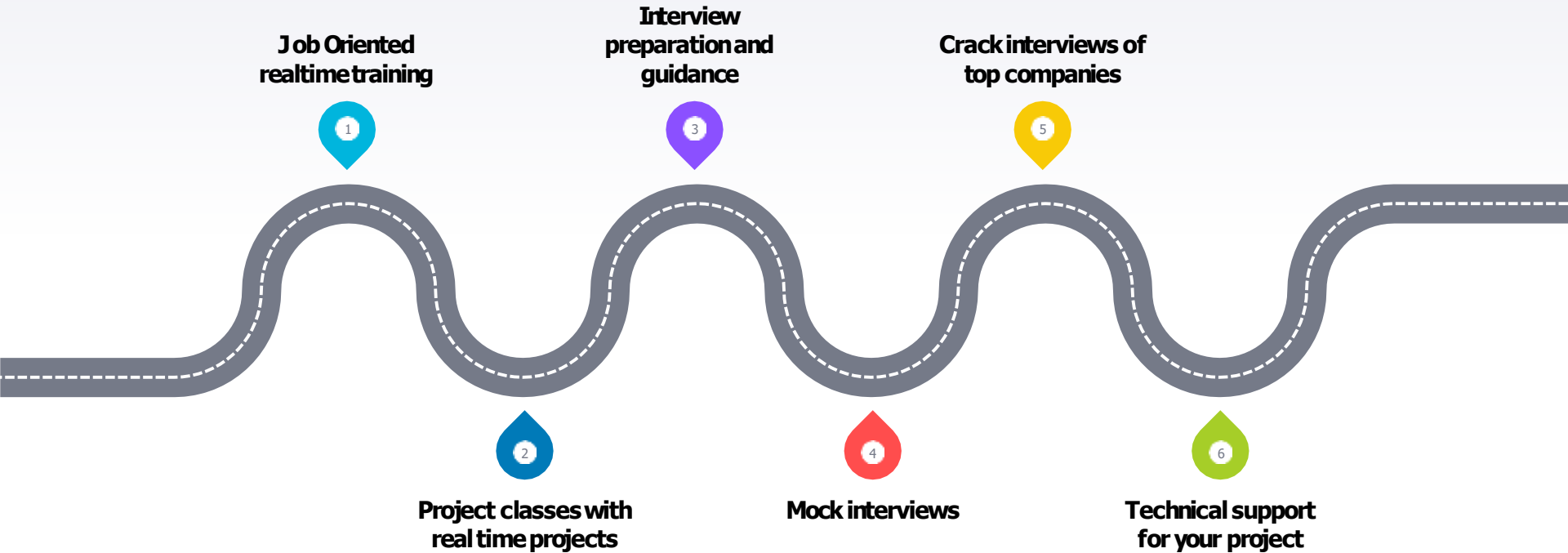
**Day by day data grows exponentially and they are going to feed us business ideas and helps us get management decisions**

**Cloud technologies are slowly ruling most of the IT tech stacks because of their huge advantage and reliability.**

**Here we are going to learn the top data engineering skills with the cloud tools using AWS.**

**So we are going to be part of team with leading and demanding IT skills in market.**

# Placement process



## MODE OF TRAINING

- Online Instructor-Led Live Training
- Classroom Training
- Online Self-paced Training

## Cloud data engineer certified -

Become aws cloud certified Professional Get Certified By

- AWS data analytics speciality.
- CCA SPARK AND HADOOP DEVELOPER
- Databricks Certified Associate Developer for Apache Spark 3.0
- Databricks Certified Professional Data Scientist
- HDP Certified Developer (HDPCD)
- CCA Administrator
- CCA HDP Administrator



# Prerequisites

There are no prerequisites for Cloud data engineering. If you know what is computer then that would be sufficient. Let's build ourselves from fundamentals. Only prerequisite preferred would be the dedication and consistency, we will anyway build that once we dive into the course. Preferably having basic knowledge in Linux/ SQL will put you one step ahead but never mind all topics will be covered from scratch.

