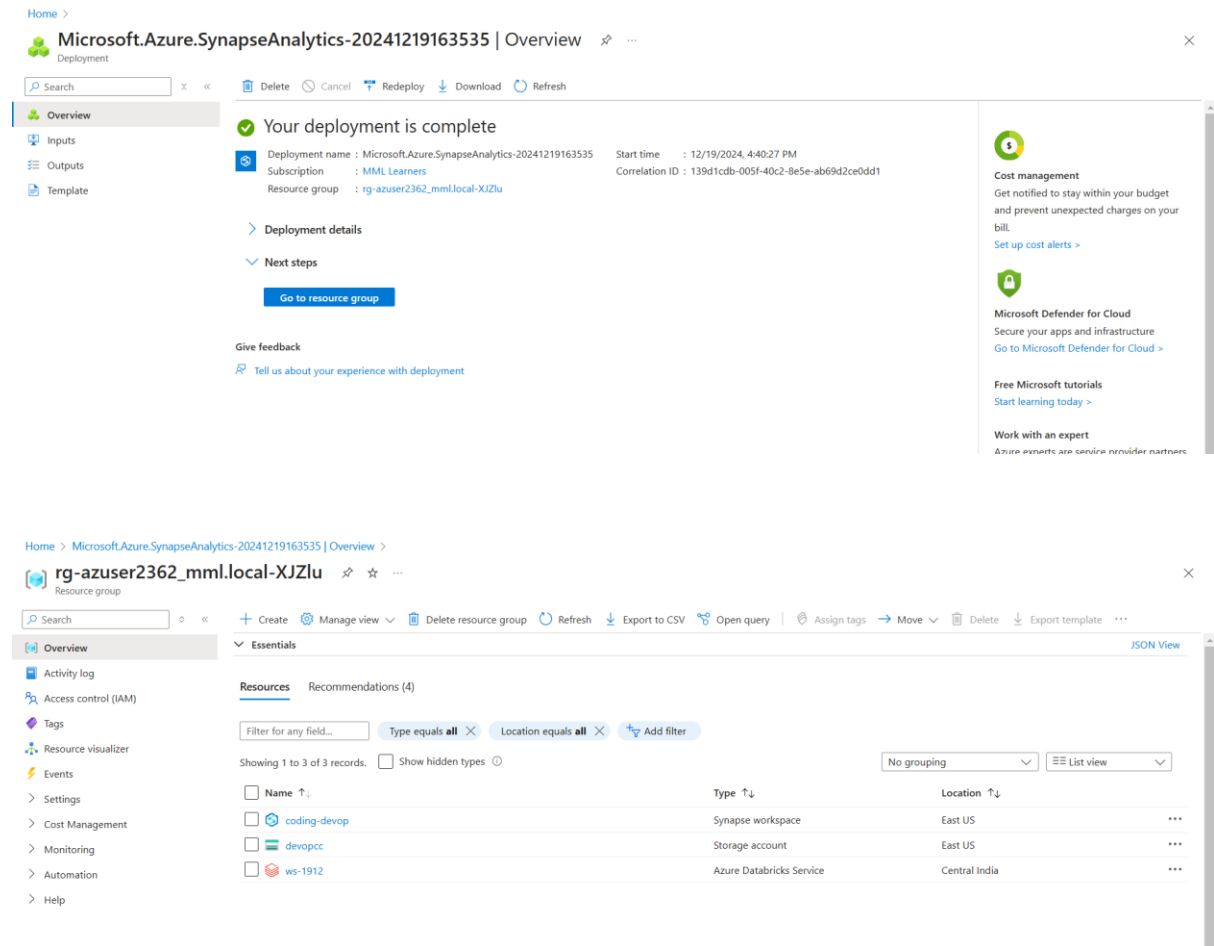


Build an Pipeline with azure synapse with dataflow running on it

Step1:Create azure synapse analytics workspace and add existing storage account and Databricks workspace



Home > Microsoft.Azure.SynapseAnalytics-20241219163535 | Overview

Deployment

Search

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20241219163535 Start time : 12/19/2024, 4:40:27 PM
Subscription : MML Learners Correlation ID : 139d1cdb-005f-40c2-8e5e-ab69d2ce0dd1
Resource group : rg-azuser2362_mml.local-XJZlu

> Deployment details

Next steps

Go to resource group

Give feedback

Tell us about your experience with deployment

Cost management

Get notified to stay within your budget and prevent unexpected charges on your bill.

