

DataStax Apache Cassandra Developer Associate Training

OEM: Apache • Duration: 3 Days (24 hrs) • Code: DSX-CASS-DEV

COURSE MODULES & TOPICS

Module 1: Cassandra Architecture & Internals

- Peer-to-peer ring topology with no single point of failure
- Virtual nodes (vnodes), gossip protocol and node states (joining, leaving, normal)
- Snitch types: SimpleSnitch, GossipingPropertyFileSnitch, RackInferringSnitch
- Partitioner and consistent hashing (Murmur3Partitioner)
- Replication strategies: SimpleStrategy vs. NetworkTopologyStrategy
- Consistency levels: ANY, ONE, QUORUM, LOCAL_QUORUM, ALL, SERIAL, LOCAL_SERIAL
- Hinted handoff (default window: 3 hours) and read repair
- Write path: CommitLog ! MemTable ! SSTable flush
- Read path: Bloom filter ! key cache ! partition summary ! partition index ! SSTable
- Compaction strategies: STCS, LCS, TWCS; tombstones and gc_grace_seconds

Module 2: Data Modeling

- Query-driven design methodology (access-pattern-first)
- Chebotko diagrams — logical and physical model notation
- Partition key, clustering columns, regular columns; composite partition keys
- Clustering column ordering (ASC/DESC); User-Defined Types (UDTs)
- Collection types: List, Set, Map (frozen vs. unfrozen); STATIC columns
- Counter tables, secondary indexes, materialized views
- Denormalization strategies and anti-patterns (wide rows, excess secondary indexes)
- One-to-many and many-to-many patterns in Cassandra

Module 3: Cassandra Query Language (CQL)

- DDL: CREATE/ALTER/DROP KEYSPACE, TABLE, INDEX, TYPE, VIEW
- DML: INSERT, UPDATE, DELETE, SELECT — upsert behavior
- WHERE clause restrictions: partition key required; ALLOW FILTERING caveats
- TTL (Time-To-Live), WRITETIME(), ORDER BY, LIMIT and paging
- Lightweight Transactions (LWT): IF NOT EXISTS, IF condition (Paxos)
- BATCH statements: logged (atomic) vs. unlogged; single vs. multi-partition
- UUID vs. TIMEUUID (uuid() vs. now()); prepared statements

Module 4: Data Loading & Bulk Operations

- DSBulk: load, unload, count subcommands; key flags and options
- COPY TO / COPY FROM (CSV import/export)
- DSBulk vs. COPY trade-offs for large datasets
- sstableloader for offline bulk load

Module 5: Application Development & Driver Integration

- Cassandra driver connection: contact points, port 9042
- Driver-level load balancing: token-aware, round-robin policies
- Retry policies and speculative execution
- Connection pooling and paging through large result sets