# Running the HDInsight Samples

A set of sample are provided to help you get started using the Windows Azure HDInsight service that can be run using Windows Azure PowerShell. These samples are made available on each of the Hadoop clusters that you create and manage with the Windows Azure HDInsight Service. Running these samples will familiarize you with the PowerShell scripts used to run MapReduce programs on clusters managed by the HDInsight Service.

MapReduce programs can also be run programmatically from an application using the Microsoft .NET API for Windows Azure HDInsight Service. For more information on using the HDInsight APIs for job submission, see [Submit Hadoop Jobs Programmatically](file:///C:\en-us\manage\services\hdinsight\submit-hadoop-jobs-programmatically\).

Much additional documentation exists on the web for Hadoop-related technologies such as Java-based MapReduce programming and streaming, as well as documentation on the cmdlets using in PowerShell scripting. For more information on these resources, see the final **Resources for the HDInsight Service** section of the [Introduction to Windows Azure HDInsight Service](file:///C:\en-us\manage\services\hdinsight\introduction-hdinsight\) topic.

**What These Samples Are**

These samples are intended to get you up to speed quickly on how to deploy Hadoop jobs and to provide you an extensible testing bed to work with the concepts and scripting procedures used by the service. They provide you with examples of common tasks such as creating and importing data sets of various sizes, running jobs and composing jobs sequentially, and examining the results of your jobs. The data sets used can be varied in size, allowing you to observe the effects that data sets of various size has on job performance.

**Prerequisites**:

* You must have a Windows Azure Account. For options on signing up for an account see [Try Windows Azure out for free](http://www.windowsazure.com/en-us/pricing/free-trial/) page.
* You must have provisioned an HDInsight cluster. For instructions on the various ways in which such clusters can be created, see [Provision HDInsight Clusters](file:///C:\en-us\manage\services\hdinsight\provision-hdinsight-clusters\)
* You must have installed Windows Azure PowerShell and the HDInsight PowerShell Tools, and have configured them for use with your account. For instructions on how to do this, see [Install and configure PowerShell for HDInsight](file:///C:\en-us\manage\services\hdinsight\configure-powershell-for-hdinsight\)

## The Samples

HDInsight ships with the following samples.

* [**The Hadoop on Azure Pi Estimator Sample**](file:///C:\en-us\manage\services\hdinsight\sample-pi-estimator\): This tutorial shows how to deploy a MapReduce program with Hadoop on Windows Azure that uses a statistical (quasi-Monte Carlo) method to estimate the value of Pi.
* [**The Hadoop on Azure WordCount Sample**](file:///C:\en-us\manage\services\hdinsight\sample-wordcount\): This tutorial shows how to use Hadoop on Windows Azure to run a MapReduce program that counts word occurrences in an uploaded text file.
* [**The Hadoop on Azure 10-GB Graysort Sample**](file:///C:\en-us\manage\services\hdinsight\sample-10gb-graysort\): This tutorial shows how to run a general purpose GraySort on a 10 GB file using Hadoop on Windows Azure. There are three jobs to run: Teragen to generate the data, Terasort to sort the data, and Teravalidate to confirm the data has been properly sorted.
* [**The Hadoop on Azure C# Streaming Sample**](file:///C:\en-us\manage\services\hdinsight\sample-csharp-streaming\):This tutorial shows how to use C# to write a MapReduce program that uses the Hadoop streaming interface.

## How to run the Samples

The samples can be run using Windows Azure PowerShell. Instructions on how to do this are provided for each of the samples on the pages linked above.

## Next Steps

From this article and the articles on each of the samples, you learned how to run the samples included with the HDInsight Service using Windows Azure PowerShell. For tutorials on using Pig, Hive, and MapReduce from the Windows Azure HDInsight Service, see the following topics:

* [Tutorial: Using Pig](file:///C:\en-us\manage\services\hdinsight\using-pig-with-hdinsight\)
* [Tutorial: Using Hive](file:///C:\en-us\manage\services\hdinsight\using-hive-with-hdinsight\)
* [Tutorial: Using MapReduce](file:///C:\en-us\manage\services\hdinsight\using-mapreduce-with-hdinsight\)
* [Submit Hadoop Jobs Programmatically](file:///C:\en-us\manage\services\hdinsight\submit-hadoop-jobs-programmatically\)