Set up cost alerts >

Microsoft Defender for Cloud

Secure your apps and infrastructure

Go to Microsoft Defender for Cloud >

Free Microsoft tutorials

Start learning today >

Work with an expert

Azure experts are available on demand.

Home > Microsoft.Azure.SynapseAnalytics-20241219163535 | Overview

rg-azuser2362_mml.local-XJZlu

Resource group

Search

Overview

Activity log

Access control (IAM)

Tags

Resource visualizer

Events

Settings

Cost Management

Monitoring

Automation

Help

Essentials

Resources

Recommendations (4)

Filter for any field...

Type equals all

Location equals all

Add filter

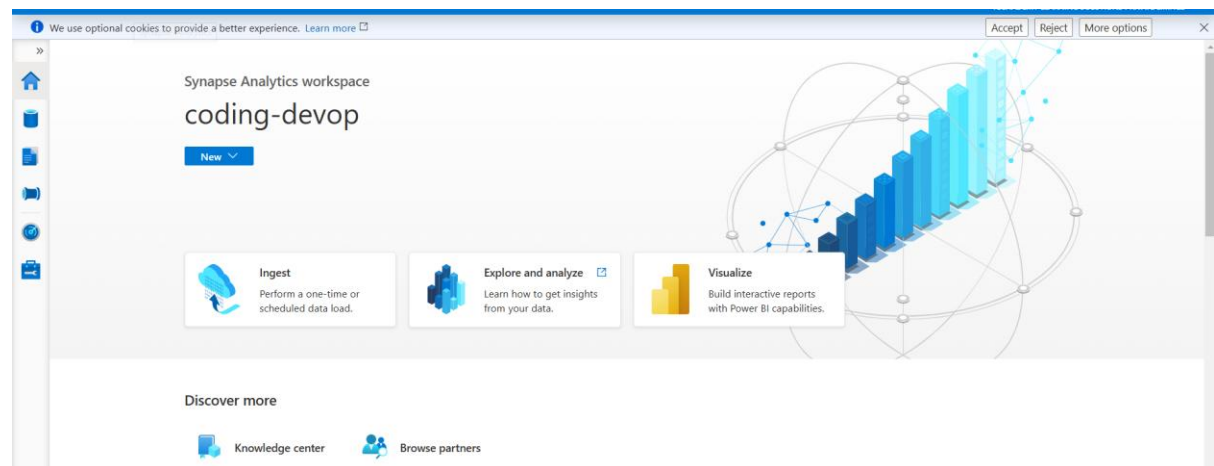
Showing 1 to 3 of 3 records. Show hidden types

No grouping

List view

Name	Type	Location
coding-devop	Synapse workspace	East US
devopcc	Storage account	East US
ws-1912	Azure Databricks Service	Central India

Step2:Go to synapse workspace



We use optional cookies to provide a better experience. Learn more

Accept Reject More options

Synapse Analytics workspace

coding-devop

New

Ingest

Perform a one-time or scheduled data load.

Explore and analyze

Learn how to get insights from your data.

