Praveen Madupu
Mb: +91 98661 30093
Sr. Database Administrator

Apache Cassandra DBA roles and responsibilities — covering **on-premises** and **cloud deployments** (AWS Keyspaces, Azure Managed Cassandra, GCP Bigtable-compatible, etc.).

1. Installation & Setup

- **Cluster Installation**
  - Install Cassandra binaries on Linux/Windows nodes.
  - Configure `cassandra.yaml` (cluster name, seeds, data directories, commit log location).
  - Choose the appropriate Java version (Cassandra is JVM-based).
- **Initial Configuration**
  - Configure **replication factor** and **snitch type** (e.g., `GossipingPropertyFileSnitch` for multi-datacenter).
  - Configure **heap sizes** and garbage collector settings.
  - Configure **commit log archiving** and compression.

2. Cluster Architecture & Capacity Planning

- **Data Modeling**
  - Design keyspaces, tables, partition keys, clustering columns based on access patterns.
  - Avoid anti-patterns like large partitions or unbounded clustering keys.
- **Capacity Planning**
  - Estimate storage needs based on replication factor and expected growth.
  - Plan for read/write throughput requirements.
  - Plan for number of nodes per data center.
- **Replication Strategy**
  - Choose between `SimpleStrategy` (single DC) and `NetworkTopologyStrategy` (multi-DC).

3. Security & Compliance

- **Authentication & Authorization**
  - Enable `PasswordAuthenticator` or integrate with LDAP/Kerberos.
  - Define roles and permissions (`CREATE ROLE`, `GRANT`).

Praveen Madupu
Mb: +91 98661 30093
Sr. Database Administrator

- **Encryption**
  - Configure **client-to-node** and **node-to-node** TLS.
  - Enable encryption at rest.
- **Audit & Compliance**
  - Enable audit logging.
  - Regularly review permission grants.

4. Performance Tuning

- **Read/Write Optimization**
  - Tune **memtable sizes**, **commit log sync policy**, and **compaction strategy** (SizeTieredCompactionStrategy, LeveledCompactionStrategy, TimeWindowCompactionStrategy).
  - Optimize **bloom filter false positive chance**.
- **GC & JVM Tuning**
  - Adjust heap size (usually 50% of available RAM, max 8GB heap).
  - Choose G1GC or CMS based on workload.

- **Query Performance**
  - Avoid full table scans.
  - Use `ALLOW FILTERING` sparingly.
  - Monitor slow queries.

5. Backup & Disaster Recovery

- **Backup Types**
  - **Snapshots**: `nodetool snapshot` (for SSTables).
  - Incremental backups for changes.
- **Restore**
  - Restore SSTables to correct directories and run `nodetool refresh`.
- **Multi-DC Disaster Recovery**
  - Configure replication across data centers.
  - Test failover scenarios.

Praveen Madupu
Mb: +91 98661 30093
Sr. Database Administrator

6. Monitoring & Maintenance

- **Monitoring Tools**
  - Use `nodetool` for real-time status (`status`, `cfstats`, `tpstats`).
  - Integrate with Prometheus + Grafana, DataStax OpsCenter.
- **Health Checks**
  - Monitor `Load`, `Owns`, `Up/Down` status per node.
  - Watch for **read repair** and **hinted handoff** backlog.
- **Routine Maintenance**
  - Run repairs (`nodetool repair`) regularly to maintain consistency.
  - Compact SSTables if needed (`nodetool compact`).
  - Clean up tombstones with `nodetool cleanup`.

7. Patching & Upgrades

- **Rolling Upgrades**
  - Upgrade one node at a time to avoid downtime.
  - Validate schema compatibility between versions.
- **Configuration Changes**
  - Apply changes gradually and validate.

8. Troubleshooting

- **Node Failures**
  - Use `nodetool status` and system logs.
  - Replace dead nodes (`nodetool removenode` / `nodetool assassinate`).
- **Performance Issues**
  - Check `nodetool tpstats` for thread pool saturation.
  - Analyze logs for GC pauses.
- **Consistency Issues**
  - Run `nodetool repair`.

Praveen Madupu
Mb: +91 98661 30093
Sr. Database Administrator

- Check with `consistency level` queries.

9. Cloud-Specific Responsibilities

- **AWS Keyspaces**
  - No node management; focus on schema design, IAM security, performance optimization.
- **Azure Managed Instance for Apache Cassandra**
  - Manage scaling, VNet integration, encryption.
- **Hybrid Cloud**
  - Configure multi-region replication and network latency optimization.

Example Daily Cassandra DBA Checklist

Task	Frequency
Check cluster health with `nodetool status`	Daily
Monitor disk usage per node	Daily
Review logs for errors/warnings	Daily
Verify backups completed successfully	Daily
Run `nodetool repair` on schedule	Weekly
Monitor read/write latency	Daily
Patch security vulnerabilities	Monthly
Test DR restore process	Quarterly

Cassandra DBA Roles & Responsibilities Comparison Table across **On-Prem**, **AWS Keyspaces**, and **Azure Managed Cassandra**:

Praveen Madupu
Mb: +91 98661 30093
Sr. Database Administrator

Area	On-Prem Cassandra	AWS Keyspaces	Azure Managed Instance for Cassandra
Cluster Provisioning	Install & configure nodes manually (OS, Java, Cassandra).	No server management — AWS handles provisioning.	Azure provisions cluster nodes automatically.
Capacity Planning	Plan node count, disk, RAM, replication factor.	No node-level planning — focus on table throughput settings.	Node-level scaling partly automated; plan replication factor.
Data Modeling	Design schema, partition keys, clustering columns.	Same — focus on modeling for throughput & cost efficiency.	Same — must align with managed service replication policies.
Replication Setup	Configure `NetworkTopologyStrategy`, seeds, DCs manually.	AWS handles replication internally; you set replication options.	Azure handles replication between DCs within region(s).
Performance Tuning	Tune heap, GC, compaction, memtables, caches.	Limited tuning; mostly query and schema optimization.	Limited tuning; can adjust compaction and replication settings.
Security	Enable TLS, auth (`PasswordAuthenticator`), role-based access, firewall rules.	Use IAM authentication, VPC access control, encryption at rest + in transit.	Use Azure AD auth, Private Link, managed TLS and encryption.
Backup & Restore	Use `nodetool snapshot`, incremental backups, manage restore process.	Point-in-time recovery; AWS handles backups automatically.	Automated backups; point-in-time restore via Azure Portal/API.
Monitoring	`nodetool`, Prometheus, Grafana, OpsCenter.	CloudWatch metrics & logs.	Azure Monitor, Log Analytics.
Upgrades & Patching	Manual version upgrades (rolling upgrade best practice).	AWS upgrades service automatically.	Azure handles patching/upgrades.
Disaster Recovery	Multi-DC replication, manual failover procedures.	Cross-region replication handled by AWS.	Multi-region replication handled by Azure.
Troubleshooting	Node replacement, repair, compaction, tombstone cleanup.	Limited to query troubleshooting & schema fixes.	Limited to query troubleshooting & schema fixes.

Praveen Madupu
Mb: +91 98661 30093
Sr. Database Administrator

Here's the **color-coded visual chart** comparing Cassandra DBA responsibilities:

Green = Fully DBA responsibility
Yellow = Shared responsibility
Red = Fully handled by cloud provider