

# **2018 Predictive Analytics Symposium**

## **Session 20: B/I - Data Science in the Cloud in Under an Hour**

[SOA Antitrust Compliance Guidelines](#)

[SOA Presentation Disclaimer](#)

# Data Science in the Cloud in Under an Hour

Guy Yollin and Aabhas Gupta

2019 SOA PREDICTIVE ANALYTICS SYMPOSIUM



## Important notices

- The views expressed in this presentation are those of the presenter, and are not necessarily shared by Milliman, Inc., Microsoft, Inc., or the Society of Actuaries
- This presentation is not sponsored or approved by Microsoft, Inc.

# Agenda

- Introduction
- Getting onto Azure
- Starting up the Data Science Virtual Machine
- Connecting to the Data Science Virtual Machine
- Q & A

# Introduction

# The public cloud

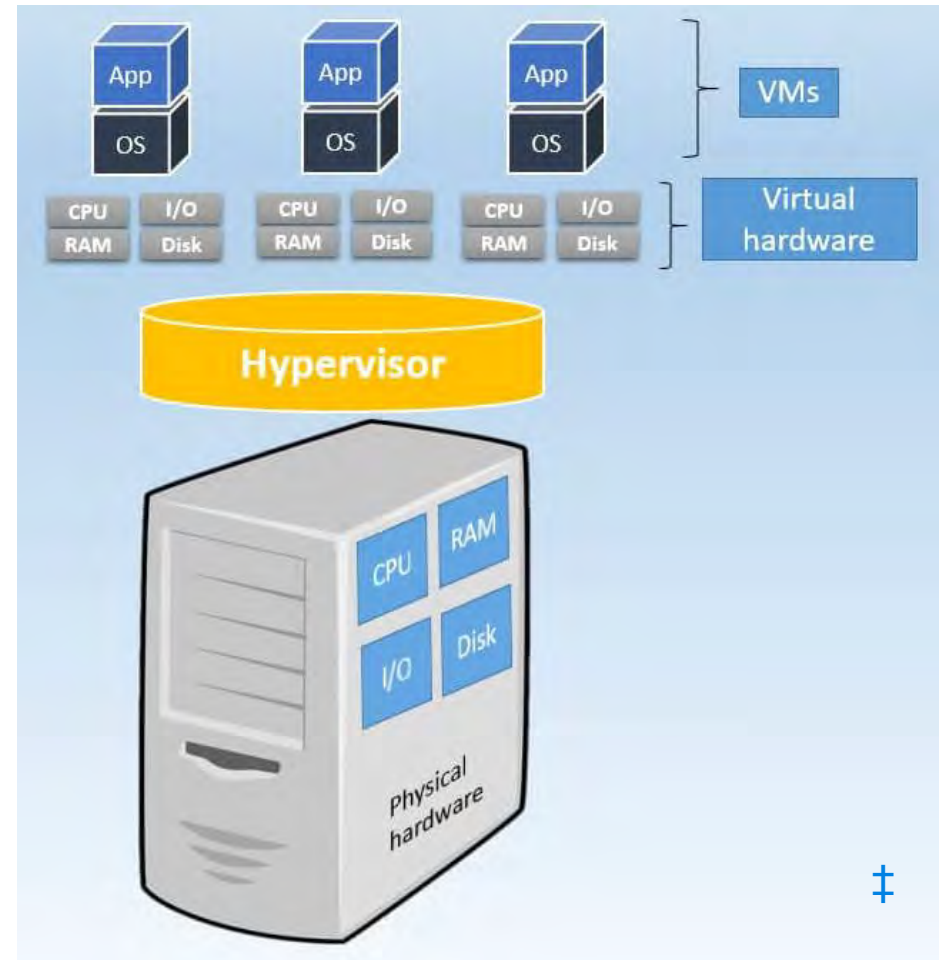
- Computing services accessible through the public internet
- Pay-for-usage pricing
  - CPU / RAM
  - Storage
  - Network bandwidth
- All of the big 3 offer free trials or free tiers

## Public Cloud Big 3



# Virtual machines

A computer file (typically called an image) that behaves like an actual computer. Multiple virtual machines can run simultaneously on the same physical computer †



† <https://azure.microsoft.com/en-us/overview/cloud-computing-dictionary/>

‡ <https://www.nakivo.com/blog/physical-servers-vs-virtual-machines-key-differences-similarities/>

# The Azure Data Science Virtual Machine (DSVM)

The Data Science Virtual Machine (DSVM) is a customized VM image on the Azure cloud platform built specifically for doing data science. It has many popular data science and other tools pre-installed and pre-configured to jump-start building intelligent applications for advanced analytics.

The tool configurations are rigorously tested by data scientists and developers at Microsoft and by the broader data science community. This testing helps ensure stability and general viability.

The DSVM is available on:

- Windows Server 2016, Windows Server 2012
- Ubuntu 16.04 LTS and CentOS 7.4

## ⓘ Note

All VM tools for deep learning have been folded into the Data Science Virtual Machine.

<https://docs.microsoft.com/en-us/azure/machine-learning/data-science-virtual-machine/overview>



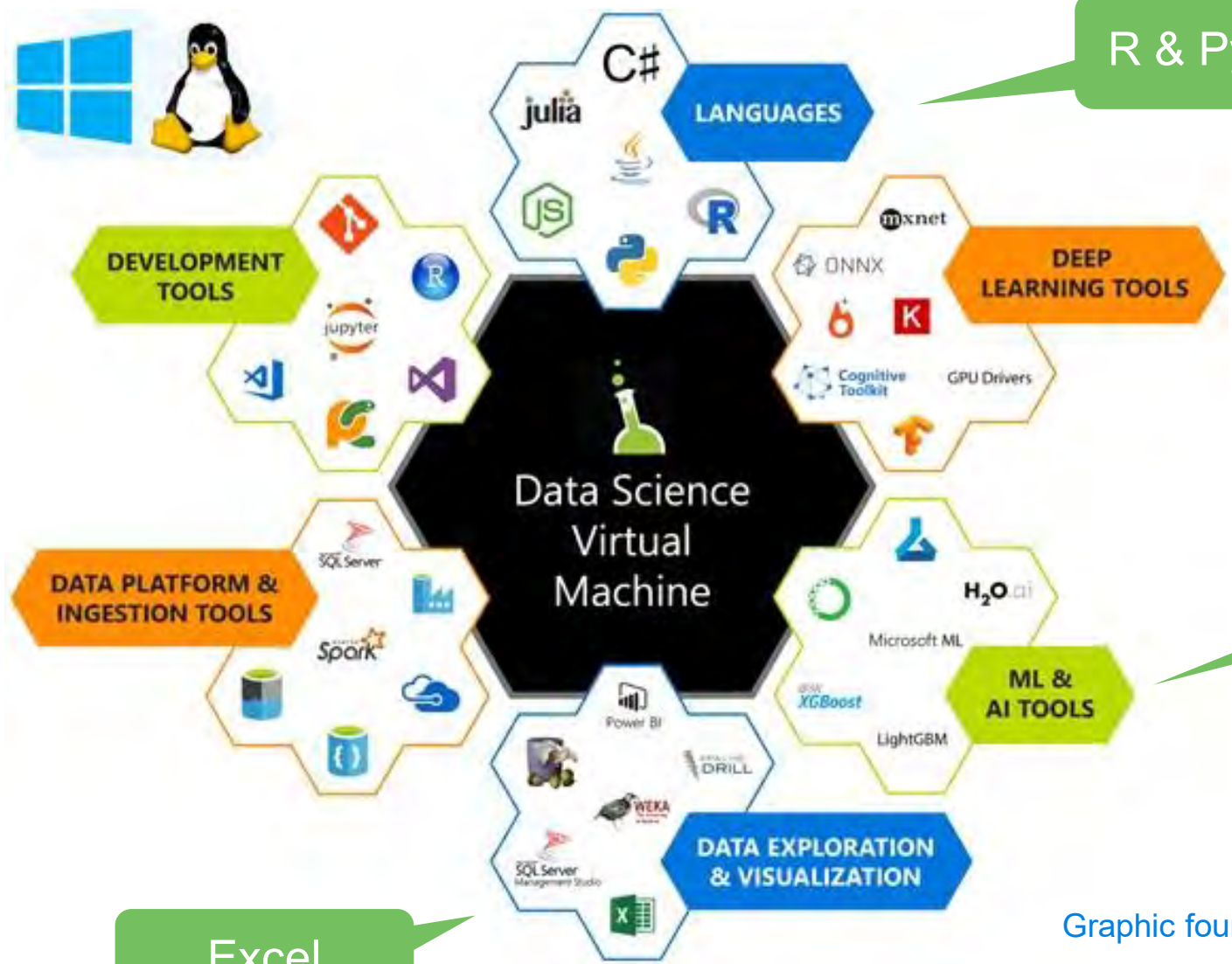
# The Azure Data Science Virtual Machine (DSVM)

