

HOL Guide for Enterprise Risk Analysis

USING DATAMEER, HDINSIGHT, TRENDMICRO DEEP SECURITY
AND CHEF

HOL Guide for Enterprise Risk Analysis

The purpose of this section is to capture all changes made to the content of the document.

Contact for Enquiries and Proposed Changes

If you have any questions regarding this document, please contact:

Email Address

azuremarketplace@avyanconsulting.com

1 Table of Contents

1	Overview	3
2	How to deploy this solution	3
3	How to configure the components.....	7
3.1	Datameer.....	7
3.2	TrendMicro	7
4	Signing into Datameer UI.....	8
5	Configure Datameer to Fetch Data from Azure Storage	10
6	Link, Clean and Prepare the Data	14
7	Perform Analysis to Identify Outliers.....	25
8	Logging in to the TrendMicro DSM	42
8.1	Server name	42
8.2	Server login.....	42
9	Perform policy configuration on the TrendMicro DSM	43
10	Exercises.....	45
10.1	Datameer – Visualize the Data	45
10.2	TrendMicro – Malware test.....	45
11	Visualize the Data	45
12	Malware Test	49
12.1	Generating Malware alert in the computer	49
12.2	Dashboard – Malware Alert	49
12.3	Malware Alert verification	50
13	References, Attachments & Definitions	51
13.1	References.....	51

1 Overview

The purpose of this document is to provide the step-by-step instructions of deploying and configuring the Enterprise Risk Analysis using Datameer Business Intelligence and TrendMicro DeepSecurity solution and lab exercises.

The exercises includes creation of credit fraud risk awareness using sample (representative) data, building powerful Infographics of the Datameer and the security intelligence in the malware detection of the TrendMicro DeepSecurity

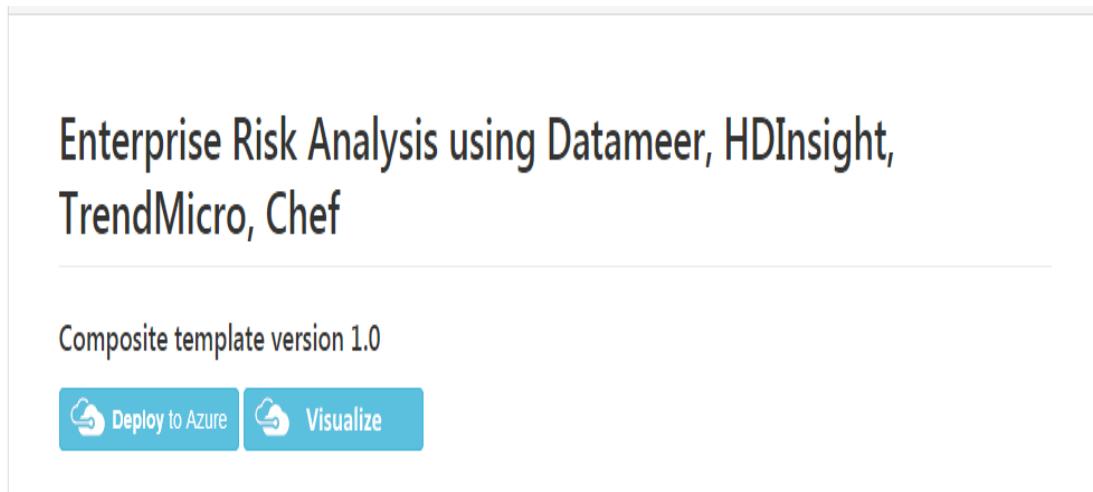
2 How to deploy this solution

This section will provide you the details of how to deploy this solution in the Microsoft Azure

- 1) Go to the below link available in the Github

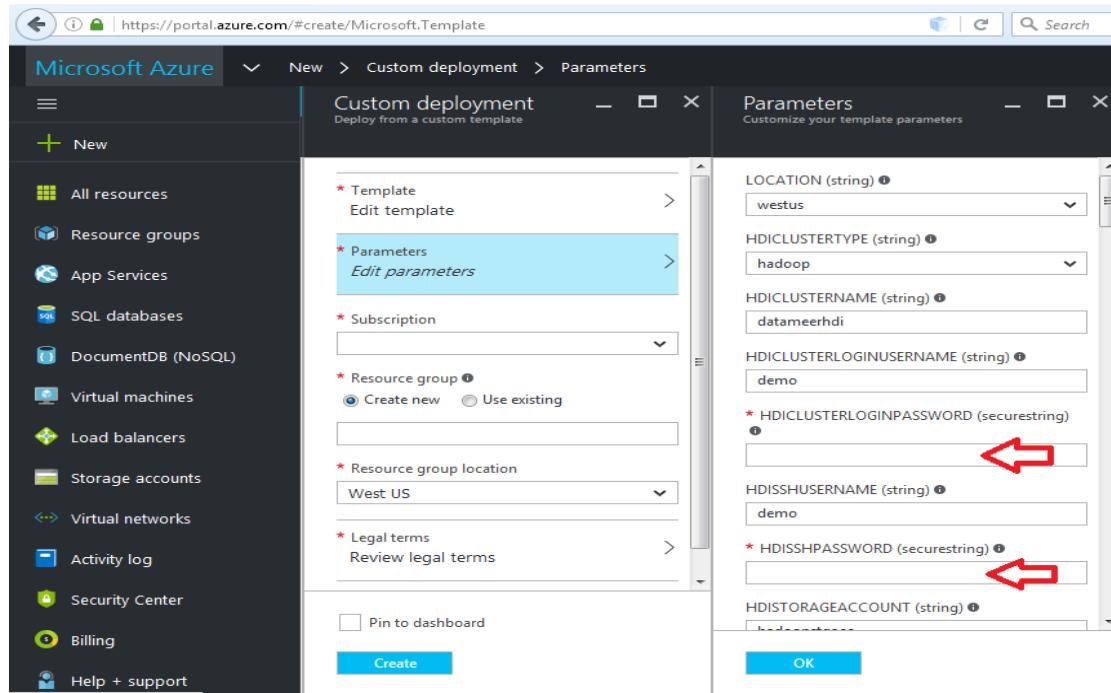
<https://github.com/AvyanConsultingCorp/azure-quickstart-templates/tree/master/datameer-trend-chef-businessintelligence>

- 2) Click on the “Deploy to Azure” in the page, this will take you to the page where you need to provide the parameters

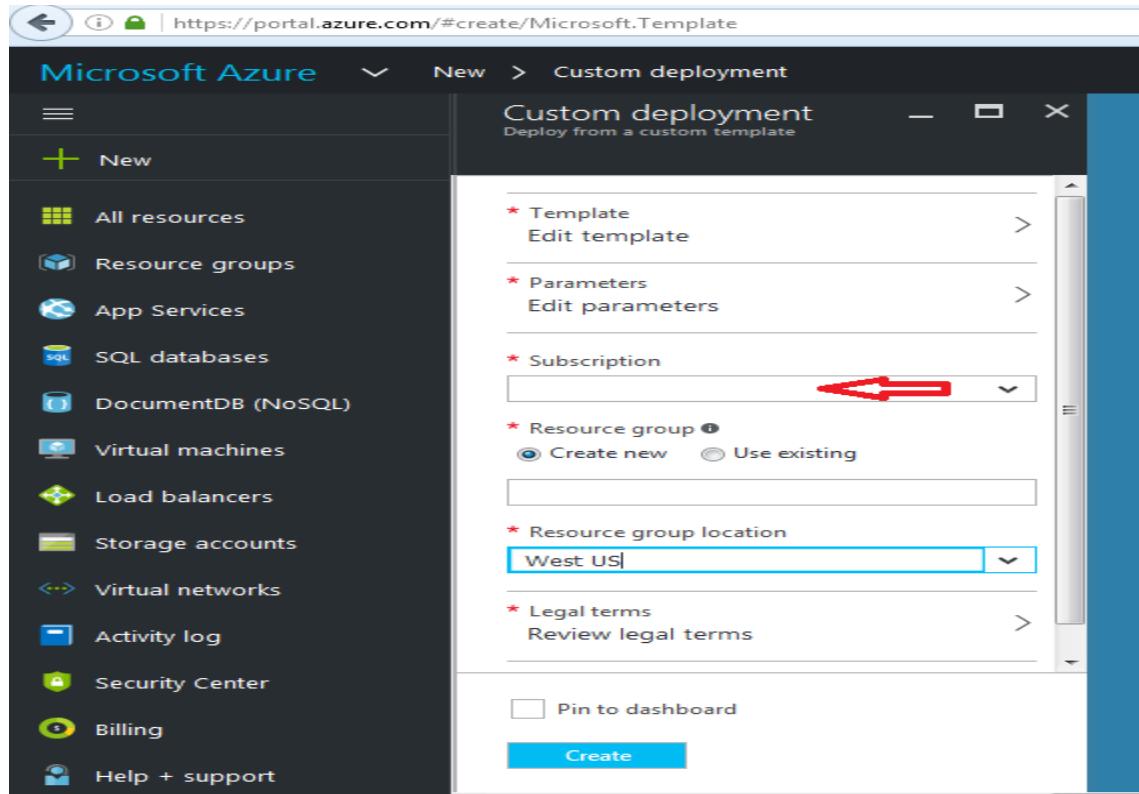


HOL Guide for Enterprise Risk Analysis

- 3) Provide the custom parameters for the solution accordingly and click “Next”

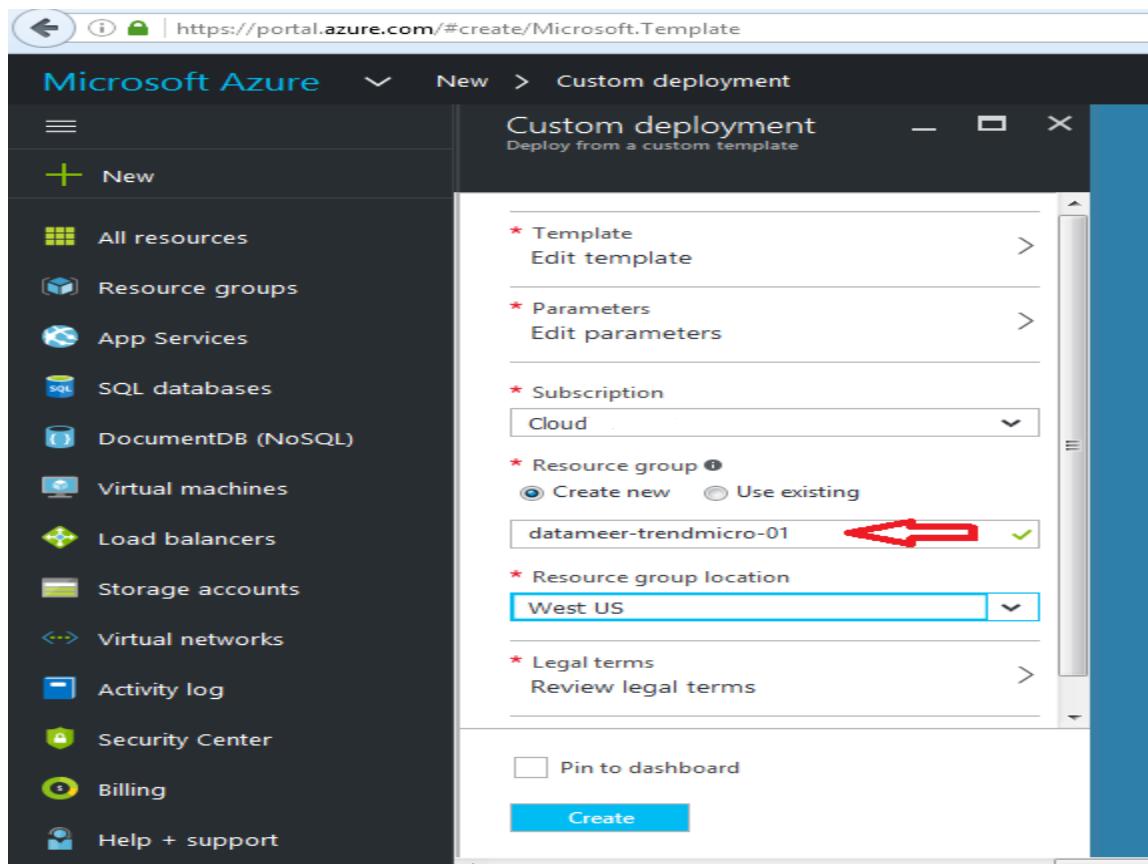


- 4) You need to select the subscription you want to deploy this solution

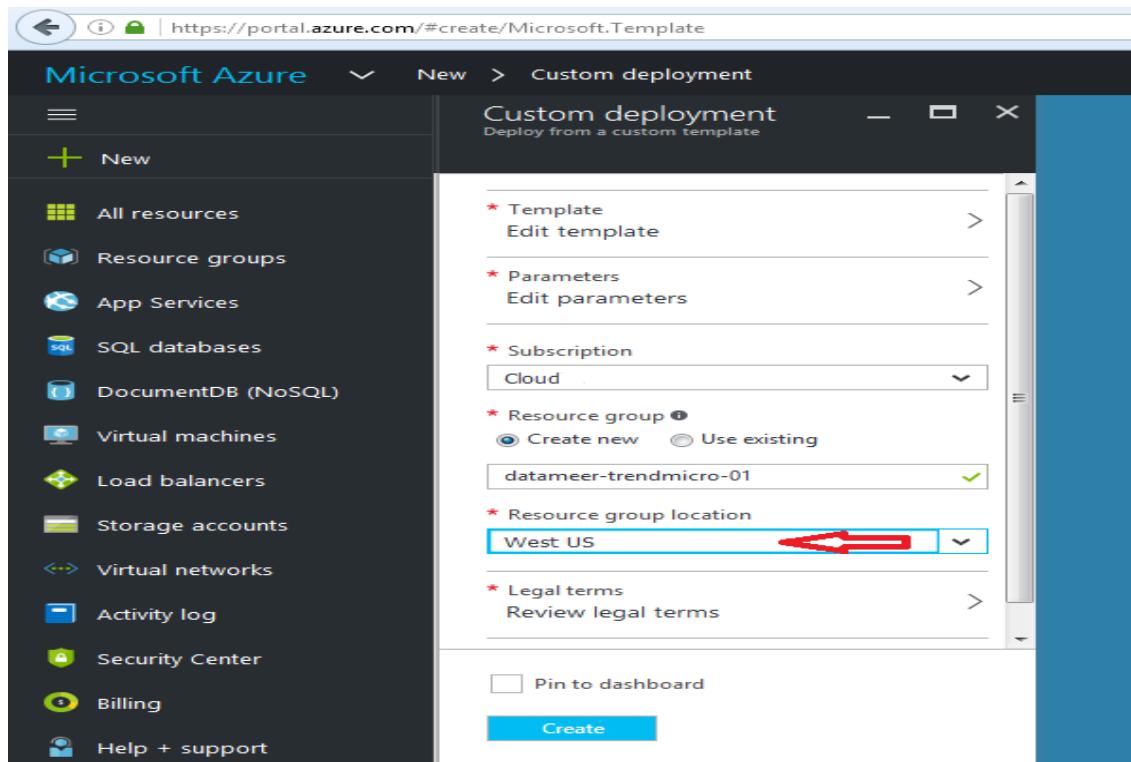


HOL Guide for Enterprise Risk Analysis

- 5) Either you can create a new “Resource Group” or use the existing resource group to deploy this solution

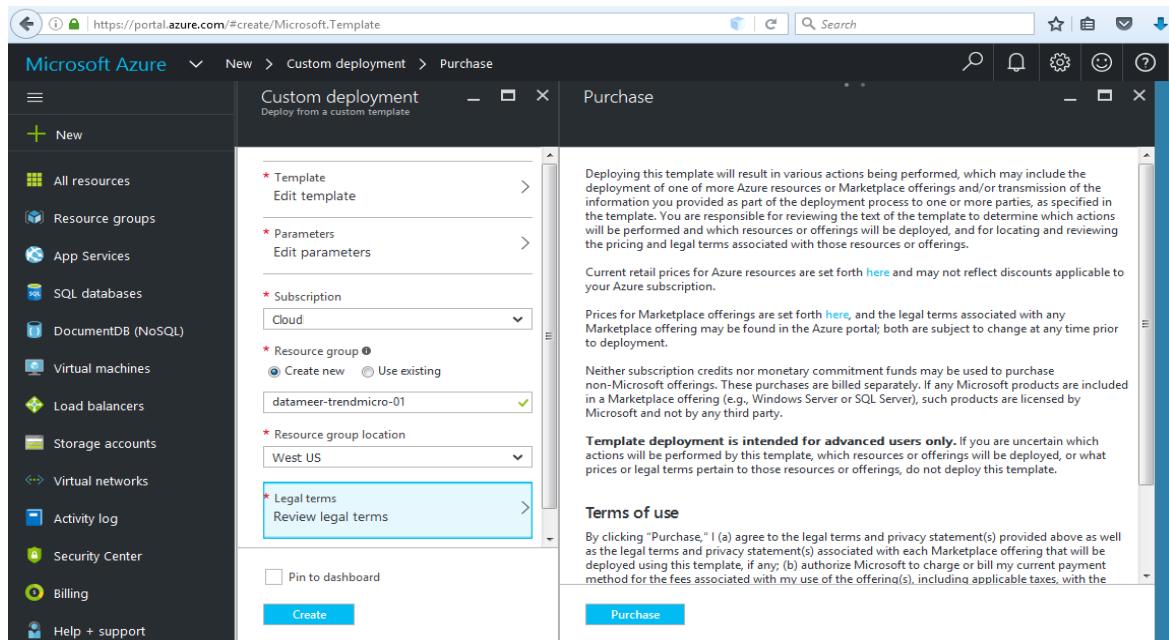


- 6) Select your choice of Region to deploy this solution,

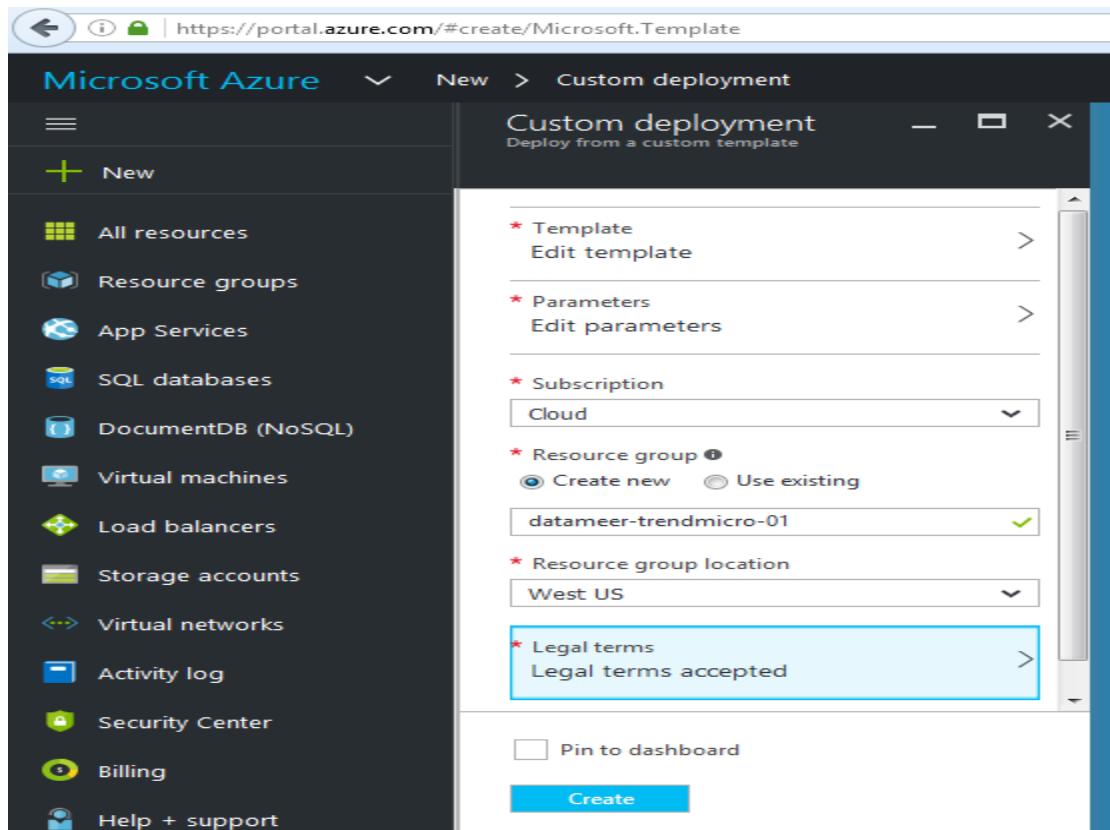


HOL Guide for Enterprise Risk Analysis

- 7) Accept the legal terms to deply the products from the Azure marketplace which includes, Datameer, TrendMicro and Chef. Click on the “Purchase” button for the same.



- 8) Click on the “Create” button to start deply the solution now



3 How to configure the components

3.1 Datameer

Datameer is the product used for the Big Data Analysis. It can be used many types of data and can connect to different data sources like storage, database etc. In this solution, the data (.csv) from the azure blog storage will be used too identify the Fraud detection using the credit card. The below sections will provide the details of the configuration of the data in the Datameer for the Big Data Analysis

3.2 TrendMicro

TrendMicro is the industry leading security product, which has the capabilities of

- Anti-Virus/Anti-malware detection and prevention.
- Web reputation
- Host based firewall
- Host based Intrusion detection and prevention
- File Integrity monitoring
- Log Inspection

TrendMicro DeepSecurity is an agent based security solution which will help the organisations to comply with all their security requirements.

