# Extracting Data and Creating Dataset for Purchase Event Dashboards

The purpose of this document is to provide the details on the dashboard for purchase metrics.

*Scenarios:*

* Purchase Event General Stats
* Purchase Event PII Data Leakage
* Purchase Event Corrupted Values
* Purchase Event Duplicate Metric

*Event Used:*

* confirmationpagedetailscleanevent

*Metrics:*

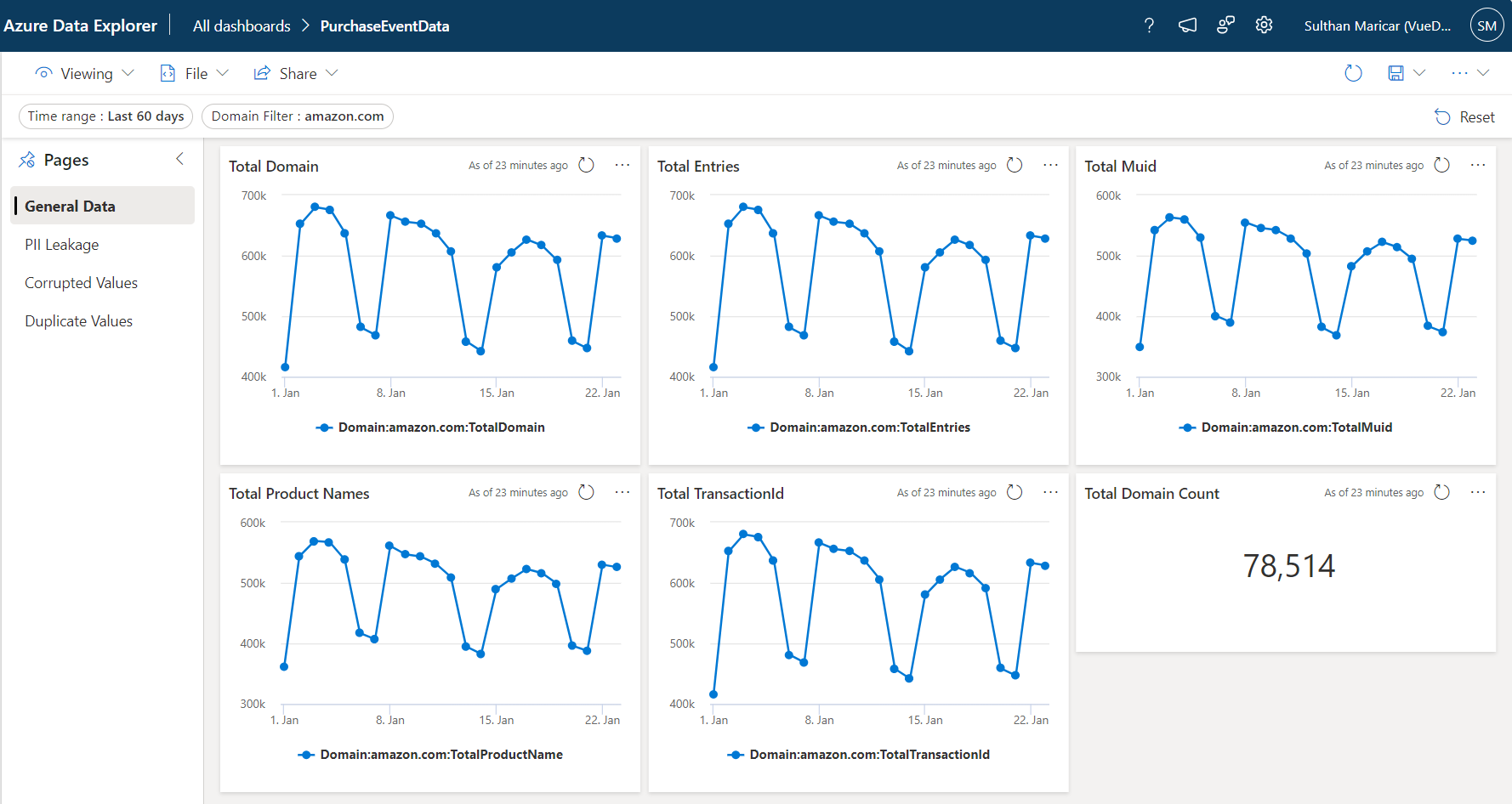
*General Stats:*

1. *Total Domain*
2. *Total Entries*
3. *Total MUID*
4. *Total Product Names*
5. *Total TransactionId*