Available in Windows or Linux



R & Python

RStudio



H2O & XGBoost

Excel

Graphic found on microsoft.com

# Tutorial roadmap



# Getting onto Azure

# Google Search

Google:  
data science vm azure

The screenshot shows a Google search for "data science vm azure". The search bar contains the text "data science vm azure". Below the search bar, there are navigation links for "All", "News", "Images", "Videos", "Shopping", "More", "Settings", and "Tools". The search results show "About 7,620,000 results (0.45 seconds)". The top result is an advertisement for Microsoft Azure VM, titled "Microsoft® Azure VM | Create Your Free Account Today". Below this are two smaller ads: "What is Azure?" and "Azure Service Fabric". The next result is "Study Data Science | A Fast-Growing Industry | galvanize.com". The following result is "Using H2O with Microsoft Azure Linux Data Science VM - H2O.ai". The final result, highlighted with a red box, is "Data Science Virtual Machines | Microsoft Azure" with the URL "https://azure.microsoft.com/.../virtual-machines/data-science-virtual-machin...". Below the red box, it says "You've visited this page 3 times. Last visit: 8/20/19".

Select this link

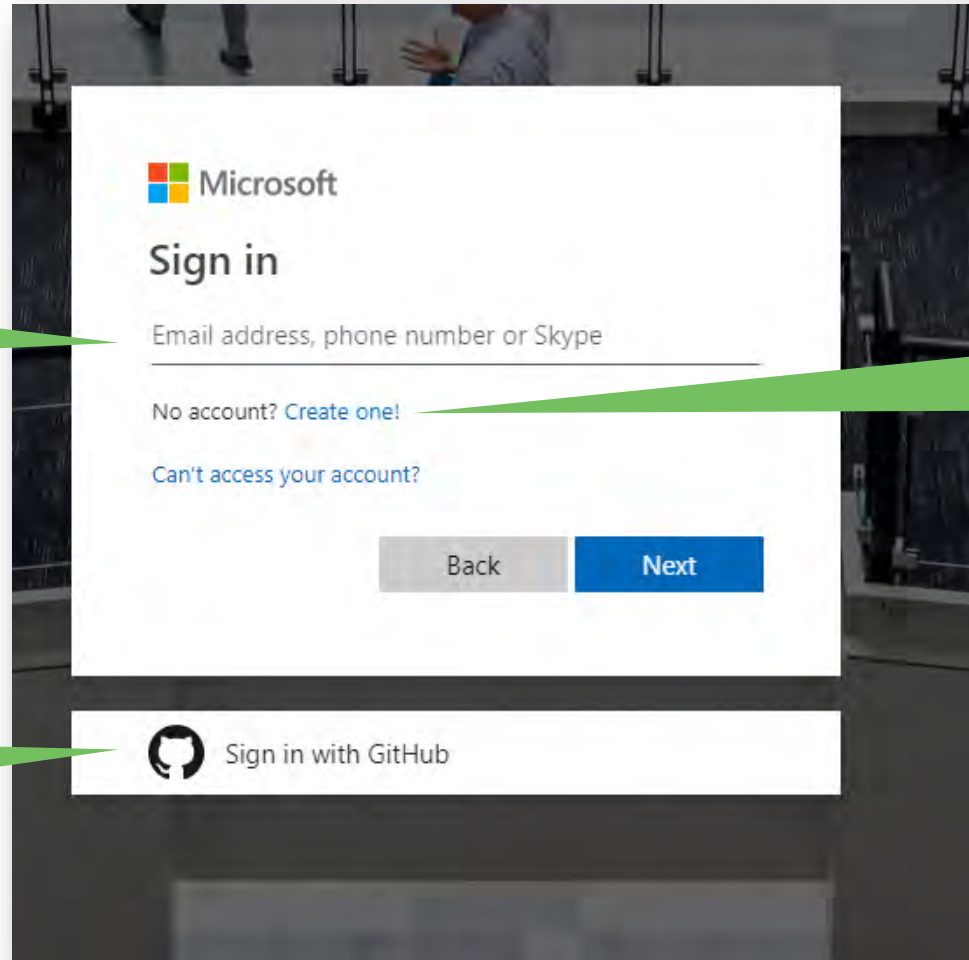
# Create a Free Account

Click:  
Start free

The screenshot shows the Microsoft Azure website interface. At the top, the Microsoft Azure logo is on the left, and navigation links for 'Overview', 'Solutions', 'Products', 'Documentation', 'Pricing', 'Training', 'Marketplace', 'Partners', 'Support', 'Blog', and 'More' are in the center. On the right, there are links for 'Contact Sales: 1-855-270-0615', 'Search', 'My account', 'Portal', and 'Sign in'. Below the navigation is a dark banner for 'Data Science Virtual Machines' with the text 'Comprehensive pre-configured virtual machines for data science modelling, development and deployment.' A green callout box with a red border points to a 'Start free >' button. Below the button is a link: 'Already using Azure? Try Data Science Virtual Machines now >'. At the bottom of the banner are links for 'Explore Data Science Virtual Machines', 'Pricing details', and 'Documentation'. Below the banner is a section titled 'Pre-Configured environments in the cloud for Data Science and AI Development'. The text reads: 'DSVMs are Azure Virtual Machine images, pre-installed, configured and tested with several popular tools that are commonly used for data analytics, machine learning and AI training.' Below this text is a diagram showing various tools categorized into 'DEVELOPMENT TOOLS', 'LANGUAGES', and 'DEEP LEARNING TOOLS'. The 'DEVELOPMENT TOOLS' category includes icons for Windows, Linux, Jupyter, and Visual Studio. The 'LANGUAGES' category includes icons for Julia, C#, JavaScript, and R. The 'DEEP LEARNING TOOLS' category includes icons for ONNX, TensorFlow, Keras, Cognitive Toolkit, and GPU Drivers. The central part of the diagram is labeled 'Data Science Virtual'.

# Sign in with a Microsoft Account (or GitHub)

You'll need a Microsoft account



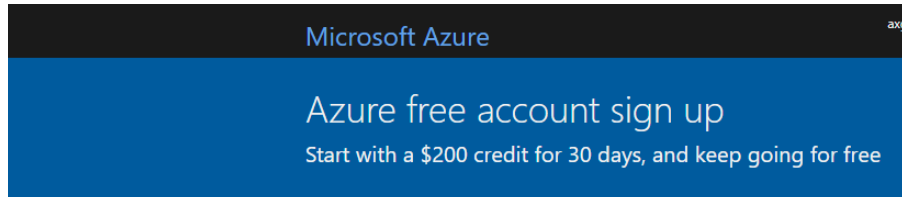
The screenshot shows the Microsoft sign-in interface. At the top is the Microsoft logo. Below it is the heading "Sign in". There is a text input field labeled "Email address, phone number or Skype". Below the input field are two links: "No account? Create one!" and "Can't access your account?". At the bottom of the sign-in section are two buttons: "Back" and "Next". Below the sign-in section is a separate button for "Sign in with GitHub" with the GitHub logo to its left.

