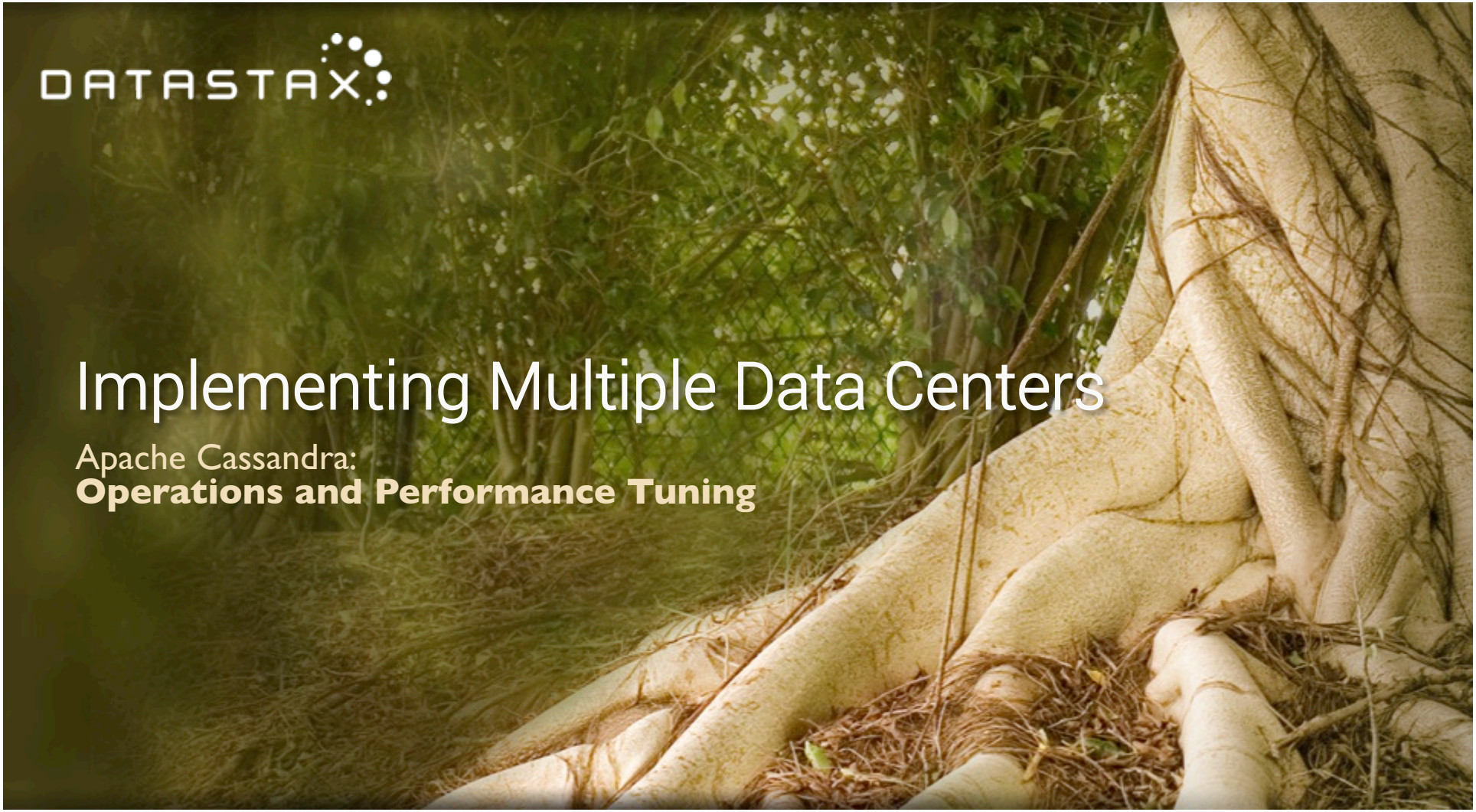


Implementing Multiple Data Centers

Apache Cassandra:
Operations and Performance Tuning

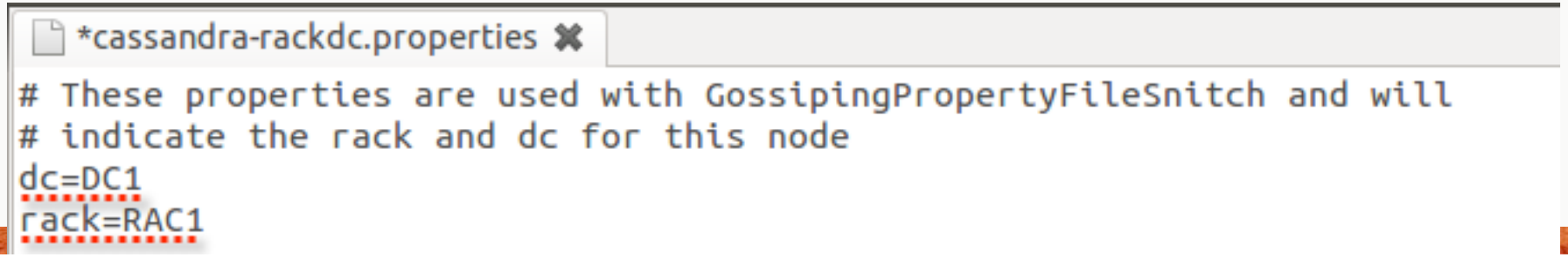
Learning Objectives

- **Implement a multiple data center cluster**

DATASTAX

How are nodes organized as racks and data centers?