Dashboard Charts:

*Dashboard daily view*

General Stats Dashboard – [PurchaseEventData (azure.com)](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=60days&p-_endTime=now&p-_domain=v-amazon.com#756a43f4-0388-406c-9865-7480aaa74329)

**

*PII Data Leakage:*

1. *Email present in TransactionId*
2. *Customer name present in TransactionId*
3. *Account number present in TransactionId*
4. *Address present in TransactionId*
5. *Phone number present in TransactionId*
6. *Email present in Product Name*
7. *Customer name present in Product Name*
8. *Account number present in Product Name*
9. *Address present in Product Name*
10. *Phone number present in Product Name*
11. *Email present in Purchase Total*
12. *Customer name present in Purchase Total*
13. *Account number present in Purchase Total*
14. *Address present in Purchase Total*
15. *Phone number present in Purchase Total*

*Dashboard daily view*

PII Data Leakage Dashboard – [PurchaseEventData (azure.com)](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=60days&p-_endTime=now#9022e5d8-9ed1-45cb-affd-8d360b99d9b9)

*A screenshot of a computer

Description automatically generated*

*Corrupted Values:*

1. *$ with no value in Purchase Total*
2. *Empty Purchase Total*
3. *0$ Purchase Total*
4. *R Count in Purchase Total*
5. *Garbled TransactionId*
6. *Invalid TransactionId*
7. *Invalid Product Names*
8. *Invalid Text PII data in Product Name*
9. *Price Extracted in Product Name*
10. *Empty TransactionId*
11. *Empty Product Names*
12. *Empty TransactionId & Product Name*

*Dashboard daily view*

Corrupted Values – [PurchaseEventData (azure.com)](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=60days&p-_endTime=now#0791547e-1e2b-40bc-971f-ab24fcb30de6)

*A screenshot of a computer screen

Description automatically generated*

*Duplicate Values:*

1. *Duplicate TransactionId*
2. *Duplicate TransactionId & Product Names*

*Dashboard daily view*

Duplicate Values – [PurchaseEventData (azure.com)](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=60days&p-_endTime=now&p-_domain=v-amazon.com#1bfd83f6-3eac-4cac-b80c-7f01da7de580)

A screenshot of a computer

Description automatically generated

*Mandatory Fields:*

1. *Domain Name Empty*
2. *MUID Empty*
3. *Product Name Empty*
4. *Purchase Total Empty*

File Location:

Cosmos path shared location:

*Cluster: cosmos08-prod-co3c*

*Virtual Cluster: ugc-prod*

*Purchase Event General Stats:*  
Path: [cosmos08 ugc-prod cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/GeneralData/TSV/2024/01/ (osdinfra.net)](https://www.cosmos08.osdinfra.net/cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/GeneralData/TSV/2024/01/)

*Path Structure*:

**Structure:** Local / Cashback / Metrics / <ScenarioName> / <Sub Scenario Name> / <File Format> / <Year> / <Month> / FileName\_YYYY\_MM\_DD.tsv  
  
**Example:** Local/ Cashback / Metrics / PurchaseEventData/ GeneralData / TSV / 2024 / 01 / general\_data\_2024\_01\_01.tsv  
  
**Safety:** Local/ Cashback / Metrics / PurchaseEventData/ GeneralData / TSV / 2024 / 01 / latest.tsv  
  
  
*General Stats Script link:*

[cosmos08 ugc-prod cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/GeneralData/Scripts/ (osdinfra.net)](https://www.cosmos08.osdinfra.net/cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/GeneralData/Scripts/)  
  
