

Big Data Service: Administration and Implementation

(customized FMIS)

Duration: 40 Hours (5 Days)

Overview

The Big Data Service: Administration and Implementation (customized FMIS) course is a hands-on enterprise training program for teams responsible for designing, administering, monitoring, securing, and operating Oracle Cloud Infrastructure Big Data Service environments for Financial Management Information System workloads.

The course focuses on practical administration and implementation capabilities across OCI Big Data Service clusters, Oracle Distribution including Apache Hadoop, Apache Spark, Hadoop ecosystem services, Object Storage integration, secure access, monitoring, performance management, update readiness, and operational handover. FMIS scenarios are used throughout the agenda to connect platform tasks with financial data ingestion, reconciliation, reporting, auditability, and production support requirements.

Audience Profile

This customized course is suitable for professionals involved in FMIS data platform implementation, big data operations, and enterprise analytics support.

- Big Data Service administrators and platform operations teams
- Data engineers supporting FMIS ingestion, transformation, and reporting pipelines
- Database administrators expanding into Hadoop, Spark, and cloud data lake operations
- Cloud infrastructure administrators responsible for OCI networking, IAM, and service operations
- Application support teams managing FMIS batch, reconciliation, and analytics workloads
- Security, governance, and compliance teams responsible for access control and audit readiness
- BI, reporting, and analytics teams consuming curated FMIS data
- Project managers and technical leads overseeing FMIS modernization or data platform rollout

Technologies & Tools Covered

- Oracle Cloud Infrastructure Big Data Service
- Oracle Distribution including Apache Hadoop (ODH)
- Apache Hadoop, HDFS, YARN, Hive, HBase, and Apache Spark concepts
- Apache Ambari for Hadoop service administration and monitoring
- Apache Hue and JupyterHub access patterns where enabled
- Apache Ranger for fine-grained data access policy management
- OCI Object Storage integration for data lake and FMIS landing-zone patterns
- OCI IAM, compartments, policies, VCNs, subnets, security rules, and load balancers
- OCI Monitoring, alarms, notifications, work requests, CLI, API, and Cloud Shell

- Oracle Cloud SQL integration concepts for SQL access across big data sources

Prerequisites & Preparation

Participants should have:

- Basic understanding of cloud infrastructure, networking, and identity concepts
- Familiarity with Linux command-line operations and SSH access
- Basic knowledge of databases, SQL, and batch data processing
- Awareness of FMIS data domains such as budget, expenditure, payment, vendor, ledger, and audit data
- Access to an OCI tenancy or training sandbox with permissions for Big Data Service labs

Recommended preparation:

- Confirm OCI service limits, quotas, and regional availability for planned lab resources
- Prepare SSH key pairs and approved access path through bastion, VPN, FastConnect, or private network
- Identify sample FMIS datasets or representative training data for ingestion and validation exercises

Day Wise Agenda

Day 1 - Big Data Service & OCI Foundations for FMIS

- Introduction to OCI Big Data Service and enterprise Hadoop/Spark use cases
- FMIS workload patterns: ingestion, reconciliation, reporting, audit, and analytics
- Big Data Service architecture: clusters, node roles, ODH stack, and supporting OCI services
- Tenancy preparation: compartments, IAM policies, quotas, service limits, SSH keys, and networking
- Cluster planning: HA versus non-HA, shapes, storage, capacity, naming, and tagging
- Hands-on: prepare OCI resources and create a training Big Data Service cluster

Day 2 - Cluster Administration, Access & Security

- Cluster lifecycle administration: details, work requests, node operations, and service health
- Apache Ambari for managing and monitoring Hadoop ecosystem services
- Secure access patterns using private endpoints, bastion/VPN/FastConnect, and load balancers
- Kerberos-enabled secure cluster concepts and Active Directory/LDAP integration options
- Apache Ranger policy model for HDFS, Hive, HBase, Spark, Kafka, and Trino services where deployed
- FMIS access-control design: role mapping, segregation of duties, privileged access, and audit evidence

Day 3 - Data Implementation & FMIS Pipeline Enablement

- Data lake layout for FMIS: landing, raw, curated, reconciliation, and reporting zones
- HDFS and OCI Object Storage integration for durable data storage and exchange
- Hive metadata, external tables, partitioning, file formats, and retention planning
- Spark-based processing patterns for validation, enrichment, aggregation, and reconciliation
- Cloud SQL and SQL access concepts for analytics and reporting consumers
- Hands-on: build an FMIS-style ingestion and transformation workflow with validation checkpoints

Day 4 - Monitoring, Performance & Operational Management

- OCI Monitoring for Big Data Service cluster and node health, capacity, alarms, and notifications
- Ambari dashboards, service checks, logs, JMX metrics, and operational troubleshooting
- YARN and Spark workload visibility, queues, resource usage, job failures, and retry analysis
- Performance tuning concepts for storage layout, partitions, file sizes, Spark jobs, and Hive queries
- Incident response for FMIS batch delays, failed reconciliations, capacity pressure, and access issues
- Hands-on: create monitoring views, investigate workload symptoms, and document runbook actions

