Contents

PROS & CONS OF OPEN-SOURCE PLATFORMS

Cnvrg.io

PROS:

Machine Learning Pipelines
AI Library
Open Compute
Dataset Management
Machine Learning Tracking
Machine Learning Model Deployment
Scalable Streaming Endpoints

FLYTE

PROS:

Data preparation and test validation splits
Model training
Model validation and scoring
Computing metrics

Configuring Machine Learning and Data workflows can get complex and error-prone with YAML. Flyte provides intuitive, user-friendly SDKs to get the workflows up and running in no time. Flyte is a Kubernetes-native workflow automation platform to unify data and ML processes.

IGUAZIO

PROS:

Iguazio is focused on accelerating the development, deployment, and management of your machine learning applications by the end-to-end automation of your ML pipelines with MLOps.

It makes the data science operational pipeline easy by enabling automation right from ingesting data from its source to training, deploying, and monitoring the machine learning model.

KEDRO

PROS:

Increases efficiency
robust pipelines
Project templates
Data management
Configuration management
Decreased operational risks for business

KUBEFLOW

PROS:

Full-featured MLOps platform

Offers a simple, scalable, and portable solution for running machine learning pipelines on Kubernetes.

Jupyter notebooks

Custom TensorFlow job operator

Simplified containerization

CONS:

Finicky to set up, unreliable, and difficult to configure. It also relied on many outdated components and libraries.

METAFLOW

PROS:

Framework for real-life data science,

Supports Python and R programming languages for managing data science projects.

Metaflow is focused on production pipeline and is designed to deploy and run at scale.

Great library support

Powerful version control toolkit

Good for large-scale machine learning development.

CONS:

Limited feature set and documentation

Not supported on Windows

Tightly integrated with AWS

Invoke via subprocess

ML FLOW

PROS:

Manages the machine learning lifecycle from experimentation, reproducibility to deployment.

It works with various machine learning libraries and programming languages.

Everything can be logged

Big Data capability

MLflow is generally a well-structured MLops solution that easily tracks, package, deploy, and manage the end-to-end lifecycle of machine learning models and we can:

monitor the ML pipeline,

store model metadata, and

pick the best-performing model.

Right now, there are 3 components provided by MLflow:

Tracking
Project
Model

Overall, users love MLflow because it's easy to use locally without a dedicated server, and has a fantastic UI where you can explore your experiments.

CONS:

You can't easily share experiments nor collaborate on them.
MLflow does not have a multi-user environment.
Role-based access is not present.
It lacks advanced security features.
The addition of extra workings to the models is not automatic.
It is not easy and ideal for deploying models to different platforms.

ML REEF

PROS:

platform for teams to collaborate and share the results of their machine learning experiments.
boosts the speed of development and makes the workflow more efficient
Data management
Script repositories
Experiment management
MLOps
Helps Newcomers, Experienced, and Enterprises equally.

ML RUN

PROS:

runs in a wide variety of environments and supports multiple technology stacks.
MLRun has a layered architecture that offers the following powerful functionality:

Feature and artifact store
Elastic serverless runtimes layer
Automation layer
Central management layer

NEPTUNE AI

PROS:

Neptune.ai is a lightweight tool for experiment management and collaboration in Data Science projects.

It is an organized place for all your experiments, data exploration Notebooks and more.

Neptune.ai is a hosted SaaS solution

Neptune comes with out-of-the-box collaboration features.

SELDON

PROS:

Solves problems faster and more effectively.

It's designed to streamline the data science workflow, with audit trails, advanced experiments, continuous integration and deployment, rolling updates, scaling, model explanations, and more.

Build scalable models

Monitor model performance.

Robust and reliable

Optimized servers provided by Seldon Core allow you to build large-scale deep learning systems without having to containerize them or worry about their security.

ZEN ML

PROS:

Preprocess data

Train your models

Conduct split testing

Evaluate the results

This framework frees you from all the troubles of delivering machine learning models with traditional tools. If you struggle with providing enough experiment data that prove the reproducibility of results, want to reduce waste and make the reuse of code simpler, ZenML will help.

