Student Guide

January 2018

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The names of manufacturers, products, or URLs are provided for informational purposes only and Microsoft makes no representations and warranties, either expressed, implied, or statutory, regarding these manufacturers or the use of the products with any Microsoft technologies. The inclusion of a manufacturer or product does not imply endorsement of Microsoft of the manufacturer or product. Links may be provided to third party sites. Such sites are not under the control of Microsoft and Microsoft is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites. Microsoft is not responsible for webcasting or any other form of transmission received from any linked site. Microsoft is providing these links to you only as a convenience, and the inclusion of any link does not imply endorsement of Microsoft of the site or the products contained therein.

© 2018 Microsoft Corporation. All rights reserved.

Microsoft and the trademarks listed at <https://www.microsoft.com/en-us/legal/intellectualproperty/Trademarks/Usage/General.aspx> are trademarks of the Microsoft group of companies. All other trademarks are property of their respective owners.

Contents

[Big data and visualization whiteboard design session student guide 1](#_Toc504517723)

[Abstract and learning objectives 1](#_Toc504517724)

[Step 1: Review the customer case study 2](#_Toc504517725)

[Step 2: Design a proof of concept solution 9](#_Toc504517726)

[Step 3: Present the solution 11](#_Toc504517727)

[Wrap-up 12](#_Toc504517728)

[Additional references 13](#_Toc504517729)

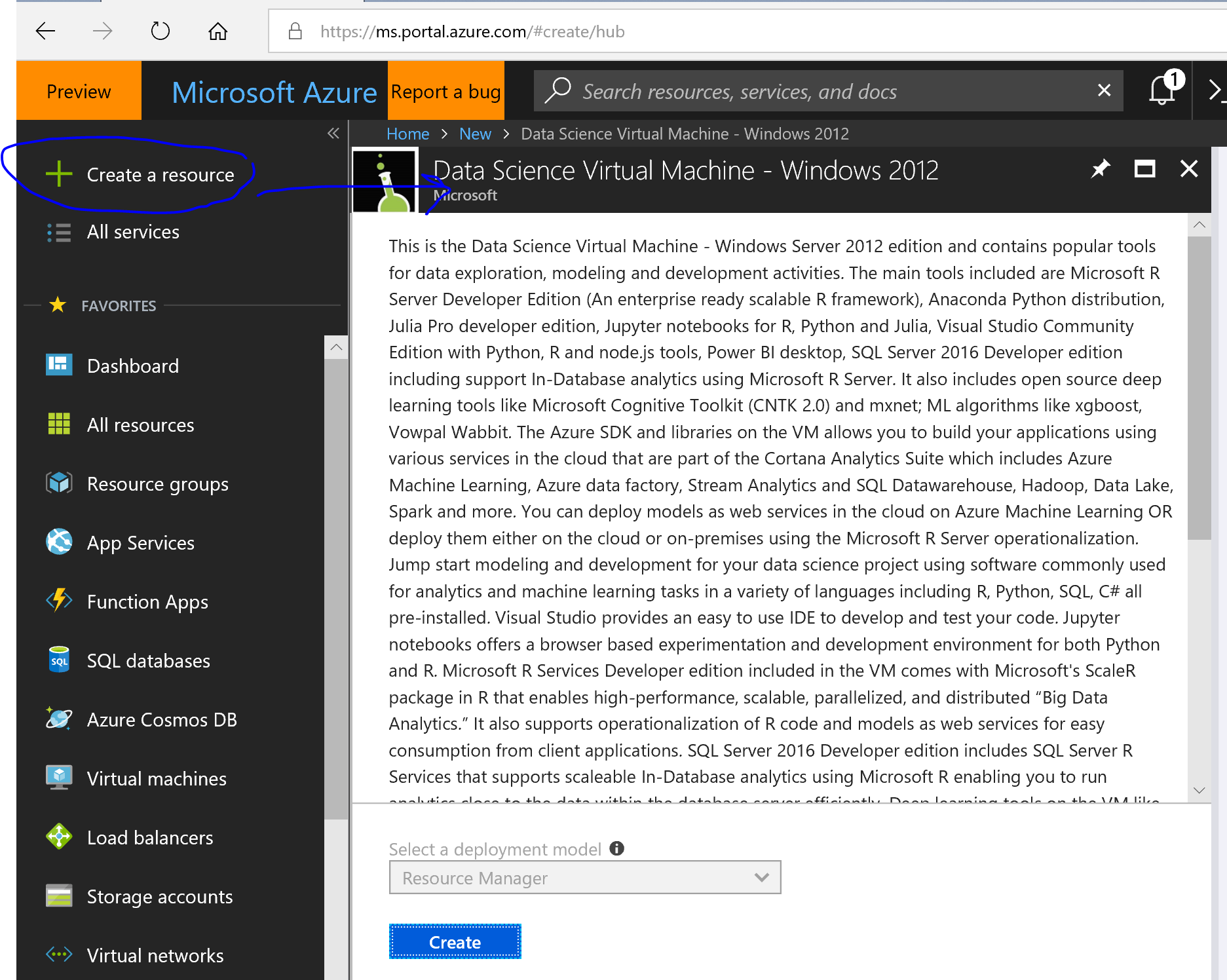
# Machine Learning in a Day student guide

