

---

# HOL Guide for Enterprise Risk Analysis

---

USING DATAMEER, HDINSIGHT, TRENDMICRO DEEP SECURITY  
AND CHEF

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](mailto: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 .....	13
7	Perform Analysis to Identify Outliers.....	24
8	Logging in to the TrendMicro DSM .....	41
8.1	Server name.....	41
8.2	Server login.....	41
9	Perform policy configuration on the TrendMicro DSM .....	42
10	Exercises.....	44
10.1	Datameer – Visualize the Data.....	44
10.2	TrendMicro – Malware test.....	44
11	Visualize the Data.....	44
12	Malware Test .....	47
12.1	Generating Malware alert in the computer .....	47
12.2	Dashboard – Malware Alert .....	48
12.3	Malware Alert verification .....	48
13	References, Attachments & Definitions.....	49
13.1	References.....	49

## 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

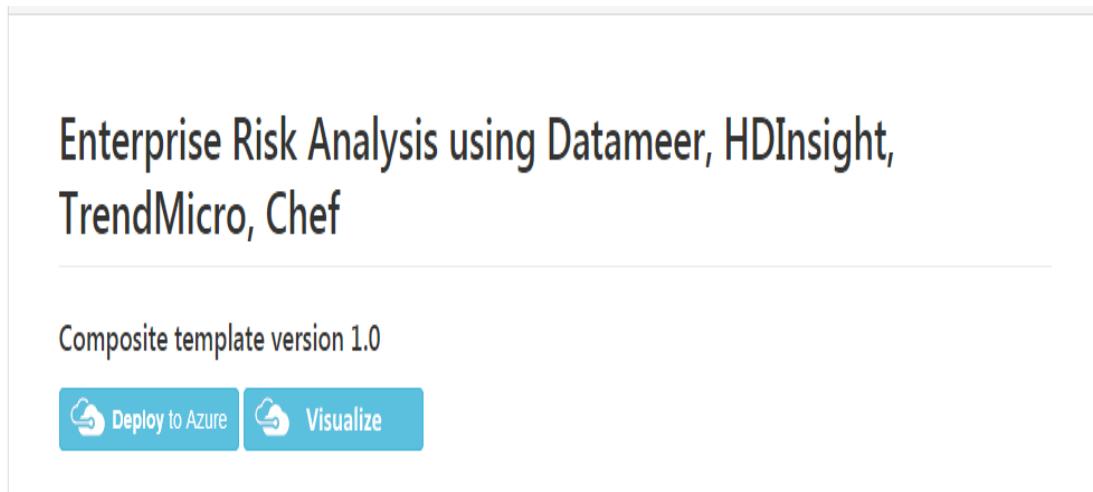
## 2 How to deploy this solution

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

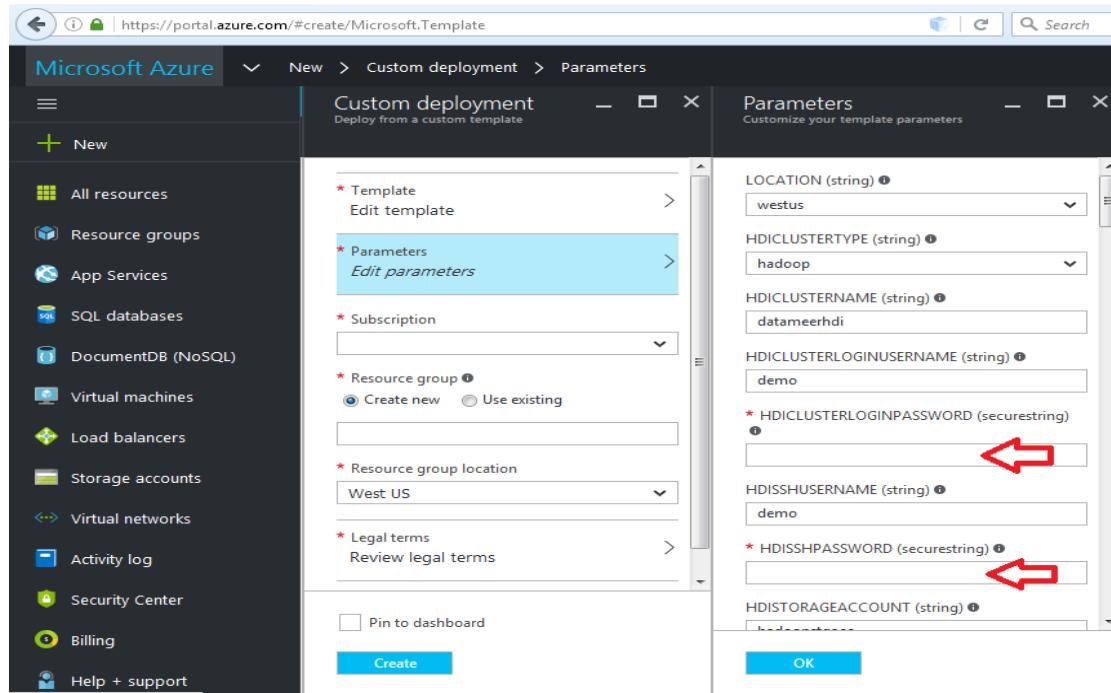
<https://github.com/AvyanConsultingCorp/azure-quickstart-templates/tree/master/datameer-trend-chef-businessintelligence>

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

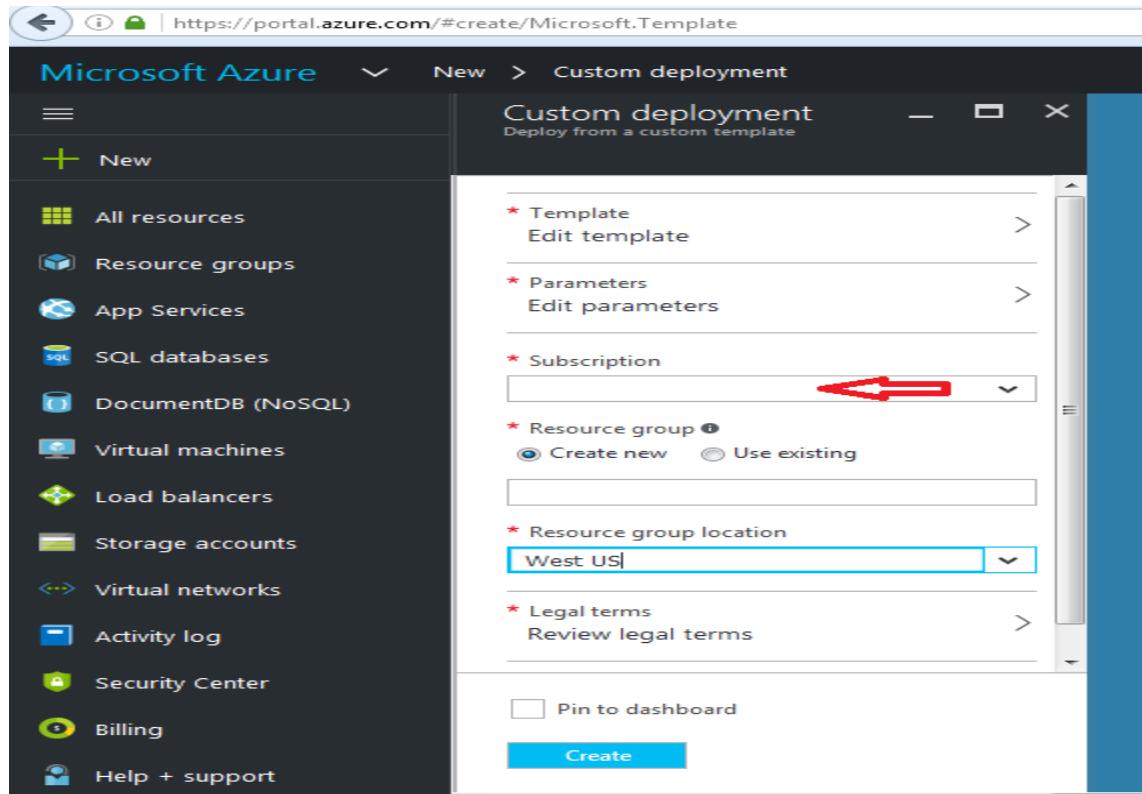


# HOL Guide for Enterprise Risk Analysis

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

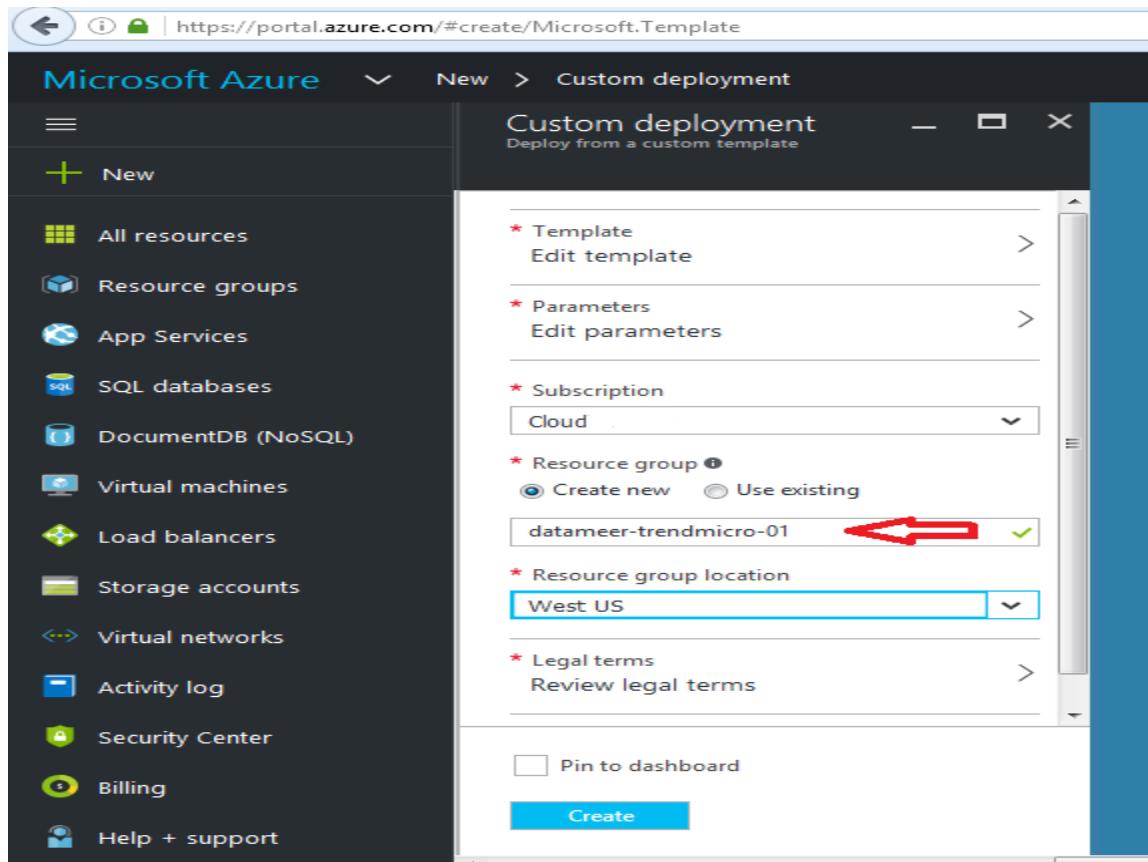


- 4) You need to select the subscription you want to deploy this solution

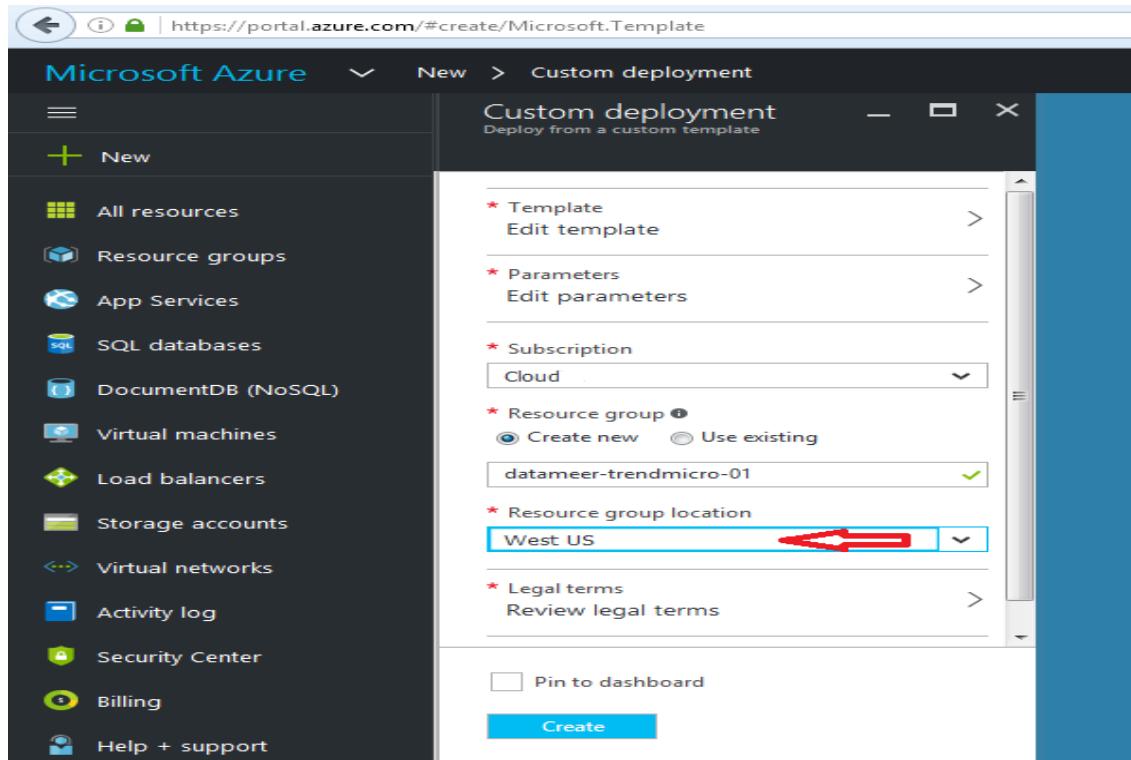


# HOL Guide for Enterprise Risk Analysis

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

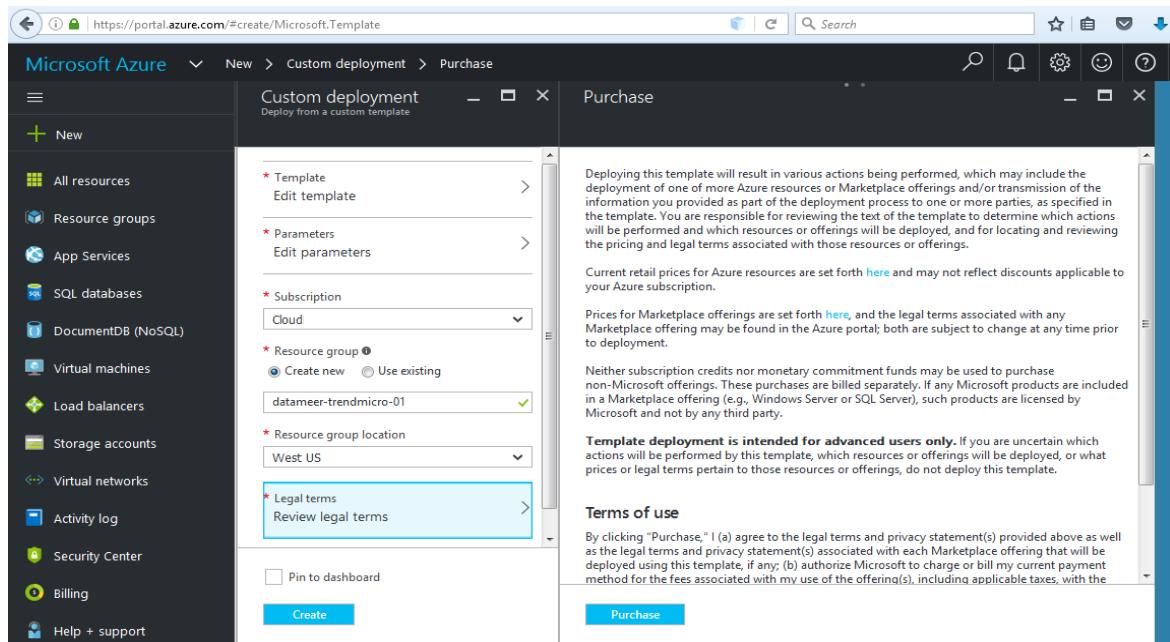


- 6) Select your choice of Region to deploy this solution,

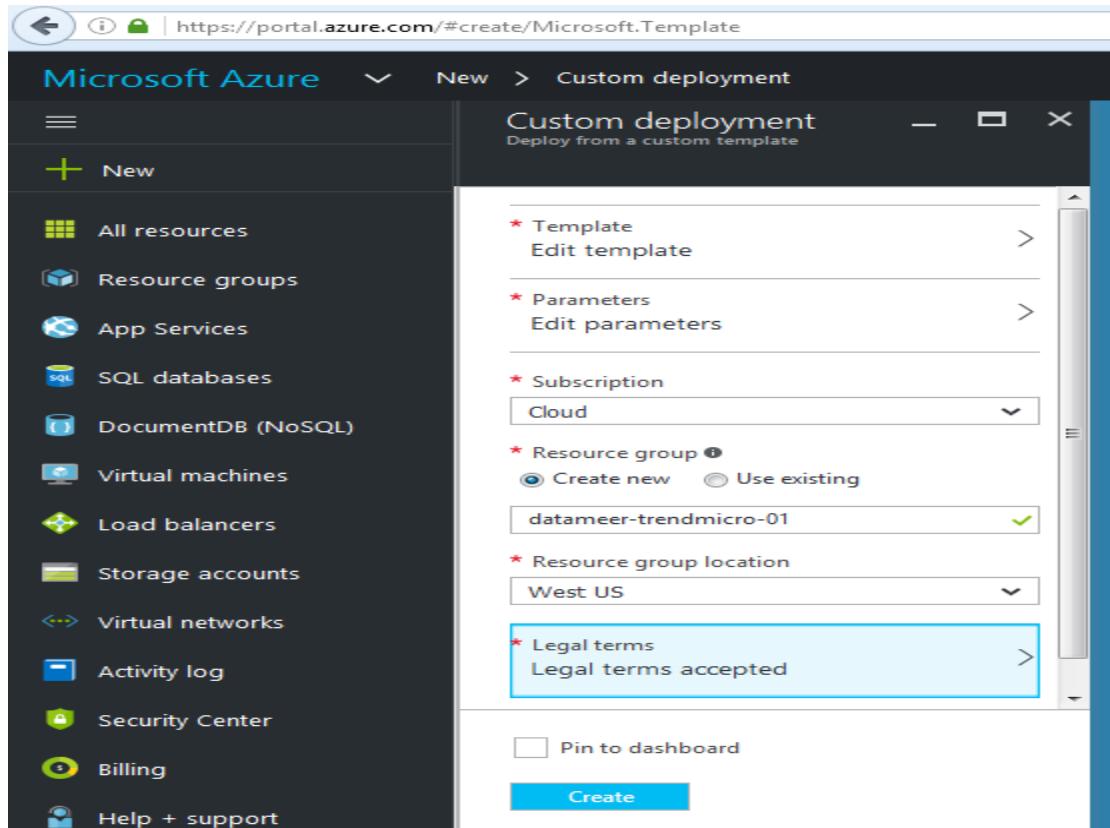


# HOL Guide for Enterprise Risk Analysis

- 7) Accept the legal terms to deploy 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 deploy 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