This solution, showcases the Anti-Malware capabilities of the TrendMicro deepSecurity and below sections will provide the details of the configuration on the same.

4 Signing into Datameer UI

Copy samples to your storage account

Typically an enterprise will send the payment and other transactions to a data lake. For the purposes of a Hand-on-lab, our team has created a samples file with approx. 1.5Million records and is made available to you during the ignite 2016 time period here

https://msignite2016stg.blob.core.windows.net/samples/online-transactions-cc_masked.csv

If for some reason this location is not accessible to you, please do not hesitate to reach out to us @ azuremarketplace@avyanconsulting.com

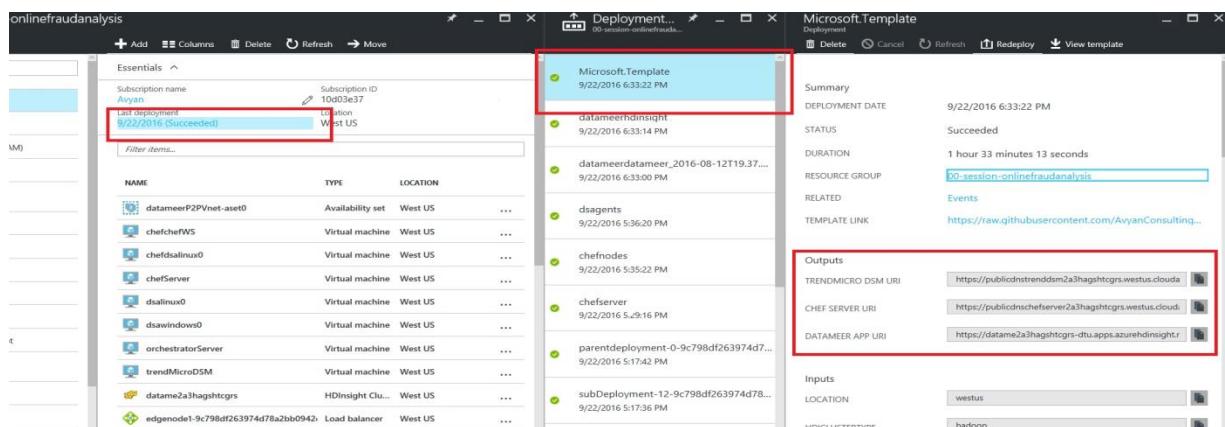
1. Download the file to your desktop
2. Navigate to the deployment resourcegroup and open the datameer storage account
3. Create a new blob container called “samples”
4. Upload the samples file to this container.

The fraud analysis is performed with the Business Analytics components of the solution, and namely Datameer and Azure HDInsight. All steps are executed in the Datameer UI. There are two ways that you can use to access the Datameer UI:

- Accessing it directly via the UI URL
- Accessing it via the Azure Management Portal

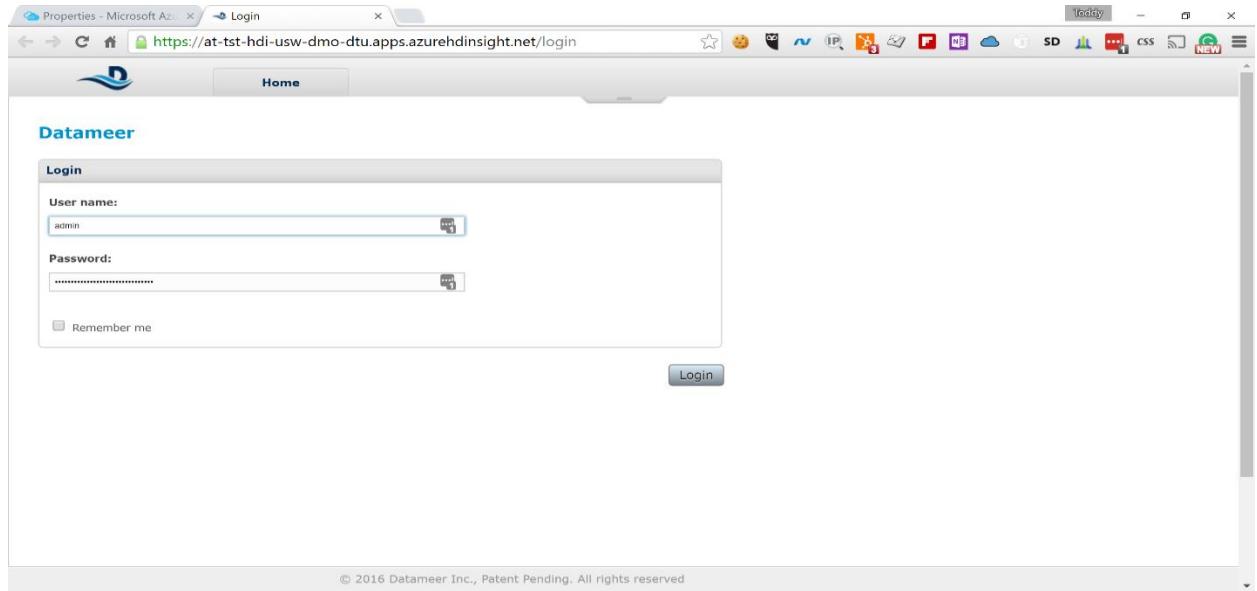
For the purposes of this HOL we will access the Datameer UI from the Azure Management Portal. Follow these steps:

1. In Azure Management Portal (<http://portal.azure.com>)
 - a. Click on specific resource group
 - b. Click on Last deployment date
 - c. Click on the Microsoft Template
 - d. Copy the Datameer URI from the Outputs



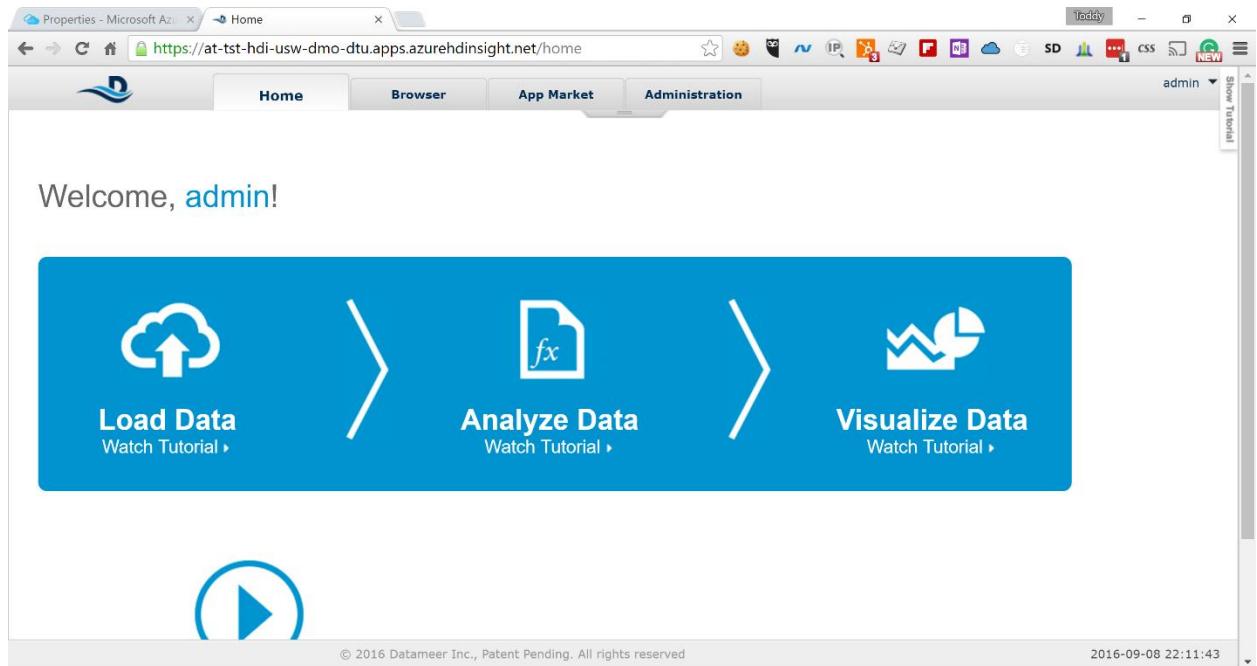
2. Paste the URI in your browser to load the Datameer UI

HOL Guide for Enterprise Risk Analysis



- You will be prompted to sign in using your Datameer username and password
- Sign into Datameer UI using the following default credentials:
username: *admin*
password: *admin*

You will see the Welcome Screen for Datameer and an introduction video will pop up



- Close the introduction video pop-up and click on the Browse tab

The screenshot shows the Datameer File Browser interface. On the left, there is a navigation tree with nodes like 'admin', 'Analytics', 'Data' (which is expanded to show 'Connections', 'DataLinks', 'ExportJobs', 'FileUploads', 'ImportJobs', 'Images', 'Library', 'Users', and 'Visualization'), and 'Fraud_POC'. The main area displays a table of connections:

Name	Type	Status	Last Processed	Records	Size	License total size	Owner
Con_AT_Sample_Da.dst		none	---	---	---	---	admin
Datameer server file.dst		none	---	---	---	---	system

To the right, an 'Information' panel shows details for the selected connection ('/Data/Connections'). It includes sections for 'Sharing', 'Owner' (set to 'admin'), 'Groups' (set to 'Not shared'), and 'Others' (with checkboxes for 'Read' and 'Write' checked). The bottom right corner shows the date and time: '2016-09-08 22:10:53'.

5 Configure Datameer to Fetch Data from Azure Storage

Datameer has more than 65 connectors built in, that allow various systems as data sources. For the purpose of this HOL we will use the Azure Storage connector and fetch the data from there. The assumption is that you have storage account data that contains the transaction data.

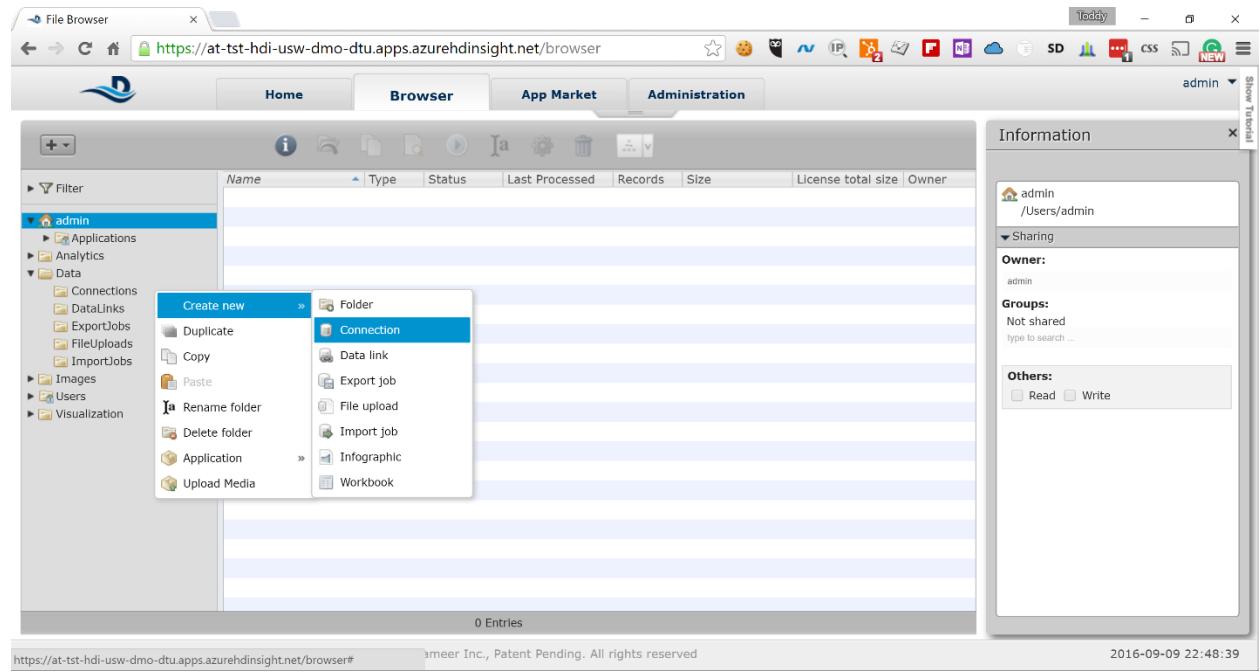
In order to configure Datameer to fetch the data from Azure Storage account you need to go through the following steps:

1. Expand the *Data* node in the left-side navigation and right-click on *Connections*

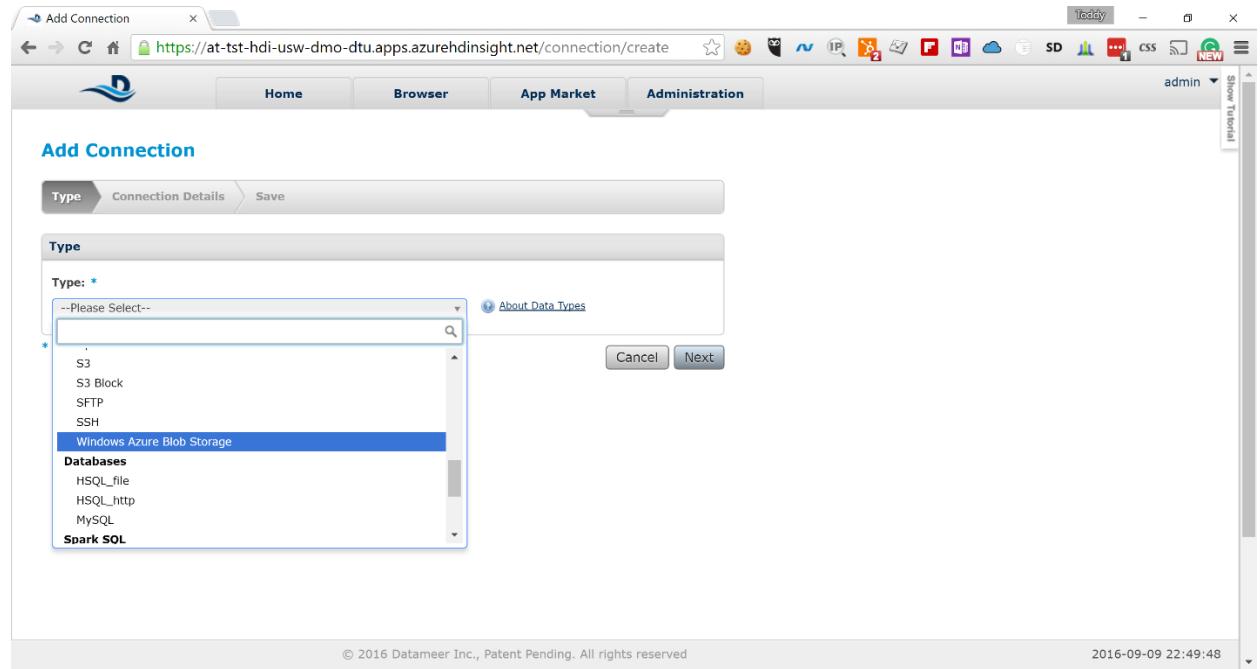
The screenshot shows the same Datameer interface as before, but with a context menu open over the 'Connections' folder under the 'Data' node in the navigation tree. The menu options include 'Create new', 'Duplicate', 'Copy', 'Paste', 'Rename folder', 'Delete folder', 'Application', and 'Upload Media'. The main area and information panel are visible in the background.

HOL Guide for Enterprise Risk Analysis

2. Select *Create new -> Connection*



3. In the *Type* drop-down, scroll down to *File* section and select *Windows Azure Blob Storage*



4. Click *Next* and fill in the following information on the next screen

HOL Guide for Enterprise Risk Analysis

The screenshot shows a web-based application interface for connecting to Windows Azure Blob Storage. The top navigation bar includes links for Home, Browser, App Market, and Administration, along with a user dropdown for 'admin'. The main title is 'Add Connection - Windows Azure Blob Storage'. The form has tabs for Type, Connection Details, and Save, with 'Connection Details' selected. The fields include:

- Storage name:** * (Input field containing 'mystorage')
Description: Name of the storage account.
Value: http://mystorage.blob.core.windows.net/
- Container name:** * (Input field)
- Access key:** * (Input field)
- Protocol:** * (Dropdown menu set to 'Secure')
- Connection usage:** (Dropdown menu set to 'Import/Export')

At the bottom, there are buttons for Cancel, Back, and Next. The status bar at the bottom indicates '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the date '2016-09-09 22:52:18'.

Storage Name: The name of the storage account where your data is

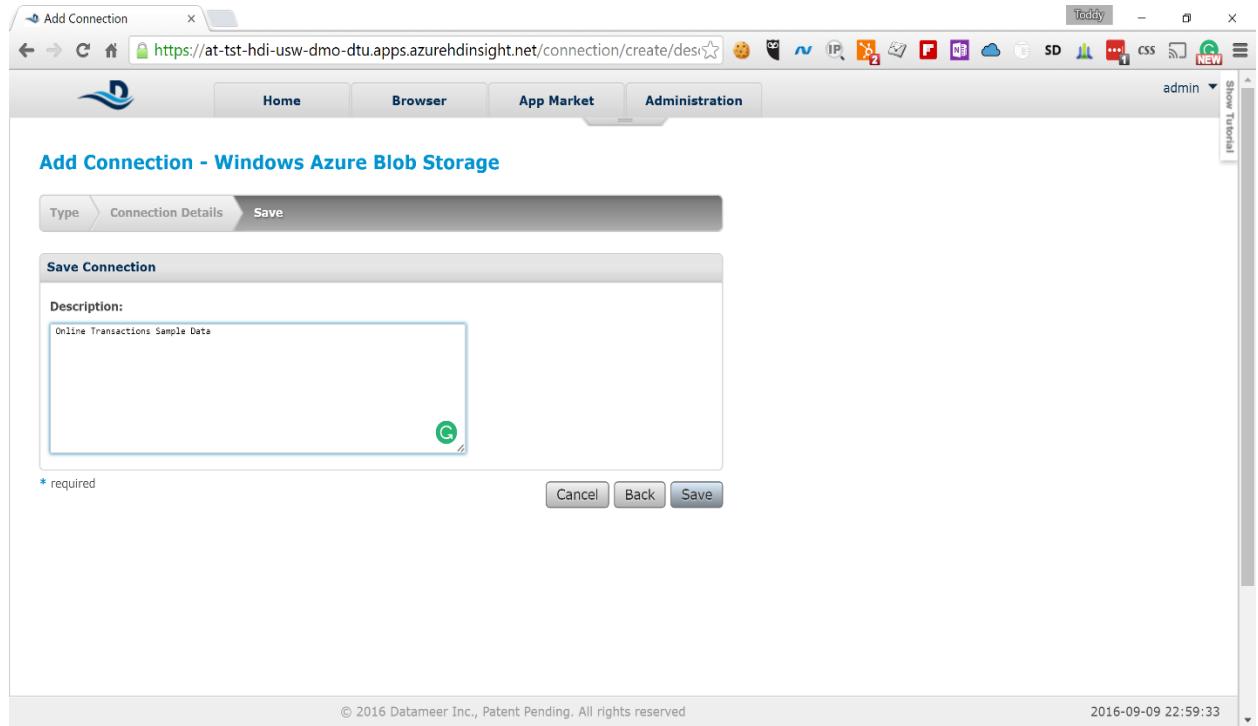
Container Name: The name of the container where your data is

Access Key: The key used to access the above storage account

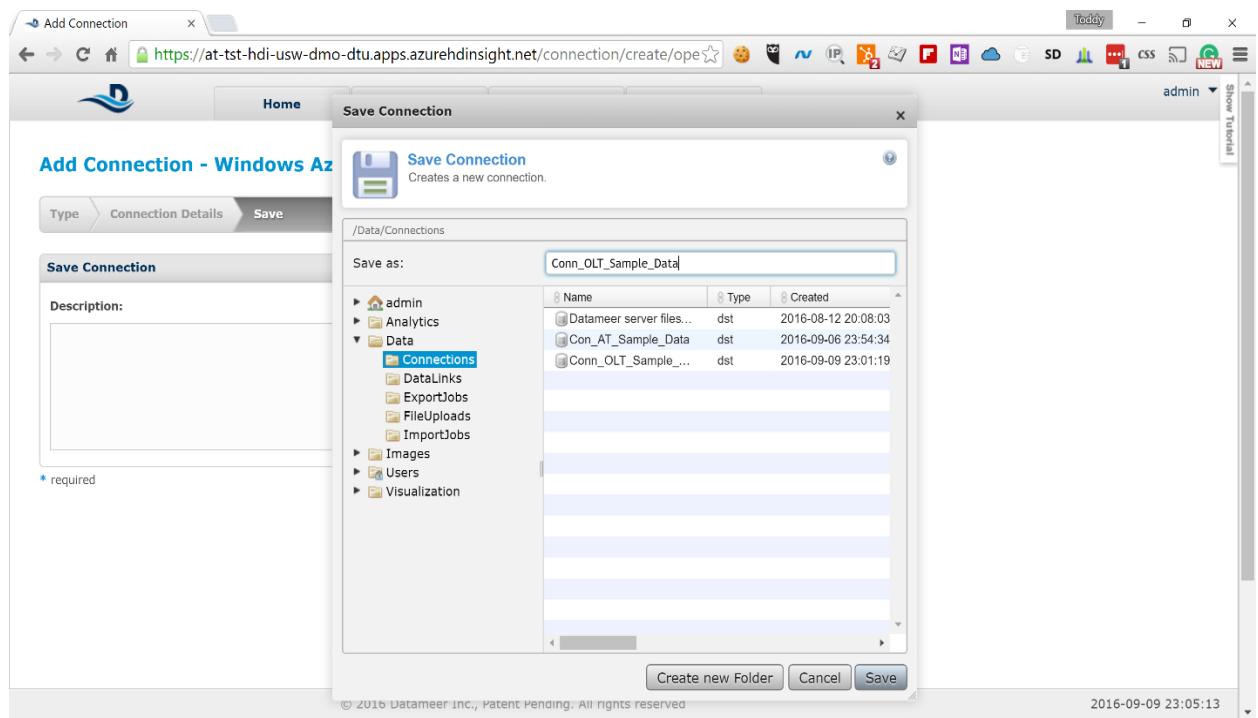
Leave the default values for *Protocol* and *Connection usage*.

