Open Source Platform for Analysis and Visualisation of Machine Data

Project Description:

Server logs are big. Really big. You just won't believe how vastly, hugely, mind bogglingly big they are. I mean you may think that the amount of project proposals you have to go through is big, but that's just peanuts to server logs.

The aim of this project is to produce an **open source** platform for the harvesting, processing and visualisation of Machine Data (logs etc), and expose this using both an interactive web interface, REST api and language bindings.

To do this we would need to build a series of components/layers; a multi-platform data agent for forwarding machine data, a rich backend system that intelligently stores log data to minimise access time and allow real time access and querying, plus an interactive web frontend that allows easy visualisation and reporting using a powerful query language.

Typical use cases:

- Analysing web server traffic to determine times of heavy load, and the specific bottlenecks this causes.
- Calculate efficiency of a distributed task by monitoring each virtual machine
- Setting up automatic alerts for emergencies such as failed hard drives
- Visualise internetworking of machines in datacenter.

Challenges

- How to efficiently store large data sets in a way that gives us the quickest access times and most redundancy
- To produce a real time web socket based reporting UI
- Visualising this data in a user-friendly and insightful manner that is dynamic for the infinite types of data that can be visualised from machine data
- How to predict certain events are going/likely to happen (failing hard drives)

Intended Group Size: 4