

AI4CI - Towards an Open Operations Data Community



Aakanksha Duggal
Data Scientist



Oindrilla Chatterjee
Senior Data Scientist



Boston, MA, USA



Emerging Tech,
Office of the CTO

Index

- Open Source Operations Data
 - What is it?
 - Where is it coming from?
 - Why is it important?
- What is AIOps?
- Project AI4CI and its tools
- Demo
- Resources

Open Source operations data – **what**, where and why?

- Data originating from **real world production systems**.
- Data made only available by **operating softwares and apps** in production.
- Eg: CI/CD data, telemetry, logs, operational dashboards.



Open Source operations data – what, **where** and why?

Operate First

- Operate First makes **operations open source**.
- An initiative centered around learning and developing code and practices in an open **production community cloud**.
- Deploy and maintain apps in an open environment leading to **open operations data** which include logs, issues, metrics.



www.operate-first.cloud/

Kubernetes

- Kubernetes testing infrastructure open sources CI/CD data originating from components such as Prow, Testgrid, Github



Open Source operations data – what, where and **why**?

- Open operations mean open sourcing **SRE best practices** leading to **better collaboration** between developers, operators, data scientists and better software.
- Open production **operations datasets are rare** and very valuable for AI communities.
- Open operations datasets can help **enable AI tools** to assist with cloud operations.



AI + Ops

Using AI tools to support Operations

AI Ops + CI

Supporting CI/CD by using AI capabilities

AI4CI

Artificial Intelligence for Continuous Integration

An collection of open source data science tools to collect and analyze CI data
built using open operations data.

AI4CI

<https://github.com/aicoe-aiops/ocp-ci-analysis>

- Collects open operations data from Kubernetes testing platforms eg: Testgrid, Github, and Prow.
- Collection of **Open Source AIOps tools** including scripts, notebooks, pipelines, dashboards and data sources.
- Collects metrics and **KPIs** and build **ML services** which can support CI/CD processes.
- Tools built and services deployed on Operate First

The screenshot shows the GitHub repository page for `aicoe-aiops / ocp-ci-analysis`. The repository is public and was generated from `aicoe-aiops/project-template`. It has 13 stars, 27 forks, and 13 watchers. The repository is currently on the `master` branch, with 6 branches and 11 tags. The repository contains a file tree with the following files and folders:

File/Folder	Description	Updated
<code>.github</code>	Update OWNERS and CODEOWNERS	2 years ago
<code>.jupyter</code>	Make precommit happy with our Python	2 years ago
<code>data</code>	Update nb with optimal stopping point	7 months ago
<code>docs</code>	Add architecture diagram	6 days ago
<code>manifests</code>	Update manifests (#167)	15 months ago
<code>models</code>	Add Optimal Stopping Point Model as a Service	4 months ago
<code>notebooks</code>	Adding github data extraction notebook	7 days ago
<code>references</code>	Initial commit	2 years ago
<code>reports</code>	tracking superset dashboard json on repo	8 months ago
<code>src</code>	Add Optimal Stopping Point Model as a Service	4 months ago
<code>.aicoe-ci.yaml</code>	Bump up base image initialized in CI.	2 months ago
<code>.env-example</code>	changed metric paths	8 months ago
<code>.gitignore</code>	Initial commit	2 years ago
<code>.pre-commit-config.yaml</code>	update pre-commit	last month
<code>.prow.yaml</code>	update pre-commit	last month
<code>.thoth.yaml</code>	Thoth's runtime environment must not use colon in name	3 months ago
<code>.zuul.yaml</code>	Initial commit	2 years ago
<code>CHANGELOG.md</code>	Release of version 0.2.0	7 months ago
<code>LICENSE</code>	add GPL-3.0 License	9 months ago

The repository also has a pull request #492 from `aakankshaduggal/master` merged 2 days ago. The repository is licensed under GPL-3.0 and has 13 stars, 27 forks, and 13 watchers. The repository is currently on the `master` branch, with 6 branches and 11 tags. The repository contains a file tree with the following files and folders:

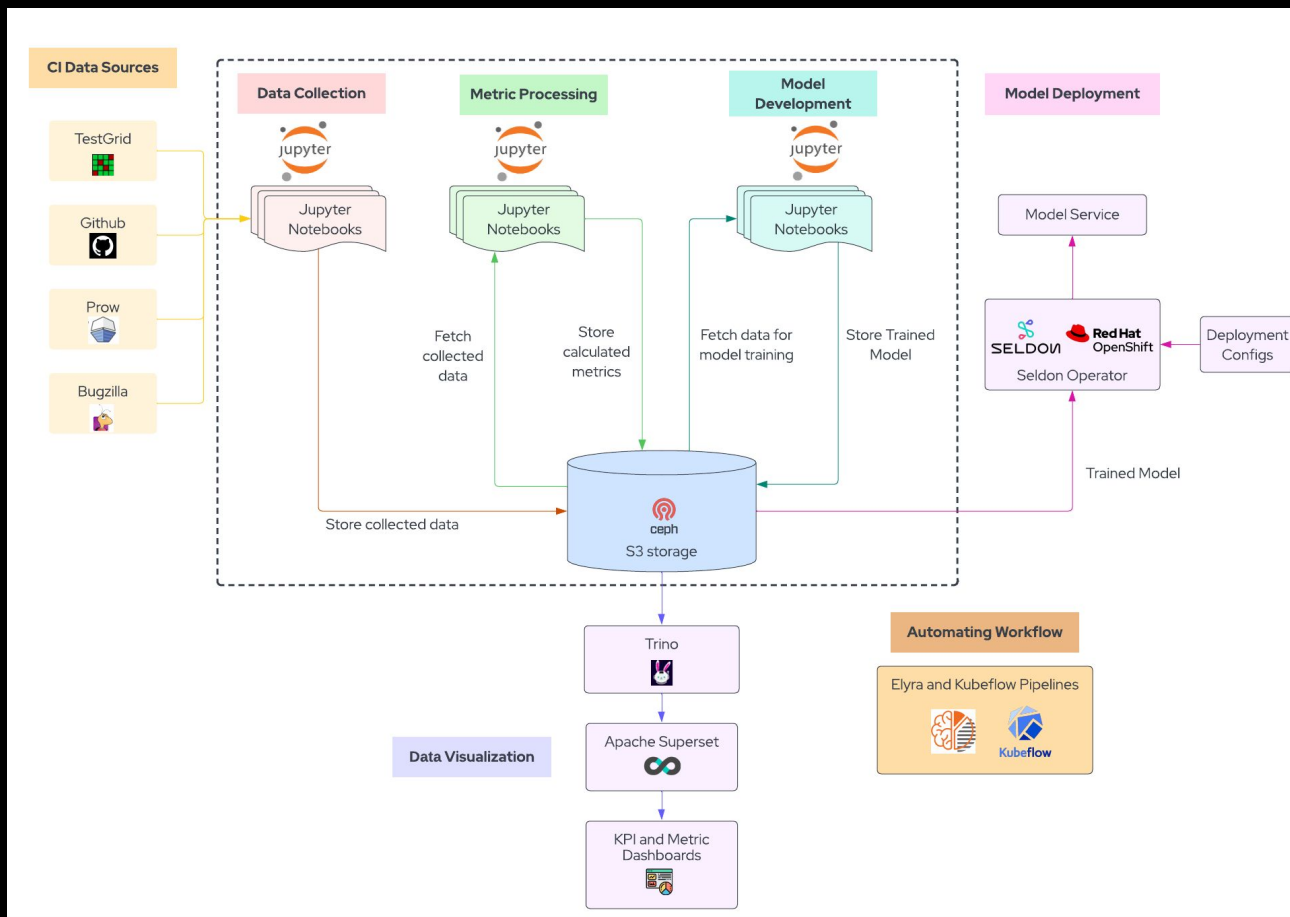
About
Developing AI tools for developers by leveraging the data made openly available by OpenShift and Kubernetes CI platforms.
www.operate-first.cloud/data-science...
Readme
GPL-3.0 license
13 stars
13 watching
27 forks

Releases 10
Tenth Release (Latest)
on Oct 1, 2021
+ 9 releases

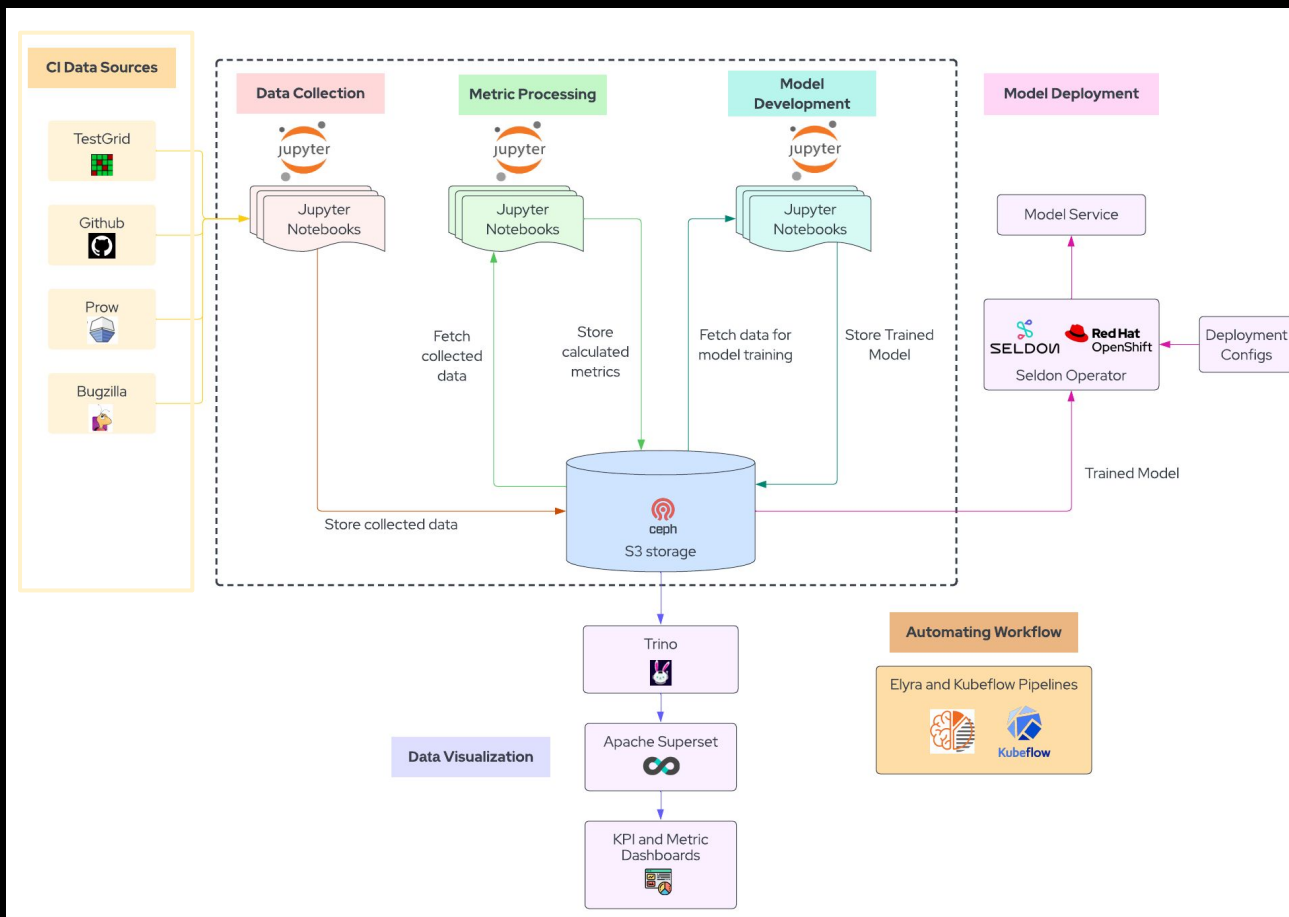
Packages
No packages published
[Publish your first package](#)

