

"Metrics"

Jeff Mack, Stephen Wolbers

February 19, 2003

Outline

- Goals
- Methodology
- Results so far
- Request for feedback
- Plans

Goals

- We (CD and the lab) should understand the utilization, loads, history, and future development of the computers that we use and support.
- There are many aspects to this, and not all of them are clear to us yet.

Goals

- Measure and track all CPU usage of all CPUs in FCC, WH7 and Muon Lab.
 - Farms
 - Servers
 - SMPs, Database machines, other.
- Think about other possible measures
 - Disk
 - Tape
 - Network
 - Power, cooling, floorspace
 - Etc.

Methodology

- Start with what already exists and expand from there.
- Accounting records currently exist for:
 - Farms (CDF, DO, CMS, General)
 - CDF central and CAF
 - DO central systems
 - FNALU
 - MISCOMP
 - KTEV

Methodology

- In addition, information from FBSNG, PBS, etc. are being looked at to see what information they contain.
- Data is in flat files, MySQL db, etc.
- Want to insert into database (ORACLE).
- Currently using MS Access for studies.
- Starting to build reports and plots using Crystal Reports.

Data Structure, Granularity

- Many systems currently provide "standard" set of useful fields such as date/time, node name, cpu time.
- Process/job ownership by group not available from all systems.
- A set of fields for all systems must be defined .
- Processor speed must be factored in in some manner.

Data Volume - Example

- CDF CAF using FBSNG:
 - Process level detail for every every job and section, including start and end times, cpu time, node name and username
 - Data for one day ~ 1 MB
 - Data rolls off ~ 7 days
 - Daily process would be needed to put summary and/or detail data into historical database.
 - It is quite likely that a monthly summary would be sufficient.

