# **Azure for Research FY14 – Global Training Curriculum/Syllabus**

**v1 – July 26, 2013**

*Wenming Ye (content lead)*

*Dennis Gannon, Kenji Takeda, Stewart Tansley (additional content and review*)

The purpose of this document is to outline the default Windows Azure training course contents for delivery at Microsoft Research workshops to our target audience – predominantly practicing research scientists with at least basic development skills. Where applicable, slide decks will be provided for all sections of the course in addition to supplementary documents which describe repeatable hands-on-labs (HOLs, i.e. practical exercises to be done during the class).

The standard course will be split into 2 days, with the first day centered on getting attendees to acquire a general understanding of Windows Azure and the second day focusing on using Windows Azure at scale.

Contents

[**Azure for Research FY14 – Global Training Curriculum/Syllabus** 1](#_Toc362871252)

[Day 1: Windows Azure for Research Introduction 2](#_Toc362871253)

[Overview of Cloud Computing and Windows Azure [1.5 hours] 2](#_Toc362871254)

[1. Getting started with Windows Azure Websites and Virtual Machines [2.0 hour] 2](#_Toc362871255)

[2. HOL: Creating a virtual machine and running simple data analysis with IPython or Wakari.io [1 hour] 3](#_Toc362871256)

[3. Windows Azure Storage [1 hour] 3](#_Toc362871257)

[4. DEMO: Building a Classic Scalable Applications with Visual Studio and the Windows Azure SDK (BLAST) [45 min] 4](#_Toc362871258)

[5. Day 1 Summary 4](#_Toc362871259)

[6. Take home: Creating a simple website for publishing documents, present dataset in the Storage – wiki/CMS (0 minutes) 4](#_Toc362871260)

[Day 2: Windows Azure for Research: Compute and data beyond the desktop 4](#_Toc362871261)

[7. Recap and Welcome to Day 2 [15 minutes] 4](#_Toc362871262)

[8. From Desktop to Cloud - Visualization Using Excel, data explorer, Power Map for Data Analysis on Windows Azure [1 hour] 4](#_Toc362871263)

[9. Introduction to Big Data and Big compute on Windows Azure [1 hour] 4](#_Toc362871264)

[10. Data Acquisition for devices and always-on Services [1h] 4](#_Toc362871265)

[11. Demo: Running parametric sweeps using R, Python and HPC [1 hours] 5](#_Toc362871266)

[12. HOL: IAAS Creating an HPC Cluster and run parametric sweep, MPI applications, submit jobs and share results using Websites [0 hours] Add during refresh. 5](#_Toc362871267)

[13. Creating an HDInsight Cluster (Demo) [30 minutes] 5](#_Toc362871268)

[14. Demo: Architecture patterns for the cloud STORM, Messaging, and Data caching [30 minutes] 5](#_Toc362871269)

[15. Run Spark and Shark in a Linux Cluster on Windows Azure (optional/ demo only) [30 minutes] 5](#_Toc362871270)

[16. HOL:  Putting it together, Build a service using live data acquisition and analysis [1h] 6](#_Toc362871271)

[17. Discussion and wrap-up [1 hour] 6](#_Toc362871272)

[Optional: 6](#_Toc362871273)

[Other Information 6](#_Toc362871274)

[Target Audience 6](#_Toc362871275)

[Prior Knowledge/Pre-requisites 7](#_Toc362871276)

[Learning Objectives 7](#_Toc362871277)

[Teaching Context 7](#_Toc362871278)

# Day 1: Windows Azure for Research Introduction

# Overview of Cloud Computing and Windows Azure [1.5 hours]

Overall introduction to the cloud and why you would use it for research, and types of workload you would use it for. Go through the traditional 4 graphs from continuous (always on), spikes, incremental or predictable citing examples of how corporations and scientists are using the cloud. Describe where and how data is stored and services are provisioned within a data center with a virtual tour of the data center through video and commentary.

