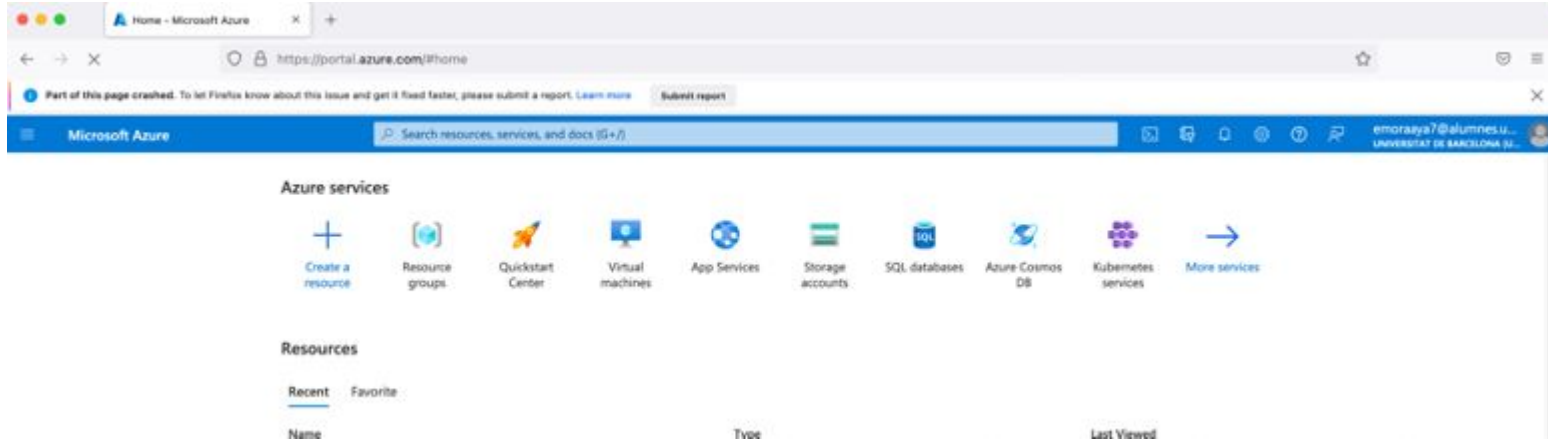


# Azure hands-on

# Azure Portal lab



1. Create a \$100 [Azure for Students](#) account for free with your UB Credentials.
2. Log in the [Azure Portal](#)
3. Azure [Calculator](#)

# Azure 101. Storage Account

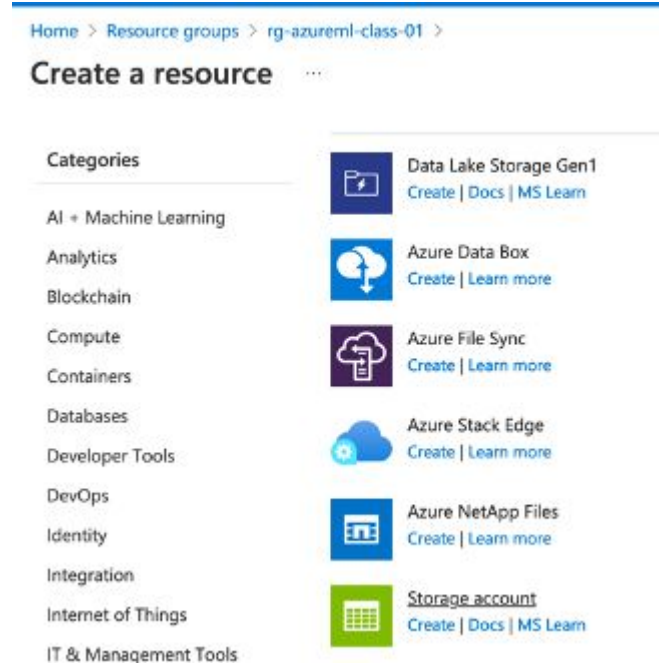
**Durable and highly available.** Redundancy ensures that your data is safe in the event of transient hardware failures. Data replicated in this way remains highly available in the event of an unexpected outage.

**Secure.** All data written to an Azure storage account is encrypted by the service. Azure Storage provides you with fine-grained control over who has access to your data.

