Contribution ID: **38** Type: **not specified**

Monitoring at scale: a needle in the haystack

Thursday 21 April 2016 17:00 (25 minutes)

Many of today's opensource monitoring tools have grown to distributed, horizontally scaling solutions.
When designing a new infrastructure, choosing and configuring the right software stack to analyze and record logs and metrics can
admittedly still be a challenge, but we are no longer restricted to the vertically scaling rrdtool-type timeseries storage.

The real challenge is the amount of data a monitored system can produce, and the difficulty to process it without classifying and tagging it appropriately.
We explain the necessity to attach relevant metadata to monitored events in order to offer a solution to the needle-in-the-haystack problem
that affects large datasets. Through practical ideas and use-cases at CCIN2P3 we underline the capital importance of metadata to leverage
the power to query the underlying indexing backend. Aggregating metrics and querying logs against technical or business-oriented key-value pairs
are a powerful way to answer questions, and provide high-level alerts.

We present the current solution of managing log events at CCIN2P3 as well as the upcoming metric solution.
The primary focus is on the tool stack and experience we gathered during
the last decade.
The current monitoring stack is based on facter (puppet), collectd, riemann, syslog-ng and elasticsearch and has successfully been used in production at CCIN2P3 on its 2 datacenters with roughly 1500 monitored nodes and 12'000 events per second on average.

Length of presentation (minutes, max. 20)

45

Author: WERNLI, Fabien (CCIN2P3)

Presenter: WERNLI, Fabien (CCIN2P3)

Session Classification: Basic IT services

Track Classification: Basic IT Services