High Density Visualizations

Chuck Boeheim

SLAC

Clusters and Scaling

A cluster is a very large error amplifier

You pay for capacity and bandwidth separately

Cheap hardware isn't

GUIs don't scale
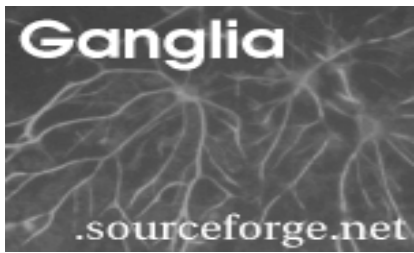
Typical Display Problems

Diffuse information

Few variables displayed at once

Context and detail mutually exclusive

Excessive white space and decoration

LSF Status Display

Ganglia
.sourceforge.net

Batch Domain Report for Fri, 26 Oct 2007 14:54:19 -0700

Last [hour] Sorted [descending]

Batch Domain > [--Choose a Source]
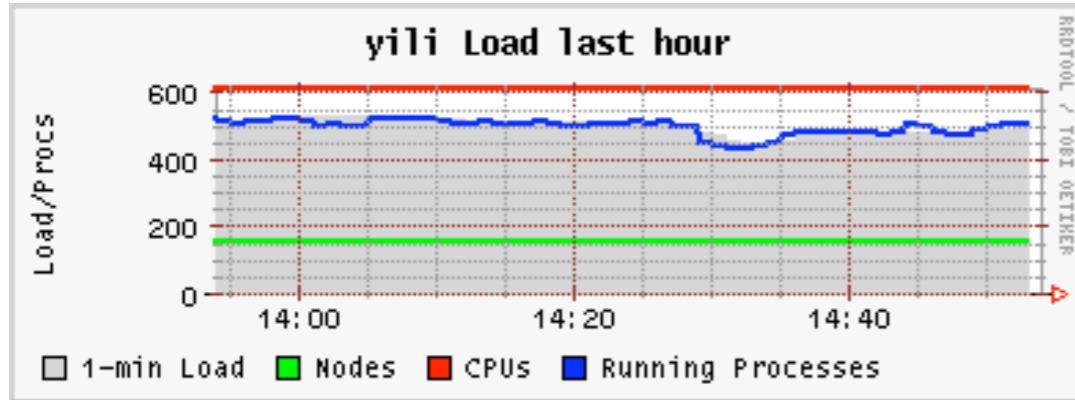
yili (physical view)

CPUs Total: **612**
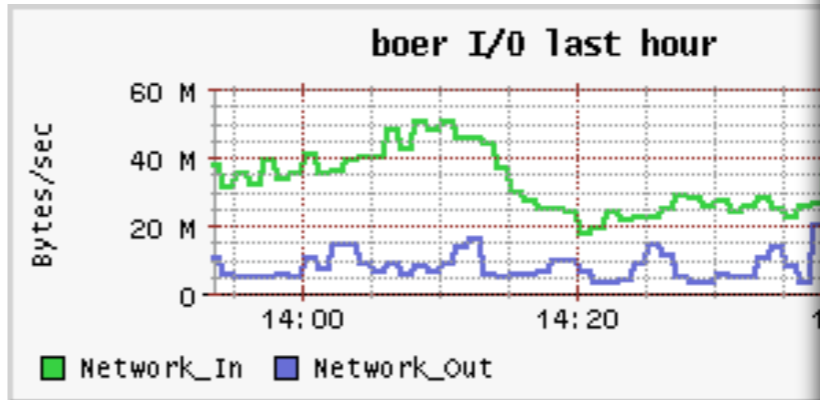
Hosts up: **153**

Hosts down: **0**

Avg Load (15, 5, 1m):
79%, 80%, 82%

Localtime:
2007-10-26 14:53
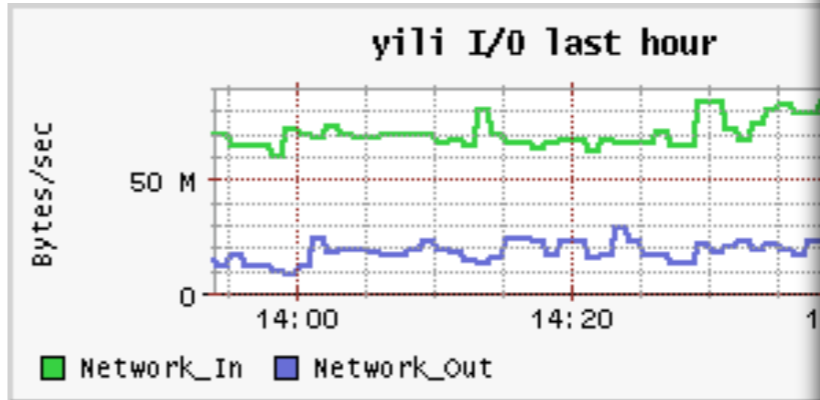
boer (physical view)

