

EECS 345 Quarter Project

Monitoring Large-Scale Distributed Systems

Winter 2008

Important Dates

Project out: February 8, 2008.

Presentation: March 13, 2008.

Demo and report: March 18, 2008.

Project Overview

There is one single project for this class. You will work on it, throughout the rest of the quarter. As part of this project you will develop a simple monitoring service for large-scale distributed systems including a visualization tool for it. Monitoring is a fundamental abstraction for large-scale distributed systems. It serves as basic building block for a variety of applications – from financial services to online gaming. You will do your work in PlanetLab (www.planet-lab.org), a global-scale testbed for distributed computing experimentation consisting of over 800 nodes.

Specifications

Your monitoring service should include an extensible set of sensors, a data aggregation and management component and, and a graphical user interface. You can implement it in your language of choice.

The service should support the monitoring of both, network conditions (e.g. latency, bandwidth) and machines attributes (e.g. CPU load, free memory). It should include sensors for at least one example of each type of information, but be easily extensible.

The data aggregation and management component collects measurements from the individual sensors and aggregates them in a scalable manner. There is a number of approaches to collecting, maintaining and aggregating measurement data and you are free to use whatever approach you prefer (e.g. gossiping). Remember, however, that your design document must justify this and all other design decisions.

The user interface for your monitoring service does not need to be particularly complicated. It should simply act as a client application, providing access to the functionality of the service in an interactive manner.

Additional Details

1. You can work alone or in teams of 2 people.
2. I am expecting you to work 7-10hr per week on this course; this includes everything, from assigned readings to working on the term project. Of course, this is only an average and some of you may need to put a bit more/less time than others. To ensure the workload remains manageable, in particular when it comes to the project, is important you make all your meetings.
3. Each team will meet weekly with the instructor – you are expected to make the majority of the meetings.

Grading

Please remember that your project counts for 45% of your final grade! The grade of the project will be determined roughly as follows:

1. Design document - 15%
2. 4/5 weekly meeting - 15%
3. Working implementation - 50%
4. Visualization - 20%

General Advice

Perhaps the first thing to do is to develop a simple ping-all like tool – an easy tool that let's you ping one or a set of nodes and aggregates the results.

It may be a good time investment to read some of the existing literature on the topic. Some well known examples include Ganglia [1], Nixes [2], S^3 [7], PIER [3], Sophia [5], SDIMS [6], and Astrolabe [4].

Given the open-ended nature of the project, you should make a point of meeting regularly with me and **the earliest the beter**.

References

- [1] <http://ganglia.sourceforge.net>.
- [2] <http://www.aqualab.cs.northwestern.edu/nixes.html>.
- [3] HUEBSCH, R., HELLERSTEIN, J. M., LANHAM, N., LOO, B. T., SHENKER, S., AND STOICA, I. Querying the Internet with PIER. In *Proc. of the VLDB Conference* (May 2003).
- [4] VANRENESSE, R., BIRMAN, K. P., AND VOGELS, W. Astrolabe: A robust and scalable technology for distributed system monitoring. *ACM TOCS* (2003).
- [5] WAWRZONIAK, M., PETERSON, L., AND ROSCOE, T. Sophia: An information plane for networked systems. In *HotNets* (2003).

- [6] YALAGANDULA, P., AND DAHLIN, M. A scalable distributed information management system. In *ACM SIGCOMM* (August 2004).
- [7] YALAGANDULA, P., SHARMA, P., BANERJEE, S., BASU, S., AND LEE, S.-J. S3: A scalable sensing service for monitoring large networked systems. In *SIGCOMM Workshops* (September 2006).