□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Level 0 - Foundations

Mile 0: Syllabus for whole course

Mile 01: Introduction to Databricks

Mile 02: Core Components of Databricks

Mile 03: Architecture of Databricks

Mile 04: Medallion Architecture

Mile 05: Creating Your First Data Pipeline

Level 1 – Spark & Delta Lake Fundamentals

Core Apache Spark concepts, DataFrames, Spark SQL, Delta Lake, jobs, ETL pipelines, and beginner performance tuning.

Mile 06: Introduction to Notebooks

Mile 07: Using Databricks Utilities (dbutils)

Mile 08: Managing Secrets Securely

Mile 09: Exporting and Importing Notebooks

Mile 10: Integration with Git Repositories

Mile 11: Performance Tips for Beginners

Mile 12: Navigating the Databricks Workspace

Mile 13: Creating and Managing Clusters

Mile 14: Cluster Configuration Basics

Mile 15: Parallelism and Partitions

Mile 16: What is Apache Spark

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 17: Understanding Spark Architecture

Mile 18: Spark DataFrames Explained

Mile 19: Reading and Writing Data

Mile 20: Basic Transformations

Mile 21: Aggregations and Grouping

Mile 22: Joins in Spark

Mile 23: Caching Data for Speed

Mile 24: Introduction to Spark SQL

Mile 25: Monitoring Spark Jobs

Mile 26: Creating DataFrames from Files

Mile 27: Understanding Schema and Data Types

Mile 28: Handling Missing Data

Mile 29: Using Temporary Views

Mile 30: Writing Data to Tables

Mile 31: Querying Tables with SQL

Mile 32: Partitioning Data

Mile 33: Working with Parquet Files

Mile 34: Displaying and Visualizing Data

□ Mile 35: Data Exploration Best Practices

Mile 36: Introduction to Delta Lake

Mile 37: Writing and Reading Delta Tables

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Mile 38: Understanding ACID Transactions

Mile 39: Schema Enforcement

Mile 40: Schema Evolution

Mile 41: Time Travel in Delta Lake

Mile 42: Optimizing Delta Tables

Mile 43: Data Compaction and Cleanup

Mile 44: Managing Delta Table History

Mile 45: When to Use Delta Lake

□ □ Practice / Hands-on:

Mile 46: Building a Simple ETL Pipeline

Mile 47: Creating a Delta Table from CSV

Mile 48: Visualizing Data Results

Mile 49: Troubleshooting Errors

Mile 50: Final Review of Level 1

Level 2 – Data Engineering & Governance

Advanced Delta Lake operations, streaming ingestion, Unity Catalog, ETL frameworks, performance optimization, and end-to-end projects.

Mile 51: Review of Delta Lake Basics

Mile 52: Using the VACUUM Command

Mile 53: OPTIMIZE for Faster Queries

Mile 54: Z-Ordering Data

Mile 55: Understanding Delta Table History

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 56: Data Compaction Techniques

Mile 57: Data Versioning and Time Travel

Mile 58: Managing Table Metadata

Mile 59: Handling Large Tables Efficiently

Mile 60: Automating Table Maintenance

Mile 61: Introduction to Auto Loader

Mile 62: Configuring Cloud File Paths

Mile 63: Handling Schema Changes in Streams

Mile 64: Using Checkpoints in Streaming

Mile 65: Writing Stream Data to Delta

Mile 66: Managing Streaming Triggers

Mile 67: Streaming Aggregations

Mile 68: Monitoring Streaming Pipelines

Mile 69: Handling Late Data

Mile 70: Stopping and Restarting Streams Safely

Mile 71: Introduction to Unity Catalog

Mile 72: Catalogs, Schemas, and Tables

Mile 73: Assigning Permissions

Mile 74: Data Lineage in Unity Catalog

Mile 75: Managing Users and Groups

Mile 76: Setting Up External Locations

Mile 77: Secure Views and Masking

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 78: Unity Catalog Audit Logs

Mile 79: Migrating Data to Unity Catalog

□ Mile 80: Best Practices for Governance

Mile 81: Designing ETL Pipelines

Mile 82: Implementing Incremental Loads

Mile 83: Change Data Capture (CDC)

Mile 84: Handling Bad Records

Mile 85: Data Validation Techniques

Mile 86: Building Modular ETL Code

Mile 87: Automating ETL Workflows

Mile 88: Monitoring ETL Pipeline Health

Mile 89: Error Handling in Pipelines

Mile 90: Documenting ETL Processes

Mile 91: Using Caching and Broadcasting

Mile 92: Managing Cost Efficiency

Mile 93: Troubleshooting Performance Issues

Mile 94: Building an End-to-End ETL Project

Mile 95: Adding Data Quality Checks

□ □ Practice / Hands-on:

Mile 96: Scheduling with Databricks Jobs

Mile 97: Generating Reports from Gold Data

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 98: Introduction to Databricks Jobs

Mile 99: Creating Multi-Task Jobs

Mile 100: Level 2 Recap and Next Steps

Level 3 – Streaming, MLflow & Optimization

Structured Streaming, Lakeflow pipelines, MLflow, query optimization, security, governance, and real-time analytics.