## Syllabus

### Topic 1: Overview of data engineering and cloud Tech

- ❖ Overview of data engineering
- ❖ Business urge for data manipulation.
- ❖ Data pipelines
- ❖ ETL
- ❖ Data stores

### Topic 2: Linux shell scripting

- ❖ History
- ❖ Architecture
- ❖ Ubuntu setup
- ❖ Command line interface
- ❖ Development Commands
- ❖ File Management
- ❖ Directories Management
- ❖ Admin Commands

# Syllabus

## Topic 2: Linux shell scripting -contd

- ❖ Analogy with GUI
- ❖ Development tools

## Topic 3: Python Basics

- ❖ History of programming languages.
- ❖ Overview of python
- ❖ Python installation and setup
- ❖ Development kits - installation and setup
- ❖ Pycharm setup
- ❖ Jupyter notebook setup
- ❖ PEP -8
- ❖ Basic commands
- ❖ String interpolation

- ❖ Comments
- ❖ Data types
- ❖ Complex data type
- ❖ Conditional statement - if else
- ❖ Iterator - For
- ❖ Iterator - While
- ❖ Function
- ❖ Parameters
- ❖ Scope of variable
- ❖ Class
- ❖ Inheritance
- ❖ Package
- ❖ PIP
- ❖ Reading data from a file
- ❖ Data manipulation.
- ❖ Quiz
- ❖ Assessment

# Syllabus

## Topic 4: SQL basics

- ❖ History
- ❖ OLTP Vs OLAP
- ❖ Installation and setup
- ❖ Database Vs Datawarehouse
- ❖ Data warehousing concepts
- ❖ Data definition language
- ❖ Data manipulation language
- ❖ Data query language
- ❖ Data Control Language
- ❖ Data Control Language
- ❖ Transaction control language
- ❖ Data types
- ❖ Data filters
- ❖ Aggregations
- ❖ Joins

- ❖ Subqueries
- ❖ Views
- ❖ Union and intersection
- ❖ Window functions
- ❖ Transactions
- ❖ ACID properties
- ❖ String and date manipulation
- ❖ JDBC and ODBC connections
- ❖ SQL developer tools
- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 5: Distributed Computing

- ❖ History
- ❖ Architecture
- ❖ Mapreduce basics
- ❖ Evolution of technologies

# Syllabus

## Topic 5: Distributed Computing-contd

- ❖ Clusters
- ❖ Master slave architecture
- ❖ Fault tolerance
- ❖ Advantages
- ❖ Quiz

## Topic 6: Big Data overview

- ❖ History
- ❖ Architecture
- ❖ Distributed storage
- ❖ Distributed processing
- ❖ Distributed scheduling
- ❖ Overview of various BD tools
- ❖ Hadoop
- ❖ Hive

- ❖ Hdfs / hdfs commands
- ❖ Quiz
- ❖ Assessment

## Topic 7: Spark

- ❖ Introduction to spark
- ❖ Architecture
- ❖ Spark setup
- ❖ Advantages
- ❖ Spark
- ❖ File formats
- ❖ Transformation and action
- ❖ DAG
- ❖ RDD
- ❖ Dataframe
- ❖ Dataset



# Syllabus

## Topic 7: Spark - contd

- ❖ Spark context
- ❖ Spark submit
- ❖ Spark UI
- ❖ Deployment mode
- ❖ Shared variable (broadcast/ accumulators)
- ❖ Spark joins
- ❖ Spark SQL
- ❖ Spark Streaming
- ❖ partitioning/ Bucketing
- ❖ Coalesce/ répartition
- ❖ Narrow and wide transformation
- ❖ Catalyst optimiser
- ❖ Quiz
- ❖ Assessment

## Topic 8: Spark Internals

- ❖ Overview
- ❖ Jobs
- ❖ Stages
- ❖ Tasks
- ❖ Fault tolerance
- ❖ AQE
- ❖ Memory management
- ❖ Data Skewness
- ❖ Spark optimisation techniques.
- ❖ Quiz
- ❖ Assessment

## Topic 9: Spark api

- ❖ Introduction to Pyspark
- ❖ Read/ write using RDD
- ❖ Data transformation
- ❖ Spark complex data types
- ❖ Pyspark transformations and actions

# Syllabus

## Topic 9: PySpark api -contd

- ❖ Aggregations
- ❖ broadcast/ accumulator
- ❖ Monitoring using Spark UI
- ❖ Spark UDF
- ❖ Quiz
- ❖ Assessment

## Topic 10: Introduction to cloud

- ❖ History
- ❖ Capex vs opex
- ❖ Deployment models
- ❖ Various cloud suppliers
- ❖ Introduction to AWS
- ❖ Types of service
- ❖ Region and Infrastructure

