Processing Usage Metrics
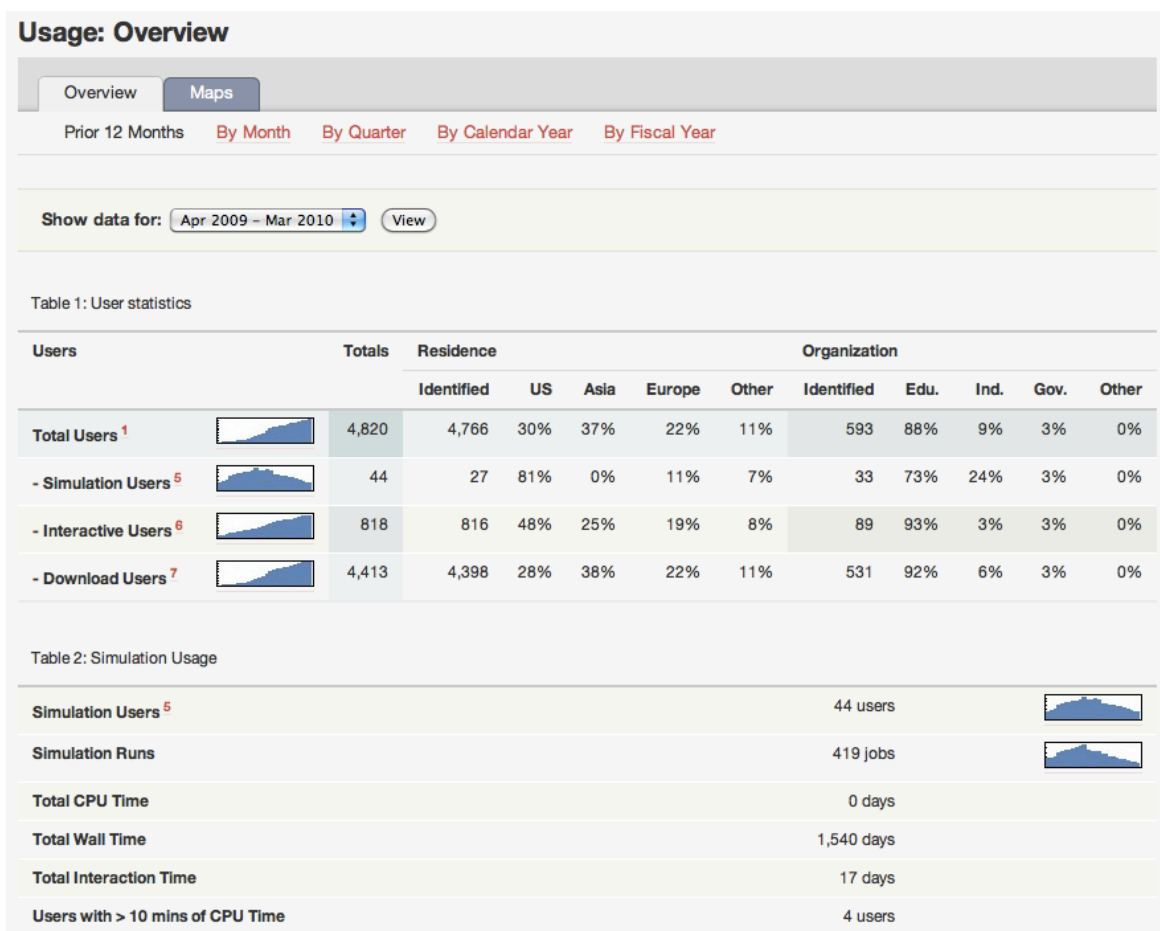
Swaroop Shivarajapura

HUBbub 2010

At the end of this presentation you will be familiar with …

- Available Metrics
- Definitions
- Log Files
- Processing Metrics
- Setup, Maintenance and Troubleshooting

- Exercise

Summary Usage - http://myHUB.org/usage

Usage: Overview

Overview | Maps

Prior 12 Months | By Month | By Quarter | By Calendar Year | By Fiscal Year

Show data for: Apr 2009 – Mar 2010 ◆ (View)

Table 1: User statistics

Users		Totals	Residence					Organization				
			Identified	US	Asia	Europe	Other	Identified	Edu.	Ind.	Gov.	Other
Total Users [1]		4,820	4,766	30%	37%	22%	11%	593	88%	9%	3%	0%
- Simulation Users [5]		44	27	81%	0%	11%	7%	33	73%	24%	3%	0%
- Interactive Users [6]		818	816	48%	25%	19%	8%	89	93%	3%	3%	0%
- Download Users [7]		4,413	4,398	28%	38%	22%	11%	531	92%	6%	3%	0%