This introduction will focus on illustrating differences between PaaS and IaaS and the different types of workloads that can be run in Windows Azure. It will describe details from the data center up and the layout (ref. Mark Russinovich: Windows Azure Internals talk) to what services are available. It will describe examples such as the genomic study examples, weather simulations in HPC and other highly scalable descriptions of workloads. We’ll talk about and reference the WATK all Azure material from there and windowsazure.com should be a first point of reference. The talk will include 6 demos:

1. Managing storage with free tools like [Cerebrata](http://www.cerebrata.com/labs/azure-explorer) azure explorer
2. Building a hello Azure application on a Web sites, the simplest ASP.net, PHP, Node JS. Paper publishing/simple app that pulls data from storage. (Data file upload/viewer example). Commercial CMS.
3. Application patterns with web role and worker roles. BLAST Demo.
4. Fetch Climate Demo.
5. Weather Demo architecture illustrating ease of scalability and the value of a live service.
6. Mobile services: sending a notification, IOS example.

Emphasis on showing broader topics on what Windows Azure can do for researchers.

// 2 days by Richard. 1 day by Wenming

## HOL: Getting started with Windows Azure Websites and Virtual Machines [2.0 hour]

The overview will describe how both Linux and Windows are first class citizens of Windows Azure. In this session users will first experience Windows Azure for themselves with a very quick HOL signing in, seeing the management console, and building their first Windows Azure website.

We’ll describe and demonstrate:

**Websites**

Introductions will be made to Windows Azure websites showing the Gallery Images for WAWS and what applications can be run. At this point make sure participants have Azure accounts running for first HOL.

1. Building a WordPress site for publishing blogs, etc.
2. Building a website with a github/bitbucket source.
3. Overview of Windows Azure Websites showing how to collect metrics, configure source control, scale, setup connections to databases, add handler mappings to run Python, node.js and PHP.

**Virtual Machines**

1. The platform image gallery.
2. The vmdepot and how to contribute.
3. IPython VM creation demo for Portal and Windows Azure CLI.
4. Different images available for science.
5. Demo: Capturing a Windows Virtual Machine with sysprep.
6. Demo: Capturing a Linux virtual machine with waagent.
7. Adding Data disks to both Linux and Windows VM.
8. We’ll cover concepts in slides and demos, **availability sets** for resilience.

// 2 days Richard // 1.5 days Wenming adding images.

## HOL: Creating a virtual machine and running simple data analysis with IPython or Wakari.io [1 hour]

1. Create an image from the VM depot using the CLI. Lab will walk through installing the Windows Azure CLI on Linux/Mac/Windows and the command to run to create a small instance vmdepot image of IPython
2. Install the Python SDK on the VM and create a script to download and upload files and create containers in blob storage
3. Run sample storage commands in IPython notebooks.
4. Run Data analysis example using Panda and Scikit-learn. // Wenming
5. Create an account on SuseStudio and create a Python VM (Optional)

// 1 day Richard //. 1 day wenming

## Windows Azure Storage, SQL Databases, and Data Management [1 hour]

Introduction to the benefits of using cloud-based storage solutions for research applications.

Overview of Brad Calder and Jai Haridas talk showing Async nature of storage. Access speed performance guidelines due to flat network storage. (Blob Meta data vs. SQL Azure) Discussion of Database as a service. This will show Windows Azure Storage at a high level. It will describe volatile storage in PaaS and persistent storage in IaaS and what underlies both.

Introduction to Blob Storage:

1. Differences between block blobs and page blobs
2. Storage Services SDK with Python (Demos)
3. Storage Services REST API and Shared Key Authentication (Demo optional or cover in slide)
4. Securing storage with Shared Access Signatures (Demo)

Introduction to Table Storage and why you would use it. Flattening out the hierarchy and picking your Rowkeys and PartitionKeys correctly. Demos will include:

1. Creating a Table for audit records in code
2. Show dynamic nature of tables and add additional columns

Description of Storage Queues and how and why you would use them.

Illustrating that PaaS notes has volatile storage.

Demo:

1. Setting up LocalStorage on a web role
2. Setting up a CloudDrive on a web role and mounting

///1 day for Richard 1 day, Wenming 1day

## DEMO: Building a Classic Scalable Applications with Visual Studio and the Windows Azure SDK (BLAST) [45 min]

1. **Demo**: Download and import of a .publishsettings file
2. **Demo**: Creating a cloud service project
3. Through the demo describe elements of ServiceConfiguration.cscfg and .csdef files and how they describe cloud services and roles.
4. **Demo**: Adding diagnostics to service and consuming using Cerebrata Azure Management Studio
5. **Demo**: Adding a startup task to the project, caching and/or remote desktop
6. **Demo**: Adding web and worker roles to a project and using Windows Azure Queue Storage

// 0, already in intro talk.