Contributors 22
+ 11 contributors

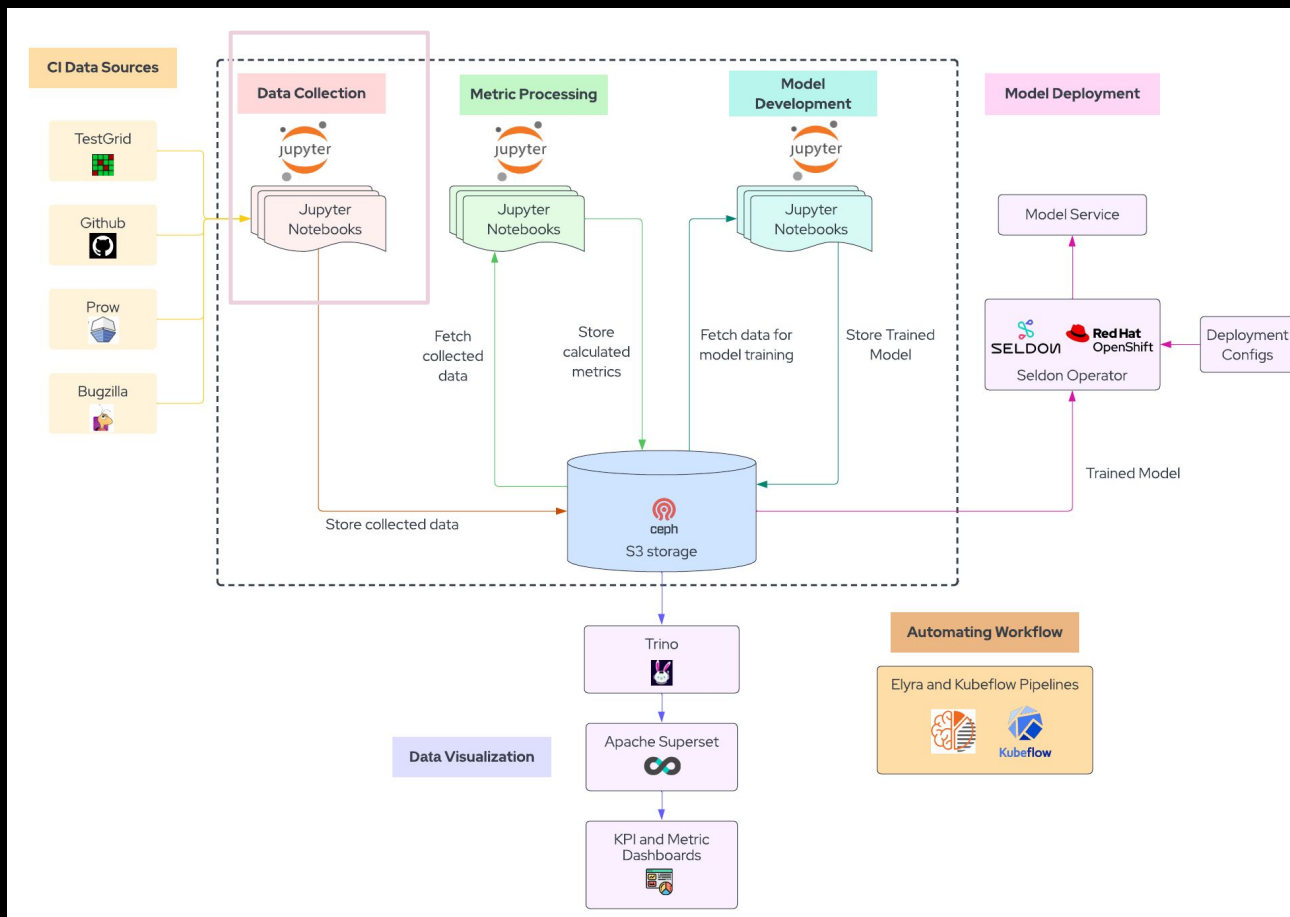
Architecture



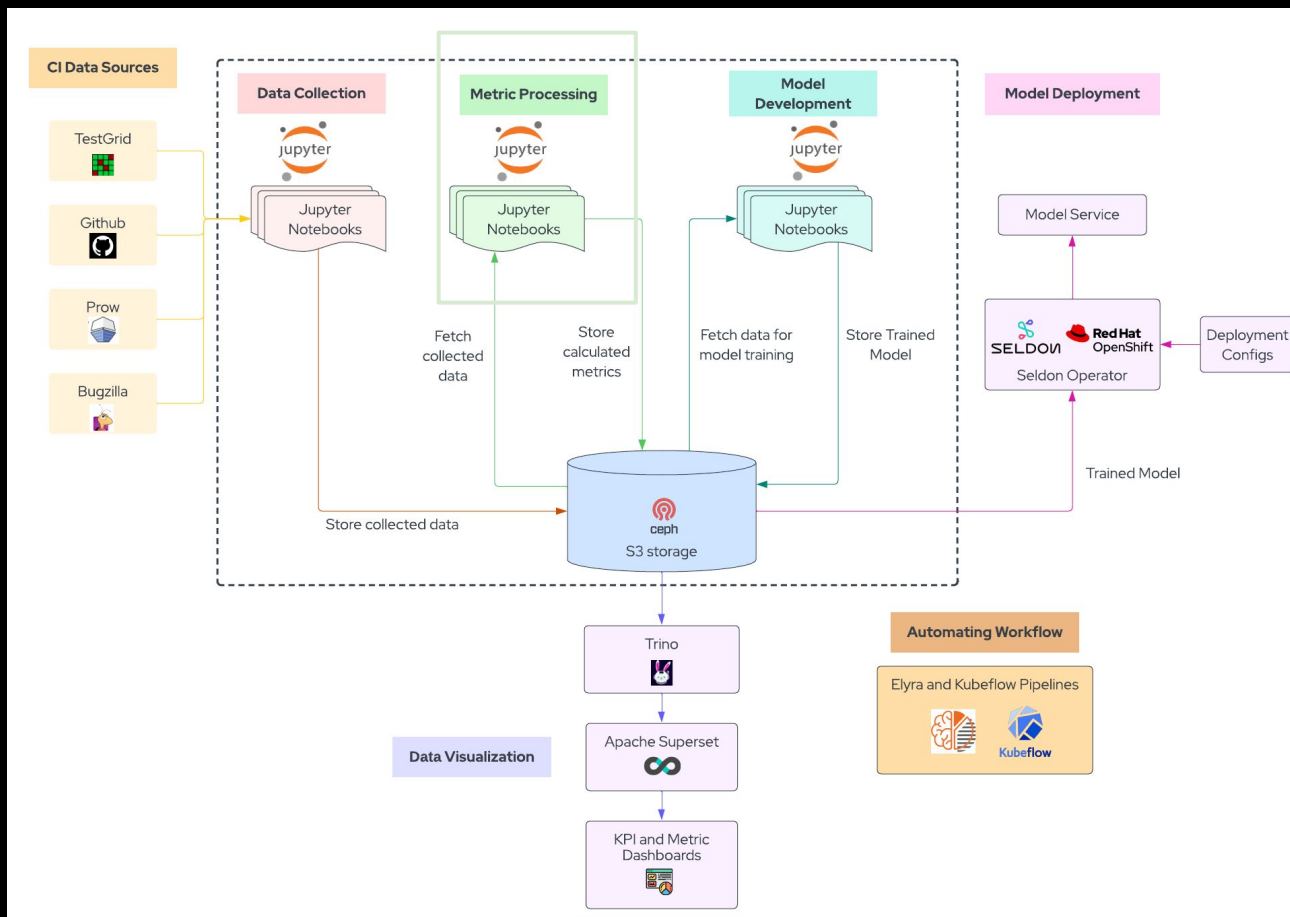
Architecture



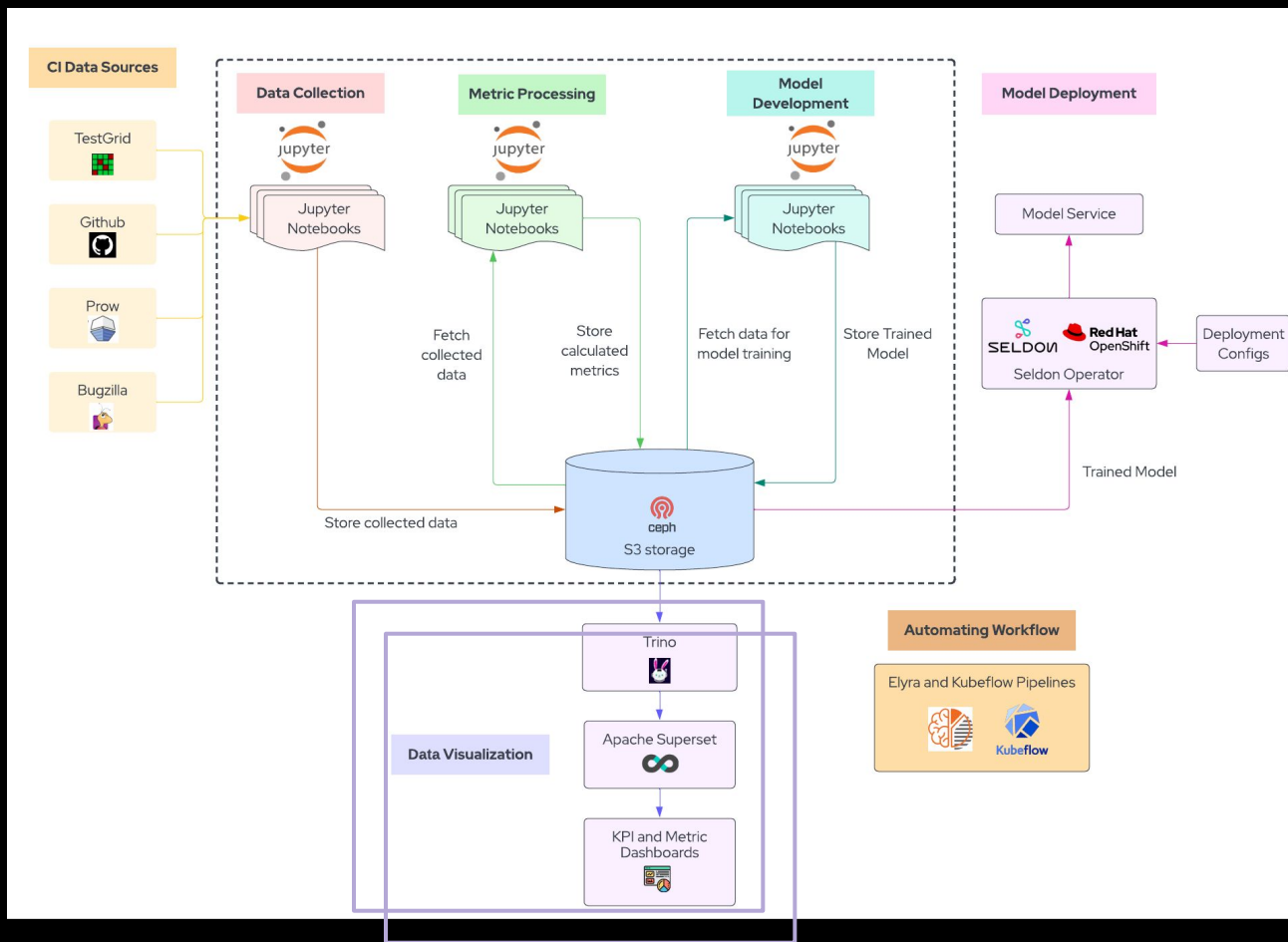
Architecture



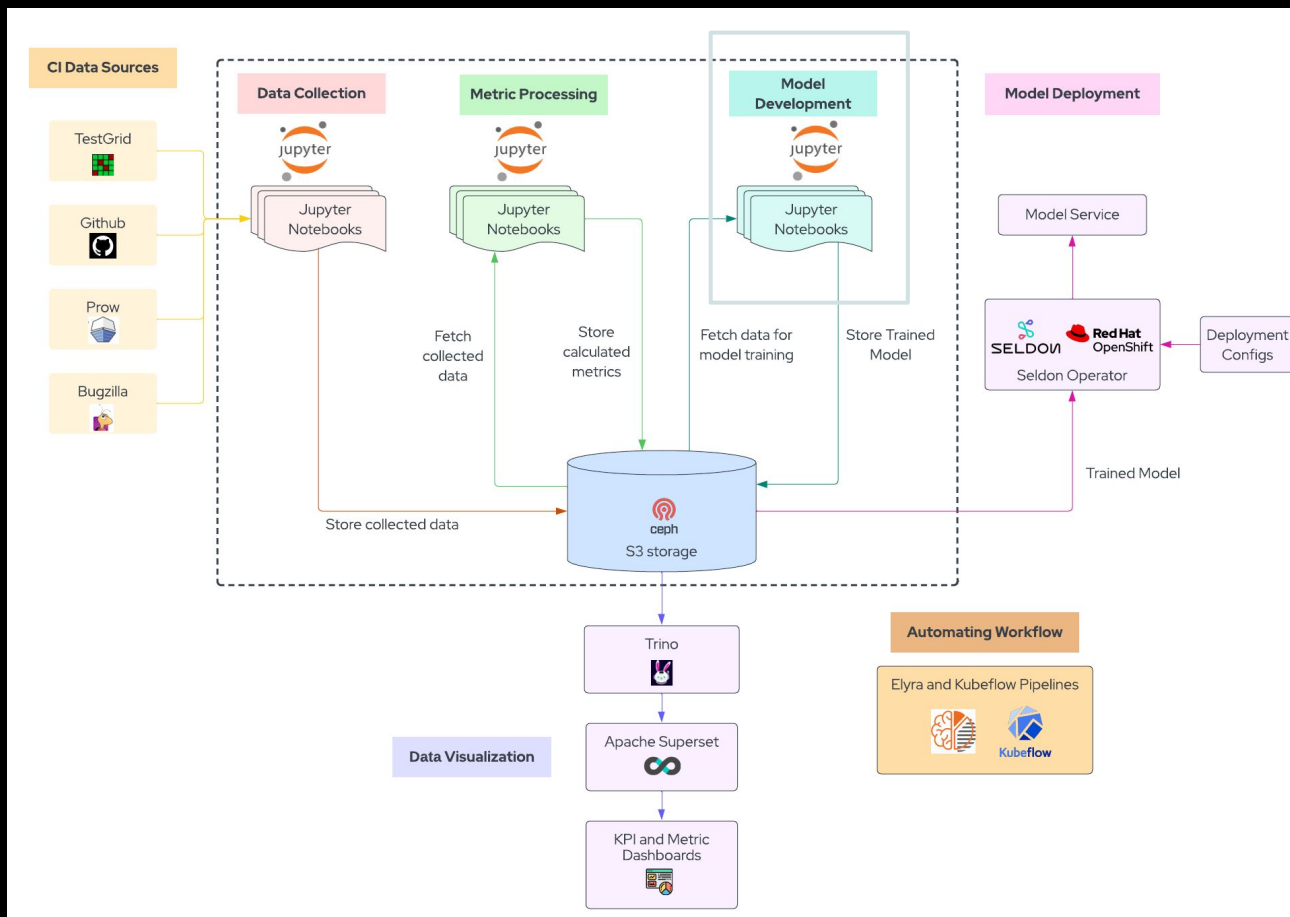