CPUs Total: **528**
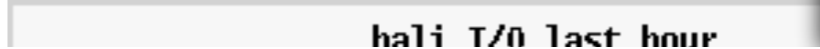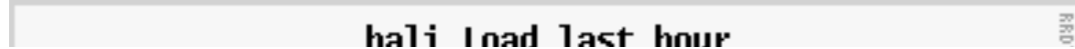
Hosts up: **132**

Hosts down: **0**

Avg Load (15, 5, 1m):
79%, 78%, 78%

Localtime:
2007-10-26 14:53

bali (physical view)

CPUs
992

bali Load last hour

bali I/O last hour

Good Practices

Allow the eye to find patterns

Provide the detail in context

Use all the space

Think beyond bar charts

Correlate variables

Allow comparisons

Provide different views

Allow exploration

A Data Exploration Tool

Magnaview (www.magnaview.nl)

Commercialization of Sequoiaview from University of Eindhoven

Can visualize 1 million observations

Support has been excellent

Also evaluated freeware packages, did not find one with as many features
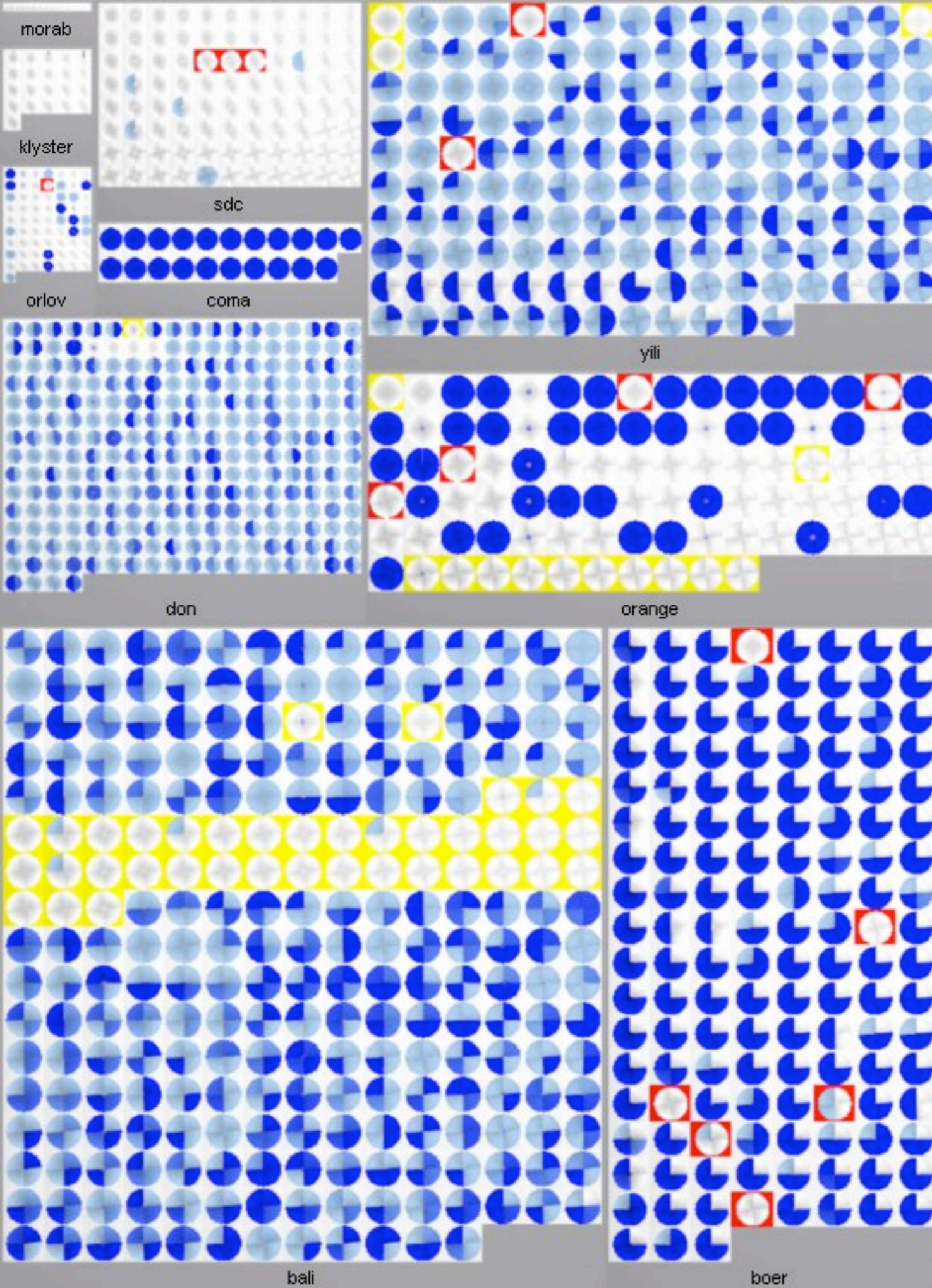
Display types in Magnaview

