CSC8110: Cloud Computing
Coursework Assignment 2013-14

Matthew Forshaw (`m.j.forshaw@ncl.ac.uk`)
Rebecca Simmonds (`r.m.simmonds@ncl.ac.uk`)

Tasks

1. Multi-tier E-commerce web application with NoSQL data store

Design and build a simple multi-tier e-commerce web application in the Microsoft Azure cloud using a NoSQL data store.

Your web application should provide the following basic functionality.

- **Register and list customers.**
  Customers should have a unique identifier, a name, and the country they are located.

- **Place and list orders.**
  Orders should have a unique identifier, a customer identifier, a *"stock keeping unit"* *(SKU)* (a unique product identifier), an order date and time, and a total order price.

- **Management Reports A and B**
  The managers of the e-commerce website have specified two "management reports" *(queries)* they wish to be able to perform on the customer and order information. These are discussed fully later in this document.
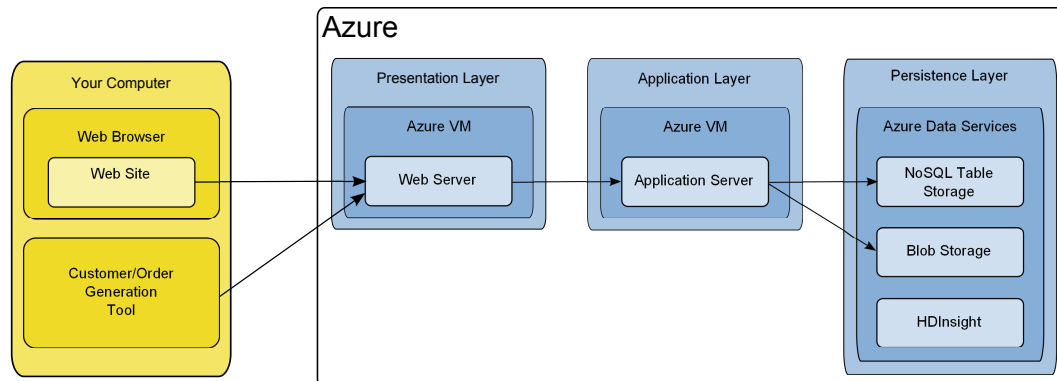
You should prove a simple form-based interface to add customers and orders to your system.

> ***Important:*** *You will not be marked on the look and feel of your web front-end. Furthermore, you are expected only to implement the functionality outlined in the coursework document, and you are not required to implement a customer authentication mechanism.*

You will use the Windows Azure Table storage service, a scalable datastore for the storage and querying of structured, non-relational data. A full description of the Azure Table storage service, instructions for creating an Azure storage account from the Azure management portal, and Java source code examples can be found at the following URL.

```
http://www.windowsazure.com/en-us/develop/java/how-to-guides/table-
service/
```

The following diagram shows the architecture you will need to implement for your application:

Azure

Your Computer

Web Browser

Web Site

Customer/Order
Generation
Tool

Presentation Layer

Azure VM

Web Server

Application Layer

Azure VM

Application Server

Persistence Layer

Azure Data Services

NoSQL Table
Storage

Blob Storage

HDInsight

Choice of Technologies

We encourage you to make your own decision over their choice of programming languages and application server (e.g. the LAMP stack you installed during Practical One, JBoss AS, Tomcat, etc) used in this coursework. However, there are a few pointers which might help guide your decision:

- **Is there an Azure SDK available for the programming language?**
  Windows Azure offers a number of SDKs in a variety of different languages including Java, Node.js, PHP, .NET, Python and Ruby. While it possible to interact with the management functions of Azure using raw REST requests, this will prove time consuming and error prone and we strongly suggest you select one of the supported languages listed above.

- **Which technologies are you most familiar and comfortable with using?**
  If there are a set of technologies you have more experience of and would feel more comfortable using in the cloud, you may want to select those for this coursework. Alternatively, if you wish to expose yourself to using new technologies, your experiences could form an interesting discussion point in your report.

- **Will the demonstrators have experience using these technologies to be able to help you?**
  While demonstrators will do their best to assist you with any issues you have, they may not be familiar with less common technologies. If in doubt, ask a demonstrator!

- **Have these technologies been used on Azure before?**
  Can you find tutorials or examples of people being successful configuring these technologies in Azure? Are there any VMDepot pre-defined instances which provide these technologies?

If you have any concerns over your choice of implementation language or application server, don't hesitate to contact a demonstrator and they will be able to advise.

2. Generating test customers and orders

Next you are required to develop a tool to generate example customers and orders and populate your system. Your application should generate **at least** *1,000* customers, and

1,000,000 orders within your system. This will provide you with the large quantity of data required later in this coursework.

We expect you to run this tool directly on your own computer, and for it to send HTTP requests either to your presentation (web layer) or directly to your application layer. Your tool should **not** insert customer and order information directly into your database.

You may develop your own tool from scratch or make use of frameworks such as Selenium (`http://www.seleniumhq.org/`) or Apache JMeter (`http://jmeter.apache.org/`).

