

Microsoft Azure Data Lake: Publishing Datasets in Data Catalog

Updated on: 12/27/2016

Introduction

In this lab, you will learn how to register data sources stored in Azure Data Lake Store in Azure Data Catalog.

Takeaways

By the end of this lab, you will be able to:

- Discover assets registered in Azure Data Catalog
- Register your own assets into the Azure Data Catalog

Prerequisites

For this lab you will need:

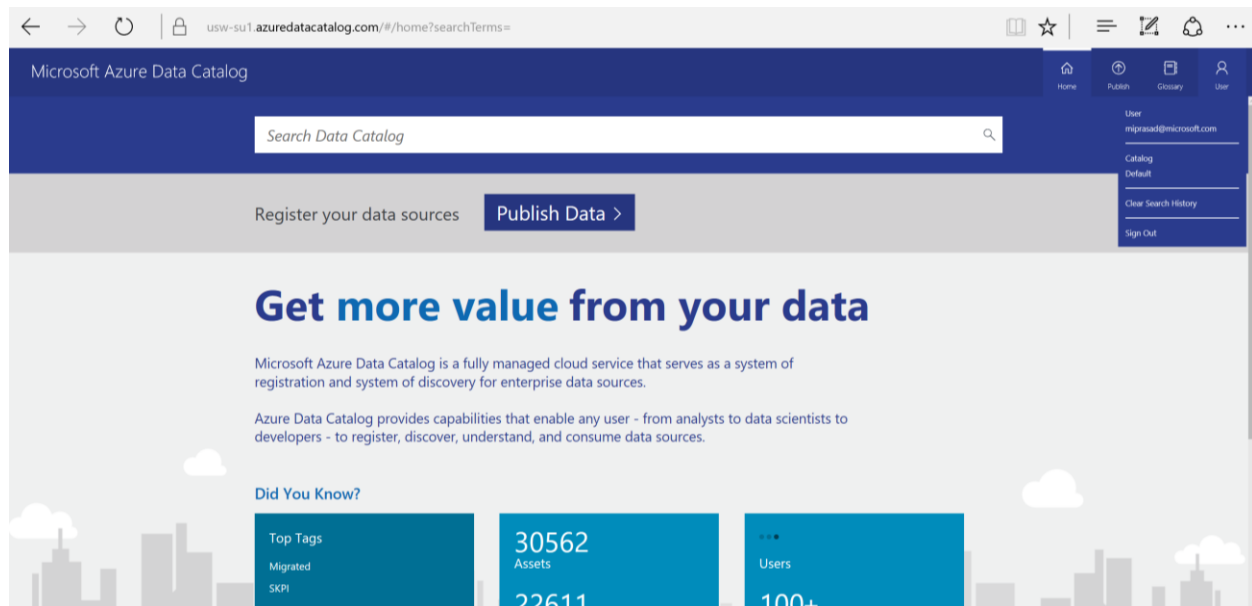
- Access to an Azure Data Lake Store account
- Access to an Azure Data Catalog instance (one is available for all Microsoft Employees)

Class

Exercise 1: Browse the Azure Data Catalog and look for data stored in Azure Data Lake Store

In this exercise, you will explore the Azure Data Catalog to discover data sources related to Azure Data Lake Store.

- 1) Go to the Azure Data Catalog portal: www.azuredatacatalog.com
- 2) After signing in, you should be able to see the below page:

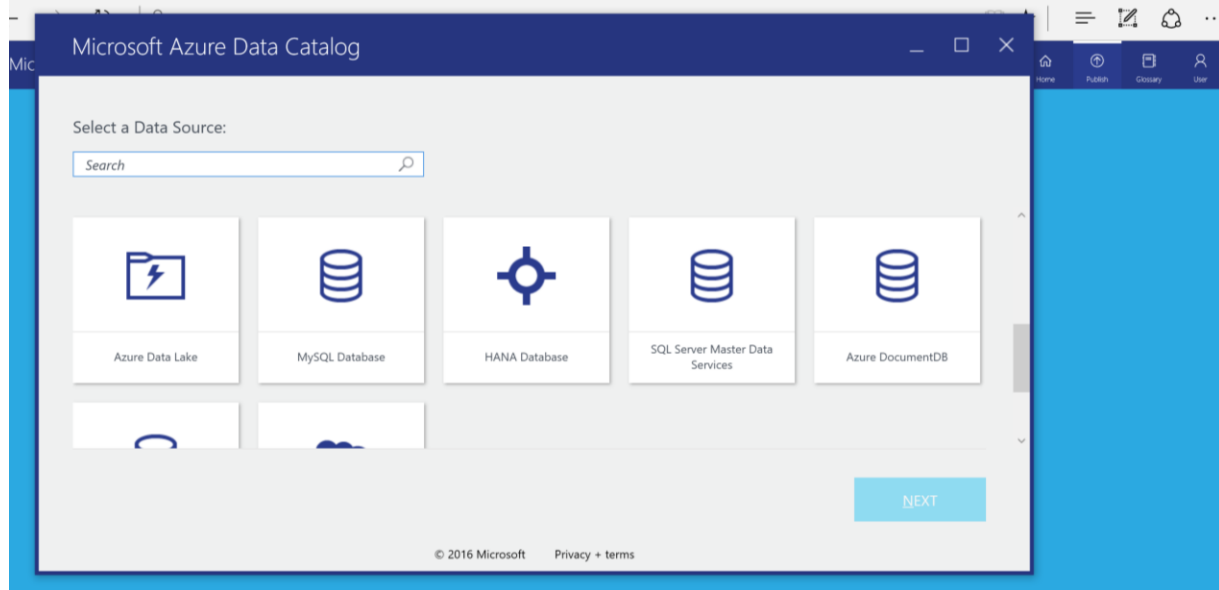


- 3) Search the Data Catalog. Select Azure Data Lake Store in the filter 'Source Type'

Exercise 2: Register a data set you're working with in the Azure Data Catalog

In this exercise, you will register your own data set into the Azure Data Catalog so other people can discover it and derive value from it.

- 1) Go to the Azure Data Catalog portal: www.azuredatacatalog.com
- 2) Click Publish Data
- 3) Click on "Launch Application" to download and run the Click-once desktop application for Azure Data Catalog
- 4) Sign-in when the application launches. You will see the below window.



- 5) Click Azure Data Lake
- 6) Type in the name of an ADLS account you have access to
- 7) Browse to your data set using the navigation tree
- 8) Double click on the data to add to the list of objects to be registered
- 9) Fill in any metadata about the objects
 - a. It's important to provide tags or an expert so other people trying to discover data will know who to consult.
- 10) Click Register

Conclusion

This lab has hopefully highlighted how to use Azure Data Catalog in discovering and publishing datasets within an enterprise.

We hope you come back and use Azure Data Lake Analytics and U-SQL for your big data processing needs!