Visualize

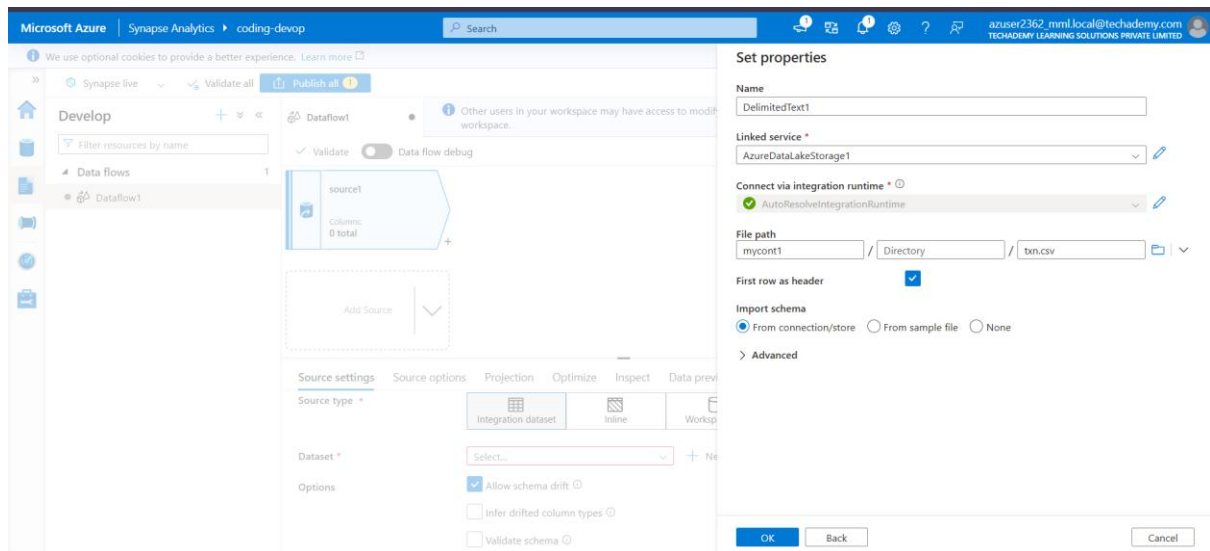
Build interactive reports with Power BI capabilities.

Discover more

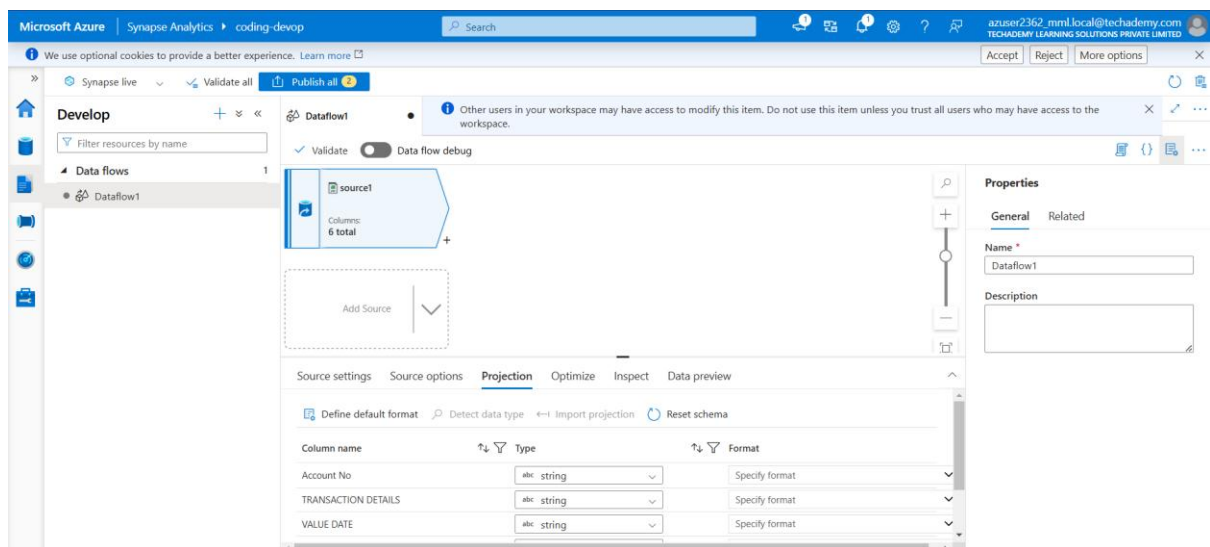
Knowledge center

Browse partners

Step3:Click on new dataflow and add source.Configure the existing storage account.



Step4:Turn on debug to detect datatype of records



Step5:Click on ‘+’ and select filter---→Filter records with Deposit amount >10000

The screenshot shows the 'Filter settings' tab for 'filter1' in the Dataflow1 Pipeline 1. The pipeline consists of the following steps: source1 (Import data from DelimitedText1), filter1 (Columns: 6 total), filter2 (Filtering rows using expressions on columns 'BALANCE AMT' > 1000000), sort1 (Sorting rows on columns 'DEPOSIT AMT'), and sink1 (Export data to DelimitedText2). The 'Filter settings' for 'filter1' are as follows:

- Output stream name:** filter1
- Description:** Filter records having Deposit amount greater than 10000
- Incoming stream:** source1
- Filter on:** { DEPOSIT AMT } > 10000