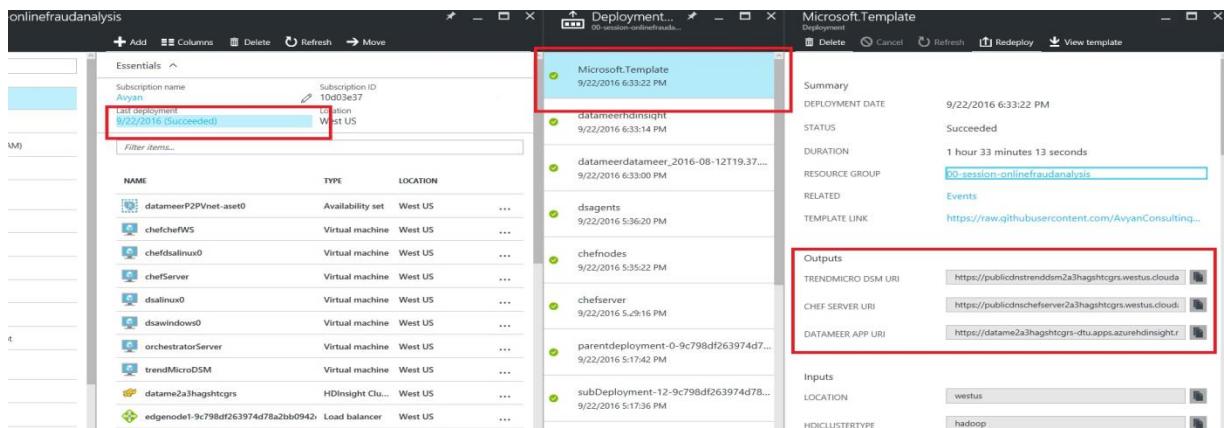
The fraud analysis is performed with the Business Analytics components of the solution, and namely Datameer and Azure HDInsight. All steps are executed in the Datameer UI. There are two ways that you can use to access the Datameer UI:

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

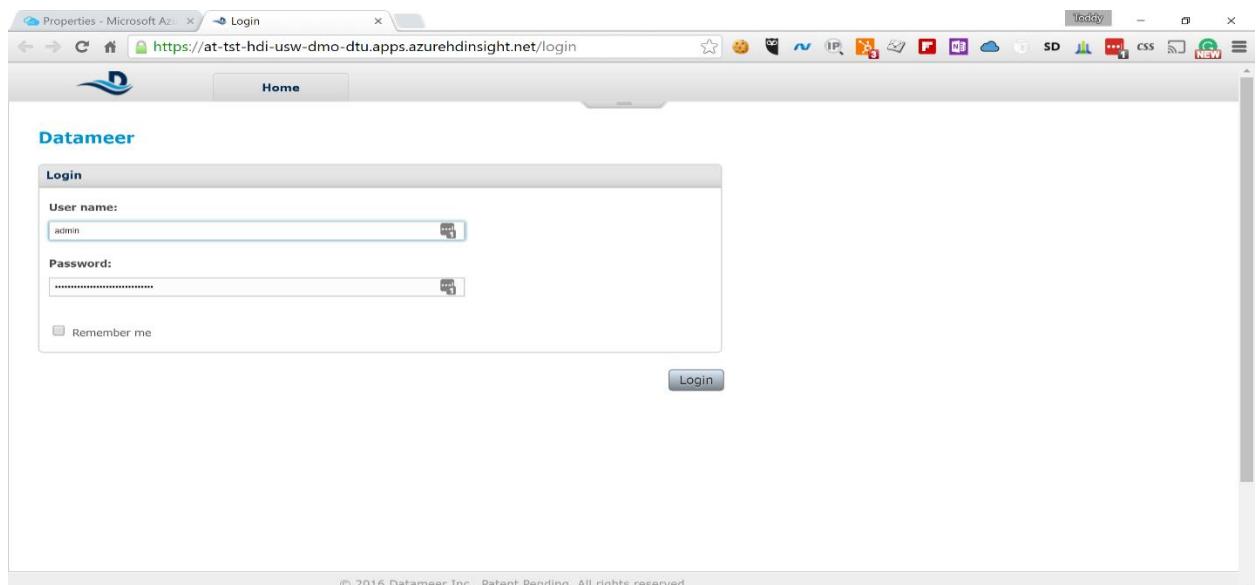
For the purposes of this HOL we will access the Datameer UI from the Azure Management Portal. Follow these steps:

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

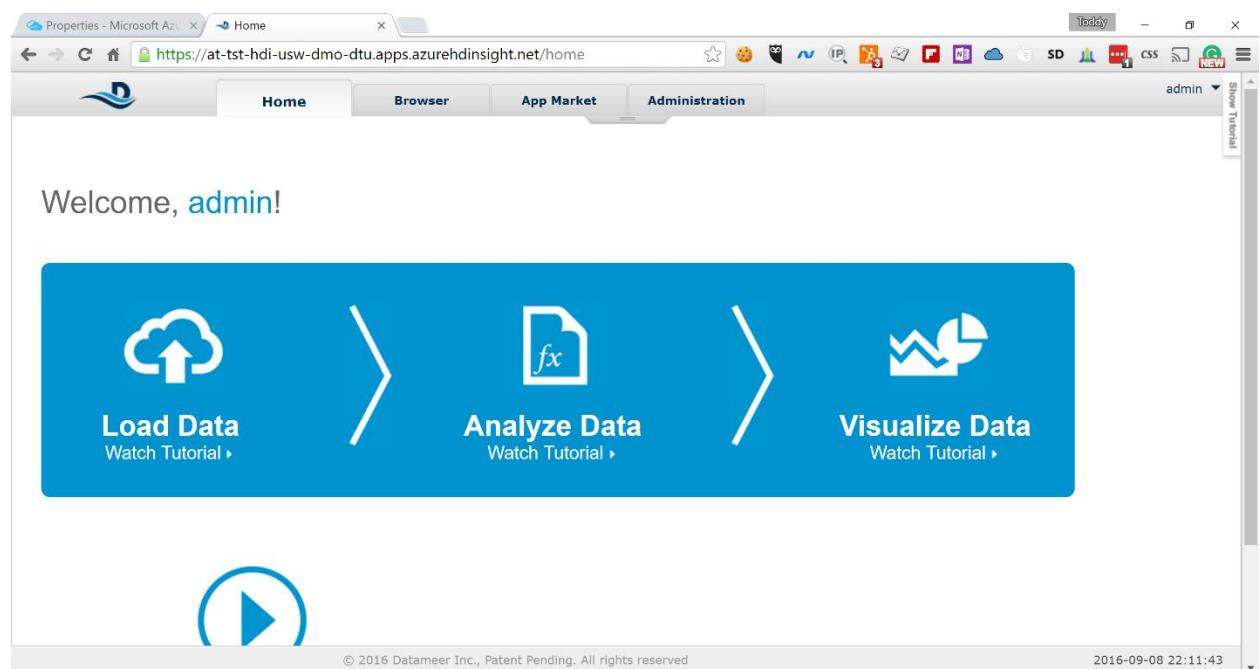
You will be prompted to sign in using your Datameer username and password

- Sign into Datameer UI using the following default credentials:

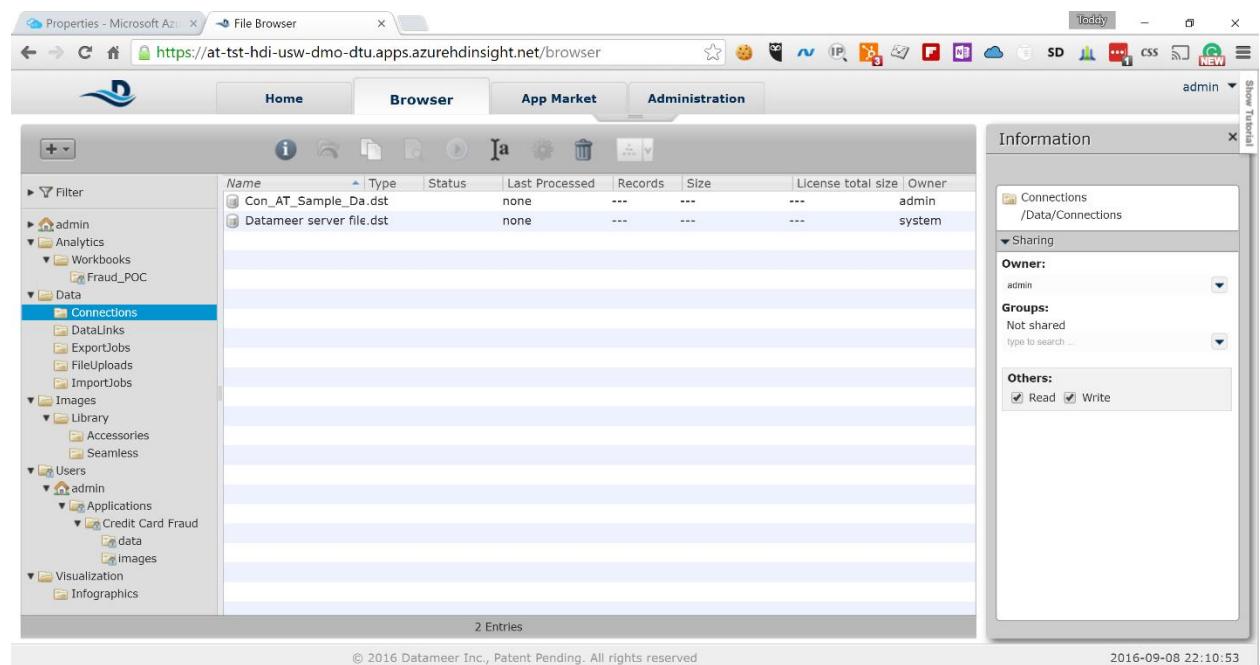
username: *admin*

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

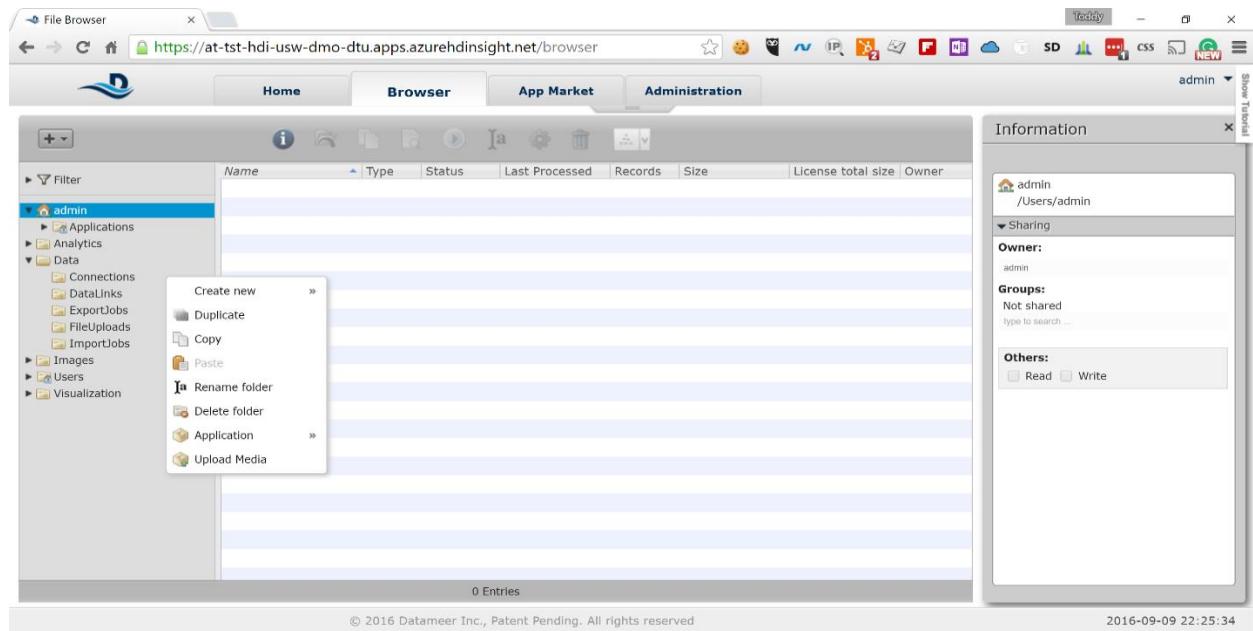


## 5 Configure Datameer to Fetch Data from Azure Storage

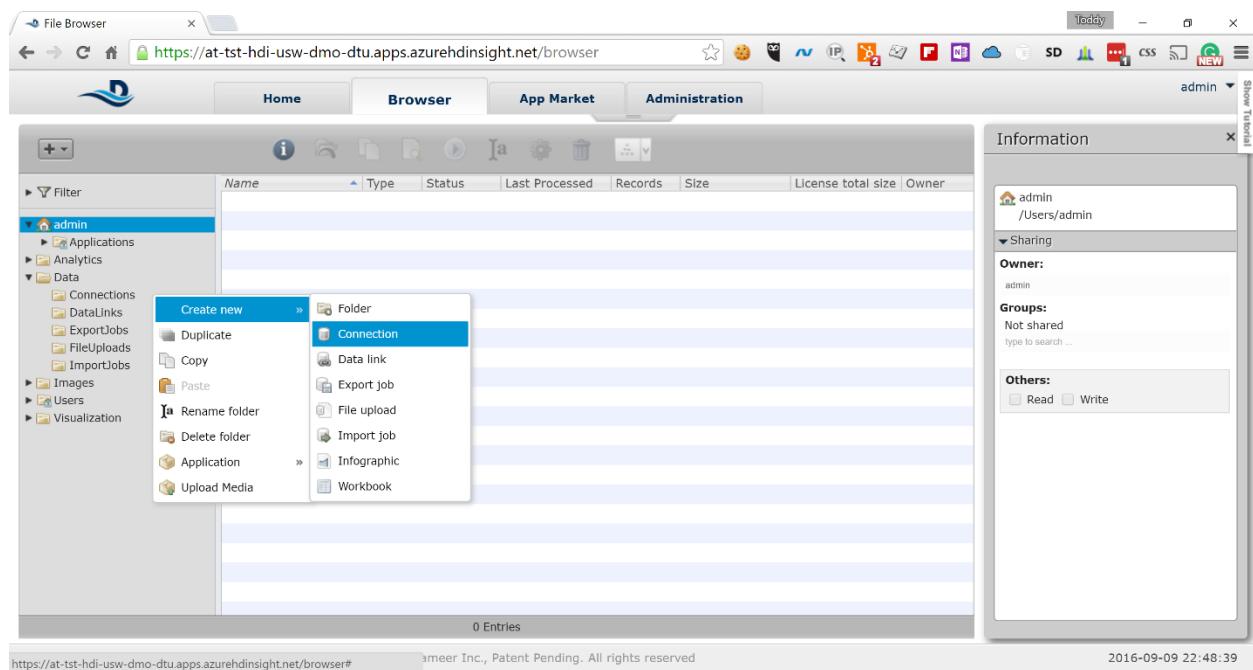
Datameer has more than 65 connectors built in, that allow various systems as data sources. For the purpose of this HOL we will use the Azure Storage connector and fetch the data from there. The assumption is that you have storage account data that contains the transaction data.