## Abstract and learning objectives

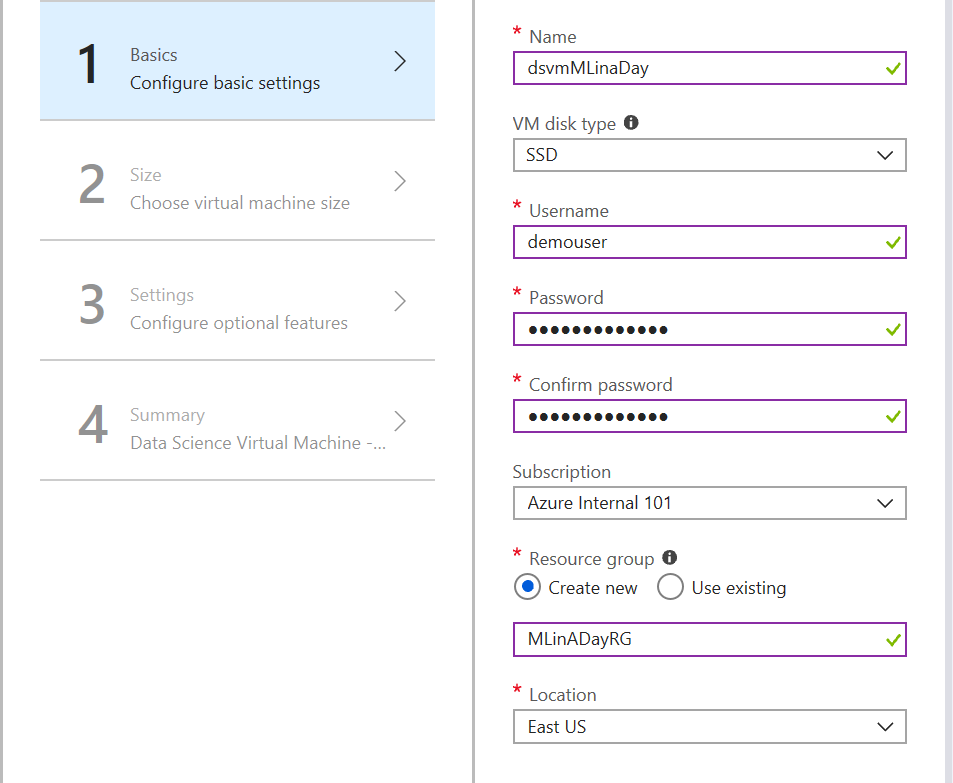
In this workshop, you will complete a web app using Machine Learning to predict travel delays given flight delay data and weather conditions, plan the bulk data import operation, followed by preparation tasks, such as cleaning and manipulating the data for testing, and training your Machine Learning model.