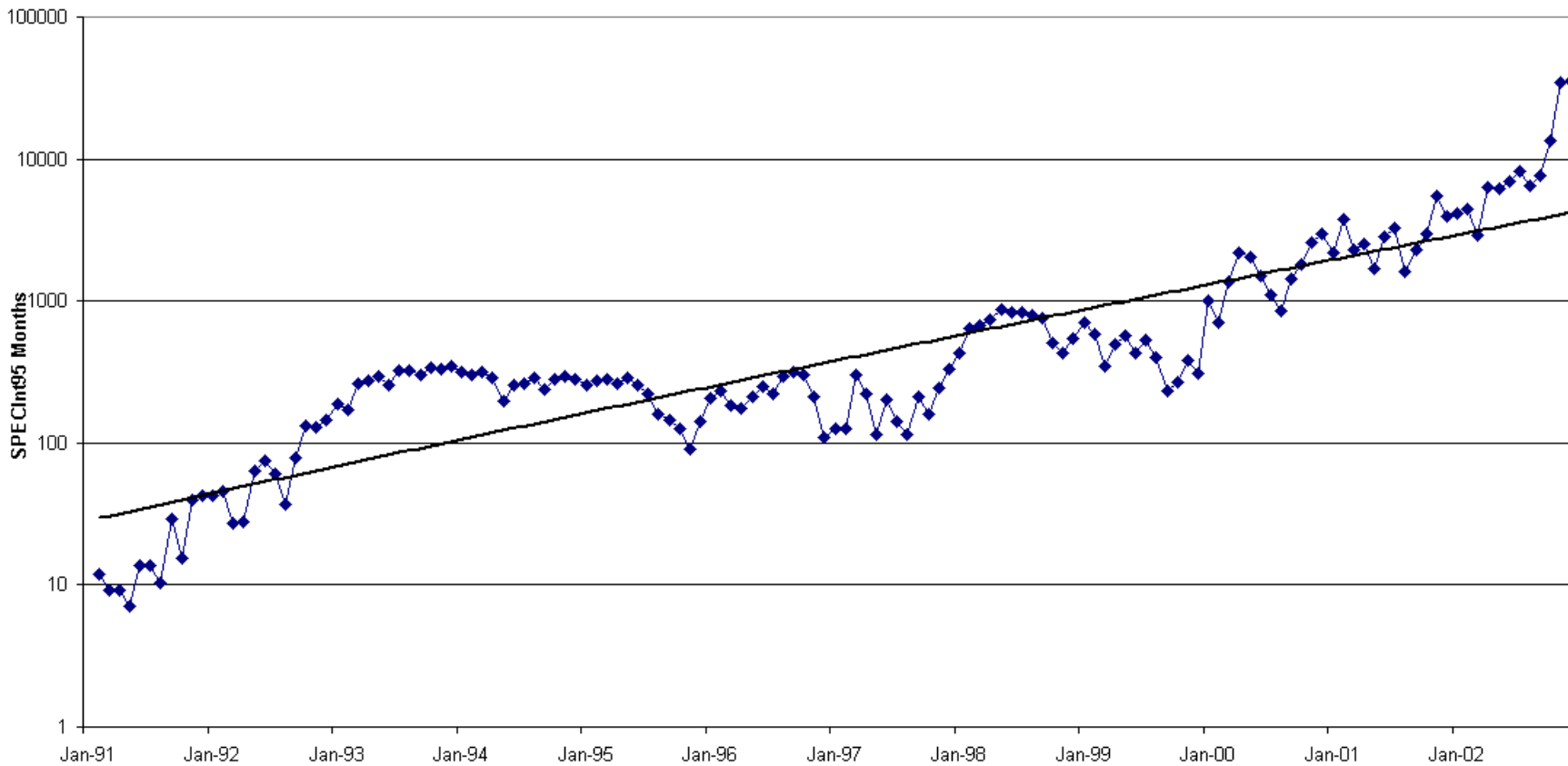
Data Volume Questions

- What are the appropriate summarizations, e.g. cpu time by user, job or group?
- What is the frequency of summarization, daily, weekly or monthly?
- These choices will significantly affect the volume of data preserved in the historical db.
- These choices will also define how we can use this data for future studies and reports.

