

**Running an Intelligent Analytical System on AWS**

**Using AWS Services & Solutions in AWS Marketplace**

*Step-by-Step Guide for Data Pipeline from AWS Marketplace*

Disclaimer:

1. The AWS Marketplace Fusion Solution showcased in this document is solely meant as a tutorial, *but with given additional customizations, it can be used for production use cases*.
2. Technologies used in this Solution can be replaced by other equivalent technologies as needed for business reasons.
3. All data used in this Solution is machine generated and fictitious.
4. For setting up this AWS Marketplace Fusion Solution, prior knowledge of the technologies used in the Solution and familiarity with Amazon AWS Cloud is recommended.
5. For most of the components, we used the region **US West (Oregon)**, but you can change it as per your choice.

**Table of Contents**

[1. Introduction 4](#_Toc467687812)

[2. Prerequisites 6](#_Toc467687813)

[3. Business Use Case 7](#_Toc467687814)

[4. Audience 8](#_Toc467687815)

[5. Setup Web App Deployment Environment 9](#_Toc467687816)

[5.1. Install Python packages 9](#_Toc467687817)

[5.2. Install R packages 9](#_Toc467687818)

[5.3. Update Project Properties 10](#_Toc467687819)

[5.4. Build the project 16](#_Toc467687820)

[5.5. Configure TOMCAT 16](#_Toc467687821)

[5.6. Publish Web Application 18](#_Toc467687822)

[5.7. Build CloudBeam Slave Application 19](#_Toc467687823)

[5.8. Publish Slave Web Application on ATTUNITY CloudBeam 21](#_Toc467687824)

[5.9. Add Firewall rule on ATTUNITY CloudBeam 22](#_Toc467687825)

[6. Step by Step Data Pipeline Guide 27](#_Toc467687826)

[6.1. Step 1: Overview of Dashboard 28](#_Toc467687827)

[6.2. Step 2: Load historical Weather Data 30](#_Toc467687828)

[6.3. Step 3: Load historical Company Data into MySQL & Stock Data into Amazon S3 30](#_Toc467687829)

[6.4. Step 4: Load Company Data from MySQL to Amazon RedShift using Attunity Cloudbeam 31](#_Toc467687830)

[6.5. Step 5: Load Stock and Weather Incremental Data 31](#_Toc467687831)

[6.6. Step 6: Start Prediction 32](#_Toc467687832)

[7. Appendix 33](#_Toc467687833)

1. Introduction

This Project is the 2nd project of the 3 Projects involved in setting up this AWS Marketplace Fusion Solution. (Projects should be executed sequentially.)

1. **Project 1**: Step-by-Step Deployment Guide – Part 1

Step-by-Step Deployment Guide – Part 2

2. **Project 2:** Data Pipeline

3. **Project 3:** Machine Learning, Reporting and BYOD

The Data Pipeline process execution can be done in 2 ways:

* Execute the Data Pipeline process using Dashboard, which was built to automate the process. (Automated process)
* Another way is to execute the Data Pipeline process manually. (Manual process)

In this document we will go through the Automated Process only.

Refer below to the high-level architecture diagram of the overall data flow and the system/solution information.

C:\Users\Abhinandan\Desktop\Diagram\aws_marketplace_immersion_project_HLA.png

2. Prerequisites

* + - 1. Make sure you have executed the steps in the following documents:
* Prerequisites document
* Project 1 –
  + Deployment Guide Part 1
  + Deployment Guide Part 2
    - 1. Make sure the following components are up and are in running status:

|  |  |  |
| --- | --- | --- |
| **Component #** | **Component Name** | **Purpose** |
| A | Amazon S3 | Storage |
| C | ATTUNITY CloudBeam | ETL Software |
| D | MySQL | Database |
| E | AWS RedShift | Storage |
| G | Tomcat 1 | To Run the Automate Dashboard |
| H | Tomcat 2 | To run the ATTUNITY CloudBeam ETL Job |

3. Business Use Case

Please refer to the Overview document for the Business Use Case used in this AWS Marketplace Fusion Solution.

4. Audience

This document is designed for different types of audiences. Therefore, each of the steps are explained in detail. Below is a list of people who are best suited to explore this document:

**Enterprise Organization**

* Developer
* IT Professional
* IT Manager

**Medium to Small Organization**

DevOps

1. Setup Web App Deployment Environment

Amazon BYOD web application will be hosted on the **TIBCO Spotfire** EC2 instance in order to establish a remote desktop connection to the **TIBCO Spotfire** instance and perform the following steps on it.

* 1. Install Python packages

Follow the steps below to install the **psycopg2** and **tinys3** Python packages on the **TIBCO Spotfire** instance.

1. Open a command prompt and execute the following command to install the **psycopg2** package.

|  |
| --- |
| pip install psycopg2 |



1. Execute the following command to install **tinys3.**

|  |
| --- |
| pip install tinys3 |



* 1. Install R packages

Follow the steps below to install **rJava**, **RJDBC,** and **devtools** R packages.

1. Open a command prompt and execute an **R** command to initiate an R shell.

|  |
| --- |
| R |



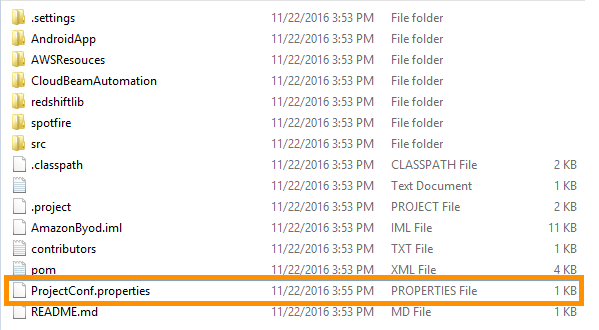
1. Execute the commands below to install the **rJava**, **RJDBC,** and **devtools** R packages.

