Microsoft Fabric DREAM PoC Accelerator

BDM + TDM Script

For demo feedback and other questions, please email dreamdemos@microsoft.com
DREAM Demos and DPoC assets for partners in **MPN Portal** at https://aka.ms/dreams

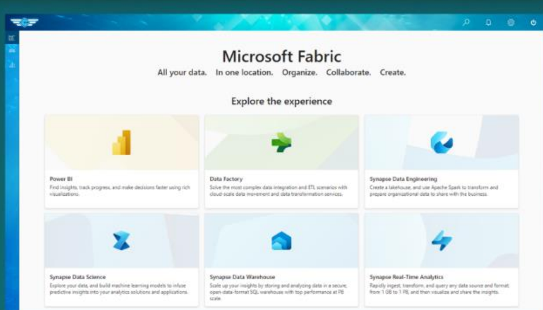


Table of Contents

1. Setting the scene

! Important Note:

This demo illustrates some compelling Microsoft Fabric features. To avoid creating multiple assets during execution, this demo has been constructed with a combination of **web app, hands-on experience, and click-by-click (embedded in the web app)** instructions. We recommend following the steps provided with the demo for a seamless experience.

This document is intended to be used as a demo script for the environment that was set up using the setup instructions located here. Make sure you follow the setup document completely before starting the demo.

1.1 Why Microsoft Fabric?

Microsoft is a leader in all four Gartner Magic Quadrants with its comprehensive portfolio of analytics products that are widely used in the industry on a massive scale. Over 5 million developers build on the Microsoft platform every single month. It has data integration tools, cloud database management systems, cloud AI and analytics, and a business intelligence platform. Microsoft makes all these capabilities work seamlessly so that developers can focus on moving the business forward versus systems integration.

However, industry is at a crossroads today. AI is causing a massive shift, but AI is only as good as the data that it can work with – if you put garbage in, you're going to get garbage out. We recognize that in today's world, it's literally awash with data, not just data from our transactional systems of record, but data from the world around us, the applications we build, the devices we use, interpersonal interactions, and so much more. And the challenge we are all struggling with is how do we translate this data into a competitive advantage? Because we recognize that in today's world, a competitive advantage starts with how we leverage data.

The good news is there's been a ton of innovation in the data and AI space. However, there's massive fragmentation of the modern data stack, and somehow our customers are expected to make sense of all its components, to go from data to insights. This is why we hear chief data officers consistently say, "Please simplify. I want to be the chief data officer, not the chief integration officer." This is why we have developed Microsoft Fabric.

Microsoft Fabric has everything to offer that lets raw data go from developers' hands to insights in the hands of business users, moving the business forward. It has been built from the ground up to be ready for AI. Microsoft Fabric makes four promises:

1. **Complete analytics platform:** To go from raw data to insights, Microsoft Fabric delivers all the capabilities as software as a service (SaaS) to each data user in the field of analytics. A developer can simply sign up within a few seconds and get real value in minutes.

2. **Lake-centric and open architecture:** OneLake as a single, unified stack data lake for the entire organization. Microsoft is moving away from proprietary formats, a storage format across an entire analytic stack, to completely open formats based on Parquet and Delta Lake.

3. **Deep integration of Microsoft Fabric with Microsoft 365:** Every piece of data in OneLake moves the business forward through Power BI which is integrated into Microsoft 365. So, automatically the data from Microsoft Fabric flows into Microsoft Teams, Excel, PowerPoint, and SharePoint. In short, it is the data showing up for business users versus them having to come to the data.

4. **Microsoft Fabric + Azure OpenAI Copilot:** Copilot in Microsoft Fabric, is now in preview. **Note:** Copilot experience won't be covered in this demo.

1.2 Demo Scenario

Contoso is a large enterprise with a presence in various industries such as retail, manufacturing, and finance. While this demo illustrates the transformation they implement in their retail business, it is equally applicable to other industries as well.

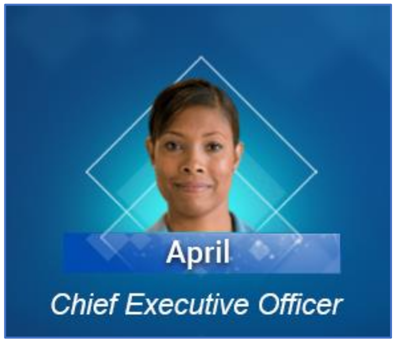
On the retail side, Contoso has thousands of brick-and-mortar stores across the world. They also have an online store. April is the new CEO of Contoso. When April takes the helm of Contoso as its new CEO, she soon discovers some discouraging numbers for some of their most important company KPIs like revenue, churn rate, operating margin, and customer experience. She talks to her Chief Data Officer Rupesh, who is a data-driven decision maker. Our story is centered around Rupesh and the transformation he brings to Contoso. His CDO KPIs directly map to those company KPIs mentioned earlier. Rupesh knows that departments in Contoso currently have their data in silos, leading to huge integration challenges.

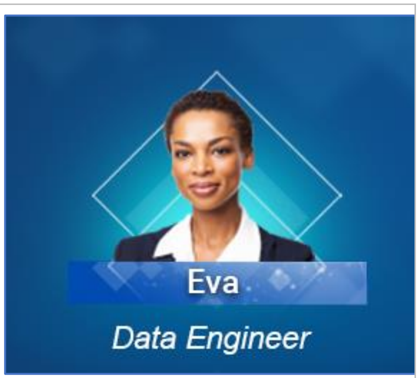
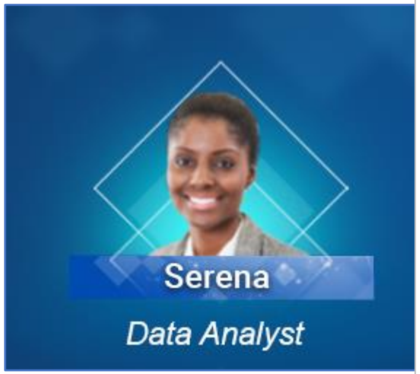
To make matters worse in terms of integration challenges, Rupesh's team is faced with yet another challenge. A while ago Contoso leadership discovered that millennials were leaving them in large numbers because they didn't have the brands and products that captivated their attention. With an eye on the future, Contoso decided to acquire LitWare Inc., an organization that carries the brands that millennials love. However, integrating LitWare Inc.'s Sales data with Contoso was a huge challenge, impacting their projected revenue.

So, Rupesh rallies his team of data engineers, data scientists, and data analysts to help overcome these integration challenges. Fortunately, they are exclusive participants of Microsoft Fabric (Preview) – a complete analytics platform for all the data users.

The demo illustrates how Rupesh and his team leverage Microsoft Fabric to literally transform his CDO KPIs. The demo shows how they build an end-to-end analytics project focusing on role-centric data management experiences. We will see how the improved CDO KPIs finally led to a green dashboard for April, their CEO.

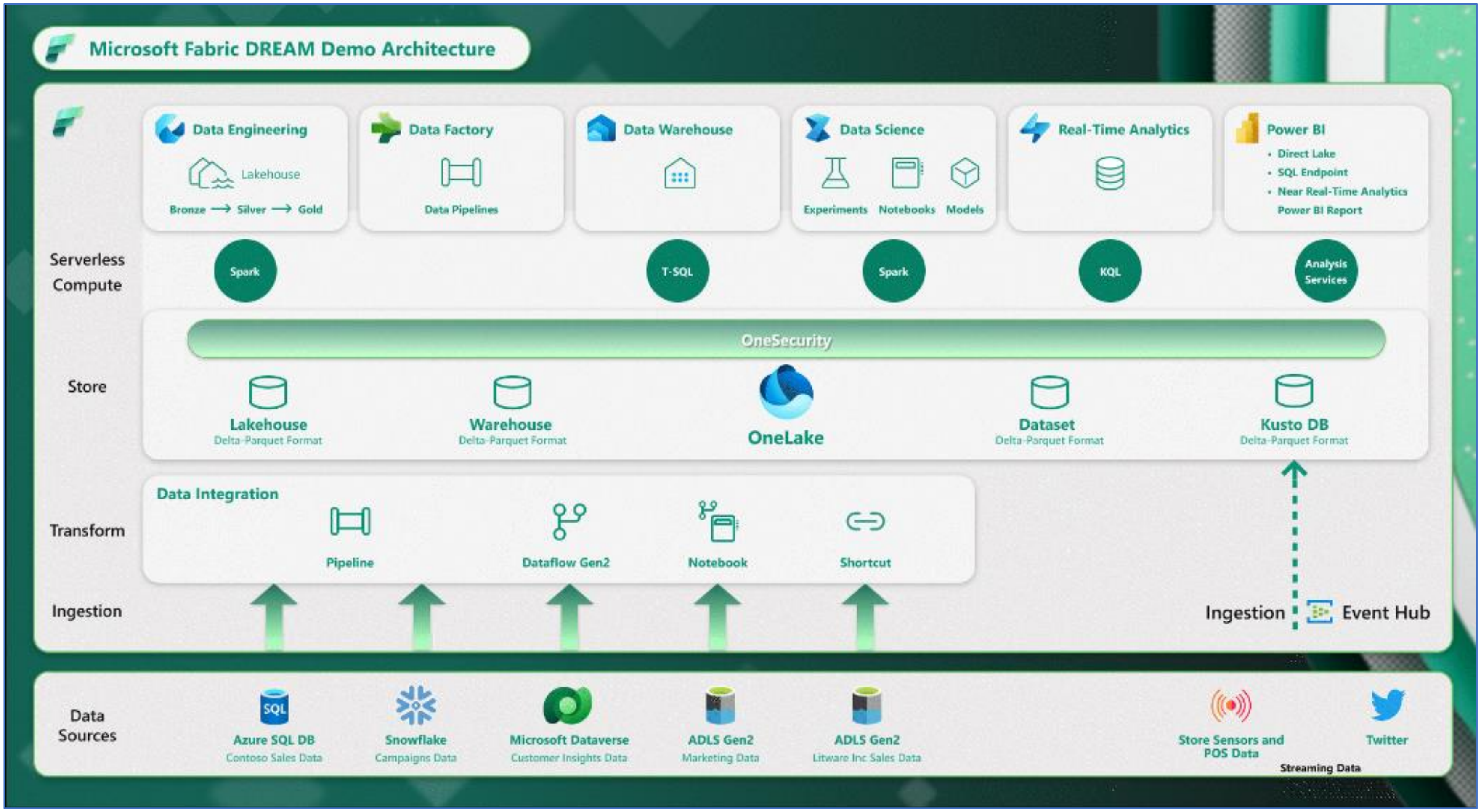
1.3 Demo characters

Introduction	Characters
April is the Chief Executive Officer in charge of driving growth across Contoso. When she took over the role of CEO, she realized that the company was not performing well. She wants to improve their **group wide KPIs, business** outcomes, and implement new ideas.	
Rupesh is the Chief Data Officer (CDO) who oversees a range of data-related functions that include data management, ensuring data quality, and creating data strategy.	
Andre is the Chief Information Officer (CIO) responsible for the management and implementation of information technologies.	

<table>
<tr>
<td>Eva is a Data Engineer who reports to Rupesh, the Chief Data Officer. She builds data pipelines and transforms data into formats that can be easily analyzed by developing, maintaining, and testing infrastructures for data generation.</td>
<td>
</td>
</tr>
<tr>
<td>Miguel is a Data Scientist. He spots machine learning techniques to improve the quality of data or product offerings. He helps the organization make better decisions by developing predictive models for theorizing and forecasting. He reports to the Chief Information Officer, Andre.</td>
<td>
</td>
</tr>
<tr>
<td>Serena is a Data Analyst. She collects, cleans, and interprets data sets to solve business problems. She also puts together visualizations like charts and graphs, writes reports, and presents them to the decision-makers in the organization. She reports to Rupesh, the Chief Data Officer.</td>
<td>
</td>
</tr>
</table>

Wendy is a Business Analyst. She is also responsible for drawing valuable insights from data using the power of analytics and AI.	

1.4 Architecture Diagram



Microsoft Fabric is a complete analytics platform that has everything on it. It's unified as one product and it's SaaS-ified.
It's lake-centric and open. It is like OneDrive for data with all documents and files.
OneLake is in an open standard delta format. It is open on all levels. Built on top of Azure Data Lake Storage Gen2, OneLake supports any type of file, structured or unstructured.
With OneCopy, once you have the data, there is no need to copy anywhere. With OneSecurity it is easy to manage and govern security across the entire organization's data estate. Microsoft Fabric offers Persona Optimized Experiences.

Each one of the personas gets an experience that is optimized for that persona. All of them work together in a single system, but each one of them gets a slightly different experience that is optimized to the things that each persona cares about the most.
In this demo we will:

- Ingest data from a spectrum of sources to get internal department data.
- Map to an external data source via a shortcut.
- Ingest real-time data from IoT sensors.
- Curate the data from bronze to silver to gold using data flows, pipelines, and notebooks.
- Leverage ML to determine key insights for decision-making.
- And arrive at compelling Power BI Reports from Direct Lake.

1.5 Products and technologies showcased:

As a single unified product, Microsoft Fabric offers purpose-built tools enabling data professionals to host all their analytics workloads in a SaaS based lake for solution, eliminating organizational silos. With the right access control at the tenant level, it comes with effortless provisioning. The existing workloads can be seamlessly integrated while exploring the new features. Microsoft Fabric caters to the complete spectrum of experiences including:

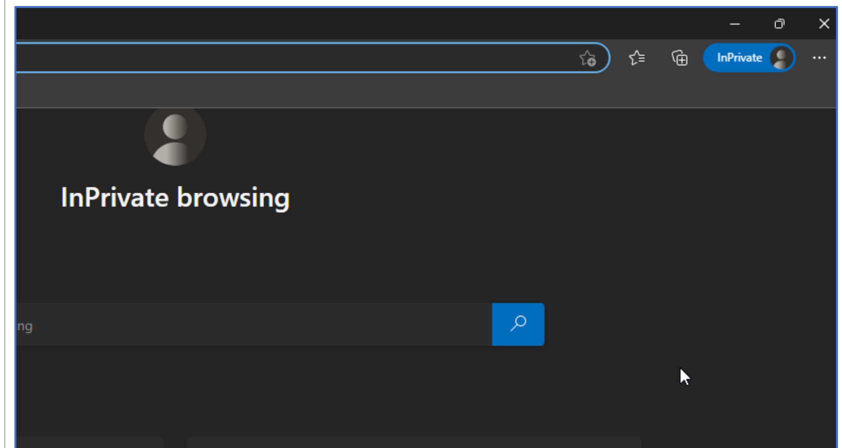
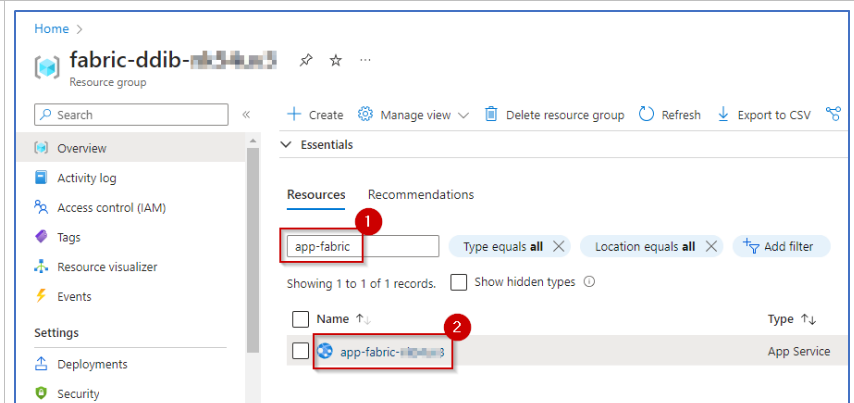
1. **Synapse Data Engineering:**
    - Combines the best of the data lake and warehouse, removing the friction of ingesting, transforming, and sharing organizational data, all in an open format.
    - Users can choose from various ways of bringing data into the Lakehouse including dataflow & pipelines, and they can even use shortcuts to create virtual folders and tables without any data movement between the storage accounts.
    - The goal is to simplify the process of working with organizational data. Rather than spending time on integrating various products, managing infrastructure, and stitching together a spectrum of data sources, Microsoft aims to empower data engineers to focus on their core responsibilities and tasks.
2. **Data Factory:**
    - Data Factory empowers users with a modern data integration experience to ingest, prepare and transform data with intelligent transformations, and leverage a rich set of activities. Data Factory primarily implements dataflows and pipelines.
    - Dataflows provide a low-code interface for ingesting data from hundreds of data sources, with 300+ data transformations.
    - Data pipelines enable powerful workflow capabilities to build complex ETL and data factory workflows that can perform many different tasks at scale, refresh dataflow, move PB-size data, and define sophisticated control flow pipelines.
    - With its fast copy (data movement) capabilities in both dataflows and data pipelines, it enables users to move data between stores blazing fast.
3. **Synapse Data Science:**
    - Microsoft Fabric offers Data Science experiences, empowering users to complete a wide range of activities across the entire data science process. All the way from data exploration, preparation and cleansing to experimentation, modeling, model scoring and serving of predictive insights to BI reports.
    - Synapse Data Science in Microsoft Fabric allows data science practitioners to work seamlessly on top of the same secured and governed data that has been prepared by data engineering teams. This eliminates the need to copy data.
    - The open Delta Lake support allows data science users to version datasets to create reproducible machine learning code.
4. **Synapse Data Warehouse:**
    - Microsoft Fabric introduces a lake centric data warehouse built on an enterprise grade distributed processing engine that enables industry leading performance at scale while eliminating the need for configuration and management.

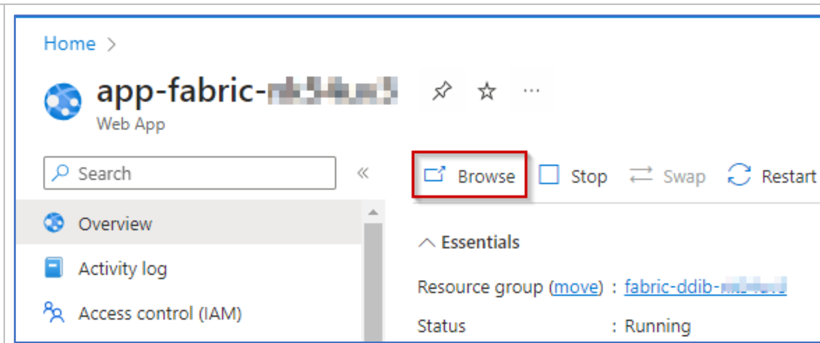
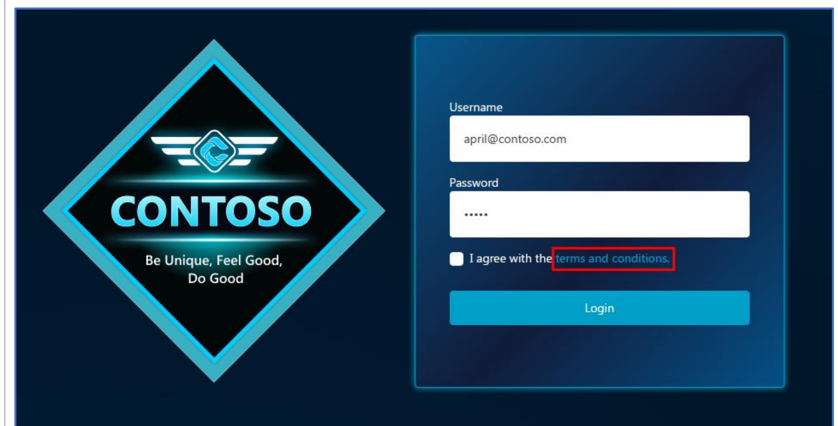
- The Warehouse is built for any skill level - from the citizen developer through to the professional developer, DBA or data engineer. The rich set of experiences built into Microsoft Fabric workspace enables customers to reduce their time to insights by having an easily consumable, always connected dataset that is integrated with Power BI in Direct Lake mode.

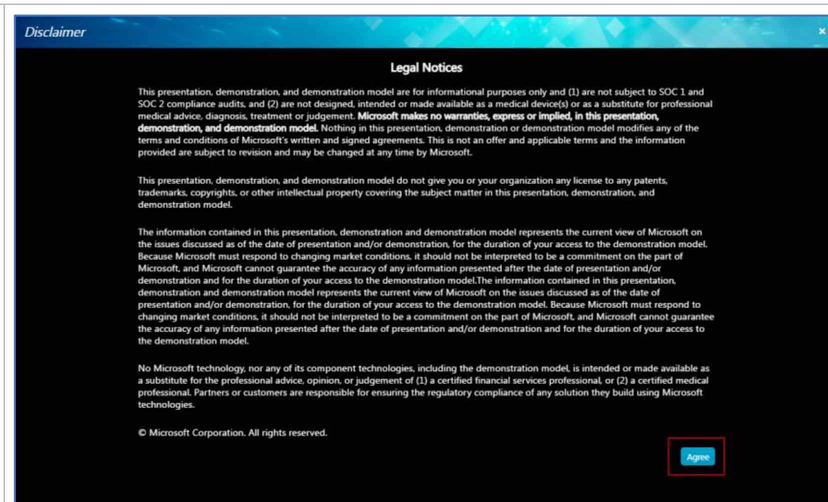
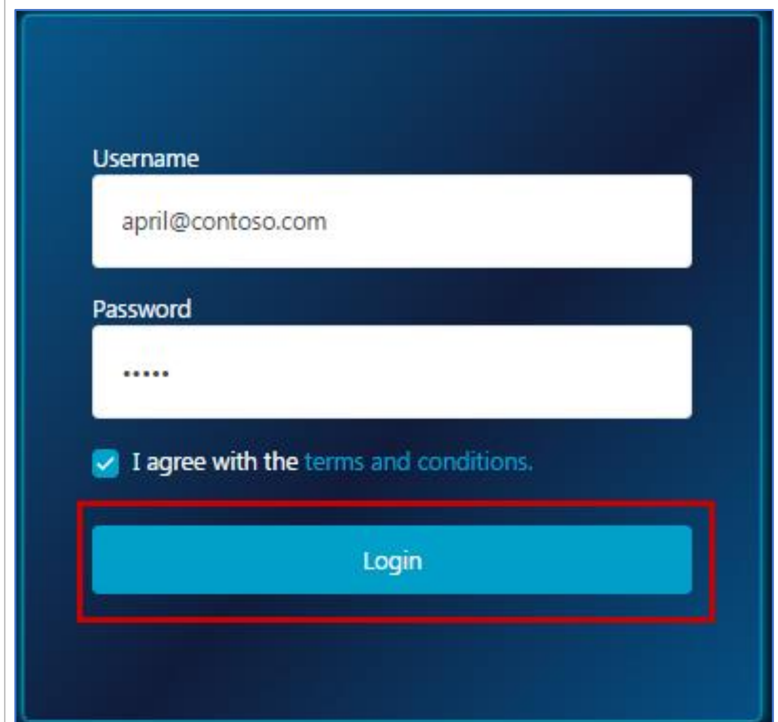
5. **Synapse Real-Time Analytics:**
    - Real-Time Analytics is fully integrated with the entire suite of Microsoft Fabric products, for both data loading, data transformation, and advanced visualization scenarios. Quick access to data insights is achieved through automatic data streaming, automatic indexing and data partitioning of any data source or format, and by using the on-demand query generation and visualizations.
    - The main items available in Real-time Analytics include:
        - **Eventstream** for capturing, transforming, and routing real-time events to various destinations with a no-code experience.
        - **A KQL database** for data storage and management. Data loaded into a KQL database can be accessed in OneLake and is exposed to other Microsoft Fabric experiences.
        - **A KQL queryset** to run queries, view, and customize query results on data. The KQL queryset allows you to save queries for future use, export, and share queries with others. It includes the option to generate a Power BI report.
6. **Power BI:**
    - Power BI is standardizing open data formats by adopting Delta Lake and Parquet as its native storage format to avoid vendor lock-in and reduce data duplication and management. Direct Lake mode unlocks incredible performance directly against OneLake, with no data movement.
    - Power BI datasets in Direct Lake mode enjoy query performance on par with import mode, with the real-time nature of DirectQuery. And the data never leaves the lake.

2. Demo

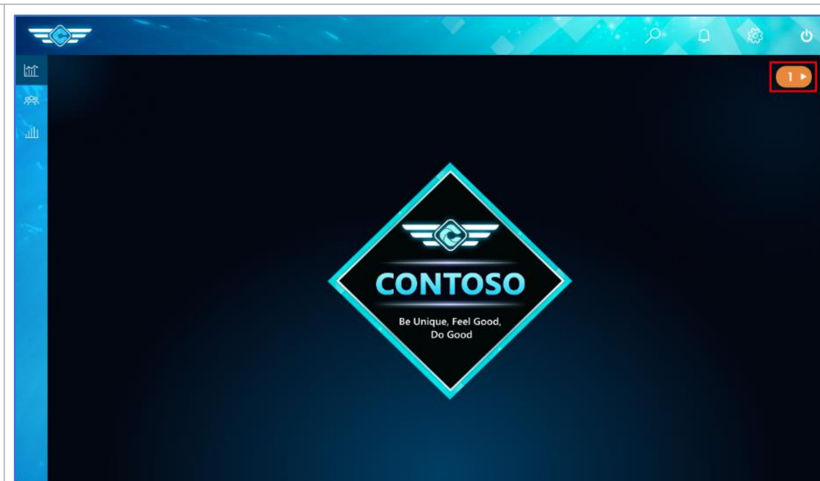
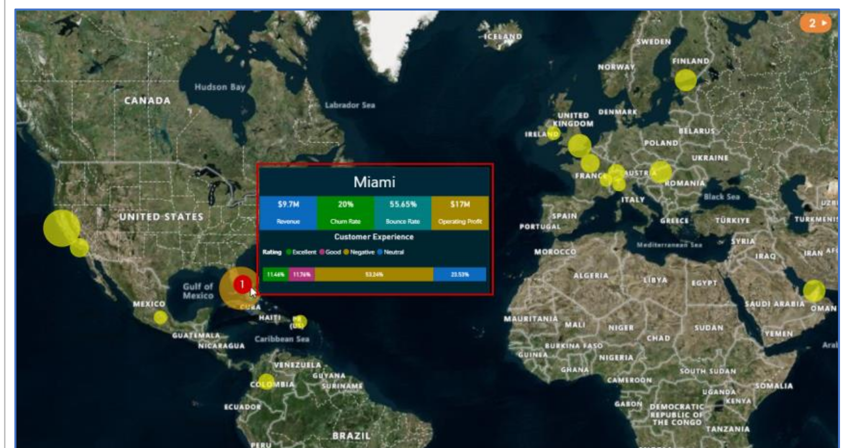
2.1 Demo Introduction

Narrative	Steps	Screenshot
For a seamless demo experience, open a new browser in **InPrivate** mode.	1. **! Open** a new browser window in InPrivate Mode. 2. **Open** portal.azure.com.	
Let's get started by navigating to the **Contoso** web application.	3. **Navigate to** the resource group you created. 4. In the search box **type** "app-fabric" and **click** on the app service resource.	

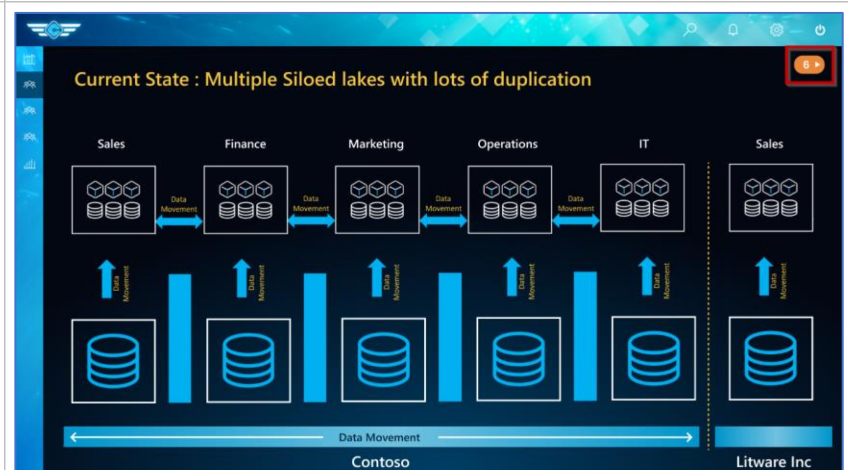
Then, click on Browse and wait for the main web app to load.	5. In the resource window **click** on Browse and wait for the main web app to load.	
Now, let's look at the terms and conditions.	6. **Click** on terms and conditions.	

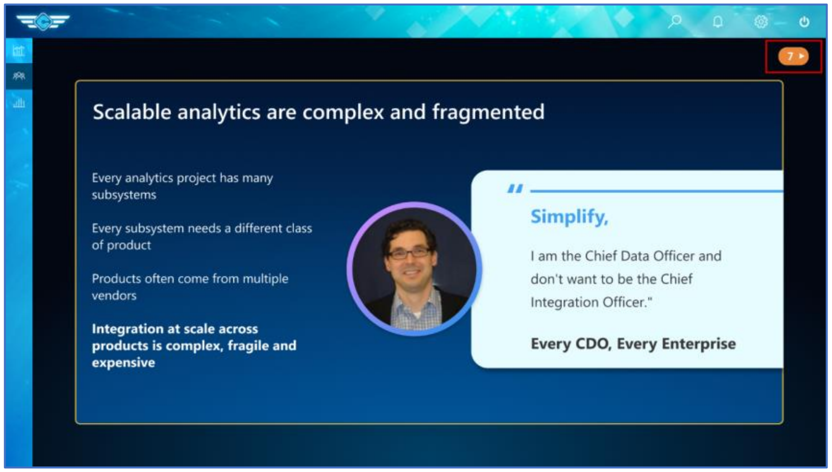
Here is the CELA approved Legal Notice. **Note:** It is important for you to show this during the demo.	7. **Show** the CELA approved legal notice. 8. **Click** Agree to accept the terms and conditions.	
Let's log in as April, the new CEO of Contoso.	9. **Click** Login.	

<table>
<tr>
<td>Contoso is a large enterprise with a presence in various industries such as retail, manufacturing, finance, and media.

While this demo illustrates the transformation they implement in their retail business, it is equally applicable to other industries. Contoso's retail business has hundreds of brick-and-mortar stores. They also have an online store.</td>
<td>10. Click on arrow 1.</td>
<td>
</td>
</tr>
<tr>
<td>Here, April is happy to see a high-level view of how all their stores are performing. She can easily hover on top of each location to see the top 5 KPIs that she cares about the most.
April hovers over the Miami store to see its 5 KPIs including:
Revenue
Churn Rate
Bounce Rate
Operating Expense
And Customer Experience</td>
<td>11. Hover over the orange bubble to show the five main KPIs for the Miami store.
12. Click on arrow 2.</td>
<td>
</td>
</tr>
</table>

And here she's in Miami, South Beach, where Contoso's headquarters is located.	13. **Click** on arrow 3.	
April is a data-driven CEO. She dreams of a dashboard like this. One that uses structured, unstructured, and semi-structured data from the past to get meaningful insights for the present moment. For example, the Campaign and Social Media Analytics pillar here. But that's been done before. She wants to utilize real-time analytics to make decisions for the next moment such as the Location analytics KPIs related to foot traffic at various locations. But even more exciting, she dreams of getting predictive and prescriptive insights into the future. She wants to see predicted customer experience including satisfaction and predicted churn. How cool is this? This demo starts on January 30th. Sadly, April can see that Contoso's KPIs are struggling.	14. **Point** to ROI and Customer Churn. **#1** 15. **Point** to Location Dwell Time and Average Foot Traffic. **#2** 16. **Point** to Forecast Customer Satisfaction and Historical and Predicted Churn Volume. **#3** 17. **Click** on arrow 4.	

