

a transformation in research workflow scheduling at genomics scale

e[hive]

Eagle Genomics has integrated highly efficient open source code (Sanger/EBI) in a new fully tested architecture. The product uses the Docker container technology to deliver market-leading efficiencies in parallel computation pipelines on any distributed, local or cloud infrastructure.

HOW IT WORKS

e(hive) coordinates “runnables” – validated processing code wrapped in an appropriate container. Any piece of analytics code can be wrapped into a validated runnable component. Runnables from an ever-growing library can be combined to create complicated workflows to support analysis in genomics, metagenomics, data visualisation etc. The platform supports sharing, versioning and a wealth of monitoring metadata necessary for experimental reproducibility (e.g. execution time, errors, user information and performance).



BUSINESS BENEFITS



Promotes collaboration

Supports collaboration and sharing of advanced analytics workflows – libraries of components and dockerised images of fully functional analytical virtual machines can be created, stored and used on demand.



Reduces time to discovery

Unlimited scalability coupled with data discovery and interactive reporting reduces “time to insights” by two order of magnitude.



Drastically reduced infrastructure costs

Autonomously managed analytics resource that ensures minimal computing costs without any infrastructure maintenance or installation investment. e(hive) automate resource instantiation and termination, even to the point where the application itself autonomously manages this process (no manual intervention required).



Streamline integration of any analytics applications

Efficient integration of analytics or knowledge management software applications: These include genomics, metagenomics pipelines, machine learning and modelling, bioinformatics, cheminformatics, visualisation and reporting.

MODULE FEATURES

Real-time monitoring capabilities » Computing resource, error tracking and resource allocation is tracked using Kibana dashboards. Running, completed jobs and any errors can all be quickly identified.

Uses docker images » Our propriety Docker Swarm Orchestrator engine builds on top of existing proven Docker technologies, such as Docker Machine and Docker Swarm to add further functionality such as elastic cluster scalability.

Enterprise development of open source code » based on high quality open source code, internal or external teams can develop and test re-usable workflow units (runnables) on GitHub. All dependencies and code is automatically tested and deployed periodically.

Modern architecture » secure, scalable resource management compatible with a variety of data aware enterprise environments.

A USE CASE - UNILEVER, THE SMART WORKFLOW

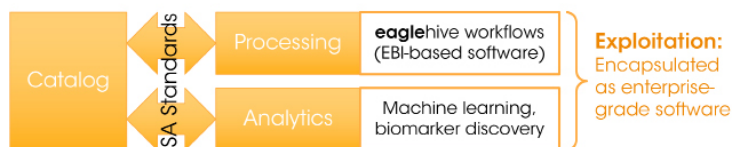
Eagle Genomics applied its bioinformatics expertise and biological experience to analyse and understand the existing Unilever approach. The solution featured a cloud-based instance of **e(catalog)** with **e(hive)** to provide data catalogue and orchestration, fully compliant with its security policies and integrated with Unilever's ELN. **e(catalog)** and **e(hive)** take full advantage of elastic cloud capacity to provide the ideal, scalable environment to deploy knowledge management and analytics to a global workforce.

"Unilever's digital data program now processes genetic sequences twenty times faster—without incurring higher compute costs. In addition, its robust architecture supports ten times as many scientists, all working simultaneously."

Pete Keeley, eScience Technical Lead - R&D IT at Unilever

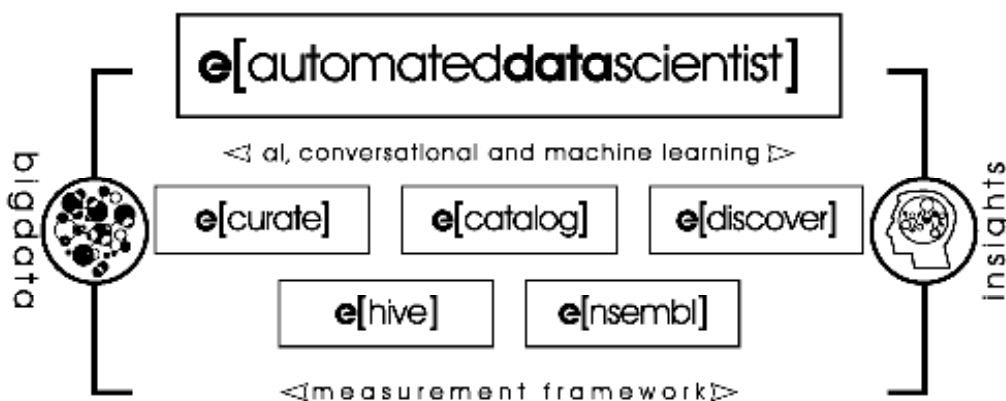
FULLY INTEGRATED WITH e(catalog)

As newer and larger datasets become available, high throughput scalable and seamless data processing and analysis is a key requirement. To address this issue, we productized eaglehive and fully integrated it with **e(catalog)**. The net result is that standardization of custom workflows should be easily accessible and usable by scientists.



OUR SMART DATA MANAGEMENT PLATFORM

e(hive) is part of our proprietary software suite that bridges the entire process from the medical data through to insight.



"This conversation between the scientist and the data sets is the next wave of innovation that we need deployed to our R&D teams, so we can quickly and systematically find and validate new compounds in this Precision Medicine era. Eagle is very much focused on solving this problem and we fully support their product plans."

Mathew Woodwork, Director of Bioinformatics, MedImmune