Step6:This can be verified in Data Preview

The screenshot shows the 'Data preview' tab for 'filter1' in the Dataflow1 Pipeline 1. The pipeline consists of the following steps: source1 (Import data from DelimitedText1), filter1 (Columns: 6 total), filter2 (Filtering rows using expressions on columns 'BALANCE AMT' > 1000000), sort1 (Sorting rows on columns 'DEPOSIT AMT'), and sink1 (Export data to DelimitedText2). The 'Data preview' for 'filter1' is as follows:

Account No	TRANSACTION ...	VALUE DATE	WITHDRAWAL ...	DEPOSIT AMT	BALANCE AMT
409000611074'	TRF FROM India...	29-Jun-17	NULL	1000000	1000000
409000611074'	TRF FROM India...	5-Jul-17	NULL	1000000	2000000
409000611074'	FDRL/INTERNAL...	18-Jul-17	NULL	500000	2500000
409000611074'	TRF FRM Indiafo...	1-Aug-17	NULL	3000000	5500000
409000611074'	FDRL/INTERNAL...	16-Aug-17	NULL	500000	6000000
409000611074'	FDRL/INTERNAL...	16-Aug-17	NULL	500000	6500000
409000611074'	FDRL/INTERNAL...	16-Aug-17	NULL	500000	7000000

Step7:Filter record with Balance amount >1000000

The screenshot shows the Databricks Dataflow interface. The pipeline consists of the following steps: **source1** (Import data from DelimitedText1), **filter1** (Filter records having Deposit amount greater than 10000), **filter2** (Filtering rows using expressions on columns 'BALANCE AMT' > 1000000), **sort1** (Sorting rows on columns 'DEPOSIT AMT'), and **sink1** (Export data to DelimitedText2). The **filter2** step is currently selected, and its configuration is shown in the bottom panel:

- Output stream name:** filter2
- Description:** Filtering rows using expressions on columns 'BALANCE AMT' > 1000000
- Incoming stream:** filter1
- Filter on:** (BALANCE AMT) > 1000000

Step8:Sort the records based on deposit amount in ascending order

The screenshot shows the Databricks Dataflow interface with the **sort1** step selected. The pipeline steps are: **source1**, **filter1**, **filter2**, **sort1**, and **sink1**. The configuration for the **sort1** step is shown in the bottom panel:

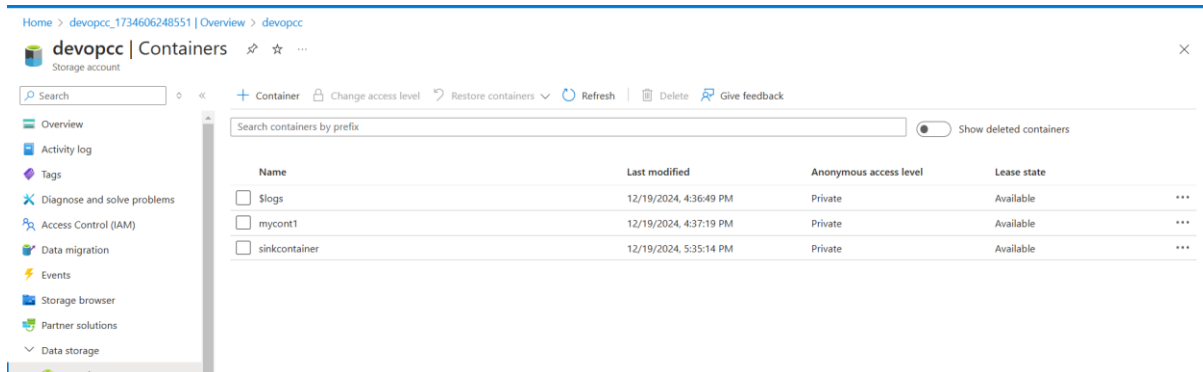
- Output stream name:** sort1
- Description:** Sorting rows on columns 'DEPOSIT AMT'
- Incoming stream:** filter2
- Options:** ☒ Case insensitive, ☐ Sort only within partition
- Sort conditions:**

filter2's column	Order	Nulls first
123 DEPOSIT AMT	Ascending	<input checked="" type="checkbox"/>