  
  
*Purchase Event Duplicate Metric:*Path: [cosmos08 ugc-prod cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/DuplicateData/TSV/2024/01/ (osdinfra.net)](https://www.cosmos08.osdinfra.net/cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/DuplicateData/TSV/2024/01/)  
*Path Structure:***Structure:** Local / Cashback / Metrics / <ScenarioName> / <Sub Scenario Name> / <File Format> / <Year> / <Month> / FileName\_YYYY\_MM\_DD.tsv  
  
**Example:** Local/ Cashback / Metrics / PurchaseEventData/ DuplicateData/ TSV / 2024 / 01 / purchaseevent\_duplicate\_data\_2024\_01\_01.tsv

*Duplicate Metric Script link:*

[cosmos08 ugc-prod cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/DuplicateData/Script/ (osdinfra.net)](https://www.cosmos08.osdinfra.net/cosmos/ugc-prod/local/Cashback/Metrics/PurchaseEventData/DuplicateData/Script/)

  
  
*Dashboard Linking:*

File Input:

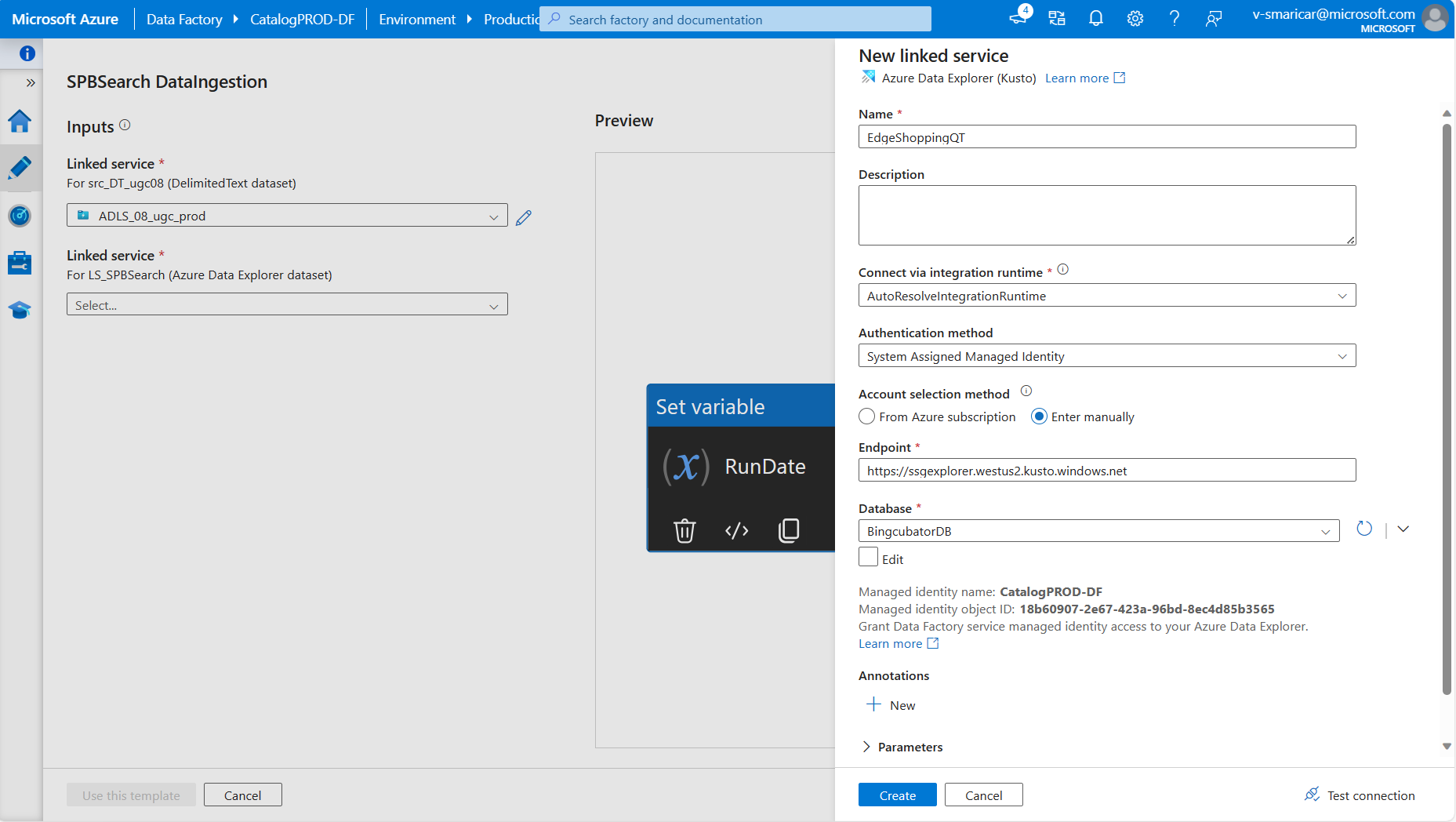
* Getting input from the cluster cosmos08 and “ugc-prod-c08” VC.
* In cosmos path, Data are in “log\_bucket” format.

