

Contents

Group 4: Members.....	1
Problem Statement Retail 360	1
Task 1.2: Data Ingestion using Azure Data Factory	3
1.2.1 . Raw data import to ADLS Gen2 storage.....	3
1.2.2 Staging ADLS Gen2 storage	5
1.2.3 Data Ingestion using Azure Data Factory from Raw to Staging storage.....	6
Creation of linked services	7
Creation of Data Sets.....	9
Creation of Data Flows	10
Configuring & Executing the pipeline.....	16
Staging data processed through ADF	19
Task 1.3: Azure SQL and Clean Data Loading using ADF.....	20
Creation of Target SQL server & database	20
Creation of Tables	22
ADF for Staging to Azure SQL Server	24
SQL Operations Post data loading.....	44
Data Visualization using Power BI.....	47

Group 4: Members

1	Harveen Kaur	harveenkaur@bharatpetroleum.in
2	Sridhar K	sridhark002@bharatpetroleum.in
3	Bidhan Sarkar	bidhansarkar@bharatpetroleum.in
4	Pooja Chaudhary	poojachaudhary@bharatpetroleum.in
5	Mahendra Chandurkar	chandurkarm@bharatpetroleum.in

Problem Statement Retail 360

Go-cart, an ecommerce ordering, and delivery application, strives to offer customers a delightful shopping journey by anticipating their product preferences and requirements throughout their interaction.

The data engineer plans to build a strong foundation for data storage, retrieval, and processing by utilizing cloud storage and managing various data sources. The primary objective is to furnish the data science team with a scalable and efficient environment, empowering them to create precise predictive models that anticipate customer preferences effectively. This data-driven strategy will also give Go-cart a competitive edge over other office supply companies as it targets specific customer segments and optimizes internal spending to boost departmental productivity at Office Depot Europe.

The first step involves is to create a comprehensive document that presents the outputs of the exploratory analysis, identify data issues requiring attention, and highlight key findings from the descriptive analysis.

Major objectives of this phase include:

- a.** Data Ingestion using Azure Data Factory: In this phase of the project, you will focus on data ingestion using Azure Data Factory (ADF) activities.
- b.** Azure SQL and Clean Data Loading using ADF: This task focuses on creating an Azure SQL Database, loading clean data, and performing SQL operations.

The key questions to address in this case study are:

1. What are the total extended price, discounted extended price, and discounted extended price plus tax for all line items shipped as of the given date?
2. How does the report calculate the average quantity, average extended price, and average discount for the line items?
3. What is the criteria for selecting the date within 60-120 days of the greatest ship date in the database for generating the summary pricing report?
4. How are the aggregates grouped by RETURNFLAG and LINESTATUS, and in what order are they listed?

Task 1.2: Data Ingestion using Azure Data Factory

1.2.1 . Raw data import to ADLS Gen2 storage

Create a storage account

Basics

Subscription	RDBF
Resource Group	RDBF
Location	eastus
Storage account name	g4storageaccountraw
Deployment model	Resource manager
Performance	Standard
Replication	Locally-redundant storage (LRS)

Advanced

Enable hierarchical namespace	Enabled
Enable network file system v3	Disabled
Allow cross-tenant replication	Disabled

Review

Deploying... Initializing deployment... Initializing template deployment to resource group 'RDBF'.

Containers

Successfully created storage container cont-stg.

Data storage

- Containers
- File shares
- Queues
- Tables

2044 29-08-2023

Screenshot of Microsoft Azure Storage account 'g4storageaccountraw' showing the 'Containers' blade. A success message 'Successfully created storage container' is displayed.

Name	Last modified	Anonymous access level	Lease state
Slogs	8/29/2023, 8:43:54 PM	Private	Available
cont-raw	8/29/2023, 8:44:48 PM	Private	Available

Screenshot of Microsoft Azure Storage account 'cont-raw' showing the 'Containers' blade. A success message 'Successfully created storage container' is displayed.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
customer.csv	8/29/2023, 8:45:12 PM	Hot (Inferred)		Block blob	233.96 kB	Available
lineitem.csv	8/29/2023, 8:45:16 PM	Hot (Inferred)		Block blob	7.04 MB	Available
nation.csv	8/29/2023, 8:45:11 PM	Hot (Inferred)		Block blob	2.19 kB	Available
orders.csv	8/29/2023, 8:45:13 PM	Hot (Inferred)		Block blob	1.57 MB	Available
part.csv	8/29/2023, 8:45:12 PM	Hot (Inferred)		Block blob	229.7 kB	Available
partsupp.csv	8/29/2023, 8:45:13 PM	Hot (Inferred)		Block blob	1.1 MB	Available
region.csv	8/29/2023, 8:45:13 PM	Hot (Inferred)		Block blob	413 B	Available
supplier.csv	8/29/2023, 8:45:13 PM	Hot (Inferred)		Block blob	13.44 kB	Available



1.2.2 Staging ADLS Gen2 storage

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

One-click to create data factory with sample pipeline and datasets. [Try it](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Resource group * Create new

Instance details

Name * Region * Version *

Previous Next Review + create Give feedback

Create Data Factory

Basics Git configuration Networking Advanced Tags Review + create

[View automation template](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

Basics

Subscription	RDBF
Resource group	RDBF
Name	g4bpclDFrawtostg
Region	East US
Version	V2

Previous Next **Create** Give feedback

1.2.3 Data Ingestion using Azure Data Factory from Raw to Staging storage.

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#@ravinsofttech.com/resource/subscriptions/5464fa55-993e-42e2-b3a4-206cba24f19e/resourceGroups/RDBF/providers>. The page title is "g4bpclDFrawtostg - Microsoft Azure". The main content area shows the "Essentials" section for the data factory, including details like Resource group: RDBF, Status: Succeeded, Location: East US, Subscription: RDBF, and Subscription ID: 5464fa55-993e-42e2-b3a4-206cba24f19e. Below this is the "Azure Data Factory Studio" interface with sections for Quick Starts, Tutorials, Template Gallery, and Training Modules.

The screenshot shows the Microsoft Azure portal with the URL <https://adf.azure.com/en/authoring?factory=%2Fsubscriptions%2F5464fa55-993e-42e2-b3a4-206cba24f19e%2FResourceGroups%2FRDBF%2Fpro...>. The page title is "g4bpclDFrawtostg - Azure Data | Microsoft Azure". The main content area shows the "Factory Resources" section, which includes Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. A central area shows a preview of data being processed by a pipeline, with a callout "Select an item" and the instruction "Use the resource explorer to select or create a new item".

Creation of linked services

The screenshot shows the Microsoft Azure Data Factory interface for creating a new linked service. The left sidebar navigation includes General, Factory settings, Connections, Linked services (selected), Integration runtimes, Microsoft Purview, Source control, Author, Triggers, Global parameters, Data flow libraries, Security, Credentials, Customer managed key, and Outbound rules. The main content area displays a 'Linked services' section with a message: 'Linked service defines the connection information to a data store or compute.' Below this is a search bar and a 'Create linked service' button. A grid of data store options is shown under the 'Data store' tab:

Icon	Name
Azure Blob Storage icon	Azure Blob Storage
Azure Cosmos DB for MongoDB icon	Azure Cosmos DB for MongoDB
Azure Cosmos DB for NoSQL icon	Azure Cosmos DB for NoSQL
Azure Data Explorer (Kusto) icon	Azure Data Explorer (Kusto)
Azure Data Lake Storage Gen1 icon	Azure Data Lake Storage Gen1
Azure Data Lake Storage Gen2 icon	Azure Data Lake Storage Gen2

Below the grid are 'Continue' and 'Cancel' buttons.

Screenshot 1: Creating a New Linked Service - Azure Data Lake Storage Gen2

Screenshot 2: Linked Services List - Azure Data Factory

Creation of Data Sets

Microsoft Azure | Data Factory | g4bpclDfrawtostg

Factory Resources

- Pipelines: 0
- Change Data Capture (preview): 0
- Datasets: 4
 - customer_orders_stg
 - customer_raw
 - lineitem_raw
 - orders_raw
- Data flows: 0
- Power Query: 0

Properties

General Related

Name * customer_orders_stg

Description

Annotations + New

Connection Schema Parameters

Linked service: customer_orders_stg Test connection Edit + New Learn more

File path: cont-stg / Directory / File name Browse

Compression type: Select...

Column delimiter: Comma (,)

Row delimiter: Default (\r\n or \n\r)

Encoding: Default(UTF-8)

Quote character: Double quote (")

Creation of Data Flows

Microsoft Azure | Data Factory | g4bpclDfrawtostg

Factory Resources

- Pipelines: 0
- Change Data Capture (preview): 0
- Datasets: 4
 - customer_orders_stg
 - customer_raw
 - lineitem_raw
 - orders_raw
- Data flows: 0
 - + dataflow1
- Power Query: 0

Properties

General Related

Name * dataflow1

Description

Actions Parameters Settings

Add New

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar lists 'Factory Resources' including Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. Under Datasets, 'customer_orders_stg' and 'customer_raw' are listed. The main area displays the 'Preview data' for 'customer.csv'. The table has columns: C_CUSTKEY, C_NAME, C_ADDRESS, C_NATIONKEY, and C_PHONE. Three rows of data are shown:

C_CUSTKEY	C_NAME	C_ADDRESS	C_NATIONKEY	C_PHONE
1	Customer#00000001	IvHzlApeRb ot,c,E	15	25-989-741-2988
2	Customer#00000002	XSTf4,NcwDvaWNe6tEgvwfmRchLxak	13	23-768-687-3665
3	Customer#00000003	MG9kdTD2WBHm	1	11-719-748-3364

The properties panel on the right shows the dataset's name as 'customer_raw' and its description as 'Retail 360 Customer Data'. The preview experience is currently off.

The screenshot shows the Microsoft Azure Data Factory interface, similar to the previous one but for a different dataset. The left sidebar lists 'Factory Resources' including Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. Under Datasets, 'customer_orders_stg' and 'customer_raw' are listed. The main area displays the 'Preview data' for 'lineitem.csv'. The table has columns: L_ORDERKEY, L_PARTKEY, L_SUPPKEY, L_LINENUMBER, L_QUANTITY, L_EXTENDEDPRICE, and L_DISCOUNT. Seven rows of data are shown:

L_ORDERKEY	L_PARTKEY	L_SUPPKEY	L_LINENUMBER	L_QUANTITY	L_EXTENDEDPRICE	L_DISCOUNT
1	1	1552	93	1	17.00	24710.35
2	1	674	75	2	36.00	56688.12
3	1	637	38	3	8.00	12301.04
4	1	22	48	4	28.00	25816.56
5	1	241	23	5	24.00	27389.76
6	1	157	10	6	32.00	33828.80
7	2	162	22	7	20.00	22666.20

The properties panel on the right shows the dataset's name as 'lineitem_raw' and its description as 'Retail 360 Line Item Data'. The preview experience is currently off.

