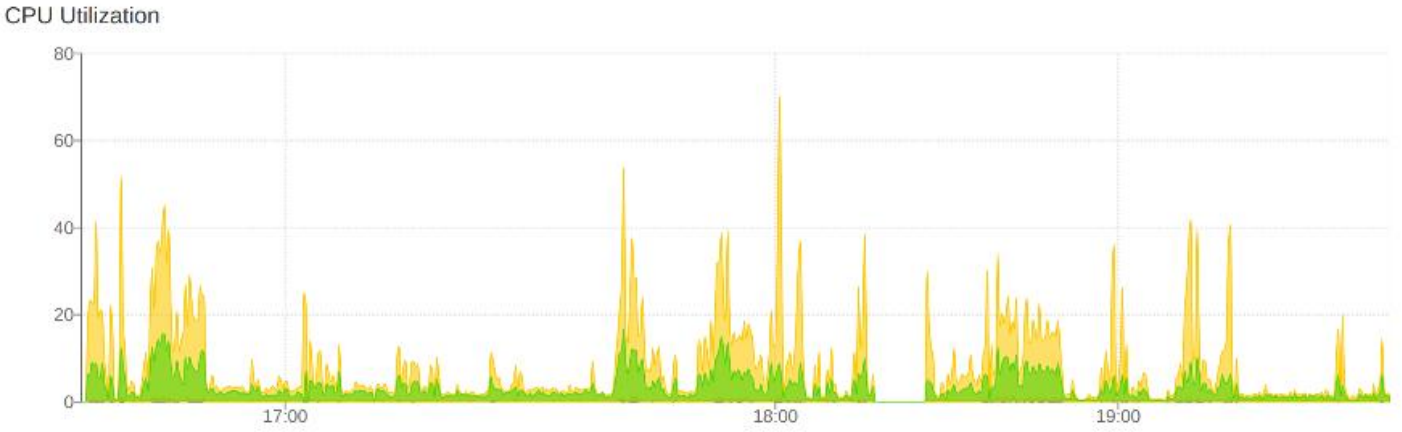
My Calico 2017 Experience

By Luke Chen Shui

Information on Infoset

- The project I worked on, called Infoset, is a high performance time series database. One example of where you might use a time series database is when you want to plot data over time on a graph, such as the amount of CPU usage of a computer every five seconds:



My Experience Working on Infoset - Mentorship

- The mentor, Jordan Jones, was easily approachable.
- He managed to communicate what he wanted me to do and how he thinks I should go about doing it in a clear fashion.
- In the rare event that I didn’t understand something, he had no issue with going through the concept as many times as necessary for me to understand

My Experience Working on Infoset – Background of Mentor

- Jordan Jones has been the mentor for Infoset since its very inception as a Calico project.
- He was the president of the IEEE Utech student branch and was instrumental in building a presence for Calico at Utech.
- He is a firm believer in the power of open source software and has made contributions to large open source projects like Babel.

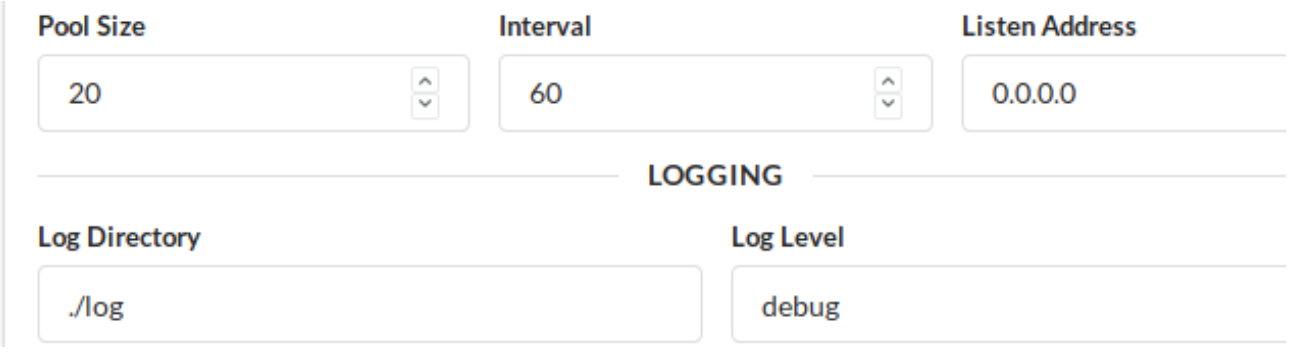
My Experience Working on Infoset – What I Did – Settings Page

- Infoset can be configured by modifying the config.yaml file, but there are numerous disadvantages:

1) The user needs to know the YAML syntax.
2) The user will need to be able to directly access the file, which may be on a remote server that doesn't support GUI text editors.
3) It's easy for people to make accidental syntax errors even if they're familiar with YAML. This could cause serious problems for Infoset

My Experience Working on Infoset – What I Did – Settings Page

- To improve the user experience and ease of use of infoset, I created a configuration/settings web page that essentially acts as a frontend for that config.yaml file:



My Experience Working on Infoset – What I Did – SQLite Conversion

- The installation procedure for infoset used to be quite complex due to the fact that the user would, amongst other things, need to install and configure MySQL, which is the database system that Infoset used.
- A database system is a special type of software used to store and organize data used by software. For example, a website like Facebook would use a database to store a user's name, email address and password for later use. The information in that database would then be queried when a user attempts to login to their account.