Conversion:  
  
Attached the scope scripts for Purchase Event, It extracts the data from cosmos path log\_bucket file into expected result in TSV format.

* Using this scope script, we can handle all the logical calculations and convert the data into required formats like TSV or structured stream.
* Once the logic part is done, we can store it in one cosmos path.

**Ingestion to Kusto: Using Azure Data Factory**

Create a pipeline of a new linked service in ADF using the below image. You can select your desired database.



In Source, we need to update the Wildcard path, where our scripts are located and update the file format like "daubydomainupdated\_\*.tsv" it will ingest the data into Kusto.

A screenshot of a computer

Description automatically generated

Once you click on the Wildcard path, here you can update the script path and click ok.

A screenshot of a computer

Description automatically generated

In Sink tab, you can update your Kusto table, where you want to ingest the data.

A screenshot of a computer

Description automatically generated

If you want to change your table, Click on “Open”.  
  
A screenshot of a computer

Description automatically generated

* Uncheck “Enter manually”, Select your table then again select checkbox.
* In Kusto, first you need to create a table with the respective column in script.  
    
  A screenshot of a computer

  Description automatically generated
* You need to verify your connection first, so select “Test connection”.
* If the connection was successful, Select Schema tab.  
    
  A screenshot of a computer

  Description automatically generated
* First you need to Clear the existing table column first, then Click on Import schema.
* Now you can see the selected table columns.
* Then click on this previous tab and now select Mapping tab.  
    
    
    
    
    
    
  A screenshot of a computer

  Description automatically generated
* In Mapping, you need to “Clear” the existing mapping columns. Then select “Import schemas” to mapping the latest column.
* You need to verify that both “Source” & “Destination” looks same, without any column or type changes.
* When you select the “Import schemas” it asks you to enter pipeline run date, Enter the run date for the script which is available on this date.  
    
  A screenshot of a computer

  Description automatically generated
* Adding scope script into the pipeline, In “Activities” you search “scope” you get the scope icon in “Data Lake Analytics”.
* Drag and drop the scope and hover the “Scope” you can able to see tick icon on right side.
* Drag and drop the tick icon into Set variable tab to connect into it.  
    
  A screenshot of a computer

  Description automatically generated
* In the “Script” tab, you can add your script path and script name. Test connection once you added the script.  
    
  A screenshot of a computer

  Description automatically generated
* In “ADLA Account”, Select your VC where you script file is located.  
    
  A screenshot of a computer

  Description automatically generated
* Then click on “Debug”, Again it asks you for Pipeline run date. You need to give the run date for which monthly date files you want to ingest.
* After that you can see the ingestion process below. Once the ingestion process was completed. You can check the records on the Kusto table.  
    
