Faculty of Engineering and Applied Science

SOFE 4790: Distributed Systems

CRN 43524

Homework: Clustering

Alexander Campbell

100703650

Clustering by Consensus

Consensus Clustering is a technique in which a consensus is gained from various iterations of a particular data set. This is very important in distributed systems as there are often many copies of the same data set, doing consensus allows us to aggregate these in an accurate way. This technique of aggregation lends itself to several types of transparency. This supports replication transparency as it allows for there to be several copies of the same data without as many conflicts and without the user being aware. Due to having consensus several copies of data, this also leads to failure transparency where even if there is a failure in one point we can still aggregate data from the other functioning sources.

Leader Election Pattern

The Leader Election Pattern is a method to reduce conflict and issues with shared resources in a distributed system through the election of one deployed instance as the leader. When there is a leader instance to coordinate the actions of other deployments, we can avoid issues such as writing data to the same entry at the same time. The main difficulty in implementing this pattern is the required leader election system must be robust and still work despite any failures. This pattern supports concurrency transparency as it aids in several processes being able to run concurrently across a distributed system without causing conflict.

Infinispan

Infinispan is an open source platform that utilizes data grids to store information. This system allows for distributed systems to store and retrieve data in an extremely performant manner. This means that the user will also be able to access services faster and improve the overall performance of the distributed system as a whole.

Gateway Routing Patterns

A Gateway Routing Pattern is when a distributed system utilizes a gateway interface so that traffic to several disparate services can be routed to one endpoint that will reroute or delegate to the multiple services. This is useful as clients will only need to access a single endpoint but be exposed to multiple services, without having to know their specific endpoints. Gateways support several types of transparencies including failure, performance, and scaling transparency. When using a gateway the control of traffic can be fine tuned, meaning we can redirect traffic to working deployments, direct them to underutilized resources, and know when scaling up or down is necessary as the gateway can measure demand. The major issue with this pattern is that it also may introduce a single point of failure within the distributed system, so the utilized gateway must be extremely robust. Outside of routing there are other ways to use specialized gateways to improve the performance of a distributed system.

Gateways can also be used in the Gateway Aggregation Pattern to increase performance in busy networks. In this pattern, we use a gateway to reduce the amount of separate backend calls to various services an application supports by aggregating multiple calls into one through the gateway. Gateways can also be configured to do administrative tasks in what is called the Gateway Offloading Pattern. Some features that involve gateways of some sort such as authentication, decryption, throttling, etc. can be offloaded to the gateway to make development of the application the gateway points to simpler and improve maintainability as a whole.