

INFOSOFT IT SOLUTIONS

Training | Projects | Placements

Revathi Apartments, Ameerpet, 1st Floor, Opposite Annapurna Block,

Infosoft It solutions, Software Training & Development Institute, 9059683947|9182540872

Bigdata Greenplum DBA

Introduction to Greenplum DBA

- Overview of Greenplum Database: Features, architecture, and advantages
- Comparison with traditional RDBMS and other Big Data solutions

Greenplum Architecture Overview

- Greenplum Components: Master node, segment nodes, and interconnects
- MPP (Massively Parallel Processing) Architecture: Distribution and parallelism concepts

Installing and Configuring Greenplum

- Preparing for Installation: System requirements and prerequisites
- Greenplum Installation: Single-node and multi-node cluster setups
- Configuration Management: gpconfigs, gpsegininstall, and gpexpand commands

Greenplum Database Objects

- Tables and Views: Creating, altering, and dropping tables/views
- Indexes: Types of indexes and their usage
- Partitioning: Range, list, and hash partitioning strategies

Data Loading and Management

- Data Loading Techniques: Using gpload, COPY command, and external tables
- Managing Data Distribution: Distribution keys and distribution policies
- Vacuum and Analyze: Maintaining data consistency and optimizing query performance

Query Optimization and Performance Tuning

- Query Execution Plans: Understanding query planning and execution
- Query Optimization Techniques: Indexes, statistics, and

High Availability and Fault Tolerance

- Greenplum Fault Tolerance: Segment mirroring and data redundancy
- Backup and Recovery: Strategies for backup, restore, and disaster recovery
- Monitoring and Alerting: Using Greenplum Command Center (GPCC) and SNMP alerts

Greenplum Security

- Authentication and Authorization: Role-based access control (RBAC) and LDAP integration
- Encryption: Data at rest and data in transit encryption options
- Auditing and Compliance: Monitoring user activity and enforcing security policies

Advanced Greenplum Features

- Advanced Analytics: Using Greenplum for machine learning and data science
- Integration with Big Data Ecosystem: Hadoop, Spark, and Kafka integration
- Greenplum Extensions: Adding functionality with PL/pgSQL, PL/Python, and UDFs

Greenplum Monitoring and Maintenance

- Performance Monitoring: Using Greenplum Command Center (GPCC) and system views
- Capacity Planning: Scaling Greenplum clusters and adding nodes
- Patching and Upgrading: Best practices for maintaining Greenplum installations

Greenplum in Production Environment

- Best Practices for Production Deployment: Configuration, tuning, and monitoring tips
- Troubleshooting Common Issues: Diagnosing performance bottlenecks and system failures
- Disaster Recovery Planning: Strategies for data protection and business continuity

Real-world Projects and Case Studies

- Implementing Greenplum Solutions: Industry-specific case studies and success stories
- Best Practices and Lessons Learned from Real-world Implementations

Career Development and Certification

- Building a Career as a Greenplum DBA: Skills development and certification paths
- Interview Preparation: Common Greenplum-related interview questions and scenarios