  A screenshot of a computer

  Description automatically generated

Dashboard:

* Once we have the data for all the datasets, we can start creating the dashboard and charts.

How to fix the errors in pipeline failure:  
  
A screenshot of a computer

Description automatically generated  
A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated  
  
Follow the steps to fix the error:

1. You need to close the edge browser.
2. Disconnect the MSFT-VPN then again Reconnect the MSFT-VPN.
3. Open command prompt in administrator mode
4. Run the command: “nslookup management.azure.com”.
5. Run the command: “ipconfig”
6. Run the command: “route -p add 40.78.196.33 mask 255.255.255.255 <<your ip>>”

A screenshot of a computer

Description automatically generated

1. You can get <<your ip>> from MSFT-VPN address in command prompt.
2. Still if you face the error, Run the command: “route -p delete 40.78.196.33”
3. Run the command: route -p add 4.150.240.10 mask 255.255.255.255 <<your ip >>
4. Open the edge browser and Redirect into [Azure Data Factory](https://ms-adf.azure.com/en/authoring/pipeline/QT%20DataIngestion?factory=%2Fsubscriptions%2Ff3f7bc33-086d-473a-ad71-cca05dc8dfb8%2FresourceGroups%2FcatalogINT%2Fproviders%2FMicrosoft.DataFactory%2Ffactories%2FCatalogPROD-DF)

How to fix this mapping issue:  
  
A screenshot of a computer

Description automatically generated  
  
Need to check the “Source” column type and “Sink” column type, If any of the column type changed.   
  
Example: timestamp as string in “Source” & datetime in “Sink”. You got mapping issue.  
  
If you’re try to map with different dataset as well, You got the mapping issue.  
  
Example: “Source” pointed to spb\_pdp dataset but in “Sink” you’re selected spb\_search dataset then Mapping will be failed.

Reference Link:

* Cluster Name: ssgexplorer
* Cluster URL: [https://ssgexplorer.westus2.kusto.windows.net](https://nam06.safelinks.protection.outlook.com/?url=https%3A%2F%2Fssgexplorer.westus2.kusto.windows.net%2F&data=05%7C01%7Cv-smaricar%40microsoft.com%7Cce2c56954bb24457575708dbe593f96f%7C72f988bf86f141af91ab2d7cd011db47%7C1%7C0%7C638356195977282446%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=RiVUBUe2fFOhbPFuREe%2B3DQl%2FJ%2F0kvpB6Bv%2FbsvMA8w%3D&reserved=0)
* Event Name: confirmationpagedetailscleanevent
* Event URL: [ConfirmationPageDetailsCleanEvent](cosmos08%20ugc-prod%20cosmos/ugc-prod/shares/ugc-prod/AriaLogs/EdgeShopping/prd/confirmationpagedetailscleanevent/2024/01/%20(osdinfra.net))
* Dashboard Name: General Stats
* Dashboard URL: [PurchaseEventGeneralStats](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=30days&p-_endTime=now&p-_domain=v-amazon.com#756a43f4-0388-406c-9865-7480aaa74329)
* Dashboard Name: PII Leakage
* Dashboard URL: [PurchaseEventPIILeakage](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=30days&p-_endTime=now#9022e5d8-9ed1-45cb-affd-8d360b99d9b9)
* Dashboard Name: Corrupted Values
* Dashboard URL: [PurchaseEventCorruptedValues](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=30days&p-_endTime=now#0791547e-1e2b-40bc-971f-ab24fcb30de6)
* Dashboard Name: Duplicate Values
* Dashboard URL: [PurchaseEventDuplicateValues](https://dataexplorer.azure.com/dashboards/e9efb5e1-e9f7-45de-bbf8-3d098a193661?p-_startTime=30days&p-_endTime=now#1bfd83f6-3eac-4cac-b80c-7f01da7de580)
* Dashboard Name: Mandatory Fields
* Dashboard URL:
* Dashboard List: Azure Data Explorer All Dashboard
* Dashboard URL: [All Dashboards](https://dataexplorer.azure.com/dashboards/)