Table 2: Simulation Usage

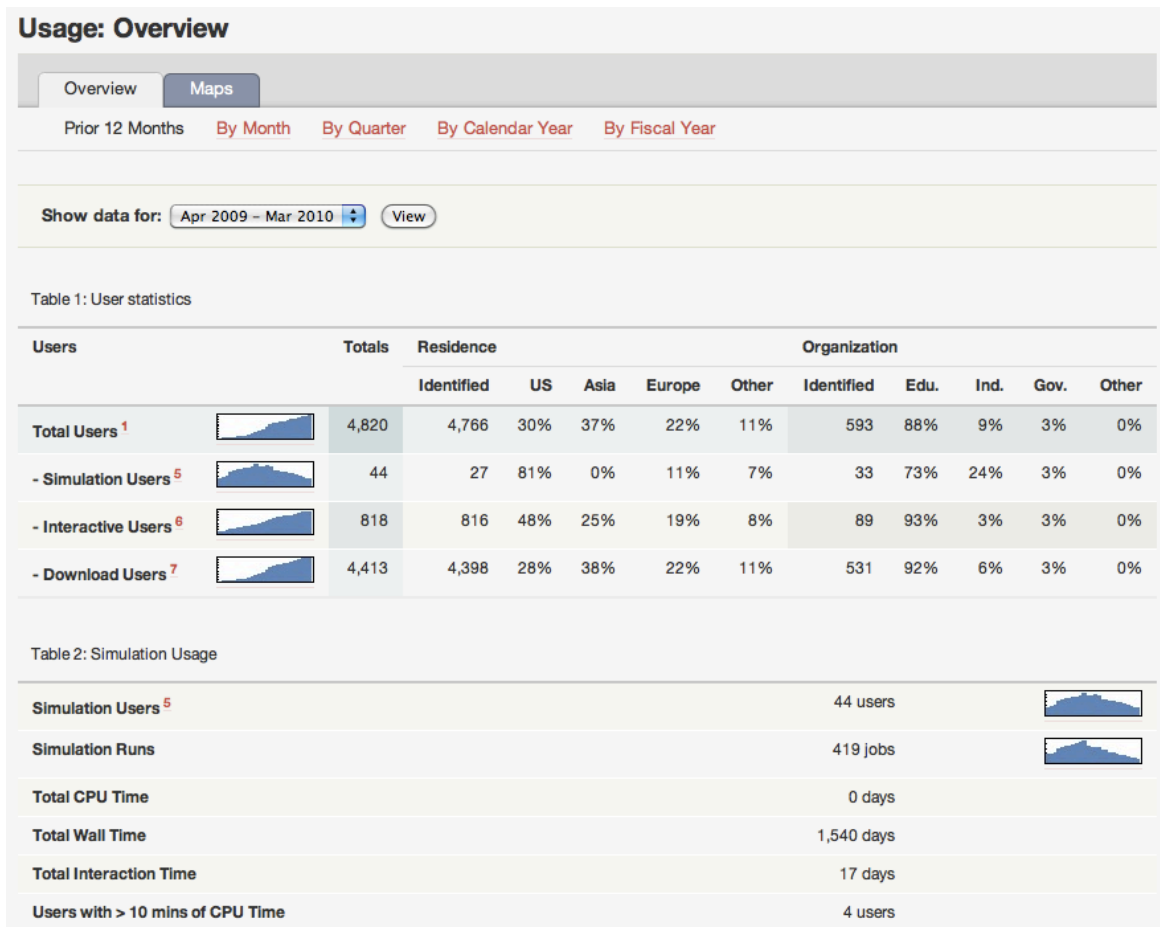|		
Simulation Users [5]	44 users	
Simulation Runs	419 jobs	
Total CPU Time	0 days	
Total Wall Time	1,540 days	
Total Interaction Time	17 days	
Users with > 10 mins of CPU Time	4 users	

Map of users currently online - http://myHUB.org/usage/maps

How we count "Users"

- Registered Users (1) - Users that logged in. User registration assigns a unique login to each individual user.

  » Simulation Users - Registered Users that ran one or more simulation runs

- Unregistered Interactive Users (2) - Identified by unique hosts/IPs, that had an active session* without logging in.

- Unregistered Download Users (3) - Identified by unique hosts/IPs, that had an active session* of less than 15 minutes without logging in and downloaded a non-interactive resource such as PDF or podcast.

- Total Users = (1) + (2) + (3)

*Active session = Begins when an IP is active on the site for at least 15 minutes. Ends when inactive for more than 30 minutes, including time spent viewing videos.