- ❖ Region and Infrastructure
- ❖ Creating first AWS account
- ❖ Quiz

## Topic 11: IAM

- ❖ Introduction
- ❖ Setting up users
- ❖ Roles
- ❖ policies
- ❖ Access and permissions
- ❖ Quiz

## Topic 12: CLI and SDK

- ❖ Introduction to CLI / SDK
- ❖ Access setup for CLI
- ❖ CLI installation and setup
- ❖ AWS CLI commands
- ❖ Quiz

# Syllabus

## Topic 13: Elastic Cloud Compute(EC2)

- ❖ Iaas - Overview
- ❖ Types of instances
- ❖ Creating EC2 instance
- ❖ Opening EC2 via SSH
- ❖ Security groups
- ❖ IAM permissions
- ❖ PEM/ PPK file security
- ❖ VPC
- ❖ Elastic block storage
- ❖ Elastic file system
- ❖ Shutting down instance
- ❖ Quiz
- ❖ Assessment

## Topic 14: Simple Storage Service(S3)

- ❖ Introduction
- ❖ File system
- ❖ Types of storage class
- ❖ Upload/ download from S3
- ❖ Versioning
- ❖ S3 cli tweek
- ❖ Quiz
- ❖ Assessment

## Topic 15: Elastic Mapreduce (EMR)

- ❖ Introduction
- ❖ EMR storage types
- ❖ EMR Configs
- ❖ Frameworks

# Syllabus

## Topic 15: Elastic Mapreduce (EMR) - Contd

- ❖ Types of executors
- ❖ Developing pyspark application
- ❖ Application steps
- ❖ SSH into master nodes
- ❖ Types of execution methods
- ❖ Quiz
- ❖ Assessment

## Topic 16: Athena

- ❖ Introduction
- ❖ Workgroup
- ❖ Security
- ❖ ACID properties

- ❖ Working session to create table
- ❖ Athena table sources and targets
- ❖ Table configs
- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 17: Glue

- ❖ Introduction
- ❖ Glue crawler
- ❖ Glue data catalog
- ❖ Partitioning
- ❖ Glue ETL
- ❖ Dynamicframes
- ❖ Transformations
- ❖ dev endpoints and tools

# Syllabus

## Topic 17: Glue- Contd

- ❖ Glue Studio
- ❖ Glue jobs
- ❖ Glue databrew
- ❖ Source and targets
- ❖ Glue job configs
- ❖ Glue Scripts
- ❖ Quiz
- ❖ Assessment
- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 18: Redshift

- ❖ Database Vs Datawarehouse Vs Datalake

- ❖ Introduction to redshift
- ❖ Architecture
- ❖ Node types
- ❖ Redshift spectrum
- ❖ Scaling strategies
- ❖ Data distribution types
- ❖ Sort keys
- ❖ Importing and exporting data
- ❖ Concurrency
- ❖ Work load management (WLM)
- ❖ Short Query acceleration (SQA)
- ❖ Cluster resizing
- ❖ VACCU command
- ❖ Advanced query accelerator (AQUA)
- ❖ Redshift antipatterns
- ❖ Redshift serverless
- ❖ Redshift Query editor
- ❖ Working session
- ❖ Quiz

# Syllabus

## Topic 19: Relational Datastores (RDS)

- ❖ Introduction
- ❖ Types of Datastore flavours
- ❖ Aws Aurora
- ❖ Read replicas
- ❖ OLTP Vs OLAP
- ❖ RDS - Postgress
- ❖ RDS Configurations
- ❖ Working session
- ❖ Quiz

## Topic 20: DynamoDb (Nosql)

- ❖ Introduction
- ❖ Nosql stacks
- ❖ Dynamodb table creating configs

- ❖ Partition and sort keys
- ❖ Read/Write Capacity Modes
- ❖ Read consistency
- ❖ DynamoDb Throttling
- ❖ Reading and writing data
- ❖ Reading data - Query
- ❖ Reading data - Scan
- ❖ Index types (LSI and GSI)
- ❖ DynamoDb PartiQL
- ❖ Optimistic Locking
- ❖ DynamoDb Accelerator (DAX)
- ❖ DynamoDB Streams
- ❖ DynamoDB -Time To Live (TTL)
- ❖ DynamoDB Operations
- ❖ Security & Other Features
- ❖ Fine-Grained Access Control
- ❖ Quiz
- ❖ Assessment