Group 4: BPCL: Retail 360

Screenshot of Microsoft Azure Data Factory interface showing the 'Preview data' view for the 'orders_raw' dataset.

Properties:

- Name: orders_raw
- Description: (empty)
- Annotations: (empty)

Preview Data:

O_ORDERKEY	O_CUSTKEY	O_ORDERSTATUS	O_TOTALPRICE	O_ORDERDATE	O_ORDERPRIORITY	O_CLERKID
1	1	370	0	172799.49	1996-01-02	5-LOW
2	2	781	0	38426.09	1996-12-01	1-URGENT
3	3	1234	F	205654.30	1993-10-14	5-LOW

Screenshot of Microsoft Azure Data Factory interface showing the 'Data flow debug' view for the 'g4dfrawtostg' data flow.

Data Flow Structure:

```

graph LR
    customer[customer] --> orders[orders]
    
```

Data preview:

C_CUSTKEY	C_NAME	C_ADDRESS	C_NATIONID	C_PHONE	C_ACCTBAL	C_MKTSEGMENT	C_COMMENT
1	Customer...	IVhztApeR...	15	25-989-7...	711.56	BUILDING	to the eve...
2	Customer...	XSTf4,NC...	13	23-768-6...	121.65	AUTOMO...	l accounts...
3	Customer...	MQ9kdTD...	1	11-719-7...	7498.12	AUTOMO...	deposits e...
4	Customer...	XV/SJHLA...	4	14-128-1...	2866.83	MACHINE...	requests f...

Group 4: BPCL: Retail 360

Screenshot of Microsoft Azure Data Factory Data Flow interface showing the 'customer' dataset being imported from 'customer_raw'. The data preview shows four rows of order data.

O_ORDERKEY	O_CUSTKEY	O_ORDERST...	O_TOTALPRI...	O_ORDERD...	O_ORDERP...	O_CLERK	O_SHIPI...
1	370	O	172799.49	1996-01-02	5-LOW	Clerk#00000...	X
2	781	O	38426.09	1996-12-01	1-URGENT	Clerk#00000...	X
3	1234	F	205654.3	1993-10-14	5-LOW	Clerk#00000...	X
4	1369	O	56000.91	1995-10-11	5-LOW	Clerk#00000...	X

Screenshot of Microsoft Azure Data Factory Data Flow interface showing the 'lineitem' dataset being imported from 'orders_raw'. The data preview shows four rows of line item data.

L_ORDERKEY	L_PARTKEY	L_SUPPKEY	L_LINENU...	L_QUANTITY	L_EXTENDED...	L_DISCOUNT	L_TAX
1	1552	93	1	17.0	24710.35	0.04	0.02
1	674	75	2	36.0	56688.12	0.09	0.06
1	637	38	3	8.0	12301.04	0.1	0.02
1	22	48	4	28.0	25816.56	0.09	0.06

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory Data Flow validation interface. The validation output indicates that the data flow has been validated successfully with no errors found. The validation summary states: "Your data flow has been validated. No errors were found."

Data Flow Validation Output:

- Description:** Export data to customer_orders_stg
- Incoming stream:** lineitem
- Sink type:** Dataset
- Dataset:** customer_orders_stg
- Options:** Allow schema drift (checked), Validate schema (unchecked)

Preview experience: Off

Projection: The projection section shows the mapping between the source columns (L_ORDERKEY, L_PARTKEY, L_SUPKEY, L_LINENUMBER, L_QUANTITY) and the target columns (short, short, short, short, double). Each column is mapped to its corresponding type and format.

The screenshot shows the Microsoft Azure Data Factory Data Flow validation interface. The validation output indicates that the data flow has been validated successfully with no errors found. The validation summary states: "Your data flow has been validated. No errors were found."

Data Flow Validation Output:

- Description:** Export data to customer_orders_stg
- Incoming stream:** lineitem
- Sink type:** Dataset
- Dataset:** customer_orders_stg
- Options:** Allow schema drift (checked), Validate schema (unchecked)

Preview experience: Off

Projection: The projection section shows the mapping between the source columns (L_ORDERKEY, L_PARTKEY, L_SUPKEY, L_LINENUMBER, L_QUANTITY) and the target columns (short, short, short, short, double). Each column is mapped to its corresponding type and format.

Group 4: BPCL: Retail 360

Pending changes (5)

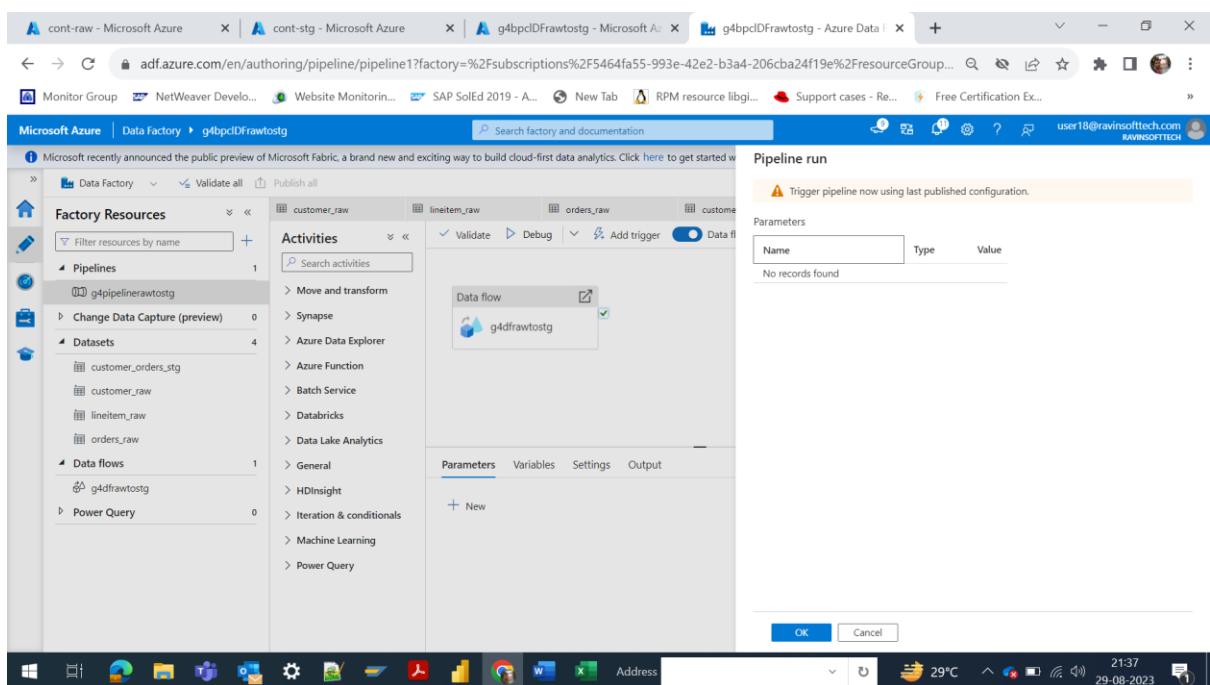
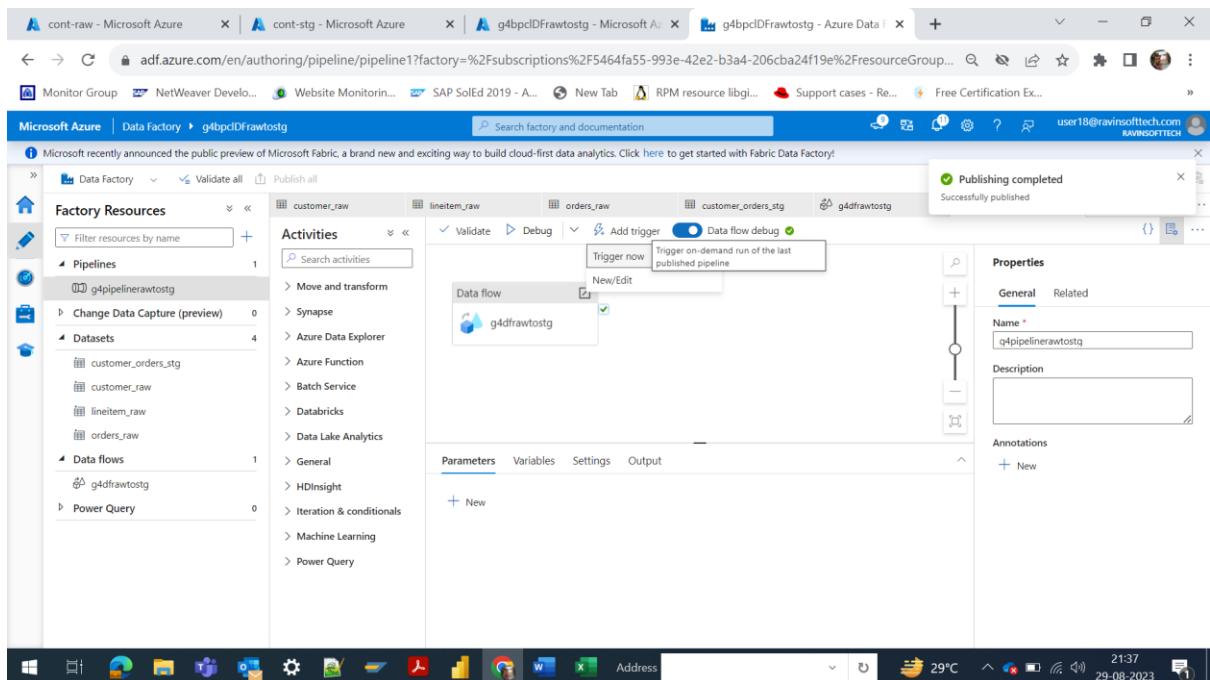
NAME	CHANGE	EXISTING
customer_raw	(New)	-
lineitem_raw	(New)	-
orders_raw	(New)	-
customer_orders_stg	(New)	-
g4dfrawtostg	(New)	-

Authentication method: Access key ([Switch to Azure AD User Account](#))
Location: cont-stg

Search blobs by prefix (case-sensitive)

Configuring & Executing the pipeline

Group 4: BPCL: Retail 360



Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory Pipeline Editor. The top navigation bar includes tabs for 'cont-raw - Microsoft Azure', 'cont-stg - Microsoft Azure', 'g4bpclDFrawtostg - Microsoft Azure', and 'g4bpclDFrawtostg - Azure Data...'. The main interface displays a 'Factory Resources' sidebar with sections for Pipelines, Datasets, Data flows, and Power Query. A central workspace shows a 'Activities' pane with options like Move and transform, Synapse, Azure Data Explorer, etc., and a 'Data flow' section containing a single step named 'g4dfrawtostg'. To the right, a 'Properties' panel shows the pipeline's name as 'g4pipelinewatstg' and its status as 'Running'. The bottom of the screen shows a Windows taskbar with various icons and the date/time as 29-08-2023.