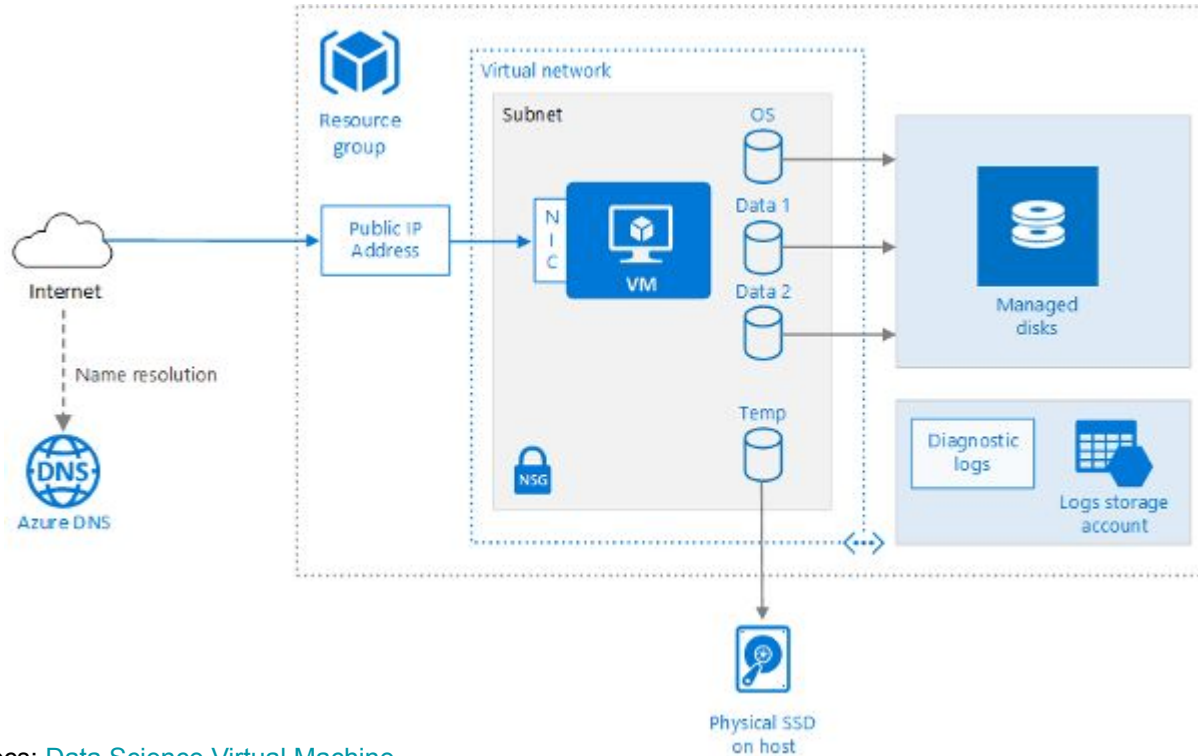
**Scalable.** Azure Storage is designed to be massively scalable to meet the data storage and performance needs of today's applications.

**Managed.** Azure handles hardware maintenance, updates, and critical issues for you.

**Accessible.** Data in Azure Storage is accessible from anywhere in the world over HTTP or HTTPS. Microsoft provides client libraries for Azure Storage in a variety of languages



# Azure 101. IaaS - Virtual Machine in the cloud



Azure docs: [Data Science Virtual Machine](#)

# Azure 101. Data Science VM

The **Data Science Virtual Machine** is an easy way to explore data and do machine learning in the cloud, since:

- are pre-configured with the complete operating system, security patches, drivers, and popular data science and development software.
- You can choose the hardware environment, ranging from lower-cost CPU-centric machines to very powerful machines with multiple GPUs, NVMe storage, and large amounts of memory.
- For machines with GPUs, all drivers are installed, all machine learning frameworks are version-matched for GPU compatibility, and acceleration is enabled in all application software that supports GPUs.

**Data Science Virtual Machine - Ubuntu 20.04**  

Microsoft

 **Data Science Virtual Machine - Ubuntu 20.04** 

Microsoft

Plan

Data Science Virtual Machine - Ubuntu...   

The "Data Science Virtual Machine (DSVM)" is a "Ubuntu 20.04" VM that has several popular tools for data science.

