Contribution ID: **147** Type: **not specified**

Harvesting Logs and Events Using MetaCentrum Virtualization Services

Thursday, 11 April 2013 14:45 (45 minutes)

Impact

As previously published, we use a common resource pool for running grid and cloud interface in the MetaCentrum. Now we demonstrate that the same principle can be applied to internal services. It represents an important benefit in the case of services with occasionally peaking resource demand. In general the layered infrastructure with a set of internal
and external services running on top of a single virtualized resource pool acts as significant internal optimization.

The described application –harvesting of huge logs –is an important step in developing MetaCentrum's continuous security processes. But it is also a new typical use case intended to show our users –researchers from various areas –how to manage large amounts of data for full-text searching or organizing large data using NGI resources. The full-text or NoSQL engines such as Elastic Search or MongoDB can be straightforwardly instantiated in MetaCentrum environment with dynamically expandable capacity and become a part of the portfolio of standard solutions provided to our users.

URL

http://www.metacentrum.cz/

Summary

The talk describes the design and implementation of MetaCentrum's (Czech NGI's) new security infrastructure service. To implement its everyday procedures, a demand emerged for a central and flexible tool to gather and analyze system logs from hundreds of nodes spread across multiple institutions in the Czech Republic. The selected solution is built on top of existing tools to gather, transfer, store and analyze logs. But we have identified several areas that the current tools do not properly cover. The new service is able to work not only in an automated mode (predefined patterns and alarms) but also in a generic mode. It allows to perform interactive queries to harvest the logs based on actual needs of operators or security officers. The whole storage, indexing and querying infrastructure is operated on top of MetaCentrum virtualization service. The resources are not decicated but allocated on-demand from the NGI resource pool.

Description

Gathering the logs over the open Internet from all our nodes is achieved by an rsyslog facility extended by a custom GSSAPI module. We decided to develop this module because our infrastructure is Kerberos-based and the only supported X.509 authentication and confidentiality services are not suitable for our needs.

From the central syslog server the logs are transferred to
distributed indexing services. We have two types of those. The first is a full text engine intended to index all the data for on-demand queries. The second is a NoSQL database intended to store a selection of the data to process it by predefined queries. For the full-text engine we use Elastic Search, a distributed parallel system which we found to be be really elastic in our environment. In its "resting" configuration (minimal set of preallocated resources) it is running just to receive and index data. It can be expanded on-demand in reasonable time scale to be big enough to process demanding queries in real time.

This expansion relies on MetaCentrum's virtual cluster
facility which is a cloud-like environment controlled by a Torque resource manager. The Elastic Search itself manages discovery of new worker nodes and transparent data (indices) replication and query distribution. An

advanced feature of the virtual cluster service - virtual networking - is used to encapsulate communication between cluster nodes to overcome lack of security in the current implementation. The role of NoSQL database, implemented by MongoDB, will be shown on a focused hunt for signs of bad guys trying to access user accounts on all of our worker nodes. As MetaCentrum worker nodes are located in the open Internet, it is important to evaluate and react to suspicious patterns in user ssh authentication.

Co-authors: SITERA, Jiri (CESNET); BODO, Radoslav (CESNET); SUSTR, Zdenek (CESNET)

Presenter: BODO, Radoslav (CESNET)

Session Classification: Operational Services