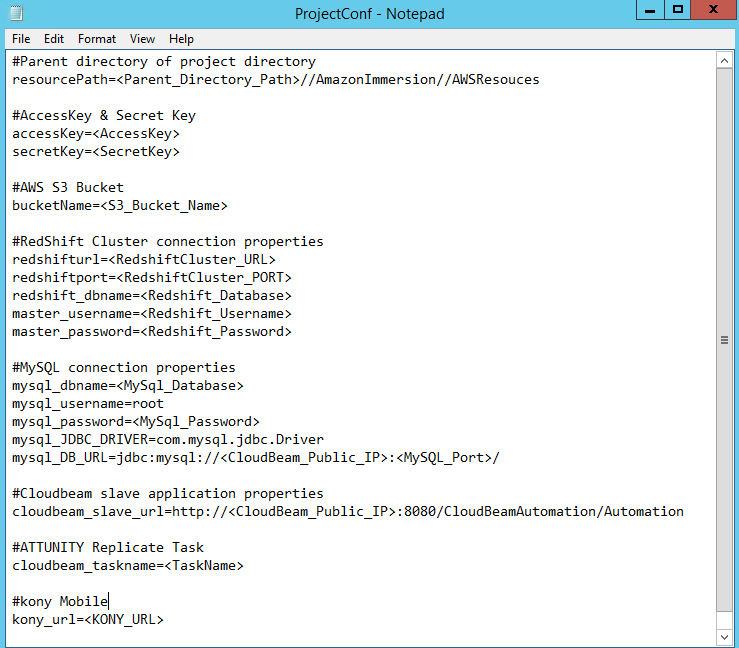
|  |
| --- |
| install.packages("rJava")  install.packages("RJDBC")  install.packages("devtools") |

While executing a command, you will be asked to select CRAN mirror for the current session in a separate window. Select an appropriate mirror and click on the **OK** button.

* 1. Update Project Properties

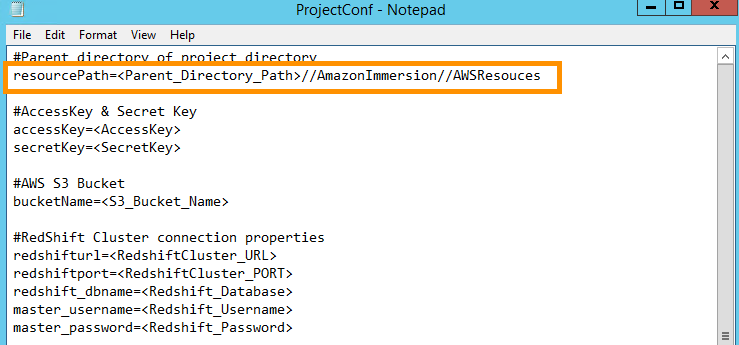
1. Open the directory where the Project Repository was cloned **in Section 4.9: Setup solution code tree & datasets** of **Deployment Guide –- Part-2 of Project 1**.
2. Open the directory **AmazonImmersion** containing the project code.
3. Open the **ProjectConf.properties** file (inside the **AmazonImmersion** directory) that contains all the project settings in the form of key value pairs.

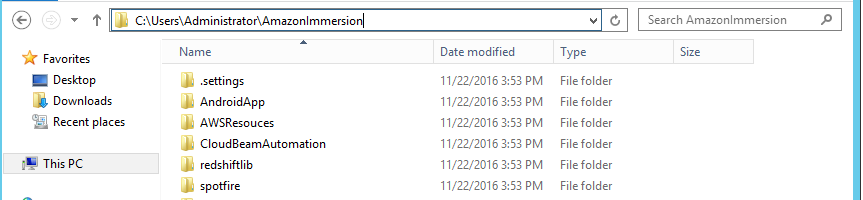




1. Replace the **<Parent\_Directory\_Path>** token from **resourcePath** Key with the path of the directory containing the cloned project. Example: Path of directory containing **AmazonImmersion** folder.

* Change the separator of the directory from backslash (‘\’) to double forward slash (‘//’).

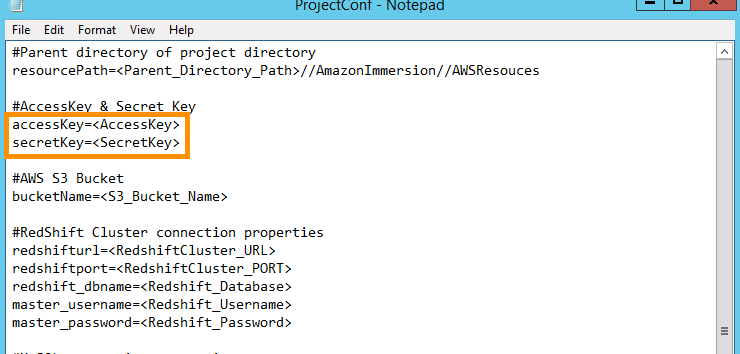




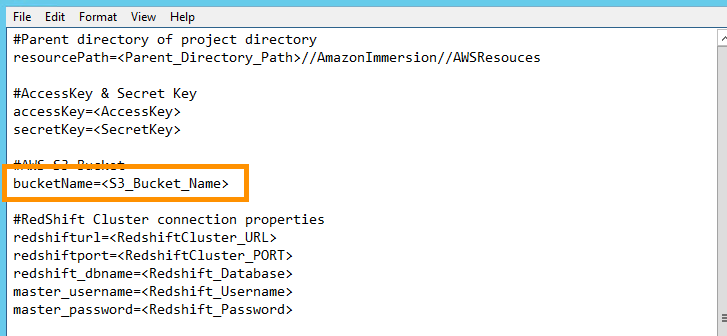
In the above example, the parent directory path is **C:\Users\Administrator.**

This example replaces the token **<Parent\_Directory\_Path>** with a value of **C://Users//Administrator.**

