PARIS 2017

12-15 June 2017 | Paris expo Porte de Versailles

THE SOCIETY OF HPC PROFESSIONALS

TITLE:
Machine Learning in Oil & Gas HPC – Opportunities and Implications

Abstract:
With three oil and gas machine learning events taking place around the EAGE conference, it is fair to say that the buzz associated with the application of artificial intelligence is right at the top of the hype curve. But are there implications for High Performance Computing? In this talk we will look at the implications on HPC system design for machine learning technologies and how to optimise performance. We will then look at how Machine Learning techniques can be applied to optimise the execution of oil and gas technical workflows (e.g. automated fault interpretation) or in a heterogeneous hybrid-cloud environment.

Presenter's Bio:
David Holmes, Chief Technology Officer – Energy, Dell EMC

Machine Learning in Oil & Gas HPC Opportunities and Implications

DELL EMC

Le Grand Hack

TOTAL

DELL EMC

Agile*

The Optimization Challenge

- There are **many workflows** being executed in **parallel**
- **Infrastructure** might vary: might even be dynamically provisioned
- **QoS metrics** vary depending on the situation:
- Execution depends on the **data and algorithm**
- We aim to achieve **global optimizations**

Big combinatorial problem

Multitenancy

With different Quality of Service (QoS) requirements

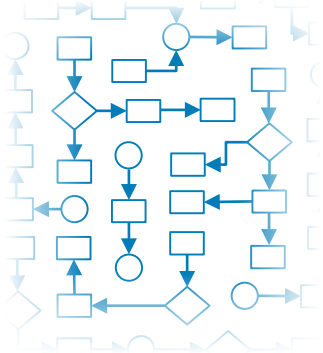
X

Big Data

Of different types, with distinct characteristics

X

Many workflows



With different programs and parameters

X

Many machines

With different settings

Expected benefits

- For customers
  - Suitable optimization for their data and goals
- For infrastructure and service providers
  - What is the optimal infrastructure for your clients?
    - With specific SLAs, different profiles and custom requirements
  - How to use the resources more efficiently and save money?
    - Looking at the big picture, with many jobs sharing the same environment
- Automated optimized scalability and elasticity
  - Dynamically set the location, number of nodes/cores, storage and data orientation per job without compromising SLA

Run more, run better

The proposed approach

- Continuous monitoring and collection of telemetry and workflow provenance data
  - Capture different circumstances varying input datasets, parameters and infrastructure
  - Real-time distributed data gathering
- Prediction models for QoS metrics
  - Take snapshots to represent the state of the environment based on the collected data
  - Based on the snapshots, train deep learning models
  - Continuous evolution of the models
- Optimization of execution settings
  - Explore changes of parameters and infrastructure settings
  - Learn policies using reinforcement learning
- Management and orchestration based on optimization policies

Conclusions

- Machine Learning can be used to deliver accurate prediction models
- Current prototype focused on workflow execution
- Possible application to any industry with complex workflows
- Heterogenous HPC incorporating hybrid HPC seems an inevitable direction of travel


DELL EMC