TFX

PROS:

Builds complete ML pipelines

Pipeline stages:

- Data collection
- Data analysis
- Data validation
- Data transformation
- Model training
- Model evaluation and validation
- Pusher

Debugging
Scalability
Library Management
Graphs

TensorFlow Data Validation
TensorFlow Transform
TensorFlow Model Analysis
TensorFlow Serving
Machine Learning Metadata

CONS:

Does not support windows
Missing Symbolic Loops
No GPU support other than Nvidia and only Python language support
Low Computation Speed
Lacks behind in both speed and usage

VALOHAI

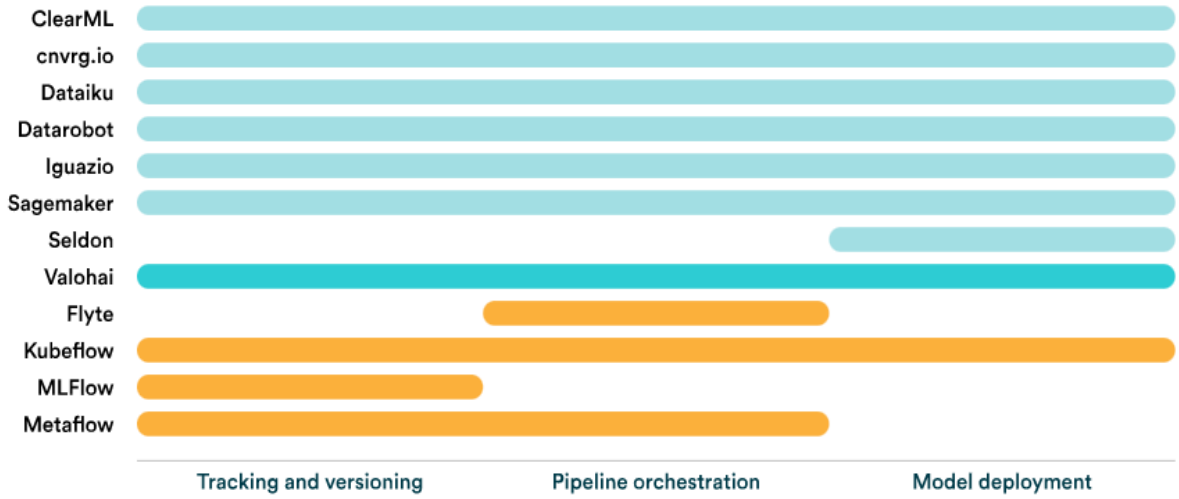
PROS:

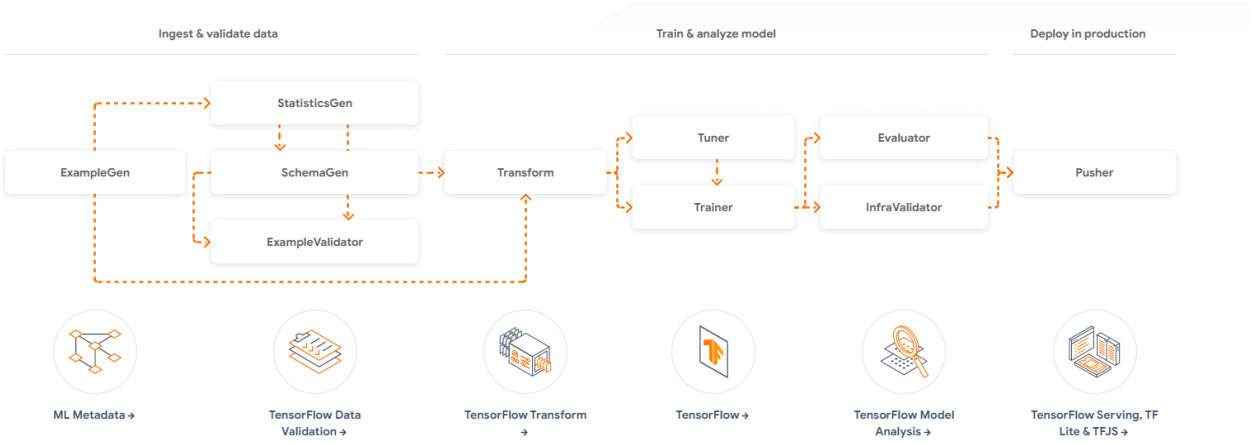
Valohai is the MLOps platform that can automate everything from data extraction to model deployment.