Revenue, Churn Rate, Bounce Rate, Operating Expense, and Customer Experience are not doing so well. She is determined to transform these KPIs for Contoso before their upcoming Memorial Day sale. So, she talks to her team of executives.		
This is April and her team. She'll work with them to improve Contoso's struggling KPIs. April asks Rupesh, the Chief Data Officer (CDO), "How can we become a data-driven organization?" Rupesh looks at the current state of their Architecture.	18. **Click** on arrow 5.	
He sees that Contoso faces a few major problems. First, the world is complex, especially around the analytics system. Each department at Contoso has its data in silos. As a result, there is a huge dependency on integration and data movement across the organization. To add to all this, Contoso has acquired a company called LitWare Inc. This was a strategic move on Contoso's part to add	19. **Click** on arrow 6.	

brands popular with millennials that are missing from their catalog. Rupesh knows that the integration challenges with ingesting LitWare Inc.'s data has already delayed the acquisition project by 6 months! So, let's see what's at the top of Rupesh's mind, given these challenges.		
One statement we hear time and again, which is true for the Contoso's CDO, Rupesh too is, "I'm the chief data officer. I don't want to be the chief integration officer." Currently, there is so much dependency on data movement for Contoso. The Contoso team knows that there are just so many potential issues they're facing. With the possibilities of: • Inconsistencies • The worry of governing and protecting warehouse data • The threat of malicious attacks • The costs associated with procuring and managing these capabilities • Rigid and costly architectures • And the reduction of business agility and speed to insights	20. **Click** on arrow 7.	

On this CDO Metrics - Current State scorecard, we can see how the major KPIs translate to the CDO level KPIs.
Let's drill down below those executive KPIs to see how CDO level metrics impact the executive level KPIs.

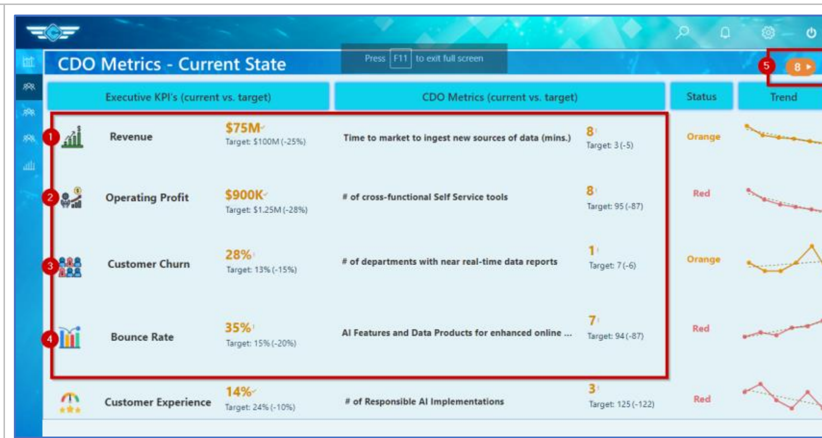
For example,

1. Slow Time to market to ingest new sources of data is affecting Revenue (due to the delayed acquisition of LitWare Inc).
2. The lack of cross-functional self-service tools is increasing the operational expenses, thereby reducing Operating Profit.
3. Since there is a lack of near real-time reports, customers' needs in stores can't be met. For example, temperature anomalies in stores go undetected. This causes customers to churn.
4. And the lack of AI Features and Data Products for enhanced online experience is causing a high Bounce Rate!

Rupesh knows he needs to solve this quickly. All before their Memorial Day sale!

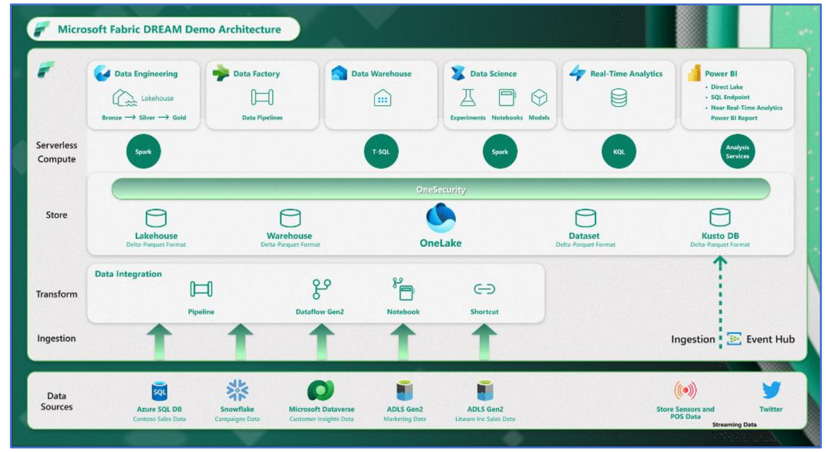
Let's see the challenges that he and his team face.

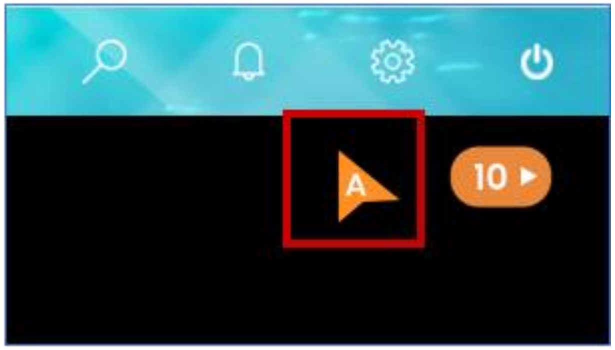
21. **Point** to Revenue (marked #1).
22. **Point** to Operating Profit (marked #2).
23. **Point** to Customer Churn (marked #3).
24. **Point** to Bounce Rate (marked #4).
25. **Click** on arrow 8.

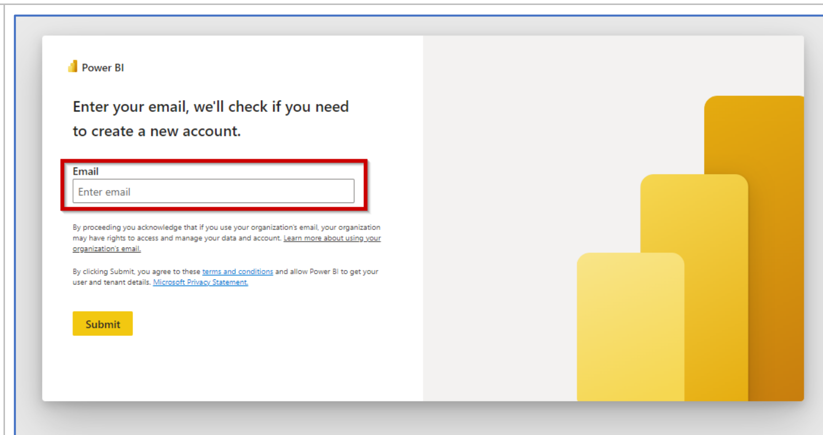
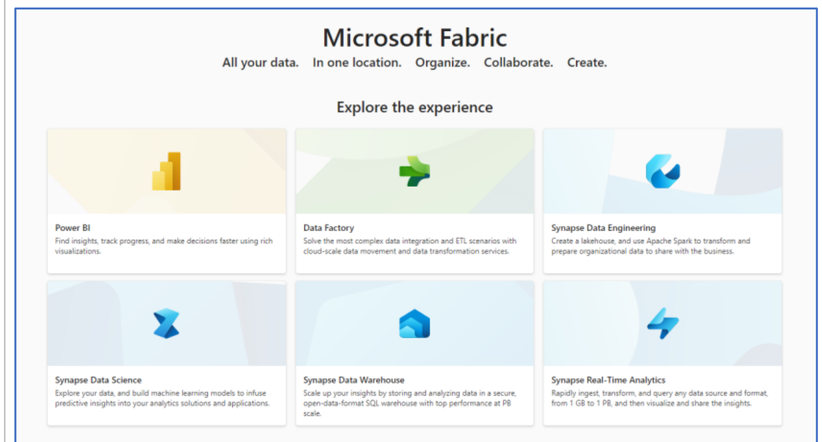


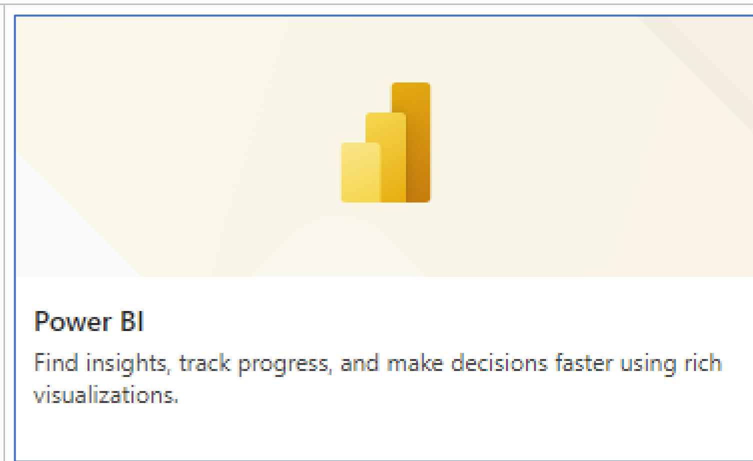
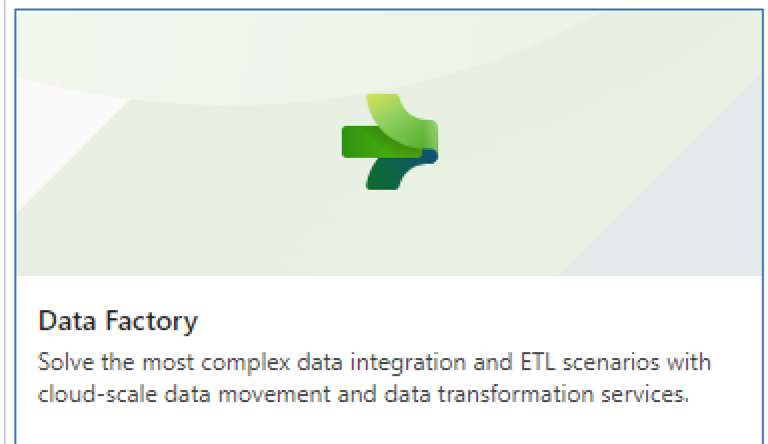
2.2 Introducing Microsoft Fabric

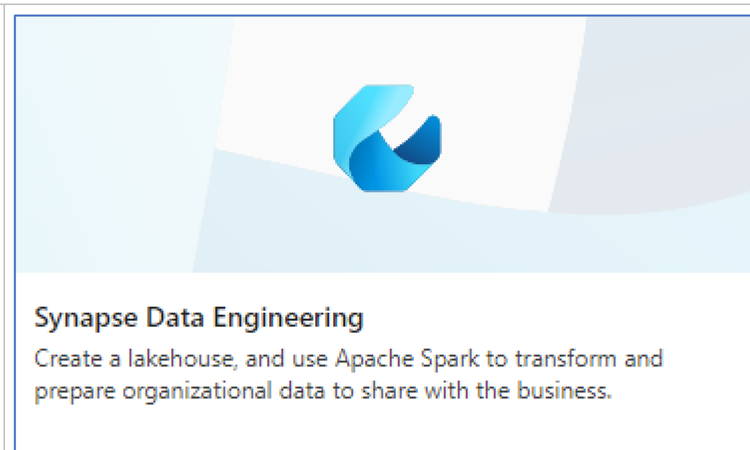
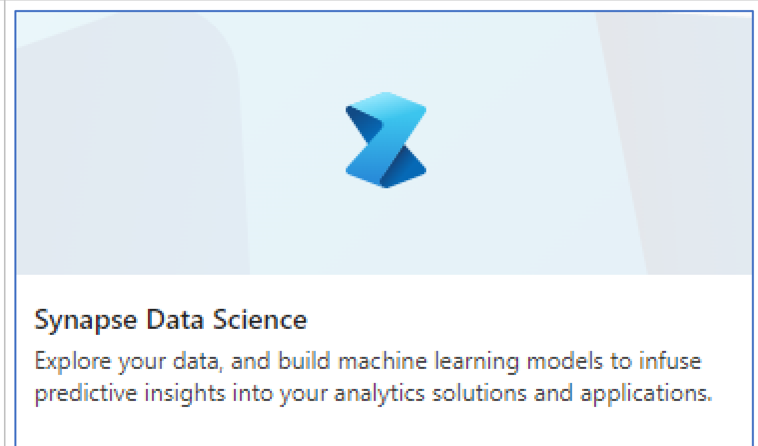
Narrative	Steps	Screenshot
When Rupesh analyzes the block diagrams here, he can see that the same components show time and time again. He sees Data Integration, Data Engineering, Data Warehousing, Real-time Analytics, Data Science, BI, Data Lake, Governance, and Administration. Today, Contoso has all these products, but it's putting them together that's holding the team back. Rupesh knows there's a lot of wiring to do because each product was designed with different design points, architectures, and so on. Luckily, Contoso got an opportunity to be part of the Microsoft Fabric public preview. Microsoft Fabric takes all these products involved in analytics and puts them together in a single, unified analytics fabric. Microsoft Fabric enables data professionals to host all their analytics workloads in a SaaS-based, lake-first solution—eliminating organizational siloes. It covers the complete spectrum of services including data ingestion, data lake, data	1. **Point** to the blocks in the diagram. 2. **Click** on arrow 9.	

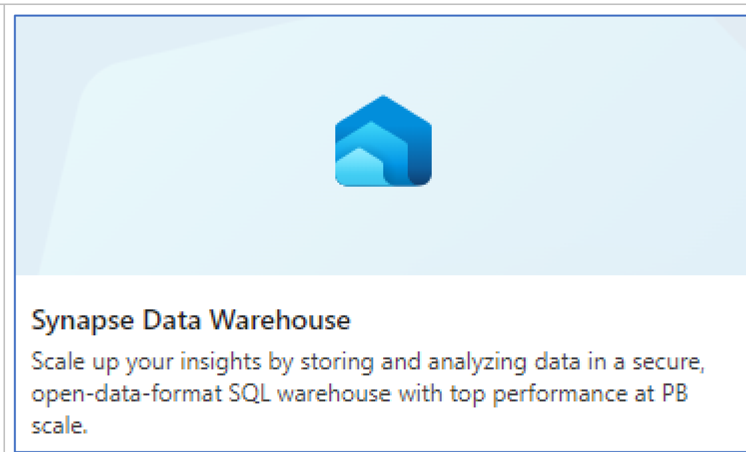
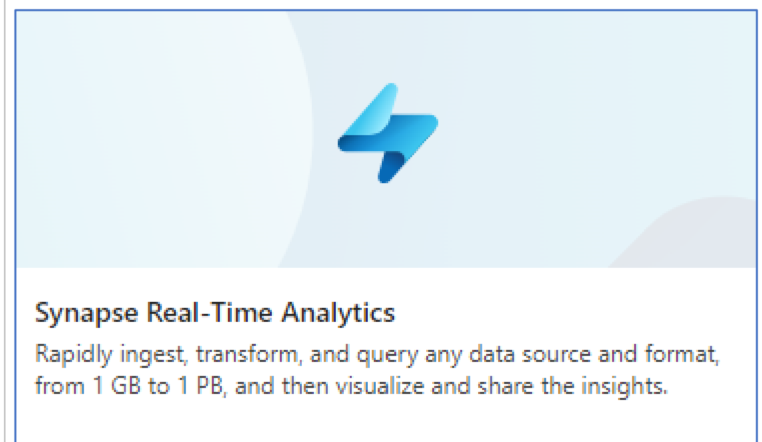
<table>
<tr>
<td>engineering, data integration, data science, real-time analytics, and BI.
Contoso can integrate their existing workloads seamlessly while exploring the new features.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Here’s the architecture of Microsoft Fabric used in this DREAM Demo.
Microsoft Fabric is a complete analytics platform that has everything in it. It's unified as one product and it’s SaaS-ified.
It’s lake-centric and open. It is like OneDrive for data with all documents and files.
OneLake is in an open standard delta format. It is open at every level. Built on top of Azure Data Lake Storage Gen2, OneLake supports any type of file, structured or unstructured. With OneCopy, once you have the data, there is no need to copy anywhere. With OneSecurity it is easy to manage and govern security across the entire organization’s data estate.
Microsoft Fabric offers Persona Optimized Experiences.
Each one of the personas gets an experience that is optimized for that persona. All of them work together in a single system, but each one of them gets a slightly different experience that is optimized to the things that each Persona cares about the most.
The CDO’s team leverages this architecture to do the following for Contoso:
1. Ingest data from a spectrum of data sources.</td>
<td>3. Talk about the architecture diagram.</td>
<td>
</td>
</tr>
</table>

2. Map to LitWare Inc.'s data via a shortcut. 3. Ingest real-time data from Contoso's stores. 4. Curate the data from bronze to silver to gold using data flows, pipelines, and notebooks. 5. Leverage ML to determine key insights like Customer Churn and 6. Sales Forecasting. 7. Create compelling Power BI Reports from Direct Lake. Now let's look at the Persona Optimized Experiences involved in the architecture above.		
With Contoso learning about the benefits of using Microsoft Fabric, it's no surprise that they are excited to start exploring this amazing new tool.	4. **Click** on arrow A to navigate to Microsoft Fabric. **<<Switching to Microsoft Fabric Login page>>**	

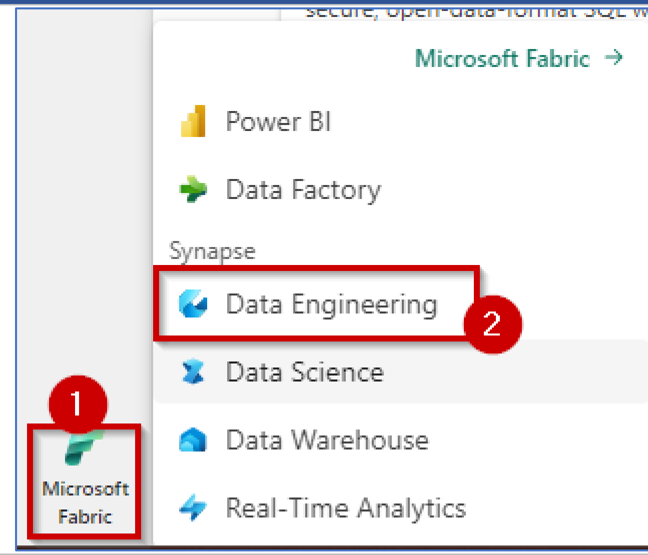
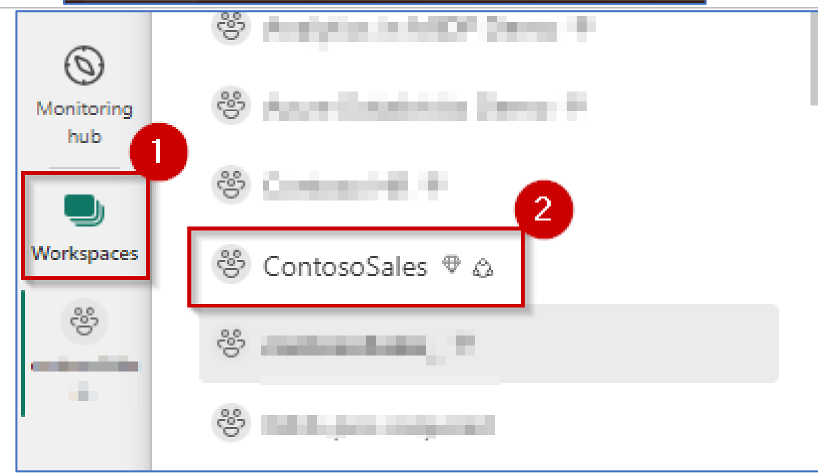
Use the Azure Credentials under the Environment details to sign in. Let’s sign into Microsoft Fabric and navigate to the Microsoft Fabric Home page.	5. **Sign in** using Azure Credentials. ***Note:*** *If you are asked to sign in follow this step otherwise you will directly land on the Microsoft Fabric Page.*	
Here we see a glimpse of the power of Microsoft Fabric, depicting a single SaaS-ified product supporting all analytical workloads for Contoso. Microsoft Fabric enables Contoso’s data professionals to host all their analytics workloads in a SaaS-based Lake solution which eliminates organizational silos. It covers a complete spectrum of experiences including Power BI, data engineering, data science, data warehouse, and real-time analytics. You can integrate your existing workloads seamlessly while exploring the new features. Before we move on to the demo, let's take a quick look at each of these experiences.	6. Welcome to Microsoft Fabric.	

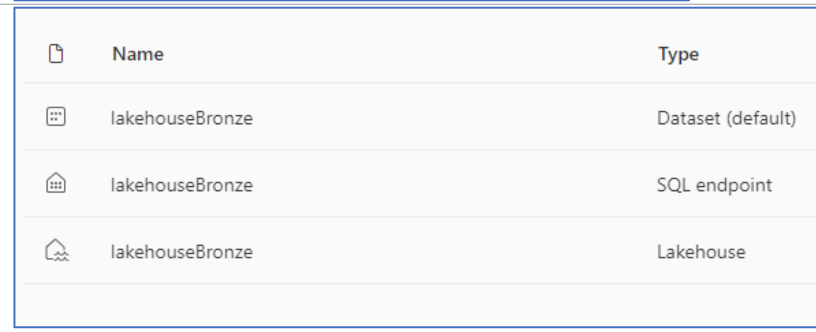
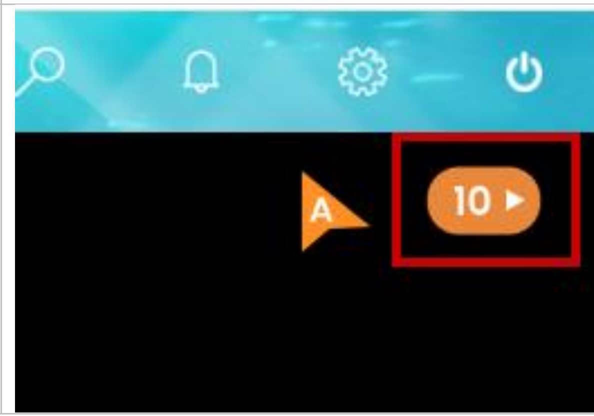
Power BI users can continue to enjoy all the functionalities Microsoft Fabric offers. Power BI Premium users can simply turn on the Microsoft Fabric tenant setting in the admin portal. With Microsoft Fabric's unified capacity model, Power BI Premium capacity can be utilized by any of the new workloads.	7. **Talk** about Power BI experience in Microsoft Fabric.	
Data Factory in Microsoft Fabric brings together the best of Power Query and Azure Data Factory into a single, modern data integration experience. This empowers both data and business professionals with capabilities to ingest and transform data as well as orchestrate data workflows. Data Factory in Microsoft Fabric empowers users with seamless connectivity to more than 170+ data stores (including on-premises data sources, cloud databases, analytical platforms, line of business applications, and more).	8. **Talk** about Data Factory experience in Microsoft Fabric.	

Microsoft Fabric provides diverse data engineering capabilities to ensure that organizations can design, build, and maintain infrastructures and systems that enable them to collect, store, process, and analyze large volumes of data. Users can: • Create and manage data using a Lakehouse. • Design pipelines to copy data into the Lakehouse. • Use Spark Job Definitions to submit batch or streaming job to Spark cluster. • Use notebooks to write code for data ingestion, preparation, and transformation.	9. **Talk** about Data Engineering experience in Microsoft Fabric.	
Microsoft Fabric offers Data Scientists a platform with complete end-to-end data science workflows for the purpose of data enrichment and business insights. They can perform data exploration, preparation and cleansing to experimentation, modeling, model scoring, and serving predictive insights to BI reports. Working on the same platform, sharing and collaboration becomes so much more seamless. Analysts can easily share Power BI reports and datasets with data science practitioners.	10. **Talk** about Data Science experience in Microsoft Fabric.	

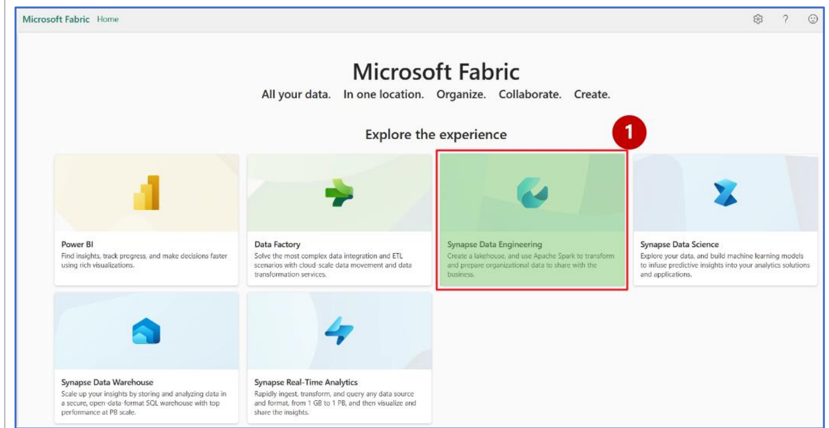
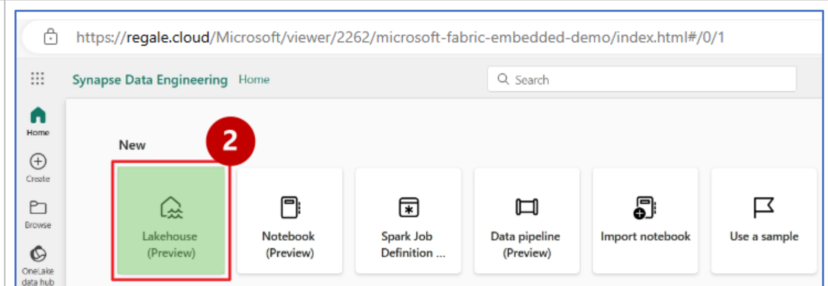
Microsoft Fabric introduces a lake-centric data warehouse built on an enterprise grade distributed processing engine that enables industry leading performance at scale while eliminating the need for configuration and management. Data warehousing workloads benefit from the rich capabilities of the SQL engine over an open data format. This allows customers to focus on data preparation, analysis, and reporting over a single copy of their data. All of which is stored in their Microsoft OneLake. The Warehouse is built for any skill level - from the citizen developer through to the professional developer, DBA, or data engineer.	11. **Talk** about Data Warehouse experience in Microsoft Fabric.	
Real-Time Analytics is a fully managed big data analytics platform optimized for streaming, and time-series data. It utilizes a query language and engine with exceptional performance for searching structured, semi-structured, and unstructured data. It is fully integrated with Microsoft Fabric, for data loading, data transformation, and advanced visualization scenarios. Let's dig deeper and see how Rupesh's team utilizes this tool to turn everything around.	12. **Talk** about Real-Time Analytic in Microsoft Fabric.	

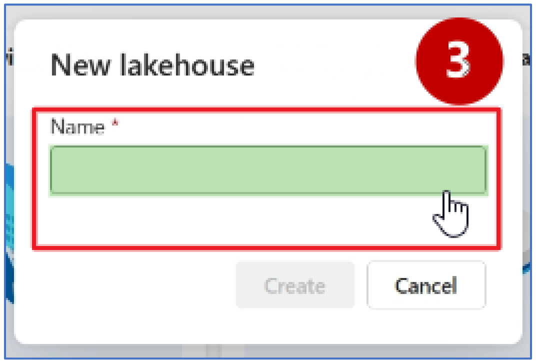
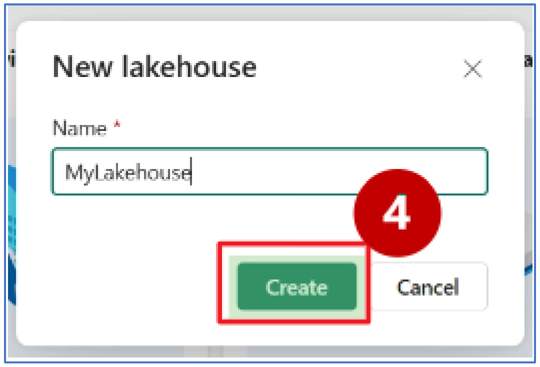
2.3 Microsoft Fabric for Data Engineering experience

Narrative	Steps	Screenshot
Let's see how Eva, the Data Engineer, begins preparing data for the merger between Contoso and LitWare Inc. With Microsoft Fabric, she logs in and simply selects the Data Engineering experience.	1. From Microsoft Fabric home **select** 'Synapse Data Engineering' experience. OR **Switch** to the Data Engineering experience using the experience switcher icon at the left corner.	
Here, we can see the workspaces for each department at Contoso like Sales and Finance. Offering a bird's eye view of the various data products. All this is ready for data ingestion without any data duplication. How cool is that. Since Eva belongs to the Sales department, she navigates to the Contoso Sales workspace.	2. From Workspaces, **select** the 'ContosoSales' workspace.	

In this workspace, Eva can either start working with a previously created asset for which she has permission, or she can create a new one.	3. **Show** listing of assets. 4. In the assets search bar, **type** "lakehouseBronze".	
When a new Lakehouse is created, it comes with 3 components, a Lakehouse, a SQL endpoint, and a default dataset. Each of these includes the same name given to the Lakehouse during creation. Now let's see how easy it is for Eva to create a Lakehouse through a click-by-click simulation in our web app.	5. **Talk** about 3 things, the lakehouse, SQL endpoint, and the dataset. **<<Switch back to the demo web app>>** 6. **Navigate** back to the web app to create a new lakehouse using the click-by-click simulation.	
Let's navigate to the next screen to create a new lakehouse using a click-by-click.	7. **Click** on arrow 10. **<<This step will open a click-by-click embedded in the web app>>**	

3.3.1 Lakehouse [Click -by - click – Lakehouse Creation]

Narrative	Steps	Screenshot
First, she goes to the Synapse Data Engineering area to create the first Data Lakehouse.	**<< Click-by-click embedded in the web app >>** 1. **Click** on the Synapse Data Engineering experience. ***Note:*** *For the purpose of this demo, lakehouses are already created in the workspace.* *To demonstrate lakehouse creation, a click-by-click is embedded in the web app to simulate the actual user experience.*	
In the Data Engineering experience, Eva starts by simply selecting the Lakehouse option to get started. The Synapse Data Engineering Lakehouse combines the best of the data lake and warehouse. This makes ingesting, transforming, and sharing organizational data, seamless. It is all in an open format and makes Lakehouse a first-class item in the workspace.	2. **Click** on Lakehouse (Preview).	

Now Eva must provide a name for the New lakehouse. So, she clicks on the name field and gives it a name.	3. **Click on** "Name*" in the New Lakehouse window.	
All Eva has to do is provide a name for the lakehouse. There's no provisioning needed, assuming Eva has all the access rights of course!	4. **Click on** Create. **Note:** This is a click-by-click step only. Please do not create an actual lakehouse in Microsoft Fabric to avoid multiple asset creation.	

Now, Eva has several options she can choose to start with data ingestion. She can start ingesting data using one of the four options. She can create a New Dataflow Gen2, New data pipeline, open a new notebook to work with code, or create a new shortcut to any existing source.

Once the data is ingested, she can access the data files and folders using the Explorer panel, on the left.

The top menu allows her to get data, create a new Power BI report, or open a notebook and start coding right away.

Once the lakehouse is created in Microsoft Fabric, it automatically creates its respective SQL Endpoint which points to the Lakehouse delta table storage and a default dataset.