The screenshot shows the Microsoft Azure Data Factory Pipeline Runs page. The top navigation bar is identical to the previous screenshot. The main content area is titled 'Pipeline runs' and shows a table of runs. The table has columns for Pipeline name, Run start, Run end, Duration, Triggered by, Status, Run, Parameters, and Annot. One row is visible, showing 'g4pipelinewatstg' as the pipeline name, '8/29/2023, 9:37:48 PM' as the run start, '--' as the run end, '17s' as the duration, 'Manual trigger' as the triggered by, 'In progress' as the status, and 'Original' as the run. The bottom of the screen shows a Windows taskbar with various icons and the date/time as 29-08-2023.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar has a tree view with 'Pipeline runs' selected. The main area displays a Gantt chart with one task named 'g4dfrawtostg'. Below the chart is a table titled 'Activity runs' showing one row:

Activity name	Activity status	Activity type	Run start	Duration	Log	Integration runtime	User properties	Activity ru...
g4dfrawtostg	In progress	Data flow	8/29/2023, 9:37:50 PM	24s				0e658fe9...

Staging data processed through ADF

The screenshot shows the Microsoft Azure Storage Container 'cont-stg' page. The left sidebar has 'Overview' selected. The main area shows a table of blobs:

Name	Modified	Access tier	Archive status	Block type	Size	Lease state
_SUCCESS	8/29/2023, 9:42:20 PM	Hot (Inferred)		Block blob	0 B	Available
part-00000-495e9dd0-87b7-43a3-9dc4-3a4e5f793b4c-c0...	8/29/2023, 9:42:17 PM	Hot (Inferred)		Block blob	3.95 MiB	Available
part-00000-d083cac6-ect2-4d1e-bd87-19f9a981806fc0...	8/29/2023, 9:42:19 PM	Hot (Inferred)		Block blob	1.6 MiB	Available
part-00000-d663c1d6-efe1-4f53-aa38-1a20991ca95-c0...	8/29/2023, 9:42:20 PM	Hot (Inferred)		Block blob	235.42 KiB	Available
part-00001-495e9dd0-87b7-43a3-9dc4-3a4e5f793b4c-c0...	8/29/2023, 9:42:16 PM	Hot (Inferred)		Block blob	2.89 MiB	Available

The screenshot shows the Microsoft Azure Storage Explorer interface. A CSV file named 'part-00000-d663c1d6-efe1-4f53-aa38-1a20991ca9f5-c000.csv' is open in the 'cont-stg' container. The file contains the following data:

C.CUSTKEY	C.NAME	C.ADDRESS	C.NATIONKEY
1	Customer#0000000001	"IVhzlApeRbt	E
2	Customer#0000000002	"XSTf4	NCwDvAWNe6tEgvwfmRchLxak
3	Customer#0000000003	MG9kdTD2WBHm	1
4	Customer#0000000004	XxVSlLAGtn	4
5	Customer#0000000005	KvpuyHCplrB84WgAiGV6syPZq7Tj	3

Task 1.3: Azure SQL and Clean Data Loading using ADF.

Creation of Target SQL server & database

The screenshot shows the Microsoft Azure portal with the 'Create SQL Database Server' wizard open. The 'Server details' step is active. The configuration includes:

- Server name:** g4sqlserver
- Location:** (US) East US
- Authentication method:** Use SQL authentication
- Server admin login:** g4sqladmin
- Password:** (redacted)

A tooltip provides validation feedback for the password:

- Your password must be at least 8 characters in length.
- Your password must be no more than 128 characters in length.
- Your password must contain characters from three of the following categories – English uppercase letters, English lowercase letters, numbers (0-9), and non-alphanumeric characters (!, \$, #, %, etc.).
- Your password cannot contain all or part of the login name. Part of a login name is defined as three or more consecutive alphanumeric characters.

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure portal with the URL portal.azure.com/#create/Microsoft.SQLDatabase. The page is titled 'Create SQL Database'. The 'Review + create' tab is selected. On the left, under 'Product details', it says 'SQL database by Microsoft' and 'Estimated cost per month' is listed as '384.63 INR'. Below that is a 'Terms' section with legal disclaimers. On the right, there's a 'Cost summary' box showing 'Basic (Basic)' with 'Cost per DTU (in INR)' at '76.93' and 'DTUs selected' at 'x 5', resulting in an 'ESTIMATED COST / MONTH' of '384.63 INR'. At the bottom, there are buttons for 'Validating...', '+ Previous', and 'Download a template for automation'. The taskbar at the bottom shows various application icons.

The screenshot shows the Microsoft Azure portal with the URL portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~overview/id/%2Fsubscriptions%2F5464fa55-993e-42e2-b3a.... The page is titled 'Microsoft.SQLDatabase.newDatabaseNewServer_afc46fcf5894fdc86ba9 | Overview'. It shows a deployment status message: '*** Deployment is in progress'. Under 'Deployment details', it lists a single resource: 'g4sqlserver' of type 'SQL server' with status 'Accepted'. A 'Give feedback' section at the bottom encourages users to share their deployment experience. The taskbar at the bottom shows various application icons.

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes tabs for 'g4sqlldb (g4sqlserver/g4sqlldb)', 'cont-stg - Microsoft Azure', 'g4bpclDFrawtostg - Microsoft Azure', and 'g4bpclDFrawtostg - Azure Data'. Below the navigation bar, the address bar shows 'portal.azure.com/#@ravinsofttech.com/resource/subscriptions/5464fa55-993e-42e2-b3a4-206cba24f19e/resourceGroups/RDBF...'. The main content area displays the 'Overview' page for the 'g4sqlldb' database. The left sidebar contains sections for Overview, Activity log, Tags, Diagnose and solve problems, Query editor (preview), Settings, Compute + storage, Connection strings, Properties, Locks, Data management, Replicas, Sync to other databases, Integrations, Azure Synapse Link, and Stream analytics (preview). The right panel shows the 'Essentials' section with details like Server name (g4sqlserver.database.windows.net), Status (Online), Location (East US), Subscription (RDBF), Subscription ID (5464fa55-993e-42e2-b3a4-206cba24f19e), and Tags (Add tags). Below this, there's a 'Getting started' tab followed by 'Monitoring', 'Properties', 'Features', 'Notifications (0)', 'Integrations', and 'Tutorials'. A 'Start working with your database' section provides links to 'Configure access', 'Connect to application', and 'Start developing'. The bottom of the screen shows the Windows taskbar with various pinned icons and the system tray indicating the date and time as 29-08-2023 at 21:49.

Creation of Tables

```
CREATE TABLE customer (
```

```
C_CUSTKEY integer NOT NULL,
C_NAME varchar(100)NULL,
C_ADDRESS varchar (200)NULL,
C_NATIONKEY varchar (200)NULL,
C_PHONE varchar (20)NULL,
C_ACCTBAL varchar (100)NULL,
C_MKTSEGMENT varchar (100)NULL,
C_COMMENT varchar (100)NULL
);
```

```
CREATE TABLE LINEITEM (
```

```
L_ORDERKEY INTEGER NOT NULL,
L_PARTKEY INTEGER NOT NULL,
L_SUPPKEY INTEGER NOT NULL,
L_LINENUMBER INTEGER NOT NULL,
L_QUANTITY decimal(20,5) NULL,
L_EXTENDEDPRICE decimal(20,5) NULL,
L_DISCOUNT decimal(20,5) NULL,
```

```
L_TAX decimal(20,5) NULL,  
L_RETURNFLAG VARCHAR(10) NULL,  
L_LINESTATUS VARCHAR(10) NULL,  
L_SHIPDATE DATE NULL,  
L_COMMITDATE DATE NULL,  
L_RECEIPTDATE DATE NULL,  
L_SHIPINSTRUCT VARCHAR(500) NULL,  
L_SHIPMODE VARCHAR(500) NULL,  
L_COMMENT VARCHAR(1000) NULL  
)
```

```
CREATE TABLE ORDERS (  
O_ORDERKEY INTEGER NOT NULL,  
O_CUSTKEY INTEGER NULL,  
O_ORDERSTATUS VARCHAR(10) NULL,  
O_TOTALPRICE DECIMAL(20,3) NULL,  
O_ORDERDATE DATE NULL,  
O_ORDERPRIORITY VARCHAR(30) NULL,  
O_CLERK VARCHAR(100) NULL,  
O_SHIPPRIORITY INTEGER NULL,  
O_COMMENT VARCHAR(1000) NULL  
)
```

The screenshot shows the Microsoft Azure portal interface. The left sidebar is the Azure portal navigation menu. The main area is titled "g4sqlldb (g4sqlserver/g4sqlldb) | Query editor (preview)". A query is being run:

```

2 C.CUSTKEY integer,
3 C.NAME varchar,
4 C.ADDRESS varchar,
5 C.NATIONKEY varchar,
6 C.PHONE varchar,
7 C.ACCTBAL varchar,
8 C.MKTSEGMENT varchar,
9 C.COMMENT varchar
10 );

```

The results pane shows a message: "Query succeeded: Affected rows: 0". The status bar at the bottom right indicates the date and time: 29-08-2023 21:57.

ADF for Staging to Azure SQL Server

The screenshot shows the Microsoft Azure Data Factory studio. The left sidebar lists "Factory Resources" including Pipelines, Datasets, and Data flows. The main workspace shows a data flow pipeline with a source dataset named "customer_orders_stg". On the right, a "New dataset" dialog is open, showing options for "Azure Database" and "Azure SQL Database Managed Instance". The Azure SQL Database option is selected. The status bar at the bottom right indicates the date and time: 29-08-2023 21:59.

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. In the main workspace, a 'Data flow' named 'stg_curated' is being configured. A 'Validate' button is visible above the data flow editor. On the right, a 'New linked service' dialog is open, specifically for an 'Azure SQL Database'. The 'Name' field is set to 'g4sqlldb'. Under 'Connect via integration runtime', 'AutoResolveIntegrationRuntime' is selected. The 'Account selection method' is set to 'From Azure subscription'. The 'Azure subscription' dropdown shows 'RDDB (5464fa55-993e-42e2-b3a4-206cba24f19e)'. The 'Server name' is 'g4sqlserver' and the 'Database name' is 'g4sqlldb'. The 'Authentication type' is 'SQL authentication'. At the bottom of the dialog, there are 'Create' and 'Cancel' buttons, along with a 'Test connection' link.

This screenshot is nearly identical to the previous one, showing the same interface and configuration steps for creating a linked service to an Azure SQL Database. The difference is in the status message at the bottom right of the dialog, which now says 'Connection successful' with a green checkmark icon, indicating that the connection has been successfully established.

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. Under 'Data flows', 'stg_curated' is selected. The main workspace displays a data flow diagram with two stages: 'stgcustomer' (Import data from 'customer_orders_stg') and 'customersql' (Sink). The 'Properties' pane on the right shows the data flow is named 'stg.curated'. The 'Mapping' tab is currently selected, showing the mapping between input columns (from 'stgcustomer') and output columns (to 'customersql').

This screenshot is identical to the one above, but the 'Mapping' tab is explicitly highlighted in blue at the top of the 'Properties' pane. The mapping details show 8 mappings: All outputs mapped.