5. Click *Next* and on the next screen type the following description for the connection:
“Online Transactions Sample Data”

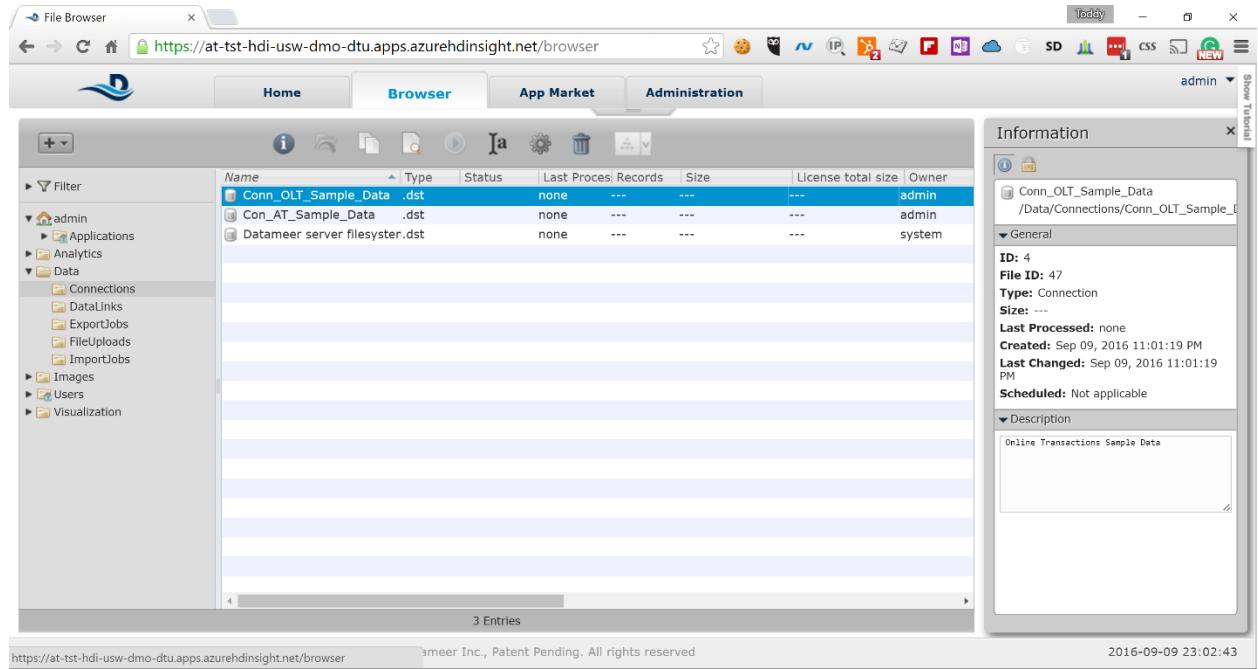
HOL Guide for Enterprise Risk Analysis



6. Click **Save** to save the connection and type the following name in the **Save as** field:
Conn_OLT_Sample_Data



7. Click **Save** again and you will see the new connection in the list



Now you have Datameer configured to look for data in the specified Azure Storage account and you can start creating your analysis.

6 Link, Clean and Prepare the Data

Before we start our analysis we need to tell Datameer which data exactly we want to analyze and make sure that it is in the correct format. The sample data we provided has the following two fields that need to be fixed before it is usable for analysis:

- The *timestamp* field is in ISO-8601 format, which needs to be converted into date/time field that Datameer can understand. We can do this conversion while we are linking the data.
- The *purchase_amount* field is a money field that is interpreted as a *STRING* by Datameer. We need to convert this to *FLOAT* in order to be able to do calculations. We will do that using Datameer formulas once we start our analysis.

Here are the steps to link the data for analysis.

1. Right-click on the *DataLinks* node in the left-side navigation and select *Create new -> Data link*

HOL Guide for Enterprise Risk Analysis

The screenshot shows the DataMeer browser interface at the URL <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/browser>. The 'Browser' tab is selected. On the left, a sidebar shows a tree structure with 'admin' as the root, containing 'Applications', 'Analytics', 'Data', 'Connections', and 'DataLinks'. Under 'DataLinks', there are 'ExportJobs', 'FileUploads', 'ImportJobs', 'Images', 'Users', and 'Visualization'. A context menu is open over the 'DataLinks' section, with 'Create new' selected. The submenu includes 'Folder', 'Connection', 'Data link' (which is highlighted in blue), 'Export job', 'File upload', 'Import job', 'Infographic', and 'Workbook'. To the right, a table lists three entries: 'DL_AT_Online_Trans.Ink', 'DL_AT_Online_Trans.Ink', and 'DL_AT_Reverse_IP .Ink'. An 'Information' panel on the right displays details for 'DataLinks' and sharing settings. The status bar at the bottom shows the URL, copyright information, and the date and time (2016-09-09 23:22:06).

- On the next screen click on the *Select Connection* button and select the *Conn_OLT_Sample_Data* connection you created previously

The screenshot shows the 'New Data Link' wizard at the URL <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/import-job/chooseCon>. The 'Connection' step is active. A 'Select Connection' dialog box is open, titled 'Select Connection'. It contains a message: 'Datameer will use the selected connection to import your data.' Below is a tree view of connections under '/Data/Connections/Conn_OLT_Sample_Data'. The 'Connections' folder is expanded, showing 'DataLinks', 'ExportJobs', 'FileUploads', 'ImportJobs', 'Images', 'Users', and 'Visualization'. Inside 'Connections', there are three entries: 'DataLink server files...', 'Con_AT_Sample_Data', and 'Conn_OLT_Sample...'. The last entry is selected. At the bottom of the dialog are 'Create new Folder', 'Cancel', and 'Select' buttons. The status bar at the bottom shows the URL, copyright information, and the date and time (2016-09-09 23:23:05).

HOL Guide for Enterprise Risk Analysis

3. Click on the *Select* button in the pop-up. Keep the default value *CSV/TSV* in the *File Type* drop down and click *Next*

The screenshot shows the 'New Data Link' interface. The top navigation bar includes tabs for Home, Browser, App Market, and Administration. The main content area has a breadcrumb trail: Connection > Data Details > Define Fields > Schedule > Save. A sub-section titled 'Choose Connection' displays a connection named 'Conn_OLT_Sample_Data' with buttons for 'Select Connection' and 'New Connection'. Below this is a 'File Type' section where 'CSV / TSV' is selected from a dropdown menu. A note indicates it is required. At the bottom right are 'Cancel' and 'Next' buttons.

4. On the next screen type the following in the *File Or Folder* field:
/samples/online-transactions-cc_masked.csv

The screenshot shows the 'New Data Link' interface on the 'Data Details' tab. The main content area has a breadcrumb trail: Connection > Data Details > Define Fields > Schedule > Save. A sub-section titled 'Basic' contains fields for 'Path Prefix' (set to '/') and 'File Or Folder' (containing the value '/samples/online-transactions-cc_non_masked.csv'). A note explains that wildcards are accepted. Another note provides instructions for using date patterns like %year%, %month%, etc. A 'Delimiter:' field is also present. At the bottom right are 'Save' and 'Cancel' buttons.

Scroll down to the bottom, keeping the default values for the rest of the fields, and click on *Next*

HOL Guide for Enterprise Risk Analysis

5. Datameer pre-fetches a representative sample of the data and shows it on the next screen

The screenshot shows the 'Define Fields' tab of the 'New Data Link' configuration. A table displays a sample of 15 rows from a dataset. Each row has a checkbox column followed by several columns representing different fields. The 'timestamp' field is currently set to 'STRING'. At the bottom of the table, there is a 'Rescan Schema' button.

all	ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_no	order_no	dasFileName	dasFilePath	dasLast
<input type="checkbox"/>	106.209.197.154	164,605	2016-08-01T00:00:00.000Z	\$3,093.82	684aa7fe-ab87-486...	3175-0372-1167-3...	973EV1	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	10.233.173.83	730,835	2016-08-01T00:00:00.000Z	\$3,844.52	af4b736c-834c-42c...	8935-620281-21432	3MS3UO	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	52.87.190.131	7,177,806	2016-08-01T00:00:00.000Z	\$3,605.26	34282d62-36f7-45...	3801-0325-0582-1...	I1QEWY	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	165.158.106.82	2,132,670	2016-08-01T00:00:00.000Z	\$2,523.34	71c4dc53-6016-4e...	8372-866834-66064	OQLQQE	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	12.15.120.180	8,596,261	2016-08-01T00:00:00.000Z	\$205.74	f7560939-a6e4-4c6...	6944-5099-7946-8...	V99HRT	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	11.216.44.6	2,946,179	2016-08-01T00:00:00.000Z	\$4,108.15	12d8a13e-5013-42f...	0283-1943-9261-5...	JLD5IC	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	185.183.151.166	6,839,961	2016-08-01T00:00:00.000Z	\$1,484.53	58e94bc0-48e7-4af...	4130-1703-0539-7...	YNHIYS	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	113.161.160.119	2,245,110	2016-08-01T00:00:00.000Z	\$969.83	1fee04ac-43c4-416...	3674-731714-47138	6BPQIW	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	28.245.64.112	1,223,107	2016-08-01T00:00:00.000Z	\$276.15	1149150d-daf5-49...	1951-397612-29310	WRXRIJ	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	35.143.156.34	2,220,765	2016-08-01T00:00:00.000Z	\$1,887.21	7309e62c-46a6-4cf...	8019-1124-7181-1...	86D7YH	online-transactions...	/at-samples/online-...	Sep 6, 2016

6. Click on the down-arrow for the field type under *timestamp* and change the type from *STRING* to *DATE*

The screenshot shows the 'Define Fields' tab of the 'New Data Link' configuration. The 'timestamp' field's type has been changed to 'DATE'. The rest of the schema remains the same as in the previous screenshot.

all	ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_no	order_no	dasFileName	dasFilePath	dasLast	
<input type="checkbox"/>	106.209.197.154	164,605	INTEGER	\$3,093.82	684aa7fe-ab87-486...	3175-0372-1167-3...	973EV1	online-transactions...	/at-samples/online-...	Sep 6, 2016	
<input type="checkbox"/>	10.233.173.83	730,835	FLOAT	\$3,844.52	af4b736c-834c-42c...	8935-620281-21432	3MS3UO	online-transactions...	/at-samples/online-...	Sep 6, 2016	
<input type="checkbox"/>	52.87.190.131	7,177,806	STRING	\$3,605.26	34282d62-36f7-45...	3801-0325-0582-1...	I1QEWY	online-transactions...	/at-samples/online-...	Sep 6, 2016	
<input type="checkbox"/>	165.158.106.82	2,132,670	BOOLEAN	\$2,523.34	71c4dc53-6016-4e...	8372-866834-66064	OQLQQE	online-transactions...	/at-samples/online-...	Sep 6, 2016	
<input type="checkbox"/>	12.15.120.180	8,596,261	BIG_DECIMAL	\$205.74	f7560939-a6e4-4c6...	6944-5099-7946-8...	V99HRT	online-transactions...	/at-samples/online-...	Sep 6, 2016	
<input type="checkbox"/>	11.216.44.6	2,946,179	BIG_INTEGER	2016-08-01T00:00:00.000Z	\$4,108.15	12d8a13e-5013-42f...	0283-1943-9261-5...	JLD5IC	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	185.183.151.166	6,839,961	DATE	2016-08-01T00:00:00.000Z	\$1,484.53	58e94bc0-48e7-4af...	4130-1703-0539-7...	YNHIYS	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	113.161.160.119	2,245,110	BOOLEAN	2016-08-01T00:00:00.000Z	\$969.83	1fee04ac-43c4-416...	3674-731714-47138	6BPQIW	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	28.245.64.112	1,223,107	BIG_DECIMAL	2016-08-01T00:00:00.000Z	\$276.15	1149150d-daf5-49...	1951-397612-29310	WRXRIJ	online-transactions...	/at-samples/online-...	Sep 6, 2016
<input type="checkbox"/>	35.143.156.34	2,220,765	BIG_INTEGER	2016-08-01T00:00:00.000Z	\$1,887.21	7309e62c-46a6-4cf...	8019-1124-7181-1...	86D7YH	online-transactions...	/at-samples/online-...	Sep 6, 2016

HOL Guide for Enterprise Risk Analysis

7. The dates in the timestamp are automatically marked in red because Datameer cannot parse the ISO-8601 date by default and an input field appears under the field type drop-down

The screenshot shows the Datameer interface for creating a new data link. The 'Define Fields' tab is active. A table lists various fields: ip_address, user_id, timestamp, purchase_amount, transaction_id, credit_card_no, order_no, desFileName, desFilePath, and desLast. The 'timestamp' field is highlighted with a red border, indicating it is an ISO-8601 date. The table contains several rows of transaction data.

ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_no	order_no	desFileName	desFilePath	desLast
106.209.197.154	164,605	2016-08-01T00:00:00	\$3,093.82	684aa7fe-ab87-486..	3175-0372-1167-3..	973EV1	online-transactions...	/at-samples/online-...	Sep 6, 20
10.233.173.83	730,835	2016-08-01T00:00:00	\$3,844.52	a4fb736c-834c-42c..	8935-620281-21432	3MS3UO	online-transactions...	/at-samples/online-...	Sep 6, 20
52.87.190.131	7,177,806	2016-08-01T00:00:00	\$3,605.26	34282d62-36f7-45..	3801-0325-0582-1..	I1QEWT	online-transactions...	/at-samples/online-...	Sep 6, 20
165.158.106.82	2,132,670	2016-08-01T00:00:00	\$2,523.34	71c4dc53-6016-4e..	8372-866834-66064	OQLQQE	online-transactions...	/at-samples/online-...	Sep 6, 20
12.15.120.180	8,596,261	2016-08-01T00:00:00	\$205.74	f7560939-a6e4-4c6..	6944-5099-7946-8..	V99HRT	online-transactions...	/at-samples/online-...	Sep 6, 20
11.216.44.6	2,946,179	2016-08-01T00:00:00	\$4,108.15	12d8a13e-5013-42f..	0283-1943-9261-5..	JLD5IC	online-transactions...	/at-samples/online-...	Sep 6, 20
185.183.151.166	6,839,961	2016-08-01T00:00:00	\$1,484.53	58e94bc0-48e7-4af..	4130-1703-0539-7..	YNHIYS	online-transactions...	/at-samples/online-...	Sep 6, 20
113.161.160.119	2,245,110	2016-08-01T00:00:00	\$969.83	1fee04ac-43c4-416..	3674-731714-47138	6BPQIW	online-transactions...	/at-samples/online-...	Sep 6, 20
28.245.64.112	1,223,107	2016-08-01T00:00:00	\$276.15	1149150d-daf5-49..	1951-397612-29310	WRXRJ1D	online-transactions...	/at-samples/online-...	Sep 6, 20
35.143.156.34	2,220,765	2016-08-01T00:00:00	\$1,887.21	7309e62c-46a6-4cf..	8019-1124-7181-1..	86D7YH	online-transactions...	/at-samples/online-...	Sep 6, 20

8. Type the following pattern in the field
yyyy-MM-dd'T'HH:mm:ss'Z'

The screenshot shows the Datameer interface for creating a new data link. The 'Define Fields' tab is active. The 'timestamp' field now contains a correctly formatted date ('Aug 1, 2016 12:00:00...'). The table contains the same transaction data as the previous screenshot.

ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_no	order_no	desFileName	desFilePath	desLast
106.209.197.154	164,605	Aug 1, 2016 12:00:00	\$3,093.82	684aa7fe-ab87-486..	3175-0372-1167-3..	973EV1	online-transactions...	/at-samples/online-...	Sep 6, 20
10.233.173.83	730,835	Aug 1, 2016 12:00:00	\$3,844.52	a4fb736c-834c-42c..	8935-620281-21432	3MS3UO	online-transactions...	/at-samples/online-...	Sep 6, 20
52.87.190.131	7,177,806	Aug 1, 2016 12:00:00	\$3,605.26	34282d62-36f7-45..	3801-0325-0582-1..	I1QEWT	online-transactions...	/at-samples/online-...	Sep 6, 20
165.158.106.82	2,132,670	Aug 1, 2016 12:00:00	\$2,523.34	71c4dc53-6016-4e..	8372-866834-66064	OQLQQE	online-transactions...	/at-samples/online-...	Sep 6, 20
12.15.120.180	8,596,261	Aug 1, 2016 12:00:00	\$205.74	f7560939-a6e4-4c6..	6944-5099-7946-8..	V99HRT	online-transactions...	/at-samples/online-...	Sep 6, 20
11.216.44.6	2,946,179	Aug 1, 2016 12:00:00	\$4,108.15	12d8a13e-5013-42f..	0283-1943-9261-5..	JLD5IC	online-transactions...	/at-samples/online-...	Sep 6, 20
185.183.151.166	6,839,961	Aug 1, 2016 12:00:00	\$1,484.53	58e94bc0-48e7-4af..	4130-1703-0539-7..	YNHIYS	online-transactions...	/at-samples/online-...	Sep 6, 20
113.161.160.119	2,245,110	Aug 1, 2016 12:00:00	\$969.83	1fee04ac-43c4-416..	3674-731714-47138	6BPQIW	online-transactions...	/at-samples/online-...	Sep 6, 20
28.245.64.112	1,223,107	Aug 1, 2016 12:00:00	\$276.15	1149150d-daf5-49..	1951-397612-29310	WRXRJ1D	online-transactions...	/at-samples/online-...	Sep 6, 20
35.143.156.34	2,220,765	Aug 1, 2016 12:00:00	\$1,887.21	7309e62c-46a6-4cf..	8019-1124-7181-1..	86D7YH	online-transactions...	/at-samples/online-...	Sep 6, 20

Datameer immediately parses the data in the field and shows it in the correct format. Scroll to the bottom of the screen and click on *Next*

HOL Guide for Enterprise Risk Analysis

9. On the next screen keep the default value for *Trigger* and click on *Next*

The screenshot shows a web browser window titled "New Data Link". The URL is <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/import-job/create/trig>. The page has a header with tabs: Home, Browser, App Market, Administration, and a user dropdown. Below the header is a breadcrumb navigation: Connection > Data Details > Define Fields > Schedule > Save. The main content area is titled "Refresh Sample Data" and contains a "Trigger:" section. It shows two radio button options: "Manually" (selected) and "On a schedule". A tooltip explains: "This determines how and when to refresh the sample data used by analysts to create workbooks against data links. If the linked data changes frequently, the refresh rate should be higher." Below this is an "Advanced" section with a "required" note. At the bottom are "Cancel", "Back", and "Next" buttons. The footer includes copyright information: "© 2016 Datameer Inc., Patent Pending. All rights reserved" and a timestamp: "2016-09-09 23:40:21".

10. On the next screen type a meaningful description for the DataLink and click on *Save*

The screenshot shows a web browser window titled "New Data Link". The URL is <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/import-job/create/desc>. The page has a header with tabs: Home, Browser, App Market, Administration, and a user dropdown. Below the header is a breadcrumb navigation: Connection > Data Details > Define Fields > Schedule > Save. The main content area is titled "Save Data Link" and contains a "Description:" field with the text "Data link for the online transactions sample data". Below this is a "Generate now:" section with a checked checkbox for "Generate sample immediately after save". An "Advanced" section with a "required" note is visible at the bottom. At the very bottom are "Cancel", "Back", and "Save" buttons. The footer includes copyright information: "© 2016 Datameer Inc., Patent Pending. All rights reserved" and a timestamp: "2016-09-09 23:41:43".

HOL Guide for Enterprise Risk Analysis

11. Type the following name in the *Save as* field for the DataLink and click on the *Save* button

The screenshot shows the 'Save Data Link' dialog box. The 'Save as:' field is populated with 'DL_OLT_Sample_Data'. The left pane shows a navigation tree with 'admin', 'Analytics', 'Data' (selected), 'Connections', 'DataLinks' (selected), 'ExportJobs', 'FileUploads', 'ImportJobs', 'Images', 'Users', and 'Visualization'. The right pane lists existing DataLinks: 'DL_AT_Online_Trans_Non_M...' (Ink, 2016-09-06), 'DL_AT_Reverse_IP' (Ink, 2016-09-07), and 'DL_AT_Online_Trans_Masked' (Ink, 2016-09-07). At the bottom are 'Create new Folder', 'Cancel', and 'Save' buttons.