- Treemaps
- Colormaps
- Pixelmaps
- Grids
- Slices
- Bar charts
- Pie charts
- Line charts

Each type can contain any other type
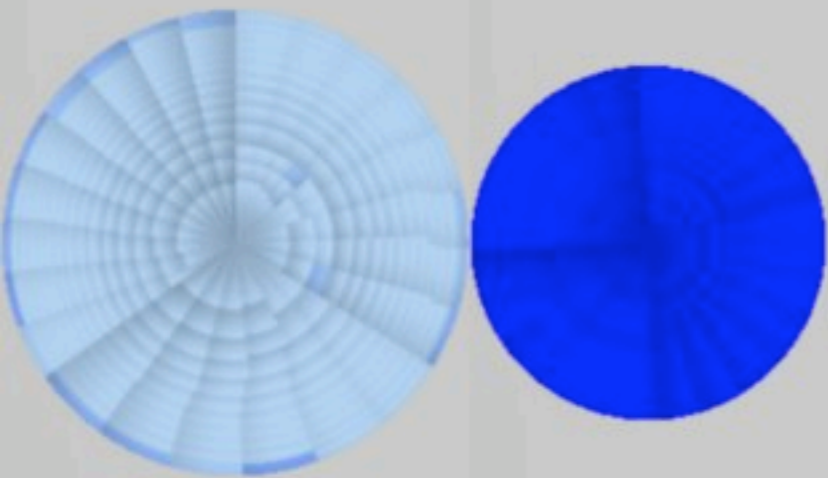
Each project can have multiple views of data

Status of batch system
Size of node
Size of cluster
Job distribution
Efficiency of job
Status of node