Group 4: BPCL: Retail 360

Factory Resources

- Pipelines: g4pipelinewrwtsg
- Change Data Capture (preview): 0
- Datasets: customer_orders_stg, customer_raw, g4azuresql_db_customer, lineitem_raw, orders_raw
- Data flows: stg_curated (selected), g4dfrawtsg
- Power Query: 0

Properties

Name: stg_curated

Data preview

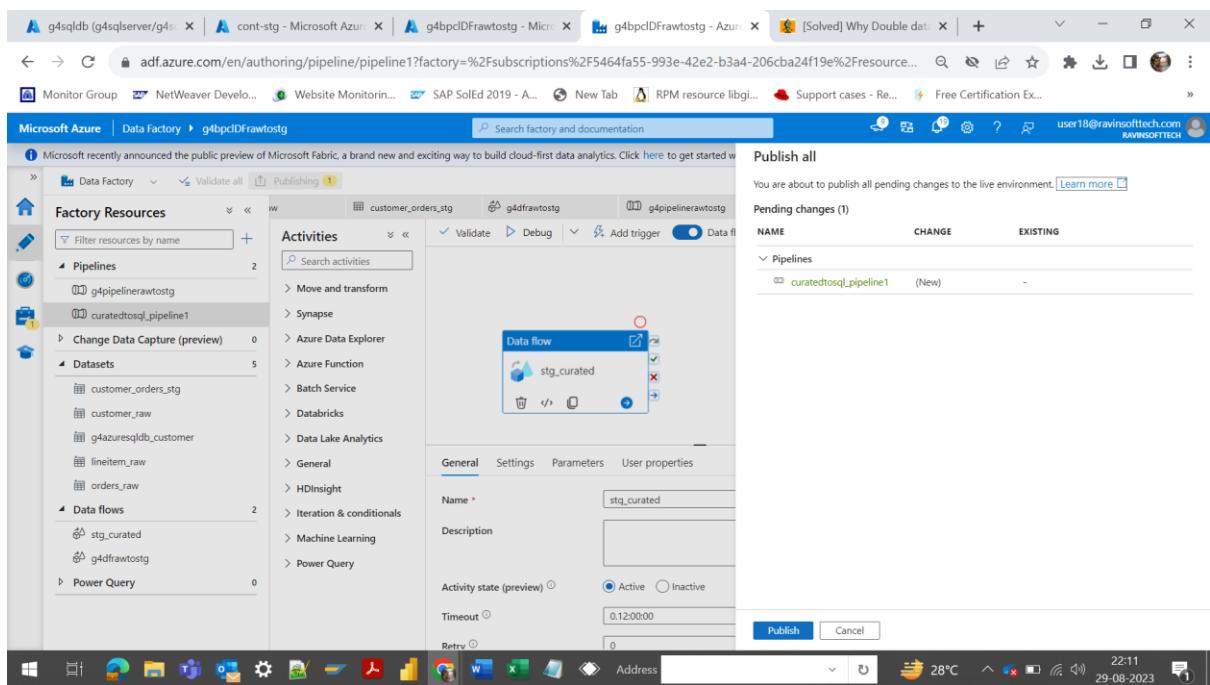
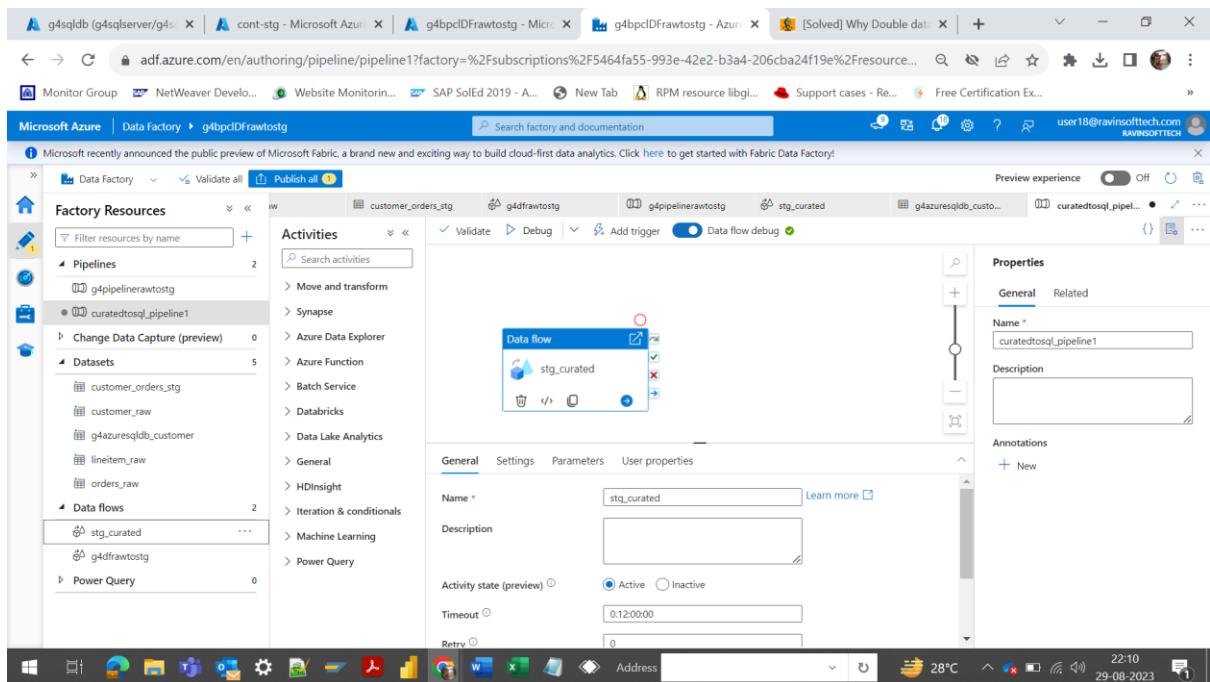
Order	Column	Type	Updated	Input column
1	C_CUSTKEY	12s short		C_CUSTKEY
2	C_NAME	abc string		C_NAME
3	C_ADDRESS	abc string		C_ADDRESS
4	C_NATIONKEY	12s short		C_NATIONKEY
5	C_PHONE	abc string		C_PHONE
6	C_ACCTBAL	12 double		C_ACCTBAL
7	C_MKTSEGMENT	abc string		C_MKTSEGMENT

Properties

Name: stg_curated

Data preview

C_CUSTKEY	C_NAME	C_ADDRESS	C_NATIONKEY	C_PHONE	C_ACCTBAL
1	Customer...	IVhzIApeR...	15	25-989-7...	711.56
2	Customer...	XSTf4,NC...	13	23-768-6...	121.65
3	Customer...	MG9k0TD...	1	11-719-7...	7498.12
4	Customer...	XxVSJhLA...	4	14-128-1...	2866.83
5	Customer...	KvpuyIC...	3	13-750-9...	794.47
6	Customer...	skZz0Csn...	20	30-114-9...	7638.57
7	Customer...	TcGe5qaZ...	18	28-190-9...	9561.95



Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, Data flows, and Power Query. The main workspace displays the 'Activities' section for the pipeline 'customer_orders_stg'. A data flow named 'stg_curated' is selected. The 'Properties' pane on the right shows the pipeline's name as 'curatedtosql_pipeline1' and its description as 'stg_curated'. The status bar at the bottom indicates 'Publishing completed'.

The screenshot shows the Microsoft Azure Data Factory monitoring interface. The left sidebar shows 'Runs' and 'Pipeline runs' under the 'All pipeline runs' section. The main area displays the 'Activity runs' for the pipeline run ID '05c9f296-1b8b-429f-bf6d-c6898fb2529e'. It shows one activity named 'stg_curated' in 'In progress' status. The status bar at the bottom indicates '22:11 29-08-2023'.

The screenshot shows the Microsoft Azure portal interface. The user is in the 'Query editor (preview)' section of the 'g4sqldb' database. The left sidebar contains navigation links like Overview, Activity log, Tags, Diagnose and solve problems, and Query editor (preview). The main area displays a table named 'dbo.customer' with columns: C_CUSTKEY, C_NAME, C_ADDRESS, C_NATIONKEY, C_PHONE, C_ACCTBAL, C_MKTSEGMENT, and C_COMMENT. A query window titled 'Query 2' shows the execution of the following SQL statement:

```
1 select * from [dbo].[customer];
```

The results pane indicates 'Query succeeded: Affected rows: 0'. The status bar at the bottom right shows the date as 29-08-2023 and the time as 22:12.

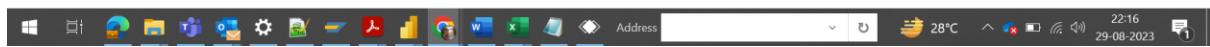
This screenshot is nearly identical to the one above, showing the same Azure portal interface and 'Query editor (preview)' section. However, the results pane now displays an error message: 'Cannot open server "g4sqlserver.database.windows.net" requested by the login. The login failed.' The status bar at the bottom right shows the date as 29-08-2023 and the time as 22:15.

Resources deleted

The screenshot shows a Microsoft Azure browser tab titled "cont-stg - Microsoft Azure". The URL is https://portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F5464fa5.... The main content area displays a large cloud icon with a raindrop, followed by the text "Not found". Below this, there's a "Summary" section with the following details:

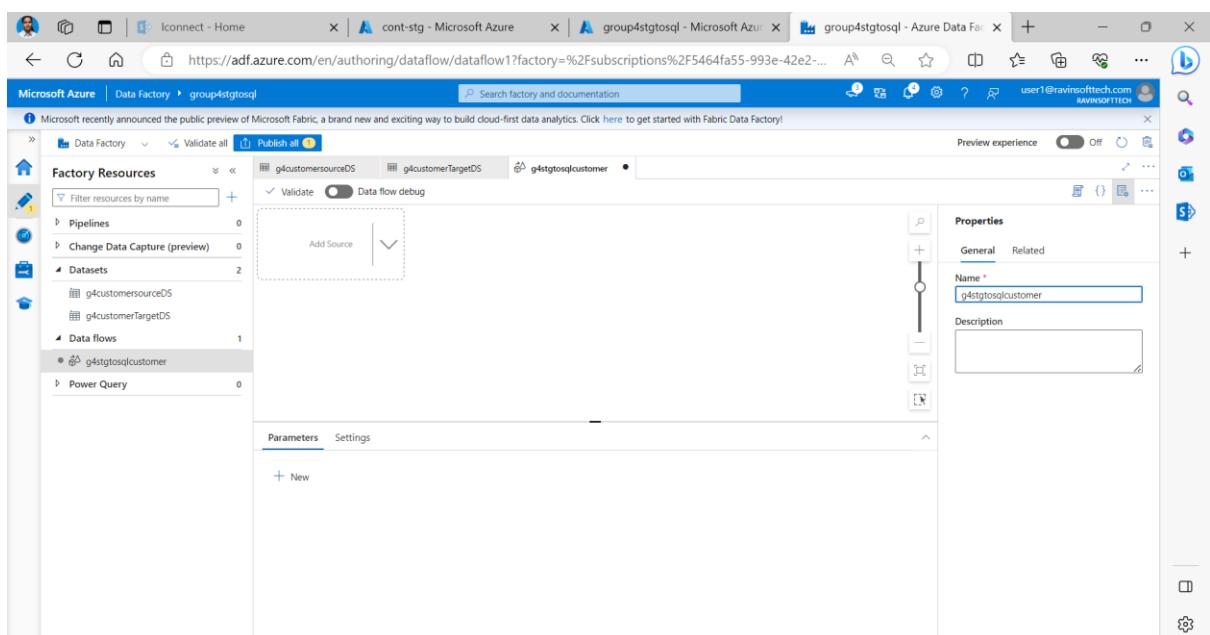
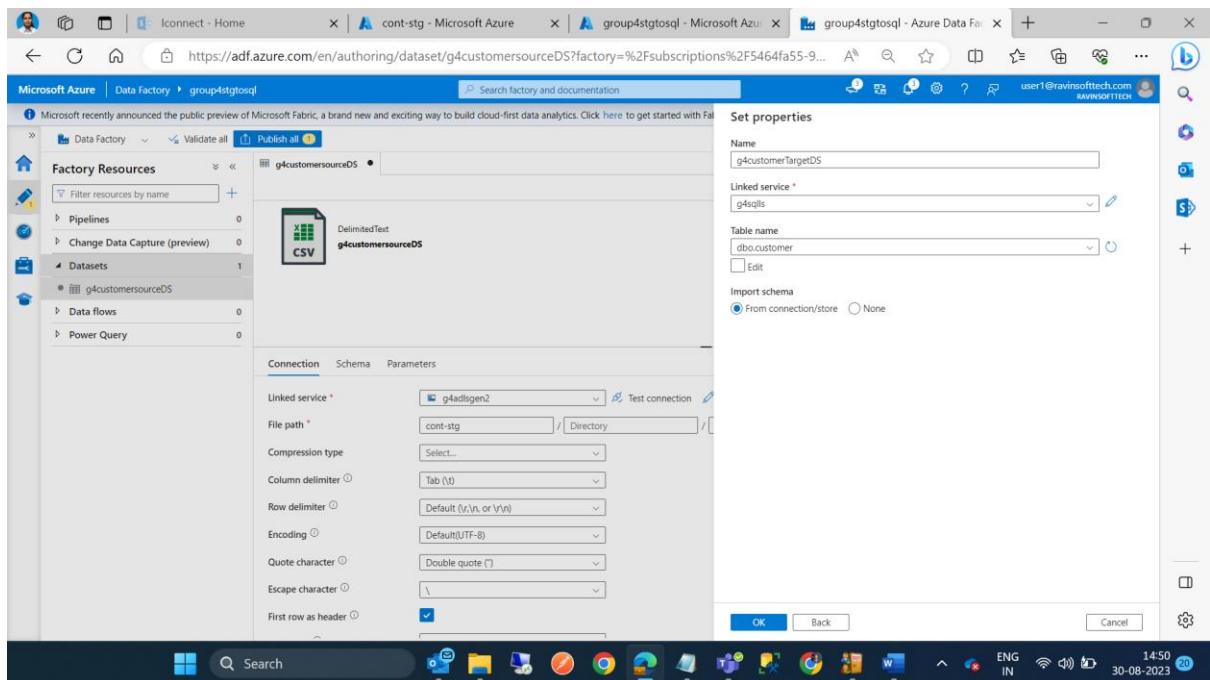
- Session ID: 6e16bc0355a479aa49dd3bcc96bc860
- Extension: Microsoft_Azure_Storage
- Error code: 404
- Resource ID: Not available
- Content: BlobsBlade

At the bottom of the page are links for "Get support" and "Perform self-diagnostics".



CUSTOMER

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar lists "Factory Resources" including Pipelines, Change Data Capture (preview), Datasets, Data flows, and Power Query. The "Datasets" section is expanded, showing one dataset named "g4customersourceDS". The main panel shows the dataset configuration for a "DelimitedText" type. The "File path" is set to "cont-stg / customer.csv". Other settings include Compression type (Select...), Column delimiter (Tab (\t)), Row delimiter (Default (\r\n, or \r\r\n)), Encoding (Default(UTF-8)), Quote character (Double quote (")), Escape character (\), First row as header (checked), and Null value (empty field). The top navigation bar shows tabs for "Data Factory", "group4stgtosql - Microsoft Azure", and "group4stgtosql - Azure Data Fair". The top right corner shows the user's email (user1@ravinsofttech.com) and the company name (RAVINSOFTTECH).



Group 4: BPCL: Retail 360

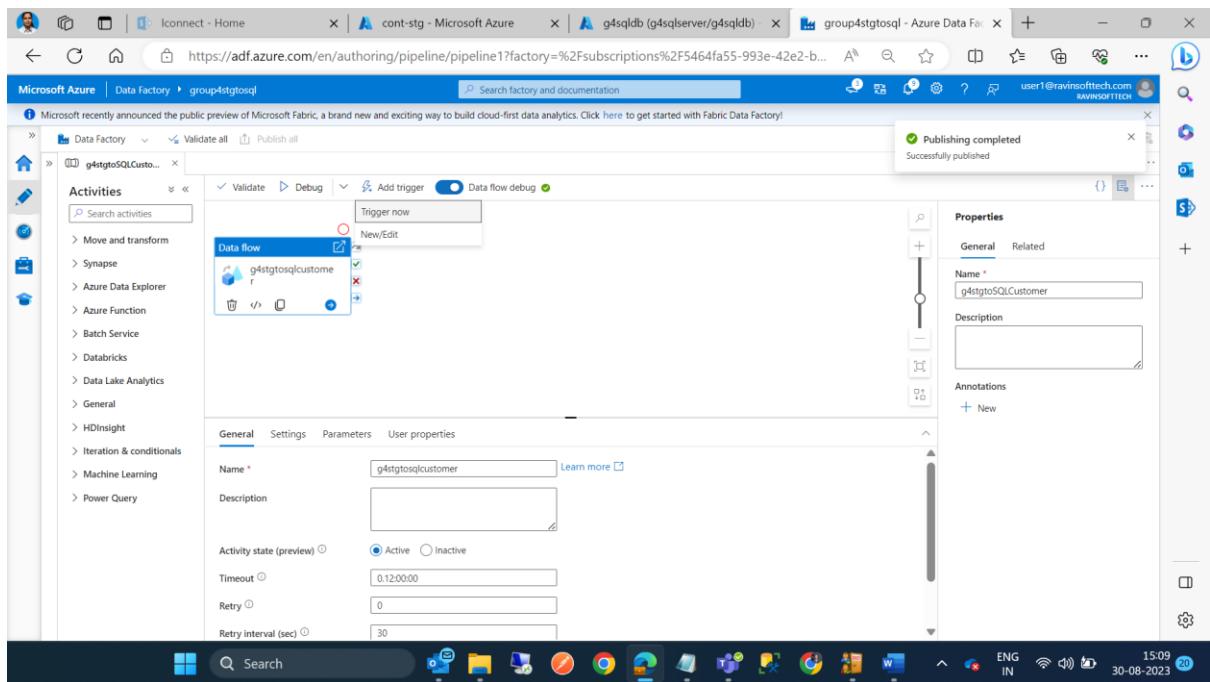
Screenshot 1: Azure Data Factory Pipeline Editor

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. Under 'Data flows', the 'g4stgtosqlcustomer' flow is selected. The main workspace displays the flow's structure: 'Customer' (sink) connected to 'CustomerTarget' (source). The source is 'g4customersourceDS'. The sink has 'C_CUSTKEY' as the primary key. The 'Data preview' tab shows 1000 rows from the source dataset.

C_CUSTKEY	C_NAME	C_ADDRESS	C_NATIO...	C_PHONE	C_ACCTBAL	C_MKTSE...	C_COMM...
1	Customer...	IvHtzApeR...	15	25-989-74...	711.56	BUILDING	to the eve...
2	Customer...	XSTf4NCw...	13	23-768-68...	121.65	AUTOMOB...	l accounts...
3	Customer...	MG9k0dTD...	1	11-719-74...	7498.12	AUTOMOB...	deposits e...
4	Customer...	XVVSJslAG...	4	14-128-19...	2866.83	MACHINERY	requests fi...
5	Customer...	KrypyuHcp...	3	13-750-94...	794.47	HOUSEHO...	n accounts...
6	Customer...	skZz0Csn...	20	30-114-96...	7638.57	AUTOMOB...	tions. even...
7	Customer...	TcGeSqaZ...	18	28-190-98...	9561.95	AUTOMOB...	ainst the ir...
8	Customer...	lO810b80A...	17	27-147-57...	6819.74	BUILDING	among th...
9	Customer...	xKiAFTIjic...	8	18-338-90...	8324.07	FURNITUR...	r the initial...

Screenshot 2: Azure Data Factory Pipeline Publishing

The screenshot shows the 'Publish all' dialog in the Azure Data Factory interface. It displays pending changes for the 'g4stgtosqlcustomer' pipeline. The 'General' tab is selected, showing details like Name (g4stgtosqlcustomer), Activity state (Active), and Timeout (0.12:00:00). The 'Publish' button is visible at the bottom right.



Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory Data Flow interface. The main pane displays a data flow named 'g4stgtosqlcustomer' with two stages: 'Customer' (source) and 'CustomerTarget' (sink). The 'Customer' stage imports data from 'g4customersourceDS'. The 'CustomerTarget' stage has 8 columns listed. Below the stages, the 'Data preview' tab is selected, showing a table with 8 rows of customer data. The table includes columns such as C_CUSTKEY, C_NAME, C_ADDRESS, C_NATIONID, and C_ACCTBAL. The preview area also shows statistics like INSERT 0, UPDATE 0, DELETE 0, and UPSERT 0. To the right of the preview, a message states 'Your factory has been validated. No errors were found.' The top navigation bar shows tabs for 'Validate all', 'Publish all', and 'Preview experience'. The status bar at the bottom indicates the date as 30-08-2023 and time as 15:05.

This screenshot is similar to the one above but includes a 'Notifications' sidebar on the right. The sidebar lists several events: 'Publishing' (Gathering your new changes, a few seconds ago), 'Successfully imported' (Successfully imported the schema for Customer (Source), 3 minutes ago), 'Publishing canceled' (Canceled, 7 minutes ago), 'Data flow debug' (Cluster is ready, Session ID: 055688d7-4b7c-4ee5-85e0-1dd3bdbe80a3, 11 minutes ago), 'Publishing completed' (Successfully published, 14 minutes ago), 'Successfully applied settings' (Successfully applied settings to g4adlsgen2 (Linked service), 17 minutes ago), and 'Successfully created' (Successfully created, 19 minutes ago). The rest of the interface is identical to the first screenshot, showing the Data Flow editor and its preview results.