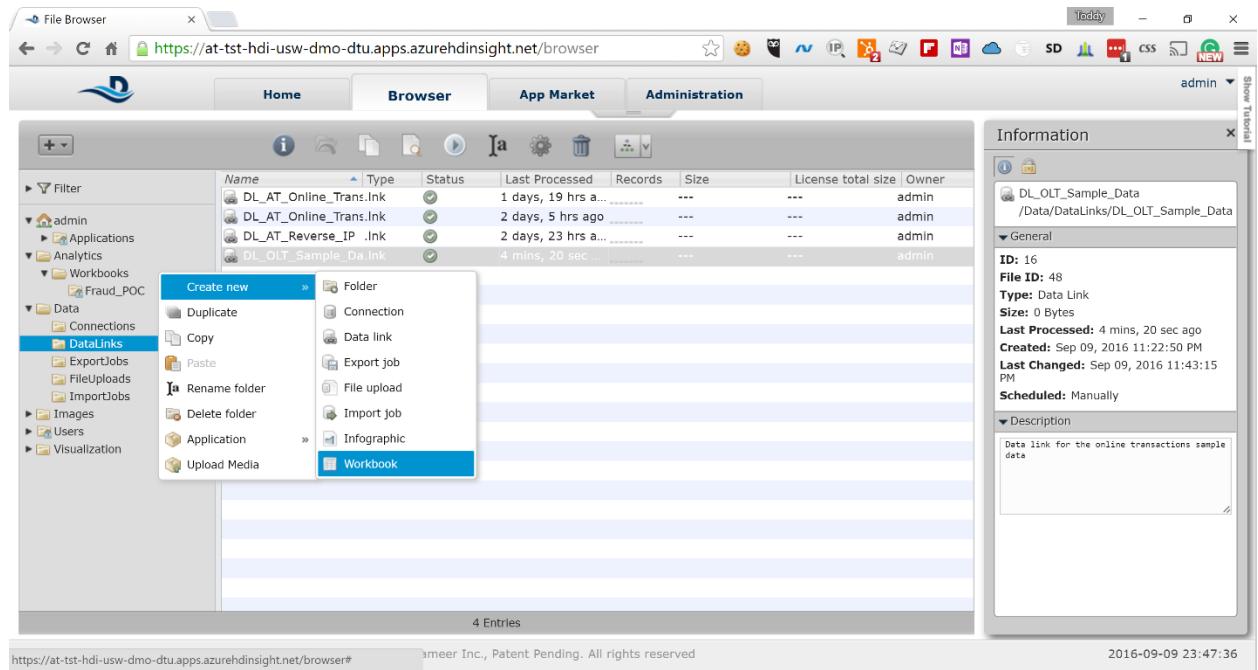
12. The new data link will appear in the list of data links on the next screen

The screenshot shows the 'File Browser' interface. The sidebar navigation tree includes 'admin', 'Applications', 'Analytics', 'Data' (selected), 'Connections', 'DataLinks' (selected), 'ExportJobs', 'FileUploads', 'ImportJobs', 'Images', 'Users', and 'Visualization'. The main area displays a table of DataLinks with columns: Name, Type, Status, Last Processed, Records, Size, License total size, and Owner. Four entries are listed: 'DL_AT_Online_Trans.Ink', 'DL_AT_Online_Trans.Ink', 'DL_AT_Reverse_IP .Ink', and 'DL_OLT_Sample_Da.Ink'. The 'DL_OLT_Sample_Da.Ink' entry is highlighted. To the right, an 'Information' panel shows detailed information: ID: 16, File ID: 48, Type: Data Link, Size: 0 Bytes, Last Processed: 1 mins, 35 sec ago, Created: Sep 09, 2016 11:22:50 PM, Last Changed: Sep 09, 2016 11:43:15 PM, Scheduled: Manually. The 'Description' field contains the text 'Data link for the online transactions sample data'.

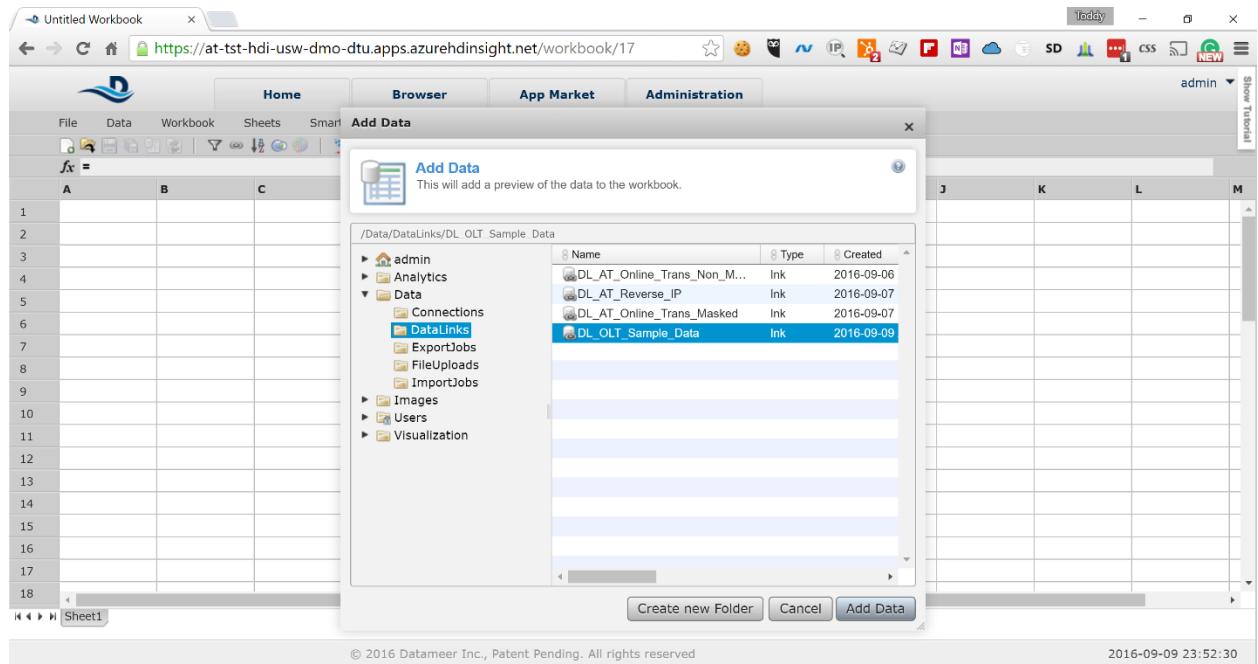
Now you have your dataset linked and have done some preliminary clean-up of the data. Next we will create a workbook where we will finish cleaning up our data and do our analysis. Here are the steps.

HOL Guide for Enterprise Risk Analysis

13. Expand the *Analytics* node in the left-side navigation, right-click on the *Workbook* node and select *Create new -> Workbook*



14. A new workbook is created and a pop-up window is shown asking you to select the dataset you want to use for analysis. Expand the *Data* node in the pop-up navigation and click on the *DataLinks* node. In the right-side window select the data link that you just created.



HOL Guide for Enterprise Risk Analysis

15. Click on *Add Data* to load a sample of the dataset in the workbook sheet

The screenshot shows the Datameer interface with a sample dataset titled "DL_OLT_Sample_Data". The dataset contains 18 rows of data with columns: ip_address, user_id, timestamp, purchas..., transacti..., credit_c..., order_no, H, I, J, K, L, M. The data includes various IP addresses, user IDs, timestamps, purchase amounts, transaction IDs, credit card numbers, and order numbers. The interface is similar to Excel, with a ribbon menu at the top and a toolbar below it.

	ip_address	user_id	timestamp	purchas...	transacti...	credit_c...	order_no	H	I	J	K	L	M
1	106.209.197.1..	164,605	Aug 1, 2016 1..	\$3,093.82	684aa7fe-ab87..	*****_*****_*...	973EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1..	\$3,844.52	af4b736c-834c..	*****_*****_*...	3MS3UO						
3	52.87.190.131	7,177,806	Aug 1, 2016 1..	\$3,605.26	34282d62-36f..	*****_*****_*...	I1QEWFY						
4	165.158.106.82	2,132,670	Aug 1, 2016 1..	\$2,523.34	71c4dc53-601..	*****_*****_*...	OQLQQE						
5	12.15.120.180	8,596,261	Aug 1, 2016 1..	\$205.74	f7560939-a6e..	*****_*****_*...	V99HRT						
6	11.216.44.6	2,946,179	Aug 1, 2016 1..	\$4,108.15	12d8a13e-501..	*****_*****_*...	JLD5IC						
7	185.183.151.1..	6,839,961	Aug 1, 2016 1..	\$1,484.53	58e94bc0-48e..	*****_*****_*...	YNHIYS						
8	113.161.160.1..	2,245,110	Aug 1, 2016 1..	\$969.83	1fee04ac-43c4..	*****_*****_*...	6BPQIW						
9	28.245.64.112	1,223,107	Aug 1, 2016 1..	\$276.15	1149150d-daf..	*****_*****_*...	WRXRJ						
10	35.143.156.34	2,220,765	Aug 1, 2016 1..	\$1,887.21	7309e62c-46a..	*****_*****_*...	86D7YH						
11	246.118.83.220	3,710,930	Aug 1, 2016 1..	\$3,880.10	526be209-479..	*****_*****_*...	I90L1K						
12	234.74.136.186	4,517,350	Aug 1, 2016 1..	\$775.60	bcd4841-6b9..	*****_*****_*...	NUKKNW						
13	246.193.132.62	5,172,602	Aug 1, 2016 1..	\$2,000.91	bfa1e8df-629c..	*****_*****_*...	MDHV8J						
14	64.245.31.254	7,438,752	Aug 1, 2016 1..	\$2,687.91	0952b115-2dd..	*****_*****_*...	Z78G96						
15	2.42.116.20	3,894,627	Aug 1, 2016 1..	\$3,313.53	52ed8371-00c..	*****_*****_*...	X1GR4C						
16	215.55.147.149	9,393,439	Aug 1, 2016 1..	\$1,433.69	6e1a9a2d-25f..	*****_*****_*...	CUP0K2						
17	173.238.152.96	1,094,163	Aug 1, 2016 1..	\$2,038.83	892c9168-362..	*****_*****_*...	R88IS9						
18	173.238.152.94	2,161,223	Aug 1, 2016 1..	\$1,000.00	*****_*****_*...						

The UI you are presented with is very similar to Excel and uses the same concepts.

16. Right-click on the *DL_OLT_Sample_Data* sheet at the bottom of the screen next and select *Duplicate*

The screenshot shows the Datameer interface with a context menu open over the "DL_OLT_Sample_Data" sheet. The menu options include "Rename", "Delete", "Duplicate", and "Move". The "Duplicate" option is highlighted. The rest of the interface is identical to the previous screenshot, showing the sample dataset in a worksheet.

https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/workbook/20#

HOL Guide for Enterprise Risk Analysis

17. Keep all of the fields selected in the pop-up and click on the *Create Sheet Copy* button

The screenshot shows the Data Miner interface with a 'Duplicate Worksheet' dialog box open. The dialog lists the columns to be copied: ip_address, user_id, timestamp, purchase_amount, transaction_id, credit_card_no, and order_no. All checkboxes are checked. Below the list are 'Cancel' and 'Create Sheet Copy' buttons.

18. A new copy of the sheet is created that contains all of the data from the original sheet. Click on the *purchase_amount* column to enable the *f_x* field for that column available on top of the sheet

The screenshot shows the Data Miner interface with the newly copied sheet. The purchase_amount column now has an *f_x* field at the top, indicating it is selected.

HOL Guide for Enterprise Risk Analysis

19. Type the following in the f_x field and press *Enter*

```
FLOAT(SUBSTITUTEALL(SUBSTR(#DL_OLT_Sample_Data!purchase_amount;1);",","))
```

The formula strips the \$ (dollar) sign in front of the amount, removes all commas and converts the string to FLOAT. Now you can use numeric functions to perform calculations on the field.

A screenshot of a DataMeer workbook titled "Untitled Workbook". The URL is https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/workbook/20. The top navigation bar includes Home, Browser, App Market, and Administration. The main area shows a table with columns: ip_address, user_id, timestamp, purchas..., transacti..., credit_c..., order_no, H, I, J, K, L, M. The first row contains the formula =FLOAT(SUBSTITUTEALL(SUBSTR(#DL_OLT_Sample_Data!purchase_amount;1);",",")). The data rows show various purchase records with timestamps from August 1, 2016, and amounts ranging from 3,093.82 to 2,038.83. The bottom of the screen shows the footer © 2016 Datameer Inc., Patent Pending. All rights reserved and the date 2016-09-10 03:01:39.

20. Right-click on the sheet name at the bottom of the screen to show the context menu for *Sheet1* and select *Rename*

A screenshot of the same DataMeer workbook as the previous one, but with a context menu open over the "Sheet1" tab at the bottom. The menu options are Rename, Delete, Duplicate, and Move. The main area of the screen shows the same table of purchase data. The bottom of the screen shows the footer © 2016 Datameer Inc., Patent Pending. All rights reserved and the date 2016-09-10 03:03:35.

HOL Guide for Enterprise Risk Analysis

21. Rename *Sheet1* to *Transaction_Data*

	ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_type	order_no	H	I	J	K	L	M
1	106.209.197.1...	164,605	Aug 1, 2016 1...	3,093.82	684aa7fe-ab87...	*****-*****-*...	973EV1						
2	10.233.173.83	730,835	Aug 1, 2016 1...	3,844.52	af4b736c-834c...	*****-*****-*...	3M3SU0						
3	52.87.190.131	7,177,806	Aug 1, 2016 1...	3,605.26	34282d62-36f...	*****-*****-*...	I1QEWTY						
4	165.158.106.82	2,132,670	Aug 1, 2016 1...	2,523.34	71c4dc53-601...	*****-*****-*...	OQLQQE						
5	12.15.120.180	8,596,261	Aug 1, 2016 1...	205.74	f7560939-a6e...	*****-*****-*...	V99HRT						
6	11.216.44.6	2,946,179	Aug 1, 2016 1...	4,108.15	12d8a13e-501...	*****-*****-*...	JLD5IC						
7	185.183.151.1...	6,839,961	Aug 1, 2016 1...	1,484.53	58e94bc0-48e...	*****-*****-*...	YNHIYS						
8	113.161.160.1...	2,245,110	Aug 1, 2016 1...	969.83	1fee04ac-43c4...	*****-*****-*...	6BPQIW						
9	28.245.64.112	1,223,107	Aug 1, 2016 1...	276.15	1149150d-daf...	*****-*****-*...	WRXRJ1						
10	35.143.156.34	2,220,765	Aug 1, 2016 1...	1,887.21	7309e62c-46a...	*****-*****-*...	86D7YH						
11	246.118.83.220	3,710,930	Aug 1, 2016 1...	3,880.1	526be209-479...	*****-*****-*...	IA9L1K						
12	234.74.136.186	4,517,350	Aug 1, 2016 1...	775.6	bcd4841-6b9...	*****-*****-*...	NUKKNW						
13	246.193.132.62	5,172,602	Aug 1, 2016 1...	2,000.91	bfa1e8df-629c...	*****-*****-*...	MDHV8J						
14	64.245.31.254	7,438,752	Aug 1, 2016 1...	2,687.91	0952b115-2dd...	*****-*****-*...	Z7BG96						
15	2.42.116.20	3,894,627	Aug 1, 2016 1...	3,313.53	52ed8371-00c...	*****-*****-*...	X1GR4C						
16	215.55.147.149	9,393,439	Aug 1, 2016 1...	1,433.69	6e1a9a2d-25f...	*****-*****-*...	CUPOK2						
17	173.238.152.96	1,094,163	Aug 1, 2016 1...	2,038.83	892c168-362...	*****-*****-*...	R88IS9						
18													

With this we are done with the clean-up of our data and are ready to perform our analysis.

7 Perform Analysis to Identify Outliers

The goal of our analysis is to identify unusual purchasing patterns that deviate from a well-established norm. If we notice something unusual this may be sign that fraud may be committed. For the purpose of this HOL we will be looking for period during the month, in which the transactions significantly deviate from the normal patters during the rest of the month. Here the steps:

1. Click on the + sign next to the *Transaction_Data* sheet to create an empty sheet for analysis

Formula Builder

Apply Formulas in Your Sheets

Easily create formulas by selecting an appropriate function. After making a selection, enter the arguments into the function.

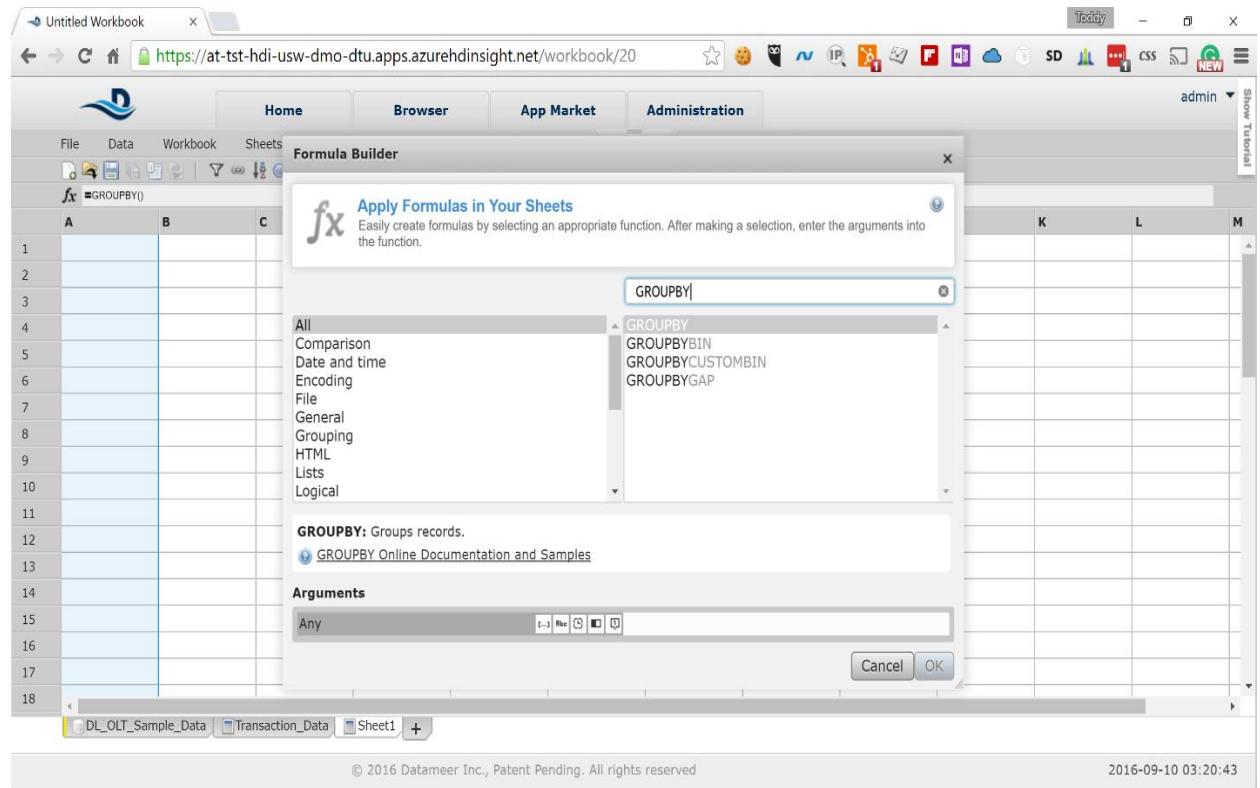
All

- ABS
- ACOS
- ACOSH
- ADD
- ADDTODATE
- AFTER
- ANALYZE_POLARITY
- AND
- ASDATE
- ASIN

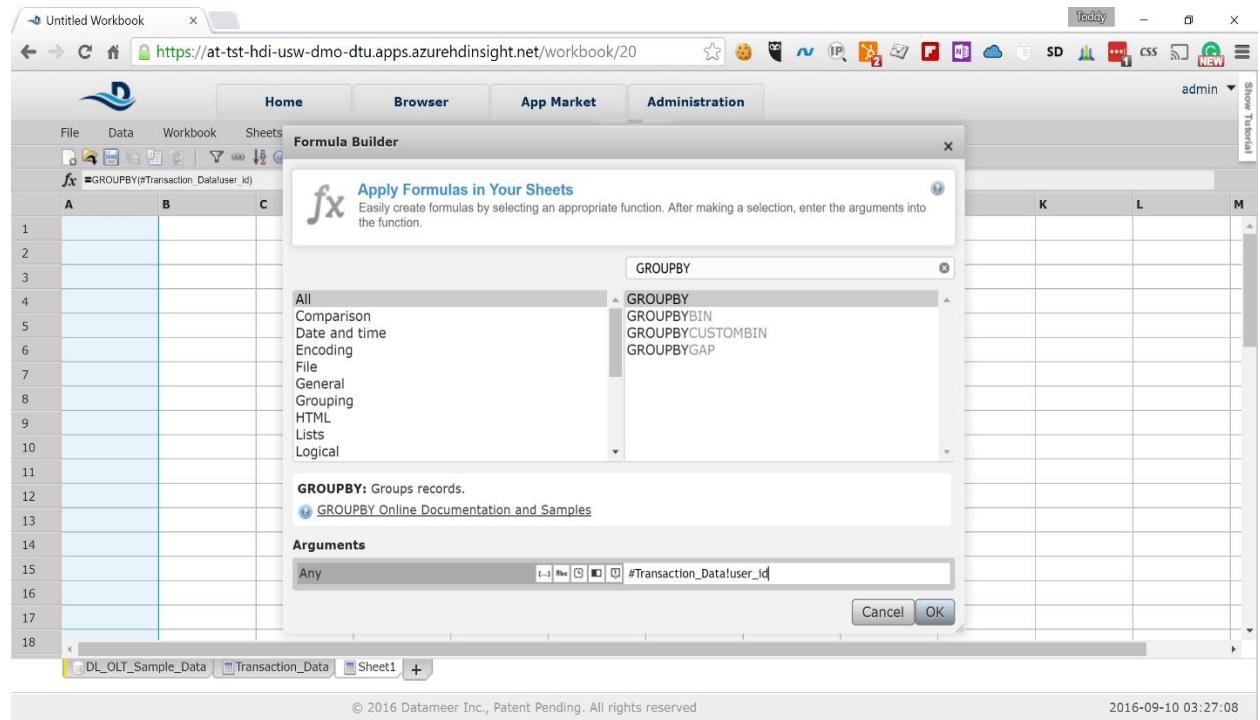
Cancel OK

HOL Guide for Enterprise Risk Analysis

2. In the input field in the pop-up type **GROUPBY** to filter the functions and select **GROUPBY** from the list



3. In the *Arguments* input field at the bottom of the pop-up type **#Transaction_Data!user_id** to group the data by user identifier and click on the **OK** button



HOL Guide for Enterprise Risk Analysis

4. The first column of the sheet will be populated with list of unique user identifiers

The screenshot shows a Data Miner interface with a table titled 'user_id' containing 18 rows of data. The columns are labeled B through M. The data is as follows:

	B	C	D	E	F	G	H	I	J	K	L	M
1	1,724											
2	1,956											
3	2,191											
4	3,400											
5	4,680											
6	5,630											
7	12,144											
8	16,170											
9	19,052											
10	19,395											
11	21,996											
12	22,990											
13	26,419											
14	28,057											
15	30,865											
16	32,986											
17	37,269											
18	20,227											

5. Click on the second column to show the functions pop-up again and type GROUPAVERAGE in the filter box and select the GROUPAVERAGE function. In the Arguments field type `#Transaction_Data!purchase_amount`

The screenshot shows a Data Miner interface with the 'Formula Builder' dialog open over a table. The table has columns A through M. The 'user_id' column is selected. The 'Formula Builder' dialog shows the 'GROUPAVERAGE' function selected under 'All' and its description: 'Returns the average of its arguments.' The 'Arguments' field contains the formula `#Transaction_Data!purchase_amount`. The dialog includes 'Cancel' and 'OK' buttons.