You can create a Microsoft account if you don't already have one

Or you can sign in with GitHub



# Setup a free Azure account



## 1 About you

Country/Region ⓘ

United States ▼

First name

This field is required

Last name

This field is required

Email address ⓘ

Phone

Example: (425) 555-0100

By proceeding you acknowledge that if you use your organization's email, your organization may have rights to access and manage your data and account. [Learn more](#)

Next

## 2 Identity verification by phone

## 3 Identity verification by card

Enter some basic info

## What's included

- ✓ **12 months of free products**  
Get free access to popular products like *virtual machines, storage, and databases* in your first 30 days, and for 12 months after you upgrade your account to pay-as-you-go pricing.
- ✓ **\$200 credit**  
Use your \$200 to experiment with any Azure service in your first 30 days—beyond the free product amounts.
- ✓ **25+ always-free products**  
Take advantage of more than 25 products, including *serverless, containers, and artificial intelligence* that are always free. Get these in your first 30 days, and always—once you choose to upgrade.
- ✓ **No automatic charges**  
You won't be charged unless you choose to upgrade. Before the end of your first 30 days, you'll be notified and have the chance to upgrade and start paying only for the resources you use beyond the free amounts.

Here's what you get when you signup

Note: No automatic charges

# Enter credit card details and accept subscription agreement

You must enter a credit card to sign up but it won't be charged unless you upgrade

1 Identity verification by card ∨

---

**2 Agreement** ∧

---

I agree to the [subscription agreement](#), [offer details](#), and [privacy statement](#).

I will receive information, tips, and offers from Microsoft or selected partners about Azure, including Azure Newsletter, Pricing updates, and other Microsoft products and services.

Setting up your account...



# Success! You just signed up for free cloud services on Azure

Click here to go to the Azure portal

Go to the portal >

You're ready to start with Azure, **User**!

Your actual name will appear here

## Join the demo to see Azure in action

Learn how to use the portal to create and manage resources and get your questions answered by Azure technical experts during a **live Q&A**. This demo covers:

- Building a virtual machine.
- Creating a web app.
- Deploying a SQL database.
- Customizing your DevOps dashboard.

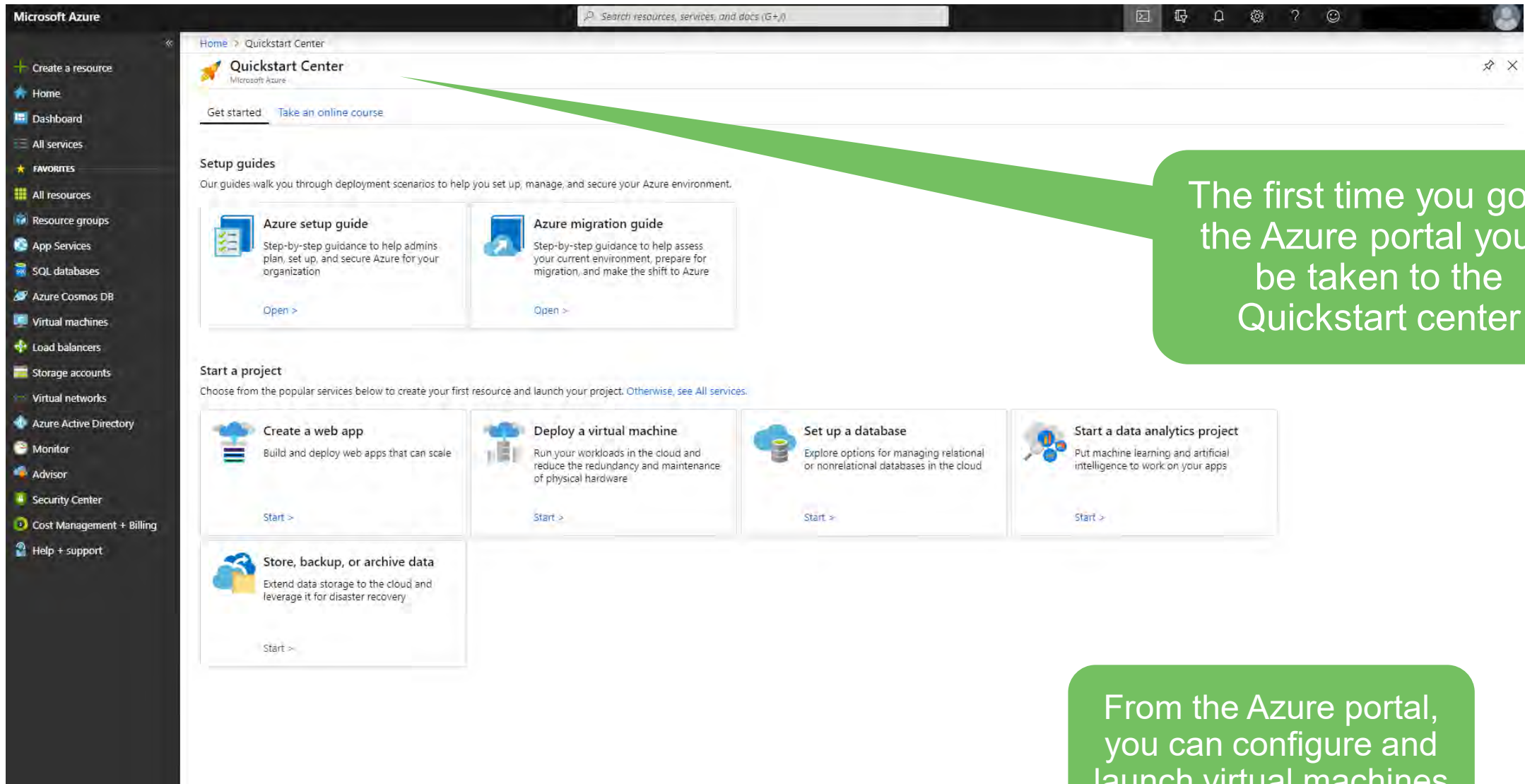
## Register to schedule a live demo

### Thank you for registering!

You will receive a confirmation email shortly. It includes a link to join the demo. If your copy of the email does not appear in your inbox in the next five minutes, please