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*

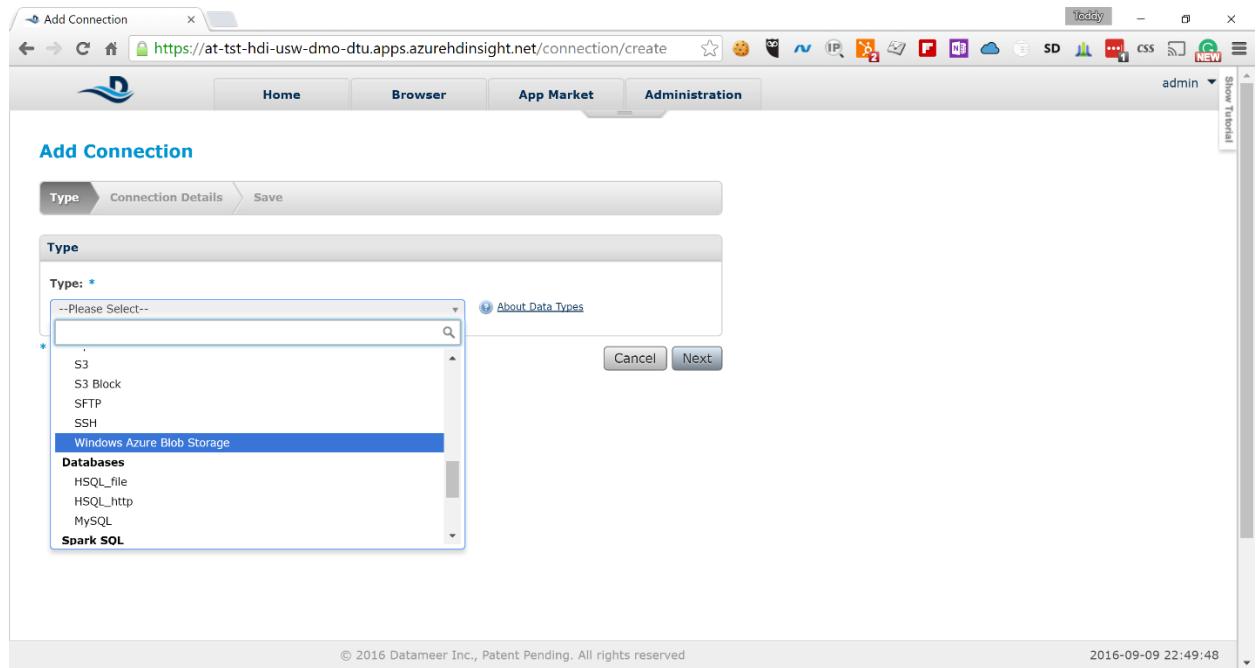


2. Select *Create new -> Connection*



# HOL Guide for Enterprise Risk Analysis

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



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

A screenshot of the 'Add Connection - Windows Azure Blob Storage' configuration screen. The title bar shows the URL 'https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/connection/create/con'. The interface is similar to the previous one, with tabs for 'Type', 'Connection Details', and 'Save'. The 'Connection Details' tab is active. It contains several input fields:

- Storage name:** A text input field containing 'mystorage'. A tooltip indicates it is the 'Name of the storage account' located at 'http://mystorage.blob.core.windows.net/'.
- Container name:** A text input field.
- Access key:** A text input field.
- Protocol:** A dropdown menu set to 'Secure'.
- Connection usage:** A dropdown menu set to 'Import/Export'.

At the bottom of the form, there is a note: '\* required' and three buttons: 'Cancel', 'Back', and 'Next'. The footer includes the copyright notice '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the timestamp '2016-09-09 22:52:18'.

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

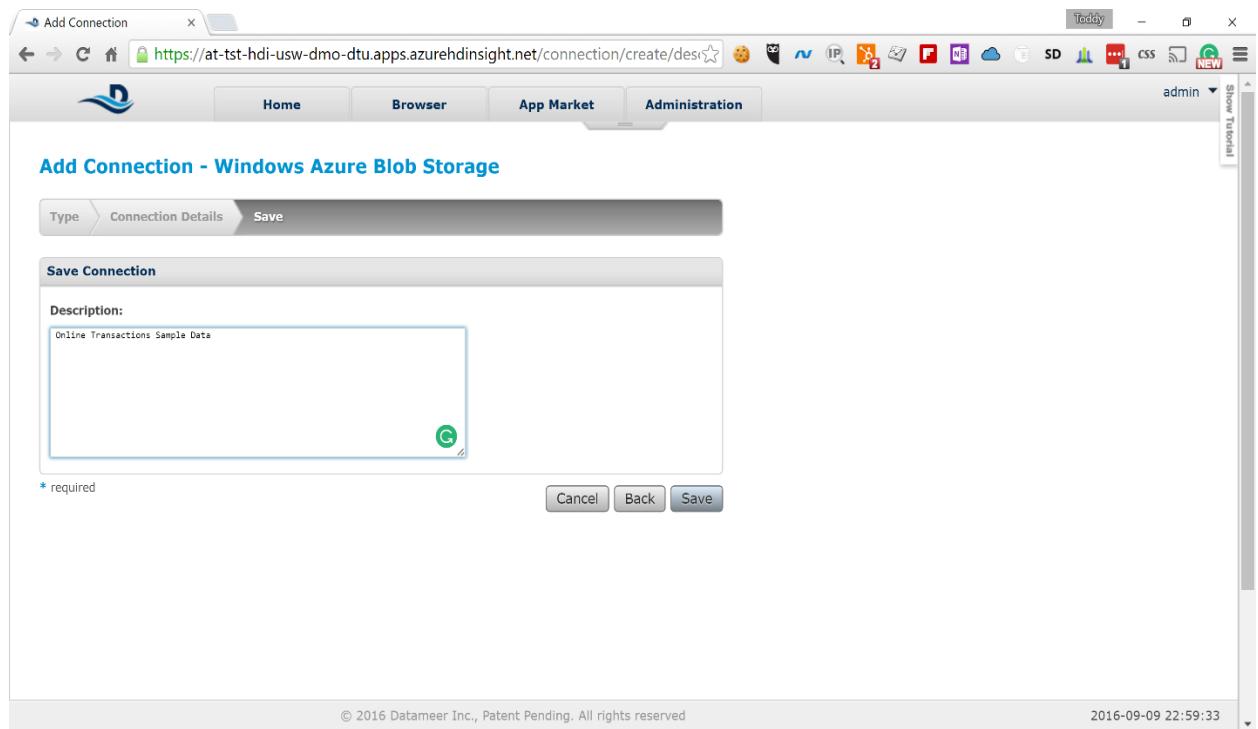
**Container Name:** The name of the container where your data is

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

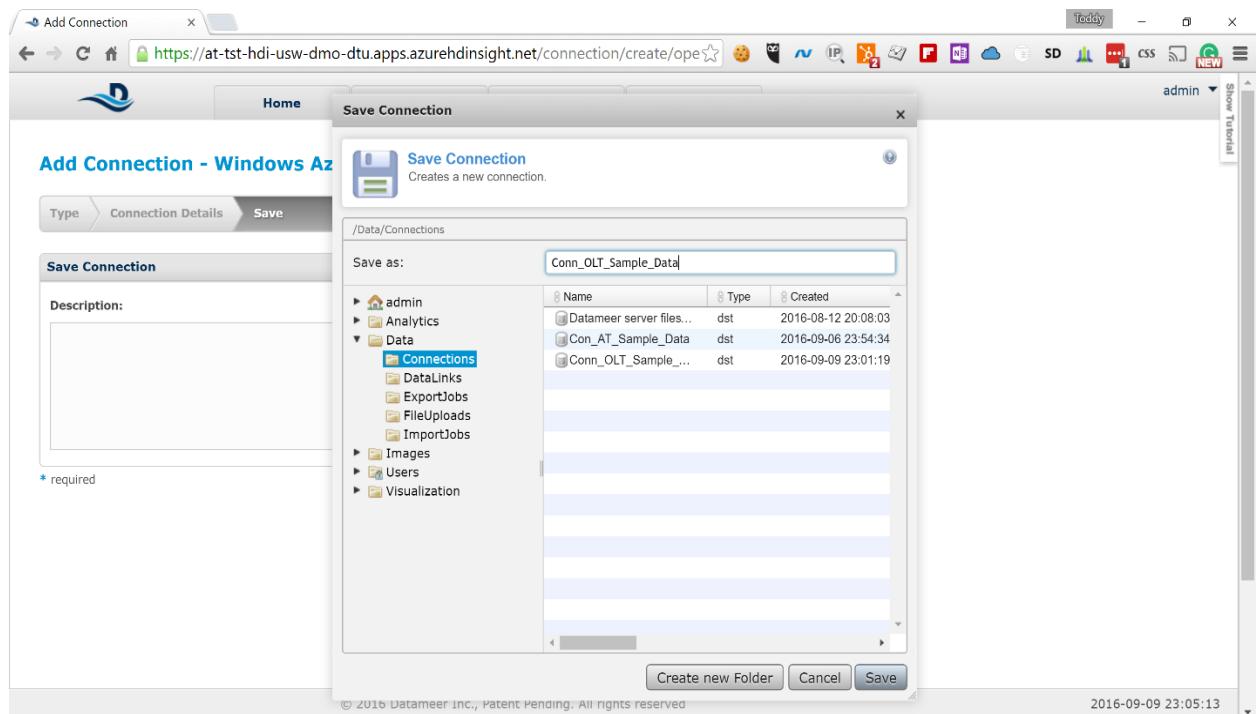
Leave the default values for *Protocol* and *Connection usage*.

# HOL Guide for Enterprise Risk Analysis

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



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

The screenshot shows the Datameer File Browser interface. The left sidebar navigation includes Home, Browser, App Market, and Administration. Under Data, there are sub-nodes: Connections, DataLinks, ExportJobs, FileUploads, ImportJobs, Images, Users, and Visualization. The main content area displays a table of connections:

Name	Type	Status	Last Processed	Records	Size	License total size	Owner
Conn_OLT_Sample_Data.dst		none	---	---	---	---	admin
Conn_AT_Sample_Data.dst		none	---	---	---	---	admin
Datameer server filesystem.dst		none	---	---	---	---	system

A right-hand panel titled "Information" provides detailed properties for the selected connection, Conn\_OLT\_Sample\_Data:

- ID: 4
- File ID: 47
- Type: Connection
- Size: ---
- Last Processed: none
- Created: Sep 09, 2016 11:01:19 PM
- Last Changed: Sep 09, 2016 11:01:19 PM
- Scheduled: Not applicable
- Description: Online Transactions Sample Data

At the bottom of the browser window, the URL is https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/browser and the page footer indicates "Datameer Inc., Patent Pending. All rights reserved". The timestamp at the bottom right is 2016-09-09 23:02:43.

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 main window displays a list of existing Data Links, including 'DL\_AT\_Online\_Trans.Ink' (1 day, 18 hrs ago), 'DL\_AT\_Online\_Trans.Ink' (2 days, 5 hrs ago), and 'DL\_AT\_Reverse\_IP .Ink' (2 days, 22 hrs ago). On the left sidebar, under 'Data' > 'Connections', 'DataLinks' is selected. A context menu is open over the third item in the list, showing options like 'Create new', 'Duplicate', 'Copy', 'Paste', 'Rename folder', 'Delete folder', 'Application', and 'Upload Media'. The 'Data link' option is highlighted. To the right, an 'Information' panel shows details for the selected 'DataLinks' entry, including its path '/Data/DataLinks', owner 'admin', and sharing settings for 'Others' (Read and Write checked). The bottom status bar shows the URL and the date/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 'Choose Connection' step is active. A 'Select Connection' dialog box is open, titled 'Select Connection', with the instruction 'Datameer will use the selected connection to import your data.' The dialog lists available connections under '/Data/Connections/Conn\_OLT\_Sample\_Data'. The 'Conn\_OLT\_Sample\_Data' connection is selected and highlighted. The dialog includes 'Create new Folder', 'Cancel', and 'Select' buttons. The bottom status bar shows the URL and the date/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 a web browser window titled 'New Data Link'. The URL is <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/import-job/chooseConnection>. The page has a navigation bar with tabs: Home, Browser, App Market, and Administration. The 'Administration' tab is selected. Below the tabs, there's a breadcrumb navigation: Connection > Data Details > Define Fields > Schedule > Save. A large section titled 'Choose Connection' contains a 'Connection' dropdown menu with 'Conn\_OLT\_Sample\_Data' selected. Below it are 'Select Connection' and 'New Connection' buttons. Another section titled 'File Type' shows a 'File Type' dropdown set to 'CSV / TSV'. A note says '\* required'. At the bottom right are 'Cancel' and 'Next' buttons. The footer includes a copyright notice '© 2016 Datameer Inc., Patent Pending. All rights reserved' and a timestamp '2016-09-09 23:27:03'.

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

The screenshot shows the same 'New Data Link' setup page, now on the 'Data Details' step. The URL is <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/import-job/create/dataDetails>. The 'Data Details' tab is selected. The 'Basic' section contains a 'Path Prefix:' input field with a single slash ('/'). A 'File Or Folder:' input field contains the path '/samples/online-transactions-cc\_non\_masked.csv'. To the right of this field is a help text explaining relative paths and wildcards, and another section about date patterns. A 'Delimiter:' input field contains a dot ('.') with a note about defining delimiter characters. A 'Schema:' input field is at the bottom. The footer includes a copyright notice and a timestamp '2016-09-09 23:29:30'.

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. The columns are: ip\_address, user\_id, timestamp, purchase\_amount, transaction\_id, credit\_card\_no, order\_no, dasFileName, dasFilePath, and dasLast. The 'timestamp' column is currently set to STRING type. At the bottom of the table, there is a 'Rescan Schema' button.

ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_no	order_no	dasFileName	dasFilePath	dasLast
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
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
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
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
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
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
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
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
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
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' column's field type has been changed to 'DATE'. The rest of the schema remains the same as the previous screenshot. The 'Rescan Schema' button is visible at the bottom.

ip_address	user_id	timestamp	purchase_amount	transaction_id	credit_card_no	order_no	dasFileName	dasFilePath	dasLast
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
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
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
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
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
11.216.44.6	2,946,179	BIG_INTEGER	\$4,108.15	12d8a13e-5013-42f...	0283-1943-9261-5...	JLD5IC	online-transactions...	/at-samples/online...	Sep 6, 2016
185.183.151.166	6,839,961	DATE	\$1,484.53	58e94bc0-48e7-4af...	4130-1703-0539-7...	YNHIYS	online-transactions...	/at-samples/online...	Sep 6, 2016
113.161.160.119	2,245,110	BOOLEAN	\$969.83	1fee04ac-43c4-416...	3674-731714-47138	6BPQIW	online-transactions...	/at-samples/online...	Sep 6, 2016
28.245.64.112	1,223,107	BIG_DECIMAL	\$276.15	1149150d-daf5-49...	1951-397612-29310	WRXRIJ	online-transactions...	/at-samples/online...	Sep 6, 2016
35.143.156.34	2,220,765	BIG_INTEGER	\$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 transaction data with columns: ip\_address, user\_id, timestamp, purchase\_amount, transaction\_id, credit\_card\_no, order\_no, desFileName, desFilePath, and desLast. The 'timestamp' column is highlighted with a red border, indicating it is an ISO-8601 date. The data table contains 15 rows of transaction details.

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	WRXR1J	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:ssZ'

The screenshot shows the Datameer interface for creating a new data link. The 'Define Fields' tab is active. The 'timestamp' column is highlighted with a red border, indicating it is an ISO-8601 date. The data table contains 15 rows of transaction details, with the 'timestamp' column now showing correctly parsed dates in the 'yyyy-MM-dd'T'HH:mm:ssZ'' format.

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	WRXR1J	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 set to "admin". Below the header is a breadcrumb navigation: Connection > Data Details > Define Fields > Schedule > Save. The main content area is titled "New Data Link" and contains a "Refresh Sample Data" section. Under "Trigger:", there are two radio buttons: "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" button with a note "\* required". 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 set to "admin". Below the header is a breadcrumb navigation: Connection > Data Details > Define Fields > Schedule > Save. The main content area is titled "New Data Link" and contains a "Save Data Link" section. Under "Description:", there is a text input field containing the text "Data link for the online transactions sample data". Below this is a "Generate now:" section with a checked checkbox "Generate sample immediately after save". An "Advanced" button with a note "\* required" is present. At the 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 sidebar shows a tree structure with 'admin', 'Analytics', 'Data' (selected), 'Connections', 'DataLinks' (selected), 'ExportJobs', 'FileUploads', 'ImportJobs', 'Images', 'Users', and 'Visualization'. On the right, a list of existing DataLinks is displayed with columns for Name, Type, and Created. The list includes '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.

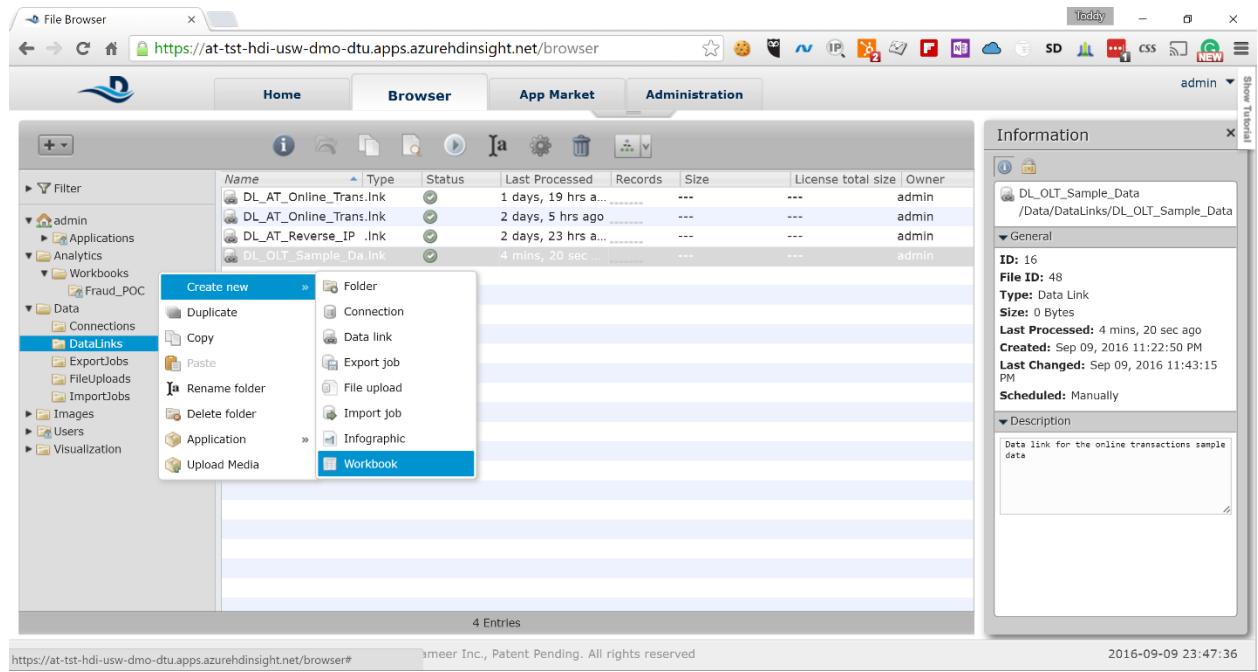
12. The new data link will appear in the list of data links on the next screen

The screenshot shows the 'File Browser' interface with the 'Data' section selected. The 'DataLinks' table has four entries: '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. The 'Information' panel on the right shows details: 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, and Scheduled: Manually. The description field contains 'Data link for the online transactions sample data'.