This will calculate the average purchase amount for each of the users.

HOL Guide for Enterprise Risk Analysis

6. Click on the third column to show the function pop-up again and type *GROUPSTDEVP* and select the *GROUPSTDEVP* function

The screenshot shows the DataMeer interface with a formula builder dialog open. The formula being built is `=GROUPSTDEVP(#Transaction_Data!purchase_amount)`. The dialog includes sections for 'Apply Formulas in Your Sheets' (describing the function), 'Arguments' (specifying 'Number' as the data type and the range `#Transaction_Data!purchase_amount`), and 'OK' and 'Cancel' buttons.

user_id	Average...	
1	1,724	394.06
2	1,956	846.75
3	2,191	692.2
4	3,400	4,092.54
5	4,680	3,177.83
6	5,630	3,809.11
7	12,144	1,263.39
8	16,170	3,649.84
9	19,052	4,064.58
10	19,395	1,939.42
11	21,996	3,136.65
12	22,990	2,712.31
13	26,419	779.64
14	28,057	197.34
15	30,865	2,069.17
16	32,986	1,661.6
17	37,269	1,251.37
18	40,007	1,050.01

In the Arguments field type `#Transaction_Data!purchase_amount` to calculate the standard deviation for the *purchase_amount* field

7. Right-click on the sheet name and select *Rename* from the context menu to rename the sheet. Choose the following name for the sheet:
Stats

The screenshot shows the DataMeer interface with the 'Stats' sheet renamed. The sheet names at the bottom are now DL_OLT_Sample_Data, Transaction_Data, and Stats. The rest of the interface is identical to the previous screenshot, showing the formula builder dialog and the data table.

user_id	Average...	purchas...	D	E	F	G	H	I	J	K	L	M
1	1,724	394.06	0									
2	1,956	846.75	0									
3	2,191	692.2	0									
4	3,400	4,092.54	0									
5	4,680	3,177.83	0									
6	5,630	3,809.11	0									
7	12,144	1,263.39	0									
8	16,170	3,649.84	0									
9	19,052	4,064.58	0									
10	19,395	1,939.42	0									
11	21,996	3,136.65	0									
12	22,990	2,712.31	0									
13	26,419	779.64	0									
14	28,057	197.34	0									
15	30,865	2,069.17	0									
16	32,986	1,661.6	0									
17	37,269	1,251.37	0									
18	40,007	1,050.01	0									

HOL Guide for Enterprise Risk Analysis

8. Next we need to join the transaction data for each user with the statistical data for each user to determine how much particular transaction differentiates from the common norm. From the menu bar select *Data -> Join* to create a joined sheet

The screenshot shows the DataMeer interface with the 'Create a Joined Sheet' dialog open. On the left, there's a preview of a sheet with 'user_id' values. The main dialog has two sections: 'Select sheet & column' and 'Drag columns to define join'. In the 'Select sheet & column' section, 'Untitled', 'DL_OLT_Sample_Data', 'Transaction_Data', and 'Stats' are listed. In the 'Drag columns to define join' section, 'Inner Join' is selected, and 'Transaction_Data/user_id' and 'Stats/user_id' are mapped. At the bottom right, there are 'Cancel' and 'Create Joined Sheet' buttons.

9. Expand the *Transaction_Data* node in the pop-up navigation tree and drag the *user_id* field to the right. Do the same with the *user_id* field from the *Stats* node.

This screenshot shows the 'Create a Joined Sheet' dialog after expanding the 'Transaction_Data' and 'Stats' nodes. The 'Select sheet & column' list now includes expanded nodes for 'Transaction_Data' (containing ip_address, user_id, timestamp, purchase_amount, transaction_id, credit_card_no, and order_no) and 'Stats' (containing user_id, Average_purchases, and purchase_amount). The 'Drag columns to define join' section remains the same as in the previous screenshot, with 'Inner Join' selected and the 'user_id' fields mapped. The bottom right shows the 'Create Joined Sheet' button.

Click on the *Create Joined Sheet* button to create the joined sheet.

HOL Guide for Enterprise Risk Analysis

10. The resulting sheet will show the joined data from both *Transaction_Data* and *Stats* sheets. For convenience let's rename few of the columns. Right-click and rename the columns as below:

	Transact...	Stats.us...	Transact...	Transact...	Transact...	Transact...	Transact...	Transact...	Stats.Av...	Stats.pu...	K	L	M
1	1,724	1,724	31.162.26.68	Aug 1, 2016 0...	394.06	5e85b3ac-c96...	*****-*****-...	M34 SHIMSY	394.06	0			
2	1,956	1,956	14.37.225.140	Aug 1, 2016 0...	846.75	c311d7e-be0...	*****-*****-...	9HBMSY	846.75	0			
3	2,191	2,191	175.233.47.33	Aug 1, 2016 0...	692.2	0b7448c6-4cb...	*****-*****-...	RV775B	692.2	0			
4	3,400	3,400	109.51.50.115	Aug 1, 2016 0...	4,092.54	44d7cc18-b51...	*****-*****-...	6H9D13	4,092.54	0			
5	4,680	4,680	66.172.58.120	Aug 1, 2016 0...	3,177.83	99dfdf7de-0f60...	*****-*****-...	ZST8WJ	3,177.83	0			
6	5,630	5,630	65.25.39.65	Aug 1, 2016 0...	3,809.11	943dfc75-e25...	*****-*****-...	G2GWLO	3,809.11	0			
7	12,144	12,144	165.192.51.226	Aug 1, 2016 0...	1,263.39	9bceaae9-710...	*****-*****-...	AS1VQS	1,263.39	0			
8	16,170	16,170	89.3.163.226	Aug 1, 2016 0...	3,649.84	116ee3fc-2a08...	*****-*****-...	S4PZP2	3,649.84	0			
9	19,052	19,052	121.60.248.238	Aug 1, 2016 0...	4,064.58	ed88bf5f-eaa5...	*****-*****-...	JHFKEKG	4,064.58	0			
10	19,395	19,395	67.67.231.179	Aug 1, 2016 0...	1,939.42	5db90cf0-995...	*****-*****-...	0OTTRW	1,939.42	0			
11	21,996	21,996	137.233.134.23	Aug 1, 2016 0...	3,136.65	02872e00-198...	*****-*****-...	6IYSU9	3,136.65	0			
12	22,990	22,990	144.122.111.2...	Aug 1, 2016 0...	2,121.31	b6c52d95-2a6...	*****-*****-...	UM57DL	2,121.31	0			
13	26,419	26,419	44.35.92.219	Aug 1, 2016 0...	779.64	e1230eed-f31...	*****-*****-...	0W1KFU	779.64	0			
14	28,057	28,057	255.70.26.115	Aug 1, 2016 0...	197.34	9c25fbac-2498...	*****-*****-...	UUNVGJ	197.34	0			
15	30,865	30,865	208.24.108.18	Aug 1, 2016 0...	2,069.17	885a1169-400...	*****-*****-...	IYFKBW	2,069.17	0			
16	32,986	32,986	88.39.61.117	Aug 1, 2016 0...	1,661.6	93117e6e-a6f...	*****-*****-...	6NBTLS	1,661.6	0			
17	37,269	37,269	117.173.105.2...	Aug 1, 2016 0...	1,251.37	ca8f3e42-5ded...	*****-*****-...	ZMD848	1,251.37	0			
18	39,267	39,267	22.22.22.22	Aug 1, 2016 0...	1,060.92	00000000-0000...	*****-*****-...	1V007Q	1,060.92	0			

For convenience let's rename few of the columns. Right-click and rename the columns as below:

Transaction_Data.user_id -> *user_id*

Transaction_Data.purchase_amount -> *purchase_amount*

Stats.Average_purchase_amount -> *average_purchase_amount*

Stats.purchase_amount_Stddevp -> *purchase_amount_deviation*

Also, right-click on the sheet name and rename it to

Joined_Data_and_Stats

11. Next, we will identify the outliers by creating a copy of the joined data and filtering it. Right-click on the *Joined_Data_and_Stats* sheet and select *Duplicate*. We will select only the data we need and ignore the rest. In the pop-up select only the following fields:

user_id

purchase_amount

Transaction_data.timestamp

average_purchase_amount

purchase_amount_deviation

	ip_address	user_id	timestamp	purchas...	transacti...	credit_c...	order_no	H	I	J	K	L	M
1	106.209.197.1...	164,605	Aug 1, 2016 0...	\$3,093.82									
2	10.233.173.83	730,835	Aug 1, 2016 0...	\$3,844.52									
3	52.87.190.131	7,177,806	Aug 1, 2016 0...	\$3,605.26									
4	105.158.106.82	2,132,670	Aug 1, 2016 0...	\$2,523.34									
5	12.15.120.180	8,596,261	Aug 1, 2016 0...	\$205.74									
6	11.216.44.6	2,946,179	Aug 1, 2016 0...	\$4,108.15									
7	185.183.151.1...	6,839,961	Aug 1, 2016 0...	\$1,484.53									
8	113.161.160.1...	2,245,110	Aug 1, 2016 0...	\$969.83									
9	28.245.64.112	1,223,107	Aug 1, 2016 0...	\$276.15									
10	35.143.156.34	2,220,765	Aug 1, 2016 0...	\$1,887.21									
11	246.118.83.220	3,710,930	Aug 1, 2016 0...	\$3,880.10									
12	234.74.136.180	4,517,350	Aug 1, 2016 0...	\$775.60									
13	246.193.132.62	5,172,602	Aug 1, 2016 0...	\$2,000.91									
14	64.245.31.254	7,438,752	Aug 1, 2016 0...	\$2,687.91									
15	2.42.116.20	3,894,627	Aug 1, 2016 0...	\$3,313.53									
16	215.55.147.149	9,393,439	Aug 1, 2016 0...	\$1,433.69	Gela9a2d-25f...	*****-*****-...	CUPOK2						
17	173.238.152.96	1,094,163	Aug 1, 2016 0...	\$2,038.83	892c9168-362...	*****-*****-...	R88IS9						
18													

HOL Guide for Enterprise Risk Analysis

Click on *Create Sheet Copy* button

12. Right-click on the *Transaction_Data.timestamp* field and rename it to *timestamp* only.
13. For the purpose of our analysis we will consider transactions with deviation two times more than standard deviation as outliers. In the new sheet select *Data -> Filter* from the menu.

The screenshot shows a data grid with the following columns: user_id, purchase_amount, average_purchase_amount, and timestamp. The timestamp column has a context menu open, with the 'Create Sheet Copy' option highlighted. The URL in the browser is https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/workbook/21.

14. Select the *Advanced* tab in the pop-up and type the following formula:

$ABS(\#purchase_amount - \#average_purchase_amount) > 2 * \#purchase_amount_deviation$

The screenshot shows the 'Apply Filter' dialog box. The 'Advanced' tab is selected, and the formula $ABS(\#purchase_amount - \#average_purchase_amount) > 2 * \#purchase_amount_deviation$ is entered in the 'Full expression' field. The URL in the browser is https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/workbook/21.

15. The resulting sheet may be empty because the representative sample that Datameer has selected may not have transactions that are considered outliers. Right-click on the sheet name and rename it to *Outliers*

The screenshot shows the Datameer interface with an 'Untitled Workbook'. The 'Home' tab is selected. A sheet titled 'Outliers' is visible in the bottom navigation bar. The main workspace is a grid with columns labeled 'user_id', 'purchas...', 'average...', and 'purchas...'. Rows 1 through 18 are present, but the data cells are empty.

16. Finally, we would like to create a summary of the data that we would like to visualize. Let's start with summary of the *Transaction_Data*. Create new sheet and in the formula pop-up select the *GROUPBY* function

The screenshot shows the 'Formula Builder' dialog box open over the Datameer interface. The dialog title is 'Apply Formulas in Your Sheets'. It displays a list of functions under the 'GROUPBY' category, including 'GROUPBY', 'GROUPBYBIN', 'GROUPBYCUSTOMBIN', and 'GROUPBYGAP'. Below this, there is a description of the 'GROUPBY' function and a link to its documentation. An 'Arguments' section contains a text input field with the formula 'YEAR(#Transaction_Data#timestamp)'. The background shows the 'Transaction_Data' sheet in the workspace.

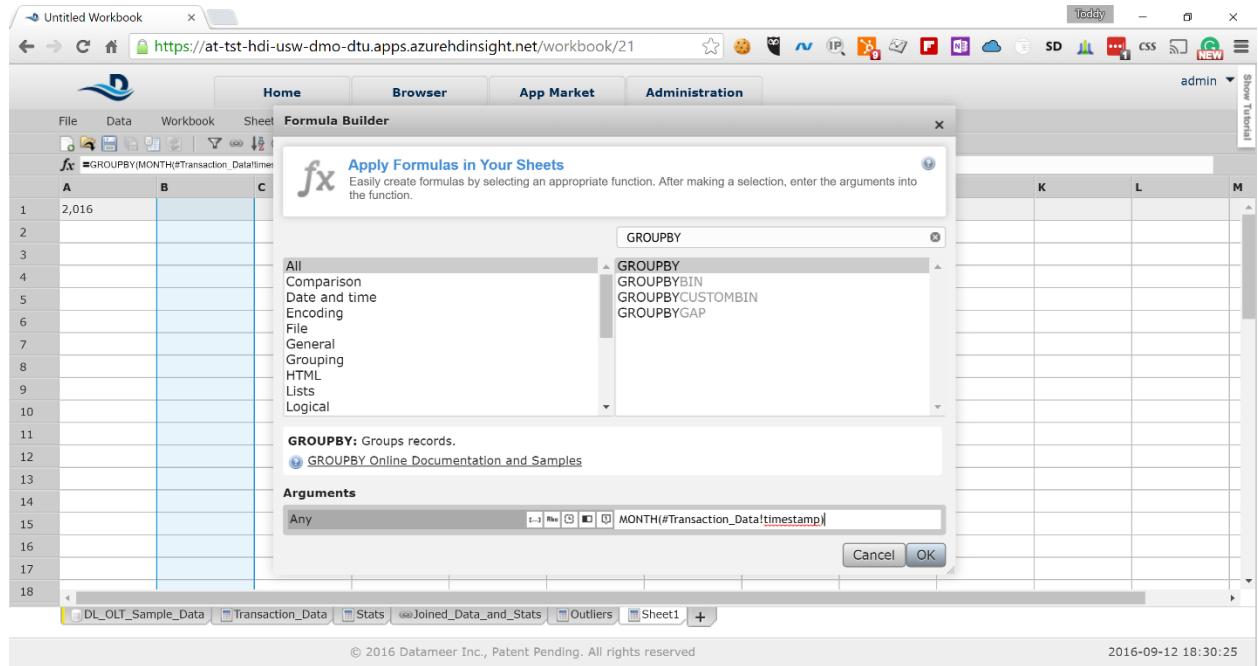
In the *Arguments* field type the following formula:

HOL Guide for Enterprise Risk Analysis

YEAR(#Transaction_Data!timestamp)

17. Click on the next column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

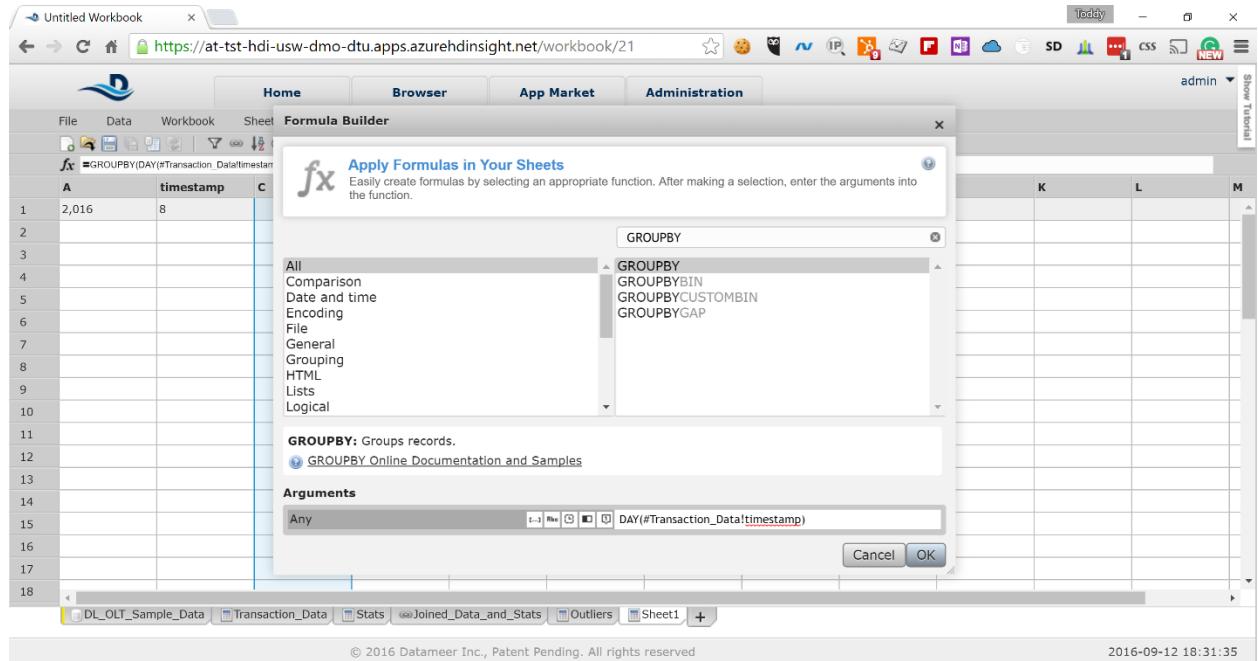
MONTH(#Transaction_Data!timestamp)



18. Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

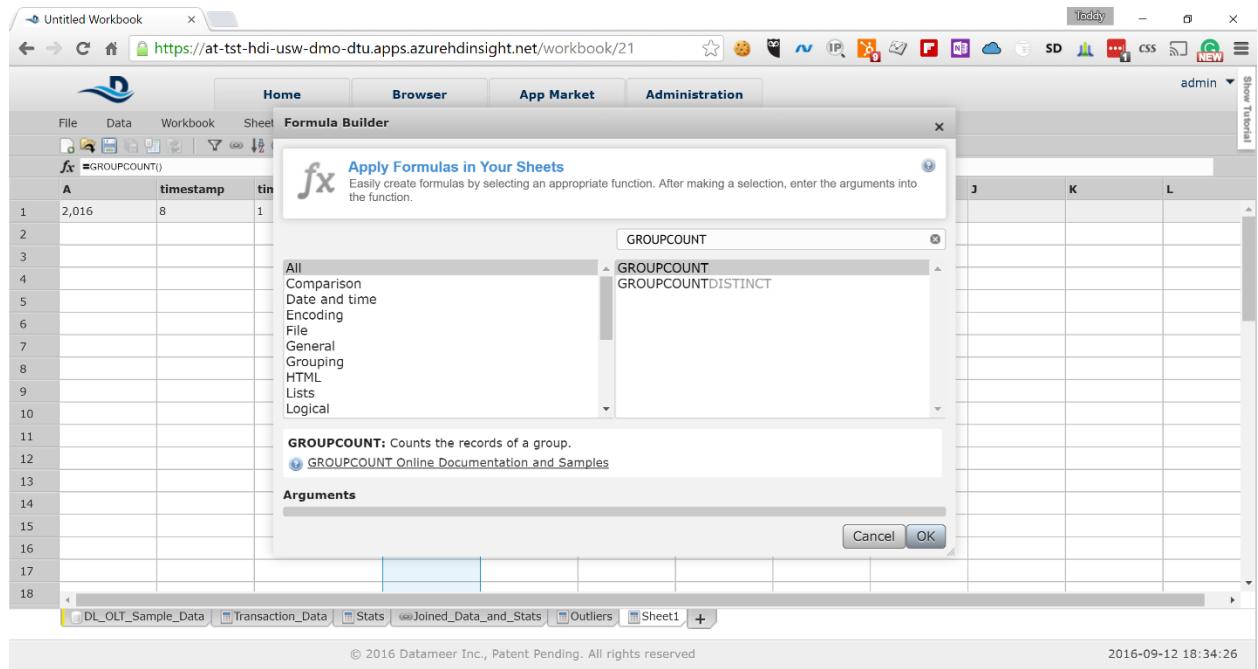
DAY(#Transaction_Data!timestamp)

HOL Guide for Enterprise Risk Analysis



The screenshot shows the DataMeer interface with the 'Formula Builder' dialog open. The formula being built is `=GROUPBY(DAY(#Transaction_Data!timestamp))`. The 'GROUPBY' function is selected from the list, and its description is visible: 'Groups records'. The 'Arguments' field contains 'Any' and 'DAY(#Transaction_Data!timestamp)'. The status bar at the bottom right shows the date and time: 2016-09-12 18:31:35.