You can also register for a live demo

# Azure portal – Quickstart center

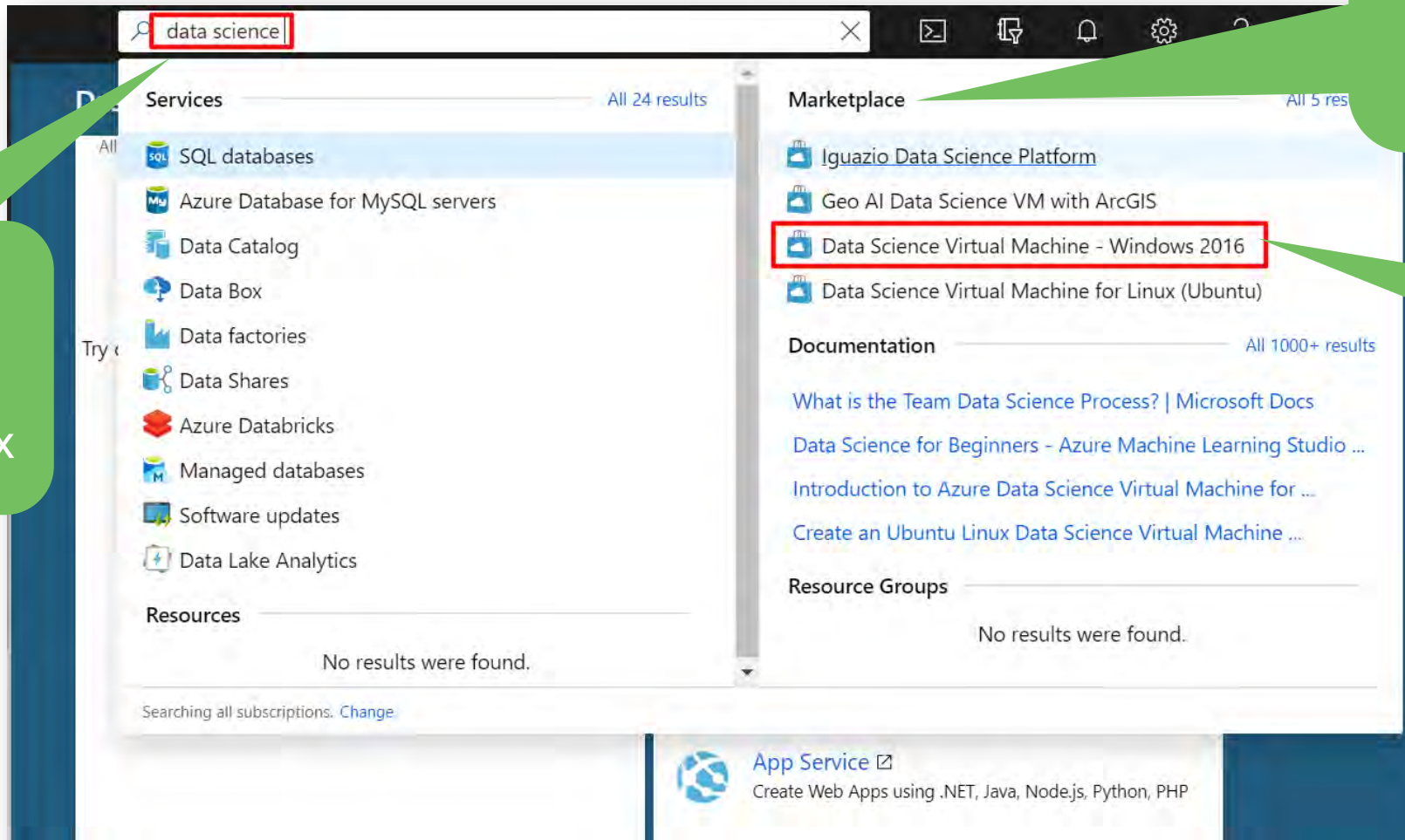


The first time you go to the Azure portal you'll be taken to the Quickstart center

From the Azure portal, you can configure and launch virtual machines

# Starting up the data science virtual machine

# Find the Data Science Virtual Machine



Search for the Data Science Virtual Machine in the search box

The different Data Science Virtual Machines images will appear under Marketplace

Choose the one for windows

# Windows DSVM versus Linux DSVM


- Both the Windows and Linux DSVMs have many popular data science and other tools pre-installed. However, some non-cross-platform tools will only be in their native environment
  - Windows
    - Microsoft office and other tools
  - Linux
    - RStudio Server and other tools
- How to access the DSVM
  - Windows
    - Remote Desktop Connection
  - Linux
    - SSH for terminal sessions
    - X2Go for graphical sessions
    - JupyterHub and JupyterLab for Jupyter notebooks

The convenience of Remote Desktop Connection and the ease of mapping local drives to the remote system is the primary reason we suggest the Windows DSVM

A Remote Desktop Connection app also exists for the Mac platform

# Ready to create DSVM

**Data Science Virtual Machine - Windows 2016**  
Microsoft



**Data Science Virtual Machine - Windows 2016** [Save for later](#)

Microsoft

[Create](#) [Start with a pre-set configuration](#)

The '**Data Science Virtual Machine (DSVM)**' is a 'Windows Server 2016 with Containers' VM & includes popular tools for data exploration, analysis, modeling & development.

Highlights:

- [Azure Machine Learning SDK](#)
- [Microsoft ML Server - Dev Edition](#) (Scalable R & Python)
- Anaconda Python
- SQL Server 2017 Dev. Edition - With In-Database R and Python analytics
- Microsoft Office 365 ProPlus BYOL - Shared Computer Activation
- Julia Pro + Juno Editor
- Jupyter notebooks
- Visual Studio Community Ed. + Python, R & node.js tools
- Power BI Desktop
- Deep learning tools e.g. Microsoft Cognitive Toolkit (CNTK, TensorFlow, Chainer, & mxnet)
- ML algorithm libraries e.g. xgboost, Vowpal Wabbit
- Azure SDKs + libraries for various Azure Cloud offerings. Integration tools are included for:

Click this option to minimize configuration questions



# General purpose options have been pre-selected

**Choose recommended defaults that match your workload**

To quickly customize your virtual machine, choose one of the following pre-set configurations. You can modify these configurations at any time.

Select a workload environment

Dev/Test	General default	Production/Mission Critical
<input checked="" type="checkbox"/> Boot diagnostics	<input checked="" type="checkbox"/> Boot diagnostics	<input checked="" type="checkbox"/> Boot diagnostics
<input type="checkbox"/> Premium disk	<input checked="" type="checkbox"/> Premium disk	<input checked="" type="checkbox"/> Premium disk
<input type="checkbox"/> High availability	<input type="checkbox"/> High availability	<input checked="" type="checkbox"/> High availability
<input type="checkbox"/> Azure backup (where available)	<input type="checkbox"/> Azure backup (where available)	<input checked="" type="checkbox"/> Azure backup (where available)

Select a workload type

General purpose (D-Series) default	Memory optimized (E-Series)	Compute optimized (F-Series)
<input checked="" type="checkbox"/> Example sizes DS2_v2: 2 CPU, 7 GB DS3_v2: 4 CPU, 14 GB	<input type="checkbox"/> Example sizes E2s_v3: 2 CPU, 16 GB E4s_v3: 4 CPU, 32 GB	<input type="checkbox"/> Example sizes F2s_v2: 2 CPU, 4 GB F4s_v2: 4 CPU, 8 GB
<input checked="" type="checkbox"/> Fast CPUs with optimal CPU-to-memory configuration	<input type="checkbox"/> High memory-to-core ratio optimized for heavy in-memory applications	<input type="checkbox"/> High CPU-to-memory ratio optimized for compute intensive workloads
<input checked="" type="checkbox"/> Workload types Enterprise applications, relational databases, analytics	<input type="checkbox"/> Workload types SAP HANA, SQL Hekaton, other large in-memory workloads	<input type="checkbox"/> Workload types Batch processing, web servers, gaming

**Continue to create a VM** Skip this step

Just click to continue

# VM Sizes

Default is:  
2 cores / 8GB RAM;  
cost \$0.15-\$0.20 per hour

Showing 11 of 188 VM sizes. | Subscription: Free Trial | Region: West US | Current size: Standard\_D2s\_v3

VM SIZE	OFFERING	FAMILY	VCPUS	RAM (GI...	DATA DISKS	MAX IOPS	TEMPORARY ST...	NUM DISK ...	COST/MONTH (...
B1ls	Standard	General purpose	1	0.5	2	200	1		US\$8.04
B1ms	Standard	General purpose	1	2	2	640	4	Yes	US\$21.43
B1s	Standard	General purpose	1	1	2	3200	2	Yes	US\$12.20
B2ms	Standard	General purpose	2	8	4	1920	16	Yes	US\$79.61
B2s	Standard	General purpose	2	4	4	1280	8	Yes	US\$42.85
B4ms	Standard	General purpose	4	16	8	2880	32	Yes	US\$159.22
D2s_v3	Standard	General purpose	2	8	4	3200	16	Yes	US\$155.50
D4s_v3	Standard	General purpose	4	16	8	6400	32	Yes	US\$310.99
DS1_v2	Standard	General purpose	1	3.5	4	2200	7	Yes	US\$93.74
DS2_v2	Standard	General purpose	2	7	8	4400	14	Yes	US\$187.49

Can upgrade to:  
4 cores / 16GB RAM;  
cost \$0.30-\$0.40 per hour



# Resource Groups

A Resource group is a container that holds related services; e.g. virtual machine, virtual network, network security group, etc.