## Step 1: Set up a Data Science Virtual Machine (DSVM) on Azure Portal

1. Visit [www.portal.azure.com](http://www.portal.azure.com) and click on **Create a resource** on the left-hand side and type in **Data Science Virtual Machine – Windows 2012** and select the **Create** button as seen in the following screenshot:

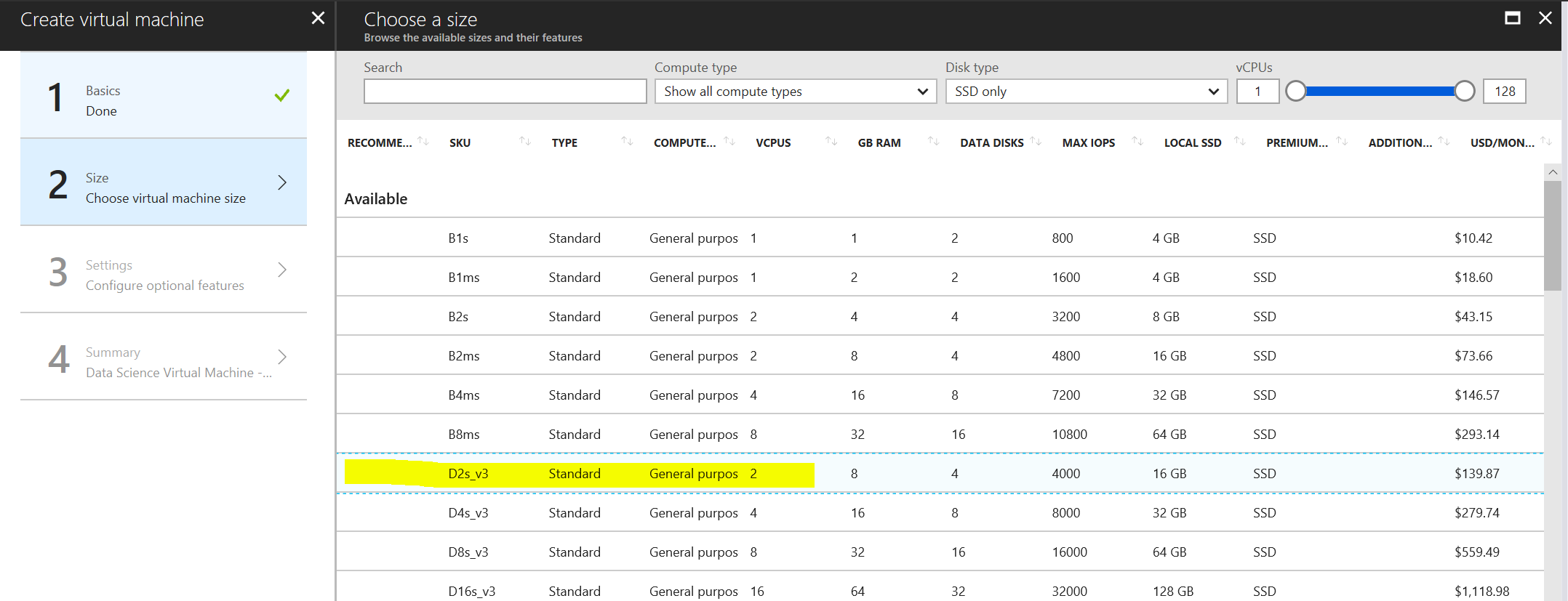


1. Assign the following basic configuration settings for your Data Science Virtual Machine and the select the **OK** button as seen in the following screenshot:

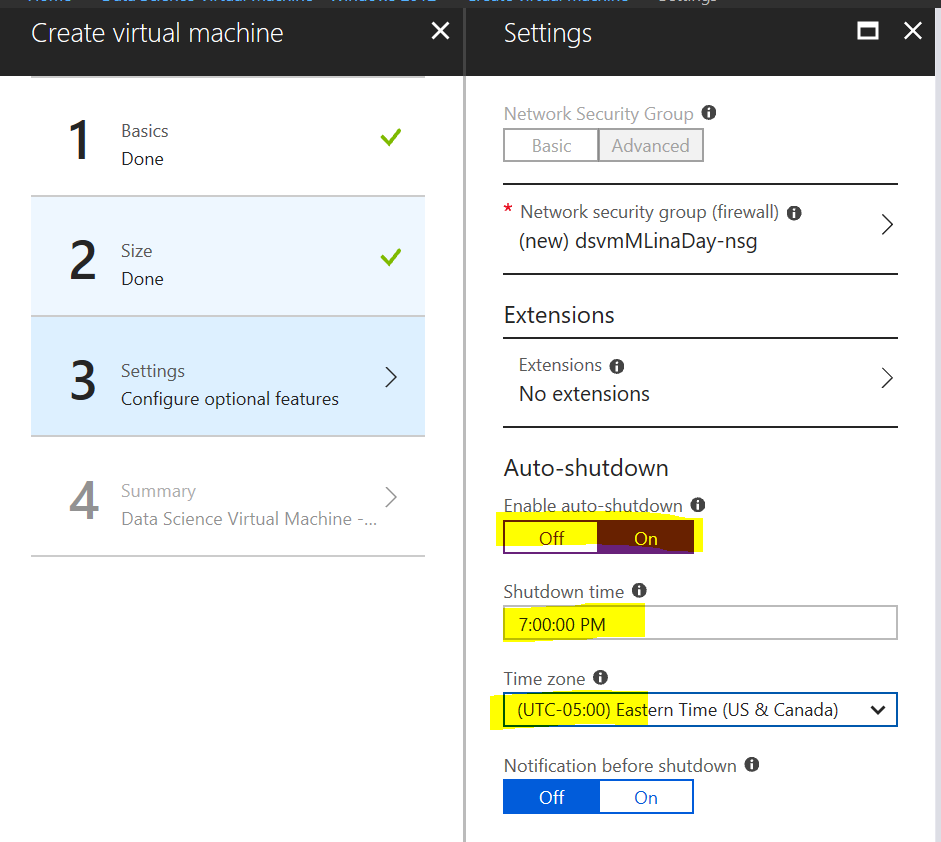


Please note that you can keep your Names and passwords different from this document as long as you remember what you used for future use. In this section, we will assign a username of **demouser** and a password of **#MLinaDay20!8**

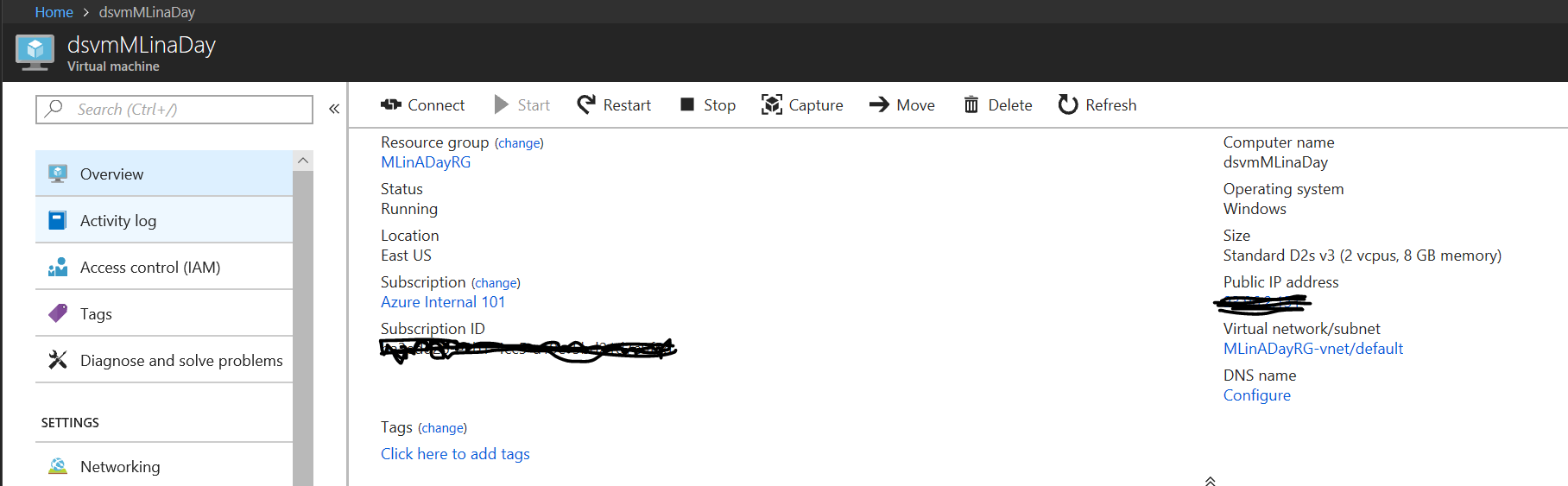
1. Choose the following size VM (**D2S\_v3**) for the purposes of this workshop as seen in the following screenshot:



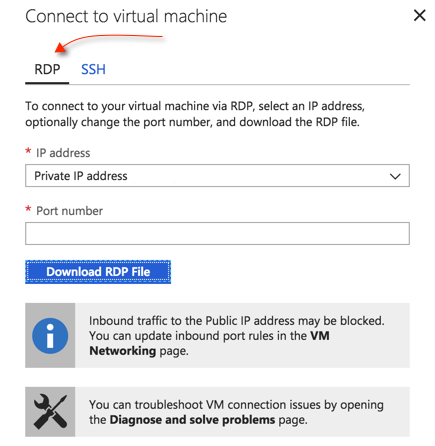
1. For settings, the only configuration that is truly needed is to enable on **auto-shutdown** every evening at **7pm EST** and then select **OK** as seen in the following screenshot:



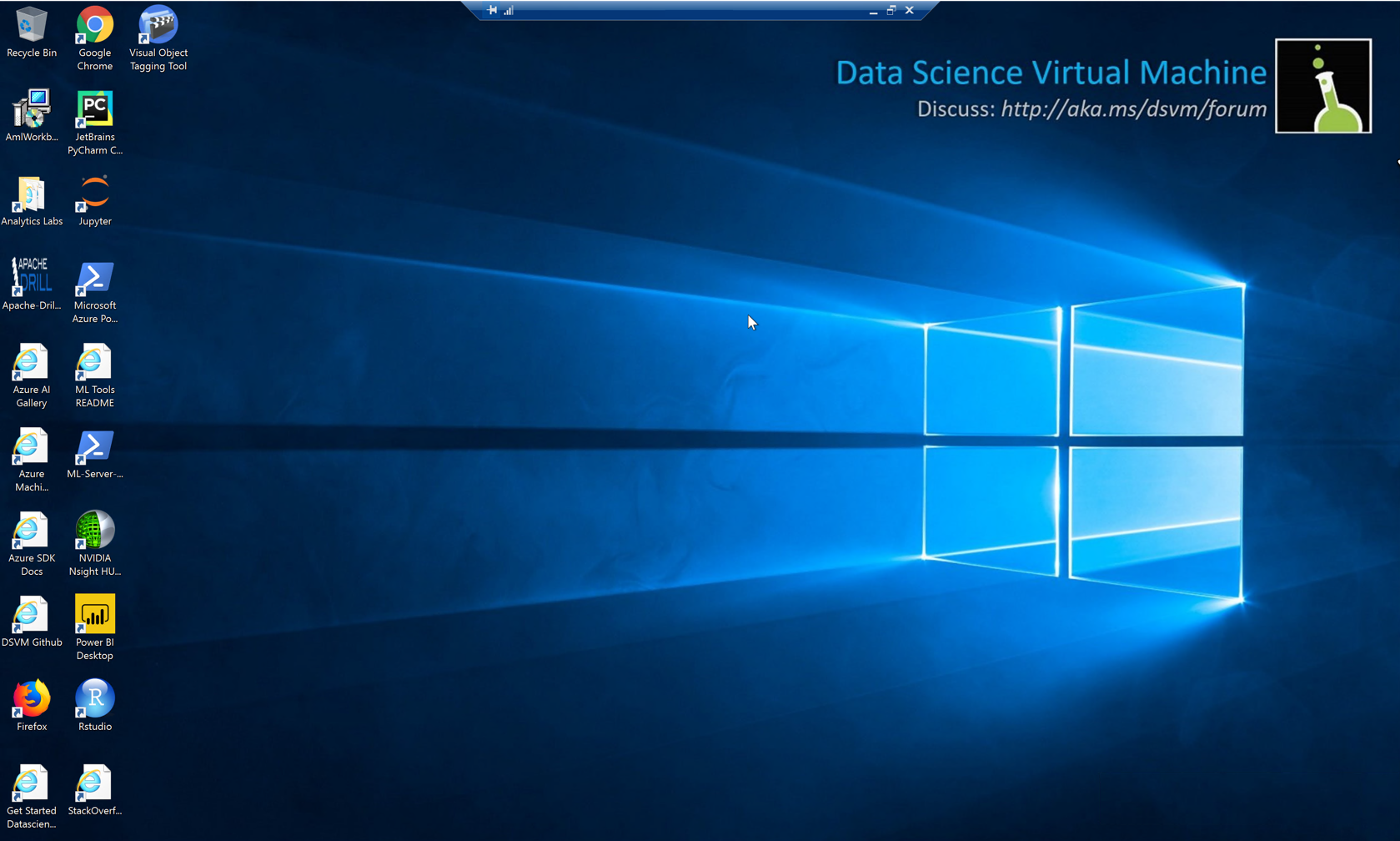
1. Click **OK** on the summary section and confirm everything that was selected is accurate and let the provisioning process of the DSVM begin.
2. Once the provisioning is complete, you can go to your resource and view your DSVM overview as seen in the following screenshot:



1. Make sure that your DSVM is currently running and click on the **Connect** button to enter your virtual machine using your credentials as seen in the following screenshot:



1. Once you log in using your credentials using the RDP approach, you should see the following remote server for the data science virtual machine as seen in the following screenshot:



## Step 2: Download Telco dataset and upload dataset to Blob Storage using Azure Data Factory

## Wrap-up

Timeframe: 15 minutes

* Tables reconvene with the larger group to hear a SME share the preferred solution for the case study.

## Additional references

|  |  |  |
| --- | --- | --- |
| **Item** | **Description** | **Links** |
| Infographic | Hi-resolution version of data analytics blueprint | [https://msdn.microsoft.com/dn630664 - fbid=rVymR\_3WSRo](https://msdn.microsoft.com/dn630664#fbid=rVymR_3WSRo) |
| Machine Learning | Azure ML algorithm cheat sheet | <https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-cheat-sheet/> |
| Azure Data Factory | What is Azure Data Factory? | <https://docs.microsoft.com/azure/data-factory/introduction> |
| HDInsight Spark | Overview: Apache Spark on HDInsight | <https://azure.microsoft.com/en-us/documentation/articles/hdinsight-apache-spark-overview/> |
| Power BI | Power BI overview | <https://support.powerbi.com/knowledgebase/articles/430814-get-started-with-power-bi> |
| Travel data | Sample data source  Bureau of Transportation Statistics, United States Department of Transportation  Database: Airline On-Time Performance Data Table: On-Time Performance Table | <http://www.transtats.bts.gov/Tables.asp?DB_ID=120> |
| Weather data | Sample REST API for weather forecasts | <http://www.wunderground.com/weather/api/d/docs> |
| ARM Templates | Understand the structure and syntax of ARM templates | <https://docs.microsoft.com/azure/azure-resource-manager/resource-group-authoring-templates> |