## Day 1 Summary

## Take home: Creating a simple website for publishing documents, present dataset in the Storage – wiki/CMS (0 minutes)

1. Create a Windows Azure Website from a Gallery image of Umbraco/Joomla and Mediawiki
2. Create a video for people to follow at home.

# Day 2: Windows Azure for Research: Compute and data beyond the desktop

## Recap and Welcome to Day 2 [15 minutes]

Refresh participants on what was covered the previous day.

Introduce agenda for Day 2.

Short Q&A.

## From Desktop to Cloud - Visualization Using Excel, data explorer, Power Map for Data Analysis on Windows Azure [1 hour]

Show how desktop application, such as Excel, can be used as a client for Cloud Services:

* 1. From Shawn Carlson Geoflow (Power Map) demo
  2. Data explorer demo.
  3. Power pivot demo (Xvelocity Column based DB).
  4. Python and Excel demo.

.// richard -> data explorer, wenming 2 days

## Introduction to Big Data and Big compute on Windows Azure [1 hour]

Introduce where Big Data and Big Compute are applicable with compelling real-world scientific examples to setup Day 2. Highlight the benefits of Azure Big Compute versus ‘traditional’ HPC systems.

1. Wenming’s deck and demos on scale out options.
2. Matlab parametric sweep Demo using HPC.
3. Demo running R and/or Python.

//1 day richard

1. Data Acquisition for devices and always-on Services [1h]
2. Catalog of examples
3. Internet of Things
4. REST
5. How to scale data acquisition
6. Sensor
7. Messaging using service bus, and service bus relays.
8. Demo: raspberry pi.

// 3 days.

## Demo: Running parametric sweeps using R, Python and HPC [1 hours]

Describe how easy it is to run parametric sweep jobs using researchers’ favorite tools: R, Python and Matlab. Explain the different types of HPC available including scheduler SDK, burst to Azure and a full IaaS implementation. Show the power of HPC as a demo.

1. Matlab parametric sweep Demo using pre-deployed HPC.
2. Running R and/or Python on Azure
3. Share results using websites

// 2 days to script out the deployment. 1. Day Wenming Matlab demo.

## HOL: IPython Cluster. [1 hour]

1. Downloading and installing the CLI running commands to install an IPython **cluster** – deploy two nodes of a four node cluster with the IPython vmdepot image
2. As above but using the Service Management API Python SDK wrapper to deploy the cluster
3. Run Python script to download files from blob storage to the cluster
4. Backup files between storage accounts using Copy Blob API in Python.

// 2 Days

## Creating an HDInsight Cluster (Demo) [30 minutes]

Show how Excel can be used to drive Big Data solutions. Explain what HDInsight is and how this is part of Microsoft’s Big Data Strategy. Explain the architecture and interfaces to create an HDInsight cluster. Introduce Spark and explain the difference between this and HDInsight. Go through a use case of when each one would be used and what the major differences are.

1. Create a four node HDInsight cluster in subscription
2. Create a simple map-reduce of an ecommerce dataset
3. Write a Hive query against the data set
4. Write a pig latin expression against the dataset
5. Using pre-deployed spark/shark instance, demo run a simple query against the same dataset as HDInsight first using Scala Map-Reduce then using Shark

// 1 day document

1. Demo: Architecture patterns for the cloud STORM, Messaging, and Data caching [30 minutes]
2. How to deal with data in the cloud – advanced topics – performance optimization
3. Prep for real-time data analysis
4. Vendor to author
5. The same demo using raspberry pi.

## Run Spark and Shark in a Linux Cluster on Windows Azure (optional/ demo only) [30 minutes]

Introduce Spark and explain the difference between this and HDInsight. Go through a use case of when each one would be used and what the major differences are.