**Create a virtual machine**

Basics | Disks | Networking | Management | Advanced | Tags | Review + create

Create a virtual machine that runs Linux or Windows. Select an image from Azure marketplace or use your own customized image. Complete the Basics tab then Review + create to provision a virtual machine with default parameters or review each tab for full customization. Looking for classic VMs? [Create VM from Azure Marketplace](#)

**Project details**

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

\* Subscription

\* Resource group

**Instance details**

\* Virtual machine name

\* Region

Availability options

\* Image

\* Size  [Change size](#)

D-series is recommended for general purpose workloads.

**Modal Dialog:**

A resource group is a container that holds related resources for an Azure solution.

\* Name:

Create a new Resource Group

# Azure regions

VMs can be deployed at data centers throughout the world:  
example:  
East US  
Central US  
North Central US  
South Central US  
West US

The screenshot shows the 'Basics' tab of the Azure portal for creating a virtual machine. The 'Region' dropdown is set to '(US) West US'. The 'Size' dropdown is set to 'Standard D2s v3', which is highlighted with a red box. A red arrow points from this box to a message: 'This particular configuration was not available in the (US) West US 2 region.' A black tooltip is also visible over the 'Virtual machine name' field, stating: 'Choose the Azure region that's right for you and your customers. Not all VM sizes are available in all regions.'

Choose a nearby region for best performance

Not all VM types are available in all regions

# Configuration review

**Create a virtual machine**

⚠ Changing Basic options may reset selections you have made. Review all options prior to creating the virtual machine. your resources.

\* Subscription

\* Resource group   
[Create new](#)

**Instance details**

\* Virtual machine name

\* Region

Availability options

\* Image   
[Browse all public and private images](#)

\* Size   
2 vcpus, 8 GiB memory  
[Change size](#)  
D-series is recommended for general purpose workloads.

**Administrator account**

\* Username

\* Password

\* Confirm password

**Save money**  
Save up to 49% with a license you already own using Azure Hybrid Benefit. [Learn more](#)

\* Already have a Windows Server license?  Yes  No

[Review + create](#) [Next : Disks >](#)

Enter a username and password for administration

Click here and the configuration will be reviewed



# Validation results

Validate passed

## Create a virtual machine

✓ Validation passed

Basics Disks Networking Management Advanced Tags Review + create

### PRODUCT DETAILS

Standard D2s v3

by Microsoft

[Terms of use](#) | [Privacy policy](#)

Subscription credits apply ⓘ

0.1620 USD/hr

[Pricing for other VM sizes](#)

This VM will cost \$0.16/hour to run (prices may vary)

### TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

### Basics

Subscription	R&D
Resource group	(new) soapas
Virtual machine name	soapas
Region	(US) West US 2
Availability options	No infrastructure redundancy required
Username	gyollin
Already have a Windows Server license?	No

### Disks

OS disk type	Premium SSD
Use managed disks	Yes
Use ephemeral OS disk	No

### Networking

Click create

Create

< Previous

Next >

[Download a template for automation](#)

# Your deployment is underway

Dashboard > CreateVm-microsoft-dsvm.dsvm-windows-server-2016-20190820161328 - Overview

## CreateVm-microsoft-dsvm.dsvm-windows-server-2016-20190820161328 - Overview

Deployment:

Search (Ctrl+/)

Delete Cancel Redeploy Refresh

### Your deployment is underway

Deployment name: CreateVm-microsoft-dsvm.dsvm-windows-serv... Start time: 8/20/2019, 4:27:52 PM  
Subscription: [Free Trial](#) Correlation ID: c62c9fd8-0651-4c49-b77b-f09682405910  
Resource group: [dsvm-rg](#)

Deployment details (Download)

RESOURCE	TYPE	STATUS	OPERATION DETAILS
No results.			

Next steps

The deployment will take about 5 minutes and you can watch as the various pieces become operational in the resource group

# Deployment complete – the VM is running

The screenshot shows the Azure portal interface for a deployment named "CreateVm-microsoft-dsvm.dsvm-windows-server-2016-20190820161328 - Overview". The deployment is complete, and the resources created are listed in a table below.

RESOURCE	TYPE	STATUS	OPERATION DETAILS
<a href="#">dsvm</a>	Microsoft.Compute/virt...	OK	<a href="#">Operation details</a>
<a href="#">dsvm114</a>	Microsoft.Network/netw...	Created	<a href="#">Operation details</a>
<a href="#">dsvm-rg-vnet</a>	Microsoft.Network/virtu...	OK	<a href="#">Operation details</a>
<a href="#">dsvm-nsg</a>	Microsoft.Network/netw...	OK	<a href="#">Operation details</a>
<a href="#">dsvm-ip</a>	Microsoft.Network/publ...	OK	<a href="#">Operation details</a>
<a href="#">dsvmrgdiag893</a>	Microsoft.Storage/stora...	OK	<a href="#">Operation details</a>

Below the table, there is a "Next steps" section with a "Go to resource" button.

Click here to go to the new resource

Here's what got created in the resource group:

- virtual machine
- network interface
- virtual network
- network security group
- public IP address
- storage account

**Connecting to  
data science  
virtual machine**



# The virtual machine resource page

Click here for details on connecting to the VM

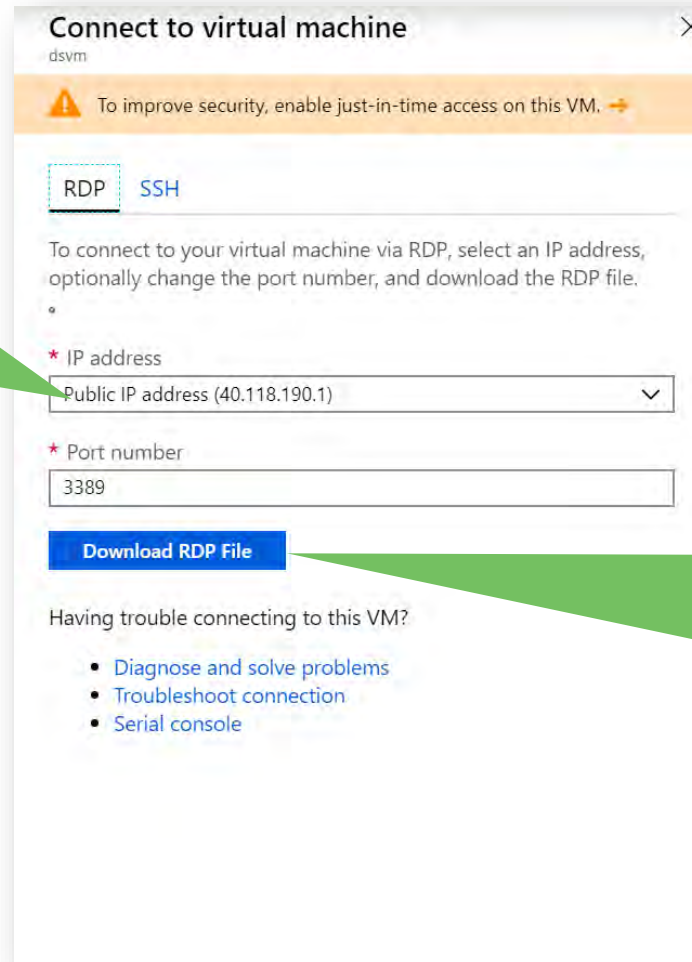
The screenshot shows the Azure portal interface for a virtual machine named 'dsvm'. The left sidebar contains navigation options such as Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and various settings like Networking, Disks, Size, Security, Extensions, Continuous delivery (Preview), Availability set, Configuration, Identity, Properties, Locks, and Export template. The main content area displays the VM's status as 'Running' and provides details for its resource group (dsvm-rg), location (West US), and subscription (Free Trial). It also lists network-related information like public and private IP addresses and DNS name. A 'Connect' button is highlighted with a red box. Below the details, there is a section for monitoring CPU usage, with a graph showing 'CPU (average)' over time, currently set to '1 hour'.