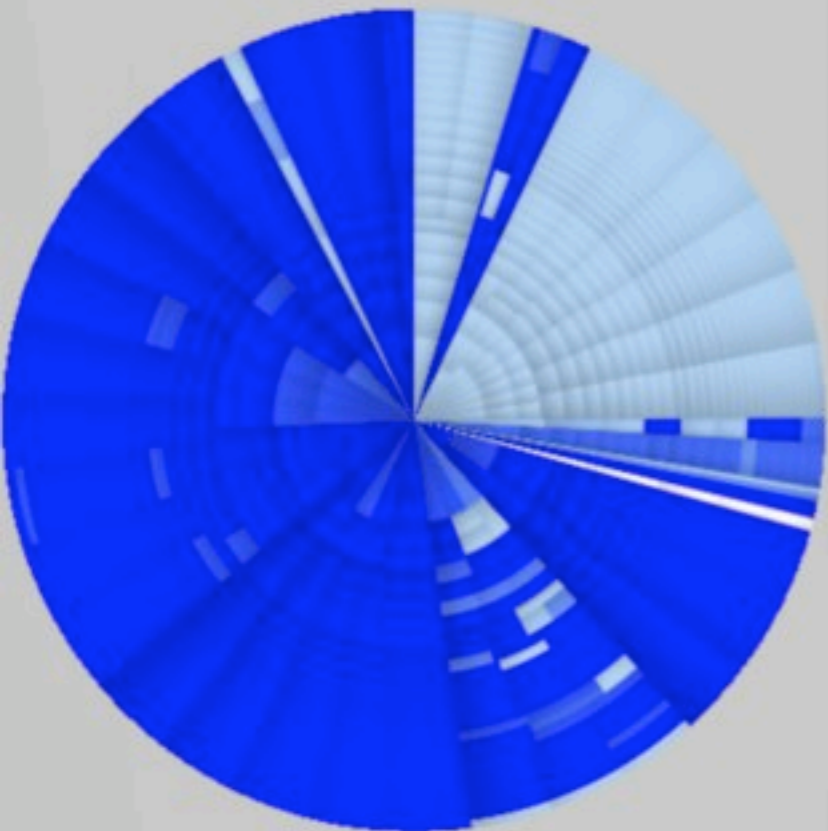jobs=306   jobs=254   jobs=2036

simuq   kipac

jobs=77   jo... jobs=3?
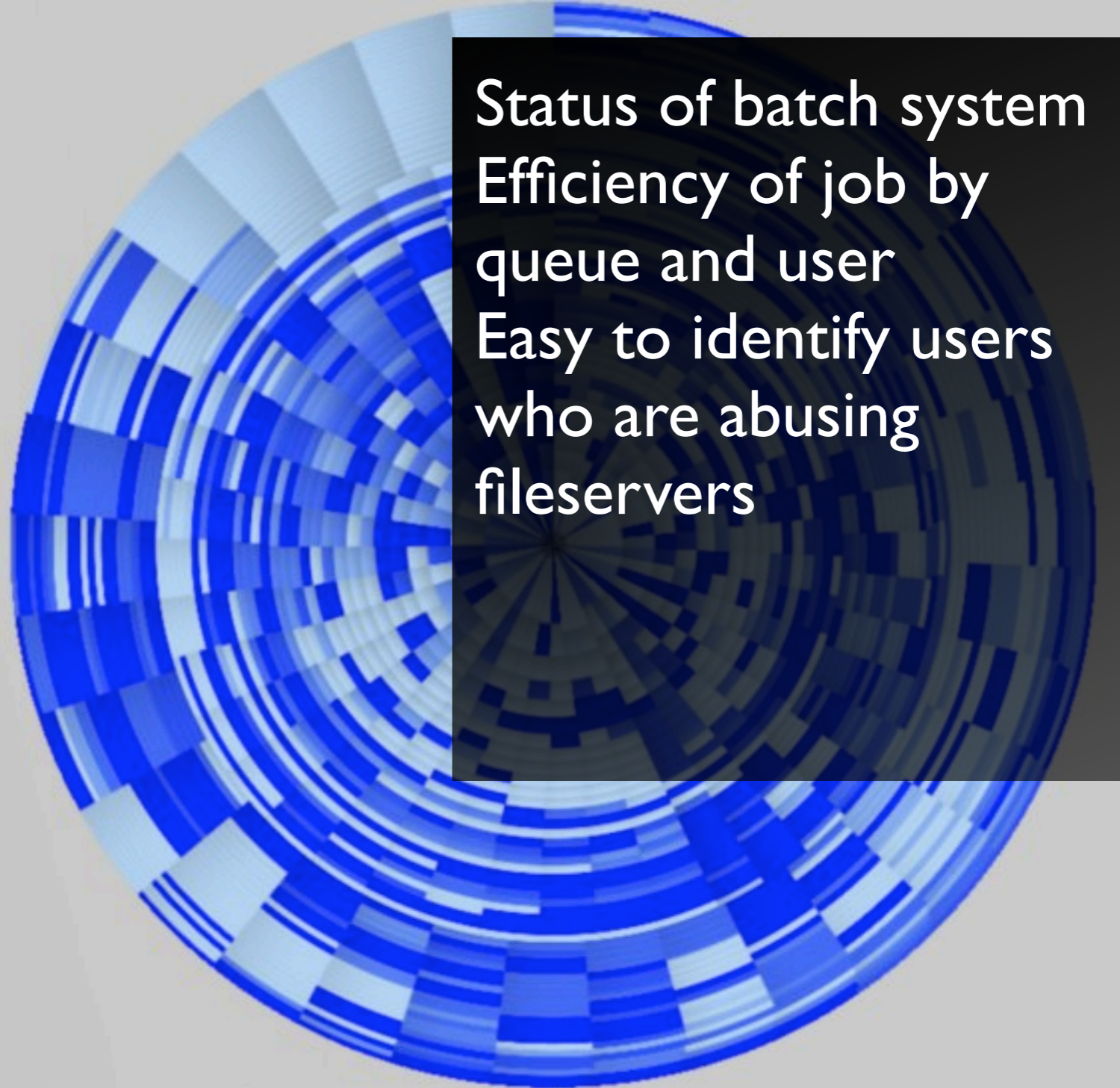
minitestq   s...   mp?