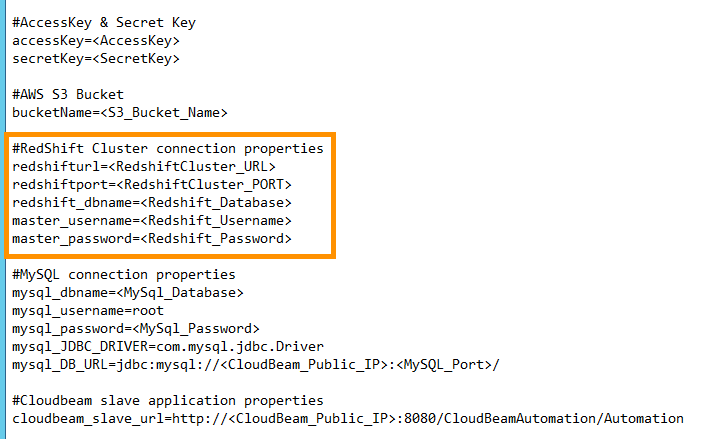
1. Replace the tokens **<AccessKey>** and **<SecretKey>** with the respective Access Key and Secret Key of the user that was used while creating the Bucket in **Section 4.2: Storage setup using Softnas** in the **Deployment Guide — Part 1** of Project 1.



1. Replace **<S3\_Bucket\_Name>** with the name of the S3 Bucket created in **Section 4.2: Storage setup using Softnas** in the **Deployment Guide — Part 1** of Project 1.

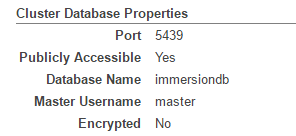


1. In the **Redshift Cluster connection properties** section, enter the appropriate values for the respective keys.

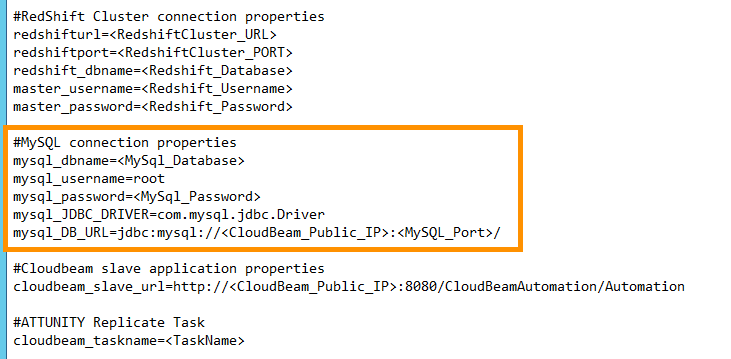


Replace the following tokens as described below:

* **<RedshiftCluster\_URL>** Replace with the endpoint of the Redshift cluster URL. Make sure port number is not part of the URL. Refer to **Step 15 Section 4.3.4** in the **Deployment Guide — Part 1** of Project 1.
* **<RedshiftCluster\_Port>**: Replace with the Redshift cluster port number. Refer to **Step 15 Section 4.3.4** in the **Deployment Guide — Part 1** of Project 1.
* **<Redshift\_Database>, <Redshift\_Username>** and **<Redshift\_Password>:** Replace these tokens with the respective details provided while provisioning the Redshift Cluster. These details can be obtained from the Redshift cluster detail page. Refer to **Step 6 Section 4.3.1** in the **Deployment Guide — Part 1** of Project 1.



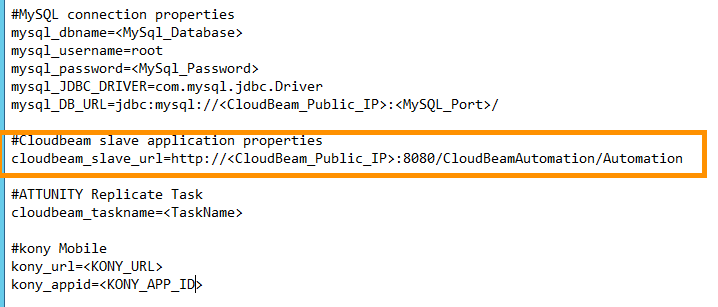
1. In the **MySQL connection properties** section, enter the appropriate values for the respective keys.



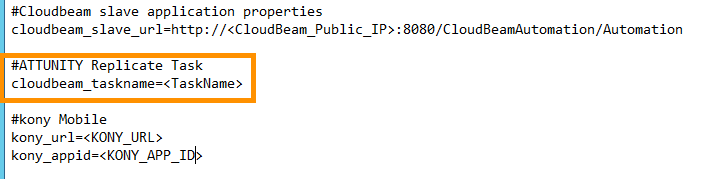
Replace the following tokens as described below:

* **<MySql\_Database>**: Replace with MySQL database name. Refer to **Section 4.4.7: Create MySQL Database** in the **Deployment Guide — Part 1** of **Project 1**
* **<MySql\_Password>:** Replace with thepassword of MySQL 5.5 Refer to **Section 4.4.5: Install MySQL 5.5** in the **Deployment Guide — Part 1** of **Project 1**
* **<CloudBeam\_Public\_IP>:** Replace with the Public IP of ATTUNITY CloudBeam EC2 instance.Refer to **Section 4.4.1. Launch ATTUNITY CloudBeam Instance** in the **Deployment Guide — Part 1** of **Project 1**.
* **<MySQL\_Port>**: Replace with the MySQL port number. By default, it is **3306.** Use the default **3306** if you have not changed it explicitly.

1. Replace the token **<CloudBeam\_Public\_IP>** present in Cloudbeam slave application properties section with the **Public IP** of the ATTUNITY CloudBeam EC2 instance you created in **Section 4.4.1. Launch ATTUNITY CloudBeam Instance** in the **Deployment Guide — Part 1** of **Project 1**.