Results so far

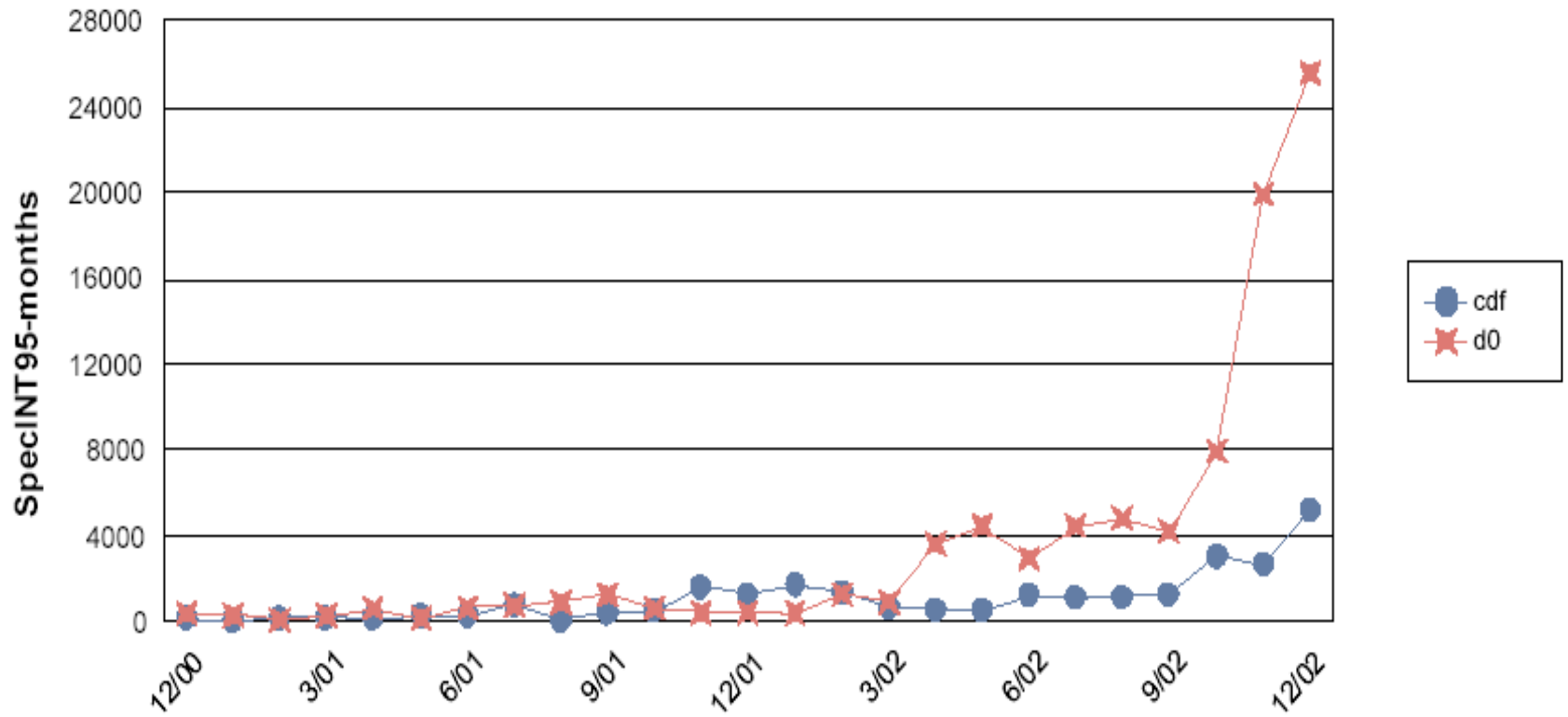
- Look at the farms (we have a 12 year history of data).
- Focus on the past 2 years.
 - Goal is to write a Fermilab TM covering the past 2 years of the farms.
- Data is in database by month and by user.
- Plots have been generated.

FARMS Usage History, 1991-2002



FARMS Usage: 2001-2002

d0,cdf



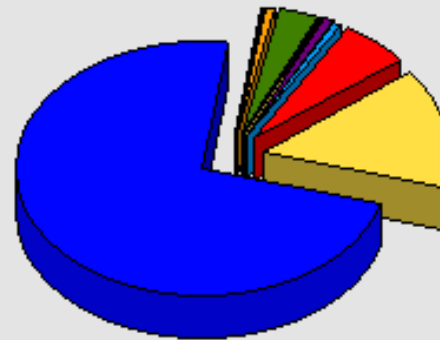
Dec-2002

**SpecINT95-
months**

35121.0

auger	2,151.6
cdf	5,261.0
d0	25,570.6
e791	120.0
e871	245.9
e872	0.0
e898	0.0
ktev	1,150.5
numi	115.1
OTHERS	271.7
sdss	234.8