Day 5 - HA, Updates, Recovery Approach & Implementation Handover

- High availability design patterns for Big Data Service and service access endpoints
- Operational backup and recovery approach for data, metadata, configurations, keys, and audit evidence
- BDS, ODH, and OS update readiness: maintenance windows, prerequisites, rollback planning, and validation
- Autoscale and capacity management concepts for predictable FMIS processing cycles
- Governance checklist: security baseline, operational controls, documentation, and support model
- Capstone: review the customized FMIS implementation plan and prepare the administration handover pack

Course Syllabus

Big Data Service Administration

Chapter 1. Introduction to Big Data Service for FMIS

- Role of Big Data Service in an enterprise data platform
- FMIS data domains, processing cycles, and analytics requirements
- BDS architecture and Oracle Distribution including Apache Hadoop overview
- Hadoop, HDFS, YARN, Hive, HBase, and Spark concepts for administrators
- Training lab environment, access model, and implementation assumptions

Chapter 2. OCI Foundations and Cluster Planning

- Compartments, IAM policies, dynamic groups, networking, and service limits
- VCN, subnet, route table, security list, NSG, and private access design considerations
- Cluster topology, node roles, HA profile, compute shapes, block storage, and Object Storage alignment
- Naming, tagging, environment separation, and governance standards for FMIS workloads
- Capacity planning for daily, monthly, quarterly, and year-end financial processing windows

Chapter 3. Provisioning and Initial Configuration

- Prepare tenancy resources, policies, keys, network paths, and approved admin access
- Create a Big Data Service cluster and monitor work requests
- Validate nodes, services, Ambari access, SSH connectivity, and cluster health
- Configure front-end access through bastion, VPN, FastConnect, or load balancer patterns
- Create a baseline inventory for operational support and change tracking

Chapter 4. Cluster Administration and Lifecycle Operations

- Review cluster details, node inventory, configuration, and linked OCI resources
- Add, restart, and manage nodes according to workload and maintenance requirements
- Use Ambari to monitor services, run service checks, and manage Hadoop components
- Manage Linux, Hadoop, and service-level user/group administration responsibilities
- Build routine administration runbooks for start-of-day, end-of-day, and month-end operations

Chapter 5. Security, Access Control and Governance

- Secure cluster concepts, Kerberos, HDFS encryption, and administrator access boundaries
- OCI IAM, compartment policies, API keys, and least-privilege administration
- Apache Ranger policy design, plugin awareness, access review, and audit monitoring
- LDAP/Active Directory integration options for authentication and group mapping
- Certificate and TLS considerations for secure service endpoints
- FMIS segregation of duties, evidence retention, and compliance control mapping

Big Data Service Implementation

Chapter 6. FMIS Data Architecture and Storage Design

- Landing, raw, curated, reconciliation, exception, and reporting-zone design

- HDFS and Object Storage roles in ingestion, retention, exchange, and downstream reporting
- File formats, compression, partitioning, lifecycle policies, and data-quality checkpoints
- Metadata conventions for FMIS datasets, batches, periods, source systems, and audit fields
- Designing for traceability from source extract to validated reporting output

Chapter 7. Data Ingestion and Integration Implementation

- Source onboarding approach for financial, budget, payment, procurement, vendor, and ledger data
- Secure file movement, Object Storage access, API key handling, and ingestion validation
- Hive table creation, schema evolution, partition registration, and data refresh patterns
- Batch control design: file receipt, load status, reconciliation counts, and exception capture
- Operational handoff between FMIS application teams, data engineers, and platform administrators

Chapter 8. Processing, Analytics and Reporting Enablement

- Spark and Hive processing patterns for cleansing, enrichment, aggregation, and reconciliation
- Workload isolation, queue planning, job execution standards, and rerun controls
- Cloud SQL integration concepts for SQL-based access across big data sources
- Analytics readiness for dashboards, regulatory reports, audit extracts, and management summaries
- Performance considerations for high-volume FMIS processing and month-end peaks

Chapter 9. Monitoring, Troubleshooting and Performance Management

- OCI cluster and node metrics, alarms, notifications, and resource-principal monitoring awareness
- Ambari health checks, logs, service restarts, JMX metrics, and service-level troubleshooting
- YARN and Spark diagnostics for failed, slow, or resource-constrained jobs
- Storage, partition, file-size, and query tuning practices for predictable FMIS workloads
- Incident response model for failed loads, delayed processing, access failures, and capacity alerts

Chapter 10. High Availability, Updates and Recovery Readiness

- HA cluster design, utility/master/worker node considerations, and service endpoint access
- Load balancer usage for Ambari, Hue, JupyterHub, and secure administrative entry points
- BDS release update awareness, OS patch prerequisites, maintenance windows, and validation steps
- Operational recovery planning for data, metadata, configurations, credentials, and documentation
- Capacity, autoscale, and DR considerations for scheduled FMIS processing windows

Chapter 11. Customized FMIS Implementation Handover

- Implementation checklist and operational acceptance criteria
- Access matrix, service inventory, environment diagram inputs, and ownership model
- Daily, weekly, monthly, and year-end runbook structure
- Monitoring dashboard checklist and alert routing recommendations
- Known-risk register, support escalation path, and continuous improvement backlog
- Final capstone review and administrator handover pack