2017

Active/Active Cloud Infrastructure

Charles Bradshaw
Virginia Commonwealth University

Edwin Lobo
Virginia Commonwealth University

Curtis Jackson
Virginia Commonwealth University

COMPUTER SCIENCE

Active/Active Cloud Infrastructure

CAPSTONE DESIGN EXPO 2017

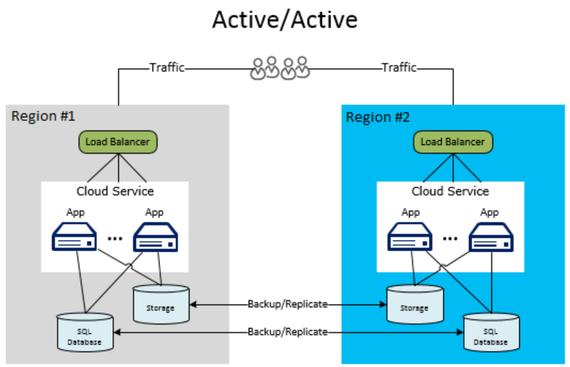
CS 308 | **Team members:** Charles Bradshaw, Edwin Lobo, Curtis Jackson | **Faculty adviser:** Robert Dahlberg | **Sponsor:** Capital One | **Sponsor adviser:** Abhi Ghosh

Leveraging the Cloud

- Since the acceptance of cloud computing, online services have been held to a higher standard of performance
- Major benefits of migrating to the cloud include massive scalability, resilient architecture, and ‘pay-as-you-go’ resources
- Infrastructure consists of several nodes distributed among different geographical locations
- Problems are encountered with multi-datacenter data replication and disaster recovery
- What’s the best strategy to work around these problems?

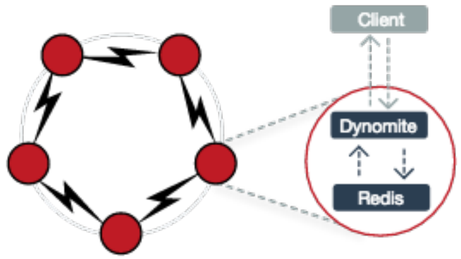
Resilient Cross-Region Replication

- The Active/Active data model (pictured below) addresses both of the previously stated problems at once
- Active/Active infrastructure should deliver cross-region replication at millisecond latencies and a high degree of fault tolerance
- Low write latencies ensure new data is available from a different region (datacenter) as quickly as possible
- High fault tolerance is provided through automatic failover — when a region experiences a service outage, it’s traffic is temporarily routed to another active region



An Initial Approach

- Cloud platforms are typically used to host content through several microservices
- Docker, an open source containerization service, seemed like it could be adapted to fit the problem
  - Benefits include easy scalability and native network protocols which shared containers among multiple Docker instances
  - A major disadvantage included too much abstraction at the database layer, which is where the Active/Active problem resides
- This approach did not solve the problem, it only made it more manageable
- Netflix’s DynomiteDB, backed by Redis, was used to test a distributed Docker based data store



A Working Solution

- The NoSQL database Cassandra is a direct fit to the Active/Active problem
- Masterless architecture, with no single point of failure, and native cross-region replication
- Nodes may be geographically aware
- Engineered with the cloud in mind
- Seamless failover (with minor latencies)
  - Cassandra operates in a cluster
  - Data written to one node is seen by the rest of the nodes in the cluster
  - If DC1 goes down, it’s traffic is routed to DC2 automatically, where replicas of DC1’s tables reside