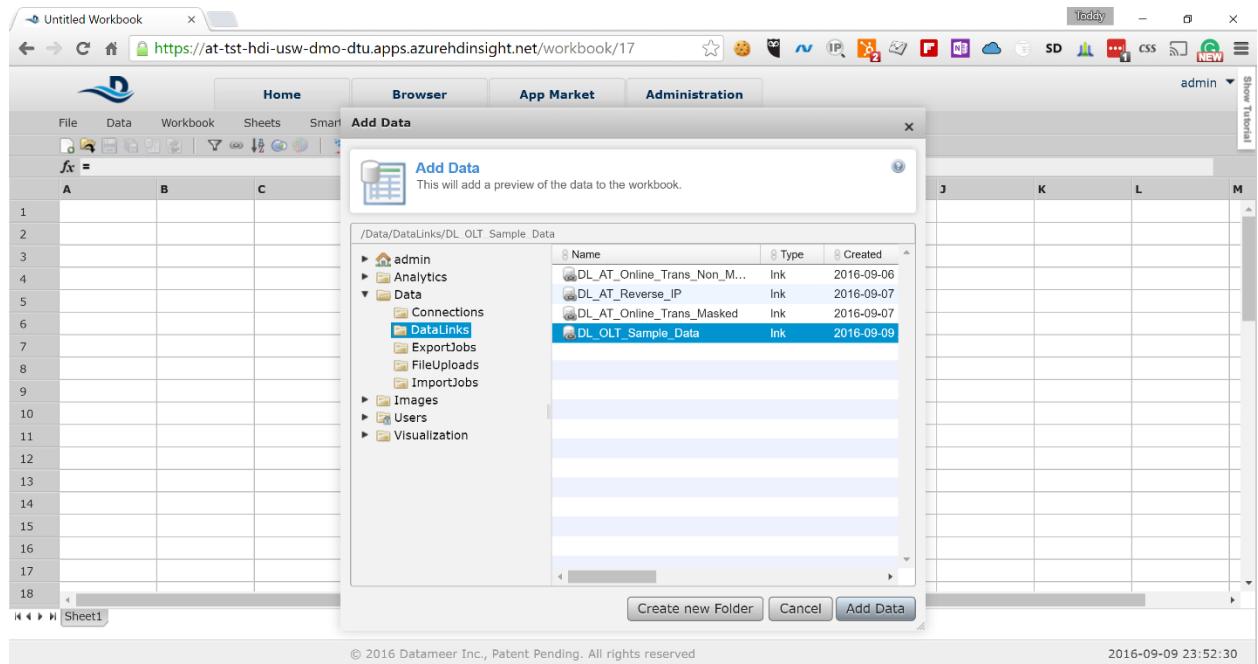
Now you have your dataset linked and have done some preliminary clean-up of the data. Next we will create a workbook where we will finish cleaning up our data and do our analysis. Here are the steps.

# 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 named '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 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...	*****_*****_**...	YNHJYHS						
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...	*****_*****_**...	WRXRIFJ						
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...	\$1,000.26	52ed8371-00c...	*****_*****_**...	MDHV8J						

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

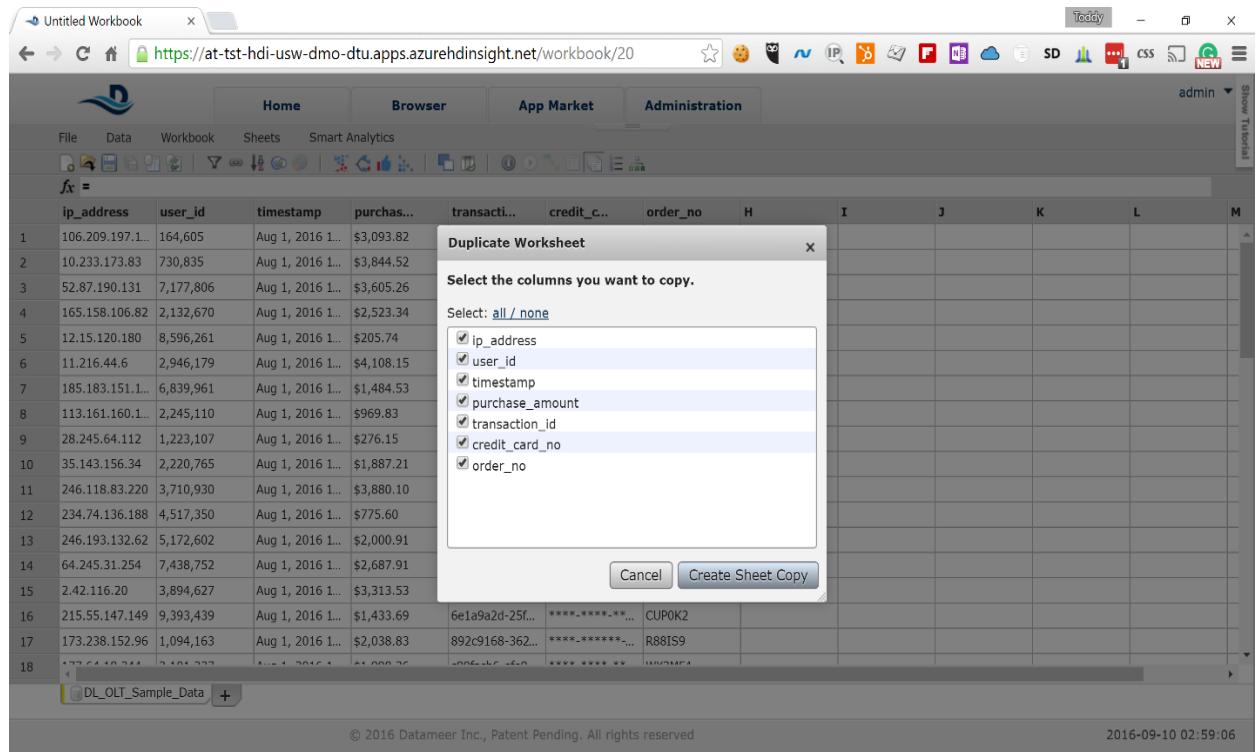
16. Right-click on the *DL\_OLT\_Sample\_Data* sheet at the bottom of the screen next and select *Duplicate*

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#

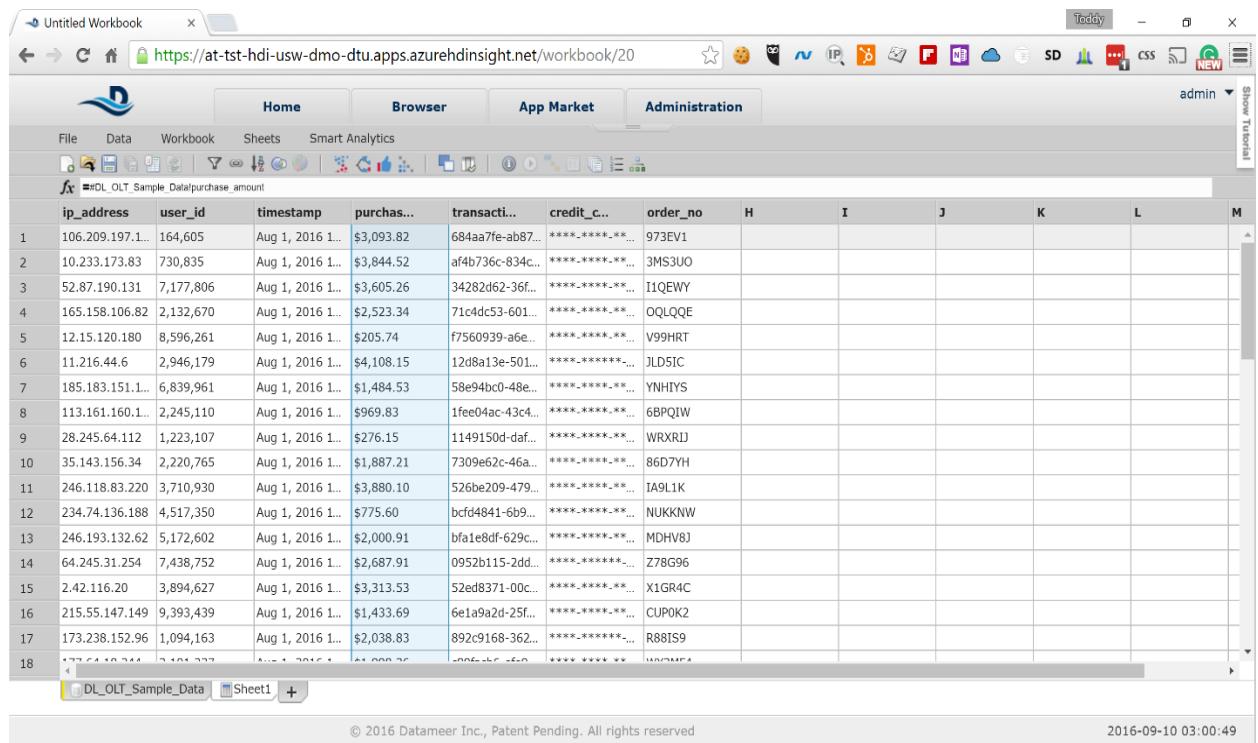
# HOL Guide for Enterprise Risk Analysis

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 box is titled 'Select the columns you want to copy.' It contains a list of columns from the original sheet: ip\_address, user\_id, timestamp, purchase\_amount, transaction\_id, credit\_card\_no, and order\_no. Each column has a checkbox next to it, and all checkboxes are checked. At the bottom of the dialog box are two buttons: 'Cancel' and 'Create Sheet Copy'.

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<sub>x</sub>* field for that column available on top of the sheet



The screenshot shows the Data Miner interface with a new sheet named 'Sheet1' visible in the bottom left corner. The sheet contains the same data as the original 'sheet1'. The 'purchase\_amount' column has an 'fx' field at the top, indicating it is now a calculated column. The rest of the columns (ip\_address, user\_id, timestamp, transaction\_id, credit\_card\_no, order\_no) do not have 'fx' fields at the top.

# 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.

ip_address	user_id	timestamp	purchas...	transacti...	credit_c...	order_no	H	I	J	K	L	M
106.209.197.1...	164,605	Aug 1, 2016 1...	3,093.82	684aa7fe-ab87...	*****_*****_**...	973EV1						
10.233.173.83	730,835	Aug 1, 2016 1...	3,844.52	af4b736c-834c...	*****_*****_**...	3MS3UO						
52.87.190.131	7,177,806	Aug 1, 2016 1...	3,605.26	34282d62-36f...	*****_*****_**...	I1QEWWY						
165.158.106.82	2,132,670	Aug 1, 2016 1...	2,523.34	71c4dc53-601...	*****_*****_**...	OQLQQE						
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						
185.183.151.1...	6,839,961	Aug 1, 2016 1...	1,484.53	58e94bc0-48e...	*****_*****_**...	YNHIYS						
113.161.160.1...	2,245,110	Aug 1, 2016 1...	969.83	1fee04ac-43c4...	*****_*****_**...	6BPQIW						
28.245.64.112	1,223,107	Aug 1, 2016 1...	276.15	1149150d-daf...	*****_*****_**...	WRXRJU						
35.143.156.34	2,220,765	Aug 1, 2016 1...	1,887.21	7309e62c-46a...	*****_*****_**...	86D7YH						
246.118.83.220	3,710,930	Aug 1, 2016 1...	3,880.1	526be209-479...	*****_*****_**...	IA9L1K						
234.74.136.186	4,517,350	Aug 1, 2016 1...	775.6	bcd4841-6b9...	*****_*****_**...	NUKKNW						
246.193.132.62	5,172,602	Aug 1, 2016 1...	2,000.91	bfa1e8df-629c...	*****_*****_**...	MDHV8J						
64.245.31.254	7,438,752	Aug 1, 2016 1...	2,687.91	0952b115-2dd...	*****_*****_**...	Z78G96						
2.42.116.20	3,894,627	Aug 1, 2016 1...	3,313.53	52ed8371-00c...	*****_*****_**...	X1GR4C						
215.55.147.149	9,393,439	Aug 1, 2016 1...	1,433.69	6e1a9a2d-25f...	*****_*****_**...	CUP0K2						
173.238.152.96	1,094,163	Aug 1, 2016 1...	2,038.83	892c9168-362...	*****_*****_**...	R88IS9						
18												

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

- Rename
- Delete
- Duplicate
- Move

# HOL Guide for Enterprise Risk Analysis

## 21. Rename *Sheet1* to *Transaction\_Data*

The screenshot shows a DataMeer workspace titled "Untitled Workbook". The "Home" tab is selected. A table titled "DL\_OLT\_Sample\_Data" is displayed, containing 18 rows of transaction data. The columns include ip\_address, user\_id, timestamp, purchase\_amount, transaction\_id, credit\_card, order\_no, and several columns with masked values. The table has a header row and 18 data rows. The bottom of the screen shows the footer: "© 2016 Datameer Inc., Patent Pending. All rights reserved" and the date "2016-09-10 03:04:51".

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

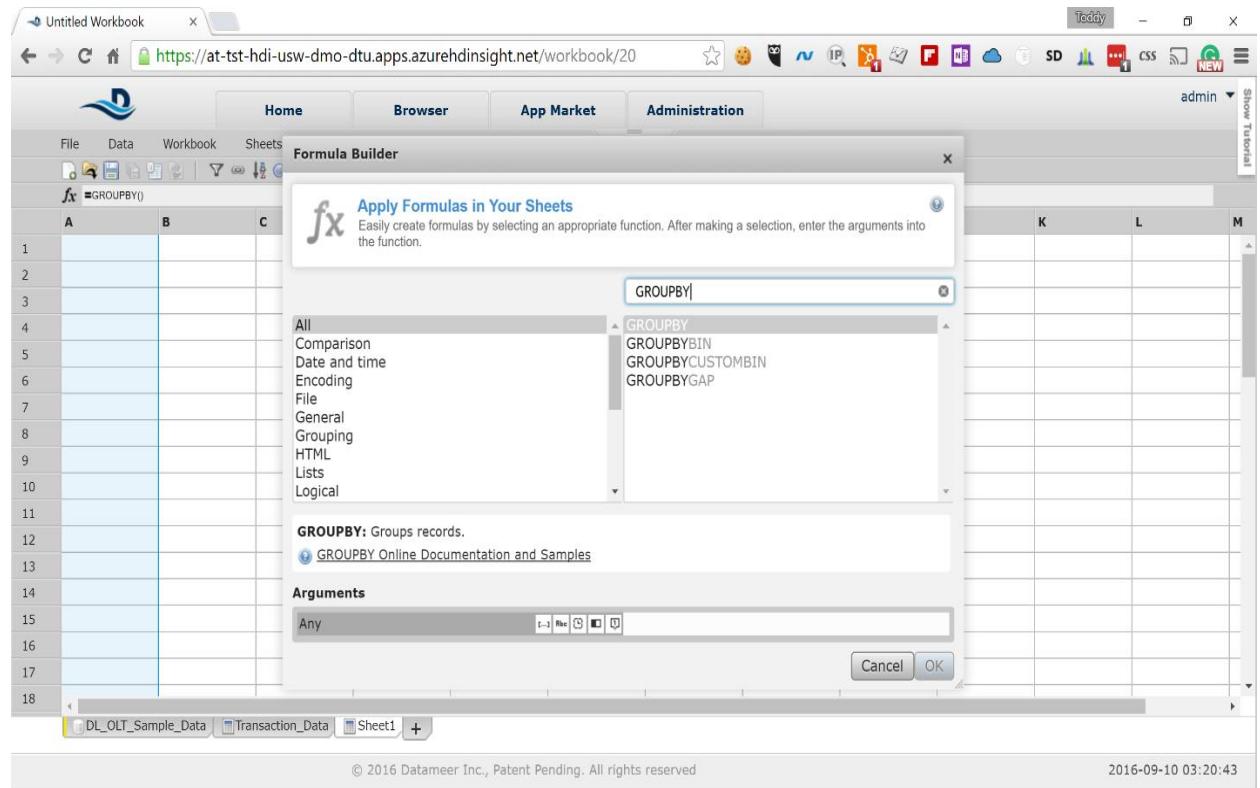
The goal of our analysis is to identify unusual purchasing patterns that deviate from a well-established norm. If we notice something unusual this may be a sign that fraud may be committed. For the purpose of this HOL we will be looking for periods during the month, in which the transactions significantly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

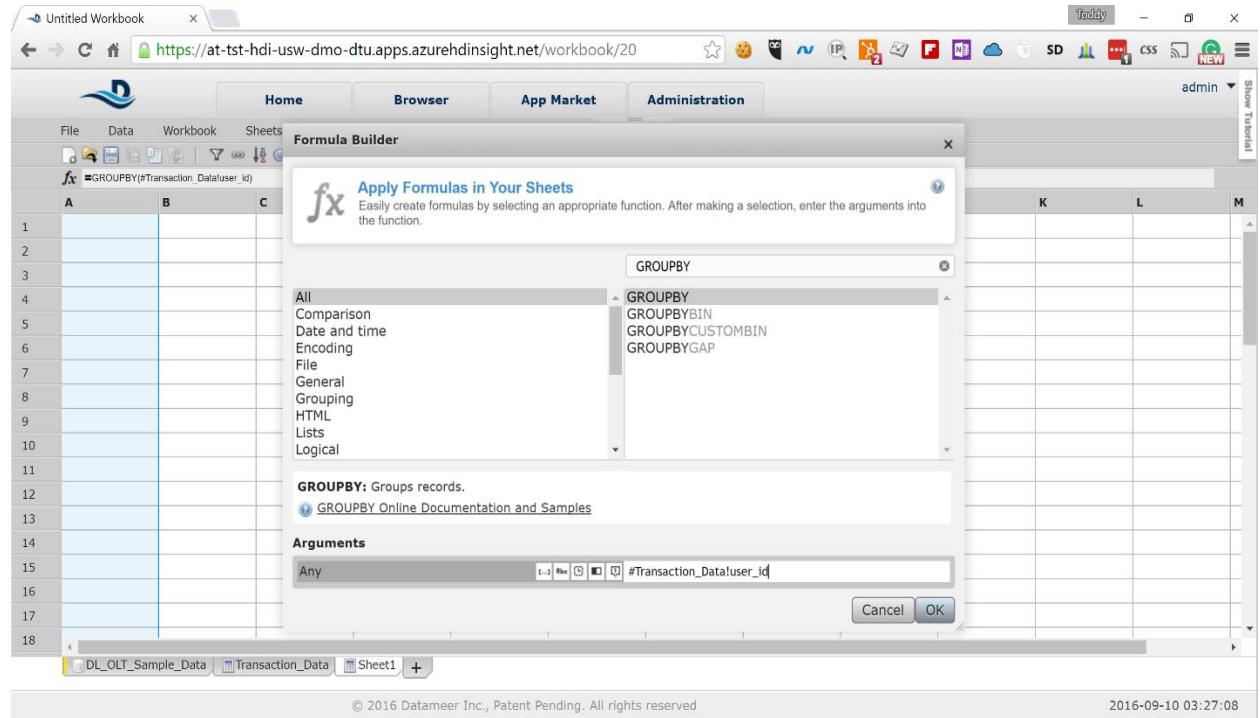
The screenshot shows the DataMeer interface with the "Formula Builder" dialog open over a blank sheet titled "Sheet1". The dialog title is "Apply Formulas in Your Sheets" and it says "Easily create formulas by selecting an appropriate function. After making a selection, enter the arguments into the function." A dropdown menu "All" is open, showing various formula categories: ABS, ACOS, ACOSH, ADD, ADDTODATE, AFTER, ANALYZE\_POLARITY, AND, ASDATE, ASIN, etc. At the bottom right of the dialog are "Cancel" and "OK" buttons. The bottom of the screen shows the footer: "© 2016 Datameer Inc., Patent Pending. All rights reserved" and the date "2016-09-10 03:17:00".