Tracking and versioning
Pipeline orchestration
Model deployment

COMPARISONS:

DIAGRAMS:





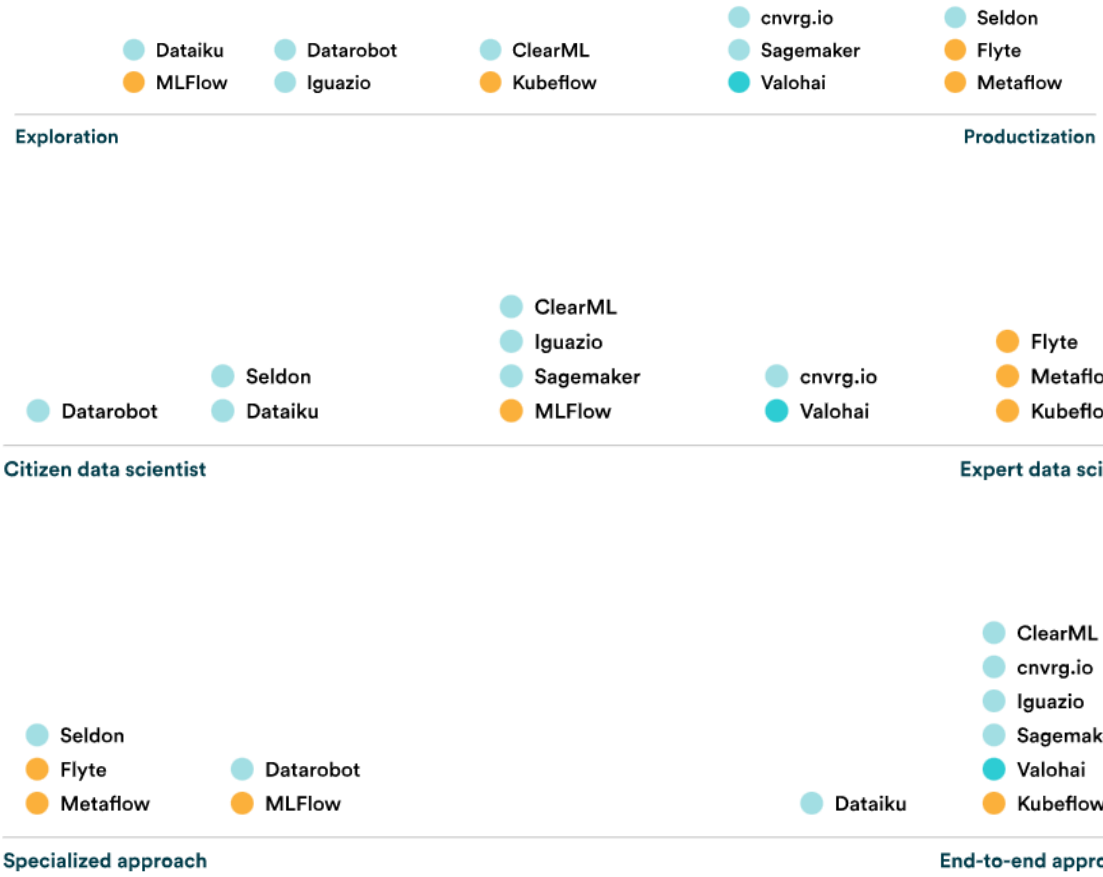


TABLE:

Name	Category	Description	Areas	Focus
AWS Sagemaker	Managed	Build, train, and deploy machine learning (ML) models for any use case with fully managed infrastructure, tools, and workflows	Tracking and versioning Pipeline orchestration Model deployment	AWS
ClearML	Managed, Open-source	MLOps with only 2-lines-of-code. Easily Develop, Orchestrate, and Automate ML Workflows at Scale.	Tracking and versioning Pipeline orchestration Model deployment	Experimentation, Structured data