My Experience Working on Infoset – What I Did – SQLite Conversion

- We want infoset's installation process to be as easy as possible and MySQL added several unnecessary layers of complexity:

1) MySQL databases aren't easily portable.
2) Using MySQL required that more configuration values exist in the config.yaml file.
3) The user would need to manually create a database user for infoset during the installation process.

My Experience Working on Infoset – What I Did

- To remedy the situation, the decision was made to switch to using SQLite as our database system because:

1) You can literally copy and paste an SQLite database when you need to create a backup or migrate your infoset installation.
2) SQLite only requires one configuration value in config.yaml.
3) Infoset is able to handle the creation of an SQLite database with ease for the user, drastically reducing the complexity of the installation process.

My Experience Working on Infoset – What I Did – Caching

- Caching involves the storing of frequently used data in a way that increases data access speed.
- This method of storing and retrieving data is generally faster than using a database because the best caching systems hold data in memory instead of on the filesystem.
- Efficient use of caching is an extremely important part of building a highly performant and scalable system.

My Experience Working on Infoset – What I Did – Caching

- Good caching important to infoset because infoset could be deployed in an environment that requires that it's able to handle a large amount of incoming time series data.
- Infoset used to use JSON files as its caching mechanism, but this was not ideal because the data would reside on the filesystem, which drastically increased the data storage/retrieval times.

My Experience Working on Infoset – What I Did – Caching

- To remedy this, I switched Infoset to Redis because:

1) It is exceptionally fast. It can do 110,000 storage operations per second and 81,000 retrieval operations per second.
2) It's relatively simple to install.
3) It's massively scalable and great for large amounts of time-based events or logs.

My Experience Working on Infoset – What I Did – CI Server

- When developing software, it is important to test it every time you make a change to the source code to ensure that no bugs surface and that old features still work.
- Manually doing this for every modification quickly becomes tedious and prone to human negligence, especially once a piece of software becomes sufficiently large.
- Unit tests are basically automated software tests that are programmed to ensure that the code that they're responsible is working as desired.
- A pull request is a method of submitting contributions to an open development project. When a contributor to the project wants to submit code that has a new feature or bug fix, they will need to open a pull request which may then be accepted or rejected into the main software repository.

My Experience Working on Infoset – What I Did – CI Server

- When a modification is made to software that has unit tests, instead of a human manually testing each feature by either clicking every button, text box, etc, all the tests can be performed by unit tests. They therefore allow programmers to speed up development time drastically because they can check whether or not a certain feature works more quickly
- If all unit tests pass, it usually means that the code is somewhat stable and may be ready for production use.

My Experience Working on Infoset – What I Did – CI Server

- A continuous integration server basically detects when a pull request is opened or when new changes are uploaded, runs the unit tests, compiles the software and optionally deploys the new version of the software if all unit tests pass.
- If the unit tests or compilation fail, the continuous integration server can indicate to the developers that there is a problem that needs to be addressed with the pull request or upload.
- Continuous integration servers may also host downloadable versions of successful builds of software for users to download.

My Experience Working on Infoset – What I Did – CI Server

- Infoset used to utilize Travis CI, but I was asked to create a Jenkins server because:

1) Jenkins can be run on our own server, whereas Travis CI is hosted on servers that we don't control.
2) As a consequence of the above point, deployments can be easily tailored to our specific server environment. Where Travis would need to use something like SSH or FTP to deploy builds to our server, Jenkins can just use regular shell commands since it's already running on our server.
3) It is a tool which is written in Java. Hence it can be portable to almost all major platforms.

The next page contains a picture that shows result of unit tests that were run when I submitted a pull request to infoset this past summer.

Pull Request, CI Server and Unit Test Example

Pr0x1m4 approved these changes on Jun 21 — View changes

Pr0x1m4 merged commit `3db029c` into `PalisadoesFoundation:master` on Jun 22 — Hide details — Revert

2 checks passed

- **continuous-integration/jenkins/pr-merge** This commit looks good — Details
- **continuous-integration/travis-ci/pr** The Travis CI build passed — Details

Pr0x1m4 referenced this pull request on Jul 1

Create a React-based web frontend to manage infoset-ng's configuration — Closed

#127

My Experience Working on Infoset – What I Learned

- I gained some more detailed knowledge of how to use Flask, which is a Python microframework for building backend servers. Infoset is written using Flask.
- I was introduced to the basics of setting up a Jenkins build server.
- I utilized Redis caching for the very first time.