

Atos Breakout session  
12 09 2019 – 10.00

Large Scale Data Analysis  
using JupyterLab and Pandas

Geert Fieremans



# Large Scale Data Analysis using JupyterLab and Pandas

With Circuit deployed for some very large customers, the instrumentation data collected for these deployments grows drastically day to day.

## ***How do you rapidly analyze millions of unstructured data records in seconds?***

Circuit R&D is leveraging JupyterLab, pandas, numpy, scikit-learn, and matplotlib to rapidly diagnose issues within very large datasets using the toolkits available in Google Cloud DataLab.

- ▶ External Benefits:
- ▶ This paper will provide – in the form of a use case on actual Circuit QoS data – insights into practical usage of best in class data science techniques & tools for real world problem solving.
- ▶ Strategic Alignment:
- ▶ This paper leverages the google strategic alliance showcasing big data, data science, and machine learning.
- ▶ Internal benefits:
- ▶ This paper will show the benefits of the a fore mentioned tools, the ease of their use for quick development purposes and the ability to detect, prevent and diagnose complex system issues.
- ▶ Technology Leadership:
- ▶ This paper leverages the best in class data science toolkits available today.
- ▶ Wow-factor:
- ▶ Showcase the analysis of millions of records in seconds.

# Large Scale Data Analysis using JupyterLab and Pandas

---

## Presenters:

**Geert Fieremans**



**Manolo Caracuel**