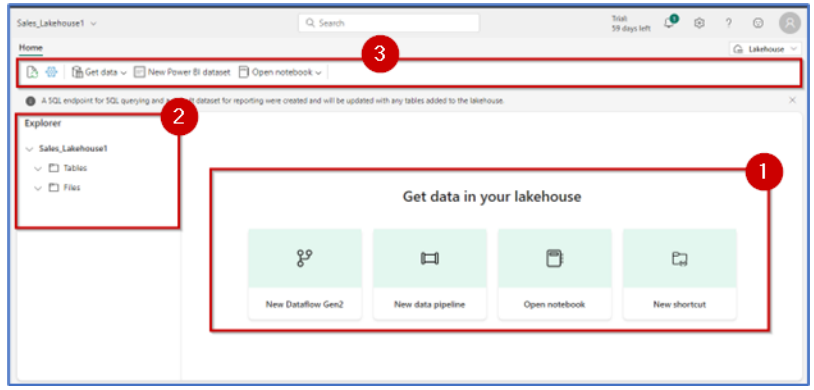
In just a few seconds the Lakehouse is ready.

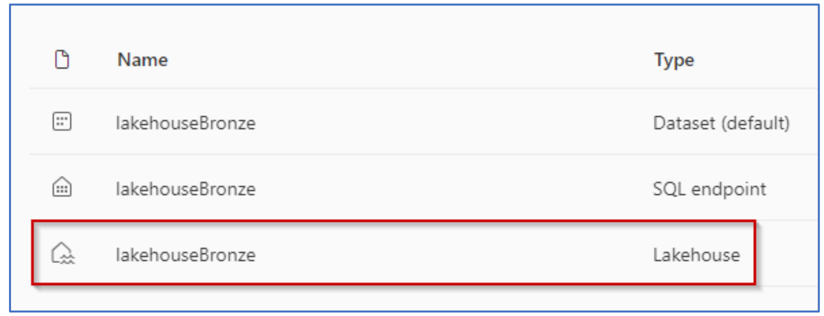
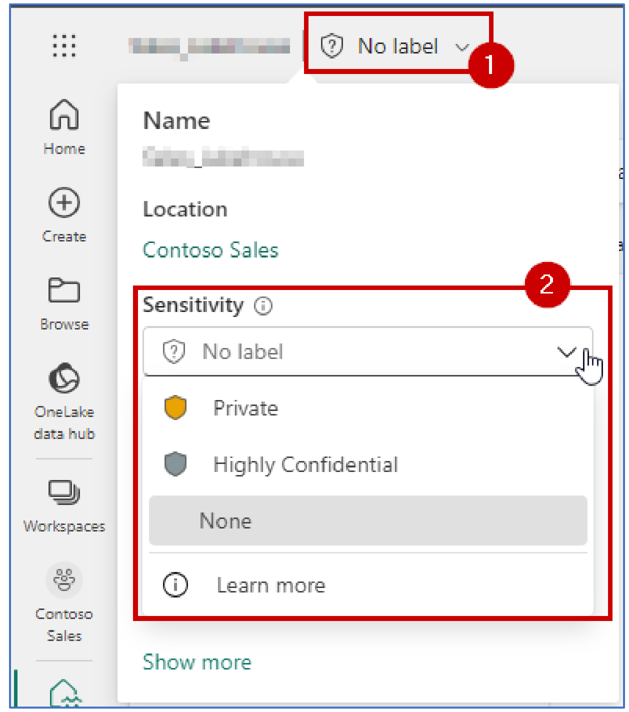
With the right access control, Eva has effortlessly created a new lakehouse. She didn't have to spin up any storage accounts, or answer questions about the network, infrastructure, key vault, Azure subscriptions, and so on.

It is so FRICTIONLESS and ready for Eva!

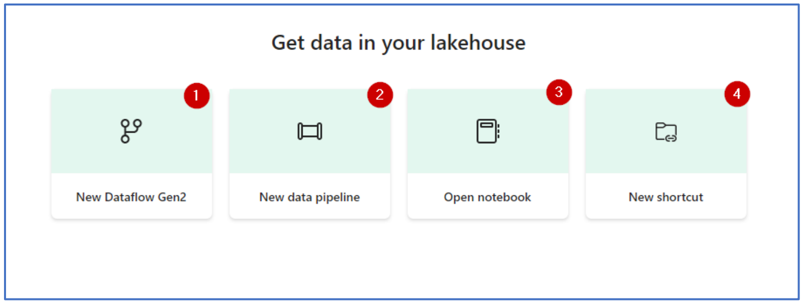
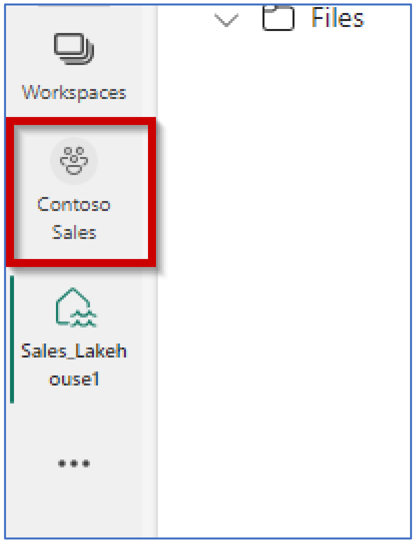
5. **View** the Lakehouse and the panels that help a user get started.

<<Switch browser tab to Microsoft Fabric *(opened earlier)*. >>

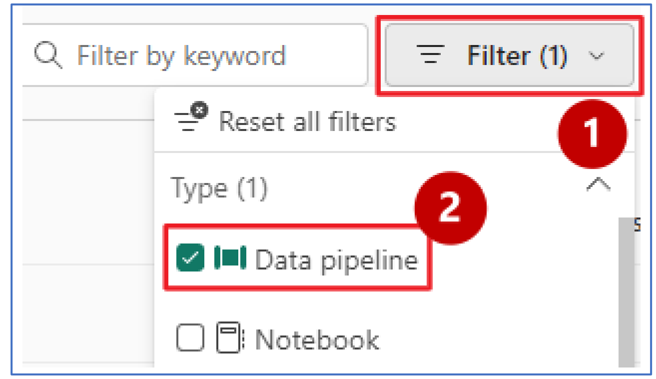
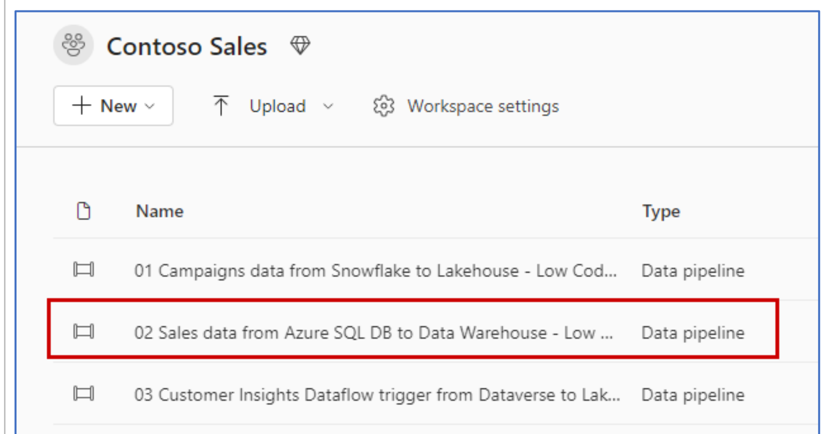


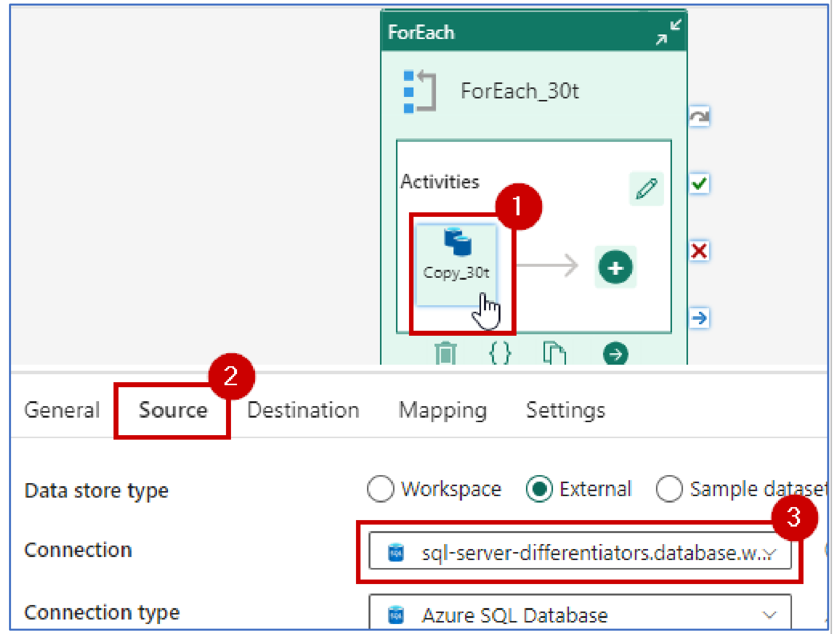
Let's switch to our pre-created lakehouse in the Contoso Sales workspace.	6. **Click** on 'lakehouseBronze'.	
Eva can assign a sensitivity label to the new lakehouse so that all the assets in the lakehouse inherit the same labels without any extra provisioning. Prior to Microsoft Fabric, Contoso had to do a lot of configuration setup to ensure proper data labelling and information protection. But not anymore! That's OneLake and OneSecurity in action for you!	7. **Click** on the dropdown to change the Sensitivity label of the lakehouse. **Note:** This screen may appear different based on labels in your environment.	

3.3.2 Get data in your lakehouse

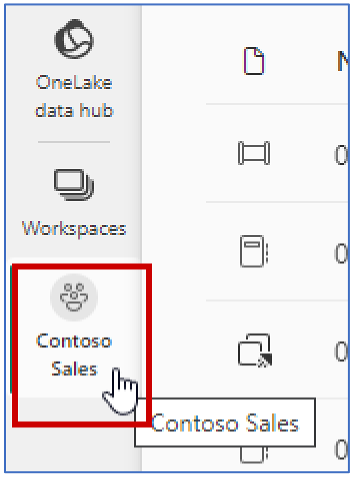
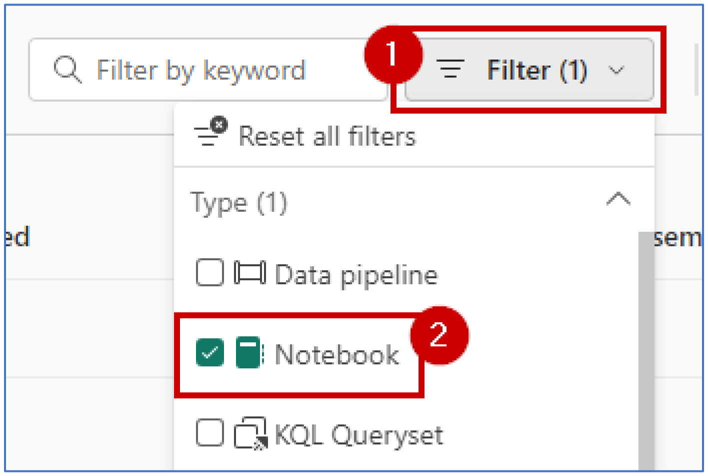
Narrative	Steps	Screenshot
Once the lakehouse is created, Eva can start ingesting data into this lakehouse from various data sources. Microsoft Fabric provides four options to ingest data. So, she can: 1. Use **Dataflows** with Low Code or No Code experience. 2. Use **data pipelines** without writing a single line of code. 3. Use **notebooks** with code-first experience. 4. Or she can opt for a **New shortcut**.	1. **Show** the get data options.	
For the purpose of this demo, we will go through the pre-created assets to look at each of these options one by one.	2. In the left navigation bar, **click** on the Contoso Sales workspace.	

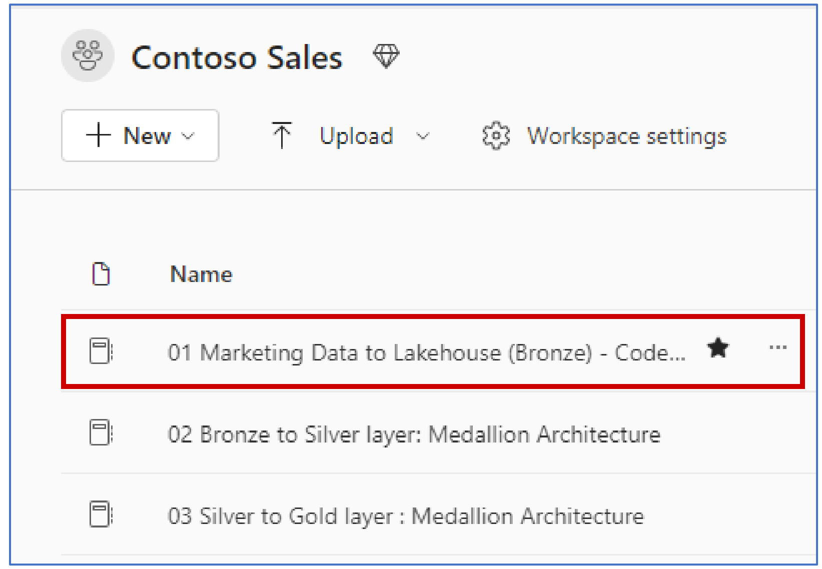
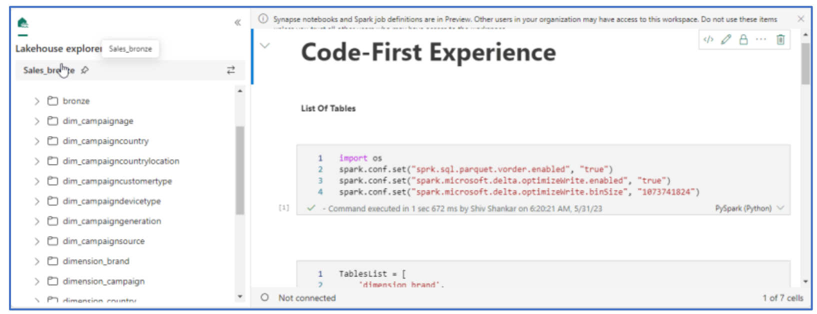
A. Data Pipelines 'No Code-Low Code experience'

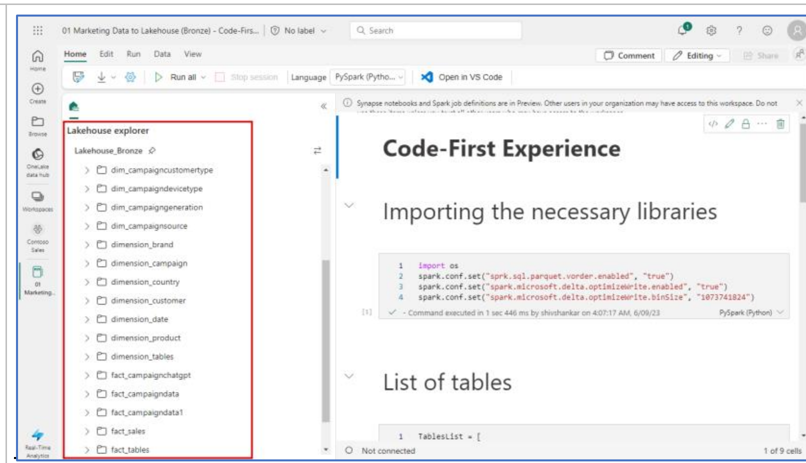
Narrative	Steps	Screenshot
Let's look at the details for ingesting the data into the lakehouse in the Data Pipeline. Eva has already created some data pipelines to ingest data from a spectrum of sources. Using the filter, she can quickly navigate to the assets.	1. **Click** on Filter. 2. **Select** Data pipeline.	
The data pipeline Eva wants to set up will be configured to bring Contoso Sales transaction data from Azure SQL DB to Data Warehouse using the Low Code Experience.	3. **Select** the pipeline "02 Sales data from Azure SQL DB to Data Warehouse – Low Code Experience".	

Here, Eva connects to Azure SQL DB using the database connection string. After establishing the connection, the Copy activity starts copying data from the database to the lakehouse. Eva has a few more data sources to ingest from before she can move on to the next process.	4. **Click** on Copy_30t activity inside the for each loop. 5. **Click** Source. 6. **View** the connection string.	

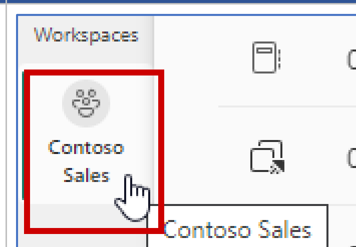
B. Spark Notebook 'Code-first experience'

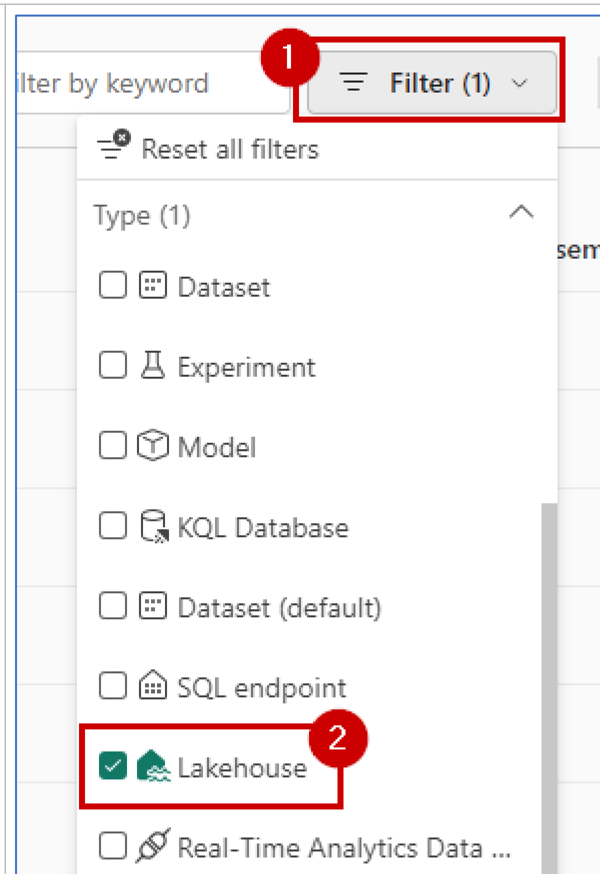
Narrative	Steps	Screenshot
Let's take a look at another option for ingesting the data. This time Eva prefers using the code-first experience. Let's go back to our workspace and navigate to that notebook.	1. **Click** on the Contoso Sales workspace in the left navigation bar.	
Let's filter our results based on notebooks.	2. **Click** Filter. 3. **Select** Notebook.	

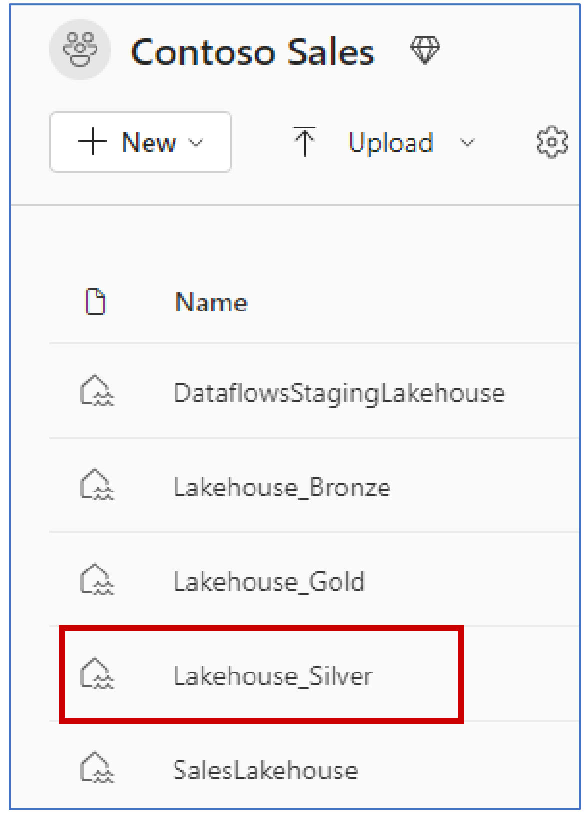
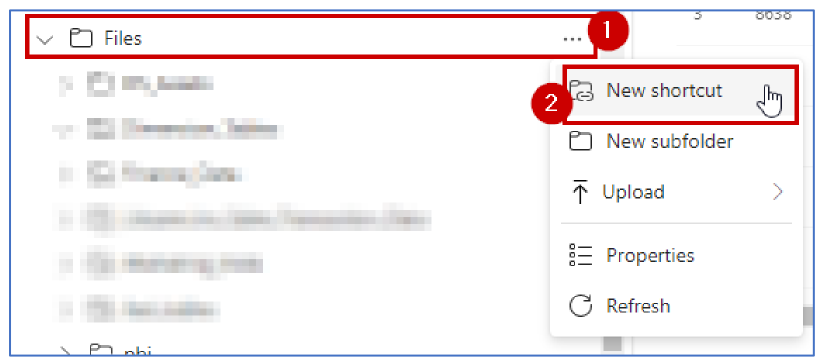
Eva would like access to the Marketing data so she can understand the campaigns and their impact on Contoso's revenue. She's created a python notebook to connect to the Marketing data. Let's open this notebook to understand the steps involved.	4. **Select** "01 Marketing Data to Lakehouse (Bronze) – Code First Experience".	
Eva is a seasoned data engineer. The code-first experience offers her a platform to apply her coding expertise. The code-first experience allows her to get started with the data and modeling. She can create new notebooks, import necessary libraries, and define her own functions to get the data from different sources. So, after importing the necessary libraries, she adds python code for the data she wants to access. Then she writes a couple of functions to read and load data from the data source to the bronze layer of the lakehouse.	5. **Review** the notebook, cell by cell.	

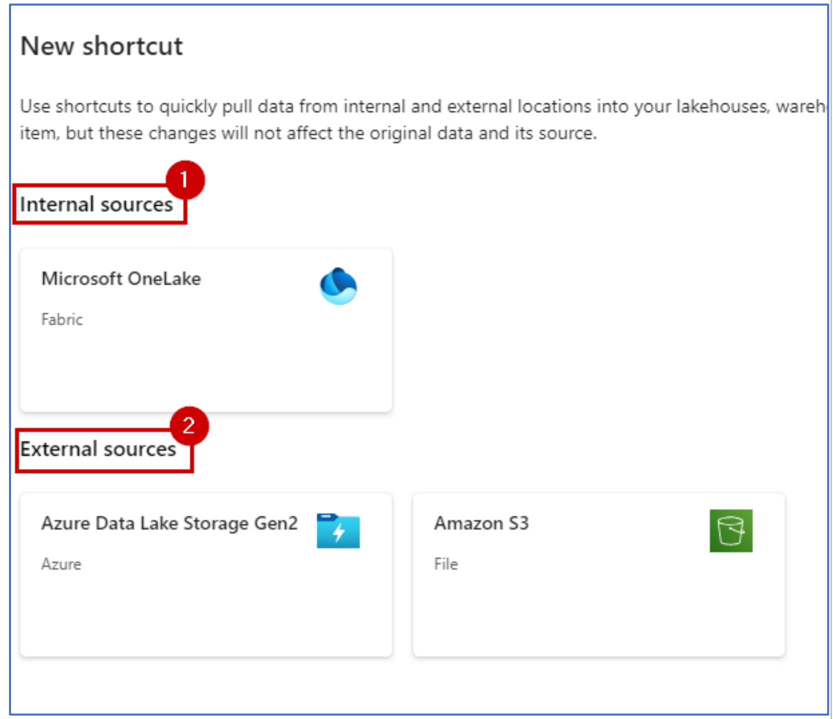
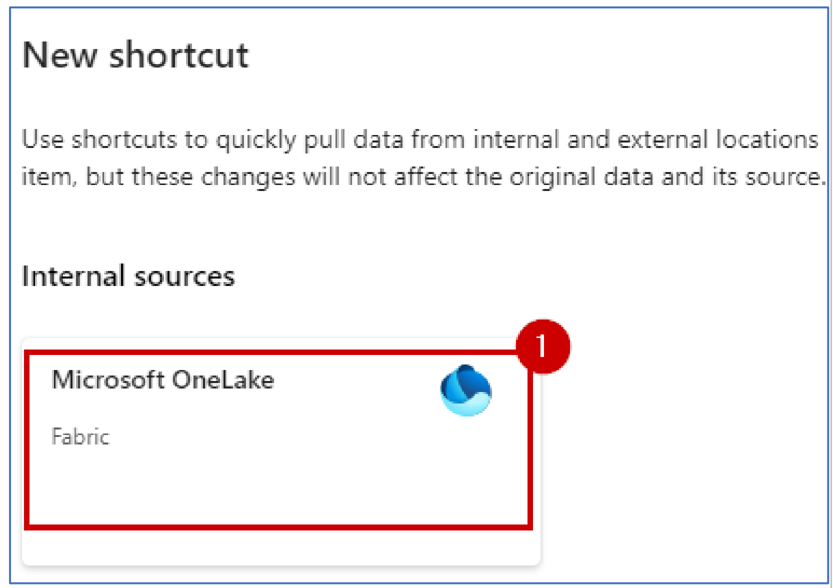
Contoso had its data located in multiple workspaces such as Marketing, IT, Operations, etc. Traditionally, if any department needs to work with data outside of their workspace, they need to create a copy. However, copying data causes data duplication, inconsistencies, and unwanted data movements across the organization, but Microsoft Fabric has a solution to this. Microsoft Fabric comes with OneLake - a single, unified, logical data lake for the whole organization. OneLake focuses on improving collaboration. It aims to give the most value possible out of a single copy of data without data movement or duplication. So far Eva has data related to campaigns, regular transactions, sales transactions, and so on in the bronze layer of the lakehouse. But she hasn't made any transformations yet.	6. In Lakehouse Explorer, **observe** the data files ingested from Contoso Marketing workspace to the bronze layer of the lakehouse.	

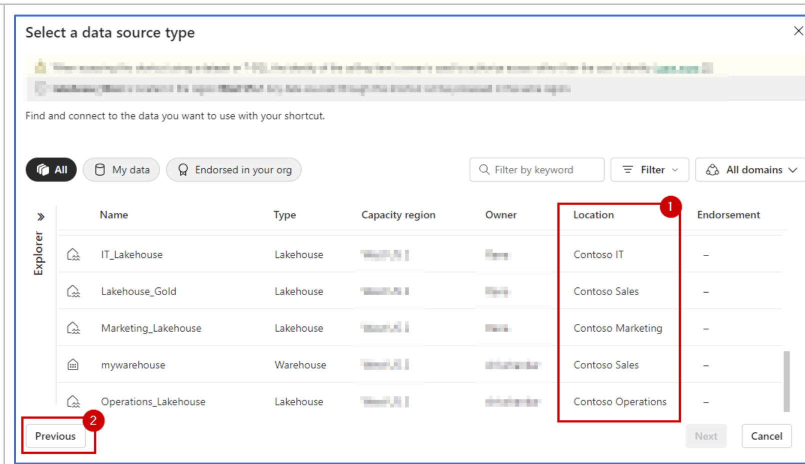
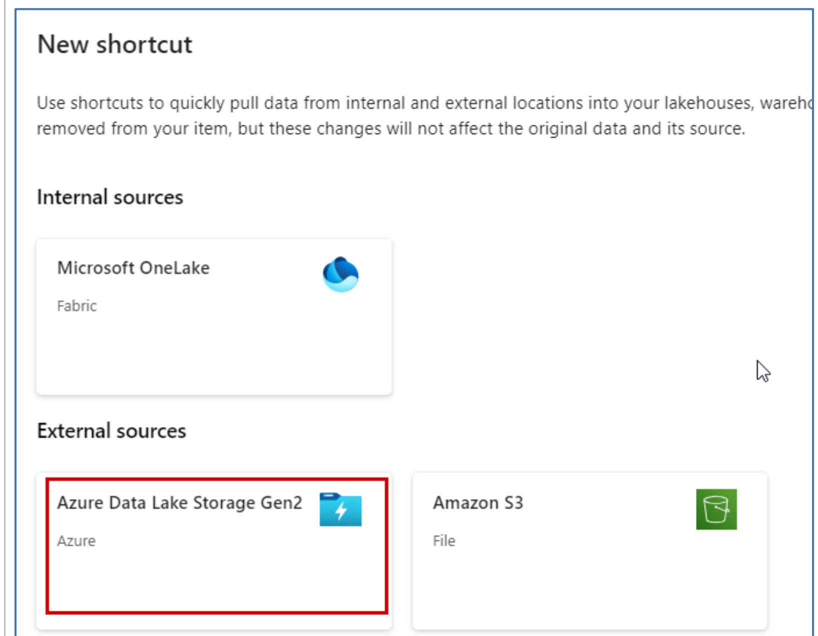
C. Using 'New Shortcut'

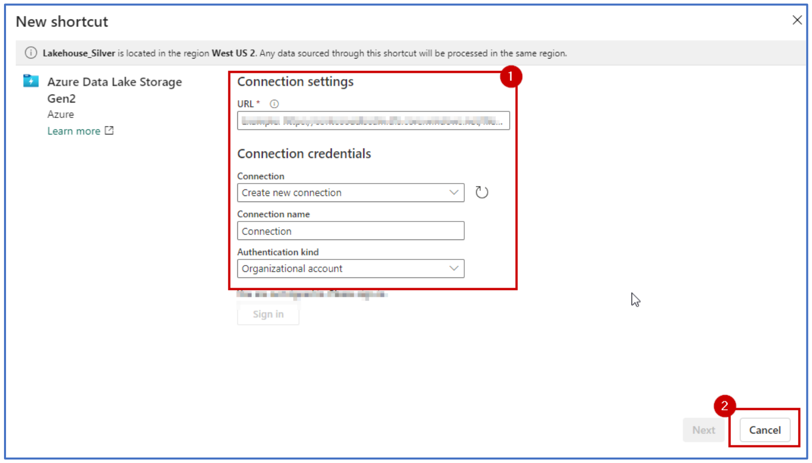
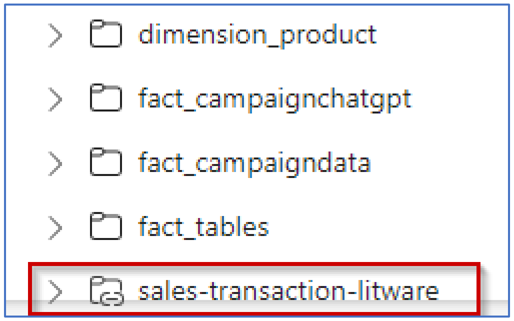
Narrative	Steps	Screenshot
Let's go back to the Contoso Sales Workspace.	1. **Select** the Contoso Sales workspace from the left navigation bar.	

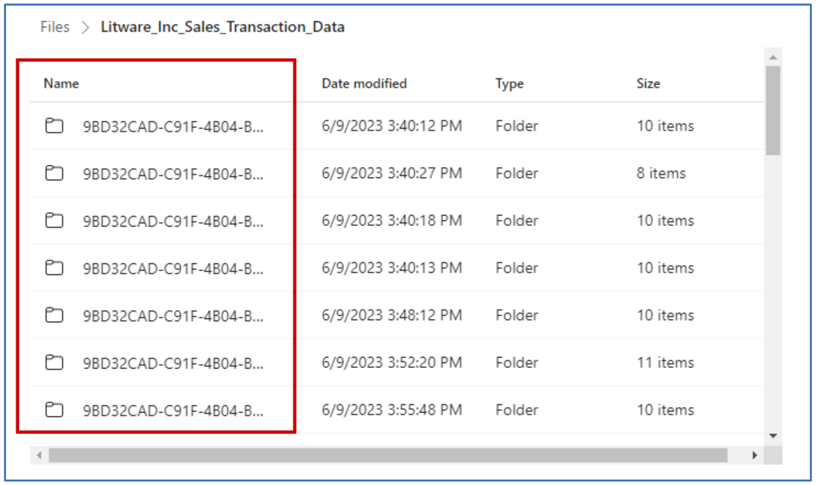
Eva uses the new shortcut option to bring data from the newly acquired company LitWare Inc., which resides in ADLS Gen2, into the silver layer of the lakehouse. Then she goes to the lakehouse.	2. **Select** Filter. 3. **Select** Lakehouse.	