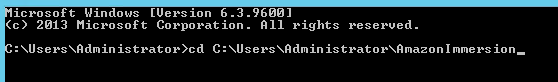


1. Replace the token **<TaskName>** with the name of the Task created in the ATTUNITY Replicate Console. (Refer to **Section 4.4.10: Create ATTUNITY Replicate Task** of the **Deployment Guide — Part 1** in Project 1.)



1. The Kony Mobile Fabric connection properties will be set in Project 3.
2. **SAVE** the **ProjectConf.properties** file.
   1. Build the project
3. Open a **Command Prompt** and navigate to the Project directory (i.e. AmazonImmersion) using the command below. Replace the token <Path\_Project\_Directory> with the path to the **AmazonImmersion** directory.

|  |
| --- |
| cd <Path\_Project\_Directory> |



1. Now execute the command below in CLI to build the project.

|  |
| --- |
| mvn clean install |



1. After successfully building the project, the **AmazonByod.war** file will be created in the **AmazonImmersion\target** directory.
   1. Configure TOMCAT

The TIBCO Spotfire server is running on port number 8080. So you need to change the port for the TOMCAT server. Perform the following steps to change the TOMCAT default ports.

1. Open the **redshiftlib** directory located in the Project Directory (i.e. **AmazonImmersion**).
2. Copy the **RedshiftJDBC4-1.1.17.1017.jar** file and paste it in the **lib** directory located inside the Tomcat installation directory (i.e. C:\Program Files\Apache Software Foundation\Tomcat 8.5\lib).
3. Navigate to the **conf** directory located inside the Tomcat installation directory (i.e. C:\Program Files\Apache Software Foundation\Tomcat 8.5\conf).
4. Open the **server.xml** file in **Notepad**.
5. Search the **Server** node and change the **port** value from **8005** to **8015.**

|  |
| --- |
| <Server port="8005" shutdown="SHUTDOWN"> |

To

|  |
| --- |
| <Server port="8015" shutdown="SHUTDOWN"> |

1. Search the **Connector** node with port number 8080 and change the **port** value from **8080** to **8088.**

|  |
| --- |
| <Connector port="8080" protocol="HTTP/1.1" connectionTimeout="20000" redirectPort="8443" /> |

To

|  |
| --- |
| <Connector port="8088" protocol="HTTP/1.1" connectionTimeout="20000" redirectPort="8443" /> |

1. Search the **Connector** node for AJP 1.3 with port number 8009 and change the **port** value from **8009** to **8019.**

|  |
| --- |
| <Connector port="8009" protocol="AJP/1.3" redirectPort="8443" /> |

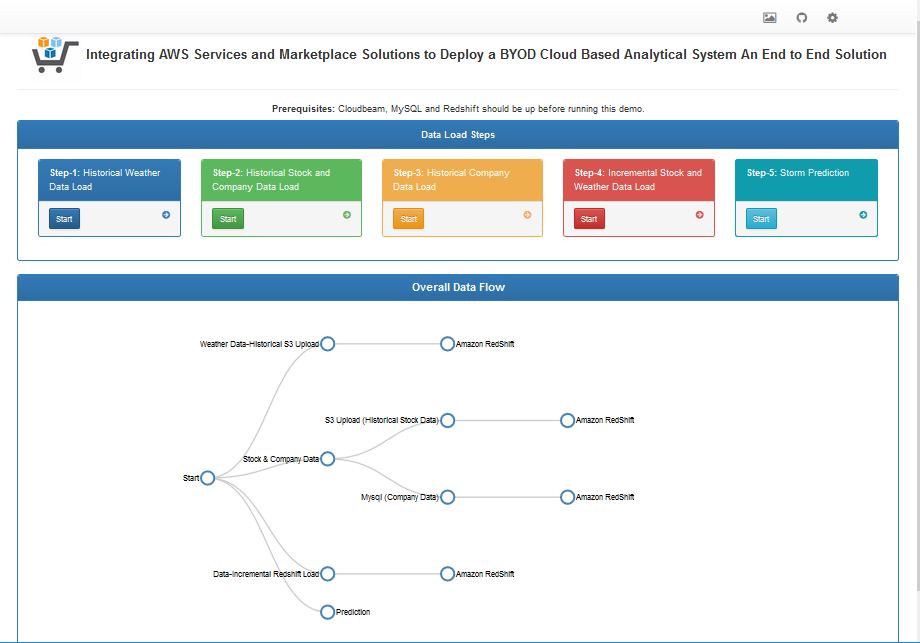
To

|  |
| --- |
| <Connector port="8019" protocol="AJP/1.3" redirectPort="8443" /> |

1. **Save and Close** the server.xml file.
   1. Publish Web Application
2. Copy the **AmazonByod.war** file located inside the **target** (i.e. AmazonImmersion\target) directory of the Project directory to the **webapps** directory located inside the Tomcat installation directory (i.e. C:\Program Files\Apache Software Foundation\Tomcat 8.5\webapps).
3. Run the **Tomcat executable** file (e.g., TomcatX.exe, where X is the version number) located inside the **bin** directory of the Tomcat installation directory (i.e. C:\Program Files\Apache Software Foundation\Tomcat 8.5\bin).

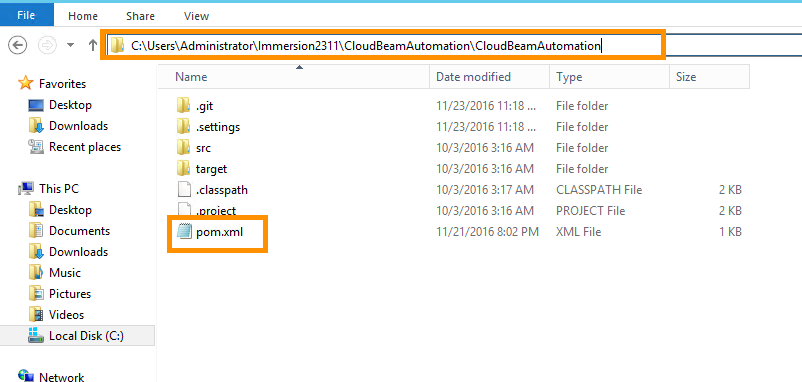
After a successful initiation of the Tomcat Server, click on the URL below to access the **AmazonBYOD** web app.

|  |
| --- |
| <http://localhost:8088/AmazonByod/> |



* 1. Build CloudBeam Slave Application

1. Open the **CloudBeamAutomation** directory located inside the Project Directory (i.e. AmmazonImmersion).
2. Copy the **CloudBeamAutomation.zip** file and paste it in the Parent directory of the Project Directory.
3. Unzip the **CloudBeamAutomation.zip** file in the Parent directory of the Project Directory.
4. Open the directory containing the **pom.xml** file inside the unzipped package directory.
5. Copy the path of the directory from the Folder Navigator.



