Geophysical Research Abstracts
Vol. 17, EGU2015-14843-1, 2015
EGU General Assembly 2015

Monitoring operational data production applying Big Data tooling

Wim Som de Cerff (1), Hotze de Jong (2), Roy van den Berg (2), Jeroen Bos (2), Rijk Oosterhoff (3), Henk Jan Klein Ikkink (4), Femke Haga (2), Tom Elsten (3), Hans Verhoef (1), Michal Koutek (1), and John van de Vegte (1)

(1) Royal Netherlands Meteorological Institute,R&D Observations and Data Technology, De Bilt, Netherlands (wim.som.de.cerff@knmi.nl), (2) TriOpSys, (3) NSpyre, (4) Centric

Within the KNMI Deltaplan programme for improving the KNMI operational infrastructure an new fully automated system for monitoring the KNMI operational data production systems is being developed: PRISMA (PRocessflow Infrastructure Surveillance and Monitoring Application).

Currently the KNMI operational (24/7) production systems consist of over 60 applications, running on different hardware systems and platforms. They are interlinked for the production of numerous data products, which are delivered to internal and external customers. All applications are individually monitored by different applications, complicating root cause and impact analysis. Also, the underlying hardware and network is monitored separately using Zabbix.

Goal of the new system is to enable production chain monitoring, which enables root cause analysis (what is the root cause of the disruption) and impact analysis (what other products will be effected). The PRISMA system will make it possible to dispose all the existing monitoring applications, providing one interface for monitoring the data production.

For modeling the production chain, the Neo4j Graph database is used to store and query the model. The model can be edited through the PRISMA web interface, but is mainly automatically provided by the applications and systems which are to be monitored. The graph enables us to do root case and impact analysis. The graph can be visualized in the PRISMA web interface on different levels.

Each ‘monitored object’ in the model will have a status (OK, error, warning, unknown). This status is derived by combing all log information available. For collecting and querying the log information Splunk is used.

The system is developed using Scrum, by a multi-disciplinary team consisting of analysts, developers, a tester and interaction designer.

In the presentation we will focus on the lessons learned working with the ‘Big data’ tooling Splunk and Neo4J.