# 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											

Below the table, there are tabs for 'DL\_OLT\_Sample\_Data', 'Transaction\_Data', and 'Sheet1'. The status bar at the bottom indicates '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the date '2016-09-10 03:30:19'.

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 of data. The table has columns A through M, with data identical to the one in the previous screenshot. The 'Formula Builder' dialog has the following details:

- Function: GROUPAVERAGE
- Arguments: Number #Transaction\_Data!purchase\_amount

The status bar at the bottom indicates '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the date '2016-09-10 03:45:46'.

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 dialog title is "Formula Builder" and the sub-section is "Apply Formulas in Your Sheets". A search bar contains "GROUPSTDEVP". Below it, a list of categories includes "All", "Comparison", "Date and time", "Encoding", "File", "General", "Grouping", "HTML", "Lists", and "Logical". The "All" category is expanded, showing the "GROUPSTDEVP" function. A description below states: "GROUPSTDEVP: Estimates standard deviation based on the entire population." There is a link to "GROUPSTDEVP Online Documentation and Samples". The "Arguments" section shows a dropdown menu set to "Number" and a text input field containing "#Transaction\_Data!purchase\_amount". At the bottom right of the dialog are "Cancel" and "OK" buttons.

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 tabs at the bottom now include "DL\_OLT\_Sample\_Data", "Transaction\_Data", and "Stats". The "Stats" tab is active. The data table has three columns: "user\_id", "Average...", and "purchas...". The "purchas..." column contains many zeros, indicating the standard deviation calculation has been applied to the purchase amount field.

# 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

Create a Joined Sheet

Creates a new sheet containing data from two or more sheets based on a key column.

Select sheet & column

Drag columns to define join

Simple Range Join

Inner Join

Choose included columns ...

Cancel Create Joined Sheet

© 2016 Datameer Inc., Patent Pending. All rights reserved

2016-09-10 04:16:56

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.

Create a Joined Sheet

Creates a new sheet containing data from two or more sheets based on a key column.

Select sheet & column

Drag columns to define join

Simple Range Join

Inner Join

Transaction\_Data/user\_id Stats/user\_id

Choose included columns ...

Cancel Create Joined Sheet

© 2016 Datameer Inc., Patent Pending. All rights reserved

2016-09-10 04:45:06

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...	*****-*****-*****...	M3J 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	3,409	3,409	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	b6c52d95-2ac...	*****-*****-*****...	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-aef...	*****-*****-*****...	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	22.22.11.111	Aug 1, 2016 0...	1,060.69	00000000-0000-0...	*****-*****-*****...	1V000000	1,060.69	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*

*average\_purchase\_amount*

*purchase\_amount\_deviation*

Select the columns you want to copy.

Select: all / none

user\_id  
 Stats.user\_id  
 Transaction\_Data.ip\_address  
 Transaction\_Data.timestamp  
 purchase\_amount  
 Transaction\_Data.transaction\_id  
 Transaction\_Data.credit\_card\_no  
 Transaction\_Data.order\_no  
 average\_purchase\_amount  
 purchase\_amount\_deviation

Click on *Create Sheet Copy* button

# HOL Guide for Enterprise Risk Analysis

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 Editor interface with a table of transaction data. A context menu is open over the first row, with the 'Filter ...' option selected. The table has columns labeled 'user\_id', 'purchas...', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', and 'M'. The data rows show various user IDs and purchase amounts.

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 in the Data Editor. The 'Advanced' tab is selected. The formula  $ABS(\#purchase\_amount - \#average\_purchase\_amount) > 2 * \#purchase\_amount\_deviation$  is entered in the 'Full expression' field. The dialog also includes a note about applying the filter to the current sheet and creating a new sheet.

# HOL Guide for Enterprise Risk Analysis

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 top navigation bar includes 'Home', 'Browser', 'App Market', and 'Administration'. The left sidebar has 'File', 'Data', 'Workbook', 'Sheets', and 'Smart Analytics' options. The main area displays a table with columns: user\_id, purchas..., average..., purchas..., E, F, G, H, I, J, K, L, M. Rows 1 through 18 are listed vertically. Below the table, tabs include 'DL\_OLT\_Sample\_Data', 'Transaction\_Data', 'Stats', 'Joined\_Data\_and\_Stats', 'Outliers', and '+'. At the bottom, a footer bar shows '© 2016 Datameer Inc., Patent Pending. All rights reserved' and the date '2016-09-12 18:22:46'.

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 Datameer interface with the 'Formula Builder' dialog box open. The dialog title is 'Apply Formulas in Your Sheets'. It contains a search bar 'fx' and a dropdown menu 'GROUPBY' which lists 'GROUPBY', 'GROUPBYBIN', 'GROUPBYCUSTOMBIN', and 'GROUPBYGAP'. Below the dropdown, a description states 'GROUPBY: Groups records.' and provides a link to 'GROUPBY Online Documentation and Samples'. An 'Arguments' section shows a field 'Any' containing the formula 'YEAR(#Transaction\_Data!timestamp)'. Buttons for 'Cancel' and 'OK' are at the bottom right. The background shows the same 'Untitled Workbook' interface as the previous screenshot.

In the *Arguments* field type the following formula:  
*YEAR(#Transaction\_Data!timestamp)*

# HOL Guide for Enterprise Risk Analysis

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)*

The screenshot shows the DataMelt interface with the 'Formula Builder' dialog open. The dialog title is 'Apply Formulas in Your Sheets'. In the 'Arguments' section, the formula `MONTH(#Transaction_Data!timestamp)` is entered. The 'OK' button is visible at the bottom right of the dialog.

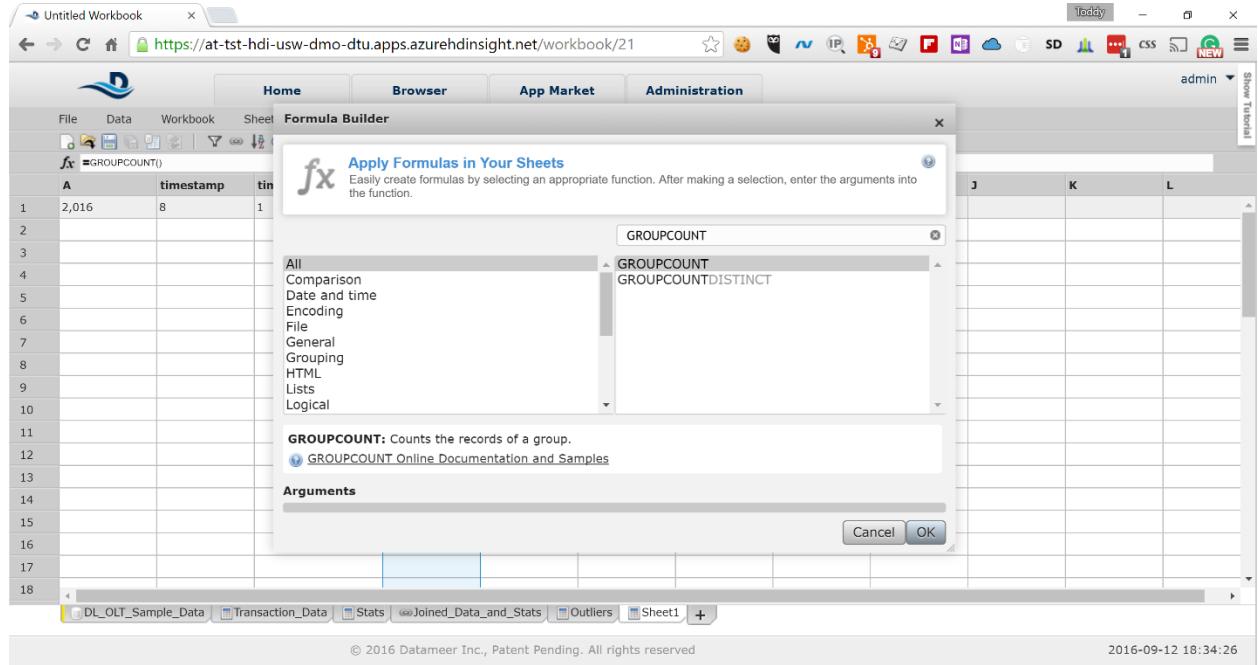
18. Click on the third column and in the formula pop-up select again the GROUPBYfunction and paste the following formula in the Arguments field:

*DAY(#Transaction\_Data!timestamp)*

The screenshot shows the DataMelt interface with the 'Formula Builder' dialog open. The dialog title is 'Apply Formulas in Your Sheets'. In the 'Arguments' section, the formula `DAY(#Transaction_Data!timestamp)` is entered. The 'OK' button is visible at the bottom right of the dialog.

# HOL Guide for Enterprise Risk Analysis

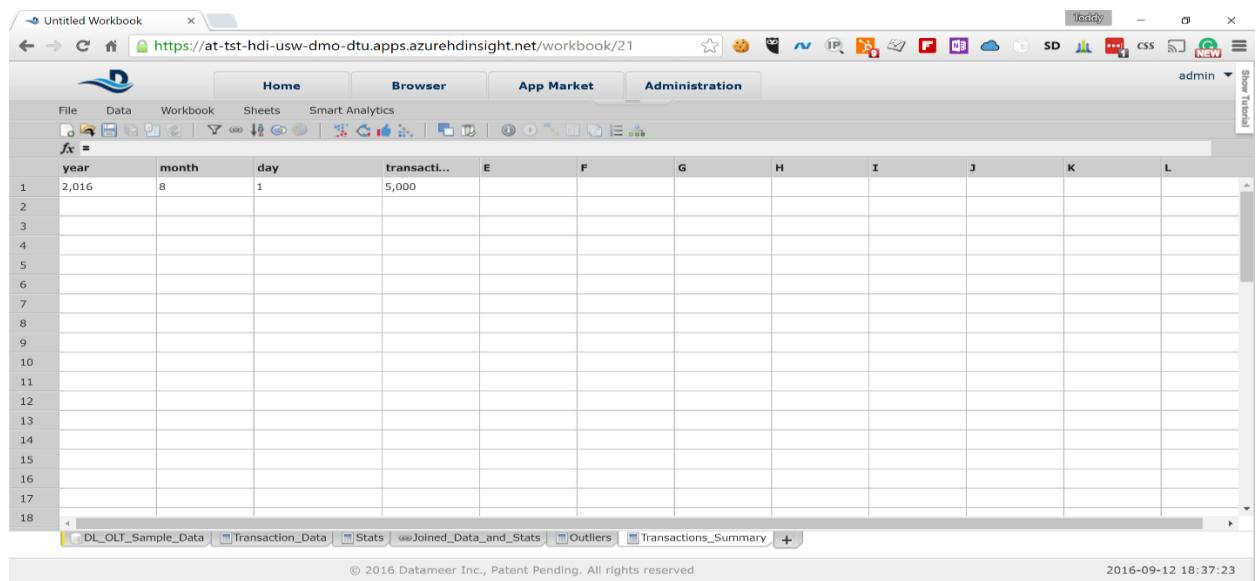
19. Click on the fourth column and in the formula pop-up select the *GROUPCOUNT* function and click the *OK* button



The screenshot shows the DataMeer interface with the 'Formula Builder' dialog open. The dialog title is 'Apply Formulas in Your Sheets'. In the center, the 'GROUPCOUNT' function is selected from a list. Below it, the description reads: 'GROUPCOUNT: Counts the records of a group.' At the bottom right of the dialog are 'Cancel' and 'OK' buttons. The background shows a spreadsheet with columns labeled 'A', 'timestamp', and 'time'.

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

*year*  
*month*  
*day*  
*transactions\_count*



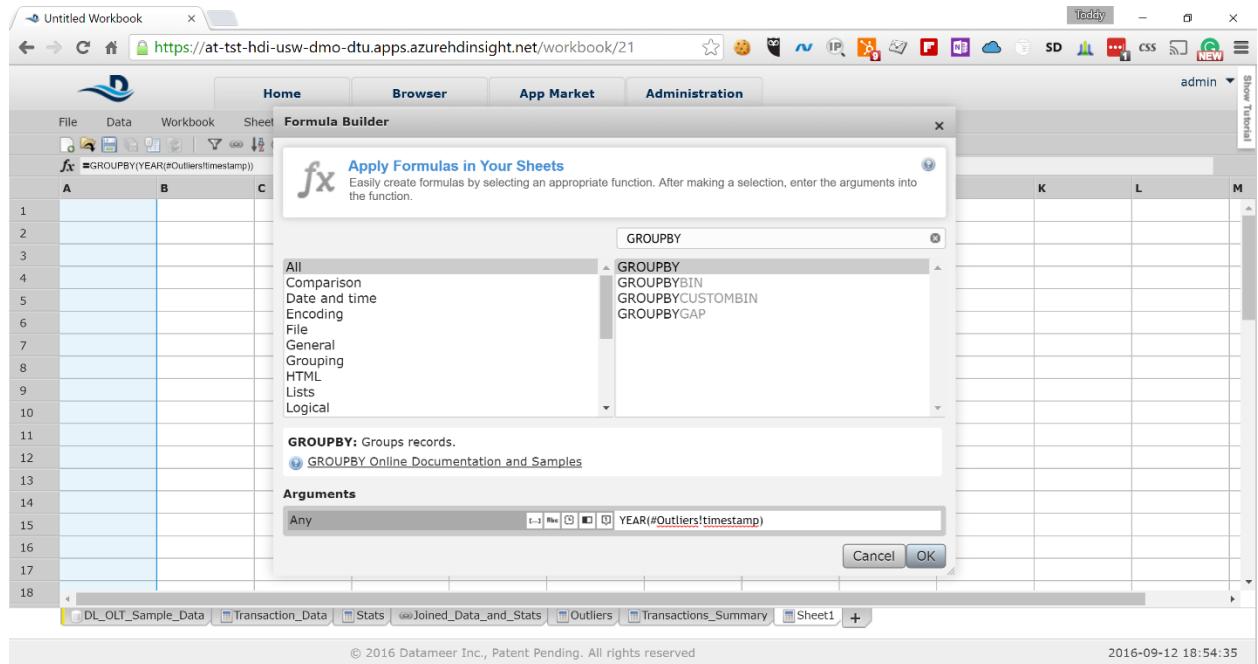
The screenshot shows the DataMeer interface with the 'Transaction\_Data' sheet selected. The columns are renamed: 'year', 'month', 'day', and 'transacti...'. The 'transacti...' column contains the value '5,000'. The background shows other sheets like 'DL\_OLT\_Sample\_Data', 'Stats', 'Joined\_Data\_and\_Stats', 'Outliers', and 'Sheet1'.

Also, rename the sheet to *Transactions\_Summary*

# HOL Guide for Enterprise Risk Analysis

21. 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:

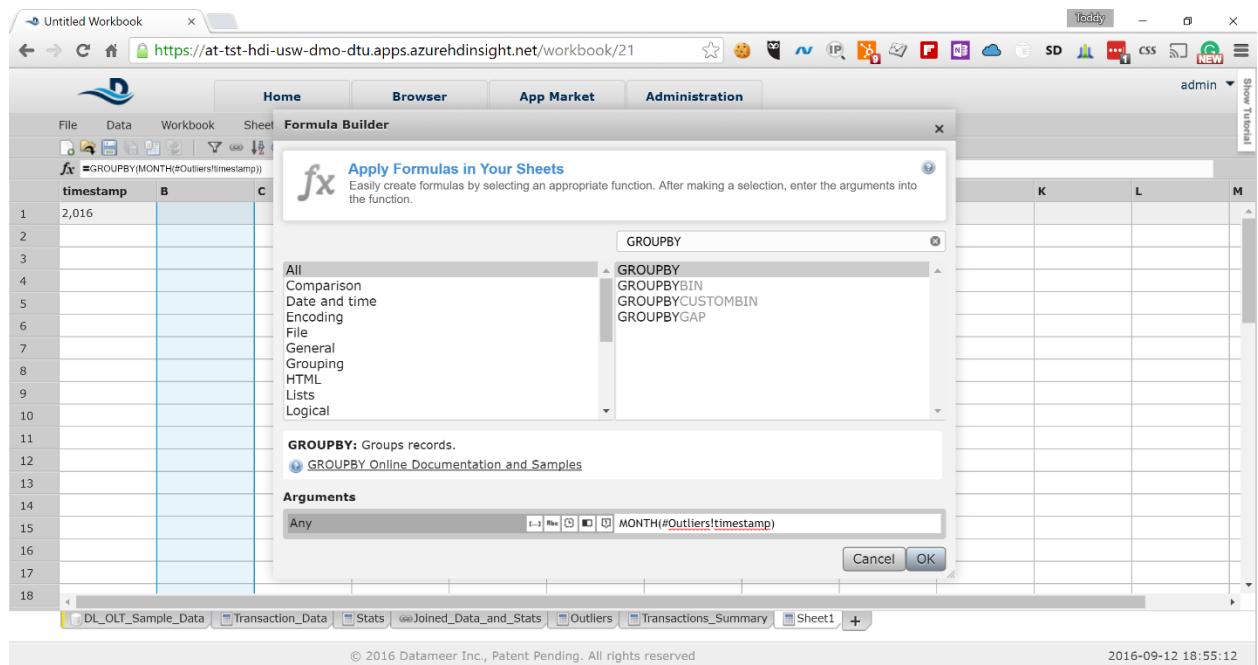
`YEAR(#Outliers!timestamp)`



The screenshot shows the DataMeer interface with the 'Formula Builder' dialog open. The formula being built is `=GROUPBY(YEAR(#Outliers!timestamp))`. The 'GROUPBY' function is selected from the list of available functions. The 'Arguments' field contains the formula `YEAR(#Outliers!timestamp)`. The DataMeer ribbon at the top includes tabs for Home, Browser, App Market, and Administration. The left sidebar shows a tree view of the workbook structure.

