Parallel Works: Accelerating Big Data and ML Workflows
Bridging the gap between complex simulation and actionable insights

Because time is money.

The **Data Works** platform from Parallel Works offers the world's fastest workflow system for machine learning and big data applications. For data scientists and data engineers, it offers an efficient, flexible, and robust data workflow scaling infrastructure to test new algorithms, deploy data pipelines, and speed time to insight.

When to use Data Works.

Model Development & Testing

Quickly shape, measure, tune, and iterate on data workflows ranging from simple structures with ultra-high-transaction rates to complex multi-language and multi-stage flow graphs.

Pipeline Deployment

Package and deploy your pipeline for production with a simple UI to run in the cloud or on in-house resources. Automatically scale up based on the needs of your job to rapidly process the workload.

Benefits across the enterprise.

Rapid Insights. Speed time to insight with Parallel Works rapid workflow execution. For massive datasets that are not scaling effectively, Data Works provides a highly scalable, robust, and low-latency workflow model that reduces computing bottlenecks and delivers insights fast.

Increase Productivity. Easily tune and test new models quickly using the Interactive Data Science Lab, a development interface for data scientists to collaborate on model code. Quickly push pipelines to the cloud for deployment at scale with the Parallel Works workflow engine.

Optimize Compute Spend. For micro-tasks that need to run at mega-scale, the Parallel Works workflow system delivers rapid speed and low latency, making optimal use of computing resources and delivering the highest return on cloud compute investment. It is robust and fault-tolerant, reducing the risk of frustrating debugging and workload failures.

Why Data Works?

It's *fast.*
Latest benchmark clocked at 1.5 billion tasks / sec using 100 microsecond tasks on 500,000 cores

It's flexible.
Tasks can be expressed in any language through embedded interpreters and run on any computing infrastructure

It's open.
Based on an open source technology from Argonne National Labs developed with support from the DOE and NSF
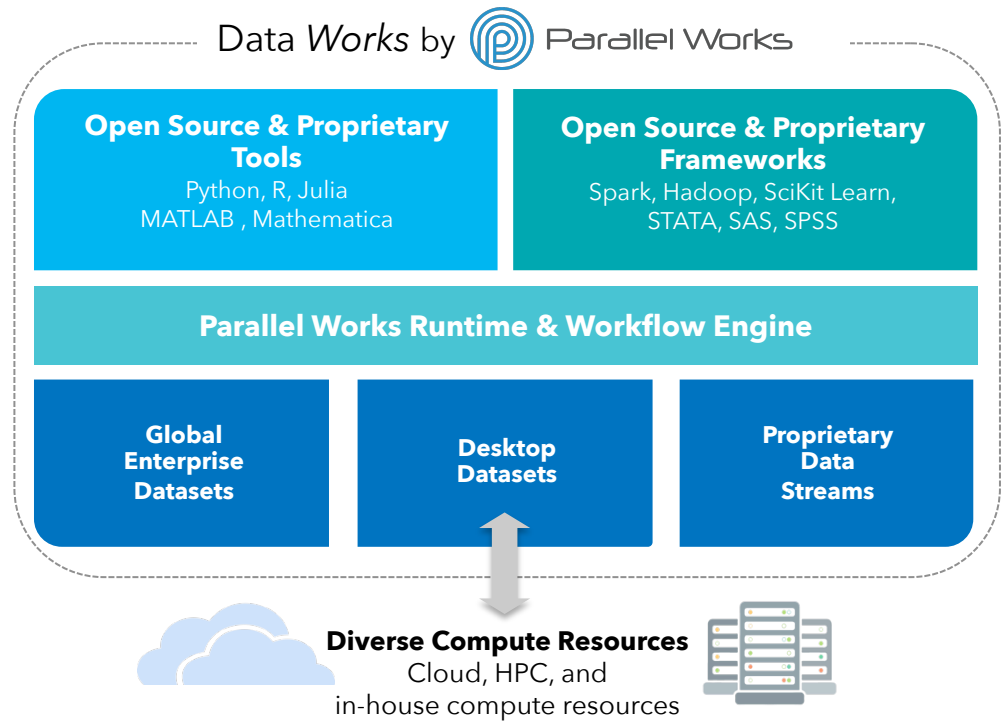
Parallel Works

Parallel Works: Accelerating Big Data and ML Workflows
Bridging the gap between complex simulation and actionable insights

Under the hood.