<table>
<tr>
<td>She selects Lakehouse_Silver.</td>
<td>4. Select Lakehouse_Silver from the list of assets.</td>
<td>
</td>
</tr>
<tr>
<td>The New shortcut option is the latest addition to the list of ways data can be replicated from storage to the lakehouse.</td>
<td>In Explorer, scroll down to the Files folder.
5. Click on the ellipses (three dots) next to Files.
Note: Drag the Explorer panel to the right to see the ellipses.
6. Select New shortcut.</td>
<td>
</td>
</tr>
</table>

Getting data from a shortcut is a fascinating method to ingest data from any source! With the New shortcut option, Eva can create a direct connection to an internal data source (data product) such as a lakehouse, warehouse, and/or KQL Database in another workspace. Alternatively, she can create a shortcut to an external data source such as ADLS Gen2 or Amazon S3. The data isn't locked behind engines or compute. All the engines can access the same data without having to move, load, import, or export the data.	7. **Point** to Internal Sources. 8. **Point** to External Sources.	
Even though Eva has already connected all of Contoso's data to the lakehouse, she is curious to explore Microsoft OneLake for shortcuts.	9. **Click** on Microsoft OneLake.	

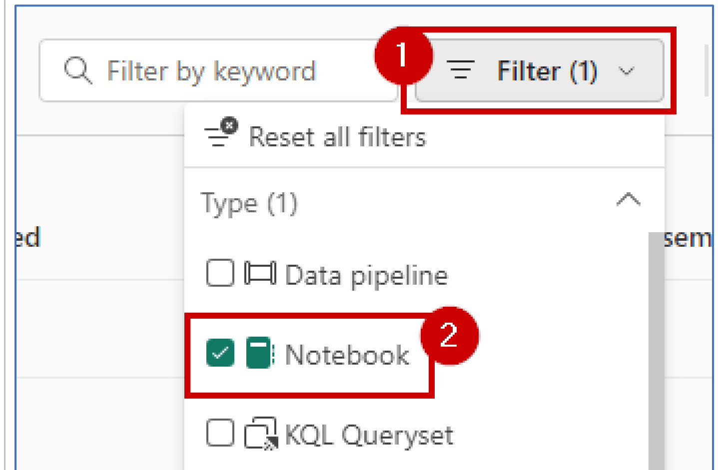
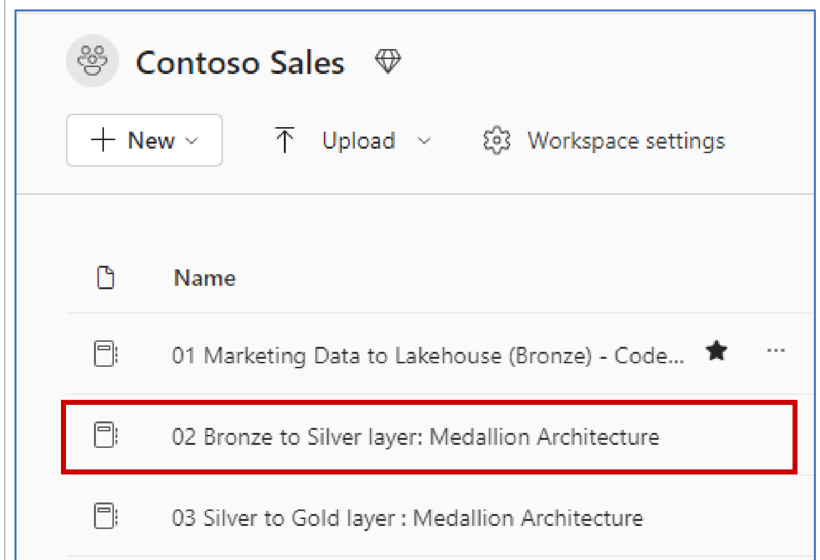
Inside Microsoft OneLake, she notices that not only can she create a reference to any workspace, but she can also create a reference to any data warehouse, or lakehouse for which she has permissions. Something Eva can consider the next time she is working with Contoso data. For now, she needs to get started with LitWare Inc.'s data as soon as possible.	10. **Show** the Location column. 11. **Click** on Previous.	
Eva navigates back to the '*New shortcut*' wizard. With the New Shortcut option, she creates (mounts) virtual pointers to directly connect to LitWare Inc.'s data located outside of the workspace, in ADLS Gen2. Using the shortcut wizard to connect to an existing data source eliminates side copies of data and reduces process latency associated with data copies and staging, ultimately saving costs on data storage.	12. **Click** Azure Data Lake Storage Gen2.	

ADLS shortcuts utilize a delegated authorization model. Here, she enters the connection URL of LitWare Inc.'s ADLS Gen2 storage, the connection name, and the authentication kind. In less than a minute, and without any data movement, the files and folders of LitWare Inc. become part of the newly created lakehouse. Let's navigate to the Sales Workspace and see what the Sales data from LitWare Inc. looks like.	13. **Show** the connection details. 14. **Click** Cancel.	
The folders with a chain link icon confirm that the data in these folders is just a reference to data stored at a different location. Eva hasn't copied any of the data. She simply connected to LitWare Inc.'s workspace. Let's take a look at the sales-transaction-source folder.	In the Explorer panel: 15. **Expand** Files**.** 16. **Click** on sales-transaction-litware.	

LitWare Inc. seems to have plenty of data from its sales transactions. Prior to Microsoft Fabric, departments in Contoso had to move the data they needed from other departments via a time-consuming ETL processes. But look, now they can get **shortcuts** to all these workspaces! No need to move any of this data. That is the power of OneLake! The bronze layer of the lakehouse has raw data from different domains and workspaces, including data from an external workspace. Now, it is time to start data transformation so that the data is ready for BI purposes.	17. **Show** the sales transaction data (raw) folders from LitWare Inc.'s workspace into the Contoso Sales Workspace.	

3.3.3 Medallion architecture in Microsoft Fabric using Spark notebooks.

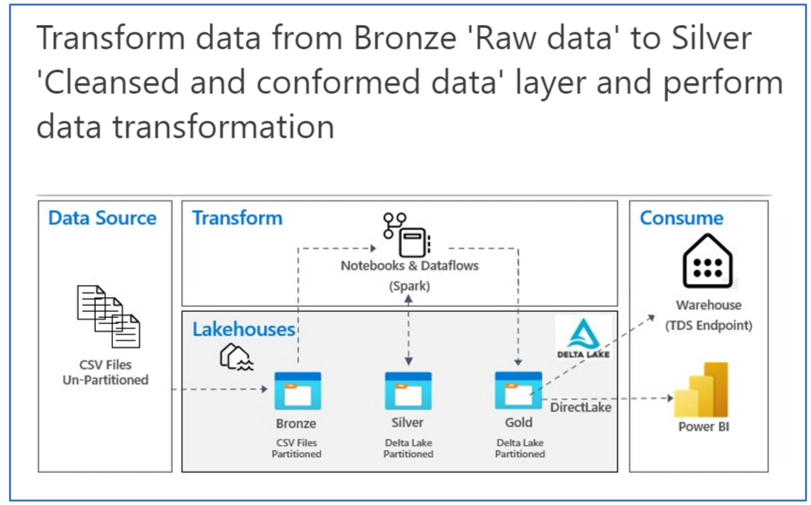
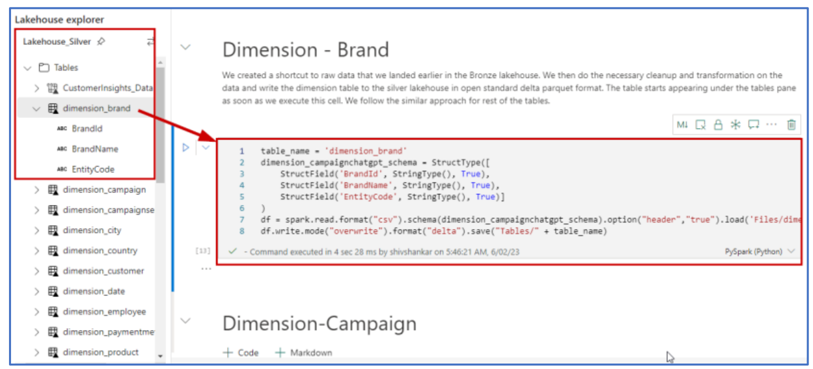
Narrative	Steps	Screenshot
The medallion architecture describes a series of data layers that denote the quality of data stored in the lakehouse. It guarantees atomicity, consistency, isolation, and durability as data passes through multiple layers of validations and transformations before being stored in a layout optimized for efficient analytics. The terms bronze (raw), silver (validated), and gold (enriched) describe the quality of the data in each of these layers. Eva thinks the best way to improve the structure and quality of data incrementally and progressively is to organize data in Medallion architecture. She has created a couple of notebooks in the Sales workspace to do that.	1. **Select** Contoso Sales workspace from the left navigation bar.	

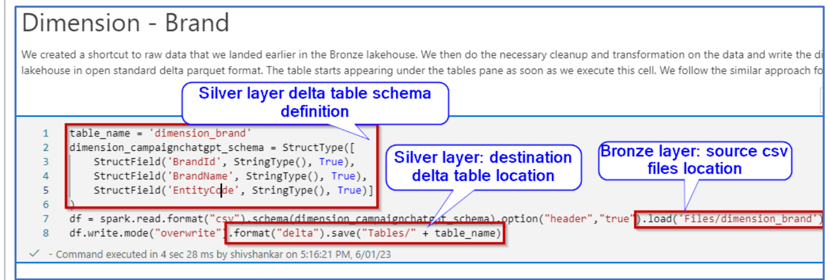
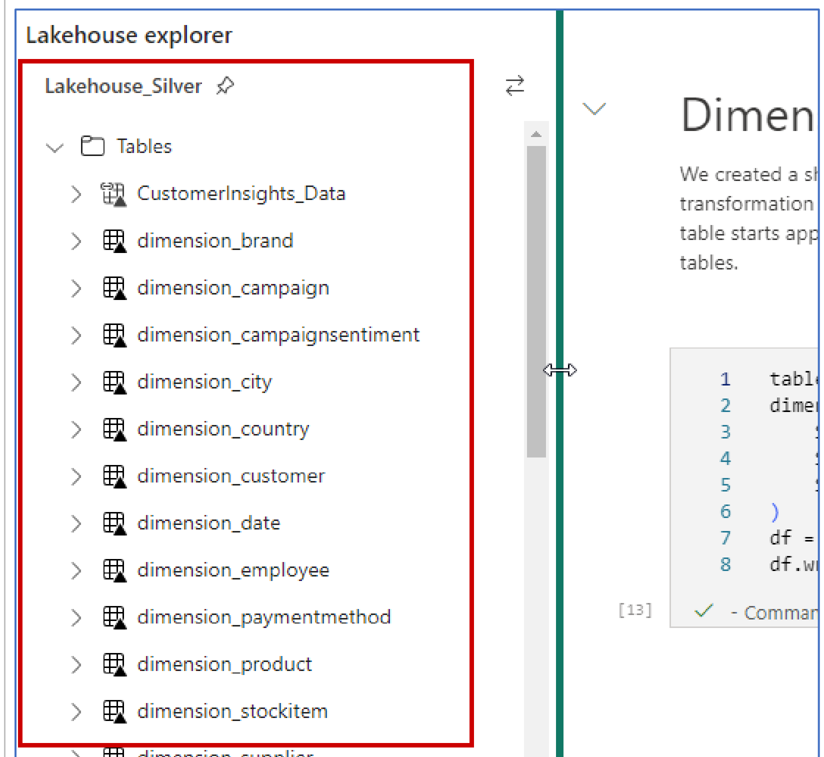
She explores these notebooks step by step.	2. **Click** Filter. 3. **Select** Notebook.	
First, she looks at the data transformation operations, i.e., convert data from raw .csv format in the bronze layer to open standard delta tables into the silver layer.	4. **Open** "02 Bronze to Silver layer: Medallion Architecture" notebook.	

<table>
<tr>
<td>In Microsoft Fabric, note that compute is isolated from the storage layer of the analytical workload. Here’s where the magic begins.

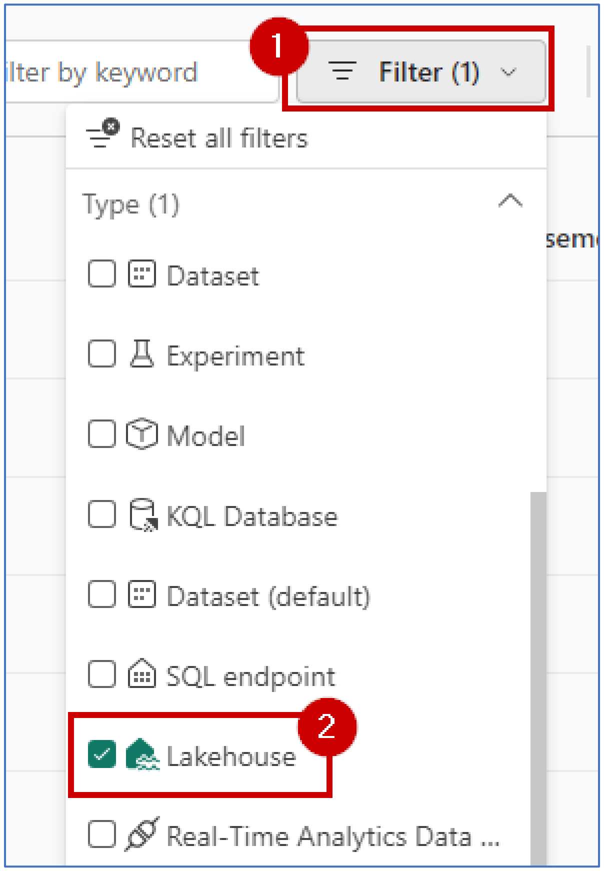
Every Microsoft Fabric workspace comes with a default Spark pool, called Live Pool. So, when Eva created the notebook, she didn’t have to specify any Spark configurations or cluster details.

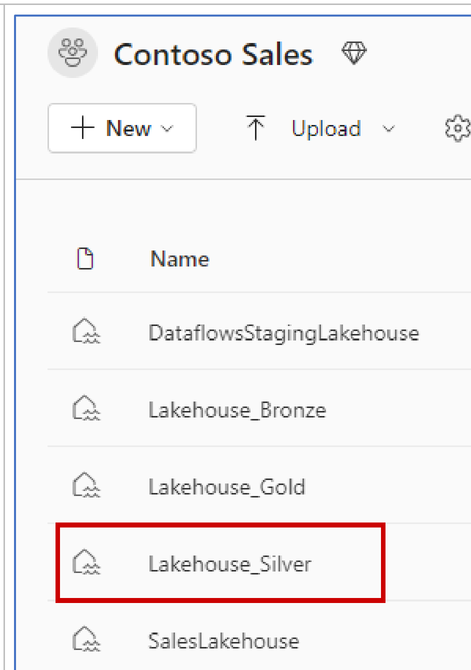
Upon executing the notebook commands, the live pool is up and running in a few seconds. The Spark session is established, and it starts executing the code.

Subsequent cell execution is nearly instantaneous in this notebook while the Spark session is active.</td>
<td>5. Talk about default Spark Pool integrated with Microsoft Fabric.</td>
<td>
</td>
</tr>
<tr>
<td>Currently the raw data is sitting in the bronze layer in .csv format.</td>
<td>6. In the notebook, scroll down to cell [13] to explain the python script that is converting the .csv files into delta tables.
For example: Dimension - Brand</td>
<td>
</td>
</tr>
</table>

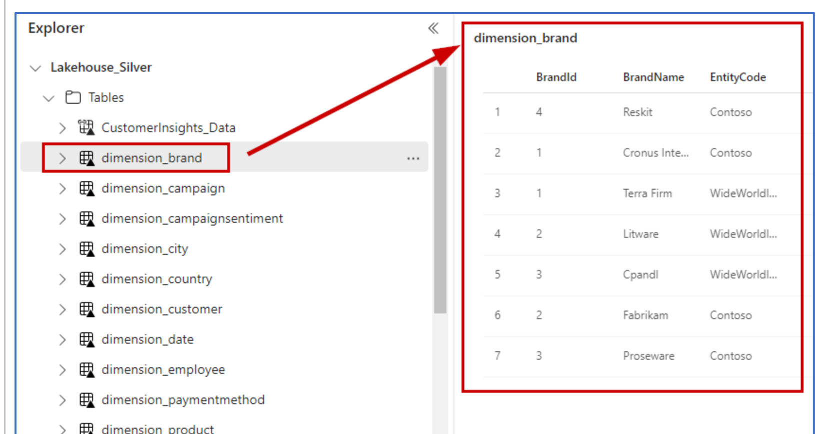
First, the code performs necessary transformations on the data, and then it writes the dimension table to the silver layer lakehouse as an open-standard delta table. The delta table appears under the tables pane as soon as she executes this cell. Then she performs the same steps for the rest of the tables.	7. **Show** delta table schema, then source and destination in the code block of 'Dimension – Brand'.	
It is fascinating to see that earlier the data was in files. Guess where Eva is right now? In the silver layer of the lakehouse, with delta tables. She has transformed the raw data files into **open-standard delta tables**. This is one copy of the data in OneLake.	8. **Point** to tables present in Lakehouse_Silver.	

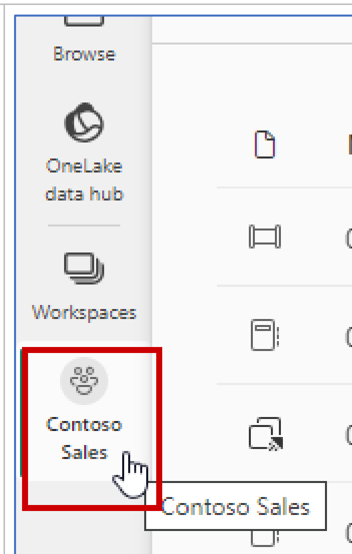
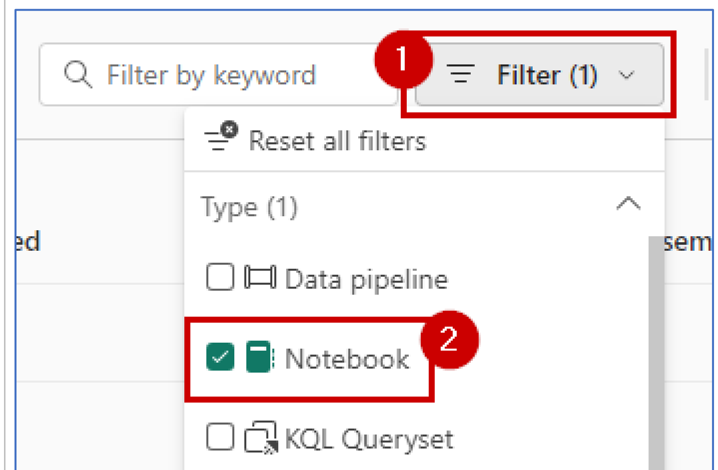
Now that the transformation operations are complete, Eva goes back to the silver layer in the lakehouse to look at the newly created tables.	9. **Select** Contoso Sales workspace from the left navigation bar.	

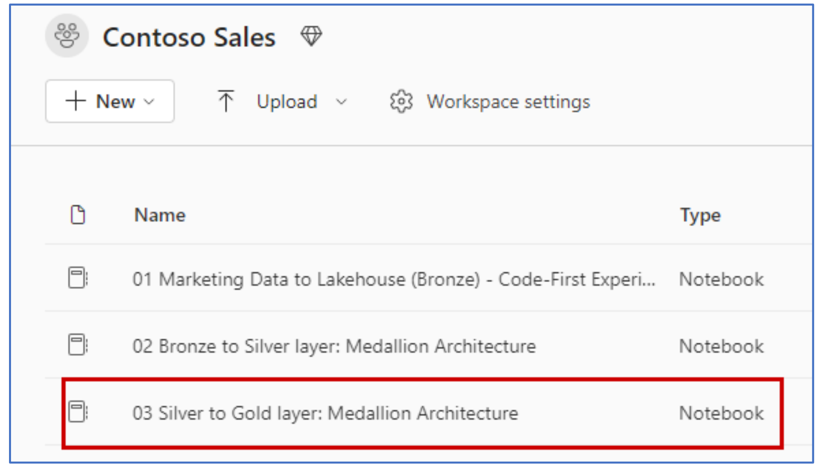
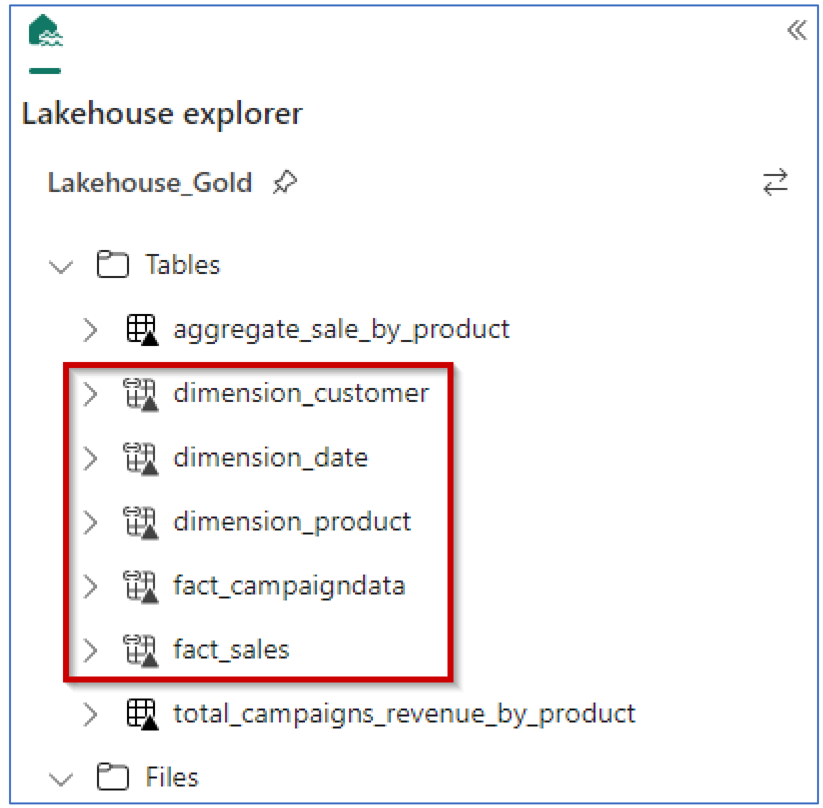
She uses the filter to select Lakehouse type assets.	10. **Select** Filter. 11. **Select** Lakehouse.	

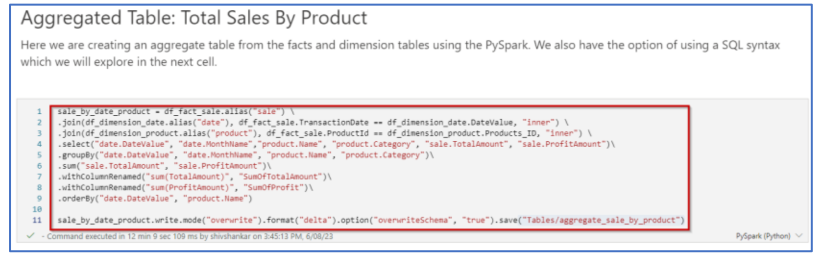
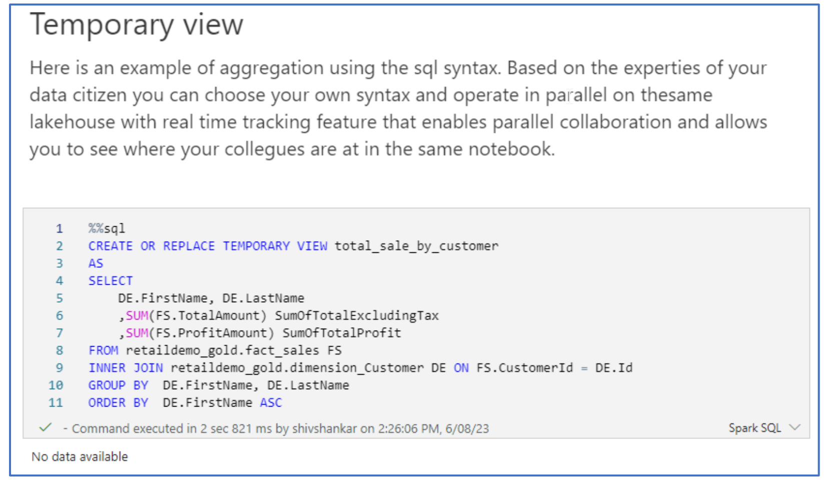
<table>
<tr>
<td>Then, selects Lakehouse_Silver.</td>
<td>12. Select Lakehouse_Silver.</td>
<td>
</td>
</tr>
<tr>
<td>Remember, the Spark notebook where Eva performed the transformations on the .csv files sitting in the bronze layer?

After executing that notebook, here she has her tables in open standard delta format.

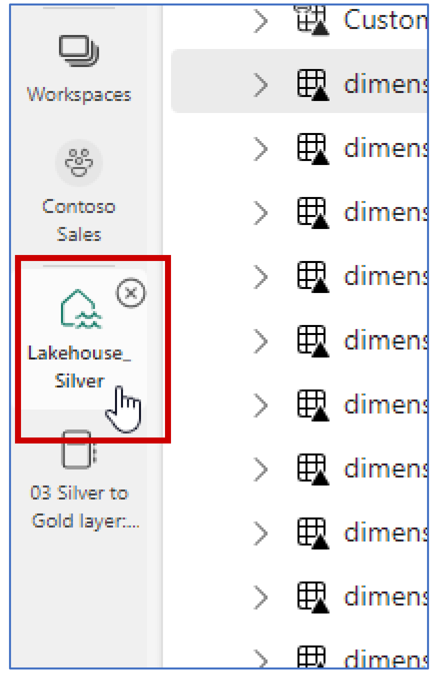
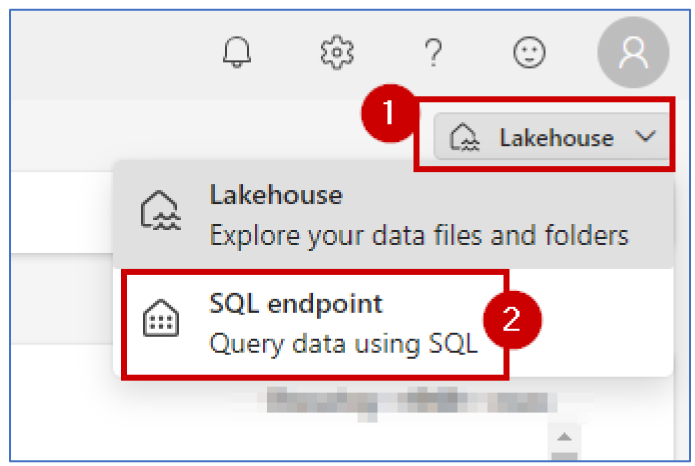
Now that the data is cleaned and transformed to open standard delta format, the next step is to refine the data further and move it from the silver layer to the gold layer.</td>
<td>13. Click on any one of the tables to verify the data.
14. For example: select dimension_brand</td>
<td>
</td>
</tr>
</table>

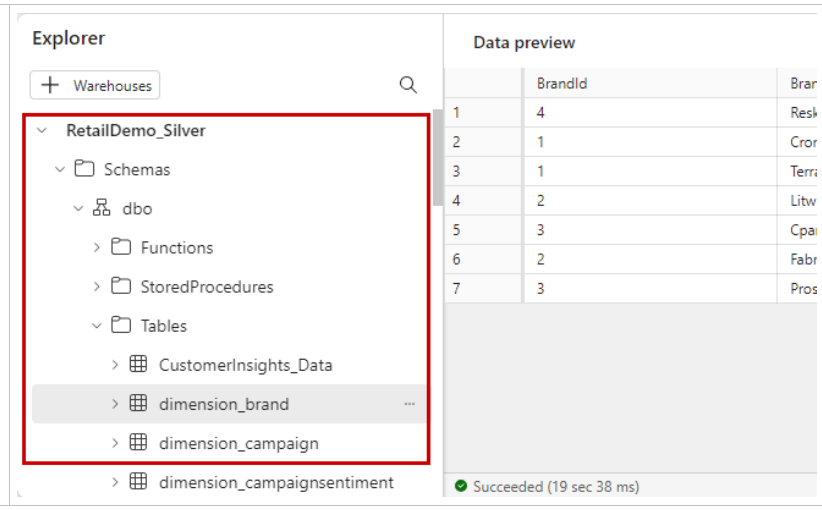
Let’s go back to Contoso Sales workspace.	15. **Select** Contoso Sales workspace from the left navigation bar.	
Select Notebook assets.	16. **Select** Filter. 17. **Select** Notebook.	