jobs=1803

jobs=1918

miniq

general   kanga

Status of batch system
Efficiency of job by queue and user
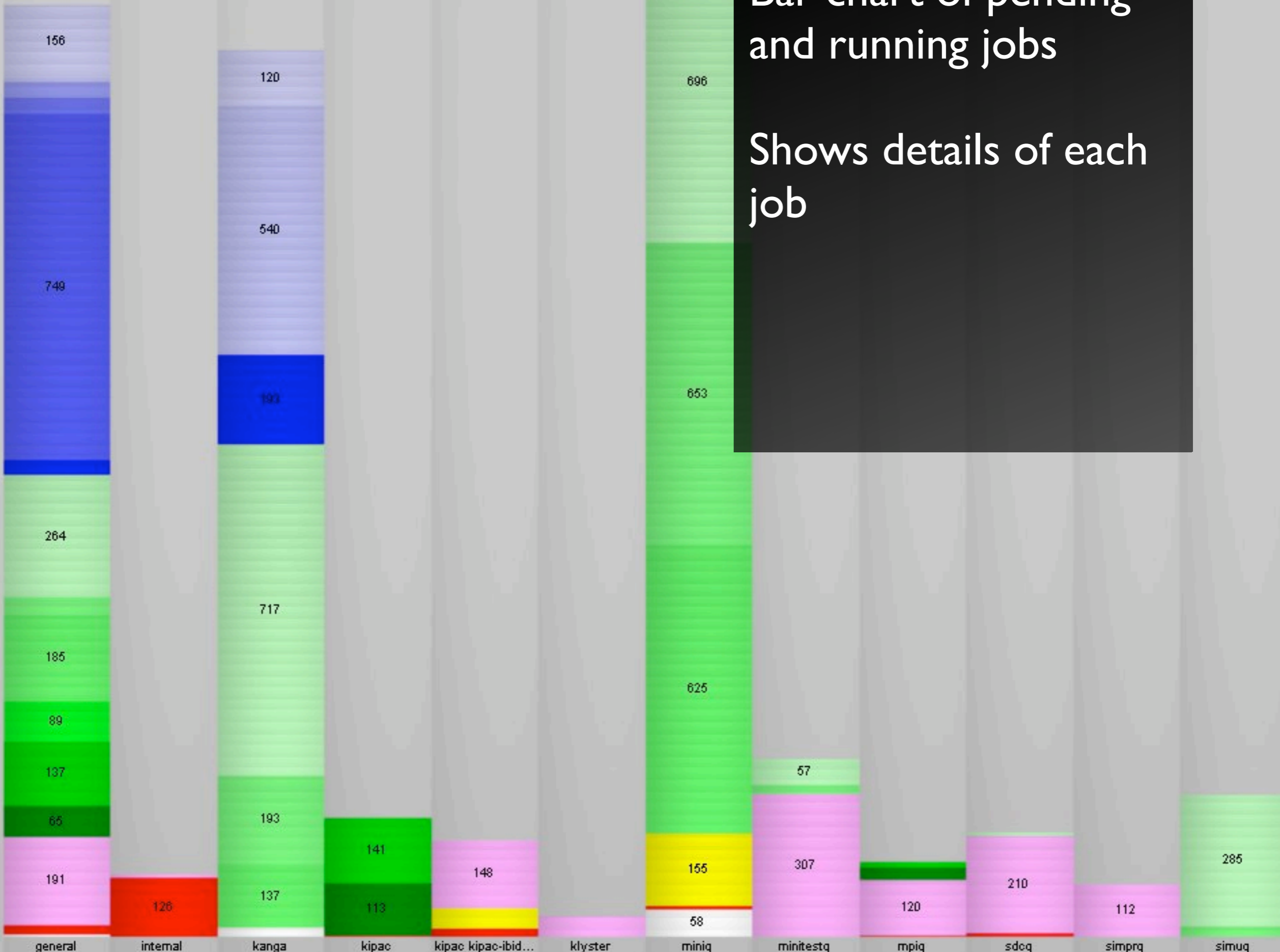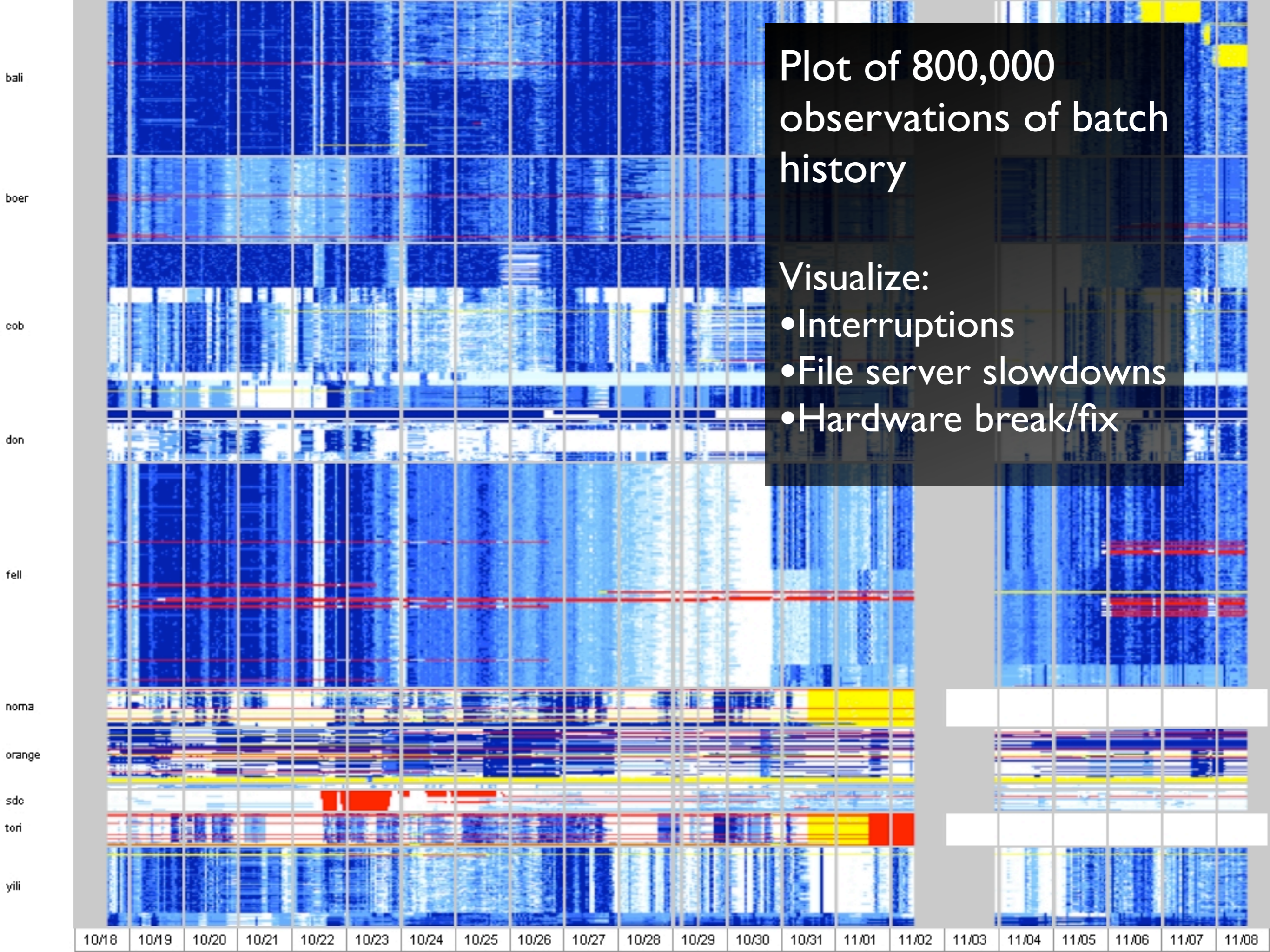Easy to identify users who are abusing fileservers