You can monitor and adjust the VM from this dashboard



# Establishing a Remote Desktop Connection

This public IP along with your admin username and password is all you need to connect



The screenshot shows a window titled "Connect to virtual machine" with a close button in the top right corner. Below the title bar is a warning banner: "To improve security, enable just-in-time access on this VM. →". There are two tabs: "RDP" (selected) and "SSH". Below the tabs is a text box: "To connect to your virtual machine via RDP, select an IP address, optionally change the port number, and download the RDP file." Below this is a list of options, with "Public IP address (40.118.190.1)" selected. There are two required fields: "IP address" (containing the selected public IP) and "Port number" (containing "3389"). Below these fields is a blue button labeled "Download RDP File". At the bottom, there is a section titled "Having trouble connecting to this VM?" with three links: "Diagnose and solve problems", "Troubleshoot connection", and "Serial console".

Downloading the RDP file can make it even easier to connect but it may not work with your enterprise networking setup (this doesn't work for me)

# Open Remote Desktop Connection

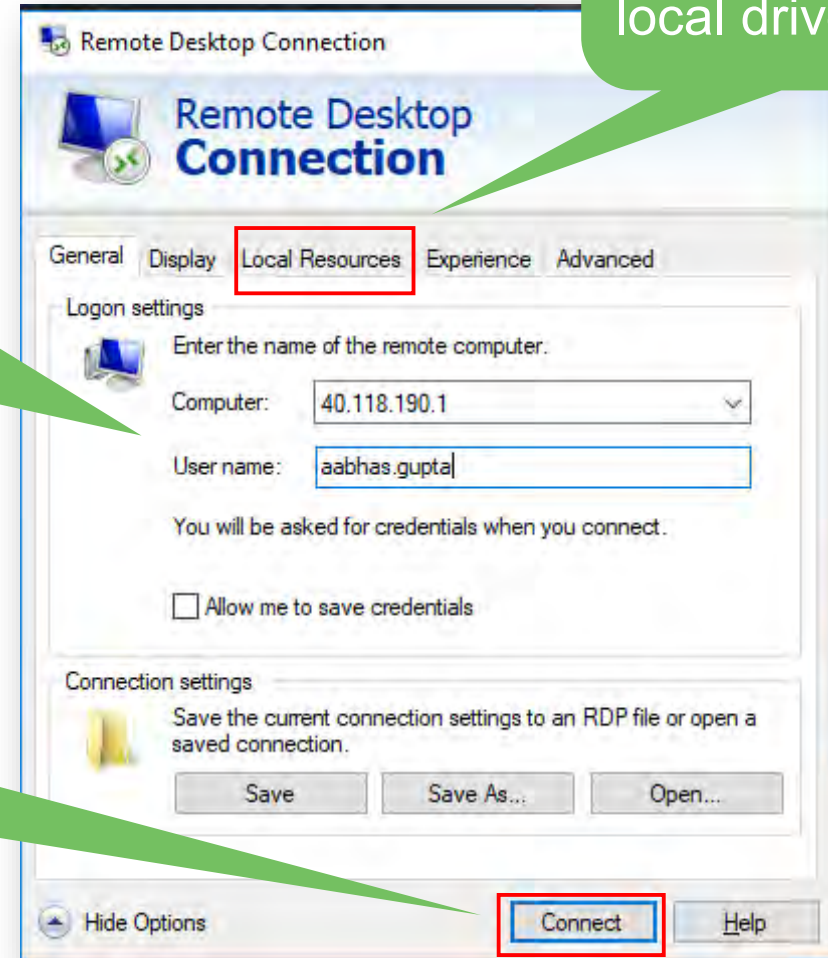
Run Remote Desktop Connection app



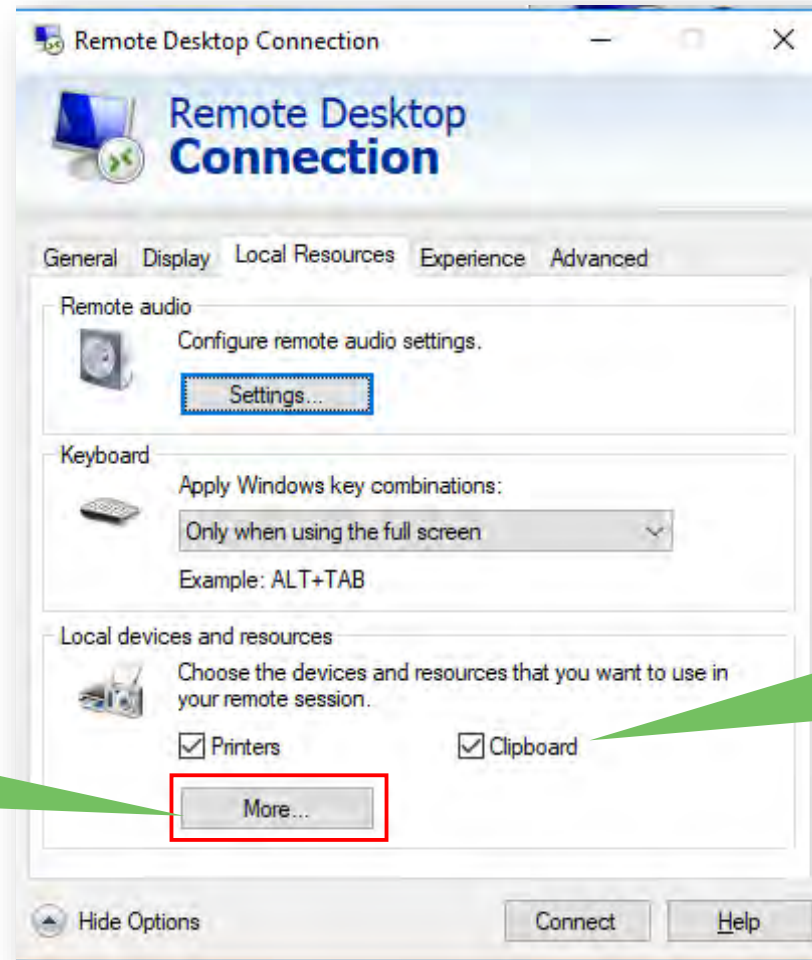
Enter the IP address of the VM along with your username