Name	Category	Description	Areas	Focus
cnvrg.io	Managed, Open-source	An end-to-end machine learning platform to build and deploy AI models at scale	Tracking and versioning Pipeline orchestration Model deployment	Technology agnostic
Dataiku	Managed	Dataiku is the platform democratizing access to data and enabling enterprises to build their own path to AI in a human-centric way.	Tracking and versioning Pipeline orchestration Model deployment	Enterprise, Data Analysis, Business Intelligence
Datarobot	Managed	DataRobot is the leading end-to-end enterprise AI platform that automates and accelerates every step of your path from data to value.	Tracking and versioning Pipeline orchestration Model deployment	AutoML, Enterprise
Iguazio	Managed, Open-source	The Iguazio Data Science Platform automates MLOps with end-to-end machine learning pipelines, transforming AI projects into real-world business outcomes.	Tracking and versioning Pipeline orchestration Model deployment	Structured data
Seldon	Managed, Open-source	Deploy machine learning models at scale with more accuracy. 85% Faster.	Model deployment	Enterprise, Deployment
Valohai	Managed	Train, Evaluate, Deploy, Repeat. Valohai is the MLOps platform that can automate everything from data extraction to model deployment.	Tracking and versioning Pipeline orchestration Model deployment	Deep Learning, API-first, Technology agnostic
Flyte	Open-source	Lyft's Cloud Native Machine Learning and Data Processing Platform, Now Open Sourced	Pipeline orchestration	Pipelines
Kubeflow	Open-source	The Kubeflow project is dedicated to making deployments of machine learning (ML) workflows on Kubernetes simple, portable, and scalable.	Tracking and versioning Pipeline orchestration Model deployment	Community, Extendibility
Metaflow	Open-source	A framework for real-life data science	Tracking and versioning	Pipelines

Name	Category	Description	Areas	Focus
			Pipeline orchestration	
MLFlow	Open-source	MLflow is an open-source platform for managing the end-to-end machine learning lifecycle.	Tracking and versioning	Experimentation, Spark

CLOUD PLATFORMS FOR MLOPs

AWS

PROS:

creates and train machine learning models
implements a human review of the machine learning models
increases the forecast accuracy
uses natural language processing for language translation
creates personal recommendations in machine learning systems
provides Deep Learning solutions
converts text into life-like speech

Extensive range of infrastructure applications
Highly flexible
Easy transition for users with existing digital infrastructure
Frequently updated and maintained
Free tier available
Greater control over security
Scalability
Cost-effective pricing model
Rapid deployment
Support for large enterprises

CONS:

Range of infrastructure options can be overwhelming for more traditional enterprises
Hybrid options available, but not a priority
Organizations operating on legacy systems may experience longer migration times

GCP

PROS:

creates, trains, and manages ML models
speech recognition system for transmitting from speech to text and it supports 120 languages.
creates machine learning models for cloud vision that detect text, etc.
speech creation system for transmitting from text to speech
natural language processing for analyzing and classifying text

Excellent integration with other Google services
Fast I/O
Strong data analytics and storage
Facilitates easy collaboration
Designed for cloud-native business
Good portability and open-source integration

CONS:

Majority of components based on Google proprietary tech; no real control over Virtual Machines
Limited choice of programming languages
Complex transition away from the platform to another vendor
Fewer features/services
Fewer global data centers

IBM CLOUD

PROS:

builds machine learning and artificial intelligence models as well as preparing and analyzing data
speech recognition system for converting speech and audio into written text
speech creation system for converting text into natural-sounding audio
natural language processing for analyzing and classifying text
search visual images and classify them
creates and managing virtual assistants

CONS:

The cloud can take time to learn, and without easy access of support, this time only increases. IBM Cloud is not the right choice if your business needs a more short-term option, as it can take multiple hours to configure a bare metal server

AZURE

PROS:

provides smart cognitive services for applications.
provides Apache Spark-based analytics
provides smart and intelligent bot services that can be scaled
mobile and web applications

creates and deploys machine learning models on the cloud

Strong focus on Security
Scalability
Cost-effective
Strong IaaS and PaaS options
Support for open source
Hybrid cloud

CONS:

Requires considerable management
Requires platform expertise
More limited backward compatibility
Comparatively more costly than other leading vendors
Additional charge for pay-as-you-go option
Customer service

SUMMARY:

Most established:
as the very first major cloud vendor, AWS is undoubtedly the most established cloud provider on the market

Best availability:
with the greatest number of regions and availability zones worldwide, AWS comes out on top for the best availability

Biggest market share:
boasting 33% of the market share, AWS remains the most popular cloud provider

Fastest growth rate:
GCP currently has the fastest growth rate of almost 100%

Number of services:
when it comes to the sheer number of services and tools available, AWS comes out on top once more.

Familiarity:
as a Microsoft product providing straightforward integration with other Microsoft tools, Azure wins the race when it comes to customer familiarity

Most cost-effective price:

with the most customer-friendly pricing and discount models, GCP leads the way for the most cost-effective pricing.