1. Open a Command Prompt and execute the command below. Replace the token **<Directory\_Path>** with the path copied in **Step 5** above.

|  |
| --- |
| cd <Directory\_Path> |



1. Execute the command below in the CLI to build the project.

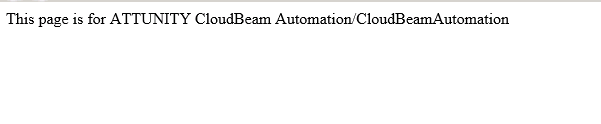
|  |
| --- |
| mvn clean install |

1. After successfully building the project, the **CloudBeamAutomation.war** file will be created in **CloudBeamAutomation\target** directory.
2. This **CloudBeamAutomation.war file** will be used to run the slave application on the **ATTUNITY CloudBeam EC2 instance**
   1. Publish Slave Web Application on ATTUNITY CloudBeam
3. Make a remote desktop connection to the **ATTUNITY CloudBeam EC2 instance**
4. Open the **webapps** directory located inside the **Tomcat installation directory** (i.e. C:\Program Files\Apache Software Foundation\Tomcat 8.5\webapps) of the ATTUNITY CloudBeam EC2 instance.
5. Copy the **CloudBeamAutomation.war** file located in the TIBCO Spotfire instance to the **webapps** directory. Open it in the ATTUNITY CloudBeam EC2 instance.
6. Start the Tomcat server in the ATTUNITY CloudBeam EC2 instance, by running the **Tomcat executable** file (TomcatX.exe, X is version number) located inside the **bin** directory of the Tomcat installation directory (i.e. C:\Program Files\Apache Software Foundation\Tomcat 8.5\bin).

* After a successful initiation of the Tomcat Server, click on the URL below to access the slave web app in the ATTUNITY CloudBeam EC2 instance.

|  |
| --- |
| <http://localhost:8080/CloudBeamAutomation/Automation> |

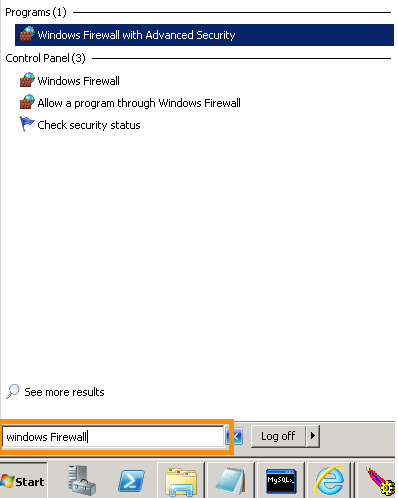
You will see following message in the web browser:



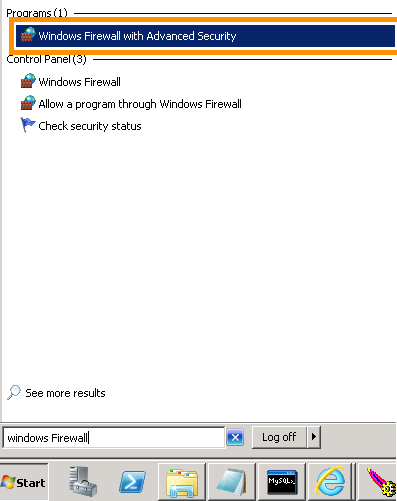
* 1. Add Firewall rule on ATTUNITY CloudBeam

Follow the steps below to add the Firewall Inbound and Outbound rule on the ATTUNITY CloudBeam EC2 instance to make the slave web application accessible by using the same public IP.

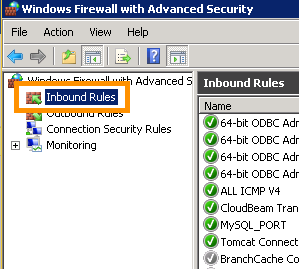
1. Click on the Start button on your taskbar and type the words **“Windows Firewall”** in the search box as shown in the image below.



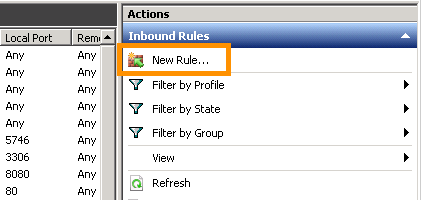
1. Select the **Windows Firewall and Advanced Security** program from the Programs list.



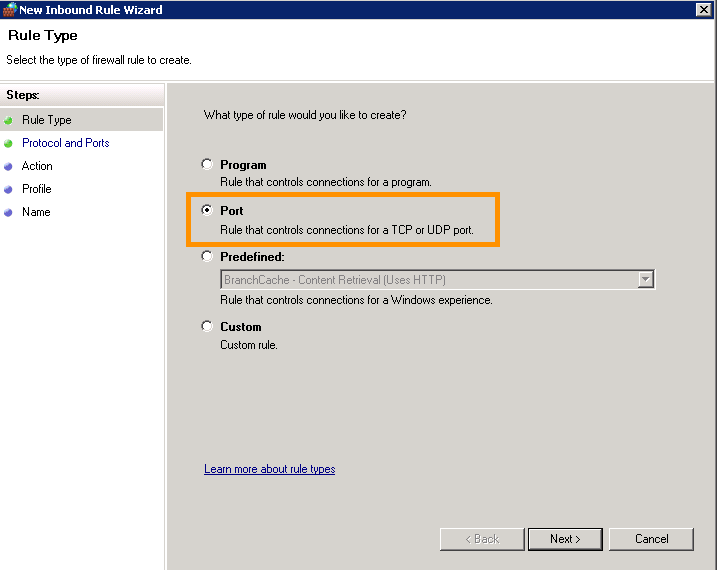
1. In the **Windows Firewall and Advanced Security** window, select **Inbound Rules** from the left side navigation pane.