After sharing local drives click Connect

Before you click connect click here to share local drives



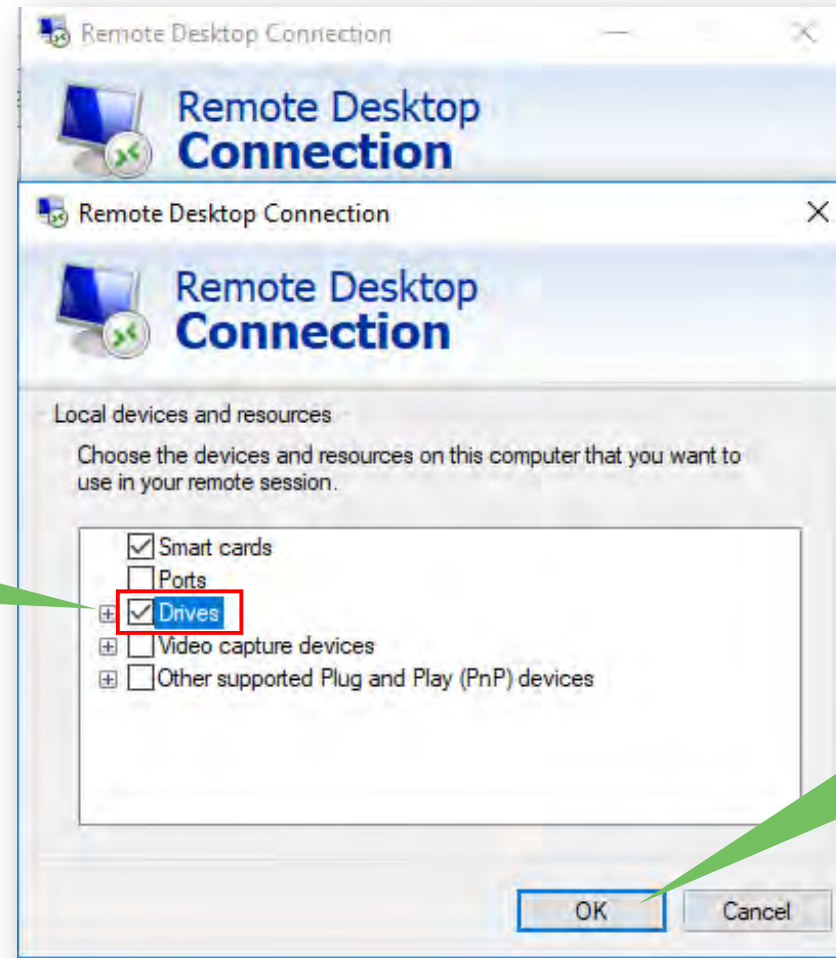
# Sharing local resources



Click More to  
add local drives

Sharing the  
clipboard is the  
default and  
very helpful

# Sharing local drives

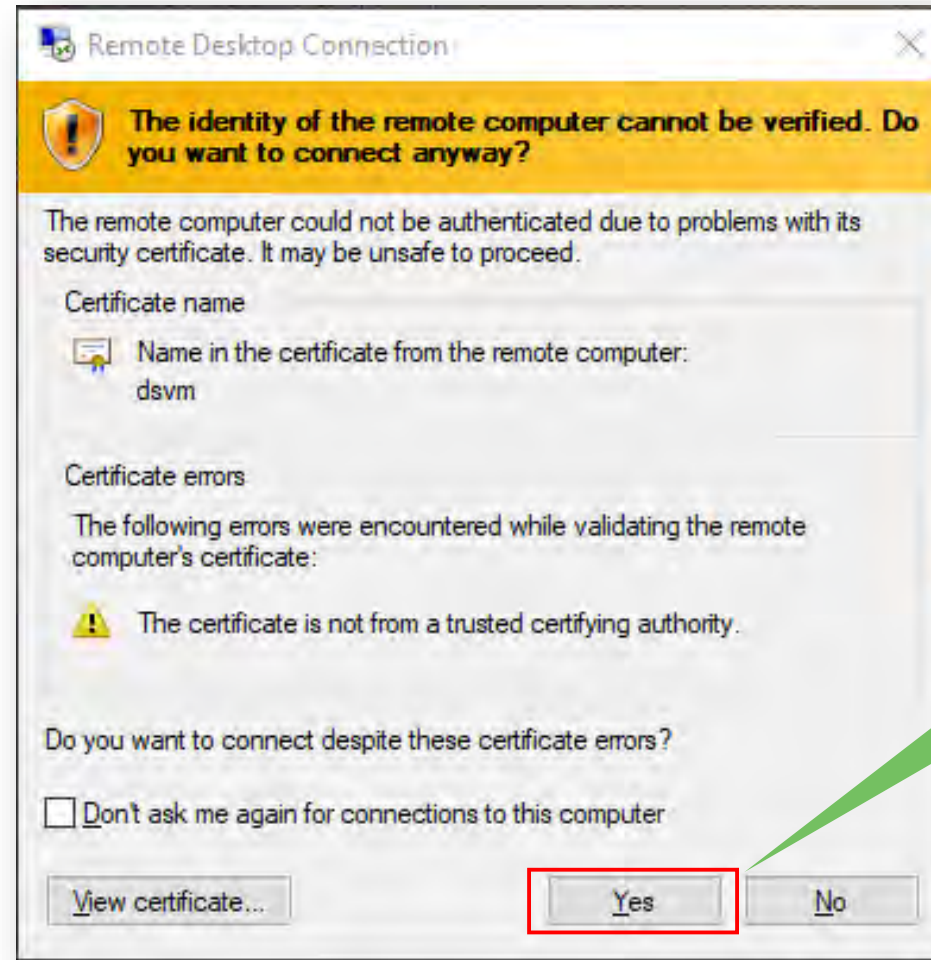


Check here to share local drives

Click OK then Connect and you'll be prompted for the password

# Security Certificate

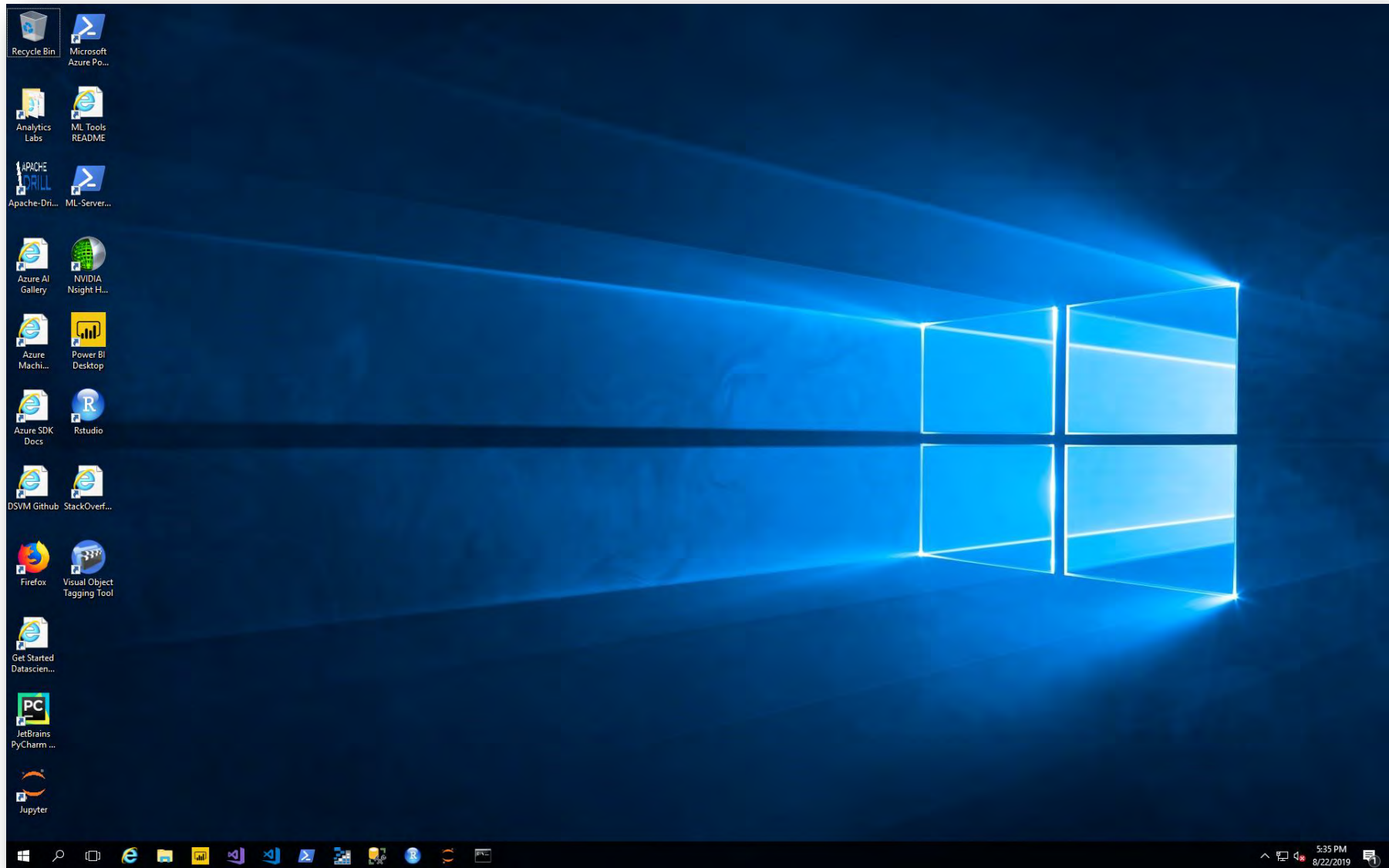
This is a newly created server that does not have a certificate from a Certificate Authority; this is expected behavior; nothing unsafe



