

More NERSC Data Collect

Cary Whitney

Outline

- Overview
- What's left?

Overview

- Many types of consumers of the data, not just system admins.
- Long data retention time with full resolution.

System Side Data Collect

- Collectd primary
- filebeats for logs
- gmond possible
- LDMS some places
- Intel snap will be investigate
- Elastic beats also being investigated

What's left?

- I fear, a lot.
- Providing data allows more questions to be asked.
 - Which means possibly more data needs to be collected. Goto previous point.
- Results: Better support for all user types.

What's left, really!

- This is a data collect, not a full blown analytics engine.
- All parts are interchangeable.
- **Important point.** This is not the only way to do the data collect. There is not a correct way.

Why present it, what are you really after?

- One. We need to work together about getting vendors to give us, their users, access to the data. This has to be in a usable format. Meaning an API or method that can allow the data to be accessed in a high frequency manner, aka 1-5 seconds or less. (Exporting from a DB is not this type of solution.)

Progress

- Talking with DDN, first suggestion came during HEPiX at BNL.
- Cray/Sonexion data gather. Progress report in 2 weeks at CUG. (Sonexion is Seagate's Lustre product.)
- DDN has a collectd module for GPFS, checking on availability.
- CSCS Lugano helping write gmond collection module for GPFS.
- LDMS gathers data within a Cray system from Sandia.
- NERSC, Oak Ridge, Sandia, Los Alamos and NCSA working together for data collection organization.

Real needs.

- Two. A common data definition so we all know what is meant by clock cycle. Load average. Percent vs actual number. Is a filesystem 90% full bad? May still have 100's TB's left? What is meant by slow? How can we define some of these things as we go? We do not want this to limit progress.

Progress

- Metrics 2.0 Tom talked about this.
- Fabien has started down this road.

Real need, more?

- Three. Visualizations and dashboards. These are probably the toughest and the most desired.
- If we as a community build viz with a common API and definition, we would not be dependent upon a single data collect or a single solution. I could use what you build and you could use things that I build.

Thanks

- Going to set up a monitoring portal.
- Come to the BOF after lunch.