Tree map of users

Identifies large users

Correlates queue time / run time with user

Bar chart of pending and running jobs

Shows details of each job

Plot of 800,000 observations of batch history

Visualize:
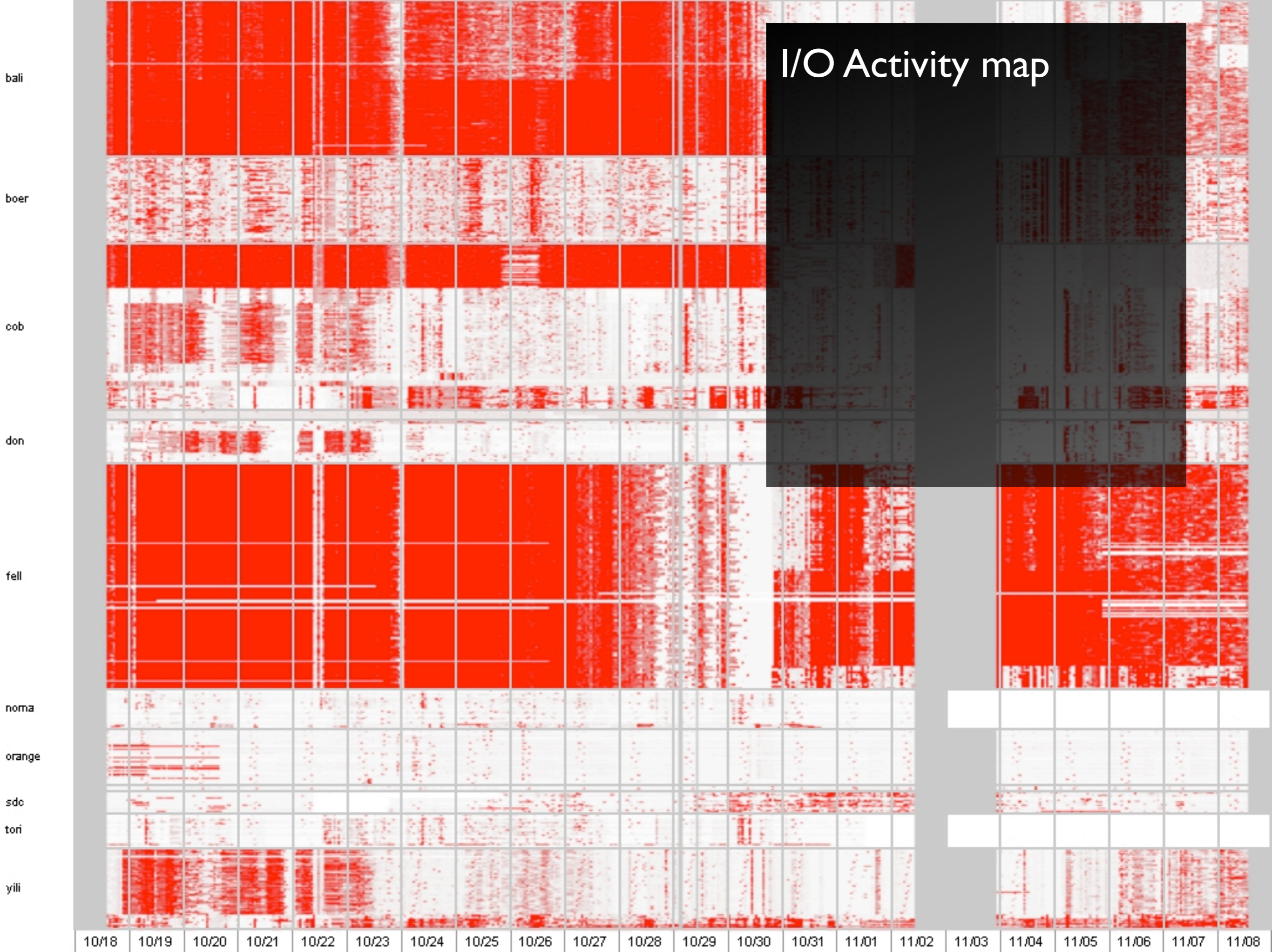- Interruptions
- File server slowdowns
- Hardware break/fix