<table>
<tr>
<td>Let’s open the notebook Eva prepared for further data transformation/aggregation operations on the silver layer data.</td>
<td>18. Click “03 Silver to Gold layer: Medallion Architecture” notebook.</td>
<td>
</td>
</tr>
<tr>
<td>Before Eva performs any aggregation, she connects to the silver layer delta tables using the internal shortcut.</td>
<td>19. In lakehouse explorer, expand Tables then show <u>mounted</u> sliver layer delta tables.</td>
<td>
</td>
</tr>
</table>

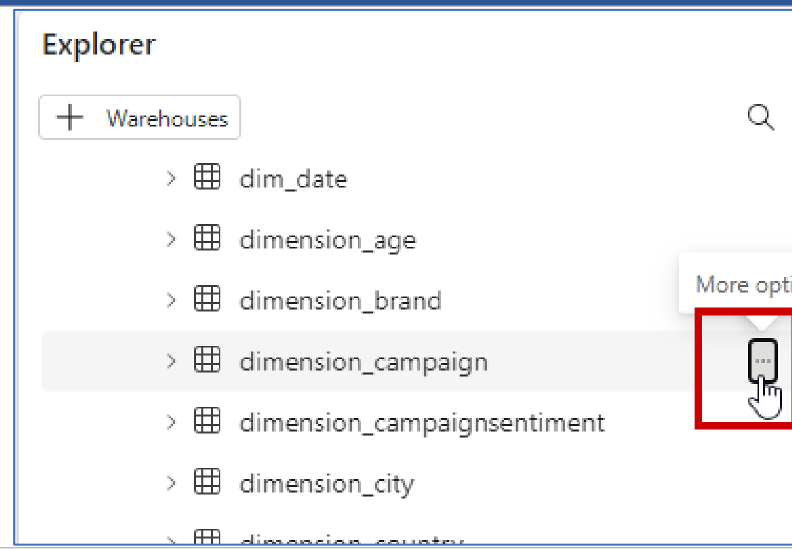
<table>
<tr>
<td>Now Eva wants to understand which products are popular among customers, so she uses PySpark to join the sales and product data tables to aggregate the data and get the total sales of each product date wise.</td>
<td>20. In the notebook, scroll down to cell [4] to demonstrate data operations being performed on silver layer data.
For example: Aggregated Table: Total Sales By Product</td>
<td>
</td>
</tr>
<tr>
<td>Because Microsoft Fabric makes it possible for people with varied experiences and preferences to work and collaborate, Eva could have performed a similar aggregation using Spark SQL instead of PySpark.
Here, Eva has created a temporary view to understand the spending pattern of Contoso's customers.
With the data ready to be consumed, Eva would like to run some queries on these delta tables.</td>
<td>21. Scroll down to cell [6] to show the SQL code for aggregating total sales by customer.</td>
<td>
</td>
</tr>
</table>

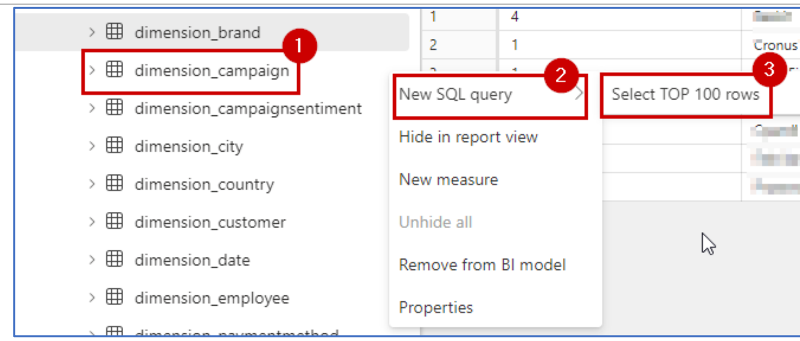
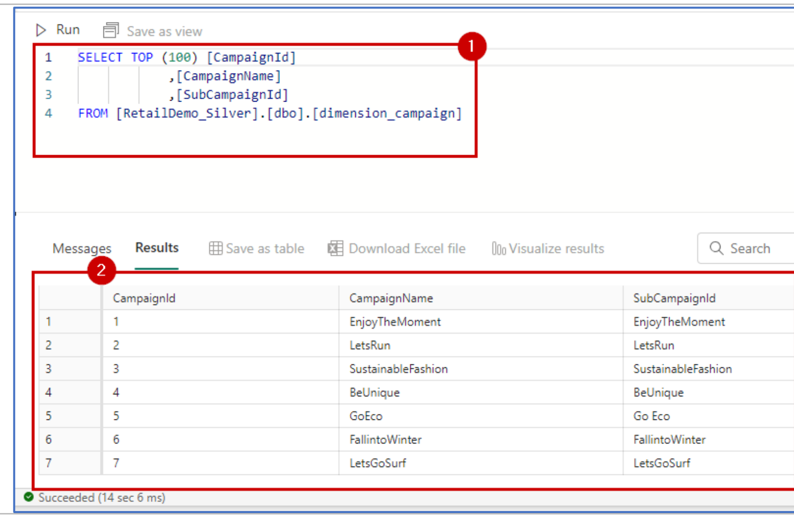
3.3.4 Lakehouse With SQL endpoint

Narrative	Steps	Screenshot
Eva has a lot of experience with SQL, and she prefers to use SQL syntax for querying purposes.	1. **Click** on Lakehouse_Silver to navigate to SQL endpoint.	
Every lakehouse comes with a default SQL endpoint. Eva can switch from lakehouse to SQL endpoint in the same window by selecting SQL endpoint from the lakehouse dropdown menu in the top right corner of the window.	2. **Click** the Lakehouse dropdown button. 3. **Select** SQL endpoint.	

Here is a list of all the tables in open-standard delta format. Now Eva can run queries on this table to get the insights she needs for the next step.	4. **Point** to the tables in the silver layer in open-standard delta table.	

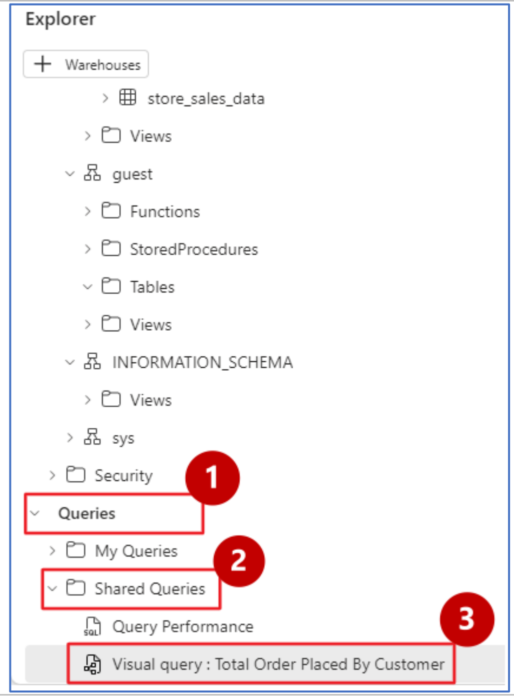
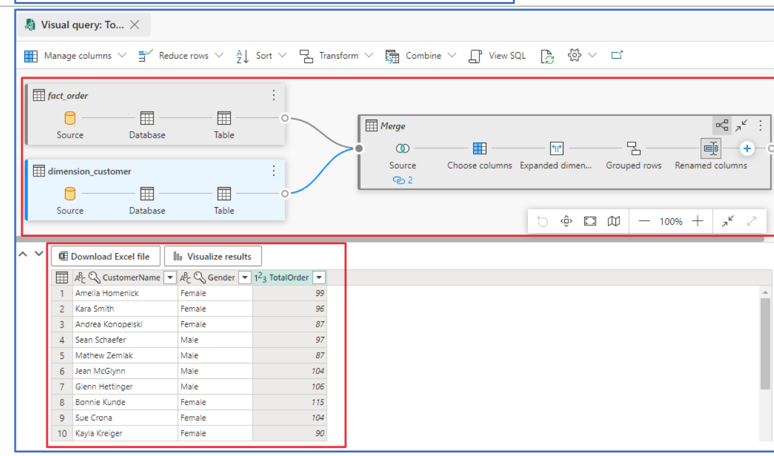
A. SQL Query

Narrative	Steps	Screenshot
Eva would like to know more about the campaigns Contoso organized.	1. **Click** on the ellipses (three dots) next to any table name, for example: dimension_campaign.	

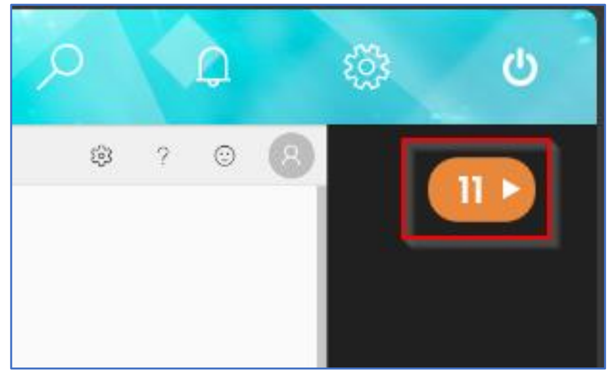
<table>
<tr><td></td><td></td><td></td></tr>
<tr>
<td>She can create the New SQL query in just a few clicks by navigating through the pop-out menu and choosing the built in Select TOP 100 rows option.</td>
<td>In the pop out menu:
2. Select New SQL query.
3. Select TOP 100 rows.</td>
<td>
</td>
</tr>
<tr>
<td>The operation opens the query panel with the result of the recently executed query. She finds out that there are a total of seven campaigns organized by Contoso.</td>
<td>4. Show the result generated after the query execution is complete.</td>
<td>
</td>
</tr>
</table>

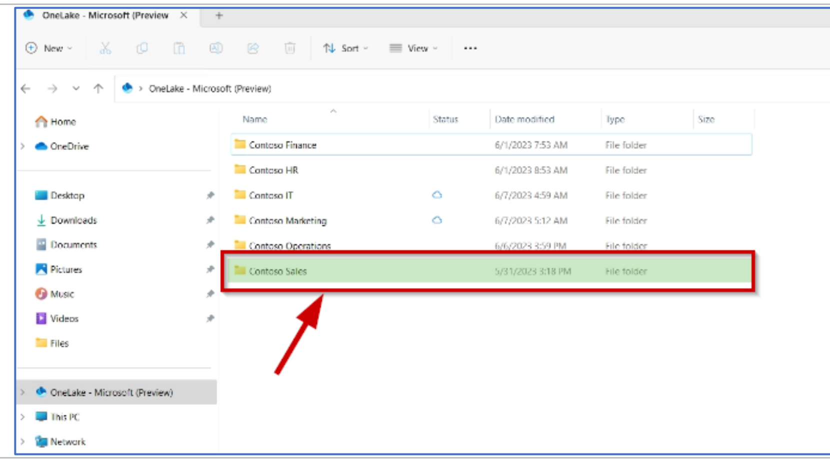
B. Visual Query

Narrative	Steps	Screenshot

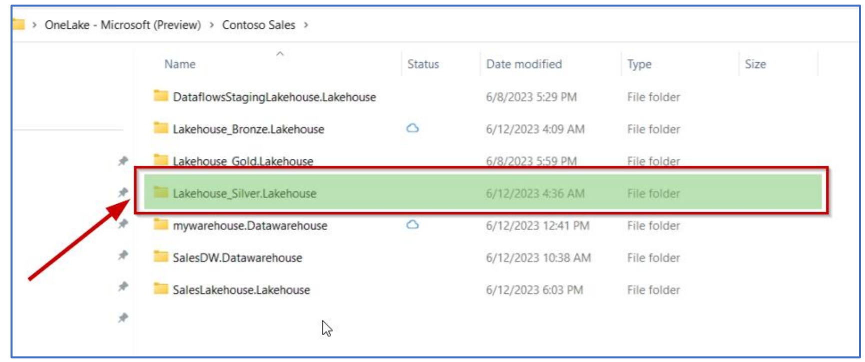
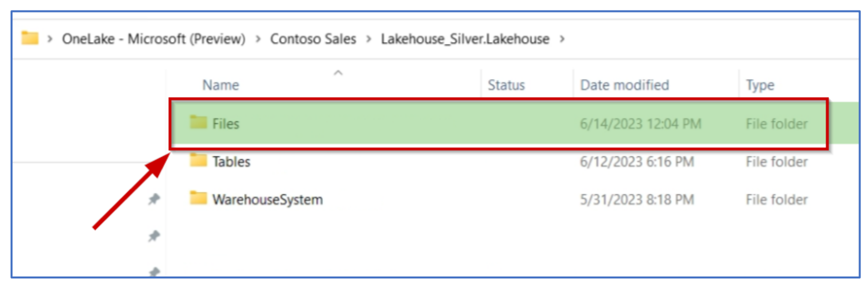
Her other option is to build a visual query. She clicks on New visual query to explore that option.	In SQL Endpoint Explorer, 1. **Expand** Queries. 2. **Expand** Shared Queries. 3. **Select** visual query ***'Visual query: Total Order Placed By Customer'.***	
A new Visual query canvas opens on her screen. Here, she can drag and drop tables from the Lakehouse and establish a relationship between them before executing the query. The next logical step for her to take is to create a Power BI report. But before that, she would like to explore the Microsoft Fabric Direct Lake with default dataset.	4. **View** the Visual Query Table & Merge activity. 5. View the result '**Total Order Placed By Customer**'. **Note:** If you have more than 10 items open in the left navigation pane, you will receive an error message when opening the next item. Make sure you close the work items on the left as you proceed to the next screens.	

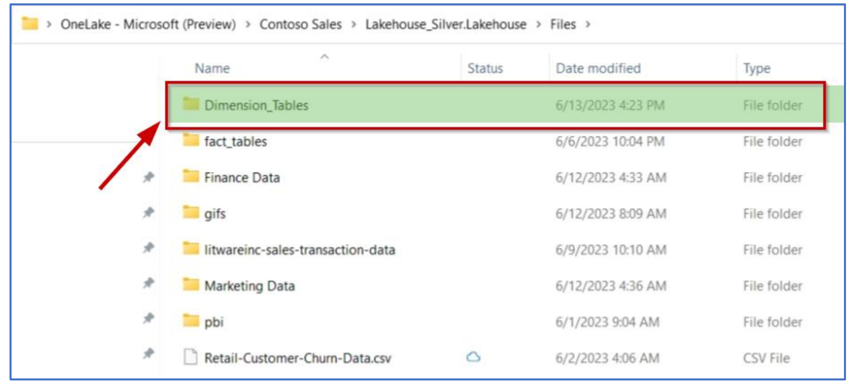
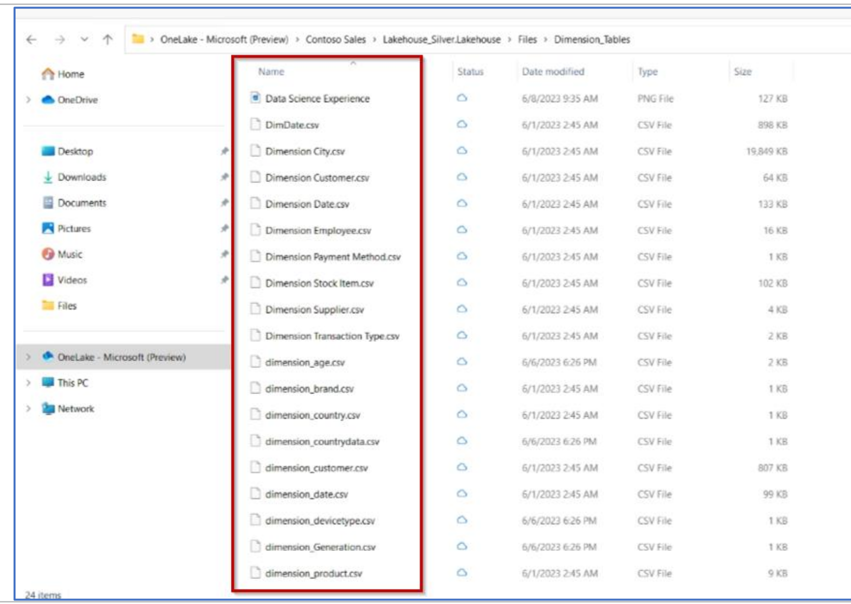
3.3.5 OneLake Explorer [Click - by - click]

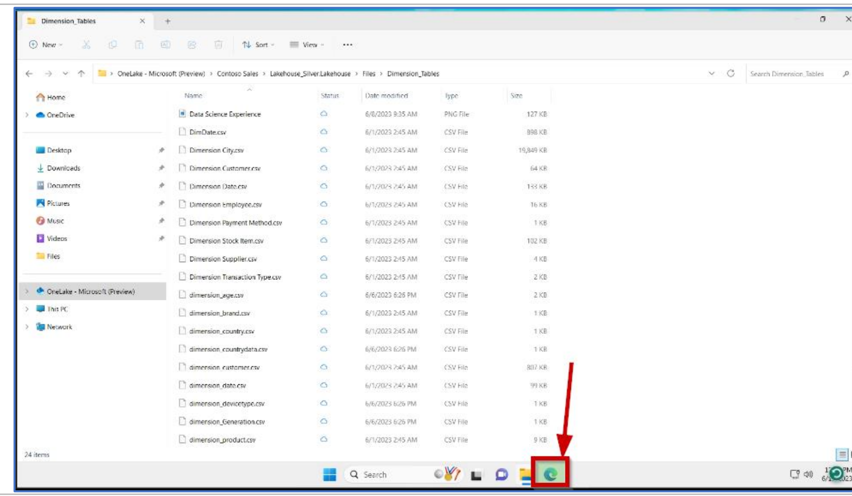
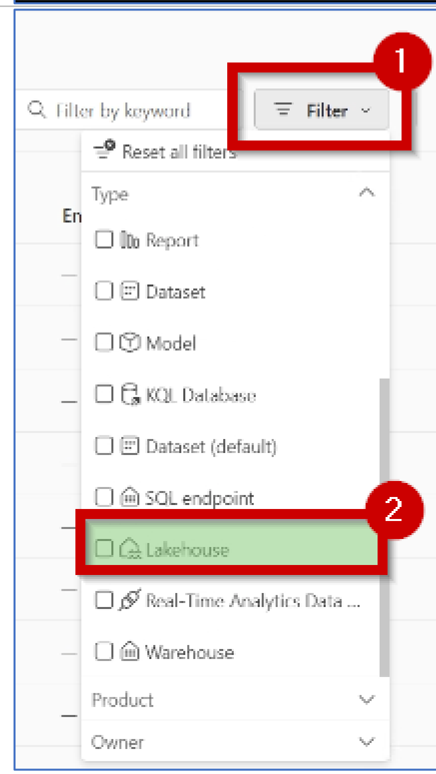
Narrative	Steps	Screenshot
OneLake, OneDrive for data. OneLake is a single, unified, logical data lake for the whole organization. Like OneDrive, OneLake comes automatically with every Microsoft Fabric tenant and is designed to be the single place for all your analytics data. OneLake brings customers One data lake for the entire organization. One copy of data for use with multiple analytical engines.	1. **Navigate** to the web app 2. **Click** on arrow 11.	
Let's navigate to OneLake explorer.	**<<This step is embedded as click-by-click in the web app.>>** 3. **Click** on File Explorer.	

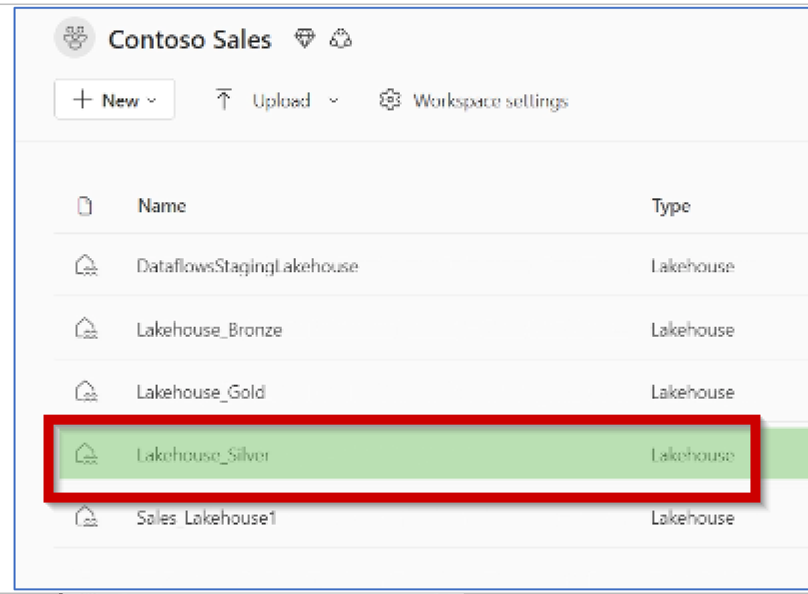
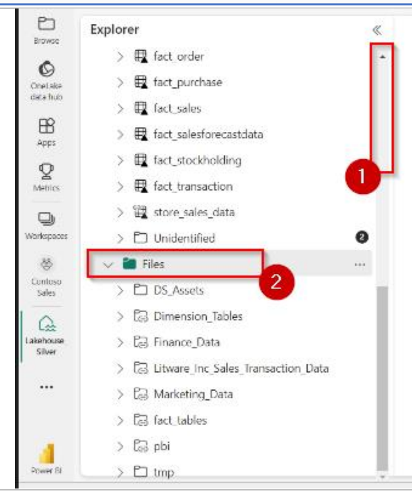
<table>
<tr>
<td>Here's the folder structure based on all the workspaces present in Microsoft Fabric. Let's go to our current workspace – Contoso Sales.</td>
<td>4. Show the folder structure.
5. Click on Contoso Sales.</td>
<td>
</td>
</tr>
<tr>
<td>Here we can see that OneLake organizes the data into a familiar folder structure like OneDrive.

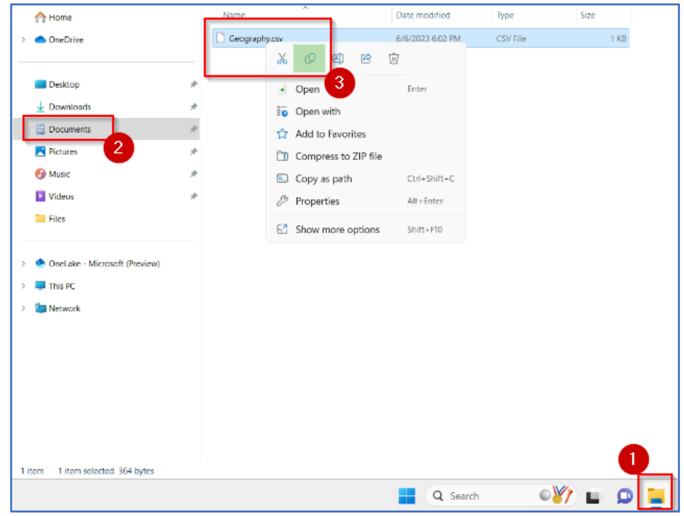
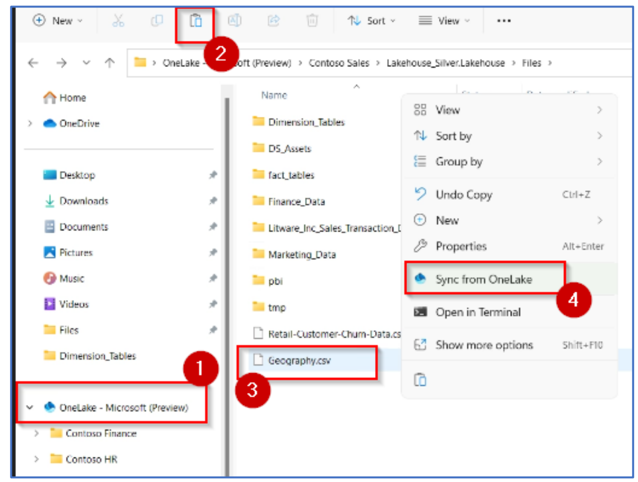
When you create, update, or delete a file via File Explorer, it automatically syncs the changes to OneLake service.

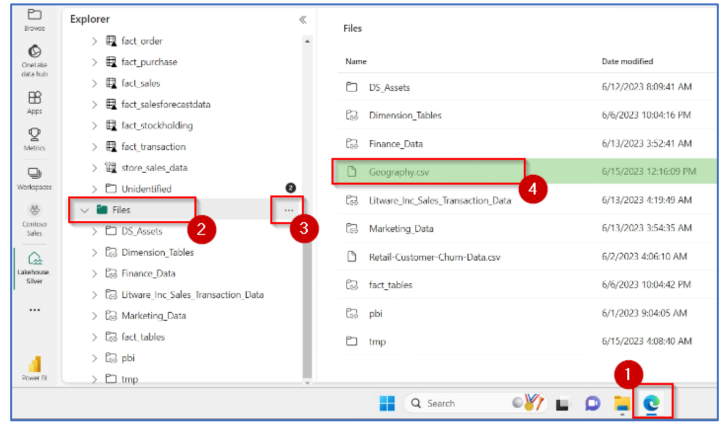
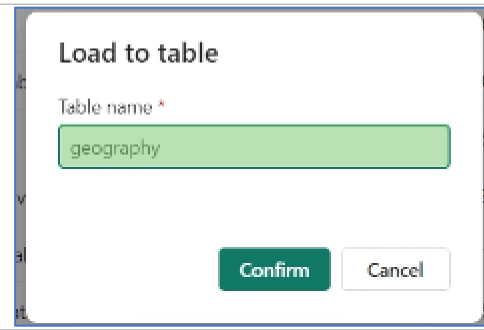
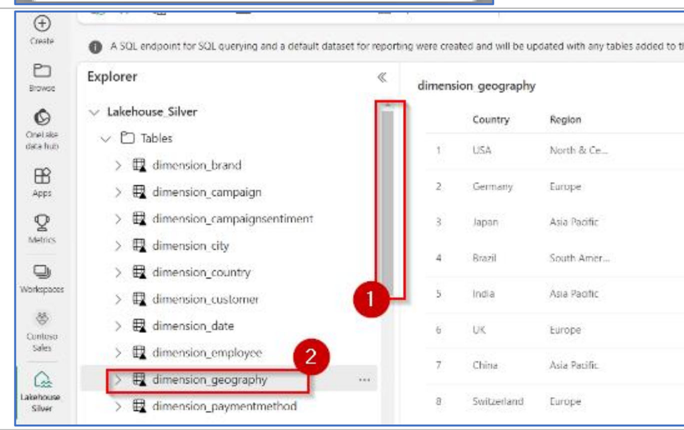
Changes made to your files outside of File Explorer aren't automatically synchronized. To synchronize these changes, you need to right click on the item or subfolder in Windows File Explorer and select Sync from OneLake.</td>
<td>6. Show the file structure in OneLake file explorer.
7. Click "Lakehouse_Silver.Lakehouse" folder.</td>
<td>
</td>
</tr>
<tr>
<td>On this screen, we can see three folders: Files, Tables, and WarehouseSystem.</td>
<td>8. Point to Files, Tables and, WarehouseSystem.
9. Click on Files.</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>All the files we loaded with pipelines and notebooks earlier can be found here in the Lakehouse_Silver.Lakehouse folder, and that includes not just the raw files, but the tables as well.</td>
<td>10. Show the files arranged in folder structure.
11. Click on Dimension_Tables.</td>
<td>
</td>
</tr>
<tr>
<td>When we open the Dimesntion_Tables, we can see all the raw files within this folder.</td>
<td>12. Show the raw files.</td>
<td>
</td>
</tr>
</table>

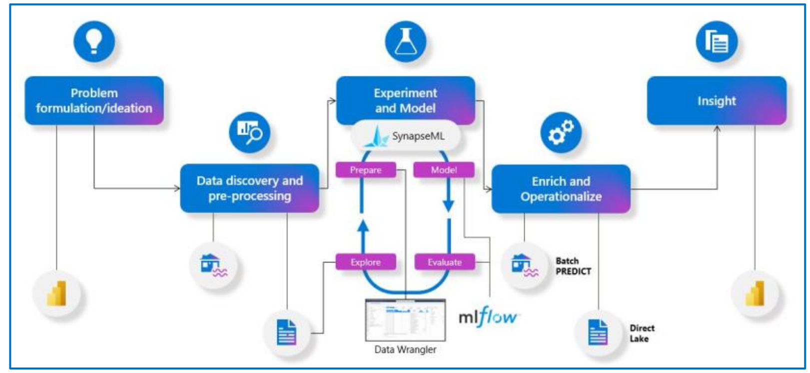
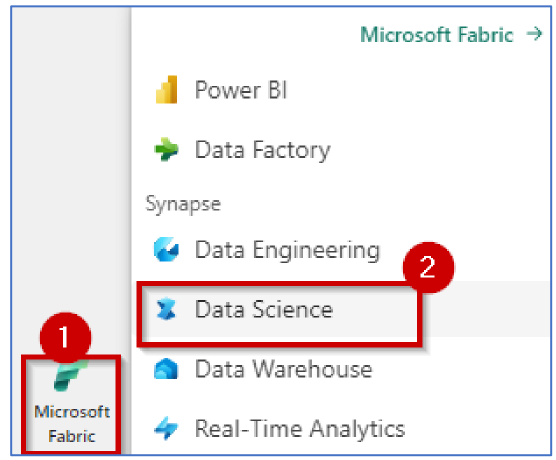
Now, let's navigate back to Microsoft Fabric, within this click-by–click to understand the next step.	13. **Click** on the embedded edge browser.	
Let's see how file synchronization happens in the lakehouse from our workspace.	14. **Click** on Filter. 15. **Click** on Lakehouse.	

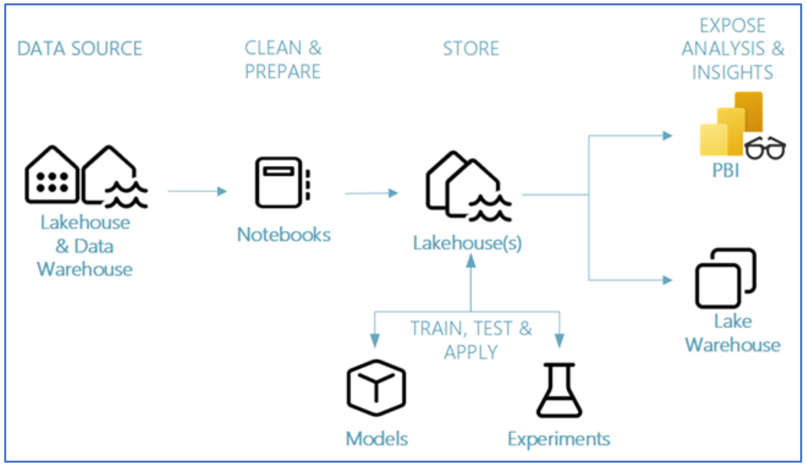
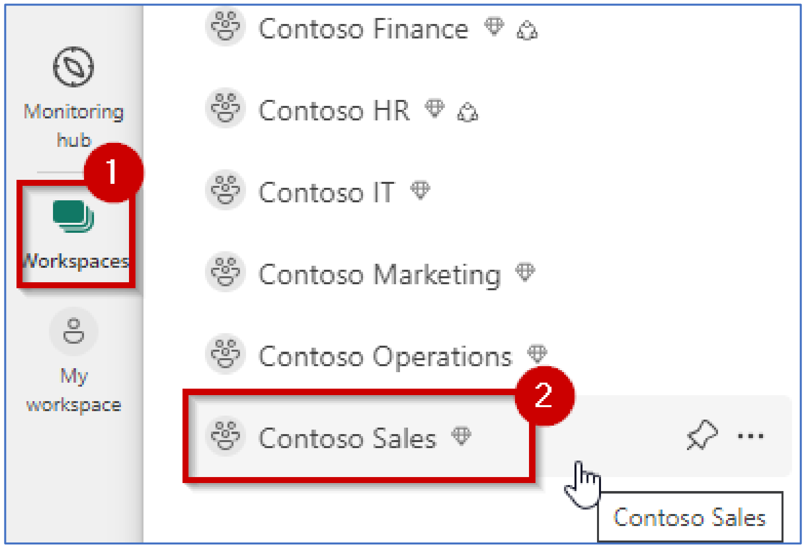
Let's open the silver lakehouse in the workspace.	16. **Click** on Lakehouse_Silver.	
We observe the same file structure in the lakehouse as we saw in the explorer.	17. **Click** on the seeker bar. 18. **Click** on files.	

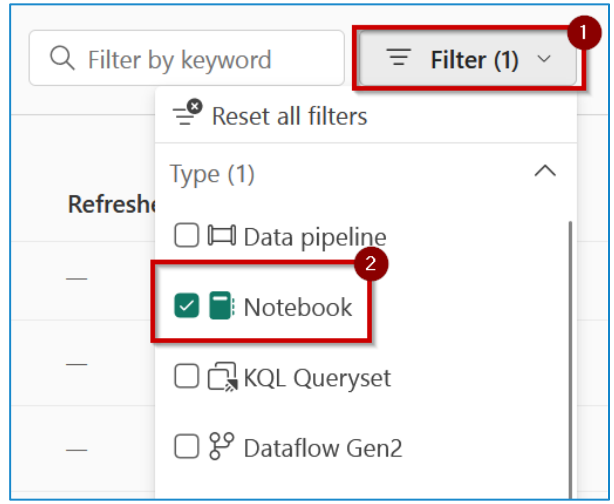
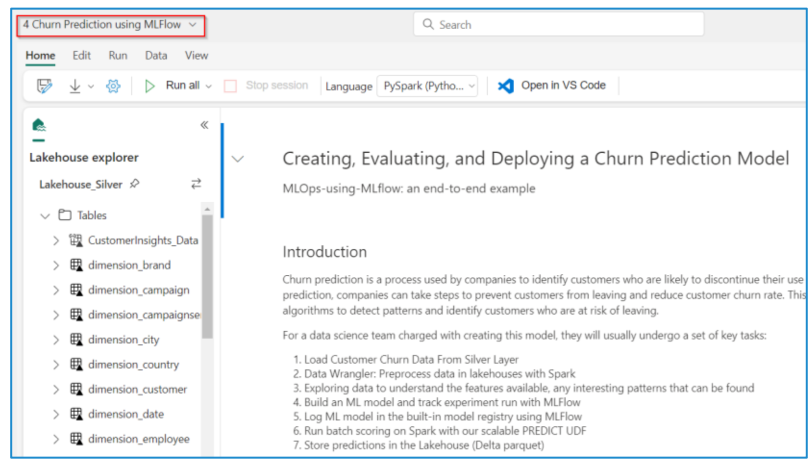
Local files from your computer can also be directly uploaded to OneLake with a simple file operation like drag and drop, or copy-paste.	19. **Click** on file explorer. 20. **Click** on Documents. 21. **Click** on the file to copy it.	
After we paste the file, we can immediately go check the Microsoft Fabric portal and see that the same file has already appeared. We can see the underlying files for our delta tables here as well. When we explore the contents of these folders, we can see that the open standard delta parquet files, along with their logs, are being stored in these folders.	22. **Click** on OneLake Explorer. 23. **Paste** the local file. 24. **Right click** on the file. 25. **Select** Sync from OneLake.	