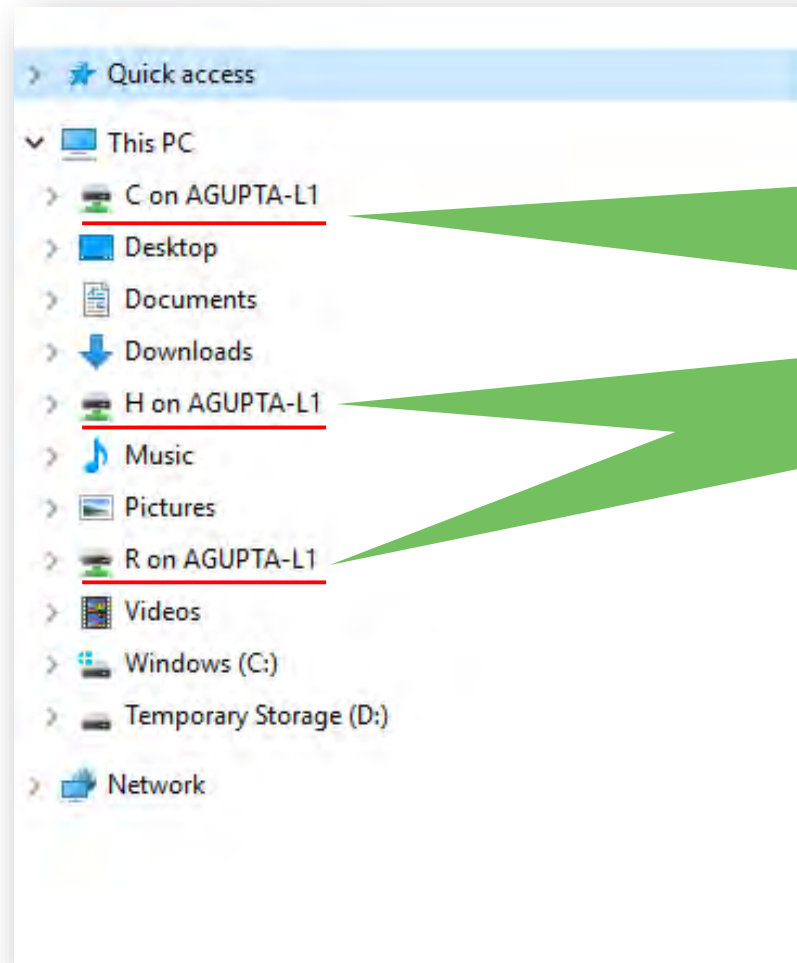
Click Yes to connect



# The desktop of the Windows Server 2016 virtual machine



# Local drive are accessible from the VM



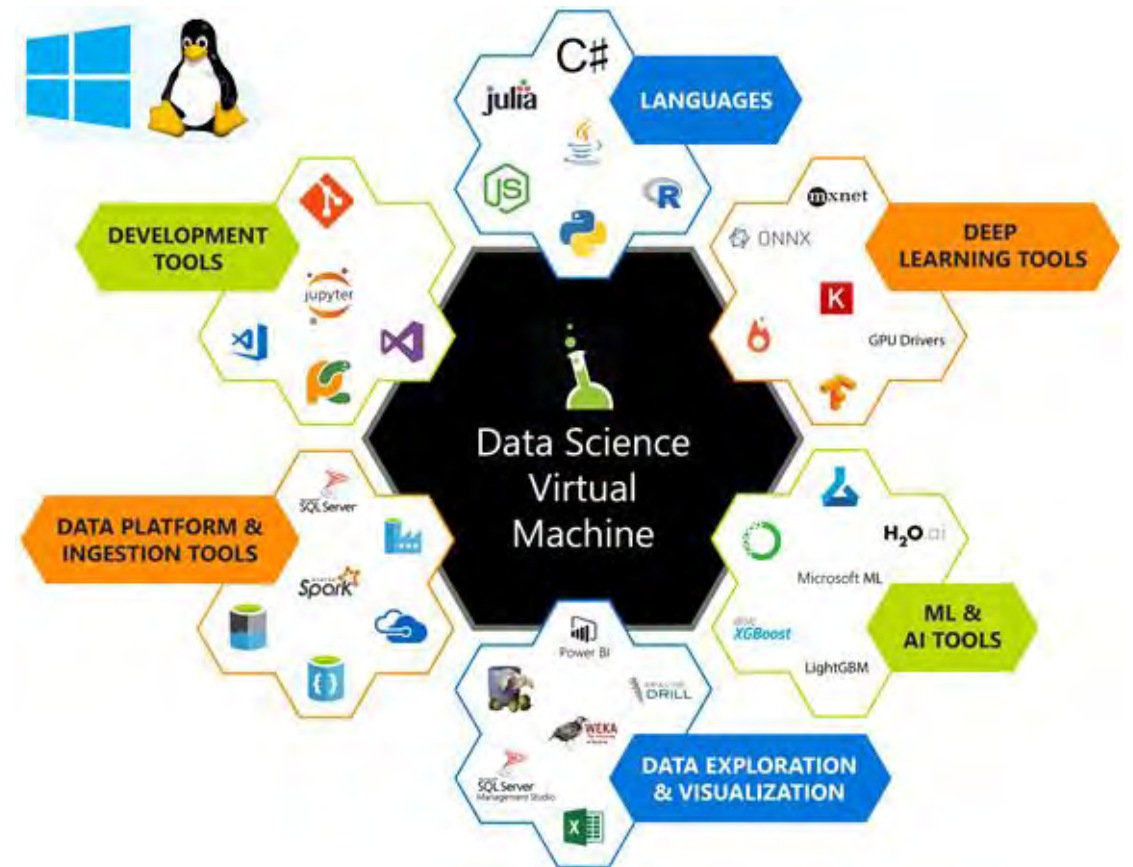
Files can be easily copied between the local computer and the virtual machine using file explorer or other windows utilities

# Data science in the cloud





# Questions & Answers





# Thank you

**Guy Yollin**

[guy.yollin@milliman.com](mailto:guy.yollin@milliman.com)