Batch Slot Occupancy

% of job slots on each host occupied

I/O Activity map

15

Heat map by rack location

Identified a faulty fan sensor
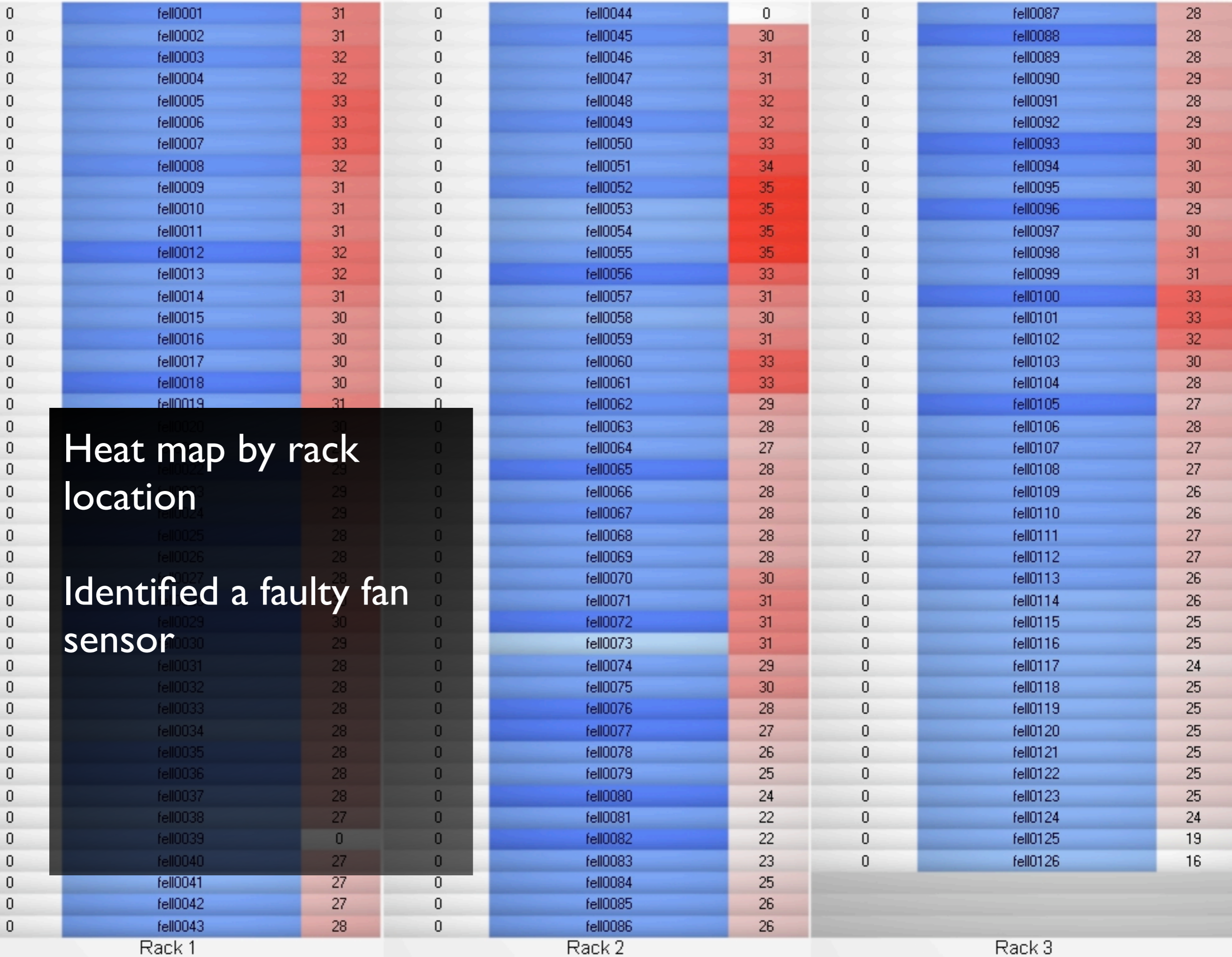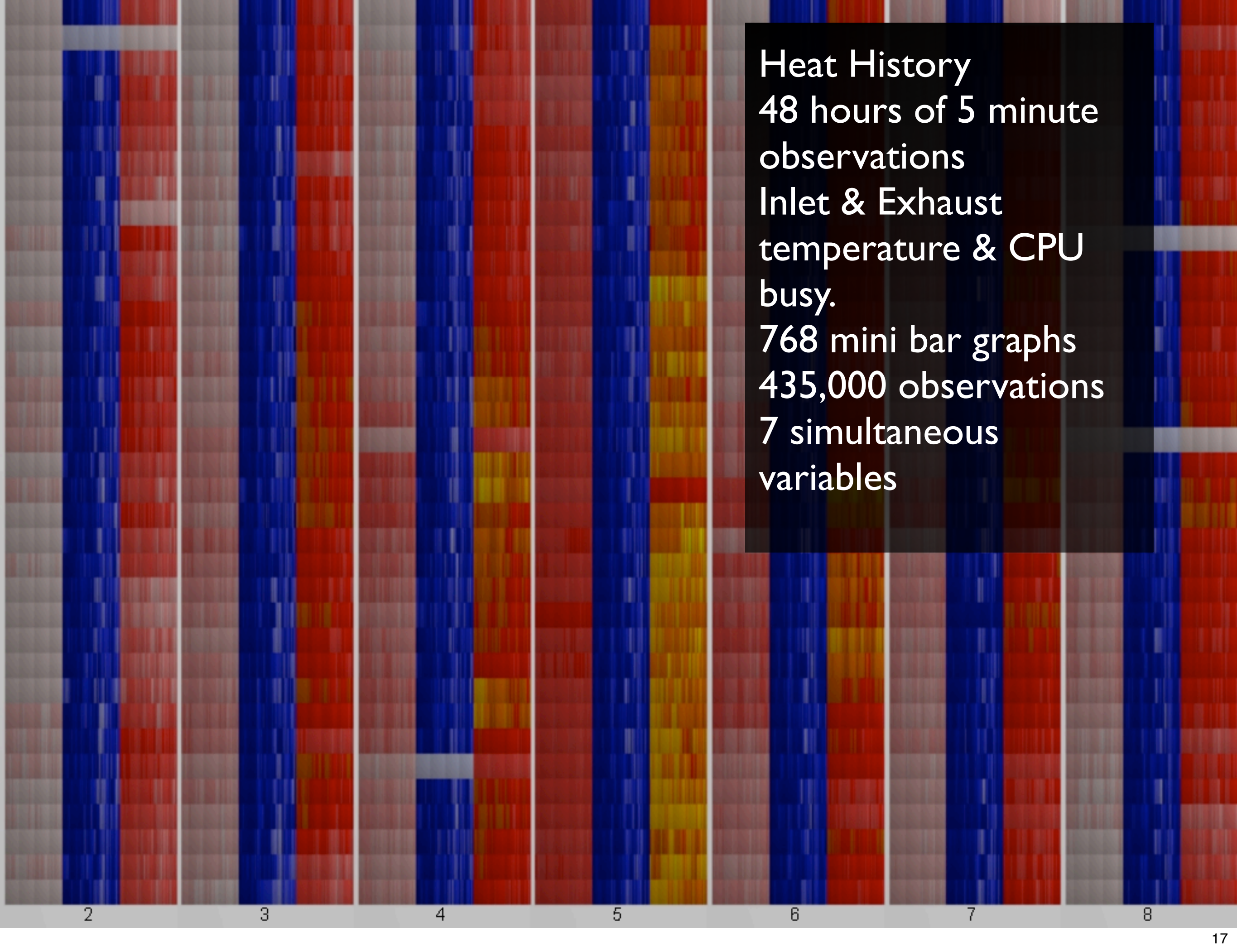
Heat History
48 hours of 5 minute observations
Inlet & Exhaust temperature & CPU busy.
768 mini bar graphs
435,000 observations
7 simultaneous variables

A map of the storage pools at SLAC

Color is coded to indicate the most dangerously full file systems

Distribution of cron job times across servers

Conclusions

Dense displays allow more meaningful information to reach the user

There are over 18 million bits on a screen - use them

GUIs can scale

Questions?