**Highlights:**

Operating System, Drivers and other base components

- "Ubuntu 20.04" image from Azure Marketplace
- Nvidia drivers, CUDA Toolkit, cuDNN (when GPU machines are used)
- Docker
- Anaconda ("conda")
- Git

Authoring Tools

- Visual Studio Code
- PyCharm Community Edition
- Jupyter, Jupyter Lab
- RStudio

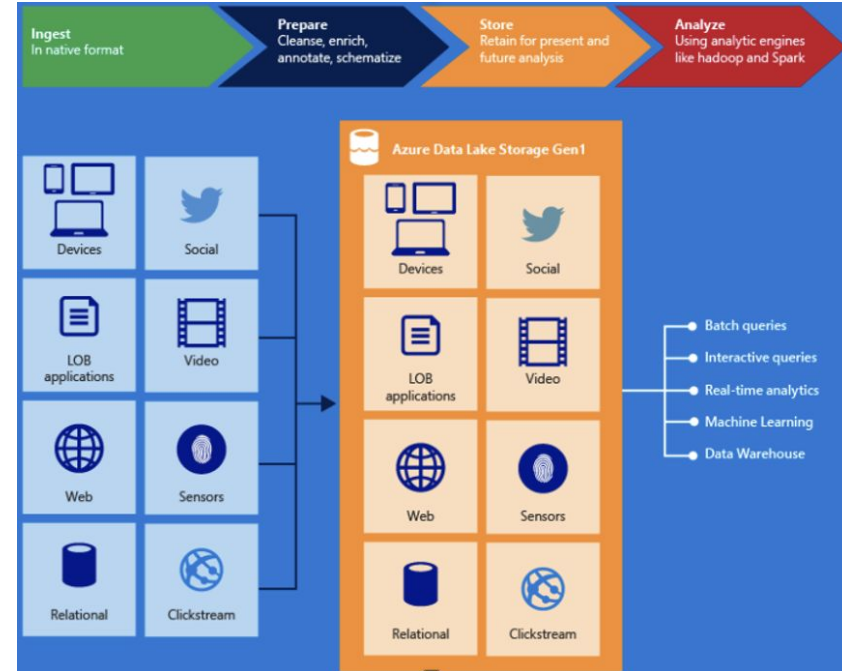
# Azure 101. Data Science VM

Tool	Windows Server 2019 DSVM	Ubuntu 18.04 DSVM	Ubuntu 20.04 DSVM	Usage notes
CUDA, cuDNN, NVIDIA Driver <sup>1/2</sup>	✓	✓	✓	CUDA, cuDNN, NVIDIA Driver on the DSVM
Horovod <sup>1/2</sup>	✗	✓	✓	Horovod on the DSVM
NVIDIA Systems Management Interface (nvdsi-ami) <sup>1/2</sup>	✓	✓	✓	nvdsi-ami on the DSVM
PyTorch <sup>1/2</sup>	✓	✓	✓	PyTorch on the DSVM
TensorFlow <sup>1/2</sup>	✓	✓	✓	TensorFlow on the DSVM
Integration with Azure Machine Learning <sup>1/2</sup> (Python)	(Python SDK, samples)	(Python SDK, CLL samples)	(Python SDK, CLL samples)	Azure ML SDK
XGBoost <sup>1/2</sup>	✓ (CUDA support)	✓ (CUDA support)	✓ (CUDA support)	XGBoost on the DSVM
Vowpal Wabbit <sup>1/2</sup>	✓	✓	✓	Vowpal Wabbit on the DSVM
Weka <sup>1/2</sup>	✗	✗	✗	
Ugn/GIM	✗	✓ (GPU, MPI support)	✓ (GPU, MPI support)	
H2O <sup>1/2</sup>	✗	✓	✓	
CarlBeast	✗	✓	✓	
Intel MKL	✗	✓	✓	
OpenCV	✗	✓	✓	
Dlib	✗	✓	✓	
Docker	✓ (Windows containers only)	✓	✓	
Nix <sup>1/2</sup>	✗	✓	✓	
Rattle	✗	✗	✗	
ONNX Runtime	✗	✓	✓	