22. 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)`

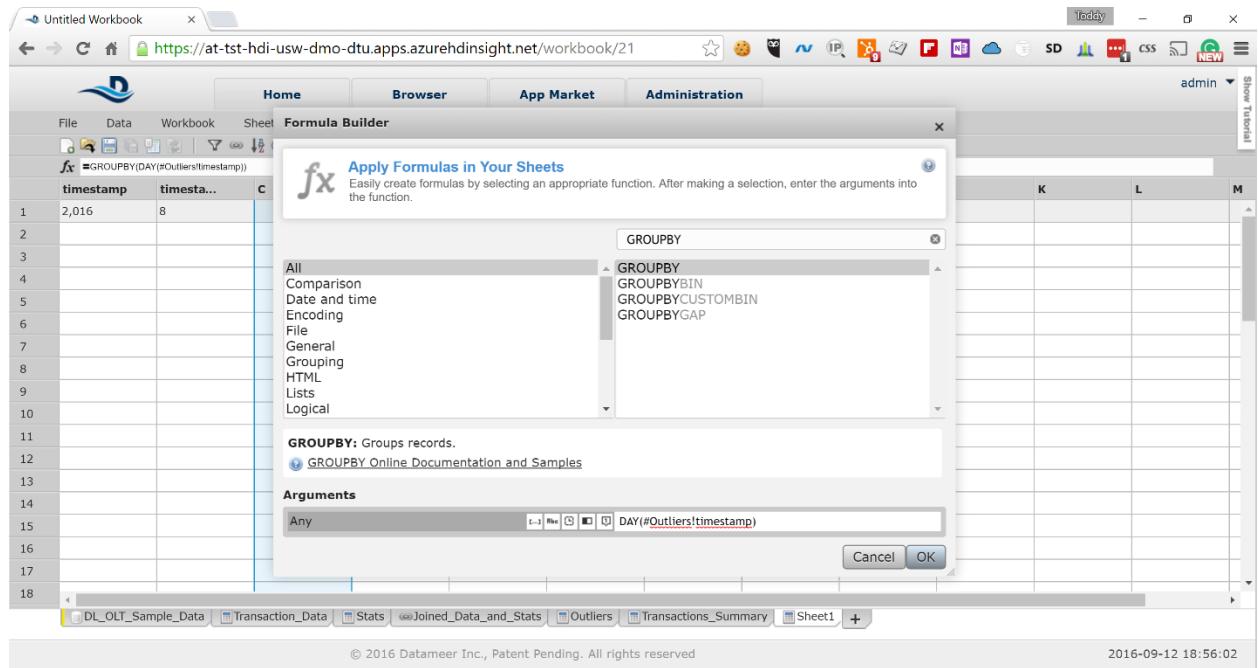


The screenshot shows the DataMeer interface with the 'Formula Builder' dialog open. The formula being built is `=GROUPBY(MONTH(#Outliers!timestamp))`. The 'GROUPBY' function is selected from the list of available functions. The 'Arguments' field contains the formula `MONTH(#Outliers!timestamp)`. The DataMeer ribbon at the top includes tabs for Home, Browser, App Market, and Administration. The left sidebar shows a tree view of the workbook structure.

# HOL Guide for Enterprise Risk Analysis

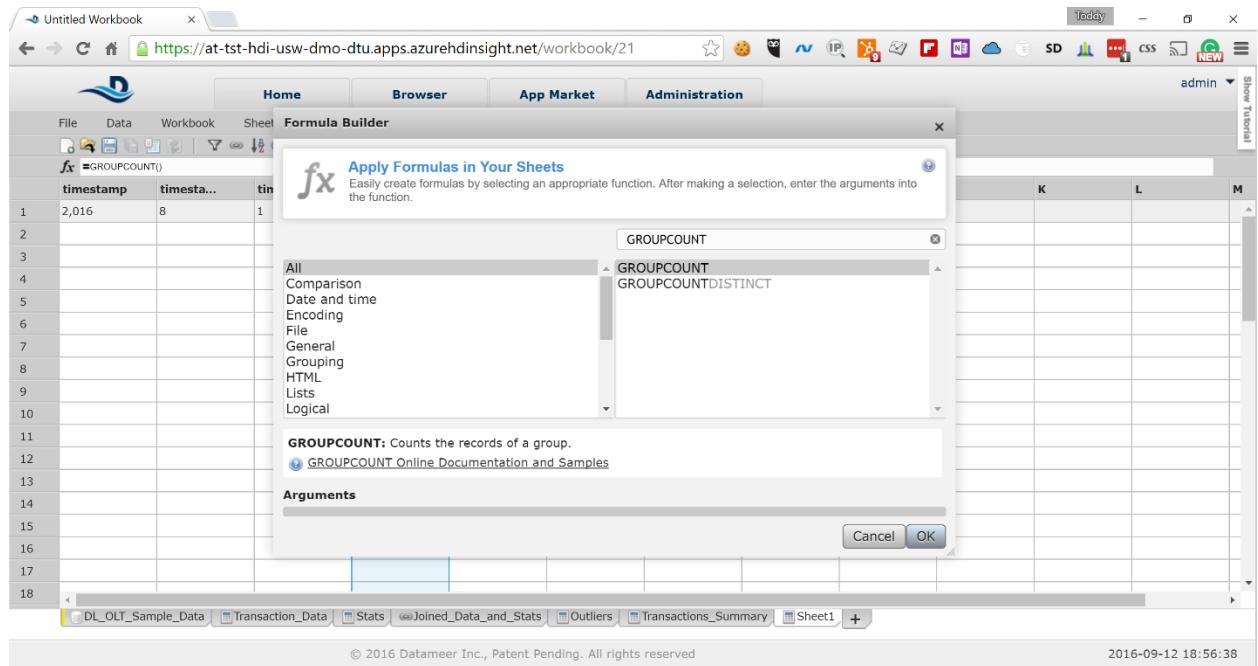
23. Click on the third column and in the formula pop-up select again the *GROUPBY* function and paste the following formula in the *Arguments* field:

*DAY(#Outliers!timestamp)*



The screenshot shows the DataMeer interface with a formula builder open. The formula bar at the top shows `=GROUPBY(DAY(#Outliers!timestamp))`. The 'Formula Builder' tab is selected. In the center, a dropdown menu for 'GROUPBY' is open, showing options like All, Comparison, Date and time, Encoding, File, General, Grouping, HTML, Lists, and Logical. Below the dropdown, a detailed description of the GROUPBY function is provided, along with a link to its online documentation. The 'Arguments' section contains the argument `DAY(#Outliers!timestamp)`. At the bottom right of the dialog are 'Cancel' and 'OK' buttons.

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

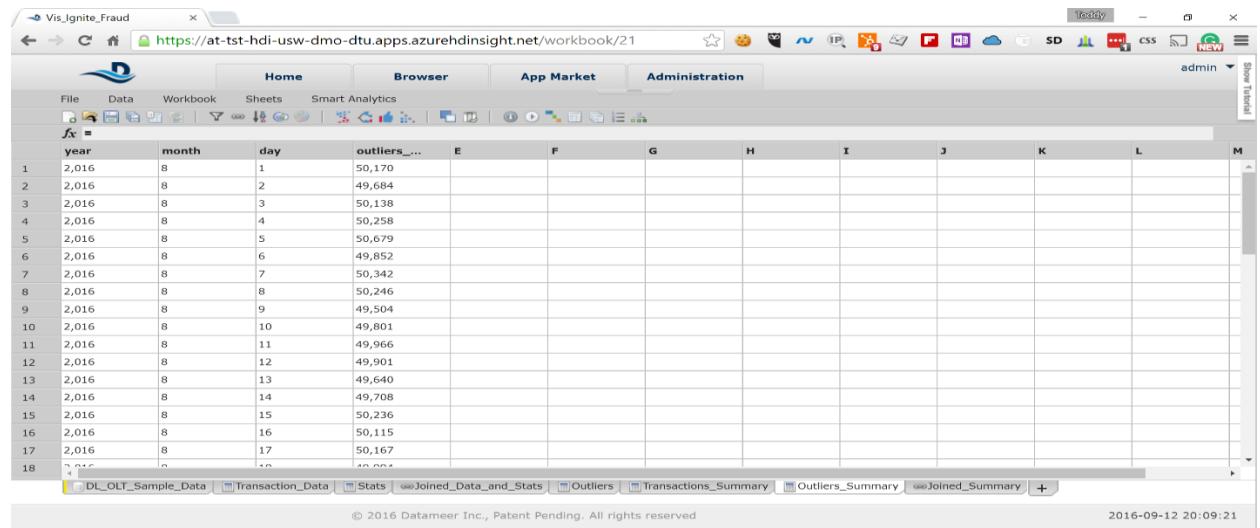


The screenshot shows the DataMeer interface with a formula builder open. The formula bar at the top shows `=GROUPCOUNT()`. The 'Formula Builder' tab is selected. In the center, a dropdown menu for 'GROUPCOUNT' is open, showing options like All, Comparison, Date and time, Encoding, File, General, Grouping, HTML, Lists, and Logical. Below the dropdown, a detailed description of the GROUPCOUNT function is provided, along with a link to its online documentation. The 'Arguments' section is empty. At the bottom right of the dialog are 'Cancel' and 'OK' buttons.

# HOL Guide for Enterprise Risk Analysis

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

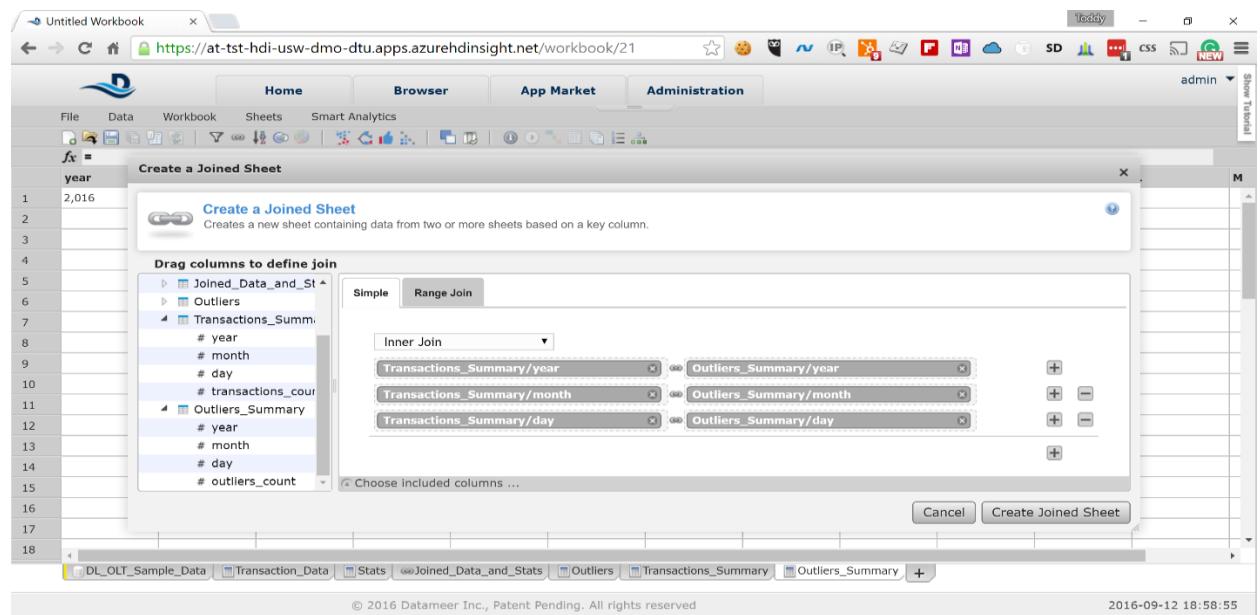
*year*  
*month*  
*day*  
*outliers\_count*



The screenshot shows a web-based data visualization tool. At the top, there's a navigation bar with tabs like Home, Browser, App Market, and Administration. Below the navigation is a toolbar with various icons. The main area displays a table with 18 rows of data. The columns are labeled: year, month, day, outliers\_count, E, F, G, H, I, J, K, L, M. The data shows daily transaction counts for August 2016. Row 18 contains a formula: =D18+C18+B18+A18.

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



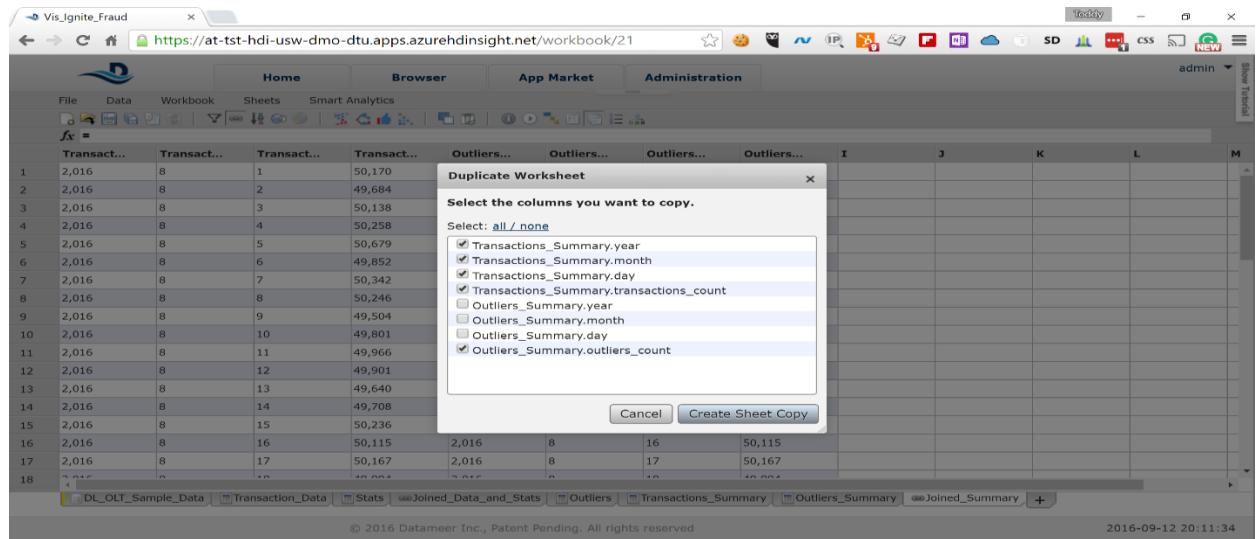
The screenshot shows the 'Create a Joined Sheet' dialog box. On the left, there's a sidebar with a tree view of available sheets: 'Joined\_Data\_and\_Stats', 'Outliers', and 'Transactions\_Summary'. Under 'Transactions\_Summary', columns 'year', 'month', 'day', and '# transactions\_count' are listed. Under 'Outliers\_Summary', columns 'year', 'month', 'day', and '# outliers\_count' are listed. The main area of the dialog has a 'Simple' tab selected, showing an 'Inner Join' setup where 'Transactions\_Summary/year' is joined with 'Outliers\_Summary/year', 'Transactions\_Summary/month' with 'Outliers\_Summary/month', and 'Transactions\_Summary/day' with 'Outliers\_Summary/day'. There are 'Cancel' and 'Create Joined Sheet' buttons at the bottom.

