CLI Design

Introduction

Cassandra is a distributed database project which was contributed to Apache by Facebook and is now maintained by contributors from a range of companies and as well as some individuals. Over the last six or so months there has been increasing interest in distributed databases on the back of the NoSQL movement started by some guy from Rackspace. This in addition to well attended talks from key contributors has led to quite a lot of people trying out Cassandra for the first time. Some of these new users have observed that Cassandra although it looks damn cool can is very hard to get started with. This is for a few reasons:

- Modeling data in Cassandra is very different from the typical RDBMS, which people are used to using.
- The tools provided need some work to address error management and overall ease of use.
- Most of the interest comes from segments of the PHP, Python and Ruby communities which are not entirely familiar with the Java Language(for example using timeuuid in php), and how to manage java based services.
- Monitoring of statistics requires an understanding of JMX and is difficult to do outside of the Java Virtual Machine.
- The API through which Cassandra is accessed can be difficult to build, quite brittle at times.

This is mainly due to a focus on tidying up and enhancing core internal functions, as apposed to working on the CLI.

This document will detail a plan to improve the initial user experience by addressing some of these issues.

Overview

In the following sections each of the previously mentioned reasons will be explored in more detail and some ideas outlined to solve them.

Data Modelling

- Need some more sample data models built up and bundled with Cassandra
- It would be handy to have some sort of visualisation to graphically illustrate the structure of the CF(s)

Tools

- Single interface to the Cassandra database, which is in itself an interesting problem as it is distributed.
    - Should be able to access statistics,
    - Modify and query data.
    - Generate reports in various ouput formats, json, xml ect.

Improve access from other Languages

- More examples with a focus on illustrating common anti patterns normally experienced when accessing Cassandra from that language.
- More documentation

Management

- Get some basic summary report interface which just pulls out all the important stuff which people from other languages may not be aware of. Use colors to indicate areas which need attention.
- More documentation about how to see what is going on inside the JVM, and Cassandra

API Improvements

- Create a low throughput easy to use API available via JSON over HTTP so new users can get up and running fast.
- Improve or replace the current API interface with a focus on better error handling and messages
- More documentation

https://c.statcounter.com/9397521/0/fe557aad/1/|stats