Since our file was in .csv format, we can also directly load it to a delta table by simply right clicking on the file. The file will then start appearing as a delta table in your tables list.	26. **Click** on the Browser icon. 27. **Click** on Files. 28. **Click** on the Refresh button to see the created file. 29. **Click** on the new file. 30. **Click** on load to table.	
We can give a new name to the delta table.	31. **Click** on the Table name* text field. 32. **Click** on confirm.	
And we immediately see the resulting delta table in the tables list.	33. **Click** on the seek bar. 34. **Click** on the new delta table.	

3.4 Microsoft Fabric for Data Science experience

Narrative	Steps	Screenshot
Microsoft Fabric offers Data Science experiences to empower users to complete end-to-end data science workflows for the purpose of data enrichment and business insights. You can complete a wide range of activities across the entire data science process, all the way from data exploration, preparation and cleansing to experimentation, modeling, model scoring and serving predictive insights to BI reports.		
To understand the cause behind Contoso's declining revenue, the team needs to dive deeper into their customers' spending patterns. Eva, the Data Engineer, has cleaned, transformed, and moved the data to the silver layer. So, let's switch our role to Miguel, the Data Scientist. It's Miguel's turn to start working with this data to build ML models to predict customer churn.	1. **Switch** to the Data Science experience using the experience switcher icon in the left corner.	

When building data science workflows, Miguel leverages data science services alongside Microsoft Fabric to handle different stages of the workflow, such as data ingestion, data processing, model training, and deployment. He stores data science models as microservices within Microsoft Fabric, taking advantage of its scalability and reliability features.	Point to: • Load data from Data Source • Clean & Prepare data • Exploratory Data Analysis • Track Experiments run • Train, test, and apply • Store Predictions in Lakehouse • Business Insights and Analysis	
Miguel navigates to the Contoso Sales workspace where he'll work with customer churn data based on customers' spending patterns with Contoso. Thanks to Eva, Miguel has this data available in the silver lakehouse. It's already in an open delta parquet format and ready for consumption.	2. From the workspaces **select** the 'Contoso Sales' workspace.	

Miguel has already created a notebook in the workspace. He can go to the existing notebooks by using the filter and selecting Notebook.	3. **Click** Filter. 4. **Select** Notebook.	
Miguel opens the Churn Prediction notebook and to look at the steps required to build a churn prediction ML model.	5. **Click** – '*04 Churn Prediction using MLflow From Silver to Gold Layer*' to open the Notebook.	
The notebook describes steps and scripts/code to create, evaluate and deploy a customer churn prediction model.	6. **Continue** to review the cells as instructed.	

<table>
<tr>
<td>Install & import required Libraries and Packages.

This command will install the necessary libraries to work with MLflow.</td>
<td>7. Point to cell 2.</td>
<td><pre><code># Install required libraries
%pip install mlflow

15 sec - Command executed in < 1 ms by raj</code></pre></td>
</tr>
<tr>
<td>Further, a few more libraries are imported to support building the model and visualizing the data in the notebooks.</td>
<td>8. Point to cell 3.</td>
<td><pre><code>import numpy as np
import pandas as pd

# Loading MLflow libraries
import mlflow
from mlflow.models.signature import infer_signature
from lightgbm import LGBMClassifier

# Loading packages for building a Machine Learning pipeline
from sklearn.model_selection import train_test_split
from sklearn import metrics
from pyspark.sql.functions import col
from pyspark.sql.types import IntegerType, DoubleType
from pyspark.ml.evaluation import BinaryClassificationEvaluator

Loading libraries for visualization
%matplotlib inline
import seaborn as sns
import matplotlib.pyplot as plt
import IPython.display as display
plt.rcParams['font.family'] = 'DeJavu Serif'
plt.rcParams['font.serif'] = ['Times New Roman']

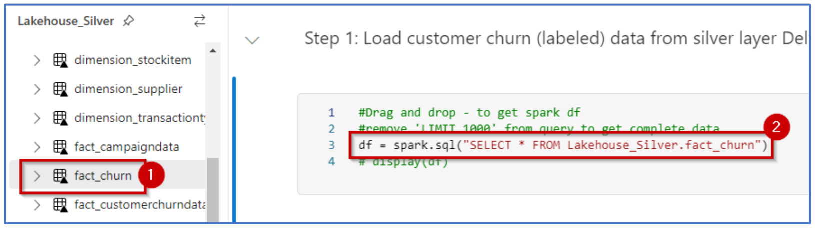
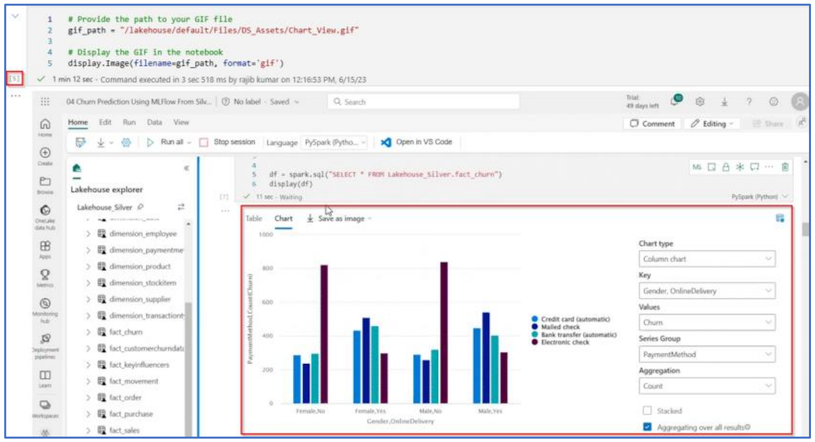
from synapse.ml.lightgbm import LightGBMClassifier
from synapse.ml.automl import TuneHyperparameters
from synapse.ml.train import ComputeModelStatistics
from synapse.ml.predict import MLFlowTransformer
import lightgbm as lgb

from synapse.ml.train import TrainClassifier
from pyspark.ml.classification import (
    LogisticRegression,
    RandomForestClassifier,
    GBTClassifier
)

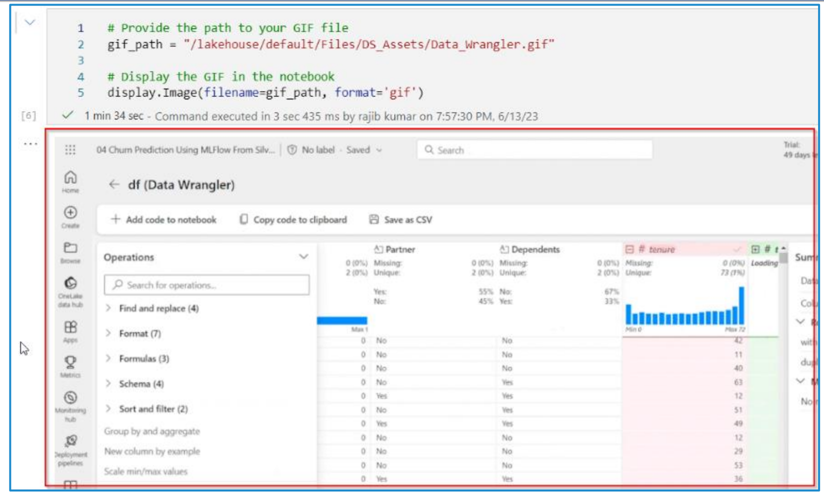
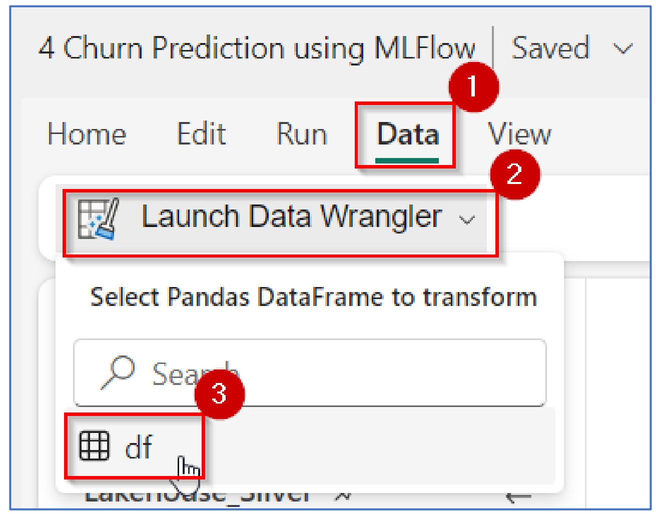
from synapse.ml.automl import *

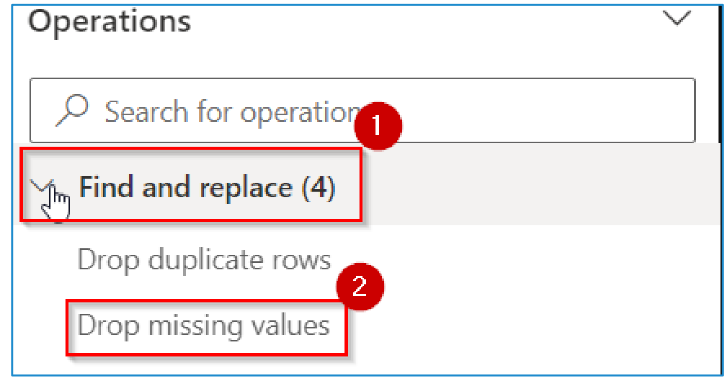
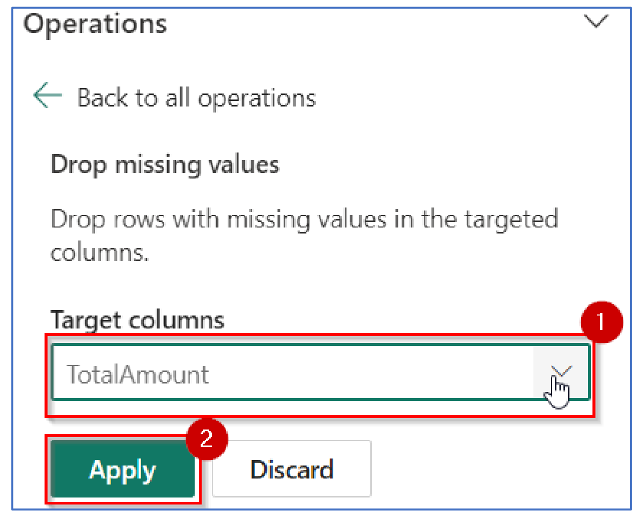
print("All Modules Loaded")</code></pre></td>
</tr>
</table>

3.4.1 Load data from sliver layer delta tables into Spark DataFrame

Narrative	Steps	Screenshot
***Load Data** using Drag & Drop* Microsoft Fabric allows an easy drag and drop feature to load the data from the delta tables. Miguel can simply drag the **fact_churn** table from Lakehouse_Silver and drop it to the notebook cell.	*Continue in the '04 Churn Prediction using MLflow From Silver to Gold Layer'* Notebook. 1. **Point** to cell 4.	
***Chart view** to explore data.* Microsoft Fabric provides an interactive visualization interface for exploring and analyzing data. It offers a variety of chart types and customization options to create insightful visualizations that help with understanding patterns, trends, and relationships within the data. *The gif simulates the drag and drop experience.*	2. **Point** to cell 5.	

3.4.2 Accelerate data prep with Data Wrangler in Microsoft Fabric

Narrative	Steps	Screenshot
***Data Wrangler** to preprocess data in lakehouses* Data Wrangler is a code tool that prepares data and generates Python code. This experience makes it easy to accelerate data cleansing and build repeatability and automation through generated code. *The gif simulates the experience of the Data Wrangling process.*	1. **Point** to cell 6. 2. **Show** the output gif to simulate data wrangler experience. ***Note:** Data Wrangler cannot be opened while the notebook kernel is busy. An executing cell must finish its execution before Data Wrangler can be launched.* *Data Wrangler currently supports only Pandas DataFrames. Support for Spark DataFrames is in progress.*	
Miguel begins pre-processing the data with Data Wrangler. He navigates to the Data tab and searches for his Pandas dataframe (df).	3. **Show** the gif.	

Now, he has a searchable list of data-cleaning steps under the Operations panel. To start with, he selects 'Drop Missing Values' from the list.	4. **Select** 'Drop Missing Values' from the Find and replace dropdown list	
Miguel uses the Data Wrangler tool to complete the data analysis. The feature combines a grid-like data display with dynamic summary statistics, built-in visualizations, and a library of common data-cleaning operations. So, Miguel applies the generated code that can be saved back to the notebook as a reusable function.	5. Steps	

<table>
<tr>
<td>To apply the newly generated code, he runs the notebook cell manually.

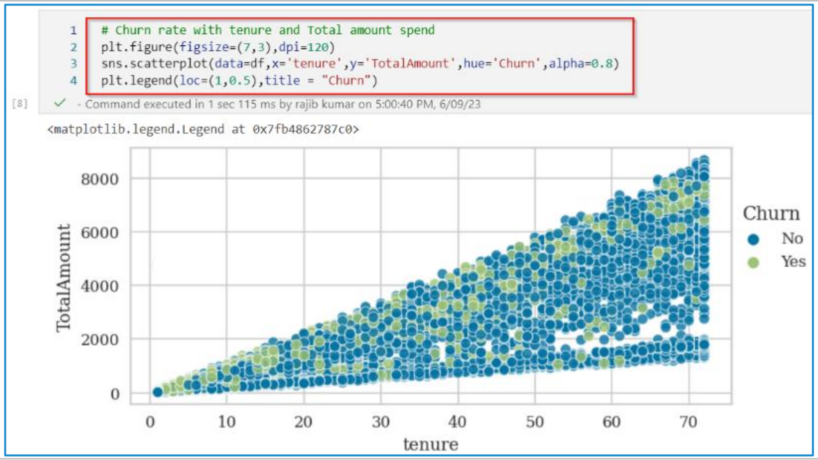
This overwrites the original Spark DataFrame. Now, the clean data is ready for further analysis.</td>
<td>6. Point to cell 7.</td>
<td><pre><code>#Data Wrangling function
df = df.toPandas()

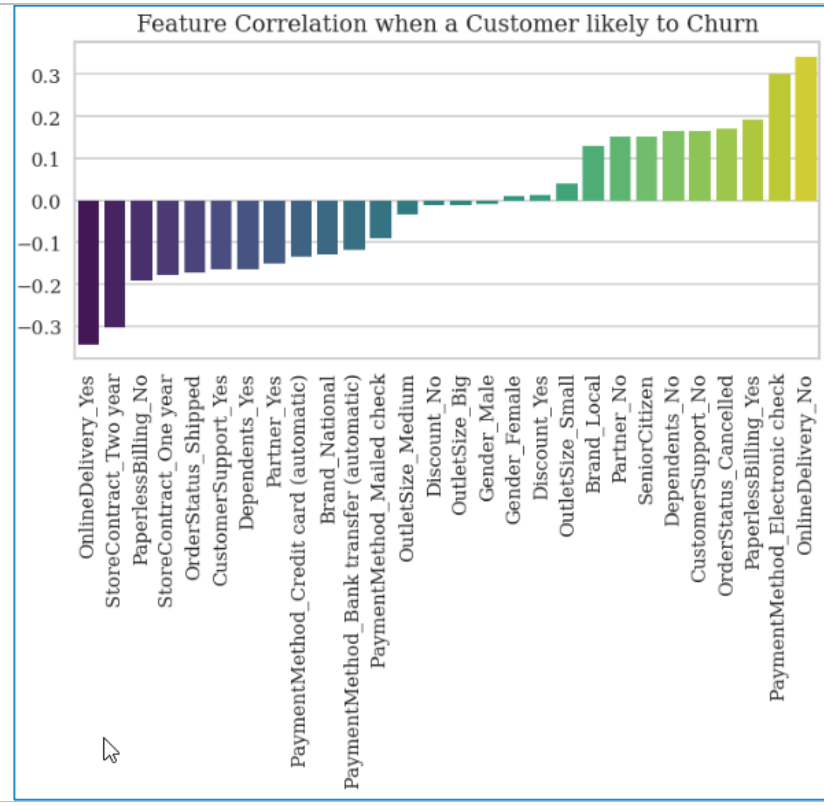
def clean_data(df):

    # Drop rows with missing data in column: 'TotalAmount'
    df = df.dropna(subset=['TotalAmount'])
    # Drop column: 'CustomerID'
    df = df.drop(columns=['CustomerID'])
    # Change column type to float32 for column: 'UnitPrice'
    df = df.astype({'UnitPrice': 'float32'})
    # Change column type to int32 for column: 'tenure'
    df = df.astype({'tenure': 'int32'})
    return df

df = clean_data(df.copy())
df['TotalAmount'] = pd.to_numeric(df['TotalAmount'], errors="coerce")
df.head()</code></pre></td>
</tr>
</table>

3.4.3 Exploring data to understand the features available, any interesting patterns in the data

Narrative	Steps	Screenshot
With the data prepared, Miguel begins exploring the data to understand the patterns it contains. On this scatter plot, he can see whether a customer is more likely to churn based on the tenure in months and the total amount spent. So, the churn rate is low if customer tenure is high, and they spend more.	1. **Point** to cell 8.	

Based on the data above, he tries to understand the correlation between data features such as Online Delivery, Payment Methods, Customer Support, etc. influencing customer churn. On this plot, Miguel can see the feature correlation when the target variable "Churn" is labeled as "Yes." He can also see that “OnlineDelivery_No” and “PaymentMethod_Electronic check” has the maximum influence on the Churn.	2. **Point** to cell 9.	

3.4.4 Build an ML model and track experiment run with MLflow

Narrative	Steps	Screenshot

Next, Miguel starts building the base model and tunes it based on the exploratory data analysis.
First, he initializes all the models that he wants to run and test.
	1. **Point** to cell 10.
	Note:
	In order to see the experiments and models you will have to run all the cells in this notebook. This may take upto 20 minutes to execute.
		4.1: Build Base Model and Hyperparameter Tune
		`logReg = LogisticRegression()`
		`randForest = RandomForestClassifier()`
		`gbt = GBTClassifier()`
		`lightgbm = LightGBMClassifier()`
		[10]

Miguel now mentions the Experiment Name, Model Name and Target column and mentions all the models that he wishes to run.	2. **Point** to cell 11.	`EXPERIMENT_NAME = "07-churn-prediction-experiment"`
		`MODEL_NAME = "08-churn-prediction-model-using-lightgbm"`
		`TARGET_COL = "Churn"`
		
		`smlmodels = [logReg,randForest,gbt,lightgbm]`
		`mmlmodels = [TrainClassifier(model=model, labelCol=TARGET_COL) for model in smlmodels]`
		[11]

The next step is to define the hyperparameters, as we know that all models have different hyperparameters, Miguel can mention which parameters he needs to tune.
Here he can see that, thanks to Synapse ML, he can add various parameters in one go.
	3. **Point** to cell 12.	`paramBuilder = (`
		`HyperparamBuilder()`
		`.addHyperparam(logReg, logReg.regParam, RangeHyperParam(0.1, 0.3))`
		`.addHyperparam(randForest, randForest.numTrees, DiscreteHyperParam([5, 10]))`
		`.addHyperparam(randForest, randForest.maxDepth, DiscreteHyperParam([3, 5]))`
		`.addHyperparam(gbt, gbt.maxBins, RangeHyperParam(8, 16))`
		`.addHyperparam(gbt, gbt.maxDepth, DiscreteHyperParam([3, 5]))`
		`.addHyperparam(lightgbm, lightgbm.baggingFraction, DiscreteHyperParam([0.8,1]))`
		`)`
		[12]

To build the final model, Miguel chooses **LightGBM,** an open-source boosting algorithm for gradient boosting machines developed by Microsoft. It uses decision-tree based learning algorithms.
 	
It is fast, efficient and produces better results than other boosting algorithms.
	4. **Point** to cell 17.	`# Best model parameters from tuned model`
		
		`lgb_model = LGBMClassifier(`
		
		`bagging_fraction=1.0, bagging_freq=7,`
		`boosting_type='gbdt', class_weight=None,`
		`colsample_bytree=1.0, feature_fraction=0.4,`
		`importance_type='split', learning_rate=0.05,`
		`max_depth=-1, min_child_samples=71,`
		`)`
		
		`lgb_model`
		
		[17]

Then he begins training the model with the training dataset.
 	
In this cell a churn_model is being fitted using X_train and y_train. The model is being evaluated using X_test and y_test with the metric "auc" and a callback with logging the evaluation every 10 steps.
	5. **Point** to cell 18.	`print(f"\n\nTraining using the train dataset Started:\n")`
		`churn_model = lgb_model.fit(`
		`X_train,`
		`y_train,`
		`eval_set=[(X_test, y_test)],`
		`eval_metric="auc",`
		`callbacks=[`
		`lgb.log_evaluation(10),`
		`],`
		`)`
		
		[18]

<table>
<tr>
<td>

Understanding Model - Feature Importance

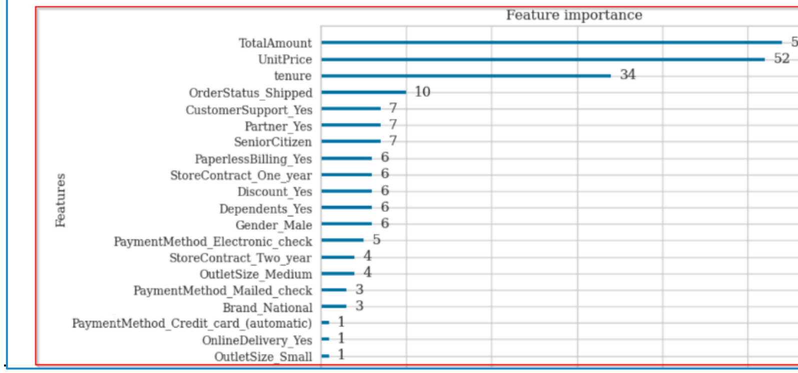
Feature Importance refers to techniques that calculate a score for all the input features for a given model — the scores simply represent the “importance” of each feature. A higher score means that the specific feature will have a larger effect on the model that is being used to predict a certain variable.

</td>
<td>

6. **Point** to cell 19.

</td>
<td>

lgb.plot_importance(churn_model)
[19] ✓ - Command executed in 451 ms by rajib kumar on 5:03:01 PM, 6/09/23
<AxesSubplot: title={'center': 'Feature importance'}, xlabel='Feature importance', ylabel='Features'>



</td>
</tr>
<tr>
<td>

Then, he defines a function that takes in a model and a test set as its parameters.

It creates a DataFrame from the target column and the predictions. The function then returns the DataFrame which contains the predictions.

</td>
<td>

7. **Point** to cell 21.

</td>
<td>

```
def prediction_to_spark(model, test):
    feature_cols = [c for c in test.columns.tolist() if c not in [TARGET_COL]]
    predictions = model.predict(test[feature_cols], num_iteration=model.best_iteration_)
    predictions = tuple(zip(test[TARGET_COL].tolist(), predictions.tolist()))
    dataColumns = [TARGET_COL, "prediction"]
    predictions = (
        spark.createDataFrame(data=predictions, schema=dataColumns)
        .withColumn(TARGET_COL, col(TARGET_COL).cast(IntegerType()))
        .withColumn("prediction", col("prediction").cast(DoubleType()))
    )

    return predictions
```

[21] ✓ - Command executed in 394 ms by rajib kumar on 5:03:03 PM, 6/09/23

</td>
</tr>
<tr>
<td>

Now Miguel is evaluating how successful a classification problem is and where it makes mistakes.

</td>
<td>

8. **Point** to cell 25.

</td>
<td>

```
cm = metrics.select("confusion_matrix").collect()[0][0].toArray()
print(cm)
sns.set(rc={"figure.figsize": (6, 4.5)})
ax = sns.heatmap(cm, annot=True, fmt=".20g")
ax.set_title("Confusion Matrix")
ax.set_xlabel("Predicted label")
ax.set_ylabel("True label")
```

[25] ✓ - Command executed in 1 sec 114 ms by rajib kumar on 5:03:10 PM, 6/09/23

```
[[4673.  501.]
 [ 872.  997.]]
```

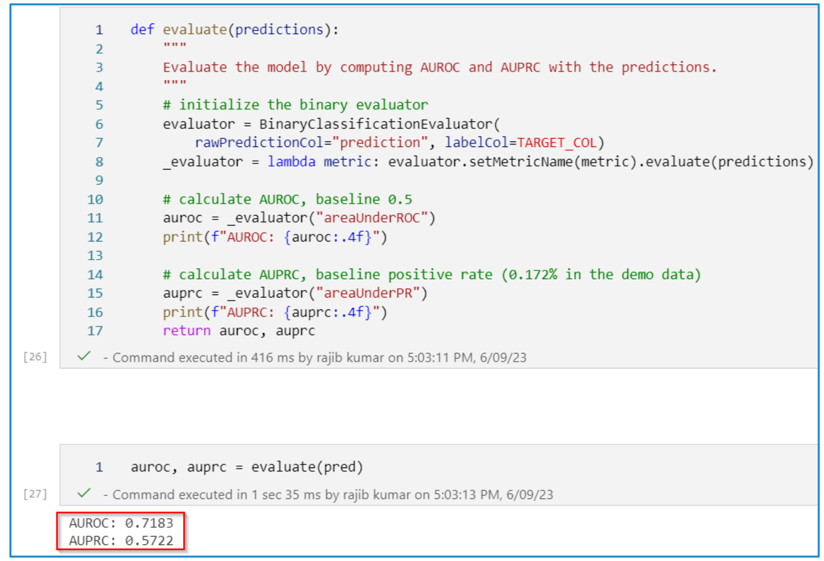
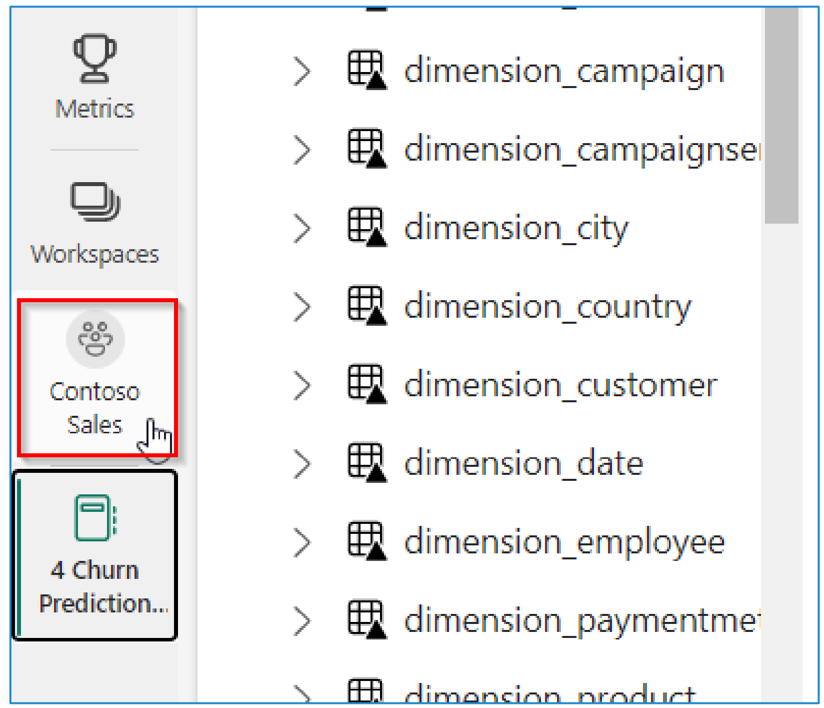
</td>
</tr>
</table>

<table>
<tr>
<td>This function is used to evaluate the model's performance. It uses BinaryClassificationEvaluator to calculate the AUROC and AUPRC metrics. The function prints out the AUROC and AUPRC scores and returns them as a tuple.

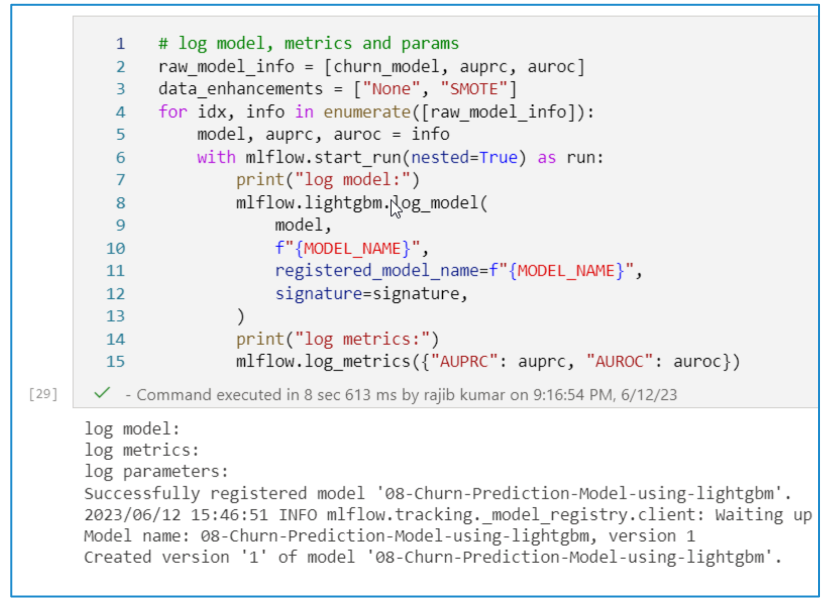
AUROC - Area Under the Receiver Operating Characteristic Curve.

AUPRC - Area Under the Precision-Recall Curve.

These metrics are used to evaluate the performance of a binary classification model.</td>
<td>9. Point to cell 26.