19. Click on the fourth column and in the formula pop-up select the **GROUPCOUNT** function and click the **OK** button



The screenshot shows the DataMeer interface with the 'Formula Builder' dialog open. The formula being built is `=GROUPCOUNT()`. The 'GROUPCOUNT' function is selected from the list, and its description is visible: 'Counts the records of a group'. The 'Arguments' field is empty. The status bar at the bottom right shows the date and time: 2016-09-12 18:34:26.

20. We have created summary sheet for our transaction data. Rename the field names as follows:

year
month
day
transactions_count

HOL Guide for Enterprise Risk Analysis

The screenshot shows a DataMeer workspace titled "Untitled Workbook". The main area displays a table with columns: year, month, day, transacti..., E, F, G, H, I, J, K, L. Row 1 contains values: 2,016, 8, 1, 5,000, and empty cells for the rest of the columns. The table has 18 rows. Below the table, there are tabs for DL_OLT_Sample_Data, Transaction_Data, Stats, Joined_Data_and_Stats, Outliers, Transactions_Summary, and Sheet1. The status bar at the bottom indicates "© 2016 DataMeer Inc., Patent Pending. All rights reserved" and the date "2016-09-12 18:37:23".

Also, rename the sheet to *Transactions_Summary*

- Let's create similar summary for the outliers. Create new sheet and in the formula pop-up select the *GROUPBY* function. In the *Arguments* field type the following formula:

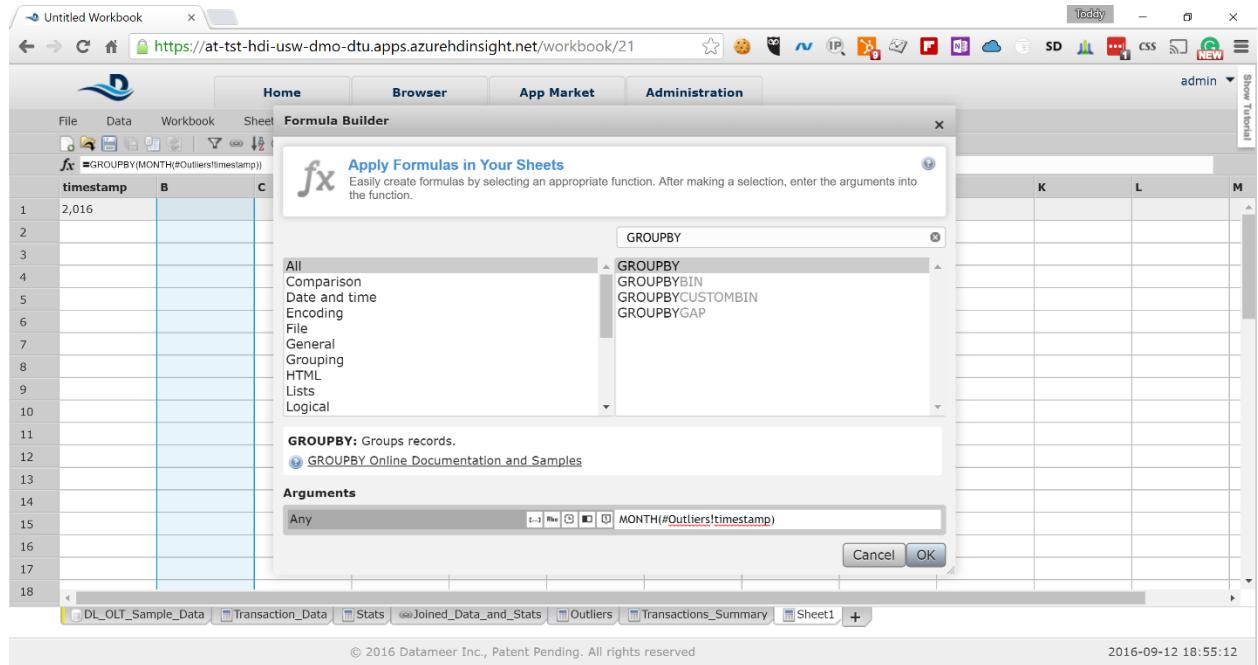
YEAR(#Outliers!timestamp)

The screenshot shows the DataMeer Formula Builder dialog box. The title bar says "Formula Builder". The main area has a "fx" icon and the text "Apply Formulas in Your Sheets". Below it is a "GROUPBY" section with a dropdown menu showing "All", "Comparison", "Date and time", "Encoding", "File", "General", "Grouping", "HTML", "Lists", and "Logical". A detailed description of the GROUPBY function is provided: "GROUPBY: Groups records." and a link to "GROUPBY Online Documentation and Samples". At the bottom, there is an "Arguments" section with a text input field containing "Any" and a formula input field containing "*YEAR(#Outliers!timestamp)*". There are "Cancel" and "OK" buttons at the bottom right.

- Click on the next column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

MONTH(#Outliers!timestamp)

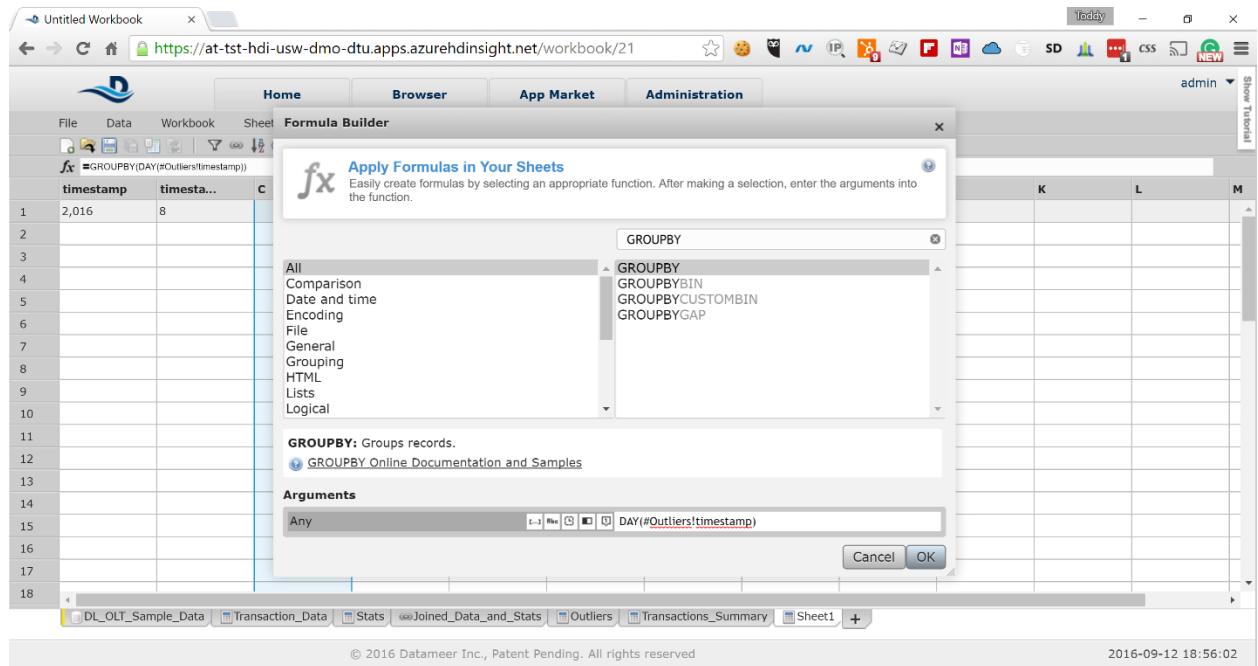
HOL Guide for Enterprise Risk Analysis



The screenshot shows the DataMeer Formula Builder dialog box. In the center, there is a list of functions under the heading 'GROUPBY'. The 'GROUPBY' function is selected. Below it, the 'Arguments' field contains the formula 'MONTH(#Outliers!timestamp)'. The background shows a spreadsheet with a single row of data: timestamp 2,016 in column A, and an empty cell in column B.

23. Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

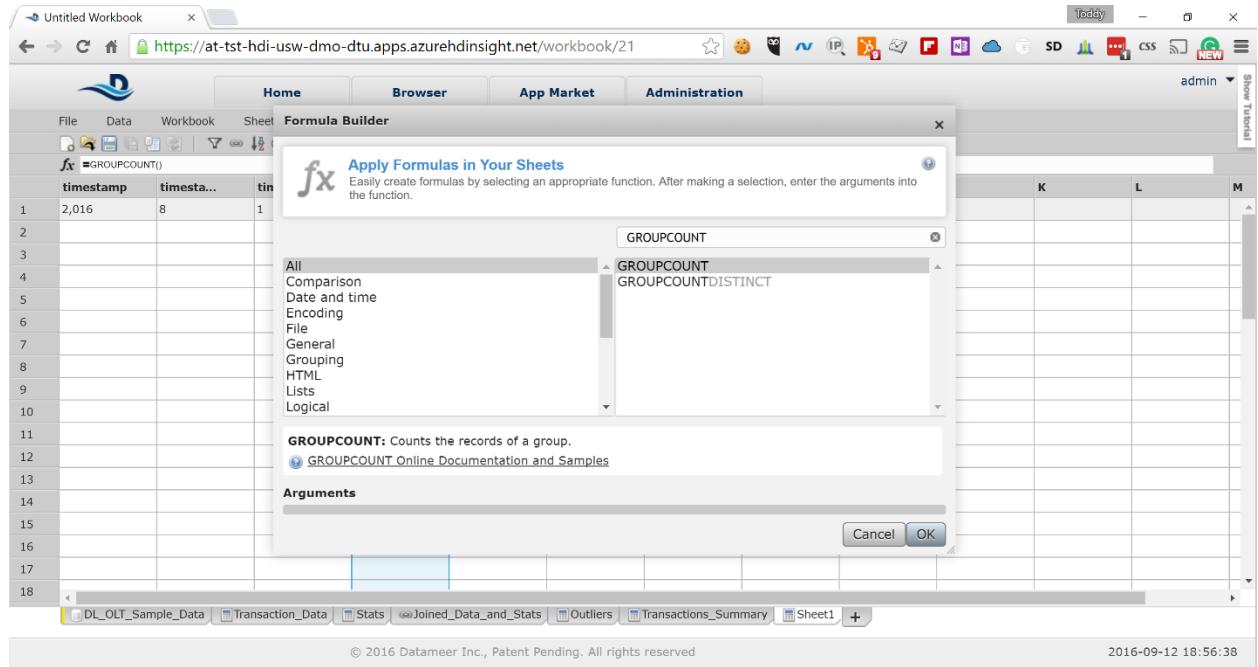
DAY(#Outliers!timestamp)



The screenshot shows the DataMeer Formula Builder dialog box. In the center, there is a list of functions under the heading 'GROUPBY'. The 'DAY' function is selected. Below it, the 'Arguments' field contains the formula 'DAY(#Outliers!timestamp)'. The background shows a spreadsheet with two columns: timestamp 2,016 in column A and value 8 in column B.

24. Click on the fourth column and in the formula pop-up select the *GROUPCOUNT* function and click the *OK* button

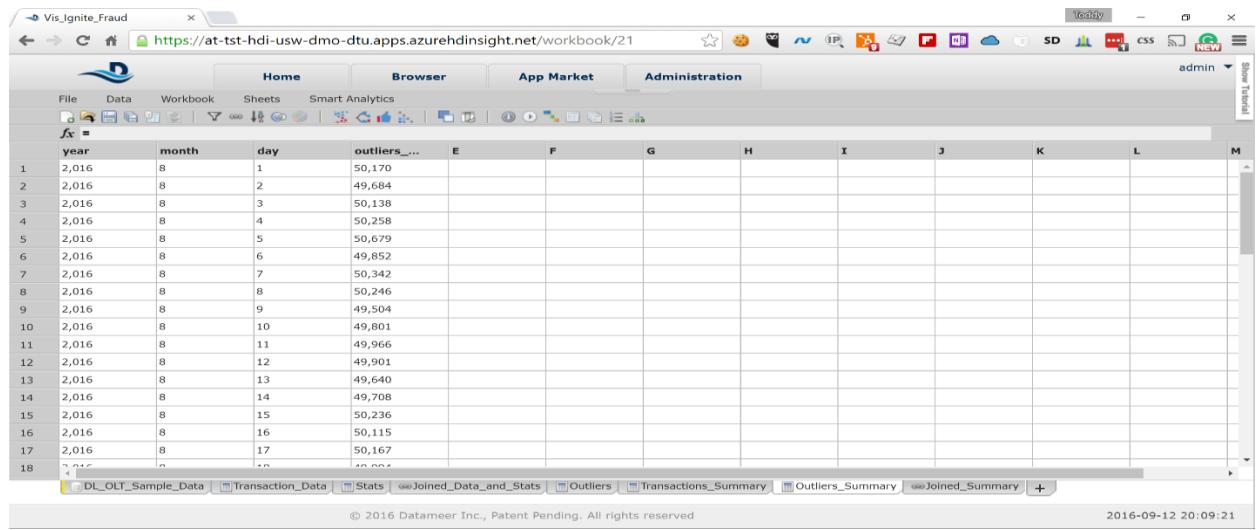
HOL Guide for Enterprise Risk Analysis



The screenshot shows the DataMeer interface with the 'Formula Builder' dialog open. The dialog title is 'Apply Formulas in Your Sheets'. In the center, the 'GROUPCOUNT' function is selected from a list. Below it, a description of the function states: 'GROUPCOUNT: Counts the records of a group.' There is also a link to 'GROUPCOUNT Online Documentation and Samples'. At the bottom of the dialog are 'Cancel' and 'OK' buttons. The background of the interface shows a spreadsheet with columns labeled 'timestamp' and 'tin'. The status bar at the bottom right indicates the date and time: '2016-09-12 18:56:38'.

25. We have created summary sheet for our transaction data. Rename the field names as follows:

year
month
day
outliers_count



The screenshot shows a spreadsheet titled 'Outliers_Summary'. It has four columns: 'year', 'month', 'day', and 'outliers_count'. The data consists of 31 rows, each representing a day in August 2016. The 'outliers_count' column contains values such as 50,170, 49,684, 50,138, etc. The status bar at the bottom right indicates the date and time: '2016-09-12 20:09:21'.

Also, rename the sheet to *Outliers_Summary*

26. We need to join the two summary sheets to have the results available in a single sheet for visualization.
Select *Data -> Join* and join the *Transactions_Summary* and *Outliers_Summary* sheets by year, month and date as on the picture below by clicking on the *Create Joined Sheet* button

HOL Guide for Enterprise Risk Analysis

The screenshot shows the DataMeer interface with the 'Create a Joined Sheet' dialog open. The dialog is titled 'Create a Joined Sheet' and describes creating a new sheet from two or more sheets based on a key column. It offers two join types: 'Simple' (selected) and 'Range Join'. Under 'Simple', an 'Inner Join' is selected, joining 'Transactions_Summary/year' with 'Outliers_Summary/year'. Other joins for 'month' and 'day' are also listed. Below the joins, there's a button 'Choose included columns ...'. At the bottom right are 'Cancel' and 'Create Joined Sheet' buttons. The main workspace shows a table with columns 'year' and 'Transactions_Summary/year'. The footer indicates '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the date '2016-09-12 18:58:55'.

Rename the joined sheet to *Joined_Summary*

27. Let's copy the joined sheet and remove the duplicate data from it. Right-click on the *Joined_Summary* sheet and select *Duplicate*. Select the following fields in the pop-up:

Transactions_Summary.year
Transactions_Summary.month
Transactions_Summary.day
Transactions_Summary.transactions_count
Outliers_Summary.outliers_count

The screenshot shows the DataMeer interface with the 'Duplicate Worksheet' dialog open. The dialog title is 'Duplicate Worksheet' and it says 'Select the columns you want to copy.' A 'Select: all / none' checkbox is checked. Below it, several checkboxes are checked: 'Transactions_Summary.year', 'Transactions_Summary.month', 'Transactions_Summary.day', and 'Outliers_Summary.outliers_count'. Other checkboxes like 'Transactions_Summary.transactions_count' and 'Outliers_Summary.outliers_count' are unchecked. At the bottom right are 'Cancel' and 'Create Sheet Copy' buttons. The main workspace shows a table with columns 'Transact...' and 'Transactions_Summary.transactions_count'. The footer indicates '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the date '2016-09-12 20:11:34'.