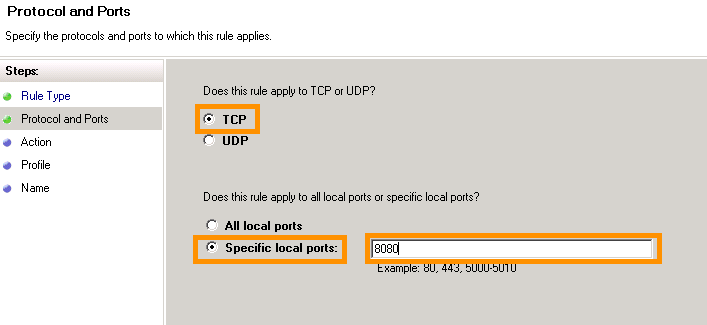
1. Click on the **New Rule** link located in the right pane of the window.



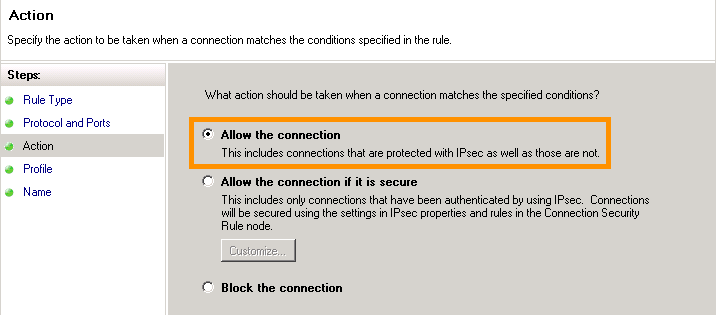
1. In the **New Inbound Rule Wizard** window, select **Port** as the Rule Type and click on the **Next** button.



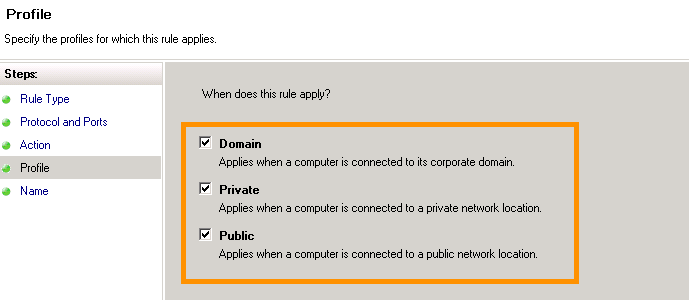
1. On the next tab, select **TCP** protocol and enter **8080** in the **Specific local ports** input box.



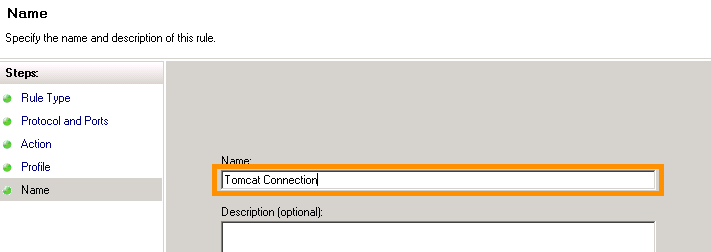
1. Click on the **Next** button at the bottom.
2. On the **Action** tab, select **Allow the connection** and click on the **Next** button.



1. On the **Profile** tab, check all the profiles and click on the **Next** button.



1. Enter a suitable name and click on the **Finish** button to complete the process.



1. Follow the same steps to add an **Outbound** rule for the **TCP** protocol and Port Number **8080**.
2. Step by Step Data Pipeline Guide

This document demonstrates how to execute each component and to move data from one component to another.

This section will walk you through:

**6.1 Step1:** Overview of Data Pipeline Dashboard

**6.2 Step 2:** Historical Weather Data Load

**6.3 Step 3:** Historical Stock and Company Data Load into MySQL & Stock Data into Amazon S3

**6.4 Step 4:** Load Company Data from MySQL to Amazon RedShift using Attunity CloudBeam

**6.5 Step 5:** Load Stock and Weather Incremental Data

**6.6 Step 6:** Start Prediction

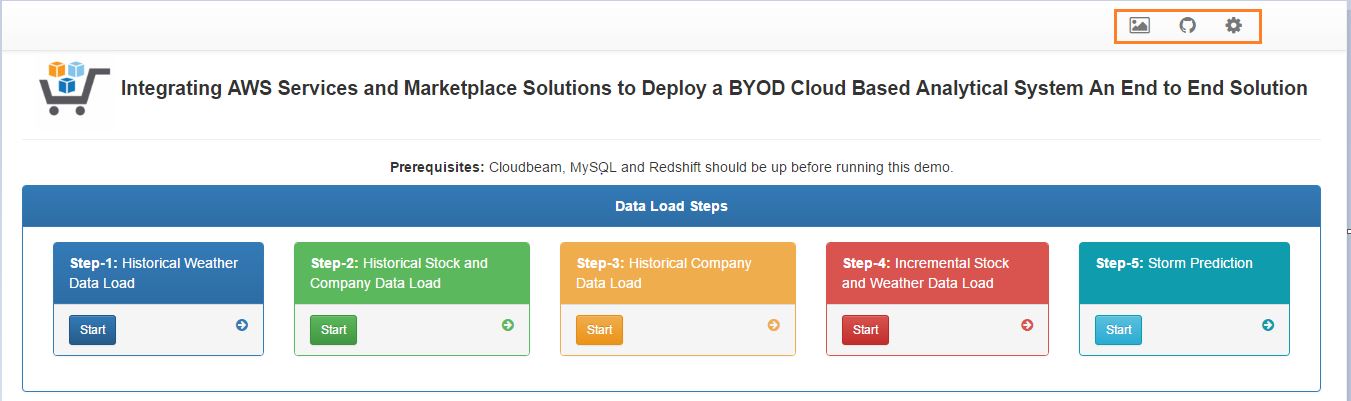
* 1. Step 1: Overview of Dashboard

The Data Pipeline dashboard is built to avoid manual execution. The Dashboard has a Start button for each step and it will execute the codes on the backend.

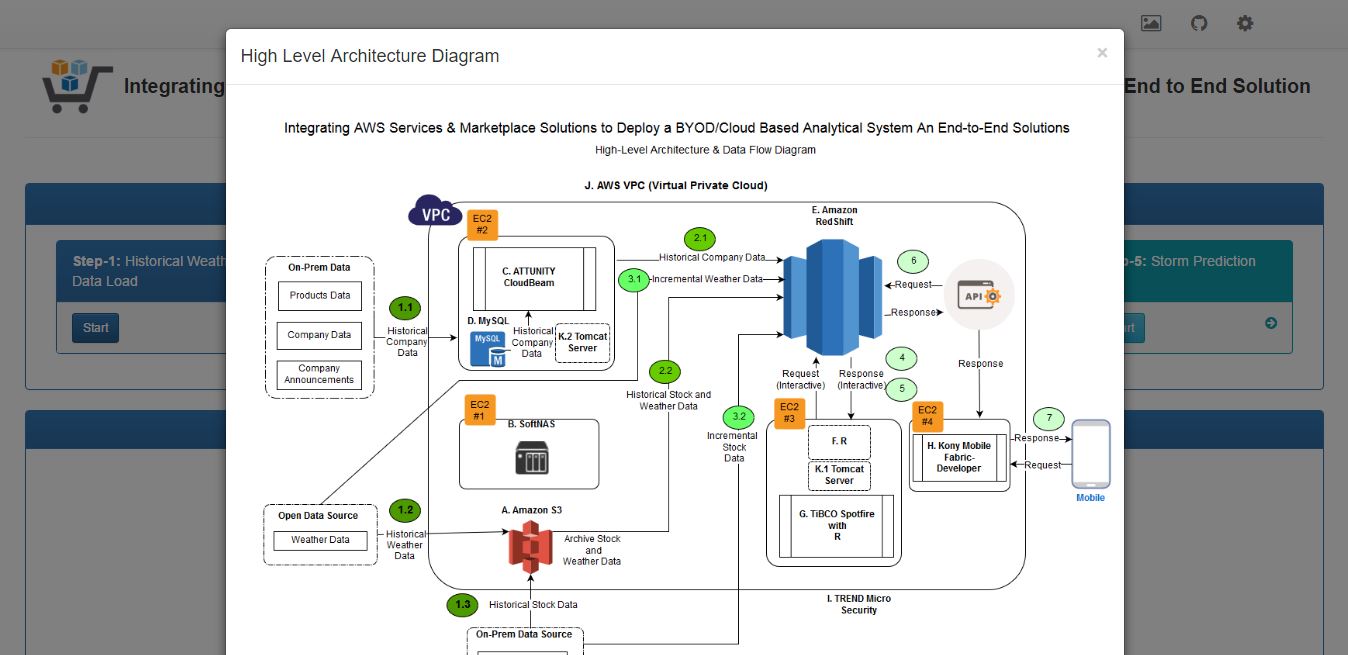
To access the Data Pipeline Dashboard, put the following IP address, with the port number, in the browser address field and hit your Enter key.

http://<your EC2 public IP where Tibco Spotfire is hosted>:8088/AmazonByod/index.html

You will see three icons on the top right side of the dashboard (highlighted with an orange box):