Summary Usage - http://myHUB.org/usage

Usage: Overview

Overview | Maps

Prior 12 Months | By Month | By Quarter | By Calendar Year | By Fiscal Year

Show data for: Apr 2009 – Mar 2010 [View]

Table 1: User statistics

Users		Totals	Residence					Organization				
			Identified	US	Asia	Europe	Other	Identified	Edu.	Ind.	Gov.	Other
Total Users [1]		4,820	4,766	30%	37%	22%	11%	593	88%	9%	3%	0%
- Simulation Users [5]		44	27	81%	0%	11%	7%	33	73%	24%	3%	0%
- Interactive Users [6]		818	816	48%	25%	19%	8%	89	93%	3%	3%	0%
- Download Users [7]		4,413	4,398	28%	38%	22%	11%	531	92%	6%	3%	0%

Table 2: Simulation Usage

Simulation Users [5]	44 users	
Simulation Runs	419 jobs	
Total CPU Time	0 days	
Total Wall Time	1,540 days	
Total Interaction Time	17 days	
Users with > 10 mins of CPU Time	4 users	

Metrics - Definitions

Usage – Simulation

Table 2: Simulation Usage

Simulation Users [5]	8,208 users	
Simulation Runs	346,741 jobs	
Total CPU Time	2,300 days	
Total Wall Time	24,148 days	
Total Interaction Time	7,749 days	
Users with > 10 mins of CPU Time	2,036 users	
Avg. Number of Simulation Runs/User	42 jobs	
Avg. Time between First and last Simualtion	187 days	
Repeat Users with > 10 Simulation Jobs	4,338 users	
Repeat Users with > 3 Months [9]	2,446 users	

Usage – Miscellaneous

Table 3: Miscellaneous Statistics

Domains Served	13,660
Cummulative Interactive User Sessions [10]	193,655
Cummulative Session Time	8,082 days
Visitors [11]	279,647
Visits [12]	909,026
New Accounts	8,131
Max User Logins	465 users on 2010-02-09
Web Server Hits	59,525,241

Metrics - Available Log Files

- User Registration Information (jos_xprofiles)
  - » Name, Email, Organization, Org. Type, Country Residence etc

- Apache log

```
2010-04-12 21:08:22 EDT - "GET /resources/169/download/2.2_Rappture_Advanced.pdf
HTTP/1.1" 200 289555 128.46.18.28 "http://hubzero.org/resources/168" "Mozilla/
5.0 (Macintosh; U; Intel Mac OS X 10_6_3; en-US) AppleWebKit/533.2 (KHTML, like
Gecko) Chrome/5.0.342.9 Safari/533.2" - 0 150812 -
210e36d84277c2f19aad082dc031adb=0667a4181a3e56b00a45b8b8481ed878
```

- Joomla's auth log

```
2010-04-06 15:45:11 swoop 128.46.16.26 detect
2010-04-08 14:45:06 2143 [swoop] 128.46.16.26 login
2010-04-08 14:55:10 swoop 128.46.16.26 logout
2010-04-09 04:03:19 swoop 128.46.16.26 invalid
```

- Middleware logs

```
Sessionlog:
Session# – Username – IP – ExecHost – Start – Appname – Wall – CPU – View – Status


Joblog:
Session# - Job# – Superjob# – Event – Start – Wall - CPU – Ncpus – Status – Venue

Viewlog:
Session# – Username – IP – Host – Time – Duration
```

Metrics Processing Steps

- Import Apache and Joomla's auth logs
- Import Middleware Logs
- Resolve hostnames and extract domain information
- Stitch Sessions
- Remove identified bots
- Look up user information for simulation logs
- Country Location lookup for IPs
- Generate Summary and Plots
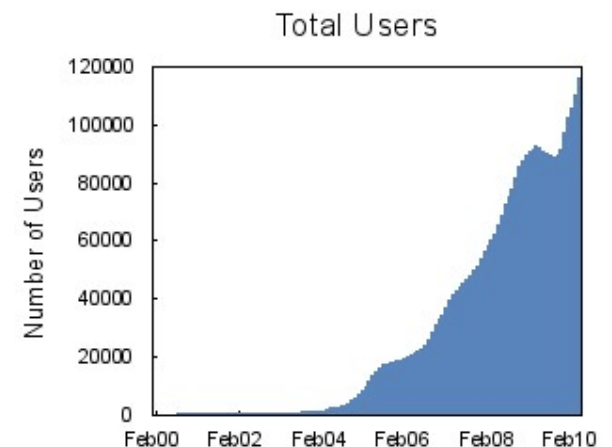
Metrics – Setup and Maintenance

Setup and Maintenance and Troubleshooting

- Check apache, Joomla and middleware logs
- Configure Cron for rotation and import of logs
- E.g. Cron setting for whoisonline module
  /etc/cron.d/crontab_hub

```
*/5 * * * *      www-data         /opt/hubzero/bin/xlogfix_whoisonline
```

- Monitor disk space
- Backup log files and database tables

Demo:
 - Signing up for Google Analytics

Exercise
 - Setting up a Google maps API Key