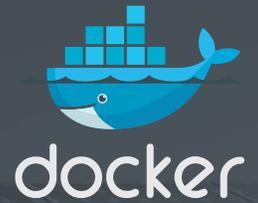# The US Dept. of Energy Joint Genome Institute Uses Docker to Deliver Better Science

To achieve their mission DOE JGI chose Docker to modernize the way applications were developed and managed

BACKGROUND

The mission of the US Department of Energy Joint Genome Institute [DOE JGI] is to advance genomics in support of clean energy generation and environmental characterization and cleanup. Key mission areas are bioenergy, carbon cycle and biogeochemistry.

CHALLENGES

DOE JGI is at the forefront of large scale sequence based science, responsible for the quality of sequencing data as it is generated by distributed production facilities. With lots of assemblers and many previous projects that are publication based, they faced challenges with data sequencing being open to subjectivity and bias. They needed to address the lack of standardization in the process of continually assembling and sequencing a high volume of genome data. The quality and performance of the assemblers made it difficult to compare results. DOE JGI was focused on addressing the following:

- Ensure that the data produced and released to the community was high quality and meeting specific standards
- Improve efficiency and reliability of application development process to protect the quality of the research
- Simplify and reduce the time to set up the traditional assembly

SOLUTION

DOE JGI adopted Docker and was able to immediately improve the assembler project by running benchmarks on assemblers. Now assemblers, crowd sourced from the bioinformatics community, can be objectively evaluated. By using Docker, all genome assemblers and associated pipelines are built within a Docker image and then hosted on Docker Hub.

The benchmarking pipeline is able to pull the image and run it against an array of reference data sets at any time. The produced assembly can be evaluated against the reference sequence using Quast, a quality assessment tool for genome assemblies. The assembly metrics and results are then posted on the site. By simplifying technology and automating processes, researchers have more time for science.

With Docker all genome assemblers and associated pipelines are built within a Docker image and then hosted on Docker Hub. By simplifying technology and automating processes, researchers have more time for science.

With Docker DOE JGI is achieving quality results:

- Docker provided a standardized pipeline with consistent APIs
- Catalog of assemblers enabling objective comparison of the tools and results
- Researchers have data driven conversation and easily share assemblers, the results and the data

The ultimate benefit – better science.

ABOUT DOCKER

Docker is the leading software container platform. Developers use Docker to eliminate "works on my machine" problems when collaborating on code with co-workers. Operators use Docker to run and manage apps side-by-side in isolated containers to get better compute density. Organizations use Docker to build agile software delivery pipelines to ship new features faster, more securely and with confidence. www.docker.com/government

www.docker.com