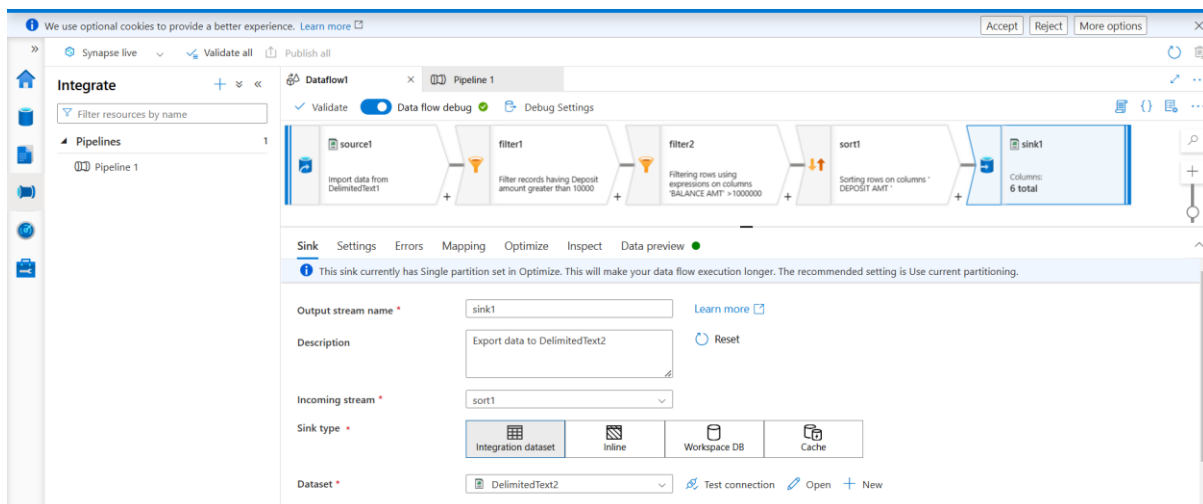
Step9:Dataflow snapshot

The screenshot shows the final Dataflow snapshot in the Databricks interface. The pipeline consists of the following steps: **source1** (Import data from DelimitedText1), **filter1** (Filter records having Deposit amount greater than 10000), **filter2** (Filtering rows using expressions on columns 'BALANCE AMT' > 1000000), **sort1** (Sorting rows on columns 'DEPOSIT AMT'), and **sink1** (Export data to DelimitedText2). The **sink1** step shows a summary: "Columns: 6 total".

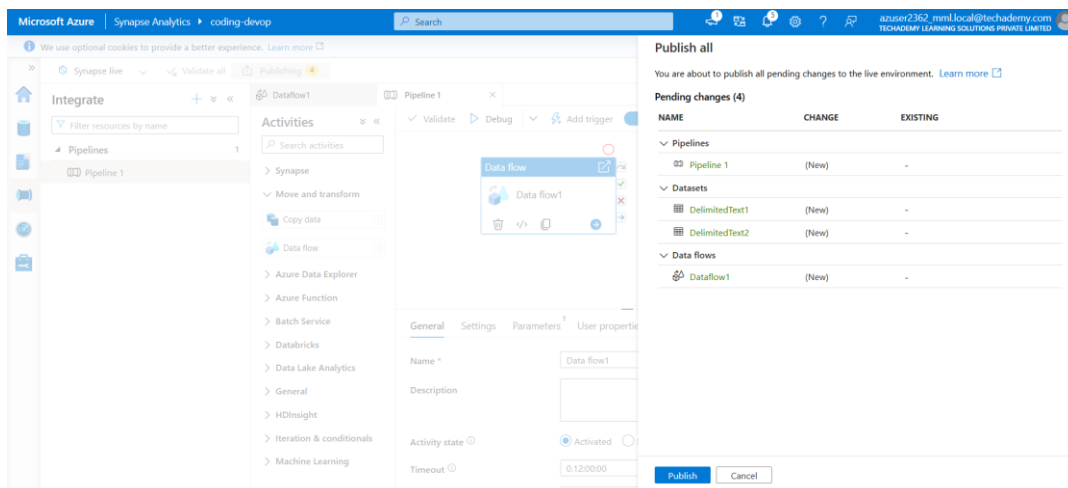
Step10:Create sink container to store transformed output



Step11:Configure the container in sink



Step12:Validate and publish all the results



Step13:Click on integrate and create a new pipeline having the data flow

The screenshot shows the Microsoft Azure Synapse Analytics interface. The 'Integrate' tab is selected, and 'Pipeline 1' is highlighted in the left sidebar. The main canvas displays a 'Data flow' activity named 'Data flow1'. The 'Pipeline status' is 'In progress'. The 'Output' tab is selected, showing a table with one row of data.