Architecture



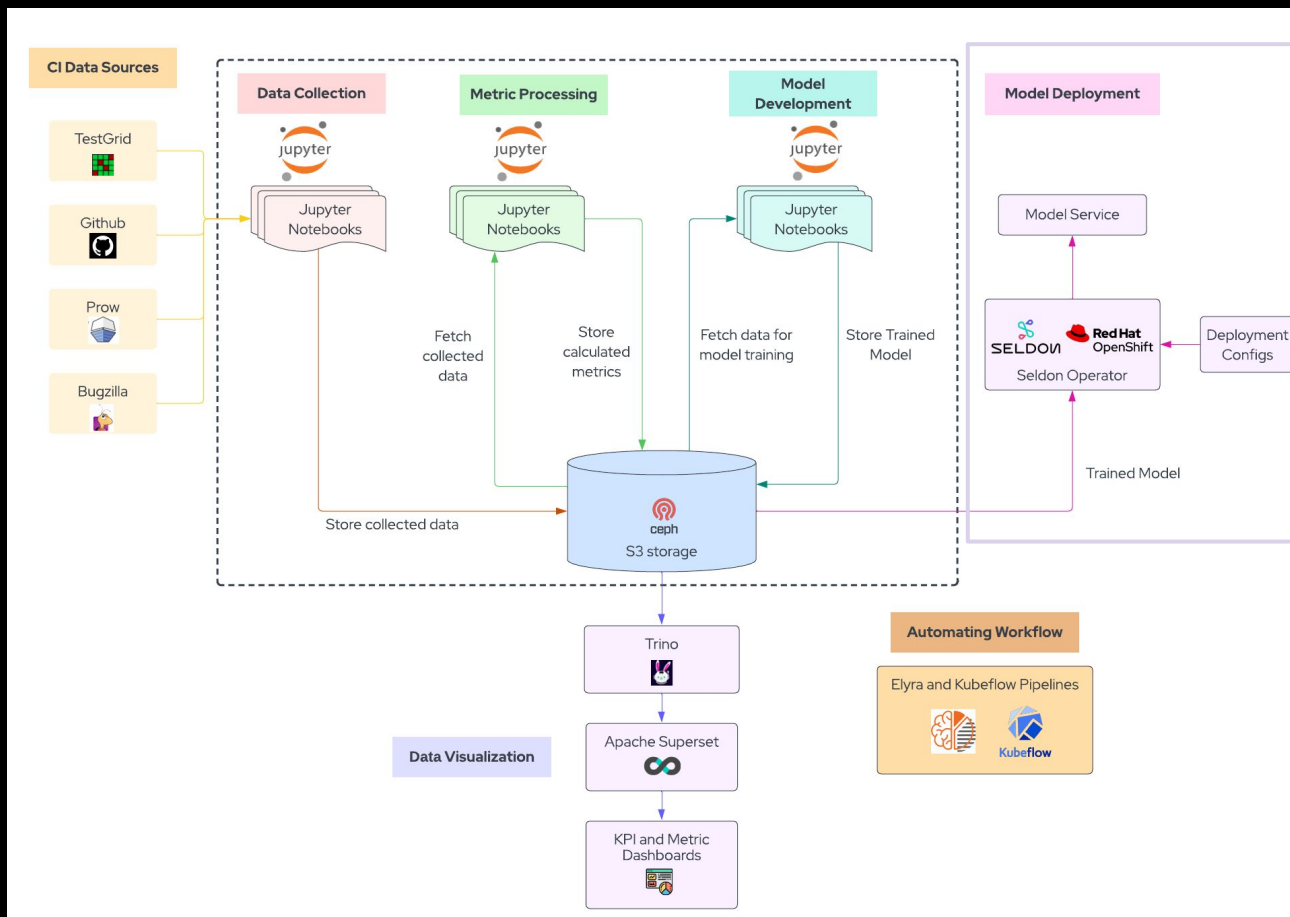
Architecture



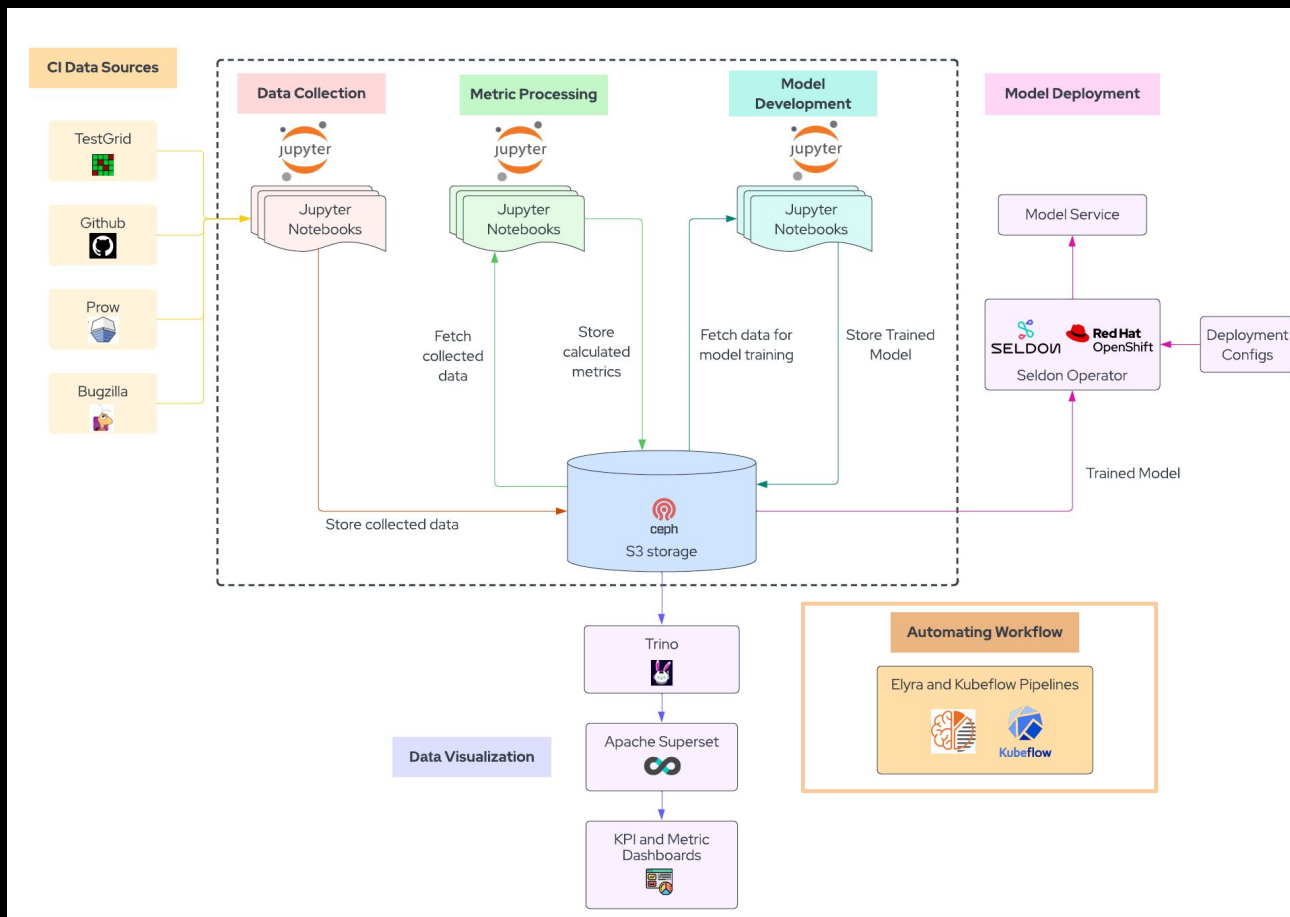
Architecture



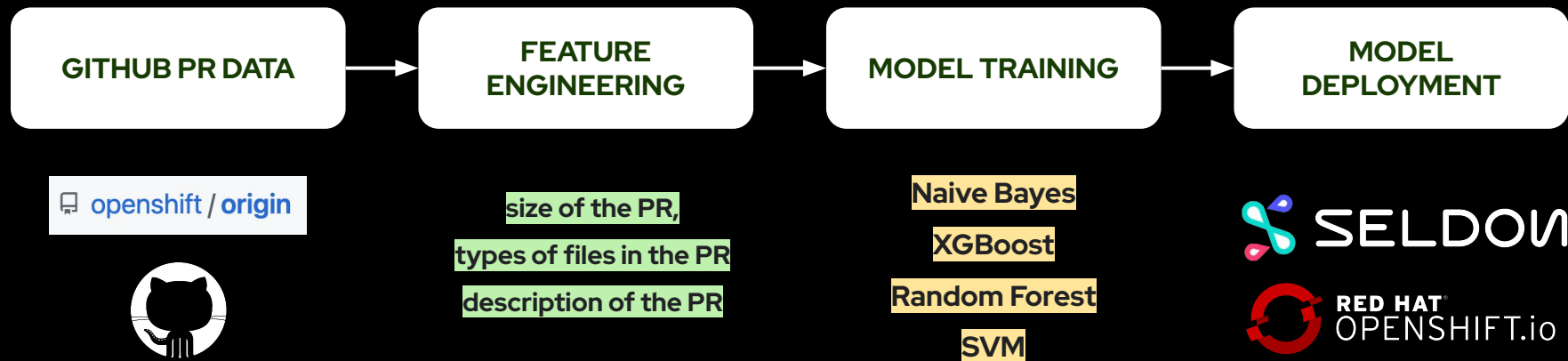
Architecture



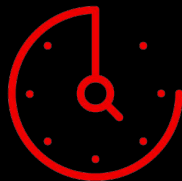
Architecture