The screenshot shows the Microsoft Azure portal with multiple tabs open. The active tab is 'g4sqlldb (g4sqlserver/g4sqlldb) | Query editor (preview)'. The query editor displays a query:

```
1 select * from [dbo].[customer]
```

The results pane shows the following data:

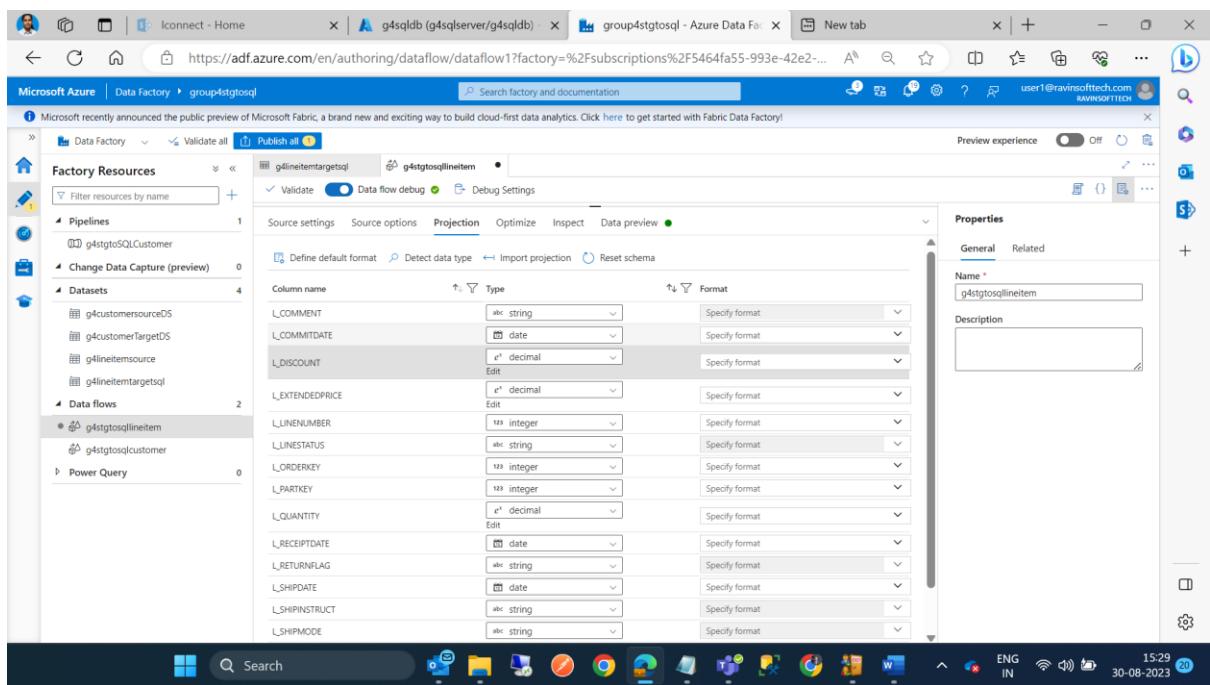
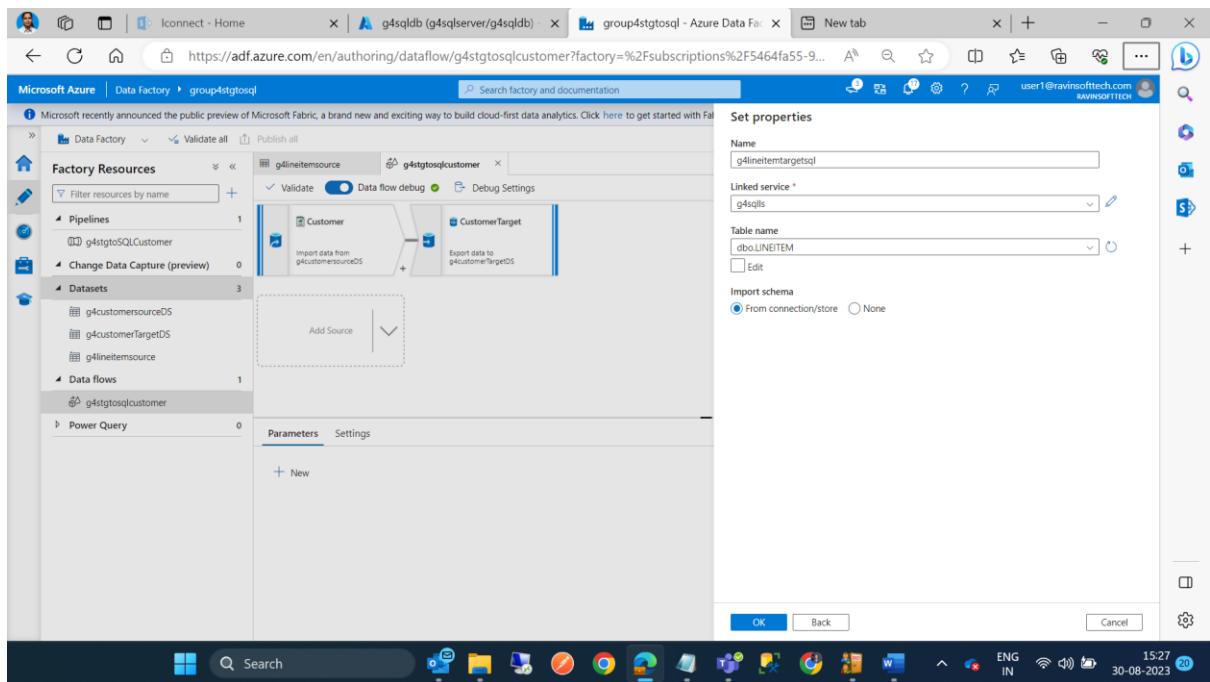
C_CUSTKEY	C_NAME	C_ADDRESS	C_NATIONKEY	C_PHONE	C_ACCTBAL	C_MKTSEGMENT
1	Customer#000000001	IVhzlApeRb ot,c,E	15	25-989-741-2988	711.56	BUILDING
2	Customer#000000002	XSTf4,NCwDVaWNe6...	13	23-768-687-3665	121.65	AUTOMOBILE
3	Customer#000000003	MG9kdTD2WBHm	1	11-719-748-3364	7498.12	AUTOMOBILE

Message at the bottom: **Query succeeded | 1s**

Line Item

The screenshot shows the Microsoft Azure portal with the URL <https://adf.azure.com/en/authoring/dataset/g4lineitemsource?factory=%2Fsubscriptions%2F5464fa55-993e-42e2-b3a4-206...>. The left sidebar shows 'Factory Resources' with 'Datasets' expanded, showing 'g4customersourceDS', 'g4customerTargetDS', and 'g4lineitemsource'. The main pane shows the 'g4lineitemsource' dataset configuration:

- Connection:** g4adlsgen2
- File path:** cont-stg / Directory / lineitem.csv
- Compression type:** Select...
- Column delimiter:** Tab (\t)
- Row delimiter:** Line feed (\n)
- Encoding:** Default(UTF-8)
- Quote character:** Double quote (")
- Escape character:** Backslash (\)
- First row as header:** checked



Group 4: BPCL: Retail 360

The screenshot shows the Azure Data Factory Data Flow preview experience. A data preview of 100 rows is displayed for the 'LineltemSource' dataset. The columns shown are L_ORDERKEY, L_PARTKEY, L_SUPPKEY, L_LINENUMER, L_QUANTITY, and L_EXTENDEDPRICE. The data preview table has 100 rows.

L_ORDERKEY	L_PARTKEY	L_SUPPKEY	L_LINENUMER	L_QUANTITY	L_EXTENDEDPRICE
1	1552	93	1	17.00000	24710.35000
1	674	75	2	36.00000	56688.12000
1	637	38	3	8.00000	12301.04000
1	22	48	4	28.00000	52016.56000
1	241	23	5	24.00000	27389.76000
1	157	10	6	32.00000	33828.80000
2	1062	33	1	38.00000	36596.28000
3	43	19	1	45.00000	42436.80000
3	191	70	2	49.00000	53468.31000
3	1285	60	3	27.00000	32029.56000
3	294	??	4	2.00000	2388.58000

The screenshot shows the Azure Data Factory Data Flow validation output. It displays a message stating 'Your factory has been validated.' and 'No errors were found.' The validation status is shown as 'Validated'.

Validation Output:

Your factory has been validated.
No errors were found.

Group 4: BPCL: Retail 360

The screenshot shows the Azure Data Factory pipeline editor. A pipeline named "g4stgtosqlitem" is selected. The pipeline consists of a single "Data flow" activity named "g4stgtosqlitem". The pipeline is set to "Active" status with a timeout of 0:12:00:00. The Properties pane on the right shows the pipeline's name as "g4stgtosqlitem" and its description field is empty.

The screenshot shows the Azure Data Factory pipeline editor with the "Publish all" dialog open. The dialog lists a pending change for the pipeline "g4stgtosqlitem". The "Publish" button is highlighted at the bottom of the dialog.

The screenshot shows the Azure Data Factory pipeline editor. A pipeline named 'g4stgtosqlitem' is selected. The 'Data flow' activity is highlighted. The pipeline status is 'Running' with a message: 'Successfully running g4stgtosqlitem (Pipeline). View pipeline run'. The pipeline properties are shown in the right panel, including the name 'g4stgtosqlitem' and general settings like timeout (0:12:00:00) and retry interval (30 seconds).

The screenshot shows the Azure portal's Query editor. A query is being run against the 'g4sqlldb' database. The query is: 'SELECT TOP (1000) * FROM [dbo].[LINEITEM]'. The results pane displays the first few rows of the LINEITEM table.

L_ORDERKEY	L_PARTKEY	L_SUPPKEY	L_LINENUMBER	L_QUANTITY	L_EXTENDEDPRICE	L_DISCOUNT
34017	224	6	1	7.00000	7869.54000	0.04000
34017	1811	12	2	43.00000	73650.83000	0.04000
34017	1153	an	3	22.00000	21785.65000	0.01000

ORDERS

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Studio interface. On the left, a sidebar menu includes 'Overview', 'Activity log', 'Tags', 'Diagnose and solve problems', 'Query editor (preview)', 'Settings', 'Compute + storage', 'Connection strings', 'Properties', 'Locks', 'Replicas', and 'Sync to other databases'. The main area displays a 'Query 3' tab with the following SQL code:

```
1 CREATE TABLE ORDERS (
2     O_ORDERKEY INTEGER NOT NULL,
3     O_CUSTKEY INTEGER NULL,
4     O_ORDERSTATUS VARCHAR(10) NULL,
5     O_TOTALPRICE DECIMAL(20,3) NULL,
6     O_ORDERDATE DATE NULL,
7     O_ORDERPRIORITY VARCHAR(30) NULL,
8     O_CLERK VARCHAR(100) NULL,
```

The 'Messages' tab shows the message: 'Query succeeded: Affected rows: 0'.

