

Distributed Data Analytics

Fast Machine Learning with ProActive and Docker

Key Benefits

- Automate your distribution of data analytic applications and workload
- Define your environment with Docker containers
- Fully utilize your infrastructure with the ProActive scheduler



Overview

Data Analytic workload is computationally intensive and comes with different stepping stones, several platforms, libraries, packages and dependencies which all need to be maintained and configured. Improving performance of data processing in various contexts is a challenging problem. Associating the ProActive Workflows & Scheduling with its R connector allows everyone to **accelerate computation** using R workload parallelization among hybrid clouds while **lowering maintenance costs** using Docker containers for R packages and configuration. This case study describes a typical machine learning application that is using the R platform: predicting the monthly returns of the S&P500 index, using Support Vector Machine (SVM) classification in a financial context.

Case study: Forecast the monthly returns of the S&P500

We forecast the monthly returns of the S&P500 with a SVM model, using historical data of S&P500 from 1900 to 2012 as features.

Reliably distribute data science application and workload

Especially in a financial context, getting faster information will benefit: parallelizing data processing on computing devices, possibly on a hybrid cloud (CPU-GPU), is a great way to achieve that. With its powerful resource management capabilities, ProActive Workflows & Scheduling allows federation of a large set of heterogeneous resources, including private and public infrastructures, to easily setup a high performance platform. Using ProActive workflows expressivity, we are able to simply and clearly define how data and task executions are orchestrated, maximizing the throughput and optimizing the resource usage. Through an intuitive GUI, users can concentrate on writing data science applications, without managing the load balancing.

Use Docker containers to reduce maintenance costs

Working with heterogeneous environments introduces complexity and costs to manage and take into account particularities of each single environment. ProActive Workflows & Scheduling comes with great mechanisms to easily handle and take benefits of this heterogeneity to increase computational platform capabilities. With Docker, containerizing R packages and configuration allows to rely on a customized environment for the R platform. That prevents compatibility issues when distributing workload among a hybrid cloud infrastructure.

Learn More



Download



Try Online

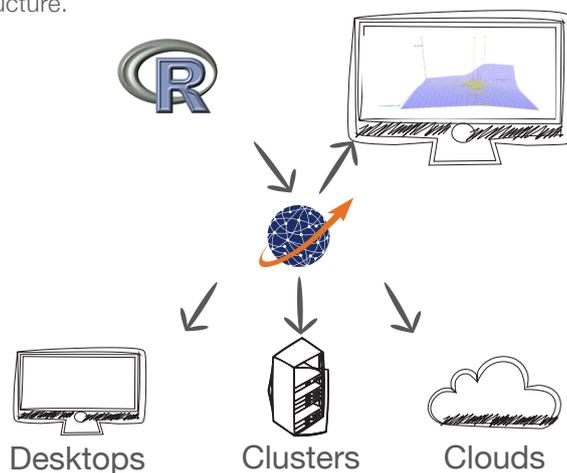


Get Demo

Register and get free support during a 10 days period

activeeon.com/parallel-scientific-toolbox

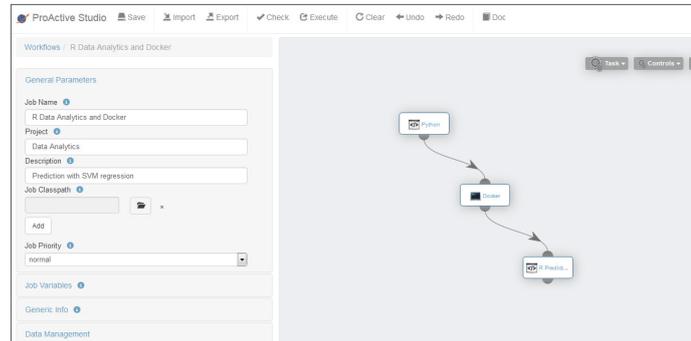
Read technical details on blog.activeeon.com



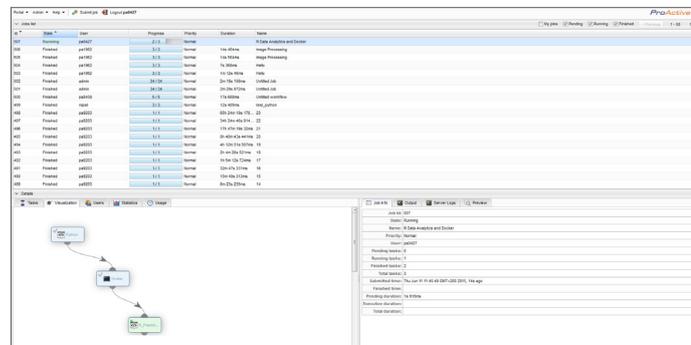
Automate your process in 4 steps

Using ProActive Studio, describe your workflow to simplify and accelerate the way you get the result you need. Integrate and automate workflow executions with connectors and flexible API. Get insights on execution with ProActive Job Scheduler and resource monitoring.

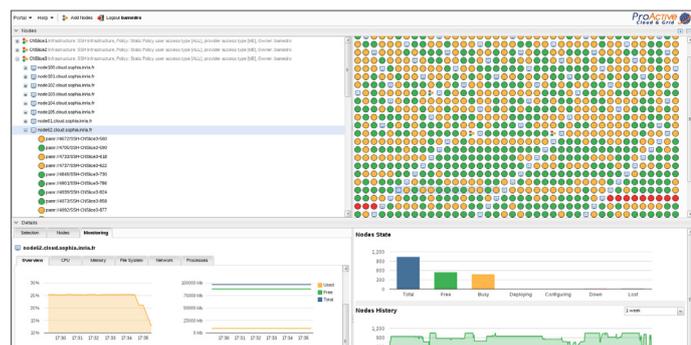
Step 1: Design the workflow with ProActive Studio



Step 2: Launch it on ProActive Scheduler



Step 3: See and monitor execution on ProActive Resource Manager



Contacts

ActiveEon

2000 route des Lucioles
Les Algorithmes - Pythagore B
06560 Sophia Antipolis - FRANCE

Tel. +33 (0) 9 88 77 76 60
Fax +33 (0) 9 88 77 76 61

contact@activeeon.com
www.activeeon.com

Step 4: Automatically get the results where you want it

