






Machine learning at scale

02/12/2018 • 3 minutes to read • Contributors     

In this article

[Model preparation and training](#)

[Model deployment and consumption](#)

[Challenges](#)

[Machine learning at scale in Azure](#)

[Next steps](#)

Machine learning (ML) is a technique used to train predictive models based on mathematical algorithms. Machine learning analyzes the relationships between data fields to predict unknown values.

Creating and deploying a machine learning model is an iterative process:

- Data scientists explore the source data to determine relationships between *features* and predicted *labels*.
- The data scientists train and validate models based on appropriate algorithms to find the optimal model for prediction.
- The optimal model is deployed into production, as a web service or some other encapsulated function.
- As new data is collected, the model is periodically retrained to improve its effectiveness.

Machine learning at scale addresses two different scalability concerns. The first is training a model against large data sets that require the scale-out capabilities of a cluster to train. The second centers on operationalizing the learned model so it can scale to meet the demands of the applications that consume it. Typically this is accomplished by deploying the predictive capabilities as a web service that can then be scaled out.

Machine learning at scale has the benefit that it can produce powerful, predictive capabilities because better models typically result from more data. Once a model is trained, it can be deployed as a stateless, highly performant scale-out web service.

Model preparation and training

During the model preparation and training phase, data scientists explore the data interactively using languages like Python and R to:

- Extract samples from high volume data stores.
- Find and treat outliers, duplicates, and missing values to clean the data.
- Determine correlations and relationships in the data through statistical analysis and visualization.
- Generate new calculated features that improve the predictiveness of statistical relationships.
- Train ML models based on predictive algorithms.
- Validate trained models using data that was withheld during training.

To support this interactive analysis and modeling phase, the data platform must enable data scientists to explore data using a variety of tools. Additionally, the training of a complex machine learning model can require a lot of intensive processing of high volumes of data, so sufficient resources for scaling out the model training is essential.

Model deployment and consumption

When a model is ready to be deployed, it can be encapsulated as a web service and deployed in the cloud, to an edge device, or within an enterprise ML execution environment. This deployment process is referred to as operationalization.

Challenges

Machine learning at scale produces a few challenges:

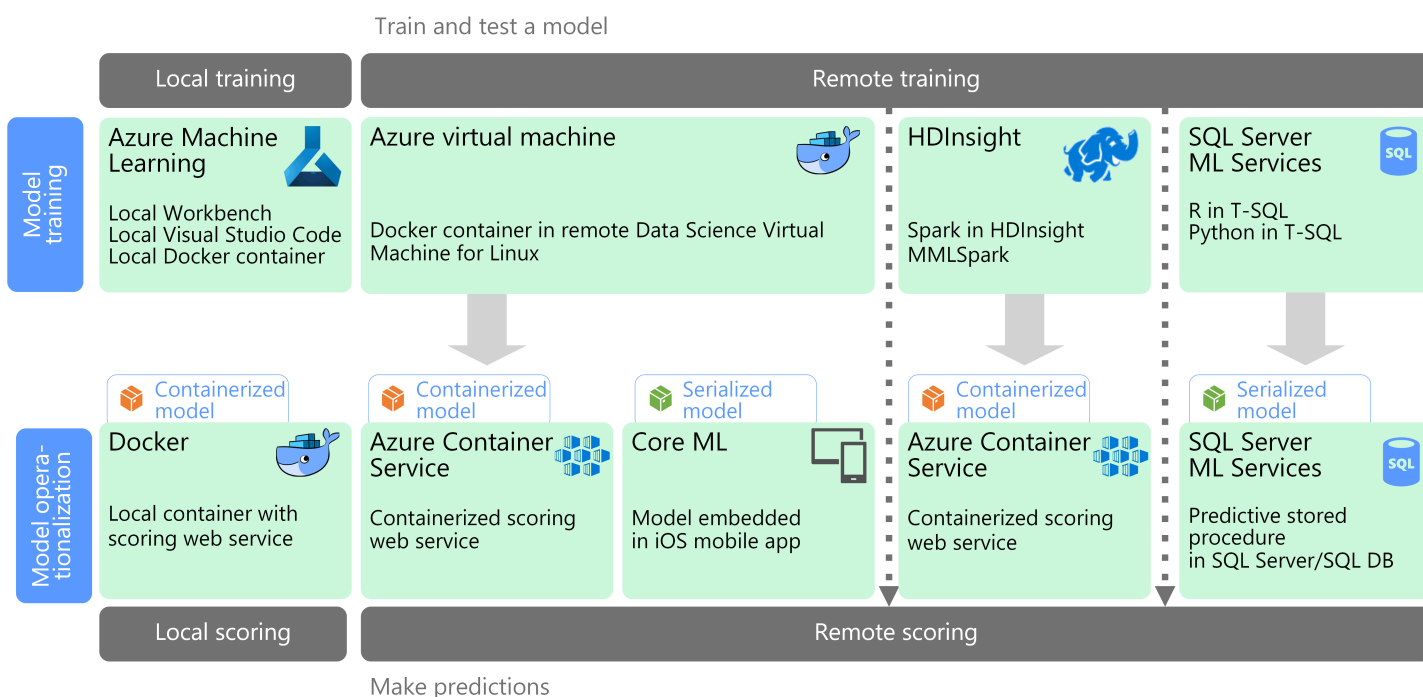
- You typically need a lot of data to train a model, especially for deep learning models.
- You need to prepare these big data sets before you can even begin training your model.
- The model training phase must access the big data stores. It's common to perform the model training using the same big data cluster, such as Spark, that is used for data preparation.
- For scenarios such as deep learning, not only will you need a cluster that can provide you scale-out on CPUs, but your cluster will need to consist of GPU-enabled nodes.

Machine learning at scale in Azure

Before deciding which ML services to use in training and operationalization, consider whether you need to train a model at all, or if a prebuilt model can meet your requirements. In many cases, using a prebuilt model is just a matter of calling a web service or using an ML library to load an existing model. Some options include:

- Use the web services provided by Azure Cognitive Services.
- Use the pretrained neural network models provided by the Cognitive Toolkit.
- Embed the serialized models provided by Core ML for an iOS app.

If a prebuilt model does not fit your data or your scenario, options in Azure include Azure Machine Learning, HDInsight with Spark MLlib and MMLSpark, Azure Databricks, Cognitive Toolkit, and SQL Machine Learning Services. If you decide to use a custom model, you must design a pipeline that includes model training and operationalization.



For a list of technology choices for ML in Azure, see:

- [Choosing a cognitive services technology](#)
- [Choosing a machine learning technology](#)
- [Choosing a natural language processing technology](#)

Next steps

The following reference architectures show machine learning scenarios in Azure:

- [Batch scoring on Azure for deep learning models](#)

- [Real-time scoring of Python Scikit-Learn and Deep Learning Models on Azure](#)