The screenshot shows the Microsoft Azure Data Factory interface. The left sidebar lists 'Factory Resources' such as 'Pipelines', 'Datasets', and 'Data flows'. The 'Datasets' section shows a 'g4OrderDetailsourceDS' dataset. The main panel displays the 'g4OrderDetailsourceDS' dataset configuration. The 'Connection' tab shows a linked service 'g4adlsgen2' selected. The 'File path' field contains 'cont-stg/orders.csv'. Other settings include 'Compression type: Select...', 'Column delimiter: Tab (\t)', 'Row delimiter: Line feed (\n)', 'Encoding: Default(UTF-8)', 'Quote character: Double quote ("')', 'Escape character: Backslash (\)', and 'First row as header: checked'. The 'Properties' pane on the right shows the name 'g4OrderDetailsourceDS' and a description field.

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. In the center, a dataset named 'g4OrderDetailssqlsourceDS' is selected. The 'Connection' tab is active, showing details for a DelimitedText CSV source. The 'Linked service' dropdown is set to 'g4sqls'. The 'Table name' dropdown is set to 'dbo.ORDERS'. The 'Import schema' section has 'From connection/store' selected. The bottom right contains 'OK', 'Back', and 'Cancel' buttons.

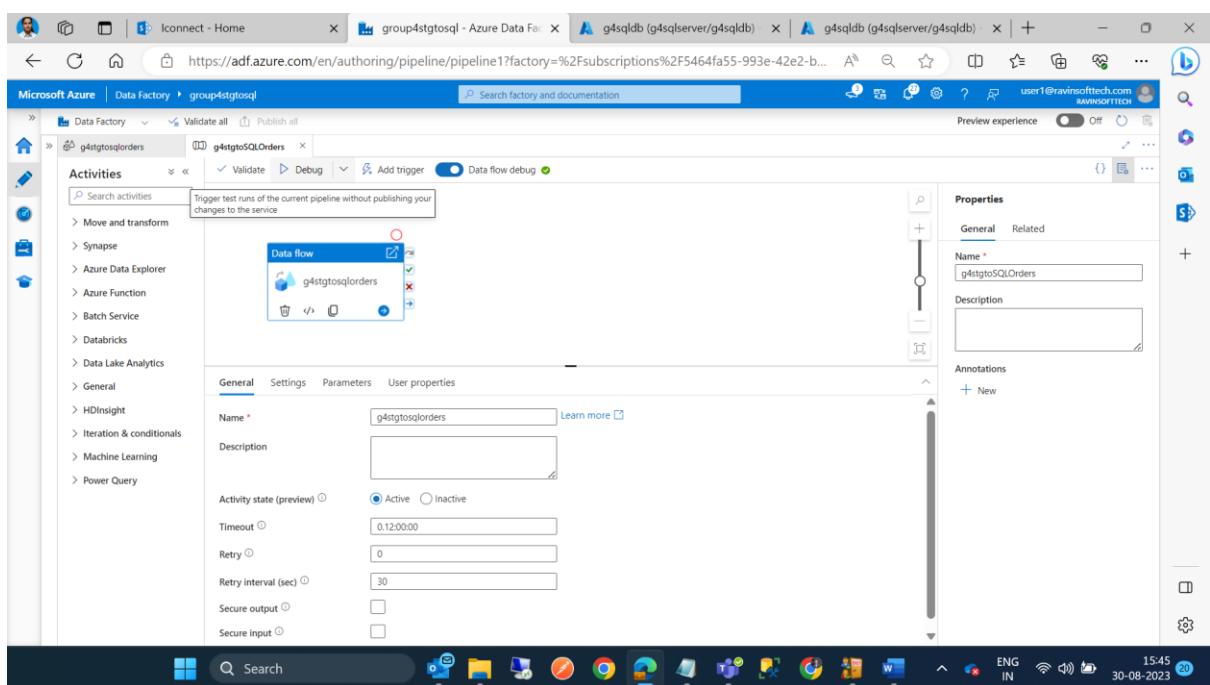
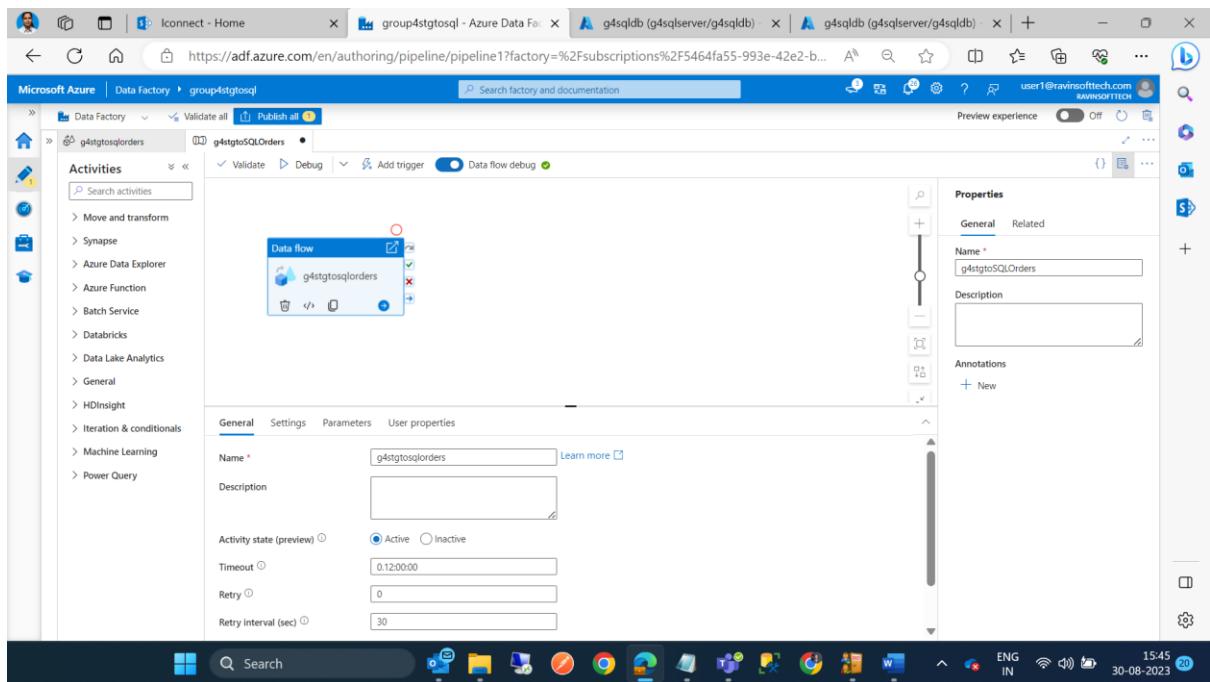
The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. A data flow named 'g4stgtosqlorders' is selected. The 'Data preview' tab is active, showing a preview of the 'OrderSource' table with 100 rows. The table has columns: O_ORDERKEY, O_CUSTKEY, O_ORDERSTATUS, O_TOTALPRICE, O_ORDERDATE, and O_ORDERPRIORITY. The data preview table is as follows:

O_ORDERKEY	O_CUSTKEY	O_ORDERSTATUS	O_TOTALPRICE	O_ORDERDATE	O_ORDERPRIORITY
1	370	O	172799.490	1996-01-02	5-LOW
2	781	O	38426.090	1996-12-01	1-URGENT
3	1234	F	205654.300	1993-10-14	5-LOW
4	1369	O	56000.910	1995-10-11	5-LOW
5	445	F	105367.670	1994-07-30	5-LOW
6	557	F	45523.100	1992-02-21	4-NOT SPEC...
7	392	O	271885.660	1996-01-10	2-HIGH
32	1301	O	198665.570	1995-07-16	2-HIGH

Group 4: BPCL: Retail 360

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Factory Resources' sidebar lists Pipelines, Datasets, and Data flows. Under 'Data flows', 'g4stgtosqlorders' is selected. The main workspace displays a data flow diagram with two components: 'OrderSource' and 'OrderTargetSQL'. The 'OrderSource' component is connected to 'g4OrderDetailsourceDS'. The 'OrderTargetSQL' component has 9 total columns. Below the diagram, the 'Mapping' tab is active, showing the mapping between 'Input columns' and 'Output columns'. The 'Output format' section indicates '9 mappings: All outputs mapped'. The status bar at the bottom shows the date and time as 30-08-2023 15:44.

The screenshot shows the same Azure Data Factory interface as the previous one, but with the 'Data preview' tab selected. This tab displays a preview of the data being processed by the 'g4stgtosqlorders' data flow. The preview table includes columns: O_ORDERKEY, O_CUSTKEY, O_ORDERSTATUS, O_TOTALPRICE, O_ORDERDATE, O_ORDERPRIORITY, O_CLERK, and O_SHIPPRIORITY. The data shows 32 rows of order details, such as OrderID 370 from Customer 781, OrderID 1234 from Customer 1369, and so on. The status bar at the bottom shows the date and time as 30-08-2023 15:44.



SQL Operations Post data loading

```
1 select * from dbo.Orders
```

O_ORDERKEY	O_CUSTKEY	O_ORDERSTATUS	O_TOTALPRICE	O_ORDERDATE	O_ORDERPRIORITY	O_CLERK
1	370	O	172799.490	1996-01-02T00:00:00.0000000	5-LOW	Clerk#
2	781	O	38426.090	1996-12-01T00:00:00.0000000	1-URGENT	Clerk#
3	1724	C	20000.000	1992-11-18T00:00:00.0000000	5-LOW	Clerk#

Query succeeded | 10s

```
select
L_SHIPDATE,
sum(L_EXTENDEDPRICE) as EXTENDED_PRICE,
sum(L_DISCOUNT) as DISCOUNT,
sum(L_TAX) as TAX,
sum(L_DISCOUNT+L_TAX) as DISCOUNT_TAX
from [dbo].[LINEITEM]
group by
L_SHIPDATE
```

```
1 select
2 L_SHIPDATE,
3 sum(L_EXTENDEDPRICE) as EXTENDED_PRICE,
4 sum(L_DISCOUNT) as DISCOUNT,
5 sum(L_TAX) as TAX,
6 sum(L_DISCOUNT+L_TAX) as DISCOUNT_TAX
7 from [dbo].[LINEITEM]
8 group by
```

L_SHIPDATE	EXTENDED_PRICE	DISCOUNT	TAX	DISCOUNT_TAX
1997-05-01T00:00:00.0000000	1593233.26000	1.94000	1.72000	3.66000
1994-08-09T00:00:00.0000000	1789405.64000	2.98000	2.02000	5.00000
1996-06-03T00:00:00.0000000	2101215.54000	2.92000	1.76000	4.68000

Query succeeded | 4s

```

select
AVG(L_EXTENDEDPRICE) as avg_EXTENDED_PRICE,
AVG(L_DISCOUNT) as avg_DISCOUNT,
AVG(L_TAX) as avg_TAX
from [dbo].[LINEITEM]