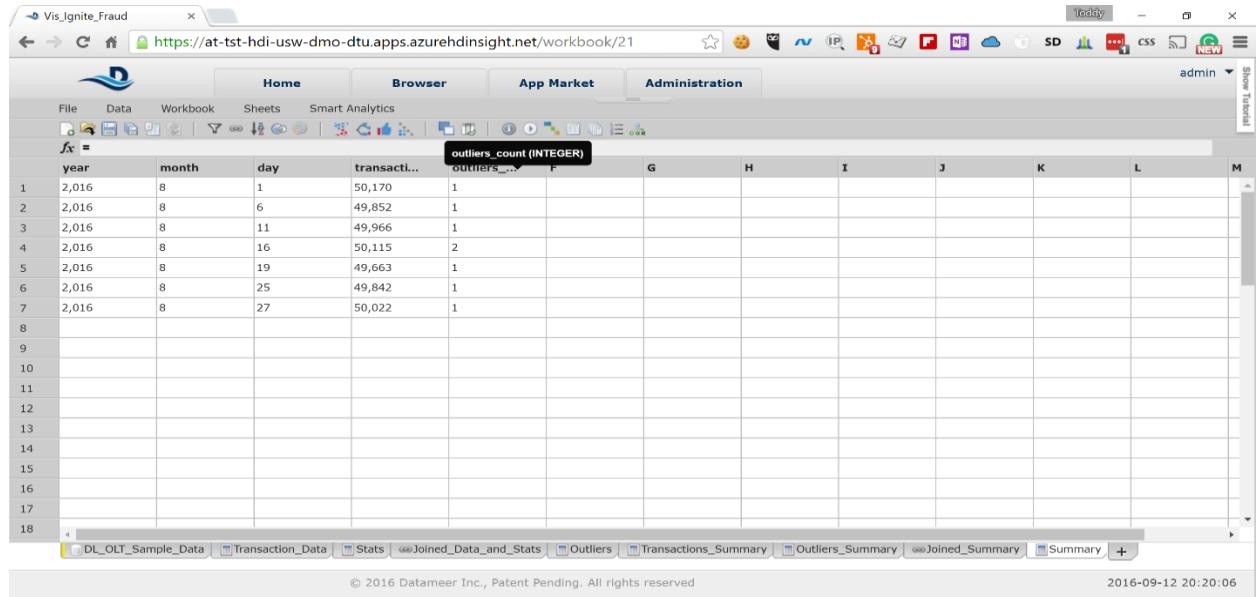
28. Rename the sheet to *Summary* and the columns as follows:

Transactions_Summary.year -> *year*
Transactions_Summary.month -> *month*
Transactions_Summary.day -> *day*

HOL Guide for Enterprise Risk Analysis

Transactions_Summary.transactions_count -> transactions_count

Outliers_Summary.outliers_count -> outliers_count

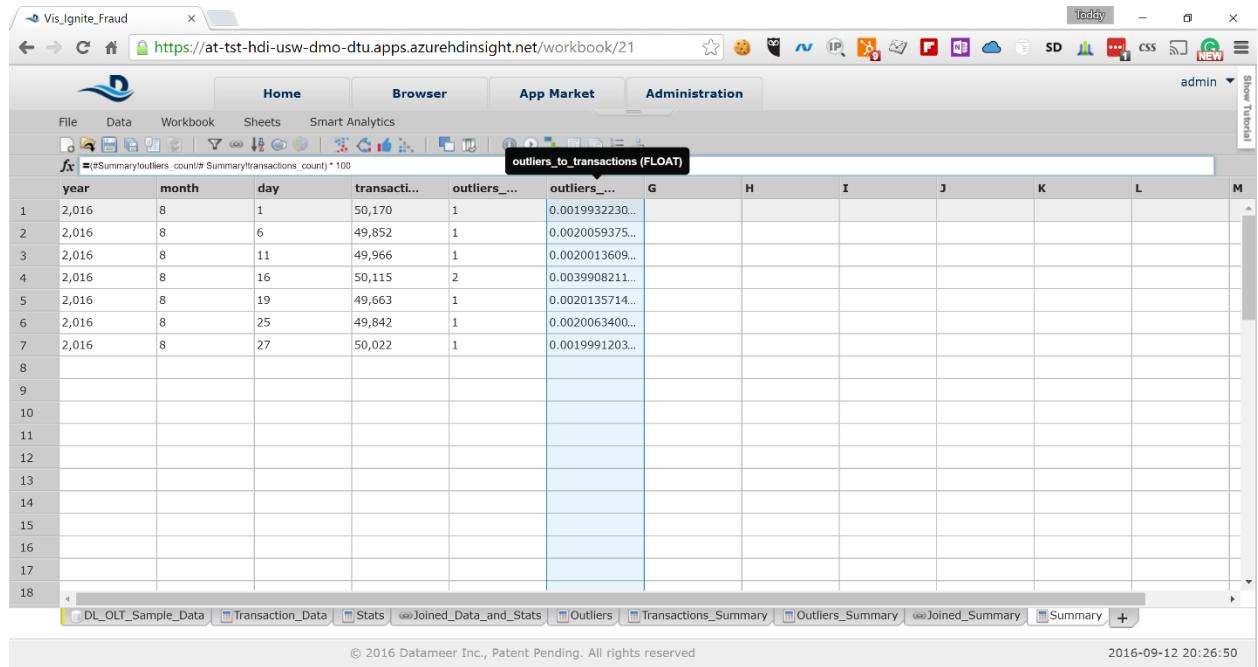


The screenshot shows a Data Miner workspace titled "Vis_Ignite_Fraud". A spreadsheet is open with the following columns: year, month, day, transacti..., outliers_..., and outliers_count (INTEGER). The outliers_count column is highlighted with a black background and white text. The data in the first few rows is as follows:

	year	month	day	transacti...	outliers_...	outliers_count (INTEGER)
1	2,016	8	1	50,170	1	1
2	2,016	8	6	49,852	1	
3	2,016	8	11	49,966	1	
4	2,016	8	16	50,115	2	
5	2,016	8	19	49,663	1	
6	2,016	8	25	49,842	1	
7	2,016	8	27	50,022	1	

29. Click on the sixth column and cancel the formula pop-up. In the f_x field on top of the sheet type the following:

$(\#Summary!outliers_count / \#Summary!transactions_count) * 100$



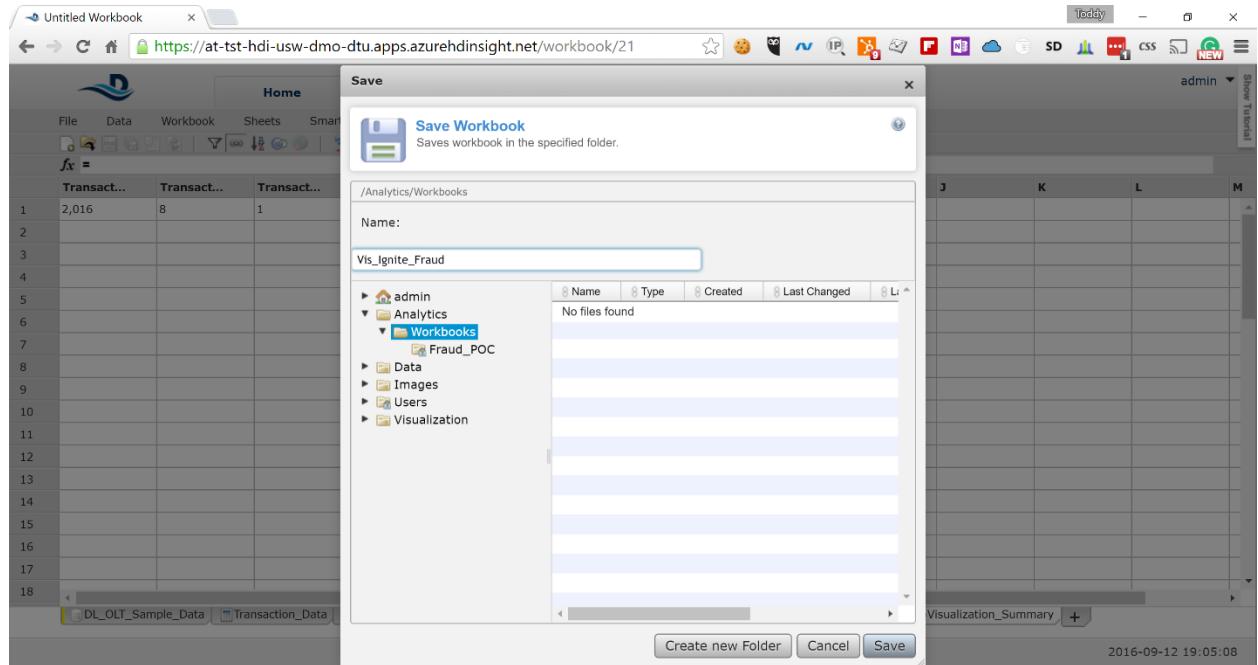
The screenshot shows the same Data Miner workspace and spreadsheet as the previous one. The formula $(\#Summary!outliers_count / \#Summary!transactions_count) * 100$ has been entered into the f_x field at the top of the sheet. The outliers_to_transactions column now contains floating-point values representing the percentage of outliers per transaction. The data in the first few rows is as follows:

	year	month	day	transacti...	outliers_...	outliers_to_transactions (FLOAT)
1	2,016	8	1	50,170	1	0.0019932230...
2	2,016	8	6	49,852	1	0.0020059375...
3	2,016	8	11	49,966	1	0.0020013609...
4	2,016	8	16	50,115	2	0.0039908211...
5	2,016	8	19	49,663	1	0.0020135714...
6	2,016	8	25	49,842	1	0.0020063400...
7	2,016	8	27	50,022	1	0.001991203...

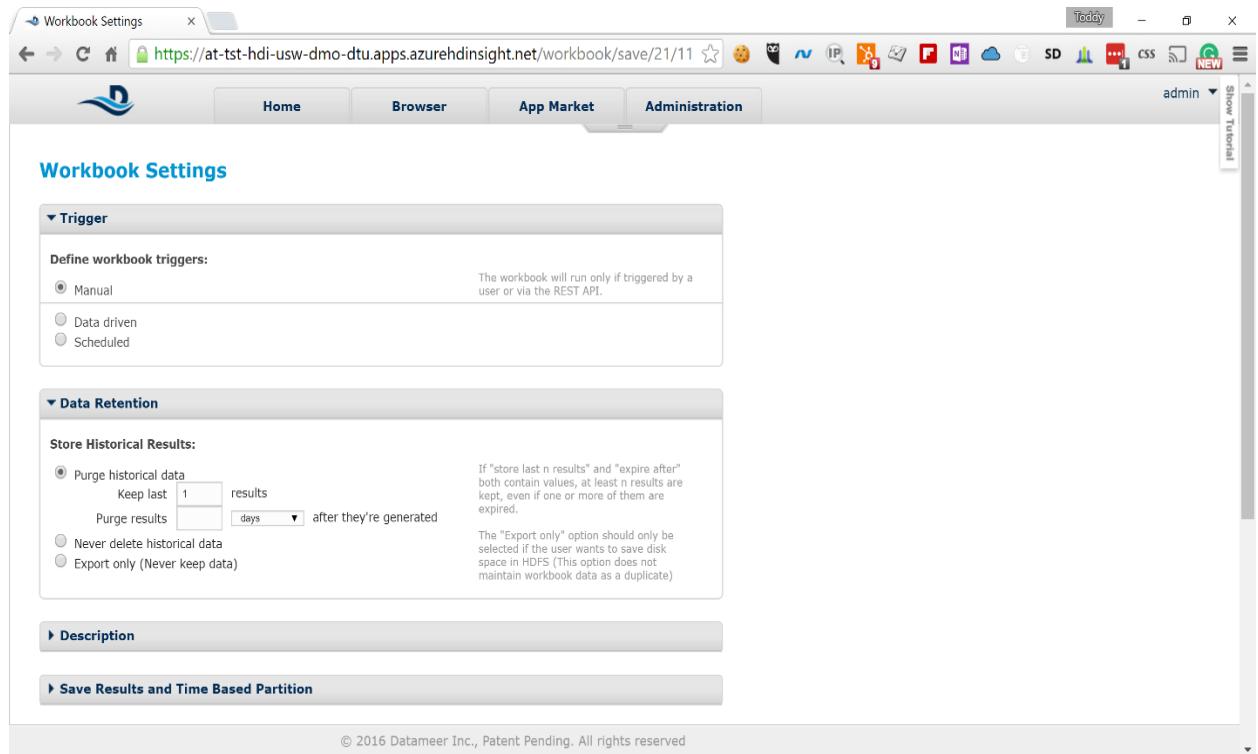
Also, rename the field to *outliers_to_transactions*

30. Select *File -> Save* from the menu and type the *Vis_Ignite_Fraud* in the *Name* field

HOL Guide for Enterprise Risk Analysis



31. On the next screen keep the default values for all the fields. Scroll down and click on the **Save** button again



32. In the list of workbooks select the newly created workbook and click on the run button from the toolbar. This will trigger the calculation on the full data set

HOL Guide for Enterprise Risk Analysis

The screenshot shows the DataMeer File Browser interface. The left sidebar contains navigation links for Home, Applications, Analytics, Workbooks, Data, Images, Users, and Visualization. The main area displays a table of workbooks with columns for Name, Type, Status, Last Processed, Records, Size, License total size, and Owner. One entry, 'Vis_Ignite_Fraud.wbk', is selected. The 'Information' panel on the right provides detailed information about this file, including its ID (21), File ID (53), Type (Workbook), Size (0 Bytes), and various timestamps for creation and last change.

You will see updates in the *Status* column, showing you how the Hadoop job is progressing.

This screenshot shows the same DataMeer File Browser interface as the previous one, but the 'Vis_Ignite_Fraud.wbk' entry in the list now includes a progress bar in the 'Status' column, indicating that the job is currently at 28% completion (Optimized MapReduce). The rest of the interface and the 'Information' panel remain identical to the first screenshot.

8 Logging in to the TrendMicro DSM

8.1 Server name

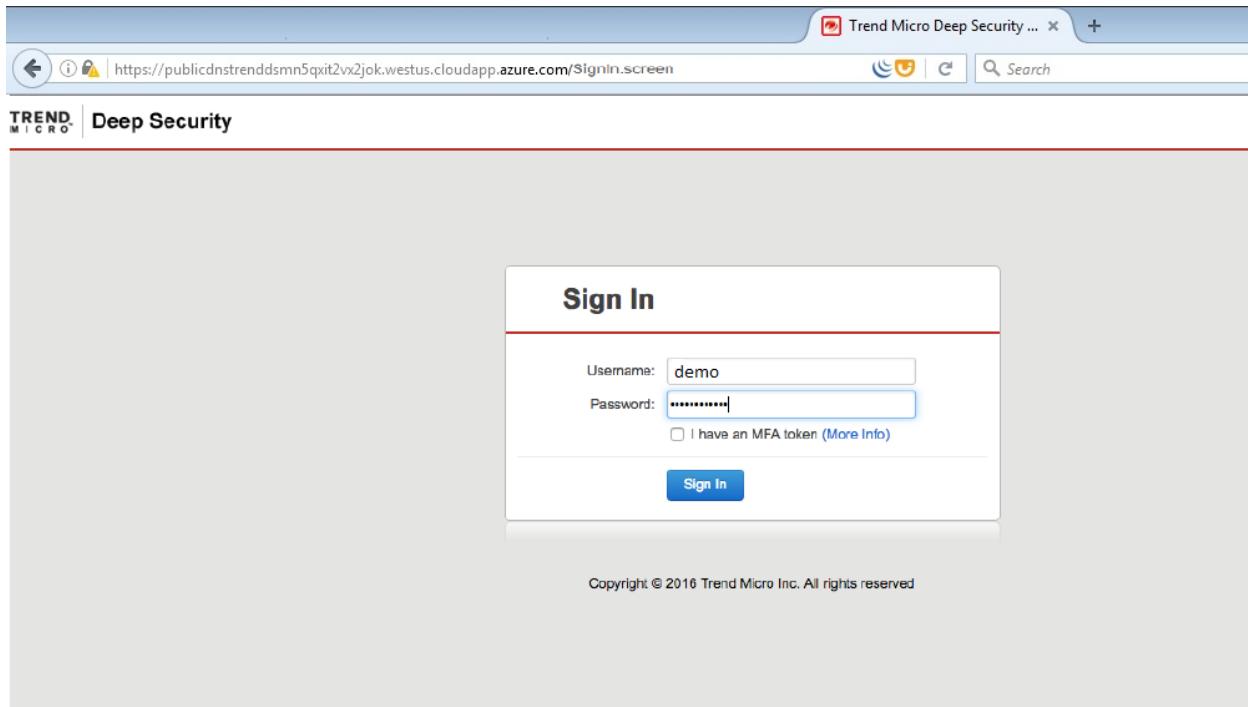
From the output section of the deployment you can get the URL for TrendMicroDSM, Splunk and Chef Server (Microsoft.Template)



8.2 Server login

To login to TrendMicro DSM

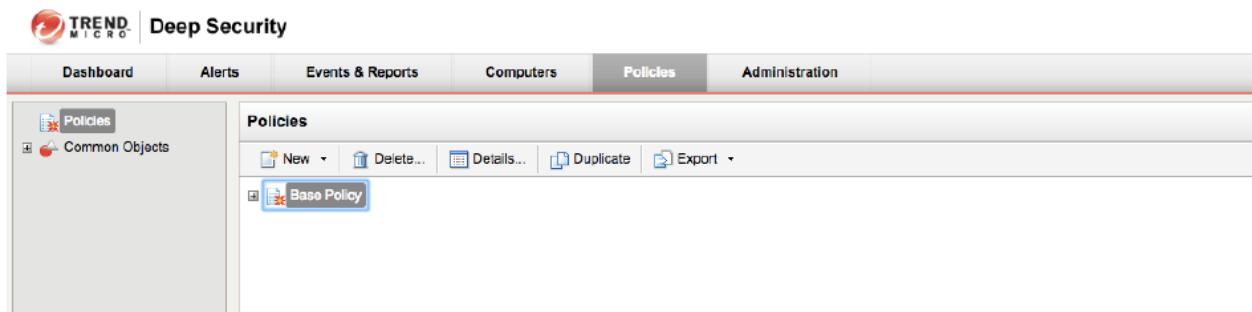
- Paste the TrendMicro DSM URL in the browser
- Enter the **Username** and **Password** provided in the parameter section during the deployment



9 Perform policy configuration on the TrendMicro DSM

1. Changing the base policy

Go to policies->Base Policy



2. Enable Anti-Malware

Go to Anti-malware->Anti-Malware State->On

Click "Save"

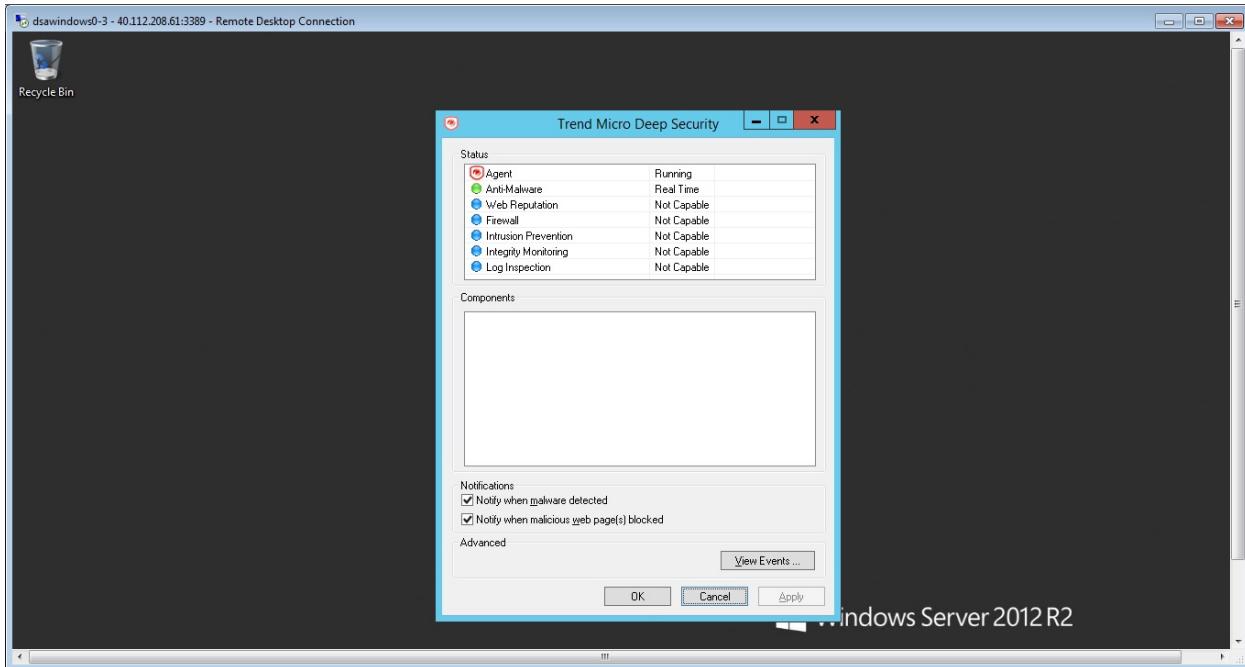
HOL Guide for Enterprise Risk Analysis

The screenshot shows the Trend Micro Policy Editor interface. On the left is a navigation sidebar with icons for Overview, Anti-Malware, Web Reputation, Firewall, Intrusion Prevention, Integrity Monitoring, Log Inspection, Interface Types, Settings, and Overrides. The main area has tabs for General, Smart Protection, Advanced, Quarantined Files, and Events. Under the General tab, there's a section for Anti-Malware where the Anti-Malware State is set to 'On'. Below that are sections for Real-Time Scan, Manual Scan, and Scheduled Scan, each with a 'Default' option and a dropdown for Malware Scan Configuration. At the bottom right are 'Save' and 'Close' buttons.

3. Applying policies to computer

The screenshot shows the Trend Micro Deep Security application. The top navigation bar includes Dashboard, Alerts, Events & Reports, Computers, Policies, and Administration. The main area displays a list of computers under the 'Computers' tab. A context menu is open over three selected hosts (40.112.208.61, 40.118.247.189, and publicdnstrenddsmn5q...). The menu options include Activate/Reactivate, Deactivate, Send Policy (which is highlighted), Cancel "Send Policy", Clear Warnings/Errors, Upgrade Agent Software..., Scan for Recommendations, Scan for Open Ports, Move To Group..., Assign Policy..., Assign Asset Value..., and Assign Relay Group... . The status bar at the bottom indicates 0 alerts.

4. Verifying policy in the computer



10 Exercises

10.1 Datameer – Visualize the Data

Datameer has powerful Infographics to Visualise the data. In this exercise, the data analysed in the above configuration will be displayed graphically.