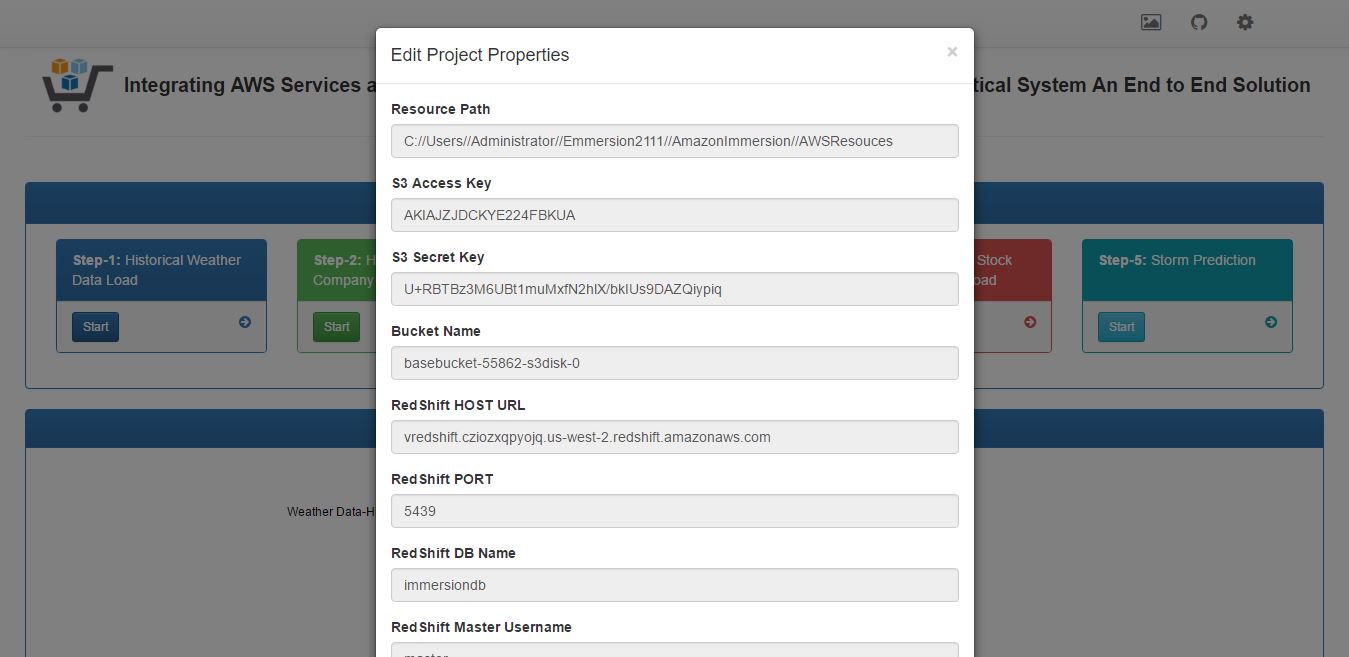


Click on the first icon from the left in order to view the High Level Architecture diagram.



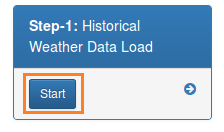
The Second icon will take you to the GitHub repository where all your codes are maintained.

The third icon shows the properties of this application, such as: **S3 Access Key, Redshift URL, MySQL user name, S3 Secret key,** etc.



* 1. Step 2: Load historical Weather Data

Click on the **Start** button of the “**Step – 1 Historical Weather Data Load**” box to load the historical weather data to Amazon S3 and to Amazon Redshift.



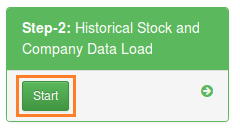
You will see the data flow in the dashboard. Please refer to Steps 1.2 and 2.2 in the architecture diagram.

Note: The data we used is open source and available for public access.

* 1. Step 3: Load historical Company Data into MySQL & Stock Data into Amazon S3

The next step is to load the company data (Stock and non-Stock). The non-stock data will load to MySQL and the company’s historical stock data will load to Amazon S3.

Once you click the Start button in the “Step – 2 Historical Stock and Company Data Load” box, it will execute the backend scripts and load the data into the targeted destination.

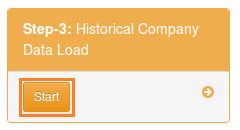


You will see the data flow in the dashboard. Please refer to the Steps 1.1, 1.3, and 2.2 in the architecture diagram.

Note: We used the company‘s “on premises” data from the company’s site.

* 1. Step 4: Load Company Data from MySQL to Amazon RedShift using Attunity Cloudbeam

Click the Start button in the “Step – 3 Historical Company Data Load” box. This will interact with the EC2 instance where Attunity Cloudbeam is hosted and will move the data to Redshift from MySQL.

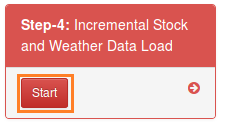


You will see the data flow in the dashboard. Please refer to the **Step 2.1** in the architecture diagram.

* 1. Step 5: Load Stock and Weather Incremental Data

Click on the Start button in the “Step – 4 Incremental Stock and Weather Data Load” box. This will load the incremental stock and weather data.

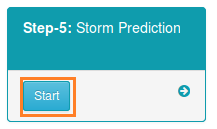
Weather data will be loaded once a day and the stock data will be loaded every 5 seconds.



You will see the data flow in the dashboard. Please refer to the Steps 3.1 and 3.2 in the architecture diagram.

* 1. Step 6: Start Prediction

Once you click on the Start button in the “Step – 5 Storm Prediction” box, it will trigger the R script and start predicting whether there is a forecast of a storm in next 5 days and load that data into Redshift.



You will see the data flow in the dashboard. Please refer to the Steps 4 and 5 in the architecture diagram.

1. Appendix

The table below gives an estimation of costs for each component used in the Data Pipeline process.

|  |  |  |
| --- | --- | --- |
| **Components Name** | **EC2 Servers** | **EC2 Charges Hourly** |
| SoftNAS | m3.xlarge | $1.616/hr. |
| Attunity CloudBeam | m4.large | $2.646/hr. |
| TIBCO Spotfire | m4.large | $2.646/hr. |
| Kony Mobile Fabric | t2.large | $0.104/hr. |
| Trend Micro | m4.large | $1.62/hr. |
| AWS RedShift | NA | $0.250/hr. |

Below are the links of the Amazon Marketplace components, please go through it to find out more details:

|  |  |  |
| --- | --- | --- |
| **Component #** | **Component Name** | **EC2 Server Type** |
| B | SoftNAS | <https://aws.amazon.com/marketplace/pp/B00PJ9FGVU?qid=1475145771249&sr=0-2&ref_=srh_res_product_title> |
| C | Attunity CloudBeam | <https://aws.amazon.com/marketplace/pp/B00LBH6GCC?qid=1475145807428&sr=0-3&ref_=srh_res_product_title> |
| G | Tibco Spotfire | <https://aws.amazon.com/marketplace/pp/B00PB74KYY?qid=1475145882225&sr=0-9&ref_=srh_res_product_title> |
| H | Kony Mobile Fabric | Developer: <https://aws.amazon.com/marketplace/pp/B010TV3U2E?qid=1475145908555&sr=0-2&ref_=srh_res_product_title>  Express: <https://aws.amazon.com/marketplace/pp/B010PHCVO0?qid=1475145908555&sr=0-1&ref_=srh_res_product_title> |
| I | Trend Micro Security | <https://aws.amazon.com/marketplace/pp/B01AVYHVHO?qid=1475145958660&sr=0-2&ref_=srh_res_product_title> |