10. Point to cell 27.</td>
<td>
</td>
</tr>
<tr>
<td>Let’s go back to Contoso Sales workspace.</td>
<td>11. Click on Contoso Sales.</td>
<td>
</td>
</tr>
</table>

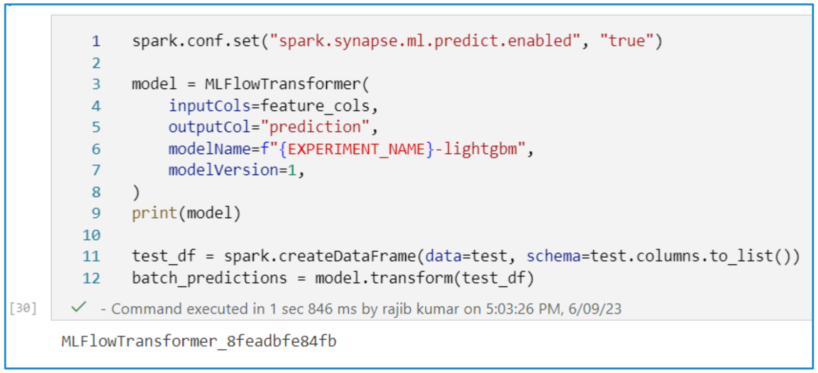
3.4.5 Log ML model in the built-in model registry using MLflow

Narrative	Steps	Screenshot
The code in the screenshot is to log the model, metrics and parameters used for the churn model. The model, metrics and parameters are stored as variables (churn_model, auprc, auroc) and are then logged using the MLflow library. The code logs the model using the log_model function, and the metrics using the log_metrics functions respectively.	1. **Point** to cell 29.	

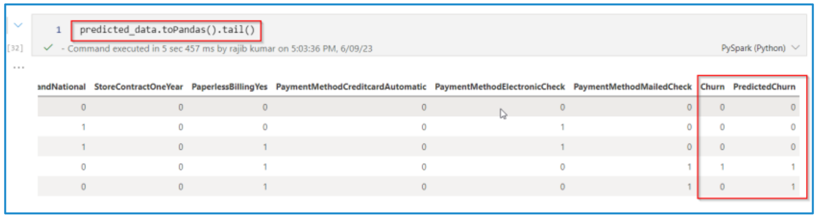
3.4.6 Run batch scoring on Spark with scalable PREDICT UDF

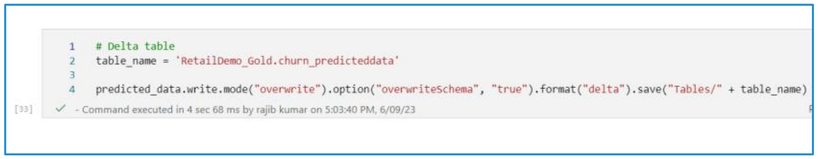
Narrative	Steps	Screenshot

<table>
<tr>
<td>PREDICT supports MLflow-packaged models in the Microsoft Fabric registry.

Creating an MLflowTransformer object to load the model for inferencing, use this object to generate batch predictions at the end of the procedure.</td>
<td>1. Point to cell 30.</td>
<td>
</td>
</tr>
</table>

3.4.7 Store predictions in the Lakehouse (Delta parquet)

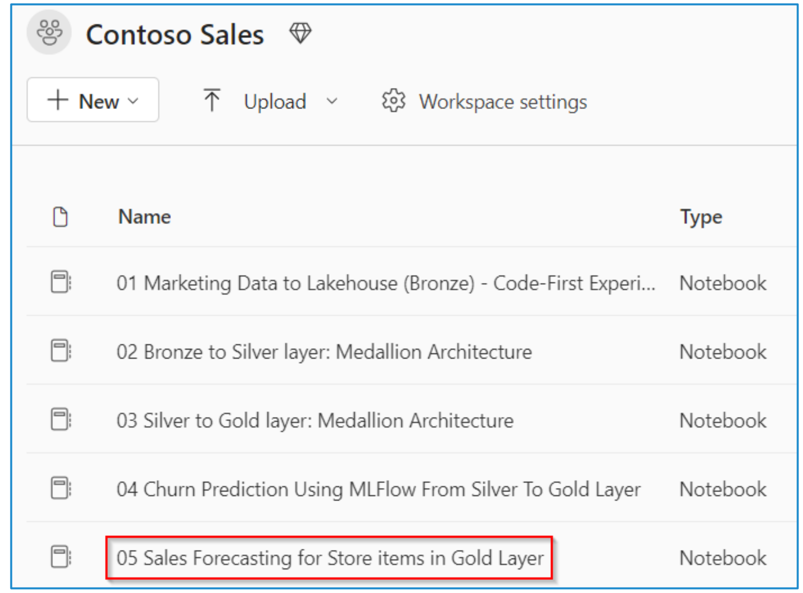
<table>
<tr><th>Narrative</th><th>Steps</th><th>Screenshot</th></tr>
<tr>
<td>This code converts the Spark DF into the pandas dataframe to see the predictions.</td>
<td>1. Point to cell 32.</td>
<td>
</td>
</tr>
<tr>
<td>Save predictions to lakehouse delta table (Gold layer) predicted_data for downstream consumption and analysis.

Storing predictions in the Lakehouse (Delta parquet) format allows for efficient and scalable storage of prediction results.</td>
<td>2. Point to cell 33.</td>
<td>
</td>
</tr>
</table>

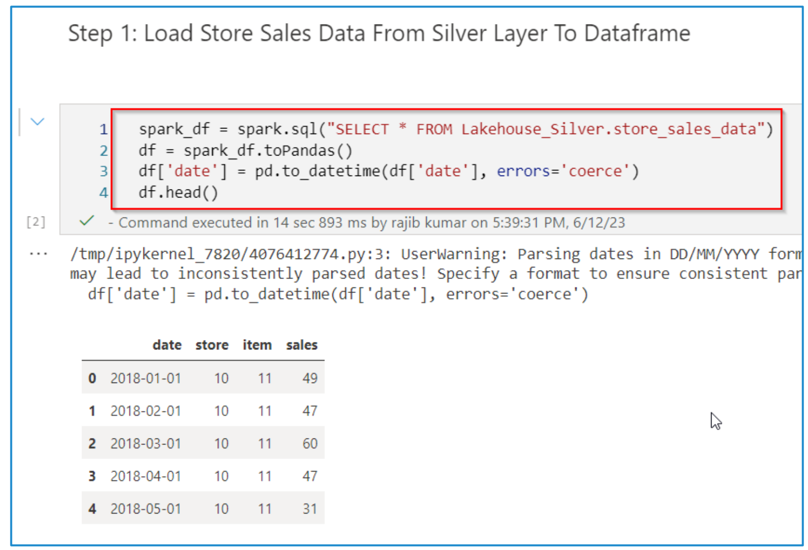
3.4.8 Sales Forecasting for Store items in Gold Layer

Narrative	Steps	Screenshot

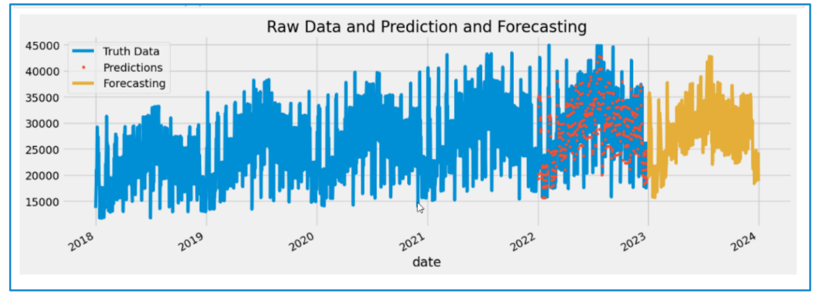
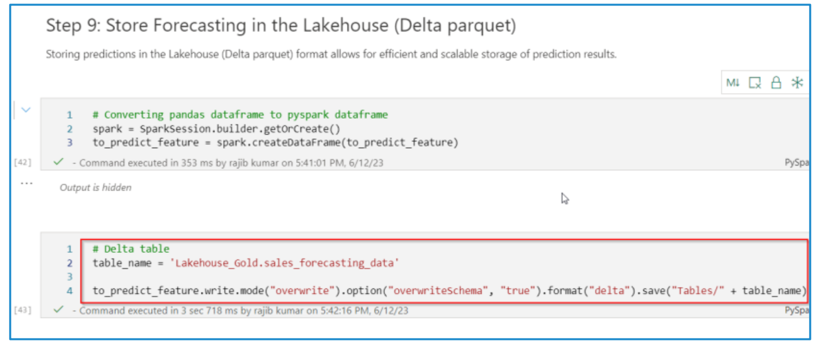
<table>
<tr>
<td>
Contoso used a sales forecasting model to validate if their approach to reduce churn and improve sales was effective.

Let’s look at a scenario where Miguel forecasts sales by taking data from the silver layer Delta table as input. This forecasted sales data will be further ingested to the gold layer Delta table for end users to consume for business purposes.
</td>
<td>
1. Click on ‘05 Sales Forecasting for Store items in Gold Layer’ Notebook.
</td>
<td>


</td>
</tr>
<tr>
<td>
Miguel loads sales data from a silver layer Delta table into a Microsoft Fabric Notebook, allowing him to leverage the capabilities of Delta Lake and perform further analysis, or use the data in downstream processes.

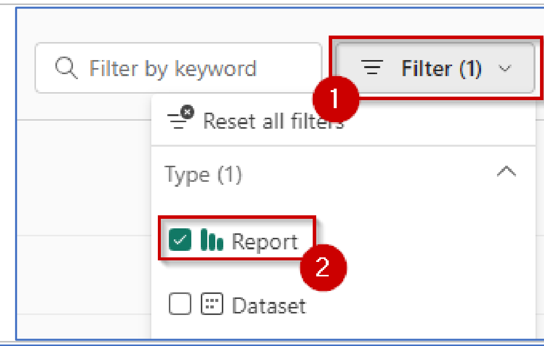
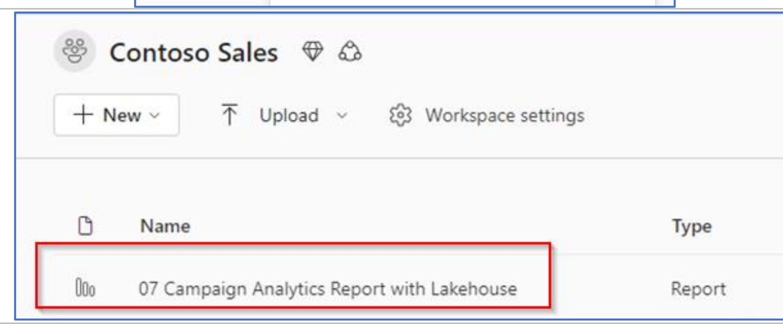
Note: Walk through notebook cells to see sales forecasting ML model creation experience.
</td>
<td>
2. Scroll down to cell [2]: Load Store sales data from silver layer to Dataframe.
</td>
<td>


</td>
</tr>
</table>

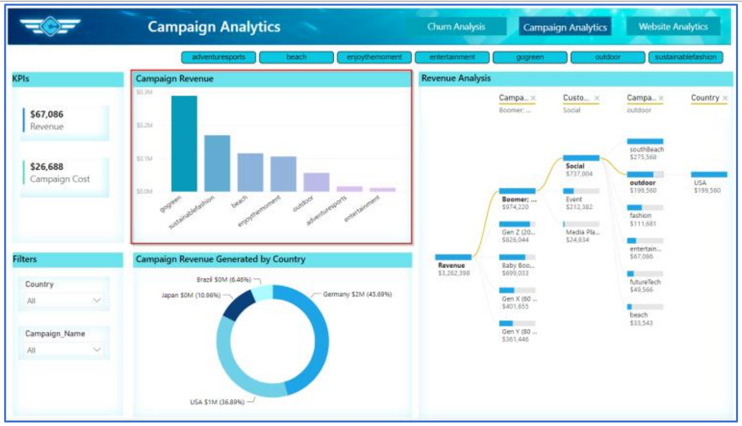
Miguel leverages the sales forecasting model to generate predictions for the next one year and use them to make informed business decisions.	3. Scroll down to cell [13]: Forecasting the sales for year 2023 – in yellow color.	
He ingests item sales forecasted data into the Gold Lakehouse in the Delta Parquet format and leverages Power BI to create interactive reports and visualizations. This enables him to gain valuable insights from the data.	4. **Go to** Store items sales forecasting in the Gold Lakehouse (Delta parquet) format.	-- 

3.4.9 Customer Churn Analysis, Campaign Analytics, Website Analytics with Power BI report

Narrative	Steps	Screenshot

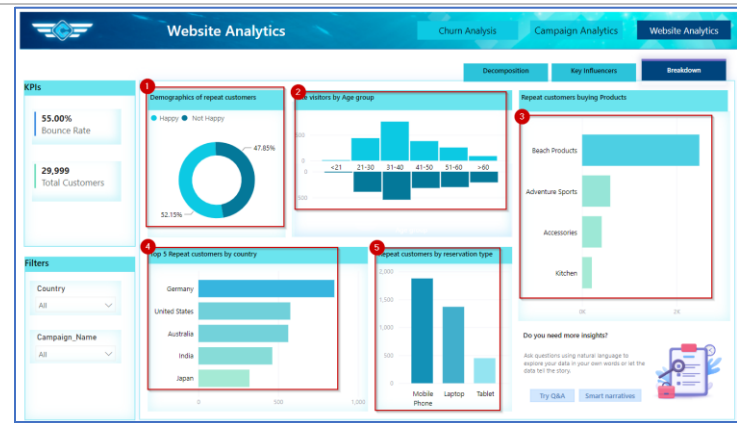
<table>
<tr>
<td>Let's go to the Contoso Sales workspace.</td>
<td>In the left navigation bar:
1. Select Workspaces.
2. Select Contoso Sales workspace.</td>
<td>
</td>
</tr>
<tr>
<td>We'll set the filter to Report.</td>
<td>3. Select Filter.
4. Select Report.</td>
<td>
</td>
</tr>
<tr>
<td>Then, let's open the Customer Churn Report that is built with the data coming from the Lakehouse.</td>
<td>5. Select "07 Campaign Analytics" report.</td>
<td>
</td>
</tr>
</table>

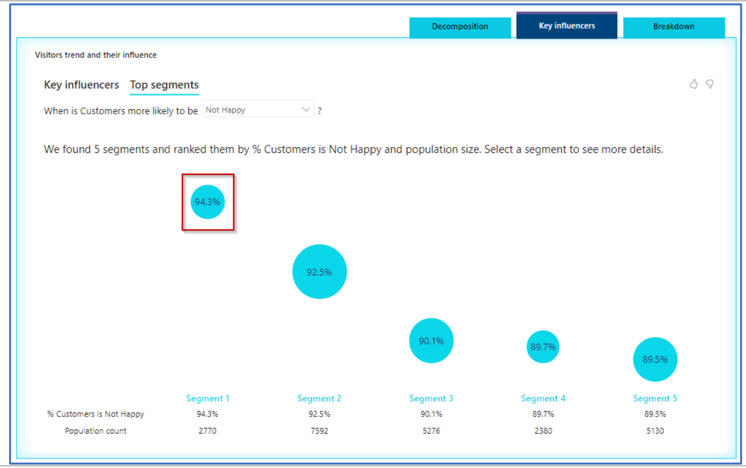
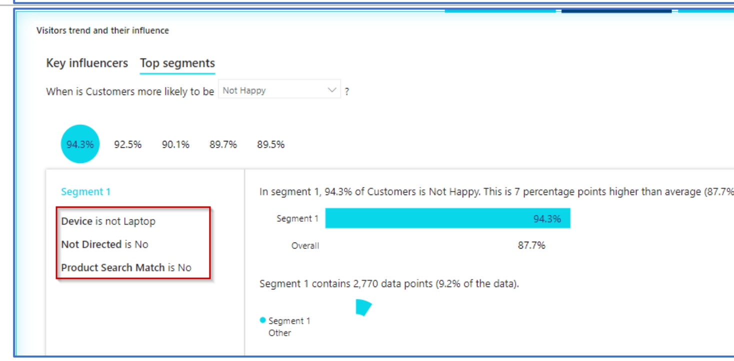
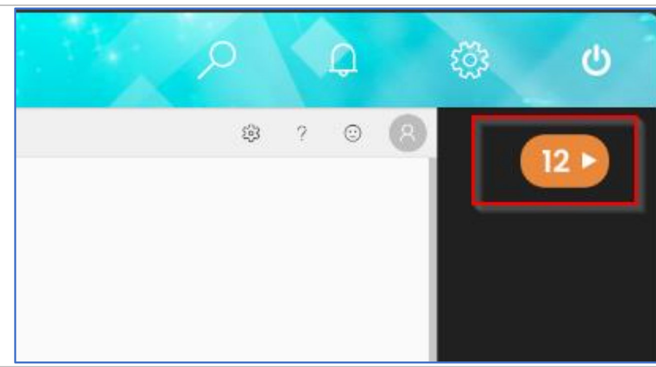
Now that Contoso's data is available in the gold layer, the data in the gold layer is in a format that is easy for business users to navigate. This data is highly performant. Wendy, the Business Analyst, is able to leverage Power BI to get actionable insights. On this report, Wendy can see which customers are more likely to churn. She can also see that customers love marketing campaigns like #GoGreen, #EnjoyTheMoment, #Beach etc.	6. **Talk** about the Power BI report built on top of the data in gold layer of the lakehouse.	
Here on the scatter plot (chart# 1) we can see that most of Contoso's loyal customers are those who have been shopping with them for a longer period of time and tend to larger purchases. On the other hand, we see that most of the customers who are likely to churn are those with less tenure (or shopping history) and tend to make smaller purchases. In the word cloud chart (chart #2), we see that #GoGreen and #SustainableFashion are the largest campaigns to generate revenue for Customers. On the bar chart (chart #3) we can confirm that the most popular hashtags associated with Contoso's campaigns on Twitter are for 'GoGreen' and 'SustainableFashion'.	7. **Talk** about the various charts and their data in the Customer Churn Report.	

On the other hand, the horizontal bar chart (chart #4) explains the probability of customer churn depending on various conditions such as the online delivery channel, and customer support experience. It seems the highest churn is in their online channel!		
Let's take a closer look at the Campaign Analytics.	8. **Click** Campaign Analytics.	
Here on this bar chart, the most popular hashtags being tweeted are 'GoGreen' and 'SustainableFashion'. So, this data insight confirms that most of the customers are environmentally conscious and want to purchase climate-friendly green products which are sustainable as well. The donut chart shows the revenue generated in each country after running the new campaigns. Clearly, these campaigns were a huge success in Germany and the USA. This decomposition tree has different levels to show campaigns reaching different age groups through social media, events, and emails.	9. **Talk** about the Revenue generated from various campaigns.	
Let's take a look at the Website Analytics.	10. **Click** Website Analytics.	

<table>
<tr>
<td>First, Wendy notices that the bounce rate for Contoso is as high as 55%. So, she clicks on the unhappy customer segment in the donut chart to understand the factors influencing the high bounce rate.

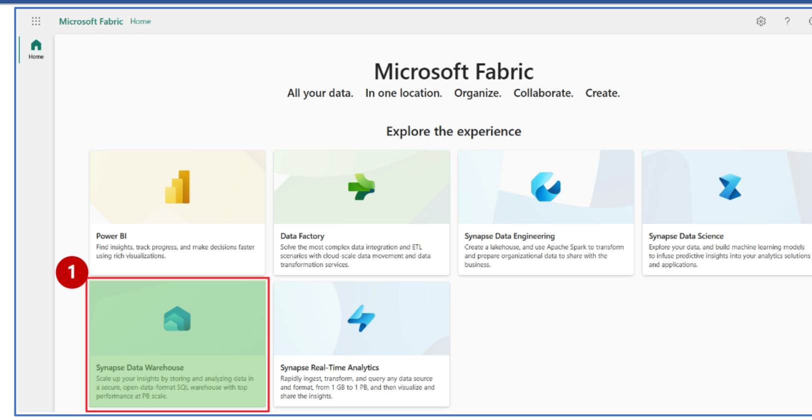
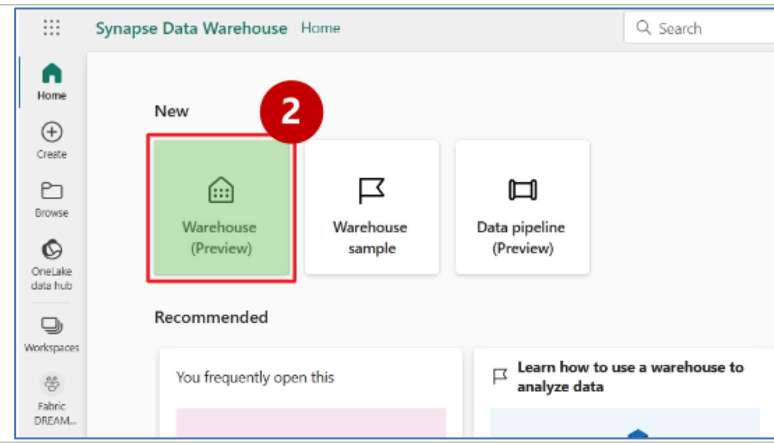
Here’s what she finds out about the customers who are navigating away from their online store:
1. They are millennials.
2. They are looking for products under the beach and sports categories.
3. They mostly reside in Germany, the USA, and India.
4. Most of them are using their mobile devices to shop online.

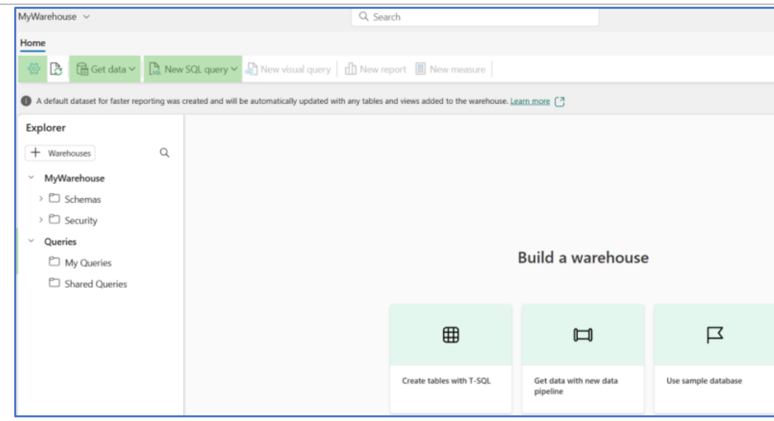
Wendy now knows where Contoso needs to focus so they can improve their Bounce Rate.</td>
<td>11. Click on the Not Happy section of the donut chart (marked #1).
12. Talk about the age group of people unhappy with Contoso website (marked #2).
13. Talk about the products they are looking for on the website (marked #3).
14. Talk about the countries where these visitors reside (marked #4).
15. Talk about the type of devices the visitors are using to surf the website (marked #5).</td>
<td>
</td>
</tr>
<tr>
<td>Now let’s navigate to the key influencers.</td>
<td>16. Click on Key Influencers.</td>
<td>
</td>
</tr>
</table>

One interesting perspective is to look at which segments are impacting the "Not Happy" set of customers to determine where more focus could be applied to improve customer satisfaction.	17. **Click** on the first bubble i.e., Segment 1.	
Segment 1 shows that 94.5 % of the "not happy" customers experienced failed product searches. By looking at the customer segment category, these are millennials who are using their mobile devices for shopping.	18. **Talk** about this segment people and the key influencing factors. 19. **Agree** to the disclaimer in the chat bot.	
Now let's see how Eva, the Data Engineer, creates a new Data Warehouse.	20. **Click** on arrow 12. **<<This step will take you to a click-by-click embedded in the web app.>>**	

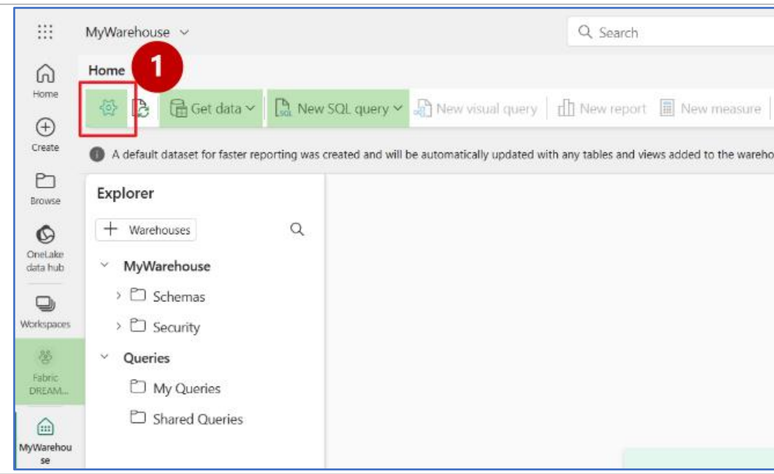
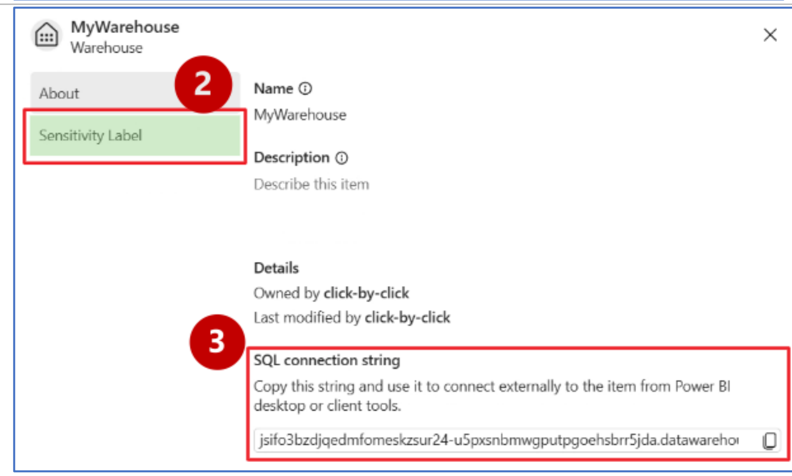
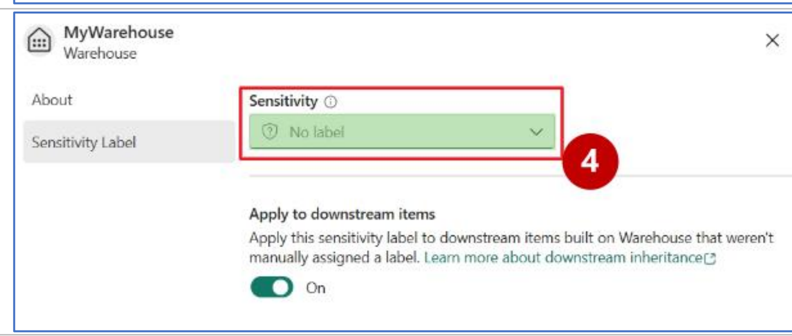
3.5 Microsoft Fabric for Data Warehouse experience

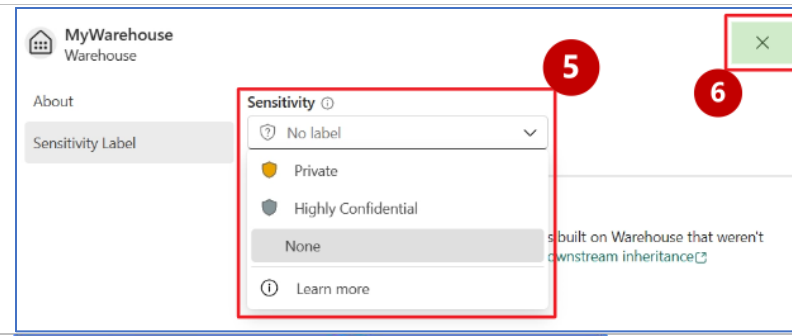
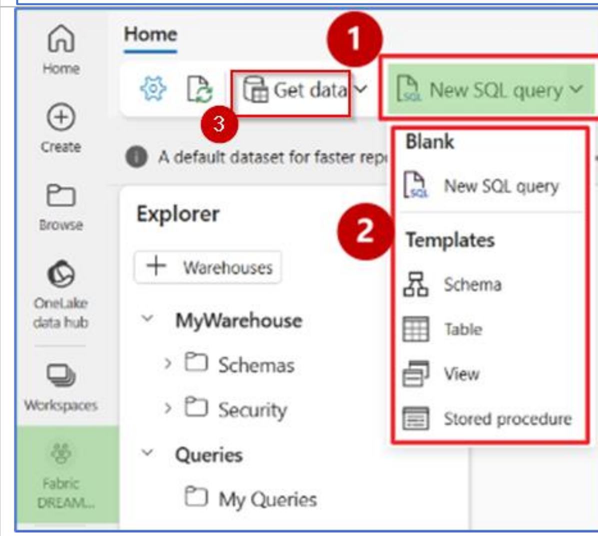
3.5.1 Create a Data Warehouse [Click – by – click]

Narrative	Steps	Screenshot
Microsoft Fabric introduces a lake centric data warehouse built on an enterprise grade distributed processing engine that enables industry leading performance at scale while eliminating the need for configuration and management. One of the pipelines Eva used earlier landed the marketing data in the Sales warehouse. Once the data is loaded, Eva is able to run large scale analytical queries using T-SQL and modify data using T-SQL commands. Let us experience how she was able to arrive at this point.	**<<This section is a click-by-click embedded in the web app>>** 1. **Click** on the Synapse Data Warehouse experience. ***Note:*** *For this demo, data warehouses are already created in the workspace. To demonstrate warehouse creation, a click-by-click is embedded in the web app to simulate the actual user experience.*	
Let's step into her shoes to act as a data engineer and create a new Warehouse. We'll go to Synapse Data Warehouse to create the first Data Warehouse. Once in the Data Warehouse experience, Eva can simply select the Warehouse option to get started.	2. **Click on** the Warehouse (Preview).	

In the past, the CDO's team had to invest a lot of effort to create simple data warehouses to support business intelligence and reporting needs. Now, with Microsoft Fabric it allows Contoso to have a completely traditional warehouse experience without any data movement! Warehouse creation is also just as simple as the lakehouse creation!	3. On the New warehouse dialog, **Click on** the name for a new data warehouse.	
All Eva has to do is provide a name for the Warehouse while creating the Data warehouse.	4. **Click on** Create. 5. **View** the data warehouse "MyWarehouse". **Note:** This is a click-by-click step only. Please do not create an actual warehouse in Microsoft Fabric to avoid multiple asset creation.	
In just a few seconds, the Data Warehouse is created. And there Eva goes! It didn't take any time to provision. Warehouse provisioning and configuration is handled at the Microsoft Fabric level, so Eva doesn't have to worry about any of it.	6. **Show** the warehouse.	

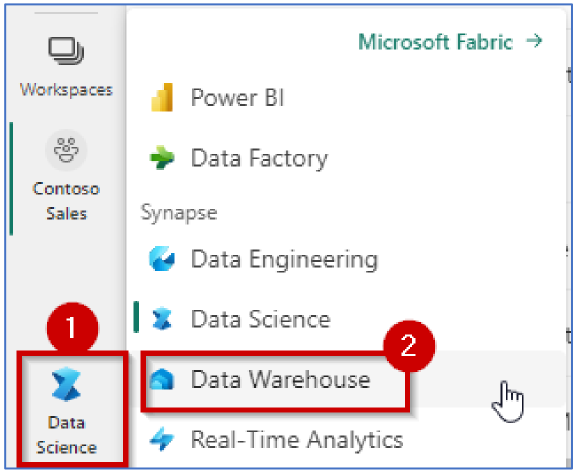
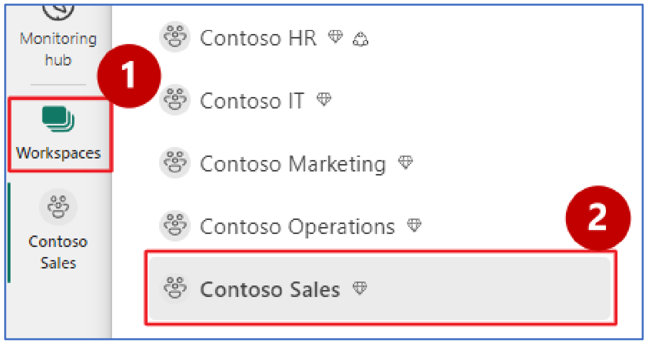
<table>
<tr>
<td>Let’s look at the Warehouse settings.</td>
<td>7. Click on the settings icon.