Mile 101: Introduction to Structured Streaming

Mile 102: Comparing Batch and Streaming Data

Mile 103: Configuring Streaming Sources

Mile 104: Using Watermarking

Mile 105: Windowed Aggregations

Mile 106: Stateful Streaming Operations

Mile 107: Checkpoints and Fault Tolerance

Mile 108: Trigger Intervals and Latency

Mile 109: Monitoring Streaming Jobs

□ Mile 110: Best Practices for Stream Stability

□ Mile 111: Practice Session: Basic Structured Streaming Hands-On

Mile 112: Introduction to Lakeflow Spark Declarative Pipelines

Mile 113: Creating a Lakeflow SDP Pipeline

Mile 114: Defining Data Quality Expectations

Mile 115: Monitoring Lakeflow SDP Pipelines

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 116: Incremental Data Processing in Lakeflow SDP

Mile 117: Managing Pipeline Dependencies

Mile 118: Combining Batch and Streaming in Lakeflow SDP

Mile 119: Versioning and Rollbacks in Lakeflow SDP

Mile 120: Deploying Lakeflow SDP with Jobs

Mile 121: Real-World Use Cases for Lakeflow SDP

□ Mile 122: Practice Session: Lakeflow SDP Pipeline Lab

Mile 123: Introduction to MLflow

Mile 124: Tracking Experiments

Mile 125: Registering and Managing Models

Mile 126: Serving Models in Databricks

Mile 127: Integrating MLflow with Databricks AutoML

Mile 128: Feature Store Basics

Mile 129: Tracking Model Versions

Mile 130: Monitoring Model Performance

Mile 131: Collaborating with Teams in MLflow

Mile 132: Real-World MLflow Use Case

□ Mile 133: Practice Session: MLflow Lifecycle Hands-On

Mile 134: Introduction to Query Optimization

Mile 135: Adaptive Query Execution (AQE)

Mile 136: Partitioning and Bucketing Data

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 137: Using Photon Engine

Mile 138: Managing Skewed Data

Mile 139: Caching and Reusing Computations

Mile 140: Optimizing Joins

Mile 141: Using the Spark UI for Debugging

Mile 142: Cost Optimization Strategies

Mile 143: Measuring Success in Optimization

□ Mile 144: Practice Session: Tuning & Optimization Exercises

Mile 145: Introduction to Data Security

Mile 146: Managing Access Controls

Mile 147: Encryption and Key Management

Mile 148: Monitoring User Activity

Mile 149: Regulatory Compliance Standards

□ Mile 150: Practice Session: Production Security & Governance Lab

Mile 151: Building a Real-Time Analytics Pipeline

Mile 152: Integrating MLflow into Your Pipeline

Mile 153: Debugging and Monitoring Jobs

Mile 154: Reviewing Optimization Results

Mile 155: Level 3 Recap and Preparation for Level 4

□ □ Capstone Project:

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

□ Mile 156: Practice Session: Level 3 Capstone Mini-Project

Level 4 – Enterprise Architecture & Certification

Certification preparation, enterprise architecture, governance, IaC, CI/CD, security, and large-scale Databricks implementations.

Mile 157: Overview of Databricks Certifications

Mile 158: Data Engineer Professional Exam Topics

Mile 159: Machine Learning Professional Exam Topics

Mile 160: Reviewing Sample Questions

□ Mile 161: Hands-On Practice with Databricks Labs

Mile 162: Exam Preparation Tips

Mile 163: Mock Test and Self-Assessment

Mile 164: Reviewing Common Mistakes

Mile 165: Final Revision Checklist

Mile 166: Scheduling and Taking the Exam

□ Mile 167: Practice Session: Certification Prep Labs

Mile 168: Introduction to Lakehouse Architecture

Mile 169: Designing Data Layers

Mile 170: Cross-Region Data Replication

Mile 171: High Availability (HA) and Disaster Recovery (DR)

Mile 172: Data Mesh Principles

Mile 173: Integration with External Systems

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 174: Managing Workspaces Across Teams

Mile 175: Governance for Multi-Tenant Environments

Mile 176: Cost Management Across Projects

Mile 177: Enterprise Case Study

□ Mile 178: Practice Session: Lakehouse Architecture Design Lab

Mile 179: Introduction to Infrastructure as Code (IaC)

Mile 180: Terraform for Databricks

Mile 181: Databricks CLI Automation

Mile 182: Databricks Asset Bundles (DABs)

Mile 183: CI/CD Pipeline Setup

Mile 184: Environment Promotion (Dev → QA → Prod)

Mile 185: Automating DLT Pipeline Deployments

Mile 186: Scheduling Deployments and Updates

Mile 187: Managing Secrets in Automation

Mile 188: Testing and Validating Deployments

□ Mile 189: Practice Session: IaC & Automation Hands-On

Mile 190: Unity Catalog for Enterprise Governance

Mile 191: Role-Based Access Control (RBAC)

Mile 192: Data Lineage and Auditing

Mile 193: Implementing Data Quality SLAs

Mile 194: Compliance and Regulatory Requirements

□ DATABRICKS LEARNING ROADMAP (Levels 0–4)

Vaanii Technology LLC | Professional Training Program

Databricks Learning Roadmap (Levels 0–4)

Mile 195: Encryption and Network Security

Mile 196: Activity Monitoring and Alerts

Mile 197: Logging and Observability Tools

Mile 198: Governance Automation

□ Mile 199: Security Best Practices Summary

□ Mile 200: Practice Session: Enterprise Governance Workshop