

Federated Secure Data Science Platform

May 2020

DataFleets is a federated data science platform that lets your global team securely and compliantly explore data and train machine learning models.

*Move algorithms,
not data*

Perform computation where data resides. Return secured and anonymized statistical insights. Simplify and accelerate data science initiatives. Reduce "time to data" from months to days.



Python API for your existing IDE

No changes to analytics workflow

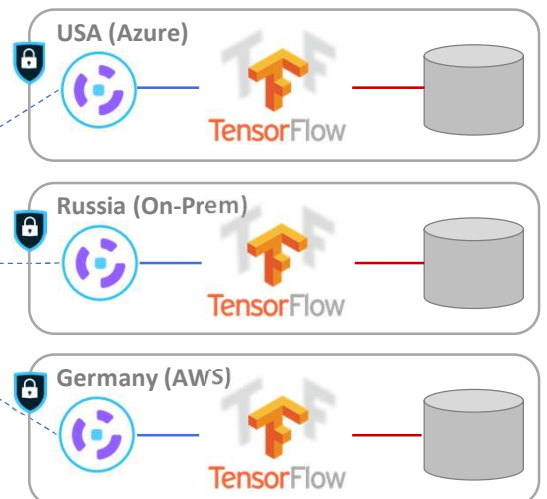


Orchestrate federation and aggregate insights



Endpoint where compute happens

No data ever leaves its home



Key Features

- Train ML models across different clouds or countries without moving data outside its jurisdiction
- Integrate with existing compute resources like Spark and TensorFlow
- Access data from a single point via Python API over a Jupyter notebook environment
- Export privacy-preserving and GDPR-compliant ML models and statistical insights
- Explore schema and data quality statistics without raw data access
- Record all data activities in audit logs, including user access and model privacy evaluation

Supported models

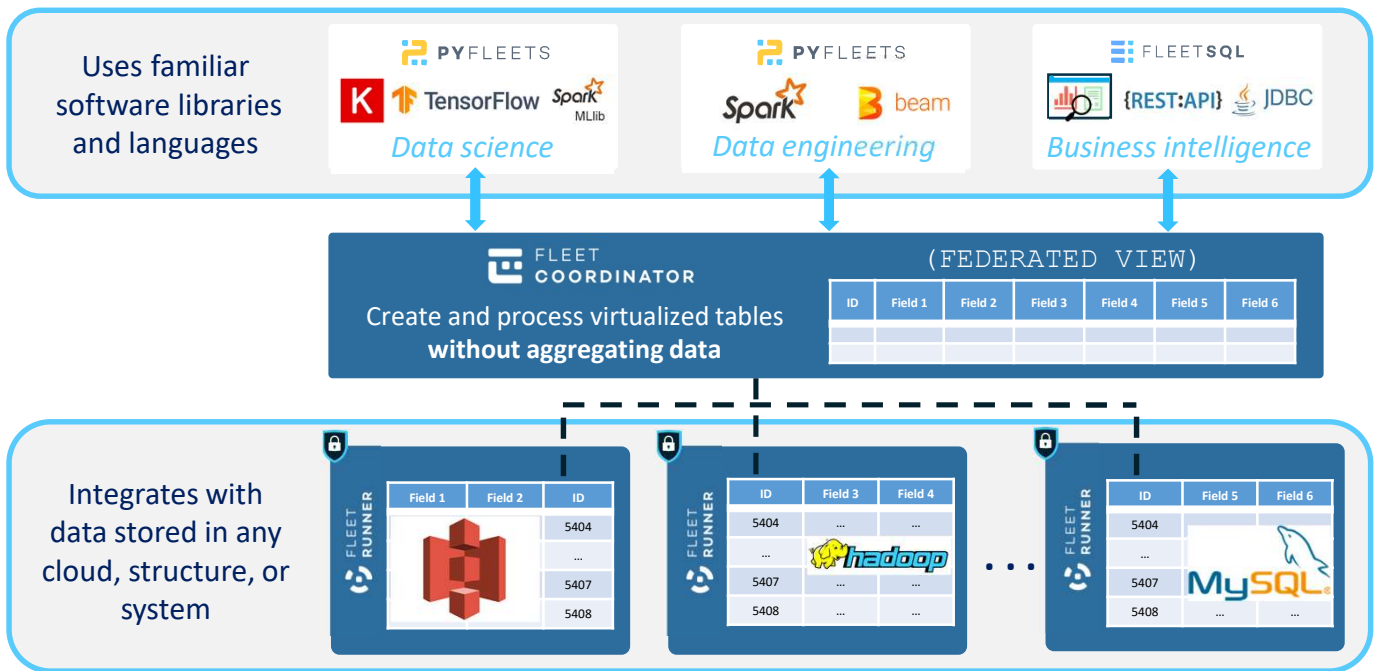
- ✓ **Linear models:** Logistic Regression, Linear Regression
- ✓ **Deep Neural Networks:** CNN, LSTM, Transformers (NLP language models)
- ✓ **Decision trees:** Gradient Boosted Trees

Deployment options

- ✓ **Cloud-native:** Azure, AWS, GCP or any Kubernetes cluster via Helm Charts
- ✓ **Docker image**
- ✓ **RedHat RPM package**
- ✓ **Standalone zip**

API/Framework support

- ✓ **Python API**
- ✓ **TensorFlow/Keras models**
- ✓ **Spark MLLib**



2020 Roadmap

Alibaba Cloud	Sklearn IBM Cloud	PyTorch, H2O Oracle Cloud
Q2	Q3	Q4