Fighting Mongooses
CIS 399
8/7/2017

Final Project Proposal

Summary: We plan to create a Cassandra database that is distributed across our AWS instances. Our planned architecture will have 1 seed node and 2 regular ring nodes.

Objectives:

- Install Cassandra on all 3 of our AWS instances
- Configure 1 seed node and 2 other ring nodes
- Generate test data to populate database
- Demonstrate that data is properly distributed across nodes

Testing Methods: Cassandra includes logs showing which node data is being written too, as well as retrieved from. There are also tools to show the current status of nodes connected to the ring, which we can use to verify that our architecture is implemented as planned.

Use Cases/User Issues:

- A distributed database is useful for applications that require a high volume of read/writes, or in cases where the database needs to be scalable on demand
- Since data is distributed across nodes, users may have issues if some nodes in the ring are offline

Security Issues: Cassandra nodes can, and in our case will, be distributed across multiple instances. This means that the database is necessarily externally accessible, creating a potential for unauthorized remote access. Additionally, a database may contain sensitive information that needs to be secured.

Maintenance/Automation:

- Cassandra is available through some package managers, much of the configuration for individual nodes may be Puppet-able
- Monitoring tools may allow for automatically adding new nodes to the ring as needed

Documentation:

- IP/configuration needed to access database from user applications
- Configuration required to add a new node to the ring

Useful Links:

- Cassandra Documentation