

What's included in the Data Science VM?

The Data Science Virtual Machine has many popular data science and deep learning tools already installed and configured. It also includes tools that make it easy to work with various Azure data and analytics products. You can explore and build predictive models on large-scale data sets using the Microsoft ML Server (R, Python) or using SQL Server 2017. A host of other tools from the open source community and from Microsoft are also included, as well as sample code and notebooks. The following table itemizes and compares the main components included in the Windows and Linux editions of the Data Science Virtual Machine.

TOOL	WINDOWS EDITION	LINUX EDITION
Microsoft R Open with popular packages pre-installed	Y	Y
Microsoft ML Server (R, Python) Developer Edition includes, * RevoScaleR/revoscalepy parallel and distributed high-performance framework (R & Python) * MicrosoftML - New state-of-the-art ML algorithms from Microsoft * R and Python Operationalization	Y	Y
Microsoft Office Pro-Plus with shared activation - Excel, Word and PowerPoint	Y	N
Anaconda Python 2.7, 3.5 with popular packages pre-installed	Y	Y
JuliaPro with popular packages for Julia language pre-installed	Y	Y
Relational Databases	SQL Server 2017 Developer Edition	PostgreSQL (CentOS), SQL Server 2017 Developer Edition (Ubuntu)
Database tools	* SQL Server Management Studio * SQL Server Integration Services * bcp, sqlcmd * ODBC/JDBC drivers	* SQuirreL SQL (querying tool), * bcp, sqlcmd * ODBC/JDBC drivers

Windows Data Science VM

- For more information on how to create a Windows DSVM and use it, see [Provision the Windows Data Science Virtual Machine](#). For more information on how to perform various tasks needed for your data science project on the Windows DSVM, see [Ten things you can do on the Data Science Virtual Machine](#).

Linux Data Science VM

- For more information on how to create an Ubuntu DSVM and use it, see [Provision the Data Science Virtual Machine for Linux \(Ubuntu\)](#).
- For more information on how to create a CentOS DSVM and use it, see [Provision a Linux CentOS Data Science Virtual Machine on Azure](#).

For a walkthrough that shows you how to perform several common data science tasks with the Linux VM, both CentOS and Ubuntu, see [Data science on the Linux Data Science Virtual Machine](#)

TOOL	WINDOWS EDITION	LINUX EDITION
Scalable in-database analytics with SQL Server ML services (R, Python)	Y	N
Jupyter Notebook Server with following kernels,	Y	Y
* R	Y	Y
* Python	Y	Y
* Julia	Y	Y
* PySpark	Y	Y
* Sparkmagic	N	Y (Ubuntu only)
* SparkR	N	Y
JupyterHub (Multi-user notebook server)	N	Y
JupyterLab (Multi-user notebook server)	N	Y (Ubuntu only)
Development tools, IDEs and Code editors		
* Visual Studio 2017 (Community Edition) with Git Plugin, Azure HDInsight (Hadoop), Data Lake, SQL Server Data tools, Node.js , Python , and R Tools for Visual Studio (RTVS)	Y	N
* Visual Studio Code	Y	Y
* RStudio Desktop	Y	Y
* RStudio Server	N	Y
* PyCharm Community Edition	N	Y
* Atom	N	Y
* Juno (Julia IDE)	Y	Y
* Vim and Emacs	Y	Y
* Git and GitBash	Y	Y
* OpenJDK	Y	Y
* .Net Framework	Y	N

TOOL	WINDOWS EDITION	LINUX EDITION
PowerBI Desktop	Y	N
SDKs to access Azure and Cortana Intelligence Suite of services	Y	Y
Data Movement and management Tools		
* Azure Storage Explorer	Y	Y
* Azure CLI	Y	Y
* Azure Powershell	Y	N
* Azcopy	Y	N
* Blob FUSE driver	N	Y
* Adlcopy(Azure Data Lake Storage)	Y	N
* DocDB Data Migration Tool	Y	N
* Microsoft Data Management Gateway : Move data between OnPrem and Cloud	Y	N
* Unix/Linux Command-Line Utilities	Y	Y
Apache Drill for Data exploration	Y	Y
Machine Learning Tools		
* Integration with Azure Machine Learning (R, Python)	Y	Y
* Xgboost	Y	Y
* Vowpal Wabbit	Y	Y
* Weka	Y	Y
* Rattle	Y	Y
* LightGBM	N	Y (Ubuntu only)
* H2O , Sparkling Water , Deep Water	N	Y (Ubuntu only)
Deep Learning Tools		
All tools will work on a GPU or CPU		

TOOL	WINDOWS EDITION	LINUX EDITION
* Microsoft Cognitive Toolkit (CNTK) (Windows 2016)	Y	Y
* TensorFlow	Y (Windows 2016)	Y
* Horovod	N	Y (Ubuntu)
* MXNet	Y (Windows 2016)	Y
* Caffe & Caffe2	N	Y
* Chainer	N	Y
* Torch	N	Y
* Theano	N	Y
* Keras	N	Y
* PyTorch	N	Y
* NVidia Digits	N	Y
* MXNet Model Server	N	Y
* TensorFlow Serving	N	Y
* TensorRT	N	Y
* CUDA, cuDNN, NVIDIA Driver	Y	Y
Big Data Platform (Devtest only)		
* Local Spark Standalone	Y	Y
* Local Hadoop (HDFS, YARN)	N	Y