- A *cluster* of *nodes* can be logically grouped as *racks* and *data centers*.
  - Node – the virtual or physical host of a single Cassandra instance
  - Rack – a logical grouping of physically related nodes
  - Data Center – a logical grouping of a set of racks
- Enables geographically aware read and write request routing
  - Cluster topology is communicated by the *Snitch* and *Gossip* (discussed ahead).
- Each *node* belongs to one *rack* in one *data center.*
  - A default Cassandra node belongs to rack *RAC1* in data center *DC1*.
- The identity of each node's rack and data center may be configured in its *conf/cassandra-rackdc.properties* file.



Why would you add a second data center?

DATASTAX

- Ensures continuous availability of your data and application
  - If one data center is affected by disaster, Cassandra keeps on working from the other data center(s) if they are geographically diverse.
  - Distributed peer-to-peer architecture guarantees this – no single point of failure due to a master server failure in a master-slave architecture.
  - Data will be accessed from a different geographical location—all the same data will be available.
- Live backup
  - If data centers are in same physical location, but are configured as separate virtual data centers, then a fallback cluster can be quickly switched on if needed.
- Improved performance
  - Latency is reduced if data is accessed from a more "local" data center.
- Analytics

How do Cassandra clusters operate between data centers?

DATASTAX

- A data center is a grouping of nodes configured together for replication purposes.
  - Different replication factors and consistency levels can be configured.
- Data replicates across data centers automatically and transparently.
  - Data replicates in each data center depending on the replication factor.
  - For example, if two data centers each have RF=2, 4 copies of the data will be written.
- Consistency level can be specified at LOCAL level for read/write operations.
  - Restricts actions to the local data center to avoid high latency.
- Consistency level can also be specified as EACH.
  - Requires data at all data centers to be read/written to receive acknowledgment.

What happens when one data center goes down?

- Failure of a data center will likely go unnoticed.
- If node/nodes fail, they will stop communicating via gossip
  - Will be marked as DOWN, but gossip does take about 10 seconds to figure it out.
  - Results in a few timeouts due to consistency level for write/read operations.
- Recovery can be accomplished with a rolling repair to all nodes in failed data center
  - If original data center is restored with the *gc_grace_seconds* period of time (10 days by default), then *nodetool repair* can be used.

How do you implement a multiple data center cluster?

DATASTAX

- Use the *NetworkTopologyStrategy* rather than *SimpleStrategy*
  - Allows for awareness of racks and data centers
  - *SimpleStrategy* is not aware of data centers, and can potentially locate all replicas at one data center, resulting in data loss in a catastrophe.
  - Can specify number of replicas per data center
- Use LOCAL_* consistency level for read/write operations to limit latency .
- If possible, define one rack for entire cluster.
- Specify the snitch.
  - Informs Cassandra about the network topology.
  - Ensures requests are routed efficiently.
  - Allows Cassandra to distribute replicas.
  - All nodes must have exactly the same snitch configuration.

What are the snitch choices?

- RackInferringSnitch – determines location of nodes by rack and data center.
  - 110.100.200.105 [node is 4th octet, rack is 3rd octet, and data center is 2nd octet]
- PropertyFileSnitch – determines location of nodes by rack and data center from cassandra-topology.properties file.
- GossipingPropertyFileSnitch – defines a local node's data center and rack; uses gossip for propagating information to other nodes.
- EC2Snitch – If all nodes are within a single region, the region is treated as the data center and availability zones are treated as racks.
- EC2MultiRegionSnitch – Used for clusters that span multiple regions; regions are treated as data centers and availability zones are treated as racks.

What are specific issues to be aware of in bringing up another data center?

DATASTAX

- Wide Area Network (WAN) bandwidth – how quickly can the data be pushed to new data center.
- Consistency Level – how much data must be pushed to establish correct amount of replication needed to meet consistency level requirements.
- Amount of data that will need to be replicated – amount of data will affect the time it takes for the new data center to be up and running.

- *OpsCenter* is a useful tool for watching the progress of a new data center as it is brought up.

Exercise: Implement a Multi-Datacenter Cluster

Summary

- Data center is a grouping of nodes often associated with a geographical location.
- Data center should be added to a Cassandra cluster if high availability is needed.
- Data center replication is automatic with Cassandra – no ETL.
- Data centers can each specify their own replication factor – different data centers can replicate according to their requirements.
- LOCAL_* consistency levels can be used to ensure that only local replicas are queried, decreasing response latency.
- Data center failure will not affect overall cluster availability – data will simply start flowing from another data center.
- NetworkTopologyStrategy and snitch choice are important to multiple data center performance.
- *OpsCenter* can be useful in monitoring a data center's power-up.

Review Questions

- When do you want to use SimpleStrategy?
- What are two reasons that you might want to use two or more data centers?
- What are the various snitch types?
- If you are setting up two data centers in a single region on AWS, which snitch would be the best choice?
- What time might be best for bringing up a new data center? What factors would you consider?
- How long do you have if a data center failed before you would have to take action?
- What recovery step is important when you need to recover a data center?