# Real World Big Data in Azure

**Module 4 Setup Instructions**

In Module 4 we analyze the events stored in Blob Storage using Apache Pig and HDInsight.

## Telemetry

This is the demo solution for the course. In this module we have the existing projects for the Web API and the Azure Cloud Service, and new projects for the DeepStorage component and Cloud Service.

### Pre-requisites

The demo solution is delivered in Visual Studio 2013, Update 4 and uses the Azure SDK version 2.5.

The solution uses NuGet package restore to load packages during the build.

To enable this in Visual Studio, open *Tools…Library Package Manager…Package Manager Settings* and ensure both options (*Allow NuGet to download missing packages* and *Automatically check for missing packages during build in Visual Studio*) are ticked.

If you want to run the sample queries on HDInsight, you will need an Azure account and you will need to preload sample data to Azure Blob Storage.

### Before

This is the solution code from the start of Module 4.

### After

By the end of the module we have added the **DeepStorage.Dashboard** project which uses D3 and C3 to visualize the output from a Pig query, and the **Pig.CountLogger** project which is used for streaming data from Pig.

### Assets

Sample Pig queries used in the module are in the **Assets\Pig** folder:

* *demo1-interactive.pig –* simple count entered into the interactive Pig shell
* *demo2-json.pig* – group by count, using local storage and the Twitter JSON loader
* *demo3-hdinsight.pig* – same query as demo2 but running over Azure Blob Storage
* *demo4-stream.pig –* streaming data from Pig into a .NET app
* *demo5-json-output.pig –* HDInsight count demo writing JSON output