# Syllabus

## Topic 21: Kinesis

- ❖ kinesis overview
- ❖ Kinesis Data Streams
- ❖ Capacity Modes
- ❖ Kinesis Data Streams Security
- ❖ Kinesis Producers
- ❖ Kinesis Data Streams Consumers
- ❖ ProvisionedThroughputExceeded
- ❖ Consumers Types
- ❖ Kinesis Client Library (KCL)
- ❖ Kinesis Operation – Shard Splitting
- ❖ Kinesis Operation – Merging Shards
- ❖ Kinesis Data Firehose
- ❖ Kinesis Data Streams vs Firehose
- ❖ Kinesis Data Analytics (SQL application)

## Topic 22: Simple queue service (SQS)

- ❖ Queue -Overview
- ❖ Queue Types
- ❖ Standard Queue
- ❖ SQS – Producing Messages
- ❖ SQS – Consuming Messages
- ❖ SQS with Auto Scaling Group (ASG)
- ❖ SQS to decouple between application tiers
- ❖ Amazon SQS - Security
- ❖ SQS Queue Access Policy
- ❖ Message Visibility Timeout
- ❖ Dead Letter Queue
- ❖ DLQ – Redrive to Source
- ❖ Delay Queue
- ❖ SQS - Long Polling
- ❖ FIFO Queue
- ❖ SQS FIFO – Deduplication

# Syllabus

## Topic 22: Simple queue service (SQS) - contd

- ❖ SQS FIFO – Message Grouping
- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 23: Simple notification service (SNS)

- ❖ SNS - Overview
- ❖ Integrations
- ❖ topics publishments
- ❖ SNS Security
- ❖ SNS & SQS Fan out
- ❖ Multiplexing

- ❖ Application use cases
- ❖ SQS vs SNS vs Kinesis
- ❖ working session
- ❖ Quiz

## Topic 24: Cloud watch

- ❖ Monitoring in AWS
- ❖ CloudWatch Metrics
- ❖ Detailed monitoring
- ❖ Custom Metrics
- ❖ CloudWatch Logs
- ❖ CloudWatch Logs - Sources
- ❖ CloudWatch Logs Metric Filter & Insights
- ❖ Logs – S3 Export
- ❖ Logs Aggregation Multi-Account & Multi Region
- ❖ CloudWatch Logs for EC2



# Syllabus

## Topic 24: Cloud watch - contd

- ❖ Logs Agent & Unified Agent
- ❖ CloudWatch Alarms
- ❖ Alarm Targets
- ❖ EC2 Instance Recovery
- ❖ CloudWatch Events
- ❖ AWS X-Ray
- ❖ AWS CloudTrail
- ❖ Quiz

## Topic 25: Lambda

- ❖ Serverless
- ❖ Feature of AWS Lambda
- ❖ Integrations
- ❖ CRON Jobs

- ❖ Pricing
- ❖ Synchronous Invocations
- ❖ Asynchronous Invocations
- ❖ S3 Events Notifications
- ❖ Event Source Mapping
- ❖ Queues & Lambda
- ❖ Lambda – Destinations
- ❖ Execution Role (IAM Role)
- ❖ Environment Variables
- ❖ Logging & Monitoring
- ❖ Best practices
- ❖ Concurrency and Throttling
- ❖ Concurrency and Asynchronous Invocations
- ❖ Cold Starts & Provisioned Concurrency
- ❖ Lambda Container Images
- ❖ Lambda Versions
- ❖ Lambda Aliases
- ❖ Limitation

# Syllabus

## Topic 25: Lambda - contd

- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 26: Managed service Kafka (MSK)

- ❖ Kafka - Overview
- ❖ Brokers , topics and shards
- ❖ Configurations
- ❖ Security
- ❖ Monitoring
- ❖ Connect
- ❖ MSK Serverless
- ❖ Kinesis Data Streams vs Amazon MSK
- ❖ Quiz

## Topic 27: Step functions and data pipelines

- ❖ Introduction to workflows
- ❖ Orchestration overview
- ❖ Step function features
- ❖ Data pipeline examples
- ❖ Data pipeline Features
- ❖ Data Pipeline Activities
- ❖ Quiz

## Topic 28: Airflow

- ❖ Introduction
- ❖ Airflow setup guide
- ❖ Components of airflow
- ❖ Operators and types
- ❖ DAG
- ❖ Schema

# Syllabus

## Topic 28: Airflow - contd

- ❖ Dependencies
- ❖ Airflow web UI
- ❖ Python Script for orchestration definition
- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 29: Databricks