1. Install Mesos on a new linux virtual machine
2. Install shark and spark
3. Create another virtual machine and add to the Spark cluster
4. Run a simple query against the same dataset as HDInsight first using Scala Map-Reduce then using Shark

// 1 day

1. HOL:  Putting it together, Build a service using live data acquisition and analysis [1h]
2. End to end exercise.
3. GPS data from China, mobile devices as clients

Vendor to author. Data feed, data analysis and dash boarding. Emulating the many devices. D3 for dash boarding.

// 3 days

## Discussion and wrap-up [1 hour]

Recap what participants have learnt. Summarise more advanced topics to follow-up indivisually with people afterwards (e.g. STORM). Q&A. Explain next steps: Visiting researcher programme, RfP applications, etc… Ask audience the type of projects they would be interested in doing.

**The END**

Optional:

1. Additional Modules in Windows Azure, Devices, Storage, and Optimized Scalable Service
2. Windows Azure Diagnostics – interesting bonus content, if time allows
3. SQL Azure, Data Caching – delete, included above
4. Understanding Messaging using Service Bus – include in a demo to at least show it
5. Multiple domain-specific examples (we will build up a portfolio of these over the year)
6. Supplementary reading material (similar to latter)
7. Guide to the community, where to get help/support, etc. How to contribute back
8. Overnight exercise between day 1 and day 2 (typically we’ll let people rest/discuss)

# Other Information

## Target Audience

* Active eScience researchers who can code, who will soon code; and
* Active CS researchers who are working with eScientists and doing the coding for those eScientists
  + Not general computer scientists, though they are welcome if they see value and if we have space
  + Not educators specifically, though also welcome (if space) if they don’t have access to cloud/Windows Azure education classes elsewhere
* Practising researchers
  + Priority to Faculty, research assistants, post-docs for physical places on courses
  + Suitable for grad students, masters students to fill extra places
  + Undergraduates as standby if course is not full
    - If they are doing undergraduate esSience research – but typically not general CS undergraduates

Note: for the “nots” we have other offerings. This class is specifically and relatively narrowly targeted by design.

## Prior Knowledge/Pre-requisites

* Can program in a modern language like Java, Python, C#, or C at minimum. Some Python and Java experience preferred – we’ll bias examples to Python, with some Java. You don’t need to be an expert programmer, but if one is programming in your eScience research today, that is ideal and preferred for getting most from the class.
* We prefer basic exposure to cloud computing (101 level), but certainly no real expertise required. No prior cloud usage experience required. We’ll cover that, as well as our particular cloud technology. We’ll take people who have heard of cloud computing and are curious or unsure about its value for their research, or equally those who are experienced in using and developing to another cloud system.
* No prior experience of Windows Azure is required – though some exposure would help us to incorporate more advanced topics. Assume none, in general.

## Learning Objectives

After attending this course you will have learned:

* What cloud computing is;
* Why and when to use cloud computing for research:
  + Including examples of where it has been successfully applied in specific domains, e.g. life sciences, environmental sciences, social sciences
* What the major design patterns for successful Cloud applications. Participants will be able to:
  + Deploy websites using Wordpress or CMS systems
  + Create virtual machines and deploy iPython
  + Manage storage using Windows Azure blobs
  + Understand how to link client applications with Windows Azure
  + Create and develop SQL Azure databases
  + Create and deploy mobile services
  + Create and run scalable computing applications using Visual Studio, Python, R and Matlab
  + Build an end-to-end cloud-based system for sensor data acquisition, data processing and REST service for access from any device
* Running researcher's own app/service on Windows Azure – with a positive training outcome directly relevant to researcher
* Next steps - personalized action plan for using Windows Azure for researcher's own work

## Teaching Context

* ~50 seats per class event is our target for the global program of classes, though there may be regional/cultural variances in practice.
* Attendees will use their own laptop. It’s primarily a “hands-on-lab”. It means they will do exercises on their laptop, accessing the cloud. Good connectivity is vital.
* It doesn’t need to be a Windows laptop. Our cloud is agnostic – though we will show both Windows and Linux VMs in the cloud. In principle, they could be remote, because of this, but the class is designed for an in-person experience.
* We will provide Windows Azure accounts for everyone that doesn’t already have an account. Our account hand-outs for the class will have sufficient resources for the class, but not for many full science projects – we have other ways to support the latter which we will highlight in the class. The account will last for a few weeks/months, not just the duration of the class.

<end>