> *Hint: Your tool may contain arrays/lists of first and last names which are selected randomly when creating customers. For your orders you may also generate random dates, times and total order prices.*

3. Management Report A: JOINs on data in NoSQL.

You now need to extend your e-commerce application to be able to generate the following report.

> *"Calculate the **total order price** of **all orders** placed by customers from a given country."*

Unlike relational databases, NoSQL does not support JOIN statements, so you will need to execute multiple queries against your NoSQL data store and perform the JOIN between *customers* and *orders* programmatically within your application logic.

4. Management Report B: MapReduce with HDInsight

You should now build upon your knowledge and experience of HDInsight gained during practical sessions, using the MapReduce features of Windows Azure HDInsight to generate the following report.

> *"For every country in your system, calculate the **average order price** for **all orders** originating from that country."*

Step 1: Extend your application to query your NoSQL database and return all customers and orders in CSV format, ensuring all required information is included in each row. For example:

```
"Joe Bloggs","United Kingdom",SKU0001,"01/12/2013","23:30",14.50
"Max Musteerman","Germany",SKU0004,"02/12/2013","08:25",52.90
"John Doe","United States",SKU0003,"02/12/2013","10:15",126.20
...
```

Step 2: Upload your CSV file to your HDInsight cluster's Windows Azure Blob Storage.

> *Hint: instructions for uploading data to HDInsight can be found here:*
> `http://www.windowsazure.com/en-us/manage/services/hdinsight/howto-upload-data-to-hdinsight/`

Step 3: Write *map* and *reduce* functions to process this data.

Hint: examples of map and reduce functions to process data in CSV format can be found here:
`http://www.ibm.com/developerworks/library/j-javadev2-15/`

Step 4: Run your MapReduce task against your CSV file, inspecting the produced output and verifying that it has run correctly.

Useful resources

This section provides links to a number of online resources you may find useful in completing this coursework assignment.

- Windows Azure Development Documentation
  `http://www.windowsazure.com/en-us/documentation/?fb=en-us`

- Windows Azure Virtual Machine Documentation
  `http://www.windowsazure.com/en-us/documentation/services/virtual-machines/`

- Windows Azure HDInsight Documentation
  `http://www.windowsazure.com/en-us/documentation/services/hdinsight/`

- Windows Azure Java Development Center
  `http://www.windowsazure.com/en-us/develop/java/`

Deliverables

Code

You are expected to submit all source code and any third-party libraries (or a maven build file with correctly specified dependencies) required for us to deploy and test your application in the Windows Azure cloud. You should also include a brief *README* document explaining any configuration steps required to run your application.

Written Report

You should prepare a written report in **PDF format** containing the following:

- A brief description of each component in your system.

- A bullet point list of any assumptions which motivated your architectural or system design decisions.

- A brief description of your NoSQL database representation of customers and orders.

- An explanation of your approach to implementing **Management Report A**. If you were to re-design your application, what could you do to make obtaining this information easier, and what would be the pros and cons of this approach?

- An explanation of your approach to implementing **Management Report B**, including a description of your *map* and *reduce* functions.

- A short user guide including a brief discussion and screenshots showing the required functionality of your e-commerce application. All screenshots should be clearly legible and cropped appropriately.

- Screenshots of the Windows Azure management portal showing your virtual machine and HDInsight cluster dashboards.

- A discussion of your personal experiences of your development process using the Azure cloud. We are interested to hear not only the parts you found easy, but any challenges you encountered and how you overcame them.

Matthew Forshaw will be happy to provide initial feedback on any written reports prior to submission. If you would like feedback on your draft report, please email a copy in **PDF format** to `m.j.forshaw@ncl.ac.uk` by **12pm** on **Thursday 12th December 2013**. He will return feedback prior to the practical session that afternoon, and will be happy to answer any further questions you have based on this feedback.

Practical Sessions

Demonstrators will be available in the computer clusters in the following CSC8110 practical sessions.

Date	Time
Thursday 4th December 2013	4-6pm
Friday 5th December 2013	11am-1pm
Monday 9th December 2013	9-11am
Tuesday 10th December 2013	3-5pm
Wednesday 11th December 2013	9-11am
Thursday 12th December 2013	4-6pm
Friday 13th December 2013	11am-1pm

Demonstration

In addition to your electronic submission, you will be expected to demonstrate your application to demonstrators during the practical session on **Friday 13th December 2013**. An online demonstration slot signup form for each cluster room will be distributed shortly.

Deadline and Submission

Completed assignments including all source code and your written report must be submitted electronically through NESS by **Friday 13th December 2013** at **5pm**.

This assignment is worth **30%** of your final mark for this module.

Questions?

If you have any further queries about the coursework assignment or your submission, you should approach one of the demonstrators during the practical sessions or on the Blackboard discussion board.

Matthew Forshaw (`m.j.forshaw@ncl.ac.uk`)
Rebecca Simmonds (`r.m.simmonds@ncl.ac.uk`)