582713 Warehouse-Scale Computing

This is the home page for the course 582713 Warehouse-Scale Computing (4 cr) held at the Department of Computer Science.

Mainly, this course involved the administration of very large-scale systems spanning hundreds of thousands of machines. Students are taught to maintain systems which maintain other systems.

The course is heavily prototype-oriented and contains no lectures. Points are given for completing weekly exercises in the NodesLab. There may be an exam, but this information is still pending.

First iteration of the course will be held during period IV of teaching year 2014, i.e., 10.3.–18.5., or beyond if necessary for completion of the exercises.

Pre-requirements

For the Spring 2014 iteration of the course, students will be cherry-picked in order to balance the amount of work done in the exercises. This means that the list below is tentative:

- 582351 Linux Fundamentals or greater scripting skills
- The networking part of 582317 Linux-ylläpito: IPv4, IPv6 interfaces and routing, ITables, tcpdump / netfilter + libpcap
  - Note that the spring 2014 iteration overlaps with this course, so you need to have gained these skills earlier

Q: I'm a foreign student and 582317 is only held in Finnish, can I still be eligible? 
A: Yes, but you need to have learned the techs listed above from somewhere and convince me about this. This is for your own benefit, as joining the course without the requirements is just a waste of time.

Exercises

- Exercises – general information
- Checking – fifth set
- Coins – sixth set

Materials


See this LISA presentation for Google’s administration system involving 40k+ mac workstations https://www.usenix.org/conference/lisa13/managing-macs-google-scale.


Contents

Primary themes include (re)installation, system monitoring, and fault-tolerance.

Emphasis is on hands-on experiments which will be completed in the NodesLab. For this purpose, students are granted physical access to the lab but encouraged to complete as much as possible of the administration remotely.

Tentative topic list:

1. PXE Installations
   - Familiarizes the students with the PXE/DHCP bootup environment and how to automate the installer
   - NodesLab already has a basic script for PXE installations, students will only need to customize it
2. Lightweight or container-based virtualization
   - Alternatives: Linux-VServer and OpenVZ
   - These tools provide application-level containment with minimal overheads due to the virtualization layer
3. Effing package management (FPM)
   - Creation of Deb/RPM packages
   - Look at examples of well known packages
   - Students will assemble add-on packages that provide the configuration stanzas for a well known package
4. Role-based automatic configuration
   - Puppet
   - Role-based tools enable reinstalling a server configuration from a similar, but unconfigured server template
5. Nagios
   - Monitoring and reporting of service failures
   - Local hardware monitoring through NRPE / NSCA
   - Fault-recovery / self-healing through response scripts
6. BitCoin client
   - The objective is to keep the client running at all times

Navigate space