ML Service: Github Time to Merge Model



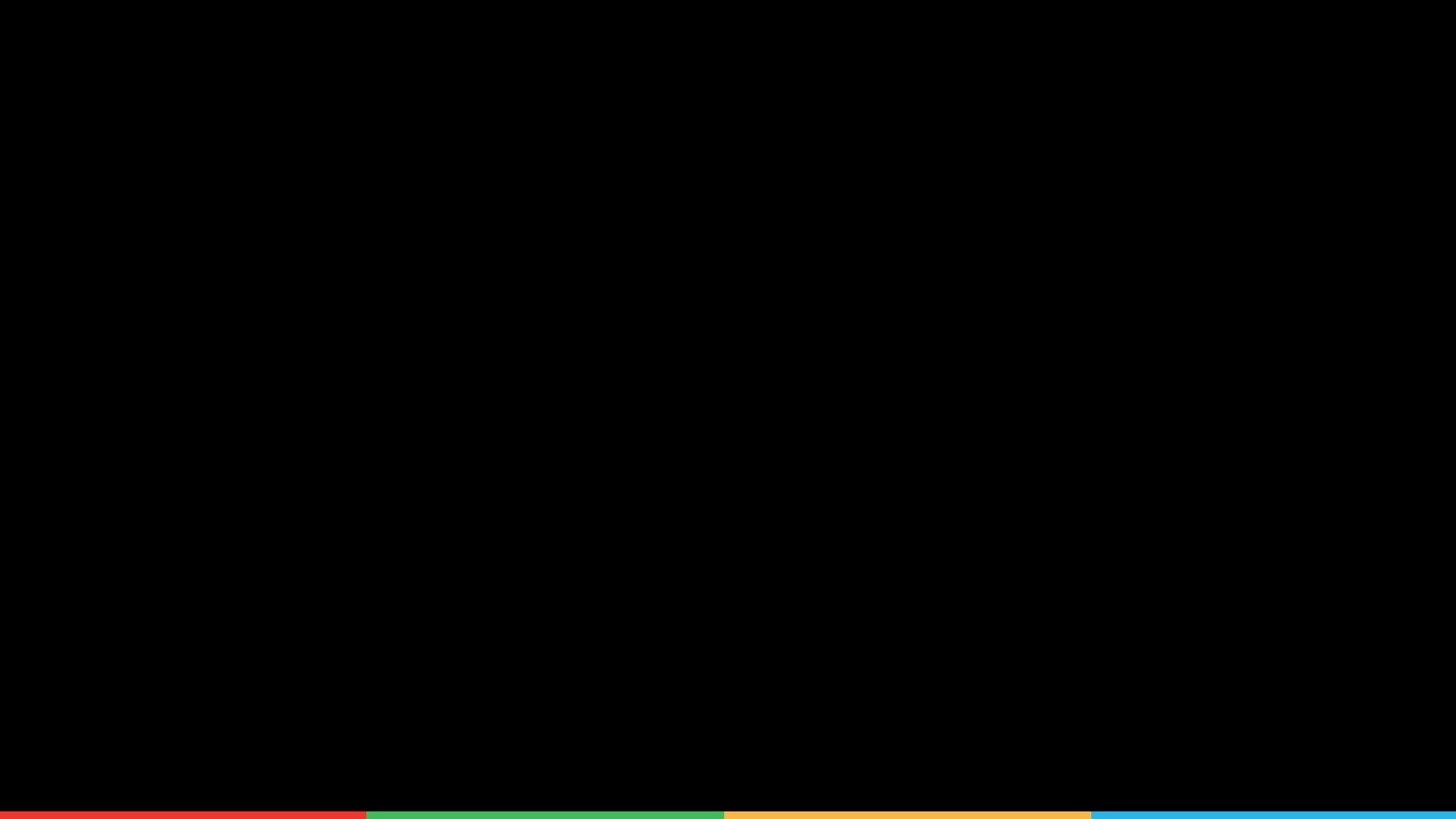
ML Service: Optimal Stopping Point Model



Sometimes
tests/builds take
longer than
expected to run

Find an Optimal
Stopping Point
after which the
test will fail.

We can better
allocate and save
resources.



Engage

Get Started

<https://github.com/aicoe-aiops/ocp-ci-analysis/blob/master/docs/get-started.md>



Open Data
Sources

Notebooks

Dashboards

Model Endpoints

Automated
Workflows

Video Playlist



<https://tinyurl.com/aiforci>

Operate First : [operate-first.cloud/](https://www.operate-first.cloud/)

- Get Started with the Operate First Cloud and Services :
<https://www.operate-first.cloud/getting-started>
- Join the Operate First Data Science Community -
<https://www.operate-first.cloud/data-science/operate-first-data-science-community/docs/meetup-landing-page.md>
- Video Playlist - <https://www.youtube.com/c/OperateFirst>



[@OperateFirst](https://twitter.com/OperateFirst)

OPERATE
FIRST>>