Tool	Windows Server 2019 DSVM	Ubuntu 18.04 DSVM	Ubuntu 20.04 DSVM	Usage notes
Relational databases	SQL Server 2019 <sup>1/2</sup> Developer Edition	SQL Server 2019 <sup>1/2</sup> Developer Edition	SQL Server 2019 <sup>1/2</sup> Developer Edition	SQL Server on the DSVM
Database tools	SQL Server Management Studio SQL Server Integration Services bcp, sqlcmd	SQuirel SQL <sup>1/2</sup> (querying tool), bcp, sqlcmd ODBC/JDBC drivers	SQuirel SQL <sup>1/2</sup> (querying tool), bcp, sqlcmd ODBC/JDBC drivers	
Azure Storage Explorer <sup>1/2</sup>	✓	✓	✓	
Azure CLI	✓	✓	✓	
AzCopy	✓	✗	✗	AzCopy on the DSVM
Blob FUSE driver <sup>1/2</sup>	✗	✗	✗	blobfuse on the DSVM
Azure Cosmos DB Data Migration Tool	✓	✗	✗	Cosmos DB on the DSVM
Unix/Linux command-line tools	✗	✓	✓	
Apache Spark 3.1 (Mandalone)	✓	✓	✓	

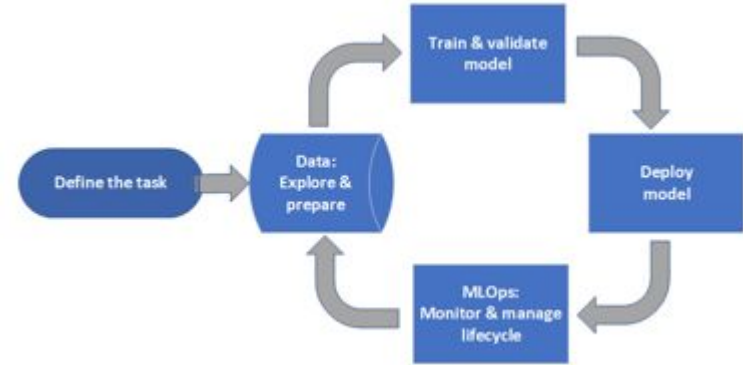
# Azure for DS. Data Lakes

- The foundation for building enterprise data lakes on Azure. Designed from the start to **service multiple petabytes** of information while sustaining **hundreds of gigabits of throughput**
- Management is easier because you can organize and manipulate files through **directories and subdirectories**.
- Security is enforceable because you can define **POSIX permissions** on directories or individual files.
- Also, Data Lake Storage Gen2 is very **cost effective** because it is built on top of the low-cost Azure Blob Storage. The additional features further lower the total cost of ownership for running big data analytics on Azure.



# Azure for DS. Azure ML

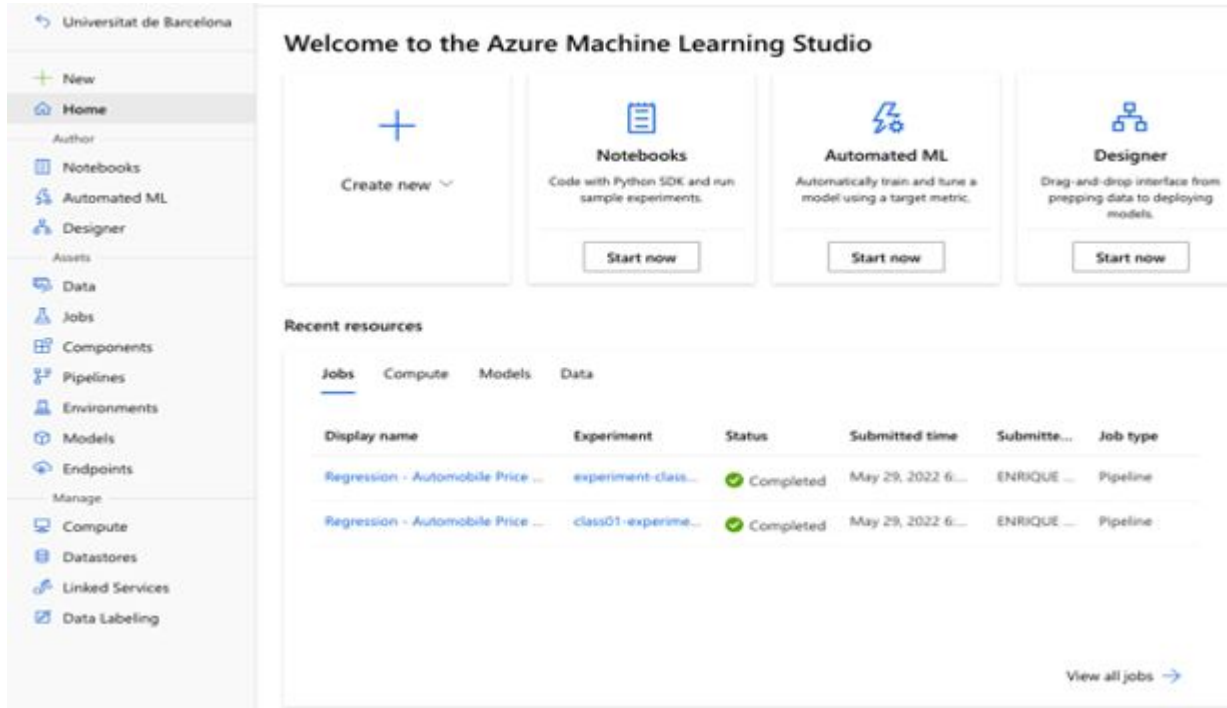
- Azure Machine Learning is a cloud service for accelerating and managing the machine learning project lifecycle..
- You can create a model in Azure Machine Learning or use a model built from an open-source platform, such as Pytorch, TensorFlow, or scikit-learn. MLOps tools help you monitor, retrain, and redeploy models.