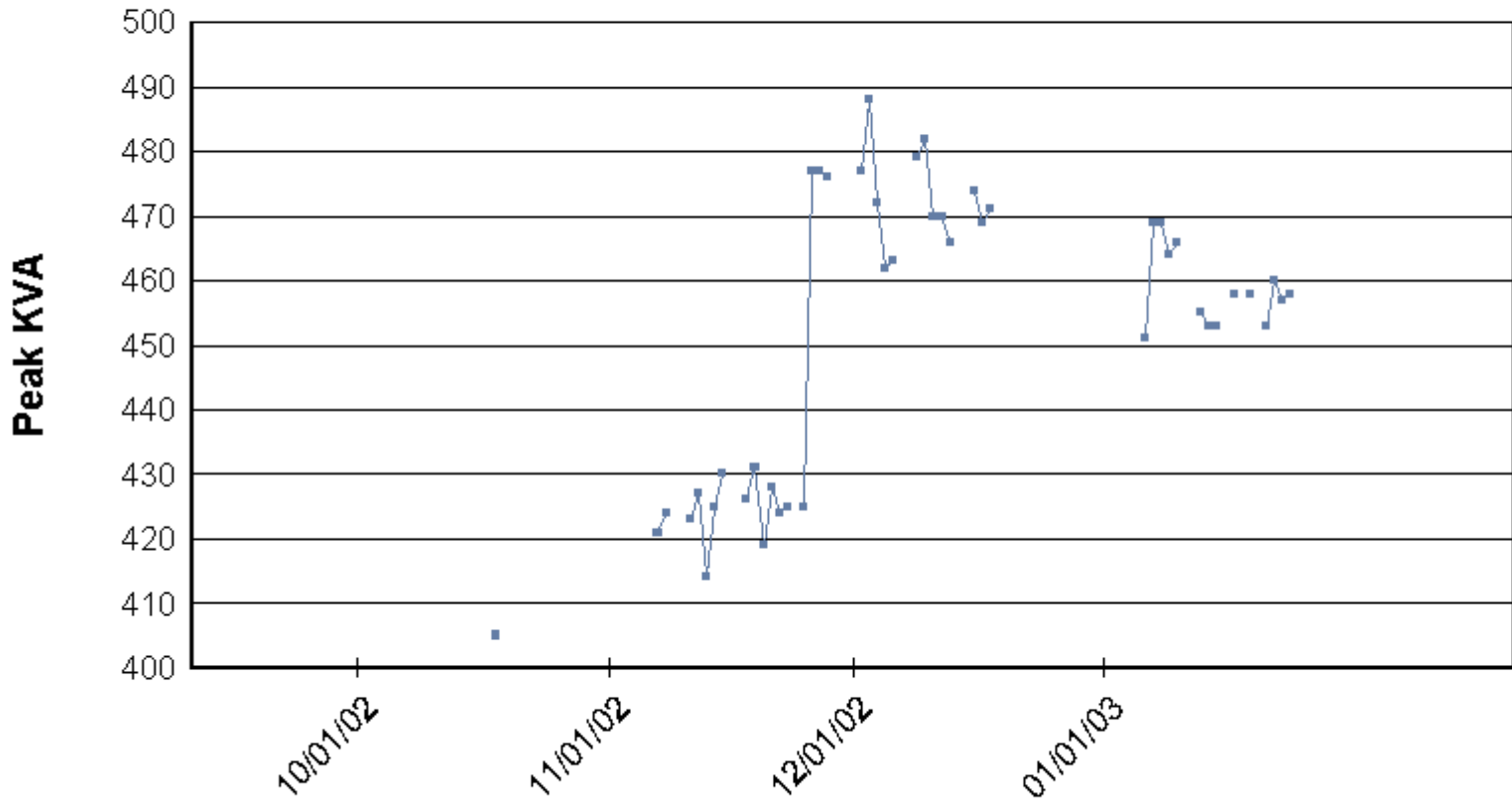
FARMS Usage



■ auger	6.1%
■ cdf	15.0%
■ d0	72.8%
■ e791	0.3%
■ e871	0.7%
■ e872	0.0%
■ e898	0.0%
■ ktev	3.3%
□ numi	0.3%
■ OTHERS	0.8%
■ sdss	0.7%
<hr/>	
Total:	100.0%

UPS Delivered Power

FCC Datacenter



Feedback

- Is this direction reasonable?
- Technology options.
- Systems to monitor.
- Reporting suggestions.
- Web/other access to data.
- Connection to or use of other monitoring tools (fbsng, grid).

Plans

- Pursue farms, cdf, d0, kteV, cms, sdss, lattice qcd accounting.
- Build tools for information collection, storage and report building.
- Create a project.