TABLE COMPARISON

Details	AWS	Azure	GCP
Compute Services	1) AWS Beanstalk 2) Amazon EC2 3) Amazon EC2 Auto-Scaling 4) Amazon Elastic Container Registry 5) Amazon Elastic Kubernetes Service 6) Amazon Lightsail 7) AWS Serverless Application Repository 8) VMware Cloud for AWS 9) AWS Batch 10) AWS Fargate 11) AWS Lambda 12) AWS Outposts 13) Elastic Load Balancing	1) Platform-as-a-service (PaaS) 2) Function-as-a-service (FaaS) 3) Service Fabric 4) Azure Batch 5) Cloud Services 6) Container Instances Batch 7) Azure Container Service (AKS) 8) Virtual Machines Compute Engine 9) Virtual Machine Scale Sets	1) App Engine 2) Docker Container Registry 3) Instant Groups 4) Compute Engine 5) Graphics Processing Unit (GPU) 6) Knative 7) Kubernetes 8) Functions
Storage Services	1) Simple Storage Service (S3) 2) Elastic Block Storage (EBS) 3) Elastic File System (EFS) 4) Storage Gateway 5) Snowball 6) Snowball Edge 7) Snowmobile	1) Blob Storage 2) Queue Storage 3) File Storage 4) Disk Storage 5) Data Lake Store	1) Cloud Storage 2) Persistent Disk 3) Transfer Appliance 4) Transfer Service

AI/ML	1) SageMaker 2) Comprehend 3) Lex 4) Polly 5) Rekognition 6) Machine Learning 7) Translate 8) Transcribe 9) DeepLens 10) Deep Learning AMIs 11) Apache MXNet on AWS 12) TensorFlow on AWS	1) Machine Learning 2) Azure Bot Service 3) Cognitive Services	1) Cloud Machine Learning Engine 2) Dialogflow Enterprise Edition 5) Cloud Natural Language 6) Cloud Speech API 7) Cloud Translation API 8) Cloud Video Intelligence 9) Cloud Job Discovery (Private Beta)
Database Services	1) Aurora 2) RDS 3) DynamoDB 4) ElastiCache 5) Redshift 6) Neptune 7) Database Migration Service	1) SQL Database 2) Database for MySQL 3) Database for PostgreSQL 4) Data Warehouse 5) Server Stretch Database 6) Cosmos DB	1) Cloud SQL 2) Cloud Bigtable 3) Cloud Spanner 4) Cloud Datastore

		7) Table Storage 8) Redis Cache 9) Data Factory	
Backup Services	Glacier	1) Archive Storage 2) Backup 3) Site Recovery	1) Nearline (frequently accessed data) 2) Coldline (infrequently accessed data)
Serverless computing	1) Lambda 2) Serverless Application Repository	Functions	Google Cloud Functions
Strengths	1) Dominant market position 2) Extensive, mature offerings 3) Support for large organizations 4) Global reach 5) Flexibility and a wider range of services	1) Second largest provider 2) Integration with Microsoft tools and software 3) Broad feature set 4) Hybrid cloud 5) Support for open source 6) Ideal for startups and developers	1) Designed for cloud-native businesses 2) Commitment to open source and portability 3) Flexible contracts 4) DevOps expertise 5) Complete container-based model

			6) Most cost-efficient
Caching	Elastic Cache	Redis Cache	Cloud CDN
File Storage	EFS	Azure Files	ZFS and Avere
Networking	Amazon Virtual Private Cloud (VPC)	Azure Virtual Network (VNET)	Cloud Virtual Network
Security	AWS Security Hub	Azure Security Center	Cloud Security Command Center
Location	77 availability zones within 24 geographic regions	Presence in 60+ regions across the world	Presence in 24 regions and 73 zones. Available in 200+ countries and territories
Documentation	Best in class	High quality	High quality
DNS Services	Amazon Route 53	Azure Traffic Manager	Cloud DNS
Notifications	Amazon Simple Notification Service (SNS)	Azure Notification Hub	None
Load Balancing	Elastic Load Balancing	Load Balancing for Azure	Cloud Load Balancing
Automation	AWS Opsworks	Azure Automation	Compute Engine Management

Compliance	AWS CloudHSM	Azure Trust Center	Google Cloud Platform Security
Pricing/ Discount Options	One-year free trial along with a discount of up to 75% for a 1–3-year commitment	Up to 75% discount for a commitment ranging from one to three years	GCP Credit of $300 for 12 months apart from a sustained use discount of up to 30%

==========