HOWTO: [Create a new Azure ML Workspace](#)



# Azure for DS. Azure ML Studio

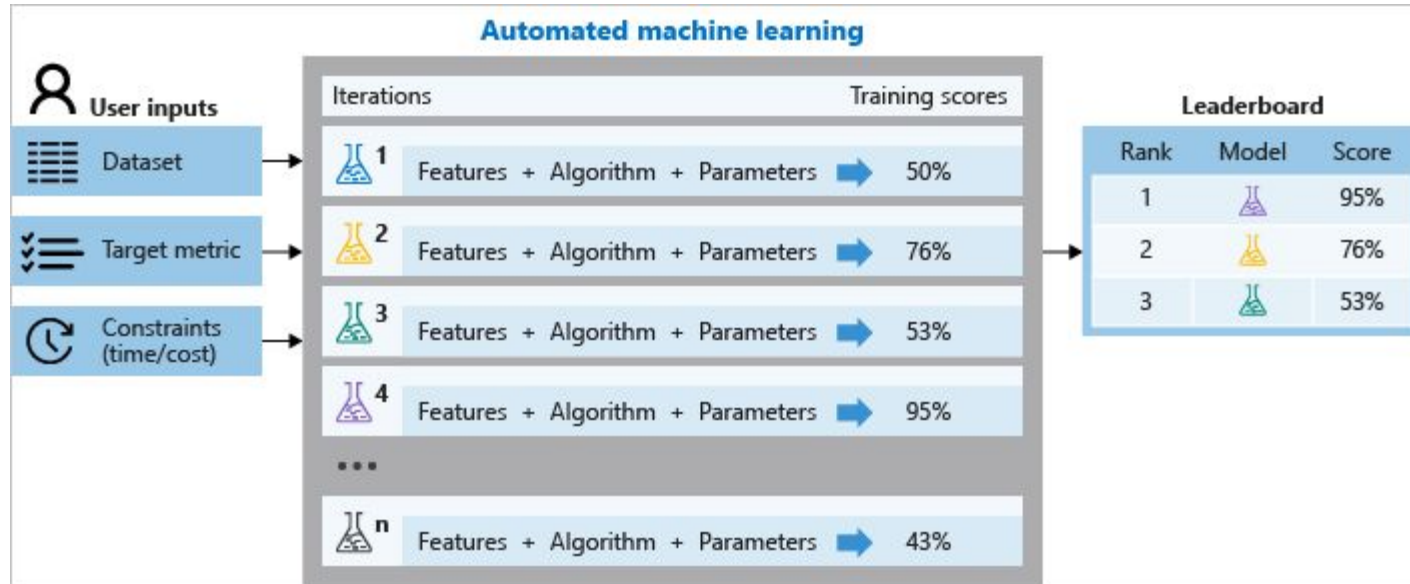


The screenshot displays the Azure Machine Learning Studio interface. On the left is a sidebar with navigation options: Universitat de Barcelona, New, Home (selected), Author, Notebooks, Automated ML, Designer, Assets, Data, Jobs, Components, Pipelines, Environments, Models, Endpoints, Manage, Compute, Datastores, Linked Services, and Data Labeling. The main area is titled 'Welcome to the Azure Machine Learning Studio' and features four cards: 'Create new' (with a plus icon), 'Notebooks' (with a document icon, description: 'Code with Python SDK and run sample experiments', and a 'Start now' button), 'Automated ML' (with a lightning bolt icon, description: 'Automatically train and tune a model using a target metric', and a 'Start now' button), and 'Designer' (with a flowchart icon, description: 'Drag-and-drop interface from prepping data to deploying models', and a 'Start now' button). Below these is a 'Recent resources' section with tabs for Jobs, Compute, Models, and Data. The 'Jobs' tab is active, showing a table of recent experiments.

Display name	Experiment	Status	Submitted time	Submitte...	Job type
Regression - Automobile Price ...	experiment-class...	Completed	May 29, 2022 6...	ENRIQUE ...	Pipeline
Regression - Automobile Price ...	class01-experime...	Completed	May 29, 2022 6...	ENRIQUE ...	Pipeline

[View all jobs →](#)

# Azure for DS. Azure AutoML



HOWTO: [AutoML Example Step-by-Step](#)

Azure docs: [Azure AutoML](#)

# Azure for DS. Azure ML - Endpoints

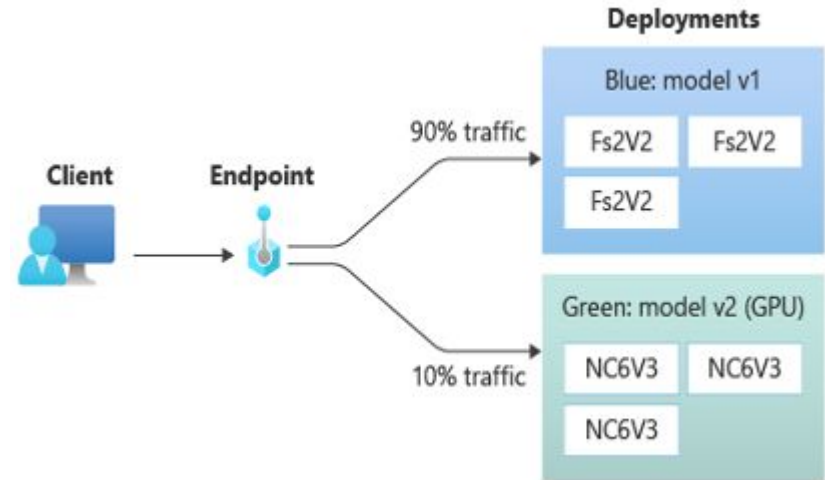
After you train a machine learning model, you need to deploy the model so that others can use it to do inferencing. In Azure Machine Learning, you can use endpoints and deployments to do so.

An **endpoint** is an HTTPS endpoint that clients can call to receive the inferencing (scoring) output of a trained model.