- ❖ Introduction
- ❖ Databricks setup - Community edition
- ❖ Notebooks
- ❖ Clusters
- ❖ Dbfs and other storage integrations

- ❖ Cluster pool
- ❖ Magic function (python, shell)
- ❖ Integration with other cloud
- ❖ Users
- ❖ Repo
- ❖ Workflows
- ❖ Jobs and scheduling
- ❖ Working session
- ❖ Quiz
- ❖ Assessment

## Topic 30: Spark performance optimisation

- ❖ Common issues can expect while running spark jobs
- ❖ Best practices
- ❖ Out of memory Exception

# Syllabus

## Topic 31: Spark performance optimisation -contd

- ❖ Spark UI debugging
- ❖ AQE Utilisation
- ❖ Realtime scenarios and organisation expectations.
- ❖ Discussion on various flavours of tech stacks

## Topic 32: Interview Questions

## Topic 33: Mock Interview and Mock Interview Answers

## Topic 34: Resume Prep session and sample resumes



# We're here to help

- ❖ Customized support addressing your needs
- ❖ Personalized approach consulting, planning and support
- ❖ Personalized career counselling
- ❖ Goal setting and establishing a career roadmap
- ❖ Focused on Emerging Technologies
- ❖ Technical Skills and personality screening
- ❖ In-depth review of a jobseeker's existing resume and creation of a customized new powerful resume.
- ❖ Establish an online presence in your job search with the latest channels.
- ❖ Networking, salary and negotiation strategies.
- ❖ Interviewing techniques and support
- ❖ Domestic and Global Placements
- ❖ Interview and Job Support
- ❖ Technical support after placement
- ❖ On-site assistance
- ❖ Re-deployment support

# Why Trendnologies?

- ❖ 10+ years of delivering exceptional service to candidates in Training and placements.
- ❖ Fast Track Placement in 30 days.
- ❖ 100% Placement assurance.
- ❖ Managed by IT Experts as Trainers
- ❖ Customized support addressing your needs.
- ❖ Personalized approach / dedicated attention.
- ❖ Personalized career counselling.
- ❖ Personalized consulting, planning and support.
- ❖ Job oriented theoretical & Practical Classes.
- ❖ More focus on hands-on and use cases.
- ❖ Focused on case studies and Real-time projects.
- ❖ Complete end to end training soft material.
- ❖ Guidance for Resume preparation and interview preparation.
- ❖ Bigdata training along with Career guidance and Certification guidance.
- ❖ Comprehensive career solutions across India, USA, UK and UAE.

# ▶ Key Features

- Real-Time Projects on Bigdata
- Training on all major concepts and make upskill
- Managed and mastered by highly skilled Industry Experts



- 200+ Hours Course Duration
- 100% Placement guarantee
- Hands-on with 30+ Case Studies
- Fast track placement mode



- Complete Career Guidance
- Placement in Top MNC Company
- Online Professional Certification: Cloudera / Hortonworks / Databricks / Cloud

# BRANCHES - CHENNAI

## **Annanagar Branch I**

F22, 2C & 2D, Second floor,  
Second Main Road,  
Annanagar East,  
Chennai - 600102  
86819 62962

## **Medavakkam Branch**

#29, M.S Complex,  
Second Floor,  
Anna Street,  
Medavakkam.  
Chennai – 600100  
84288 88962

## **Annanagar Branch II**

New no. 83, Old no. M-9-B, M Block  
Annanagar East,  
Near K4 Roundtana,  
Chennai – 600102  
86809 62962

## **Placement Division**

F22, 3D third floor,  
Second Main Road,  
Annanagar East,  
Chennai – 600102.  
91763 62962



# BRANCHES - COIMBATORE

## **Coimbatore Branch I**


#778, Mettupalyam Road,  
R S Puram  
Coimbatore – 641002  
93600 93938

## **Coimbatore Placement Division**

KB complex 105 ,  
Lokamanya Street East,  
R S Puram ,  
Coimbatore – 641002  
8428444962

## BANGALORE Branch

#421, 4th floor South  
block Manipal centre,  
MG Road  
Bengaluru – 560042  
84286 66962



Thanks Contact with AWS CLOUD  
DATA ENGINEER Experts @  
8270962962

Visit our website for more details  
[WWW.TRENDNOLOGIES.COM](http://WWW.TRENDNOLOGIES.COM)