10.2 TrendMicro – Malware test

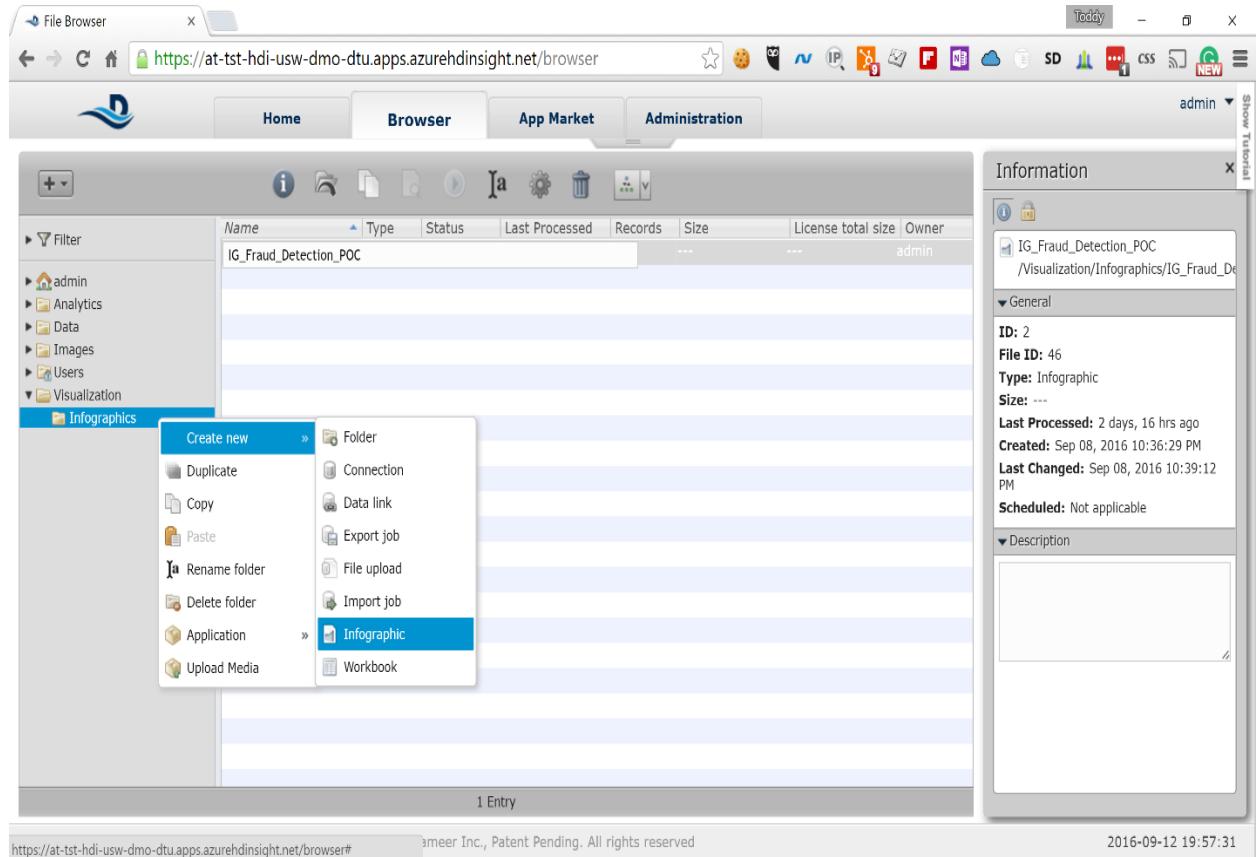
TrendMicro has security intelligence built-in to protect the systems against the malwares. In this exercise, showcases the TrendMicro DSM malware detection capability

11 Visualize the Data

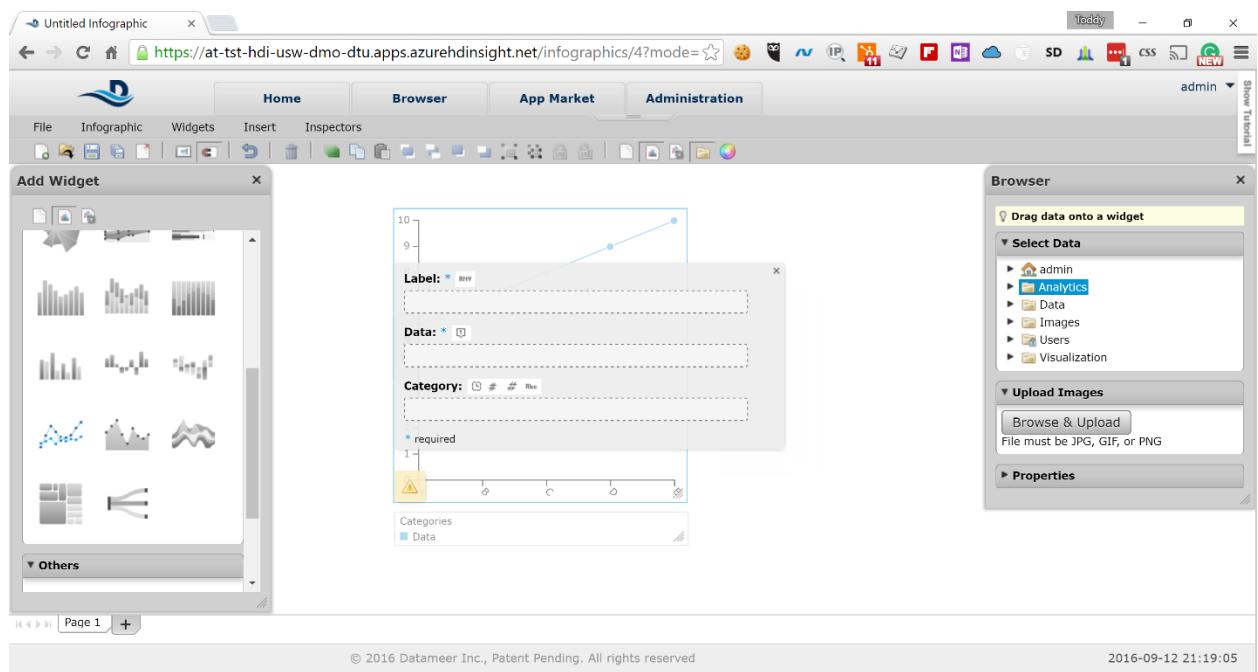
The last exercise in this HOL is to visualize the data and identify certain days when the irregular transactions have spiked. To do that we will use the following steps:

1. In Datameer's Browser view expand the Visualization node and right click on Infographics -> Create New -> Infographic

HOL Guide for Enterprise Risk Analysis

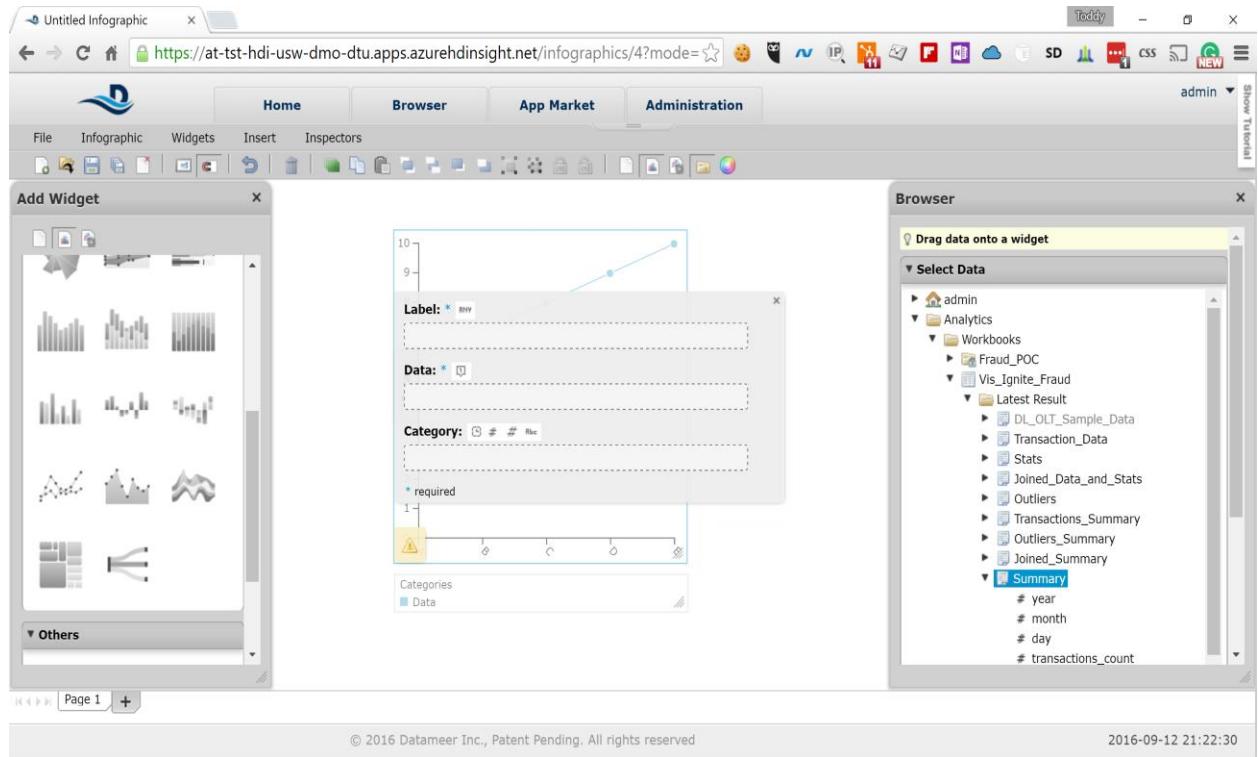


2. Drag the *Line and Area Chart* from the *Add Widget* pane on the left to the work pane in the middle

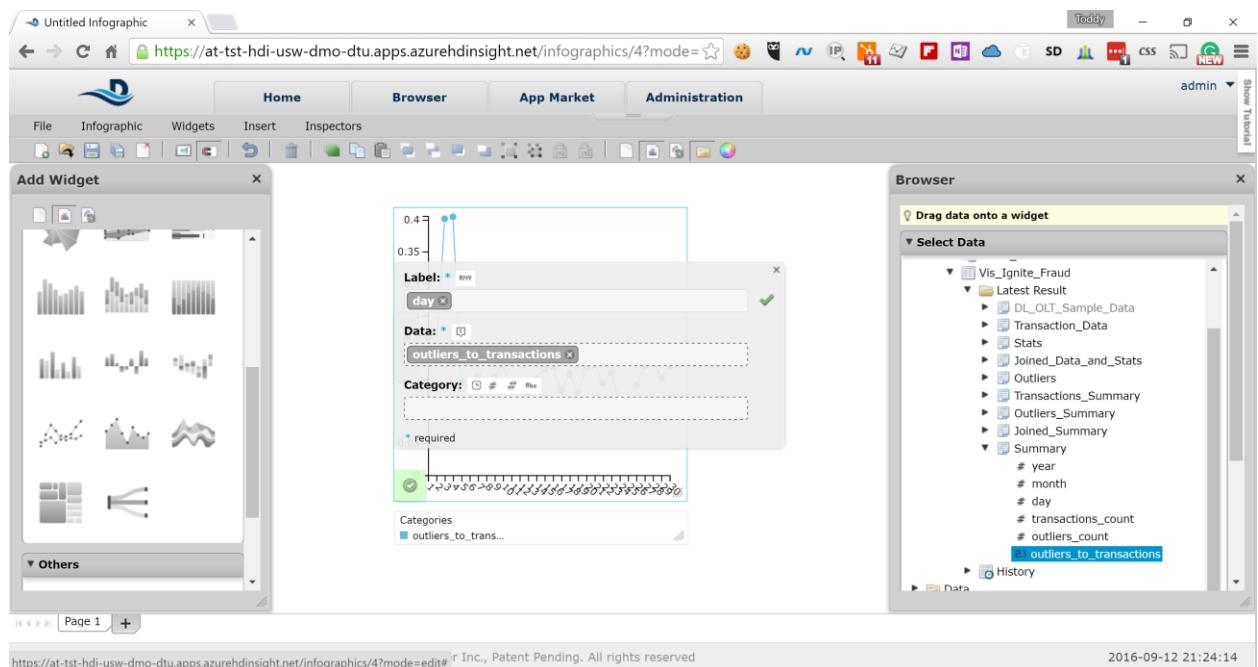


HOL Guide for Enterprise Risk Analysis

3. In the Browser pane expand *Analytics* node and then *Workbooks* -> *Vis_Ignite_Fraud* -> *Latest Results* -> *Summary*

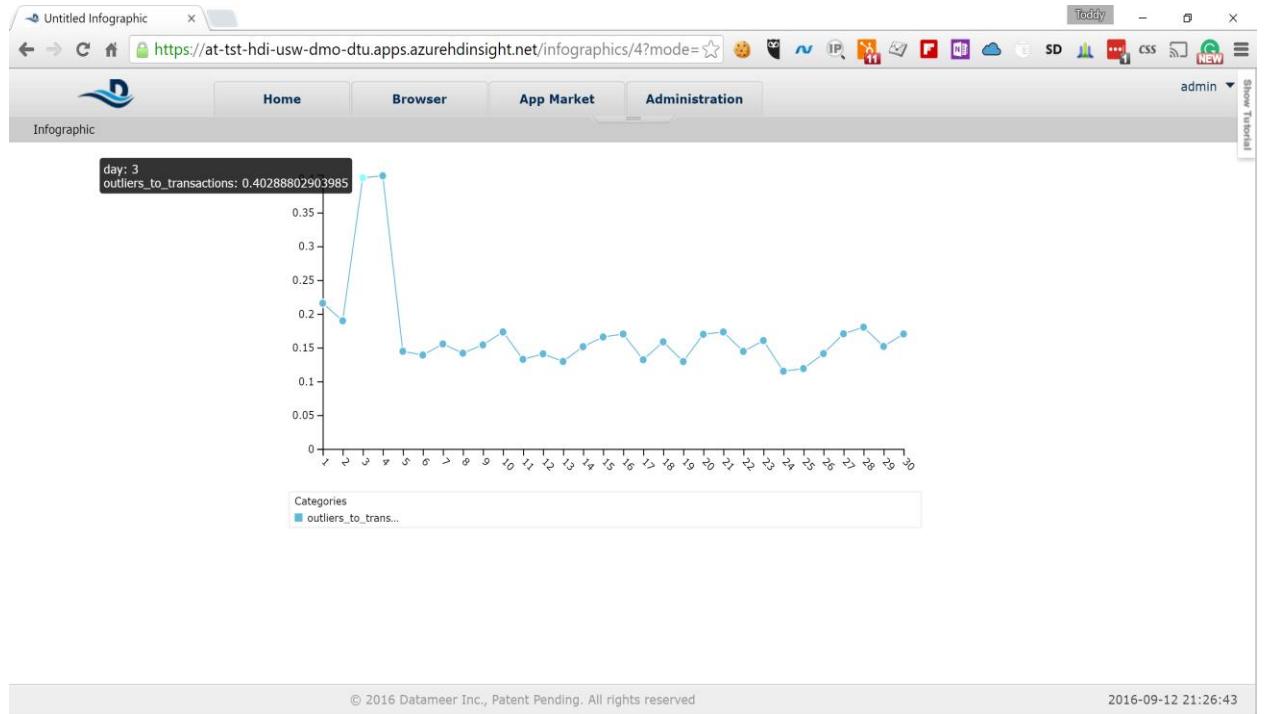


4. Drag the *day* field to the *Label* input field and the *outliers_to_transactions* field to the *Data* input field in the Work pane



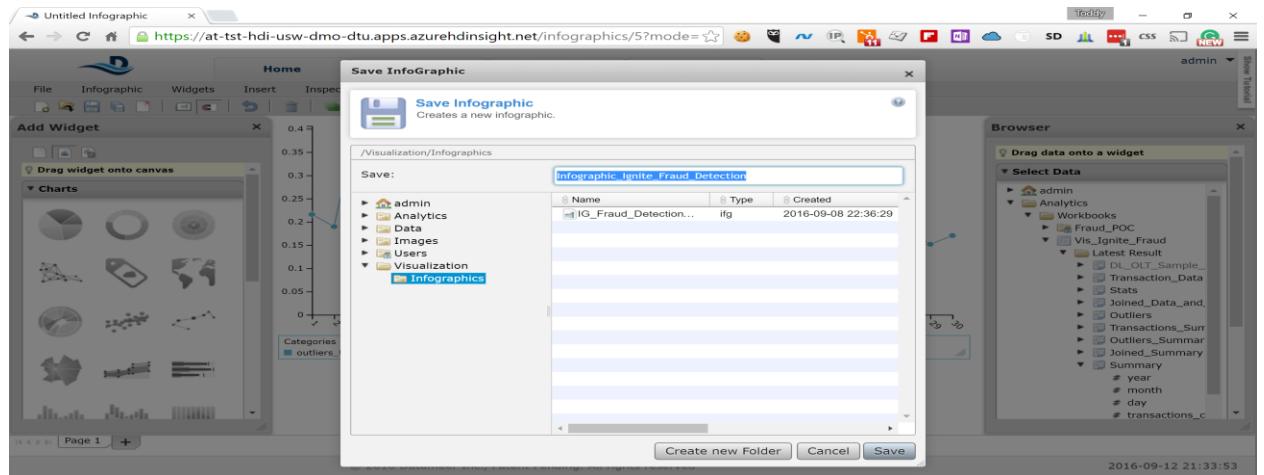
HOL Guide for Enterprise Risk Analysis

5. Select Infographic -> View from the menu to present the infographic. You can easily see that on the 3rd and 4th day of the month the outliers significantly spiked, which is a sign of something unusual going on those two days



6. Select Infographic -> Edit from the Manu and then File -> Save. Type the following in the Name field:

Infographic_Ignite_Fraud_Detection
and click on the Save button



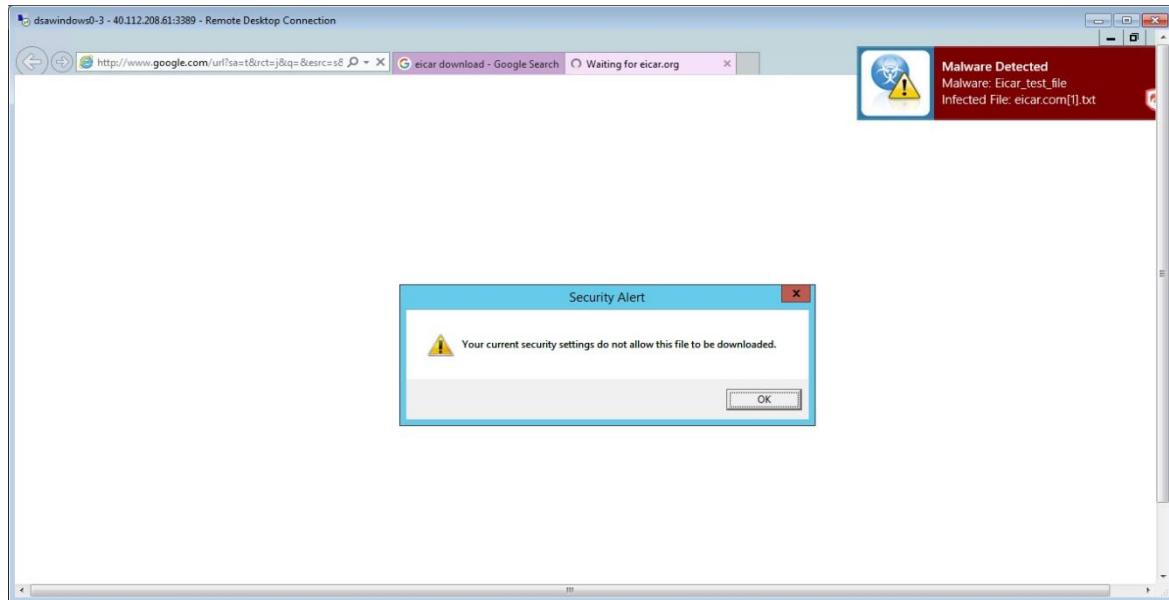
12 Malware Test

12.1 Generating Malware alert in the computer

The Malware test can be performed by going to the url

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&sqi=2&ved=0ahUKEwj6n6H1-IXPAhUSzGMKHZMGC5AQFggmMAE&url=http%3A%2F%2Fwww.eicar.org%2Fdownload%2Feicar.com.txt&usg=AFQjCNE8DvVI7BE5Nd2hq1zNDTP6hNjclA&bvm=bv.132479545,d.cGc>

this is eicar malware test



12.2 Dashboard – Malware Alert

HOL Guide for Enterprise Risk Analysis

The screenshot shows the Trend Micro Deep Security dashboard. At the top, there are several status cards: 'Alert Status' (0 Critical, 1 Warning), 'Computer Status' (0 Critical, 0 Warning, 3 Managed, 0 Unmanaged), 'My Account Status' (Username: demo, Role: Full Access, Last Sign-In: September 5, 2016 22:32, Previous Sign-In: N/A, Total Sign-Ins: 1), and 'My Sign-in History' (Success). Below these are 'Anti-Malware Event History' (2 events) and 'Anti-Malware Status (Computers)' (Top 5 Infected Computers: 40.112.208.61, 0% Uncleanable). Other tabs visible include 'Events & Reports', 'Computers', 'Policies', and 'Administration'. A bottom bar shows 'Alerts (1) (0)'.

12.3 Malware Alert verification

The screenshot shows the 'Events & Reports' section of the Trend Micro Deep Security interface. On the left, there's a sidebar with 'Events' and 'Generate Reports'. The main area displays 'Anti-Malware Events' for the period from September 5, 2016, at 00:00 to September 6, 2016, at 00:00. A single event is listed: 'September 5, 2016 22:51:23 40.112.208.61 C:\Users\demo\AppData\Local\Microsoft\Windows\NetCache\E8JDT4RUG\leicar.com[1].txt Eicar_test_file Deleted'. The interface includes standard filtering and export options.

13 References, Attachments & Definitions

13.1 References

No.	Document Title	Link/ Attachment	Comments
1			