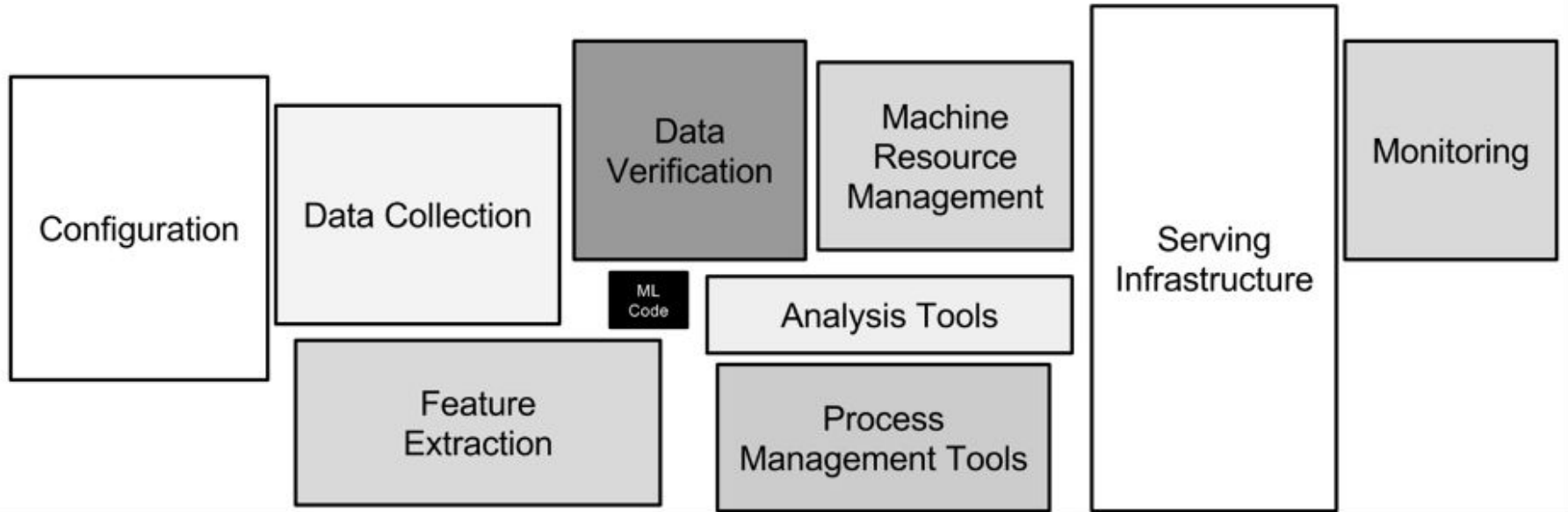
It provides:

- Authentication using "key & token" based auth
- SSL termination
- A stable scoring URI

A **deployment** is a set of resources required for hosting the model that does the actual inferencing.