```

The screenshot shows a Microsoft Azure portal window with multiple tabs open. The active tab is 'g4sqldb (g4sqlserver/g4sqldb) | Query editor (preview)'. The query entered is:

```

1 select
2 AVG(L_EXTENDEDPRICE) as avg_EXTENDED_PRICE,
3 AVG(L_DISCOUNT) as avg_DISCOUNT,
4 AVG(L_TAX) as avg_TAX
5 from [dbo].[LINEITEM]
6

```

The results section shows the following data:

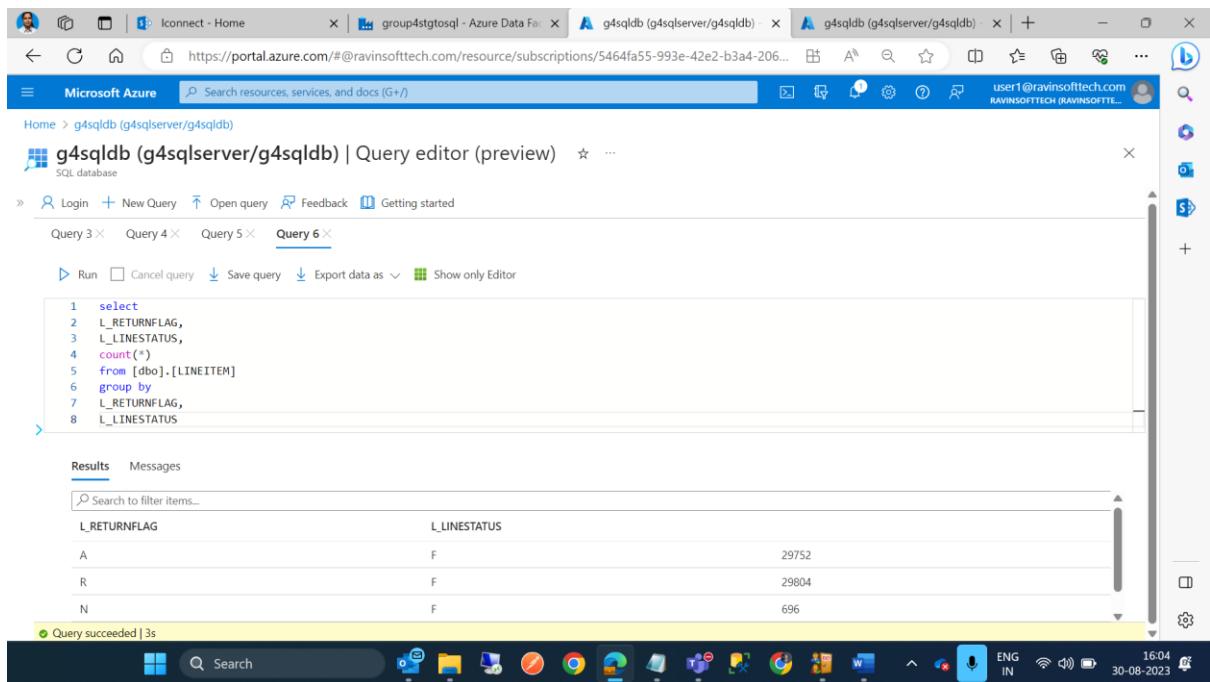
	avg_EXTENDED_PRICE	avg_DISCOUNT	avg_TAX
	35765.513260	0.049930	0.040224

A message at the bottom indicates 'Query succeeded | 1s'.

```

select
L_RETURNFLAG,
L_LINESTATUS,
count(*)
from [dbo].[LINEITEM]
group by
L_RETURNFLAG,
L_LINESTATUS

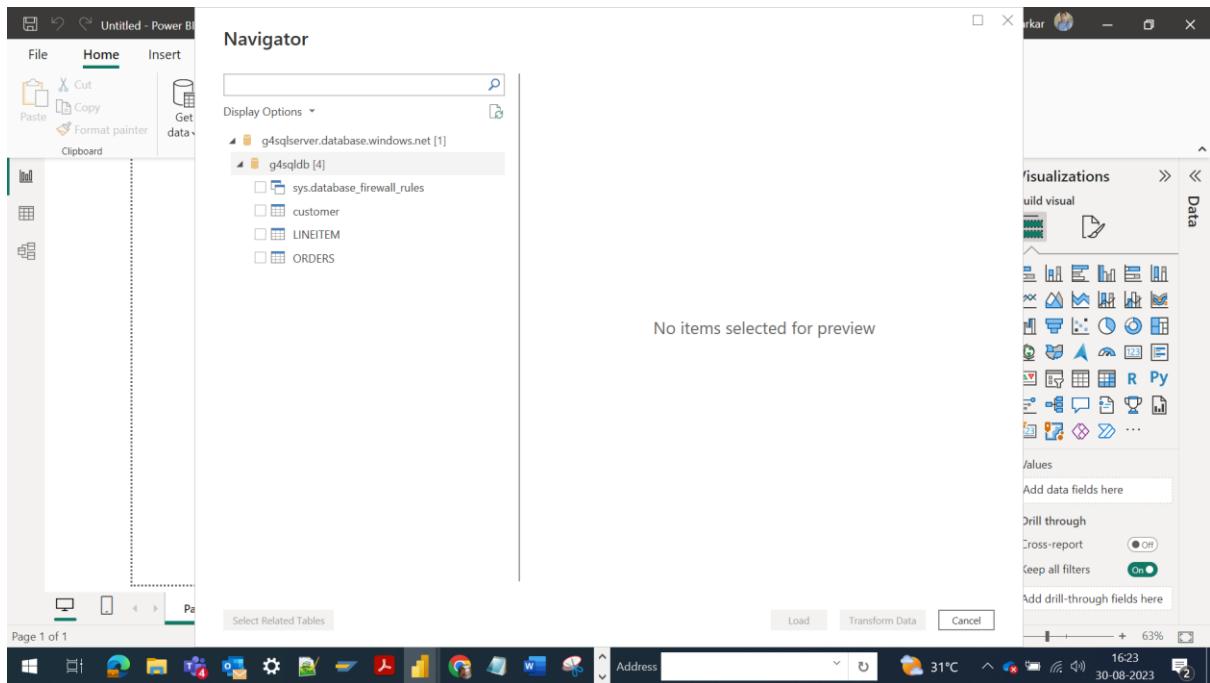
```



The screenshot shows the Microsoft Azure SQL Data Explorer interface. A query has been run against the 'g4sqldb' database, specifically against the 'LINEITEM' table. The results show the count of items categorized by 'L_RETURNFLAG' and 'L_LINESTATUS'. The data is as follows:

L_RETURNFLAG	L_LINESTATUS	Count
A	F	29752
R	F	29804
N	F	696

Data Visualization using Power BI



The screenshot shows the Power BI desktop application. The 'Navigator' pane on the left lists the tables available in the 'g4sqlserver.database.windows.net' database: 'customer', 'LINEITEM', and 'ORDERS'. The 'Visualizations' pane on the right displays various chart and report templates. The status bar at the bottom indicates the date as 30-08-2023 and the time as 16:23.

1. What are the total extended price, discounted extended price, and discounted extended price plus tax for all line items shipped as of the given date?

Untitled - Power BI Desktop

File Home Insert Modeling View Optimize Help

Clipboard Data Queries Insert Calculations

Shipment Date Total Extended Price Total discounted Extended Price Total discounted Extended Price plus Y : ...

Shipment Date	Total Extended Price	Total discounted Extended Price	Total discounted Extended Price plus Y
04 January 1992	75,584.16	75,584.04	75,584.10
06 January 1992	2,30,130.58	2,30,130.42	2,30,130.56
08 January 1992	1,23,402.72	1,23,402.58	1,23,402.58
09 January 1992	71,992.96	71,992.80	71,992.92
11 January 1992	3,94,138.96	3,94,138.54	3,94,138.74
12 January 1992	68,252.40	68,252.00	68,252.26
13 January 1992	5,10,640.38	5,10,639.52	5,10,640.06
14 January 1992	2,40,588.80	2,40,588.30	2,40,588.62
15 January 1992	3,10,170.38	3,10,170.24	3,10,170.78
16 January 1992	2,93,686.48	2,93,685.82	2,93,686.24
17 January 1992	3,40,249.28	3,40,248.90	3,40,249.18
18 January 1992	2,56,694.62	2,56,694.38	2,56,694.52
19 January 1992	3,75,981.38	3,75,980.68	3,75,981.10
20 January 1992	2,83,623.00	2,83,622.52	2,83,622.78
21 January 1992	3,11,036.00	3,11,035.62	3,11,036.02
22 January 1992	83,513.60	83,513.44	83,513.60
23 January 1992	6,33,464.88	6,33,464.18	6,33,464.74
24 January 1992	4,68,128.02	4,68,127.48	4,68,127.88
Total	4,30,43,79,520.94	4,30,43,73,511.86	4,30,43,78,352.88

Filters

Visualizations Data

Build visual

Values

Add data fields here

Drill through

Cross-report

Keep all filters

Add drill-through fields here

Storage Mode: DirectQuery (click to change)

Address: 1642 30-08-2023

2. How does the report calculate the average quantity, average extended price, and average discount for the line items?

Untitled - Power BI Desktop

File Home Insert Modeling View Optimize Help Format Data / Drill

Clipboard Data Queries Insert Calculations

L_LINENUMBER Average of L_QUANTITY

L_LINENUMBER	Average of L_QUANTITY
1	25.71
2	25.61
3	25.60
4	25.49
5	25.15
6	25.26
7	25.17
Total	25.53

Refresh

- customer
- Detecting Table Schema...
- LINEITEM
- Detecting Table Schema...
- ORDERS
- Evaluating...

Cancel

Filters

Visualizations Data

Build visual

Columns

- Line Number
- Average Quantity
- Average Extended Price
- Average Discount

Drill through

Storage Mode: DirectQuery (click to change)

Address: 1644 31°C 30-08-2023

Line Number Average Quantity Average Extended Price Average Discount

1	25.71	36,000.64	0.05
2	25.61	35,738.36	0.05
3	25.60	35,914.34	0.05
4	25.49	35,755.79	0.05
5	25.15	35,345.87	0.05
6	25.26	35,593.87	0.05
7	25.17	35,192.88	0.05
Total	25.53	35,765.51	0.05

Storage Mode: DirectQuery (click to change)

How are the aggregates grouped by RETURNFLAG and LINESTATUS, and in what order are they listed?

Return Flag Line Status Sum of Quantity Sum of Extended Price Sum of Discount Sum of Tax

A	F	7,6912.00	1,0646,96,423.30	1,490.02	1,196.16
N	F	17,942.00	2,4769,602.74	33.24	28.38
O	O	15,30502.00	2,1457,24,604.20	3,000.76	2,411.20
R	F	7,62,898.00	1,06,91,88,890.70	1,485.06	1,205.28
Total		30,72,254.00	4,30,43,79,520.94	6,009.08	4,841.02

Storage Mode: DirectQuery (click to change)