Data *Works* by Parallel Works

Open Source & Proprietary Tools
Python, R, Julia
MATLAB , Mathematica

Open Source & Proprietary Frameworks
Spark, Hadoop, SciKit Learn,
STATA, SAS, SPSS

Parallel Works Runtime & Workflow Engine

Global Enterprise Datasets

Desktop Datasets

Proprietary Data Streams

Diverse Compute Resources
Cloud, HPC, and
in-house compute resources

Features.

Low Latency Workflow. Data Works specializes in running massive volumes of small tasks for complex ML, AI, and Big Data pipelines. Its core engine can drive 100 microsecond tasks to 500,000 cores at 1.5 billion tasks/sec

Fault Tolerant. The system automatically scales to meet the needs of the pipeline and distributes work across resources, seamlessly orchestrating data flow

Collaborative Development. A workflow studio to collaboratively build, measure and tune data workflows ranging from simple structures with ultra-high-transaction rates to complex multi-language and multi-stage flow graphs

Flexible Toolkit. Use any of the language tools in the Data Science arsenal including Spark, Python, R, Java, and the tools of the Apache Hadoop ecosystem

Elastic Resources. Seamlessly connect to the cloud or run on in-house computing resources. The system automatically scales to meet the needs of the job, seamlessly orchestrating data flow. It provides lightning-fast resource allocation to responsively adapt to changes in demand

Mature Technology. Data Works is based on Swift, an advanced computing technology from the University of Chicago and Argonne National Laboratory. Parallel Works leverages this powerful model to deliver fast workflow tools for data scientists

Parallel Works

Matthew Shaxted Co-Founder
shaxted@parallelworks.com
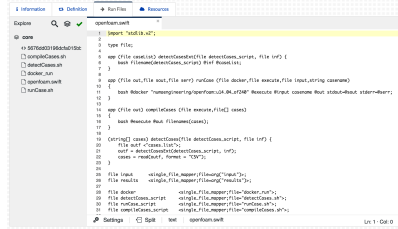
Creating and running a workflow in Parallel Works

CREATE WORKFLOW

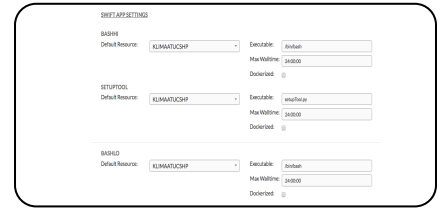**1) Select tools and bring in to development environment**

2) Write short script & bring run files into Parallel Works platform
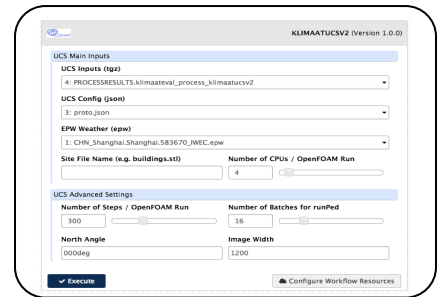
```
foreach case, i in cases {
   runCase(docker, script,
           input, case);
}
```
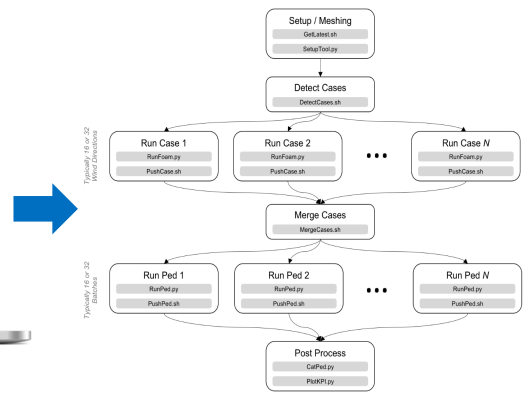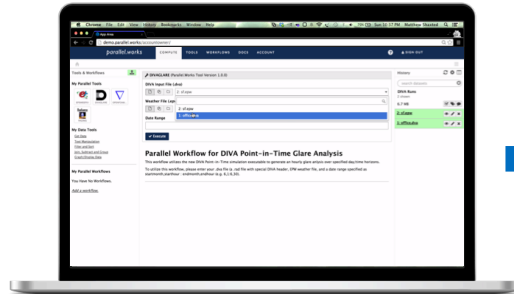

3) Select resource for each step of pipeline

4) Create job execution UI

RUN WORKFLOW

1) Link dataset input sources
2) Select desired workflow and inputs via customizable UI
3) Launch workflow
4) Monitor and track progress

VISUALIZE RESULTS

Once study is complete, view results in interactive web viewer, or download results to desktop for further analysis

Parallel Works