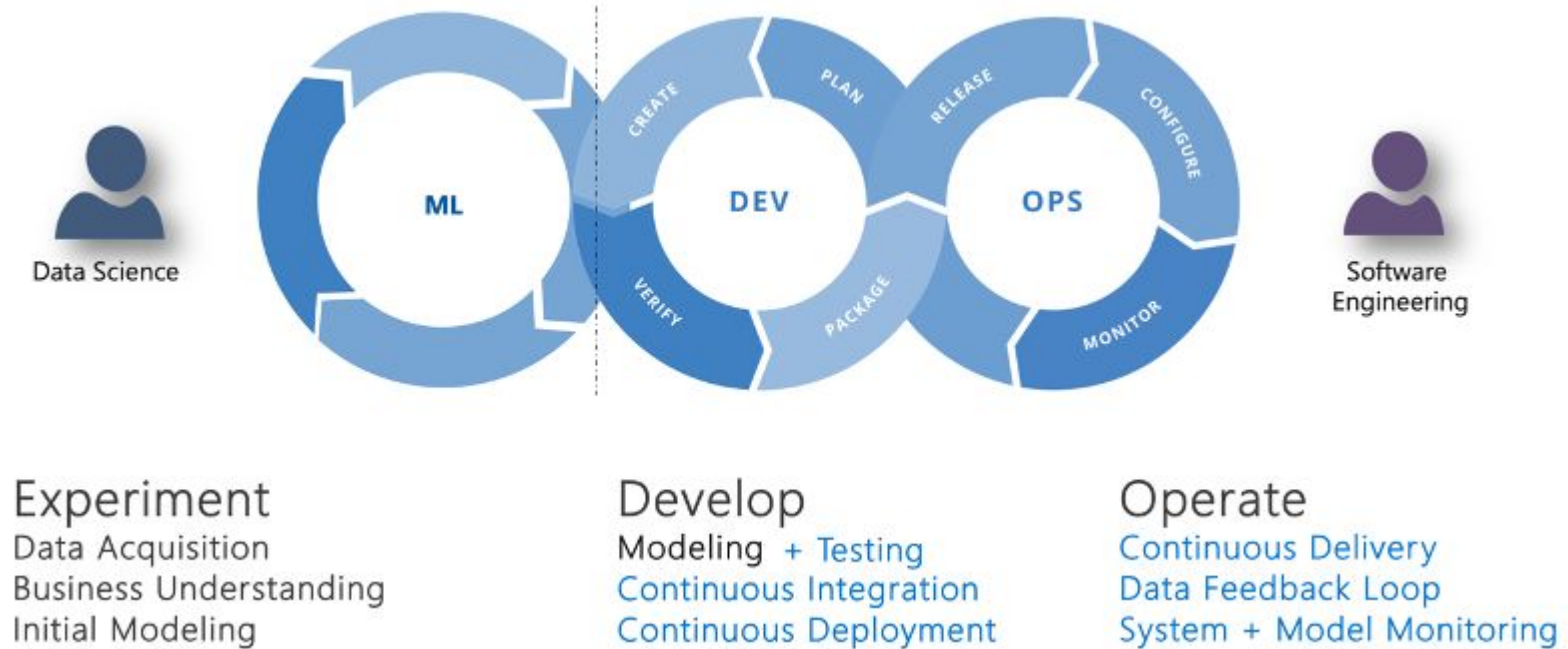


# Azure ML in Production

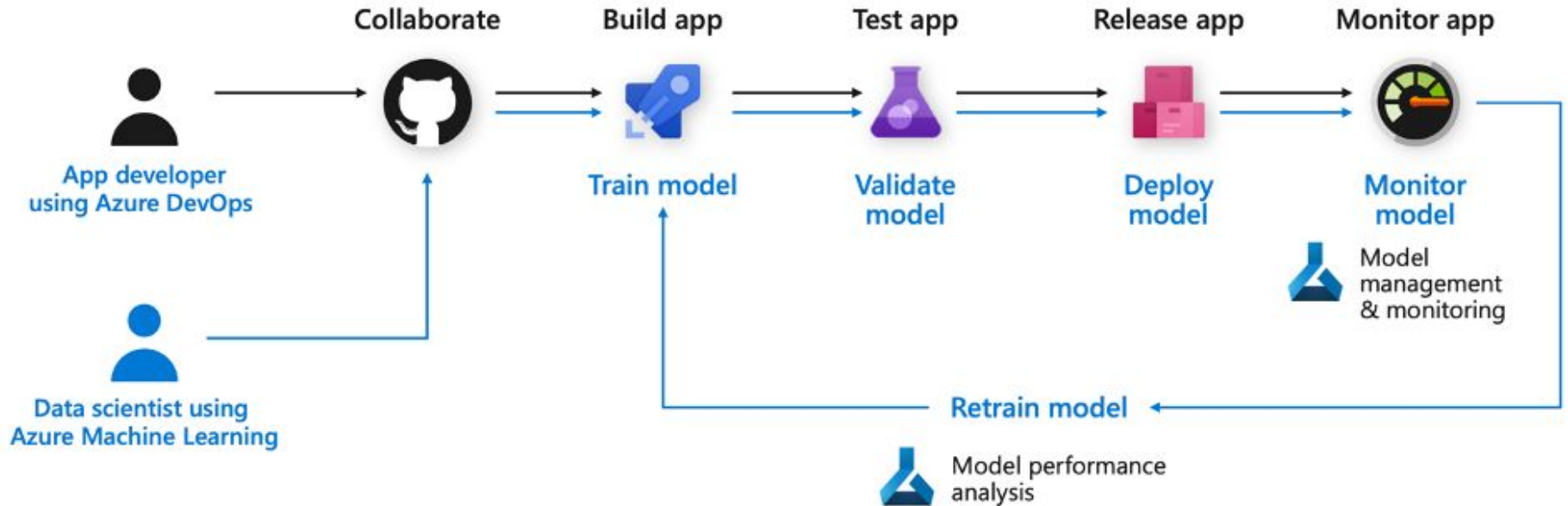


Sculley, D.; Holt, Gary; Golovin, Daniel; Davydov, Eugene; Phillips, Todd; Ebner, Dietmar; Chaudhary, Vinay; Young, Michael; Crespo, Jean-Francois; Dennison, Dan (7 December 2015). "Hidden Technical Debt in Machine Learning Systems"

# Azure ML in Production. MLOps



# Azure ML in Production. Azure DevOps



# Azure ML in Production. MLOps Lab

1. Use the Azure Subscription and Azure Machine Learning Workspace
2. Follow the pre-requisites: <https://bit.ly/3wYizJB>
3. Follow the Hands-on Step-by-Step: <https://bit.ly/3geqaxU> Until Exercise 6
4. Don't forget to stop or delete resources when you finish!