Rename the joined sheet to *Joined\_Summary*

# HOL Guide for Enterprise Risk Analysis

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:

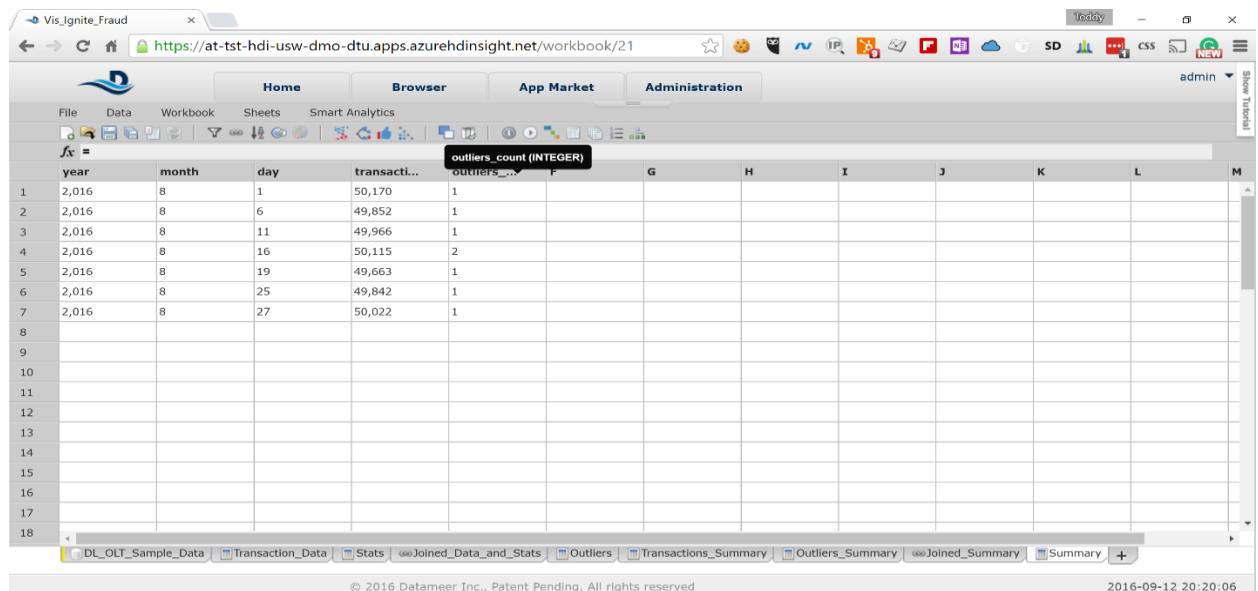
*Transactions\_Summary.year*  
*Transactions\_Summary.month*  
*Transactions\_Summary.day*  
*Transactions\_Summary.transactions\_count*  
*Outliers\_Summary.outliers\_count*



The screenshot shows the DataBlaze interface with a 'Duplicate Worksheet' dialog box open. The dialog box contains a list of columns selected for duplication. The columns listed are: *Transactions\_Summary.year*, *Transactions\_Summary.month*, *Transactions\_Summary.day*, *Transactions\_Summary.transactions\_count*, and *Outliers\_Summary.outliers\_count*. At the bottom of the dialog box are two buttons: 'Cancel' and 'Create Sheet Copy'. The background shows a spreadsheet with data and other tabs like 'DL\_OLT\_Sample\_Data', 'Transaction\_Data', etc.

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

*Transactions\_Summary.year* -> *year*  
*Transactions\_Summary.month* -> *month*  
*Transactions\_Summary.day* -> *day*  
*Transactions\_Summary.transactions\_count* -> *transactions\_count*  
*Outliers\_Summary.outliers\_count* -> *outliers\_count*



The screenshot shows the DataBlaze interface with a spreadsheet. The columns are labeled: year, month, day, transacti..., outliers\_count (INTEGER). The 'outliers\_count (INTEGER)' column is highlighted with a black border. The data rows show values such as 2,016, 8, 1, 50,170, 1; 2,016, 8, 6, 49,852, 1; 2,016, 8, 11, 49,966, 1; 2,016, 8, 16, 50,115, 2; 2,016, 8, 19, 49,663, 1; 2,016, 8, 25, 49,842, 1; 2,016, 8, 27, 50,022, 1. The bottom of the screen shows the footer with copyright information and a timestamp: '© 2016 Datameer Inc., Patent Pending. All rights reserved' and '2016-09-12 20:20:06'.

# HOL Guide for Enterprise Risk Analysis

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 a DataMeer workspace titled 'Vis\_Ignite\_Fraud'. A formula editor is active over the 'outliers\_to\_transactions' column, displaying the formula  $=\text{(#Summary!outliers\_count}/\#\text{Summary!transactions\_count}) * 100$ . The formula bar at the top of the sheet also contains this formula. The table has columns for year, month, day, transaction count, outliers count, and the calculated outliers\_to\_transactions value.

year	month	day	transaction count	outliers count	outliers_to_transactions
1	2,016	8	1	50,170	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
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					

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

The screenshot shows a 'Save Workbook' dialog box. The 'Name' field is populated with 'Vis\_Ignite\_Fraud'. The save location is specified as '/Analytics/Workbooks/Workbooks/Fraud\_POC'. At the bottom of the dialog are 'Create new Folder', 'Cancel', and 'Save' buttons. The background shows a portion of the DataMeer interface with a table and a formula editor.

# 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

The screenshot shows the 'Workbook Settings' page. At the top, there are tabs for Home, Browser, App Market, and Administration. The Administration tab is selected. On the left, there's a sidebar with a 'D' logo and navigation links for Home, Analytics, Workbooks, Fraud\_POC, Data, Connections, DataLinks, ExportJobs, FileUploads, ImportJobs, Images, Users, Visualization, and Infographics. The main content area has sections for 'Trigger' and 'Data Retention'. Under 'Trigger', 'Manual' is selected. Under 'Data Retention', 'Purge historical data' is selected, with 'Keep last 1 results' and 'Purge results 1 days after they're generated'. Other options like 'Never delete historical data' and 'Export only (Never keep data)' are also shown. Below these sections are buttons for 'Description' and 'Save Results and Time Based Partition'. At the bottom, a copyright notice reads '© 2016 Datameer Inc., Patent Pending. All rights reserved'.

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

The screenshot shows the 'File Browser' page. The top navigation bar includes Home, Browser, App Market, and Administration. The Administration tab is selected. The left sidebar contains the same navigation links as the previous screenshot. The main area displays a table of workbooks. One row is selected, showing 'Vis\_Ignite\_Fraud.wbk' with columns for Name, Type, Status, Last Processed, Records, Size, License total size, and Owner. A 'Run' button is located above the table. To the right, an 'Information' panel is open, showing details for the selected workbook: ID: 21, File ID: 53, Type: Workbook, Size: 0 Bytes, Last Processed: none, Created: Sep 12, 2016 5:51:31 PM, Last Changed: Sep 12, 2016 7:07:34 PM, and Scheduled: Manually. There's also a 'Description' section with a large text input field. The footer of the browser shows '1 Entry' and the copyright notice '© 2016 Datameer Inc., Patent Pending. All rights reserved'. The timestamp '2016-09-12 19:09:26' is also visible.

# HOL Guide for Enterprise Risk Analysis

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

The screenshot shows the DataMeer File Browser interface. The URL in the address bar is <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/browser>. The top navigation bar includes Home, Browser, App Market, and Administration tabs. The Administration tab is selected. On the left, a sidebar menu lists categories like admin, Analytics, Workbooks, Data, Images, Users, and Visualization. The main content area displays a table with one entry:

Name	Last Processed	Records	Size	License total size	Owner
Vis_Ignite_Fraud.wbk	none	.....	0 Bytes	---	admin

A progress bar at the top indicates "28% (100% Optimized MapReduce)". The right side of the screen features an "Information" panel for the selected file "Vis\_Ignite\_Fraud.wbk". The "General" section contains the following details:

- ID: 21
- File ID: 53
- Type: Workbook
- Size: 0 Bytes
- Last Processed: none
- Created: Sep 12, 2016 5:51:31 PM
- Last Changed: Sep 12, 2016 7:07:34 PM
- Scheduled: Manually

The "Description" section is empty. At the bottom of the browser window, a copyright notice reads "© 2016 Datameer Inc., Patent Pending. All rights reserved" and the timestamp is "2016-09-12 19:14:42".

## 8 Logging in to the TrendMicro DSM

### 8.1 Servername

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

The screenshot shows a Windows application window titled "Microsoft.Template Deployment". In the top bar, there are five buttons: Delete, Cancel, Refresh, Redeploy, and View template. Below the title bar, there is a section labeled "Outputs" containing two entries:

- TRENDMICRO DSM URI: <https://publicdnstrenddsrn5qxit2vx2jok.westus.cloudapp.azure.com> (with a copy icon)
- CHEF SERVER URI: <https://publicdnschefservercbm3slhj5vvc4.westus.cloudapp.azure.com> (with a copy icon)

### 8.2 Serverlogin

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

The screenshot shows a web browser window with the address bar displaying the URL: <https://publicdnstrenddsrn5qxit2vx2jok.westus.cloudapp.azure.com/SignIn.screen>. The page title is "Trend Micro Deep Security ...". The main content is a "Sign In" form with the following fields:

Username:	demo
Password:	*****
<input type="checkbox"/> I have an MFA token (More Info)	
<input type="button" value="Sign In"/>	

At the bottom of the page, there is a copyright notice: "Copyright © 2016 Trend Micro Inc. All rights reserved".

## 9 Perform policy configuration on the TrendMicro DSM

### 1. Changing the base policy

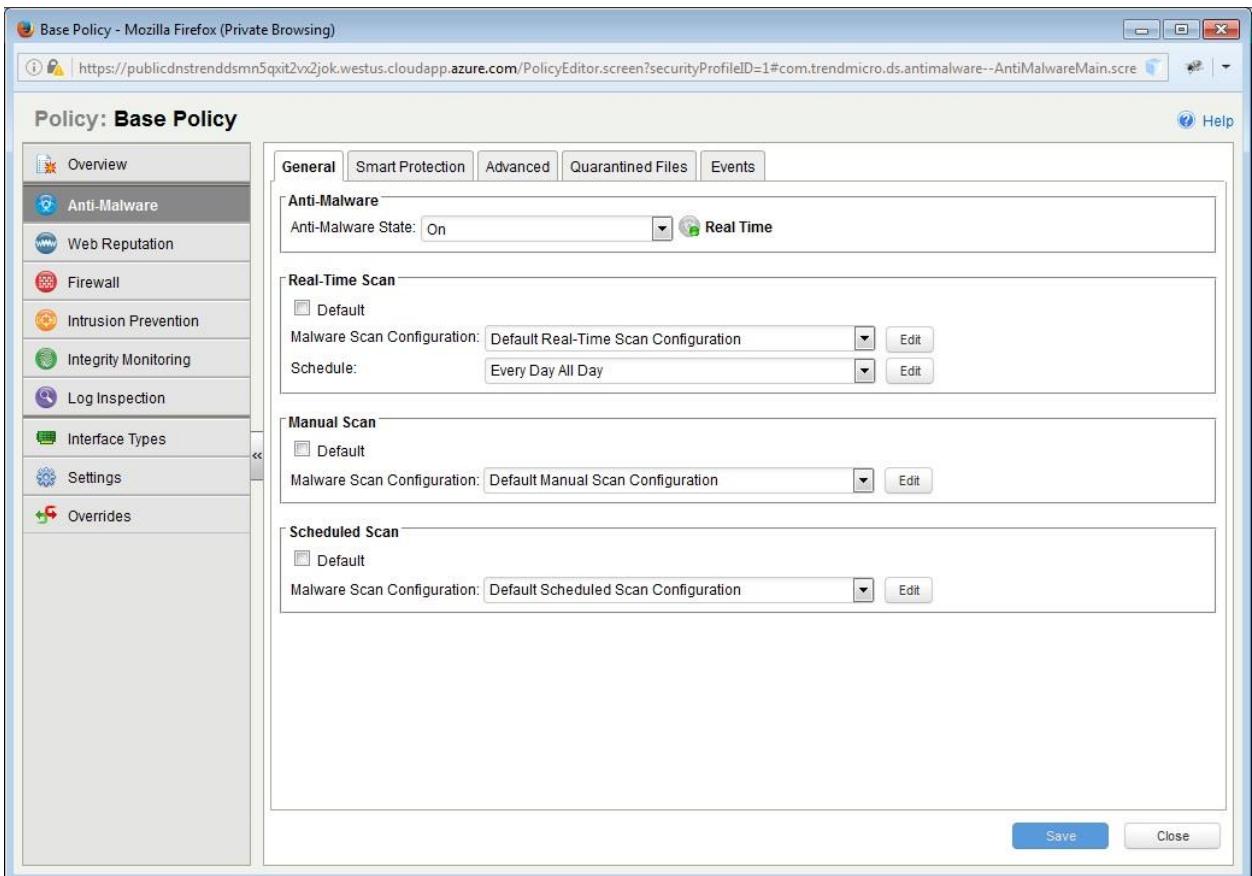
Go to policies->Base Policy



### 2. Enable Anti-Malware

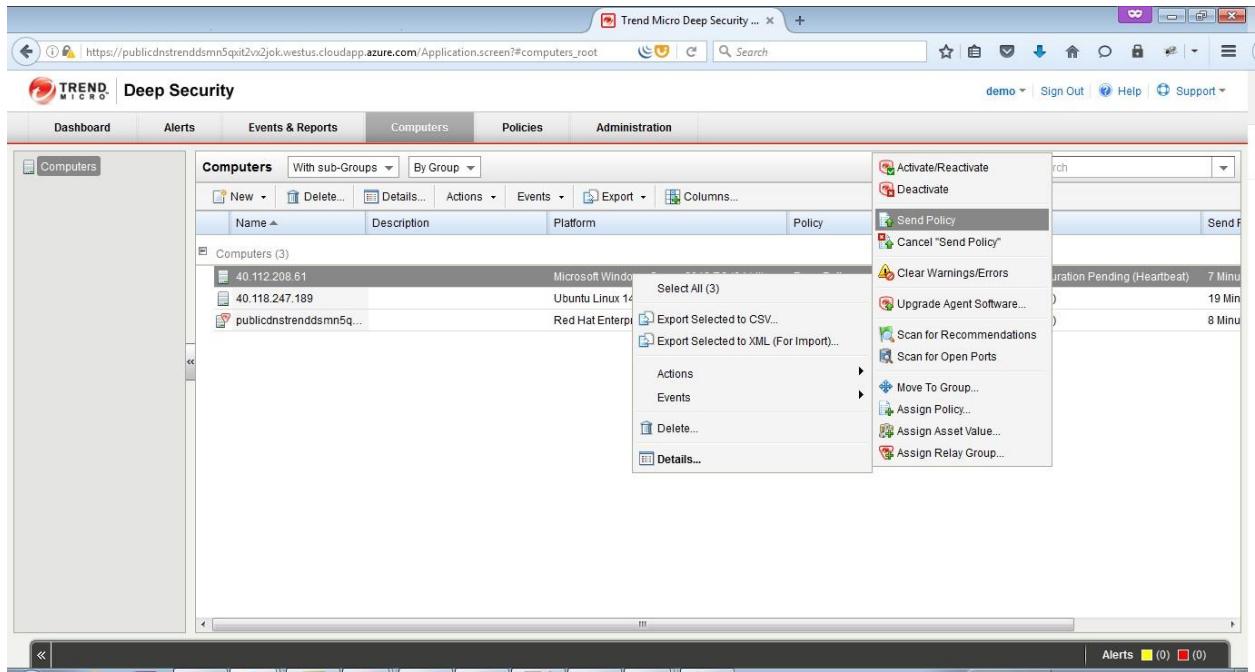
Go to Anti-malware->Anti-Malware State->On

Click "Save"

