Enabling data science against legacy software application data

Sebastian Grebe
Cegal Ltd

Summary

"Over the past 30 years software companies have provided the geoscientist with a huge suite of software applications to support their subsurface workflows. There are specialised applications for geophysics, rock physics, geology, log interpretation, seismic interpretation, modelling, simulation, history matching and more, something for every niche of the sub-surface domain. Typically all of these applications are desktop based and software updates are provided periodically by the software vendor. This means that getting access to a feature update you have requested or the latest algorithms provided by the vendor usually requires some sort of wait by the user, which in turn hampers creativity and productivity.

Although Python has been around for 30 years – its only in relatively recent years that the subsurface community has started to realise the potential Python has for allowing geoscientist to break out from traditional application silos and express their creativity with immediate effect.

Python is comparatively easy to learn and the possibilities for Python in geoscience are endless. With a stable data connection to your traditional application data from within a data science working environment we are able to give the geoscientist the power to use the latest AI/ML algorithms, analytical techniques or in house python developments as soon as they are ready. Parallel access to existing software application data allows the geoscientist or data scientist to work with these data in new and innovative ways to solve problems that may not be possible in a desk-top application and continue to be creative outside the release cycle of the software vendor.

Theory / Method / Workflow

Cegal will present a new workflow to allow users to easily and efficiently connect into Petrel for dynamic access to Petrel data from an external Python installation or Integrated Data Environment (IDE) on platforms like Anaconda, to enable a full interactive IDE experience against live Petrel project data.

This workflow provides many benefits such as updating Petrel objects live when running Python code from an external Python environment for immediate data visualisation and QC of integrated results in Petrel. The workflow allows the geoscientist to take full advantage their desired data visualisations libraries and use state-of-the-art ML libraries such as TensorFlow – directly against live project

data. Furthermore, the workflow opens up the option to distribute Python applications developed by specialized inhouse data science teams behind a UI teams to a wider audience enabling them to run advanced data science algorithms against Petrel data without having to necessarily learn the details of Python coding thus removing the barrier to entry for less python enabled users."