Activity name	Activity status	Activity type	Run start
Data flow1	Queued	Data flow	12/19/2024, 5:38:57 PM

Step14:Monitor the pipeline status in the pipeline itself

The screenshot shows the Microsoft Azure Synapse Analytics interface. The 'Integrate' tab is selected, and 'Pipeline 1' is highlighted in the left sidebar. The main canvas displays a 'Data flow' activity named 'Data flow1'. The 'Pipeline status' is 'Succeeded'. The 'Output' tab is selected, showing a table with one row of data.

Activity name	Activity status	Activity type	Run start
Data flow1	Succeeded	Data flow	12/19/2024, 5:38:57 PM

Step15:After the pipeline showed succeeded check the sink container for the transformed data

Microsoft Azure

Search resources, services, and docs (G+)

Capilot

Home > devopcc_1734606248551 | Overview > devopcc | Containers >

sinkcontainer Container

Search

Upload + Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: sinkcontainer

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
part-00000-34145bed-56b4-4cc6-80c0-b63b2be2...	12/19/2024, 5:39:11 ...	Hot (Inferred)		Block blob	40.36 KiB	Available

Step16:The transformed data can be downloaded as csvfile

B2

TRF FROM Indiaforensic SERVICES

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Account No	TRANSACTION DETAILS	VALUE DATE	WITHD	DEPOS	BALANCE	AMT						
327	409000611074	Indfor INCM INDO REVREMI0303	20-Mar-18		1320	1202734							
328	409000611074	INDO GIBLOXGN REVSTL03031	20-Mar-18		155500	1358234							
329	409000611074	Indfor INCM INDO REVREMI0103	20-Mar-18		202799	1561033							
526	409000611074	FDRL/INTERNAL FUND TRANSFE	13-Jul-18		200000	1629620							
533	409000611074	FDRL/INTERNAL FUND TRANSFE	17-Jul-18		200000	1129020							
537	409000611074	FDRL/INTERNAL FUND TRANSFE	17-Jul-18		200000	1107940							
538	409000611074	FDRL/INTERNAL FUND TRANSFE	17-Jul-18		200000	1307940							
541	409000611074	FDRL/INTERNAL FUND TRANSFE	18-Jul-18		200000	922080							
542	409000611074	FDRL/INTERNAL FUND TRANSFE	18-Jul-18		200000	1122080							
543	409000611074	FDRL/INTERNAL FUND TRANSFE	18-Jul-18		200000	1322080							
544	409000611074	FDRL/INTERNAL FUND TRANSFE	18-Jul-18		200000	1522080							
545	409000611074	FDRL/INTERNAL FUND TRANSFE	19-Jul-18		200000	1722080							
546	409000611074	FDRL/INTERNAL FUND TRANSFE	19-Jul-18		200000	1922080							
783	409000611074	FDRL/INTERNAL FUND TRANSFE	17-Oct-18		200000	844356							
785	409000611074	FDRL/INTERNAL FUND TRANSFE	17-Oct-18		200000	1344356							
803	409000611074	FDRL/INTERNAL FUND TRANSFE	23-Oct-18		200000	1163908							
811	409000611074	FDRL/INTERNAL FUND TRANSFE	29-Oct-18		200000	1037994							
812	409000611074	FDRL/INTERNAL FUND TRANSFE	29-Oct-18		200000	1237994							
813	409000611074	FDRL/INTERNAL FUND TRANSFE	29-Oct-18		200000	1437994							
820	409000611074	FDRL/INTERNAL FUND TRANSFE	30-Oct-18		200000	788649							
824	409000611074	FDRL/INTERNAL FUND TRANSFE	31-Oct-18		200000	1011527							
850	409000611074	FDRL/INTERNAL FUND TRANSFE	12-Nov-18		200000	1392684							
851	409000611074	FDRL/INTERNAL FUND TRANSFE	12-Nov-18		200000	1592684							