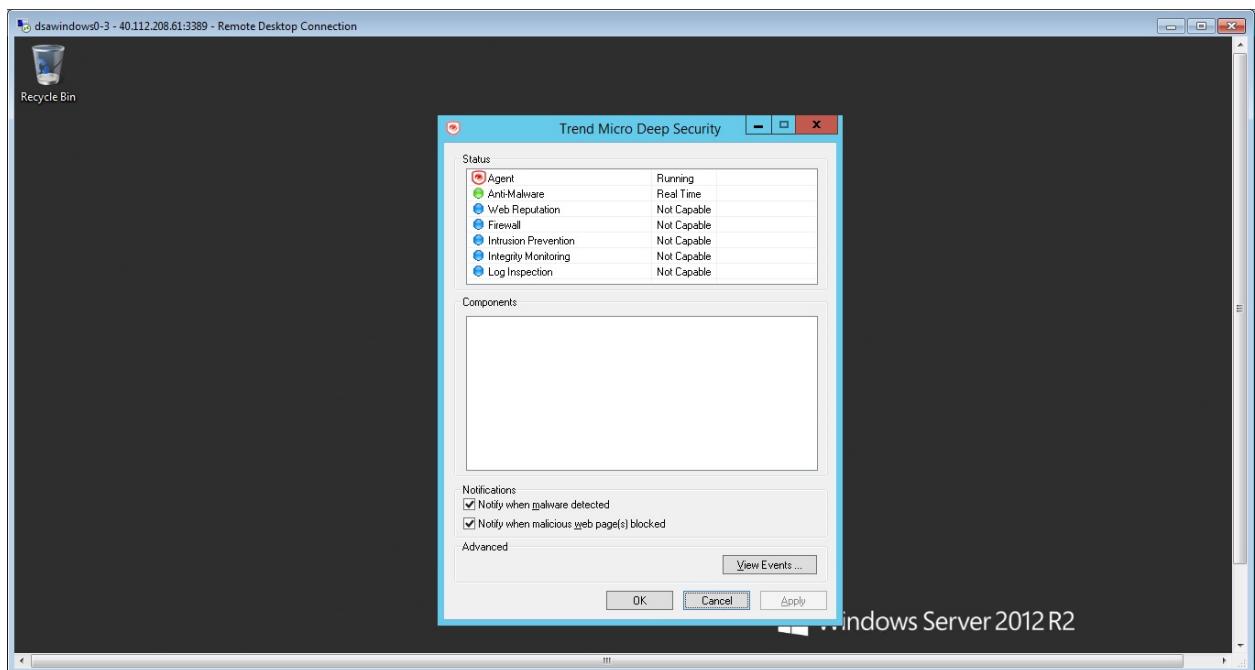


# HOL Guide for Enterprise Risk Analysis

## 3. Applying policies to computer



## 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.

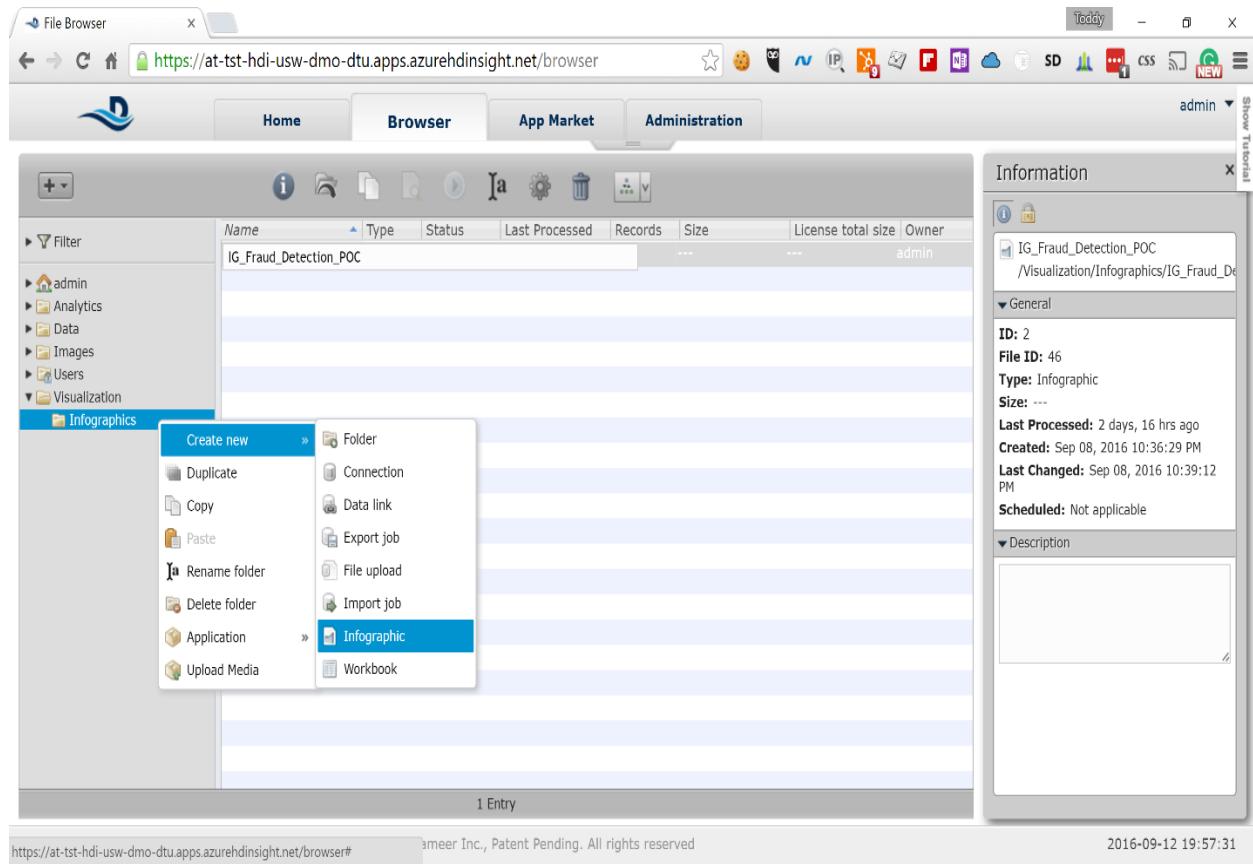
### 10.2 TrendMicro – Malware test

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

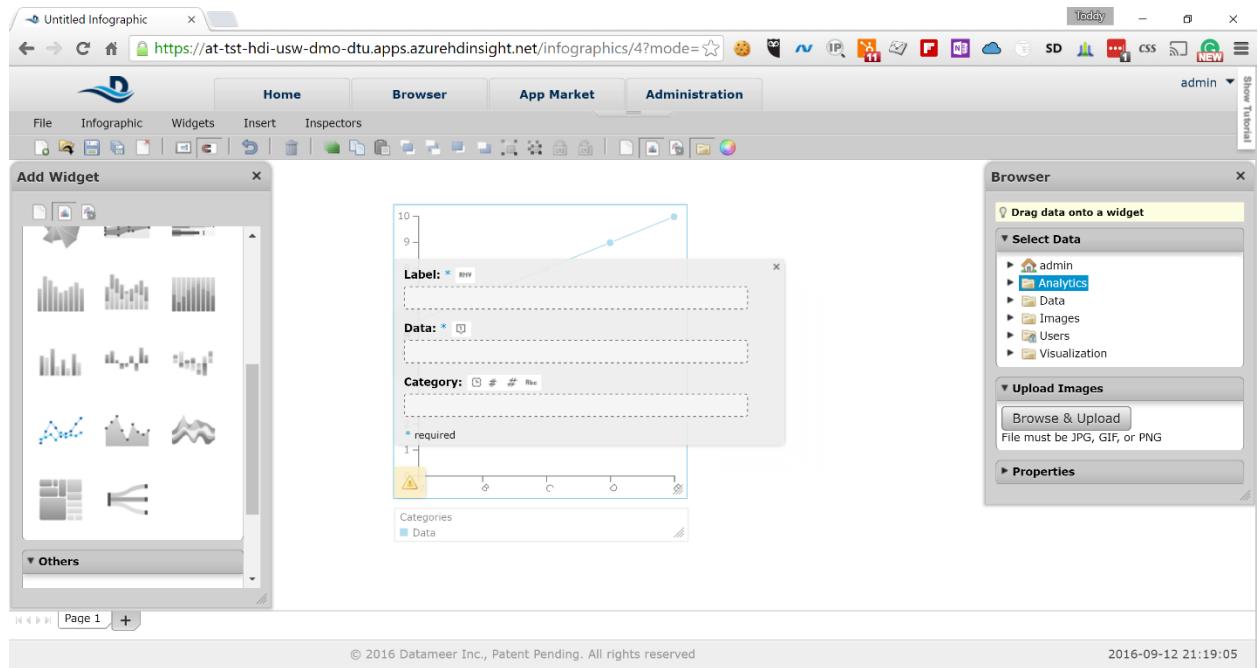
The last exercise in this HOL is to visualize the data and identify certain days when the irregular transactions have spiked. To do that we will use the following steps:

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

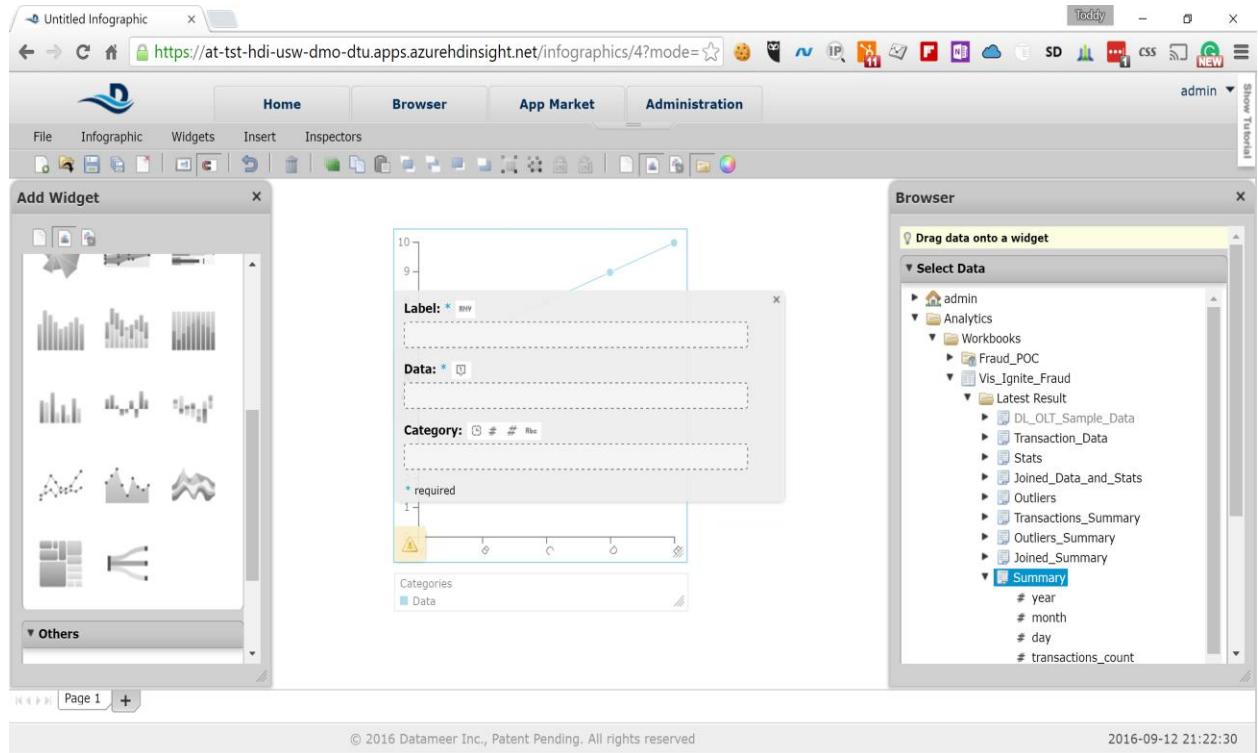


# HOL Guide for Enterprise Risk Analysis

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



3. In the Browser pane expand *Analytics* node and then *Workbooks* -> *Vis\_Ignite\_Fraud* -> *Latest Results* -> *Summary*



# HOL Guide for Enterprise Risk Analysis

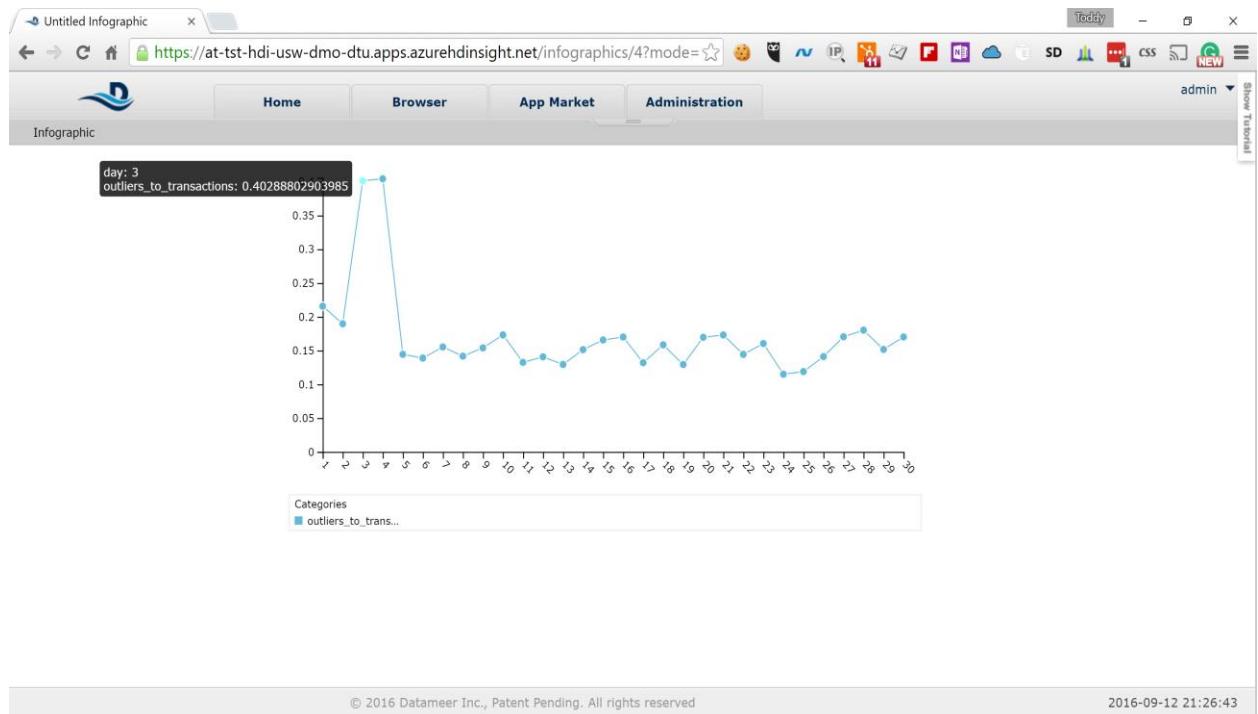
4. Drag the *day* field to the *Label* input field and the *outliers\_to\_transactions* field to the *Data* input field in the Work pane

The screenshot shows the Infographic builder interface with the 'Add Widget' dialog open. A line chart is being configured with the following settings:

- Label:** day
- Data:** outliers\_to\_transactions
- Category:** # day

The 'Browser' pane on the right displays a tree view of available data sources, including 'Vis\_Ignite\_Fraud' and its sub-folders like 'Latest Result' and 'Transactions\_Data'. The URL in the browser bar is <https://at-tst-hdi-usw-dmo-dtu.apps.azurehdinsight.net/infographics/4?mode=edit>.

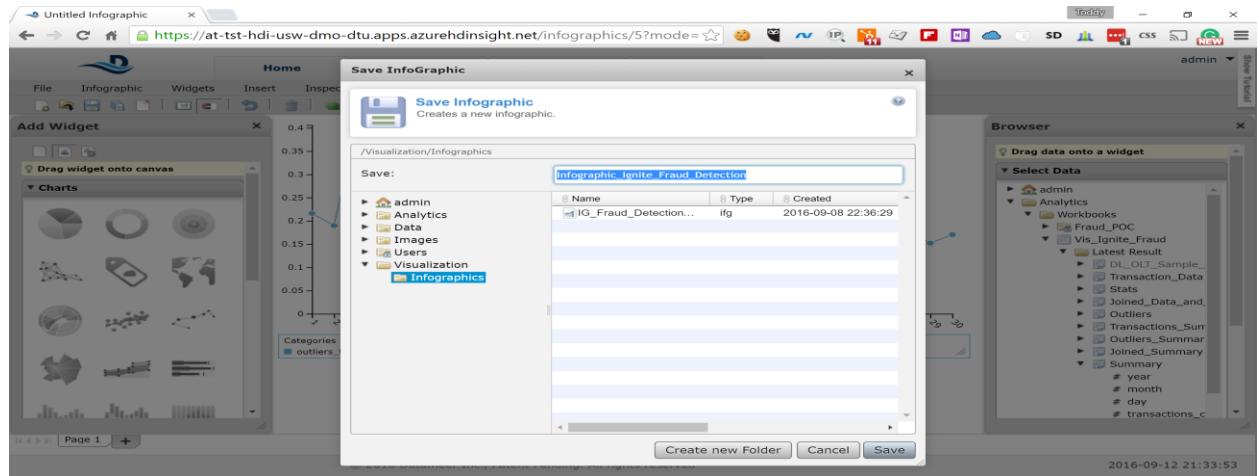
5. Select Infographic -> View from the menu to present the infographic. You can easily see that on the 3<sup>rd</sup> and 4<sup>th</sup> day of the month the outliers significantly spiked, which is a sign of something unusual going on those two days



# HOL Guide for Enterprise Risk Analysis

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



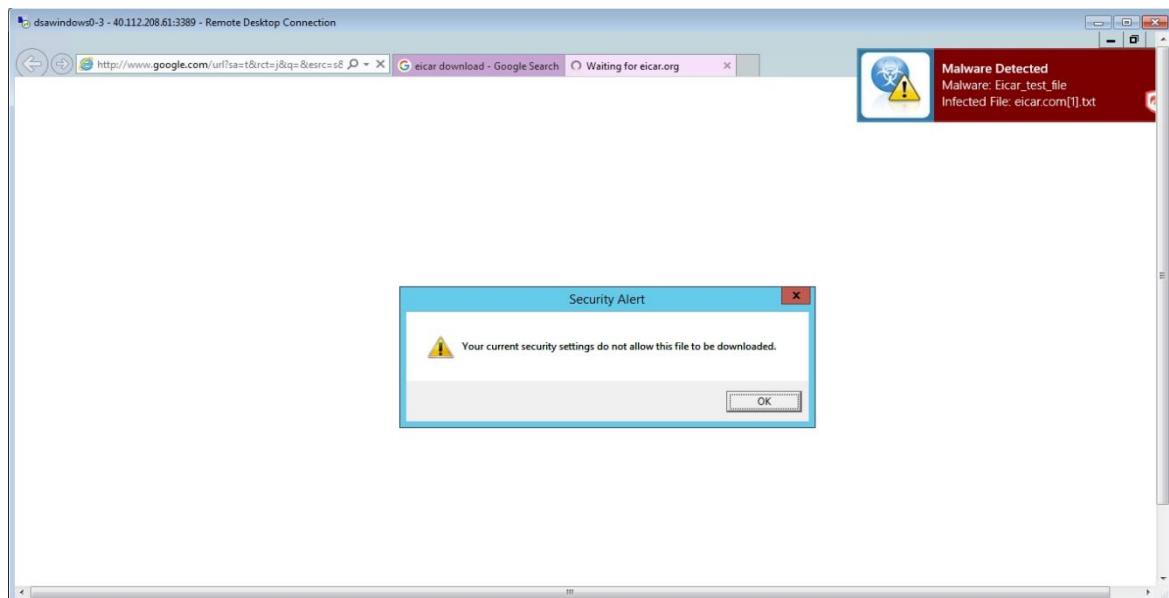
## 12 Malware Test

### 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



# HOL Guide for Enterprise Risk Analysis

## 12.2 Dashboard – Malware Alert

The screenshot shows the Trend Micro Deep Security dashboard with the following components:

- Alert Status:** Critical: 0, Warning: 1. Latest Alerts: Anti-Malware Alert - 40.112.2... 1 Minute.
- Computer Status:** Computer Status: Critical: 0, Warning: 0, Managed: 3, Unmanaged: 0.
- 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.
- My Sign-in History:** Last 1 Sign-in Attempts: September 5, 2016 22:32 Success.
- Anti-Malware Event History:** A chart showing Events vs Hour from 01:00 to 22:00. The count fluctuates between 0 and 2.
- Action Taken:** Legend: Cleaned (green), Quarantined (orange), Deleted (blue), Passed (yellow), Access Denied (grey), Uncleanable (red).
- Anti-Malware Status (Computers):** Top 5 Infected Computers: Computer name 40.112.208.61, Number of Uncleanable Total 0 (0%).
- Web Reputation Event History:**
- Web Reputation Computer Activity:**
- Web Reputation URL Activity:**

At the bottom right, there are alerts: (1) yellow, (0) red.

## 12.3 Malware Alert verification

The screenshot shows the Trend Micro Deep Security Events & Reports section with the following details:

**Anti-Malware Events**

- Period: Custom Range: From: September 5, 2016 00:00 To: September 6, 2016 00:00
- Computers: Computer: 40.112.208.61
- Table Headers: Time, Computer, Infected File(s), Tag(s), Malware, Action Taken
- Table Data:

Time	Computer	Infected File(s)	Tag(s)	Malware	Action Taken
September 5, 2016 22:51:23	40.112.208.61	C:\Users\demo\AppData\Local\Microsoft\Windows\NetCache\E8JDT4RUG\elcar.com[1].txt		Eicar_test_file	Deleted

At the bottom right, there are alerts: (1) yellow, (0) red.

## 13 References, Attachments & Definitions

### 13.1 References

No.	Document Title	Link/ Attachment	Comments
1			