



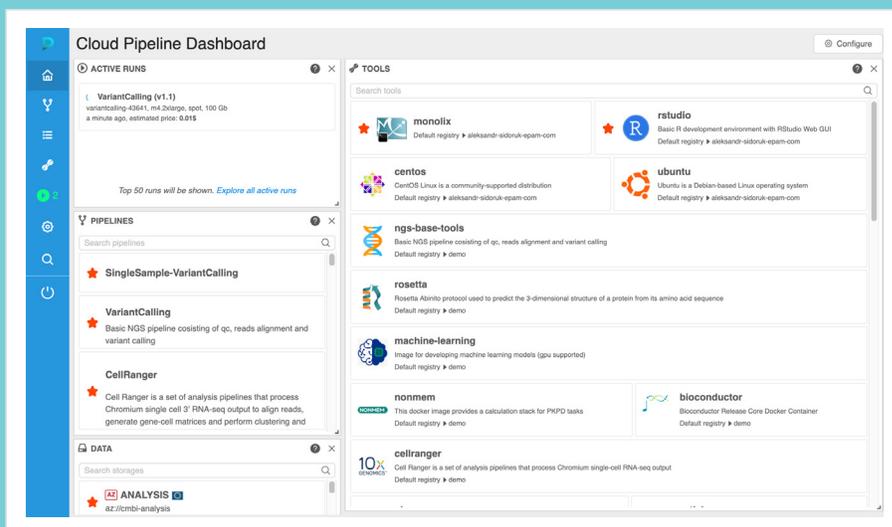
EPAM Cloud Pipeline: Democratizing the Cloud for Scientists

Discover EPAM and Microsoft's Integrated Solutions for Life Sciences Companies

Pharma and biotech companies need easy-to-use, self-service solutions that address the unique needs of their scientists, especially when it comes to the cloud. While the cloud is an excellent solution in terms of scalability, security and performance, it often requires specialized knowledge and IT skills that not all scientists possess. EPAM Cloud Pipeline is a proven solution that helps scientists take full advantage of the power of cloud computing.

WHAT IS EPAM CLOUD PIPELINE?

EPAM Cloud Pipeline is an easy-to-use, self-service cloud platform that enables scientists to perform tasks without the need for specialized cloud IT skills. These tasks include building and running customized scripts and workflows required to support genomics analysis, modeling and simulation, as well as machine learning.



As an open source platform, the solution allows rapid deployment, maximum flexibility and adaptability to include new tools and algorithms into the environment. The intuitive, web-based front-end provides access to highly scalable compute and storage resources that are necessary to support parallel processing and accelerate drug discovery research. EPAM Cloud Pipeline enables scientists to flexibly build and execute their own pipelines/models, while preserving the highest levels of safety for the data and application.

EPAM Cloud Pipeline operates in an organization's virtual private cloud. It utilizes a scalable cluster orchestrating mechanism based on Docker/Kubernetes that can be operated via a simple web UI. It can be integrated with both internal systems and external cloud applications. Best of all, it is a self-service solution that allows users to build their own compute stacks and pipelines from their tools of choice.

LIFE SCIENCES BUSINESS CHALLENGES DRIVING DEMAND FOR EPAM CLOUD PIPELINE



The volume of genomics data is expected to double every seven months, and by 2025, the capacity needed for genomics data is estimated to reach 40 exabytes¹



Modeling and simulation (M&S), also known as biosimulation or model-informed drug discovery and development (MID3), is growing as it delivers significant business, scientific and clinical value to firms that fully utilize it as an integral part of their drug development strategy



Storing and processing this volume of data on-premise isn't suitable or timely, as internal requests for data processing take a long time to complete



On-premise data centers can't increase storage or processing power at the needed pace, and storage costs are becoming very high



There's a lack of scalability as internal resources for genomics analysis are limited and expensive

¹ Stephens ZD et al. (2015) Big Data: Astronomical or Genomical? PLoS Biol 13(7): e1002195. <https://doi.org/10.1371/journal.pbio.1002195>



How EPAM Cloud Pipeline Can Help

EPAM Cloud Pipeline helps scientists handle constantly changing demands in analytics and modeling. It enables them to develop, reuse, share and run processing pipelines with a few clicks. The solution provides this ease of use while still allowing enterprise IT to define and control access and security policies.

Furthermore, new pipelines and tools can be added quickly at any time. By having complete control, flexibility and direct integration between Azure and on-premise solutions, scientists can optimize for any type of study.

EPAM CLOUD PIPELINE'S KEY CAPABILITIES

VIRTUAL PRIVATE CLOUD INFRASTRUCTURE

- Data storage (Azure BLOB storage, Azure files)
- Scalable computational cluster
- Networking: a hub of clusters to effectively orchestrate computing task across multiple nodes
- Security: user authentication, access control, encryption

BATCH ANALYSIS

- Workflows for bioinformatics and modeling & simulation analysis
- Experimental research reporting and reproducibility of scientific results

INTERACTIVE ANALYSIS

- Personal research environment for interactive analysis using the most recent versions development/analysis tools (RStudio, Jupyter Notebook, MATLAB, etc.)
- Fully utilizing the computational capacity of Azure for massive simulation tasks

DATA SHARING

- Managed data storage
- Shared URLs can be used to browse and upload/download data from the web-based interface or Command Line interface can be used for larger data sets

GLOBAL SEARCH

ElasticSearch to index all managed content, including:

- Projects and their attributes
- Pipelines (including code, configurations, attributes)
- Tools (docker images)
- Storage (object and file storage metadata, including objects/files labels)

ABOUT EPAM'S LIFE SCIENCES EXPERTISE

EPAM and Microsoft Genomics offer industry-leading cloud-enabled solutions for life sciences organizations. EPAM's scientific informatics and enterprise technology solutions improve outcomes through every stage of the drug development lifecycle, speeding research, discovery and time-to-market while enhancing collaboration, knowledge management and operational excellence.

8 **OUT OF THE TOP** 10

**PHARMACEUTICAL
COMPANIES WORK
WITH EPAM**

1,900+

**LIFE SCIENCES &
HEALTHCARE PROFESSIONALS
WORK FOR EPAM**

MICROSOFT GENOMICS BRINGS THE POWER OF AZURE TO GENOMIC COMPUTATION BY PROVIDING THE ON-DEMAND PERFORMANCE, SECURITY AND SCALABILITY OF A WORLD-CLASS SUPERCOMPUTING CENTER.



TRY EPAM CLOUD PIPELINE TODAY

Want to try EPAM Cloud Pipeline for yourself? Download it now at github.com/epam/cloud-pipeline or contact Sales@EPAM.com to schedule a demo with one of our experts.