Geophysical Research Abstracts
Vol. 18, EGU2016-5263, 2016
EGU General Assembly 2016

Exploring the "what if?" in geology through a RESTful open-source framework for cloud-based simulation and analysis

Jens Klump and Jess Robertson
CSIRO, Mineral Resources Flagship, Kensington, Australia (jens.klump@csiro.au)

The spatial and temporal extent of geological phenomena makes experiments in geology difficult to conduct, if not entirely impossible and collection of data is laborious and expensive – so expensive that most of the time we cannot test a hypothesis. The aim, in many cases, is to gather enough data to build a predictive geological model. Even in a mine, where data are abundant, a model remains incomplete because the information at the level of a blasting block is two orders of magnitude larger than the sample from a drill core, and we have to take measurement errors into account. So, what confidence can we have in a model based on sparse data, uncertainties and measurement error?

Our framework consist of two layers: (a) a ground-truth layer that contains geological models, which can be statistically based on historical operations data, and (b) a network of RESTful synthetic sensor microservices which can query the ground-truth for underlying properties and produce a simulated measurement to a control layer, which could be a database or LIMS, a machine learner or a companies' existing data infrastructure. Ground truth data are generated by an implicit geological model which serves as a host for nested models of geological processes as smaller scales.

Our two layers are implemented using Flask and Gunicorn, which are open source Python web application framework and server, the PyData stack (numpy, scipy etc) and Rabbit MQ (an open-source queuing library). Sensor data is encoded using a JSON-LD version of the SensorML and Observations and Measurements standards. Containerisation of the synthetic sensors using Docker and CoreOS allows rapid and scalable deployment of large numbers of sensors, as well as sensor discovery to form a self-organized dynamic network of sensors.

Real-time simulation of data sources can be used to investigate crucial questions such as the potential information gain from future sensing capabilities, or from new sampling strategies, or the combination of both, and it enables us to test many "what if?" questions, both in geology and in data engineering. What would we be able to see if we could obtain data at higher resolution? How would real-time data analysis change sampling strategies? Does our data infrastructure handle many new real-time data streams? What feature engineering can be deducted for machine learning approaches? By providing a 'data sandbox' able to scale to realistic geological scenarios we hope to start answering some of these questions. Faults happen in real world networks. Future work will investigate the effect of failure on dynamic sensor networks and the impact on the predictive capability of machine learning algorithms.