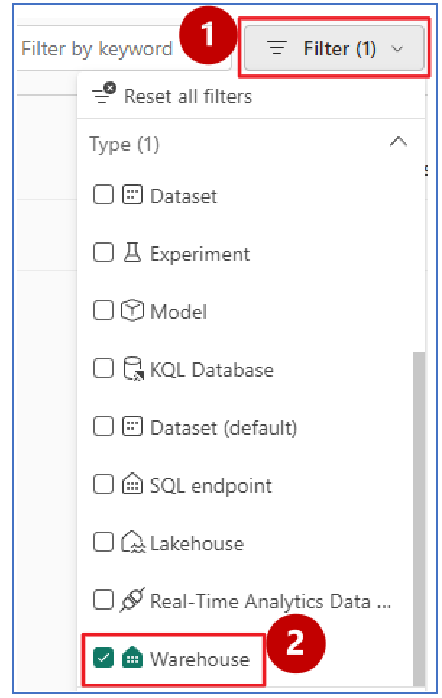
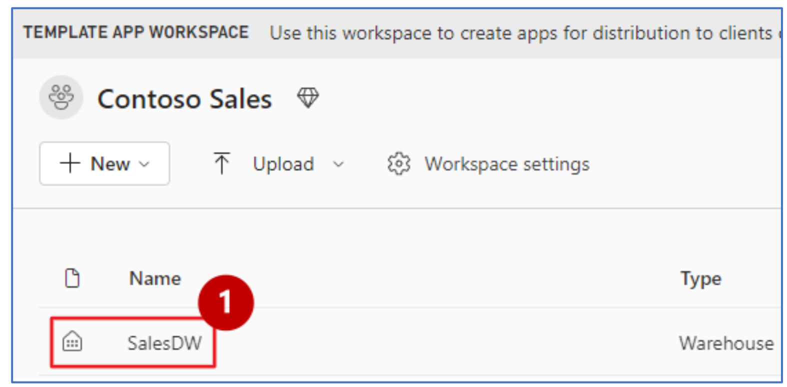
Note: Wait till you see the Settings icon blink.</td>
<td>
</td>
</tr>
<tr>
<td>Eva can connect a client tool to the workspace’s SQL endpoint for authoring T-SQL scripts and executing SQL Queries like Azure Data Studio for T-SQL development.</td>
<td>8. Click on Sensitivity Label.
9. View SQL connection string.</td>
<td>
</td>
</tr>
<tr>
<td>Just like the lakehouse, the Warehouse is also a part of OneSecurity and Eva can apply the necessary sensitivity labels to the Warehouse.</td>
<td>10. Click on Sensitivity to view the existing labels.</td>
<td>
</td>
</tr>
</table>

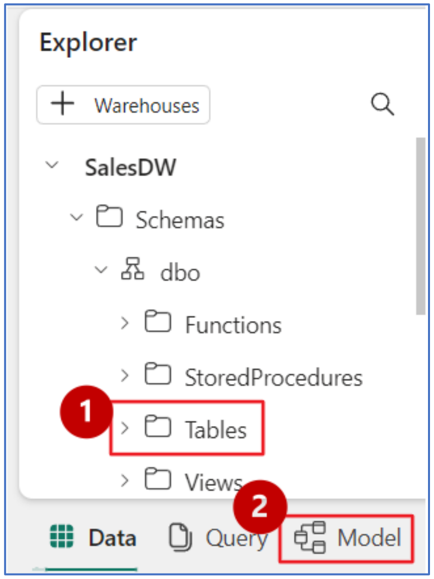
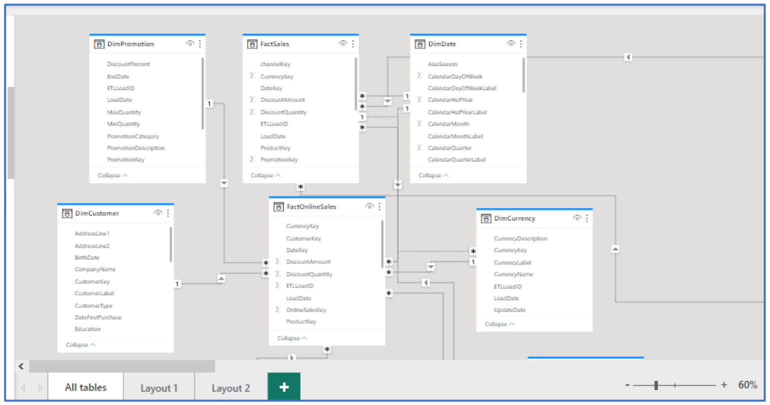
These labels will be inherited by all the related components of this lakehouse.	11. **View the** existing labels 'Private, Highly Confidential & None'. 12. **Click** on the close icon.	
The Data Warehouse will allow Eva to build model data using tables and views, run **T-SQL** to query data across the Data Warehouse and Lakehouse, use T-SQL to perform DML operations on data inside the Data Warehouse, and serve reporting layers like Power BI. In the data warehouse, data is organized in Schemas, View, Stored procedure, and Table.	13. **Click on** New SQL query and show the available menu. 14. **Show** the 'Get data' option. **<<Switch tab to Microsoft Fabric *(opened earlier)*.>>**	

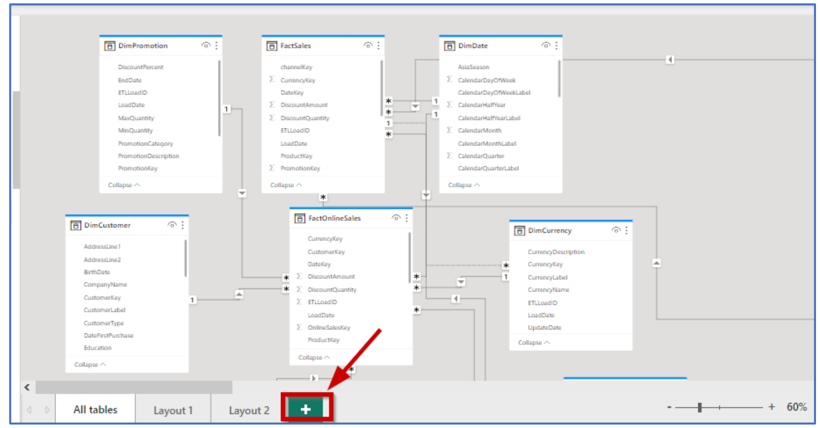
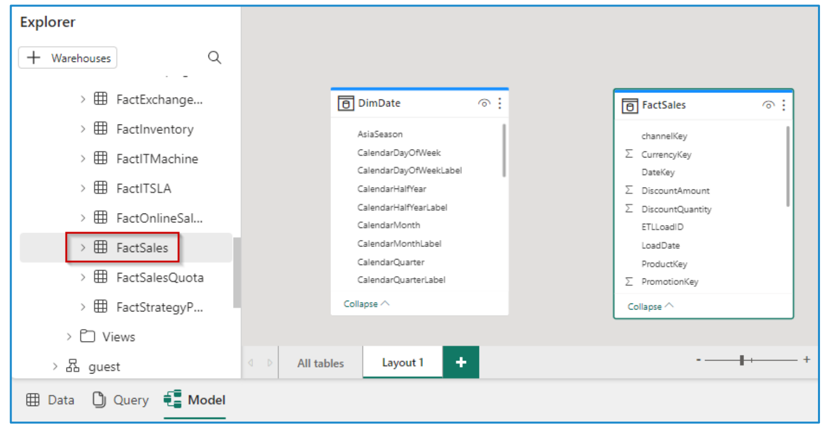
3.5.2 Show data model creation

Narrative	Steps	Screenshot

Let's navigate to the Data Warehouse using the experience switcher.	**<<The following steps are performed in Microsoft Fabric>>** **In the left navigation bar,** 1. **Click** on experience switcher (marked #1). 2. **Select** Data Warehouse.	
We'll switch to the "Contoso Sales" workspace.	3. **Click** on Workspace. 4. **Select** 'Contoso Sales' workspace.	

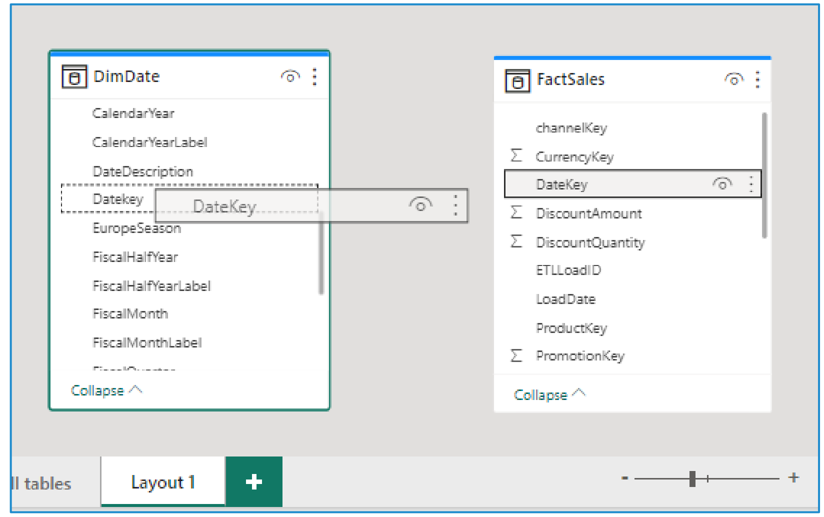
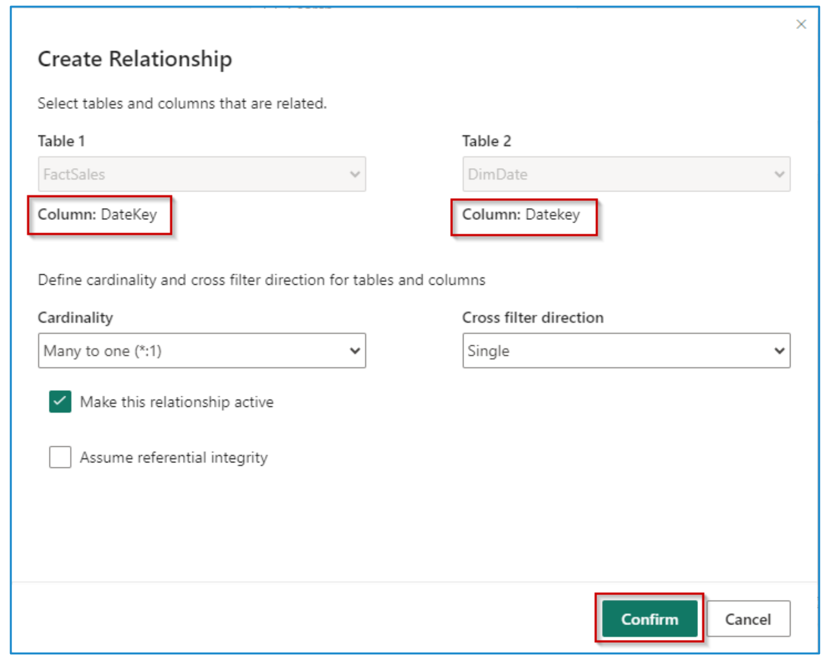
Now we will set the filter to Warehouse.	5. **Click** on Filter. 6. **Select** Warehouse.	
And click on pre-created data warehouse SalesDW.	7. **Click** on Warehouse 'salesDW'.	

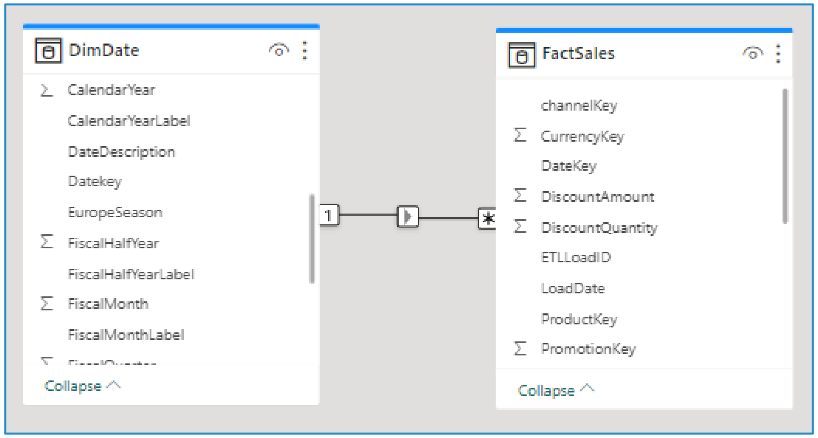
Eva and Serena, the data analysts can conduct their analysis on the tables in this warehouse. For Data modeling, Serena can analyze the process and define all the different business data, as well as the relationships between those bits of data. Let's view the existing data model.	8. **Expand** table. 9. **Click** on Model.	
The star schema model, which is optimized for querying large data sets, is a multi-dimensional data model used to organize data in a database so that it is easy to understand and analyze.	10. **Select** All tables.	

Let’s create the new model.	11. **Click** on + icon to create new model.	
We’ll search for and select one fact table and one dimension table and drag them to the canvas. Then we showcase an example to create a basic relationship between two tables.	12. **Search** for table “DimDate”. 13. **Search** for table “FactSales”. 14. **Drag** it to the canvas area.	

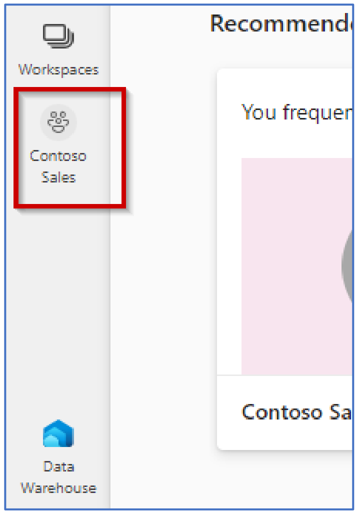
<table>
<tr>
<td>We establish the relationship between these tables with the matching keys.</td>
<td>15. Scroll down and search for Datekey in “DimDate” table.

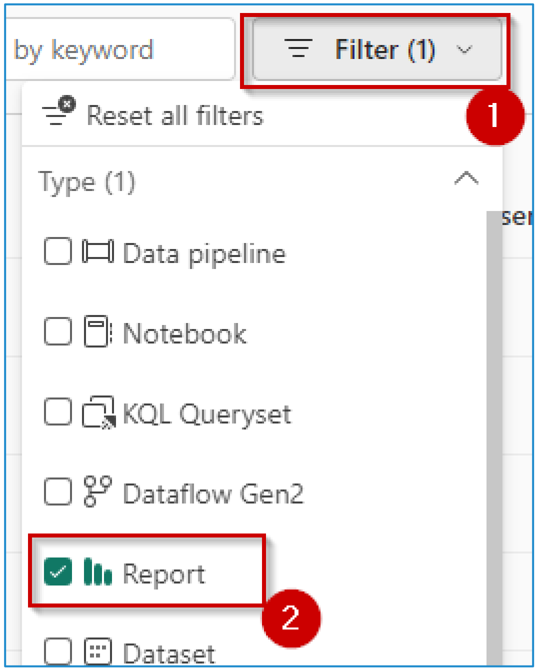
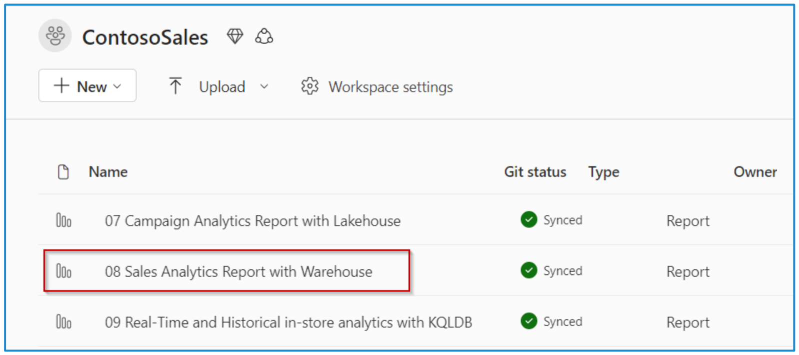
16. Click on DateKey from “FactSales” table and drag it to Datekey on “DimDate” table.

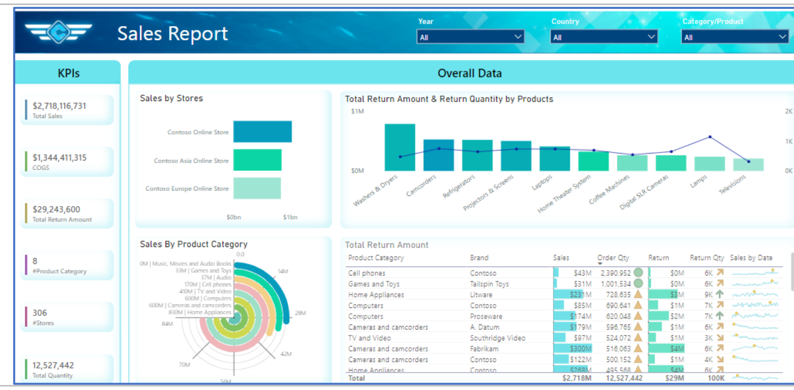
Note: Refrain from creating relationships between tables if relationships have already been established to prevent the error: There’s already a relationship between these two columns.</td>
<td>
</td>
</tr>
<tr>
<td>We verify the tables and respective column names match, then click confirm.</td>
<td>17. In the pop-up window verify the column names and click on Confirm.</td>
<td>
</td>
</tr>
</table>

The data analyst can make these relationships active, create new measures and also use these models to build exciting reports. Using this model we create a Power BI report for visualization, let's have a look at the Power BI which are coming from the warehouse.	18. We can **see** a relationship created between the two tables.	

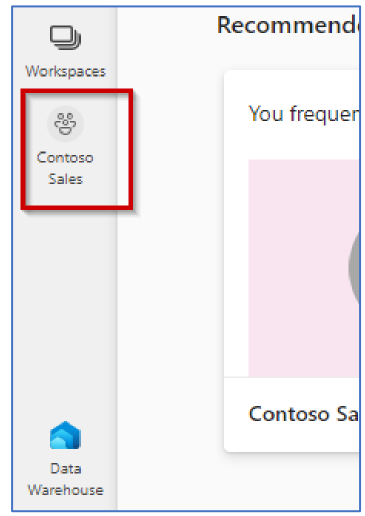
3.5.3 Show Power BI Report

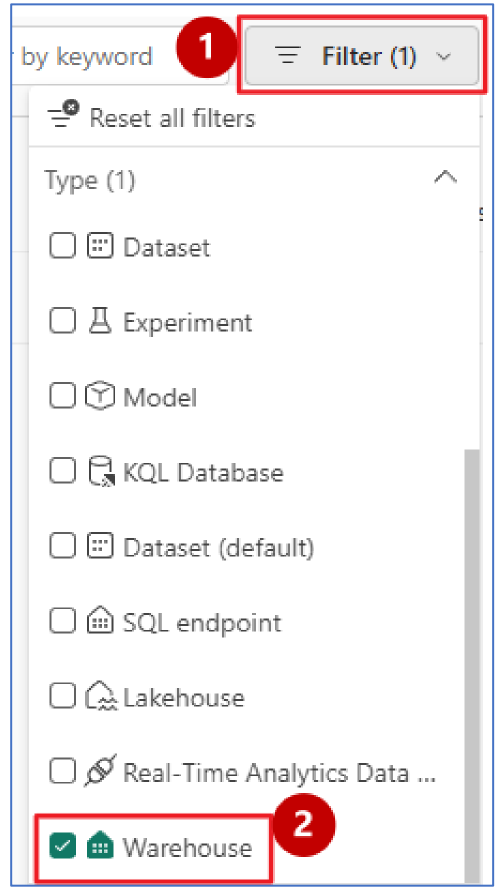
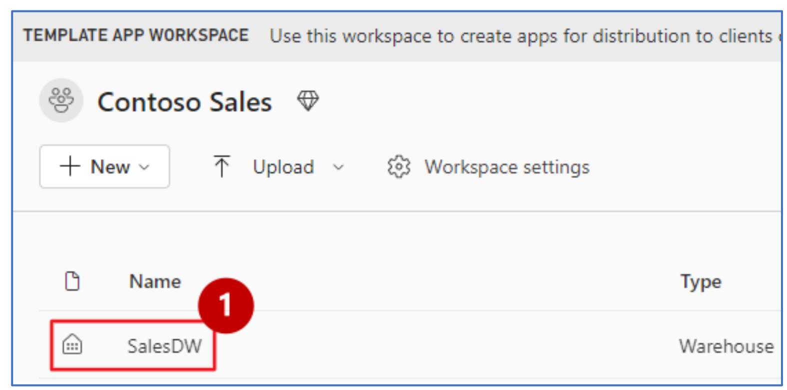
Narrative	Steps	Screenshot
Now let's navigate to the Contoso Sales workspace.	1. In the left navigation bar, **Select** 'Contoso Sales' workspace.	

Filter the assets for Report.	2. **Click** on Filter. 3. **Select** Report.	
Open the report "08 Sales Analytics Report with Warehouse."	4. **Click** on the report "08 Sales Analytics Report with Warehouse".	

This report has various KPIs, namely Sales by Stores, Sales by Product Category, Total Return Amount by Month etc. The report works like any other report built over a relational database with options to have import mode as well as direct query mode.	5. **Show** the report.	

3.5.4 Visual Query

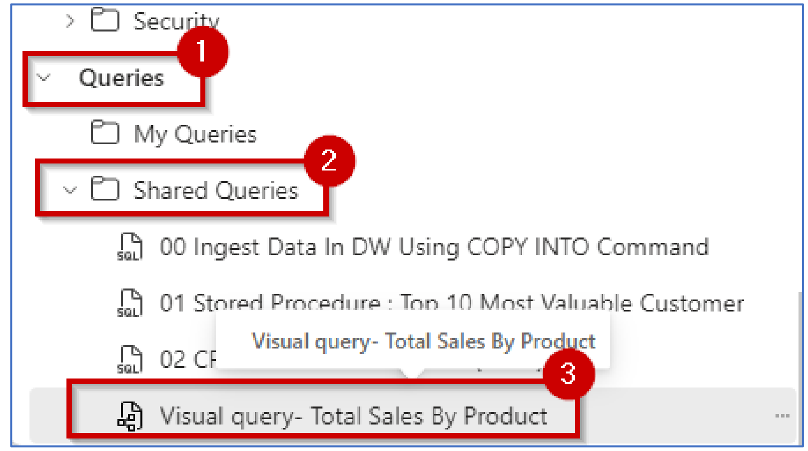
Narrative	Steps	Screenshot
Let's see how Serena, the data analyst, creates Visual Query with Power Platform Experience.	1. In the left navigation bar, **Select** 'Contoso Sales' workspace.	

<table>
<tr>
<td>Back to using our handy filter to find the relevant asset.</td>
<td>2. Click on Filter.
3. Select Warehouse.</td>
<td>
</td>
</tr>
<tr>
<td>Open the SalesDW warehouse.</td>
<td>4. Click on Warehouse ‘SalesDW’.</td>
<td>
</td>
</tr>
</table>

<table>
<tr>
<td>Now Serena, as the Data Analyst, can choose to write T-SQL queries for creating aggregate data or use visual queries based on the comfort.</td>
<td>

5. Under Explorer section, **expand** Queries.
6. **Expand** Shared Queries.
7. **Select** Visual Query 'Visual query- Total Sales By Product'.

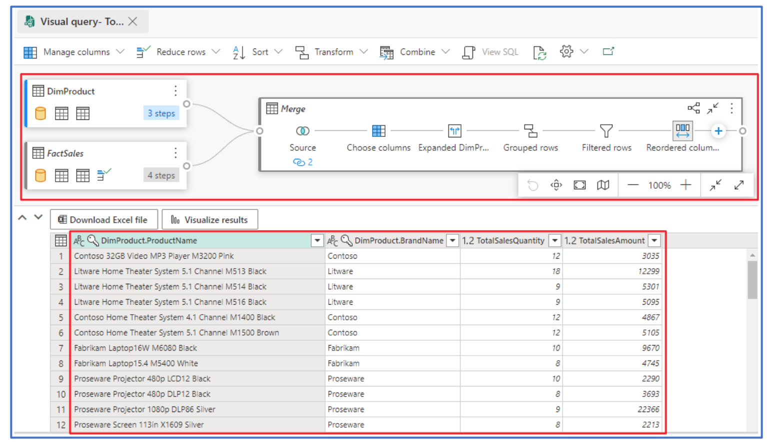
</td>
<td>



</td>
</tr>
<tr>
<td>The visual query editor will open with a preview of the records in the table. Here, the data analyst can choose from an ample number of tools to establish relationships between the tables and match and merge data to derive relevant insights.</td>
<td>

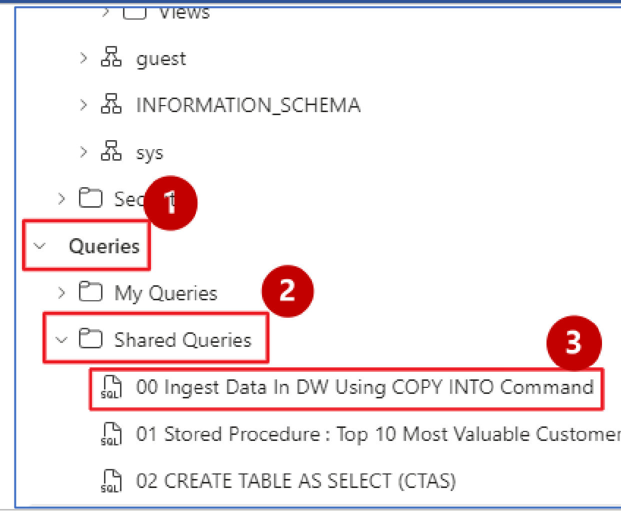
8. **View** the Visual Query Table & Merge activity.
9. **View the result** 'TotalSaleAmount & TotalSalesQuantity by Product.'

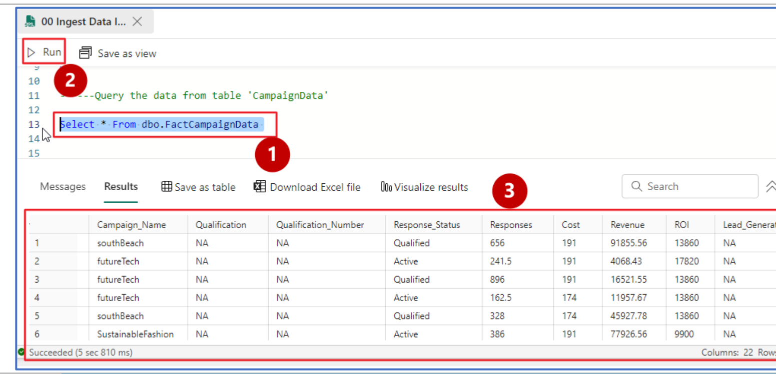
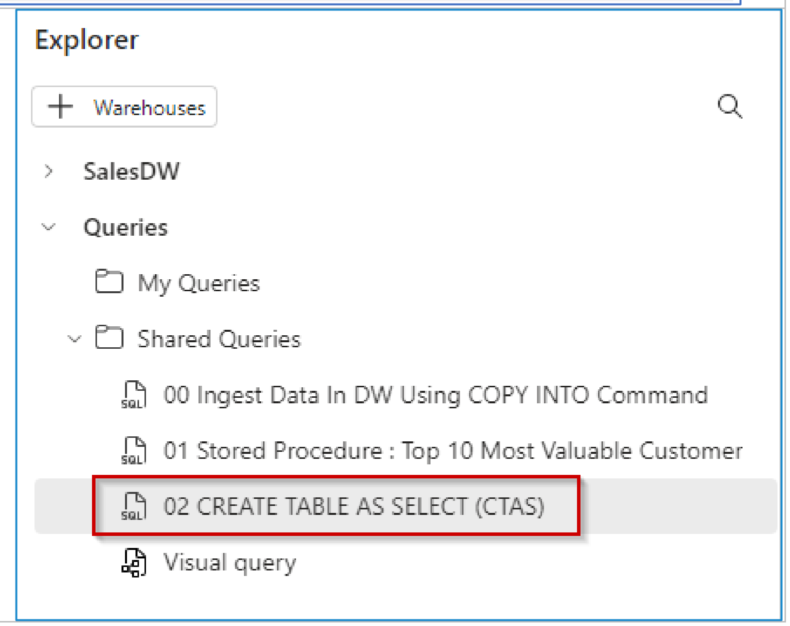
</td>
<td>

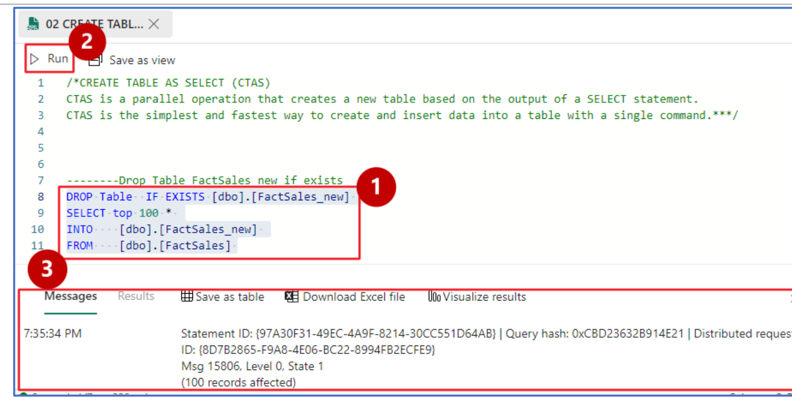


</td>
</tr>
</table>

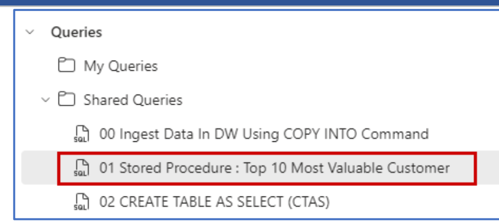
3.5.5 Create a table in Data Warehouse using COPY into syntax

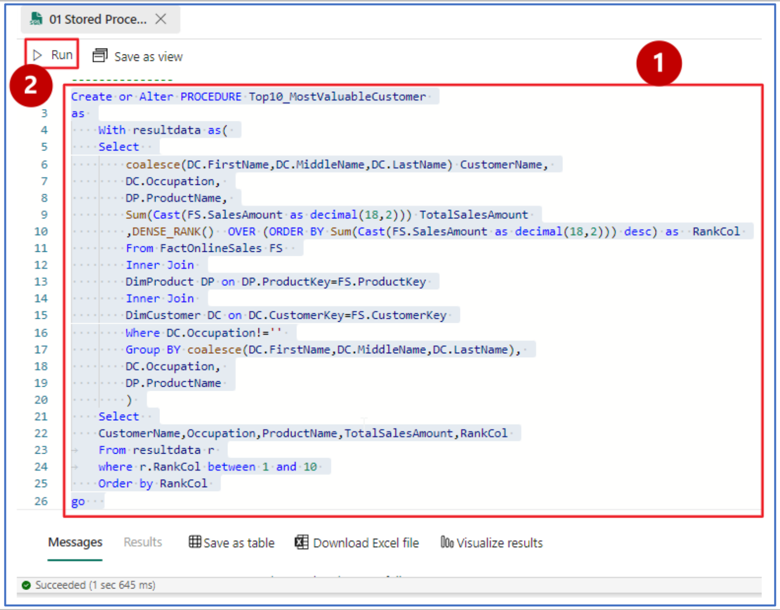
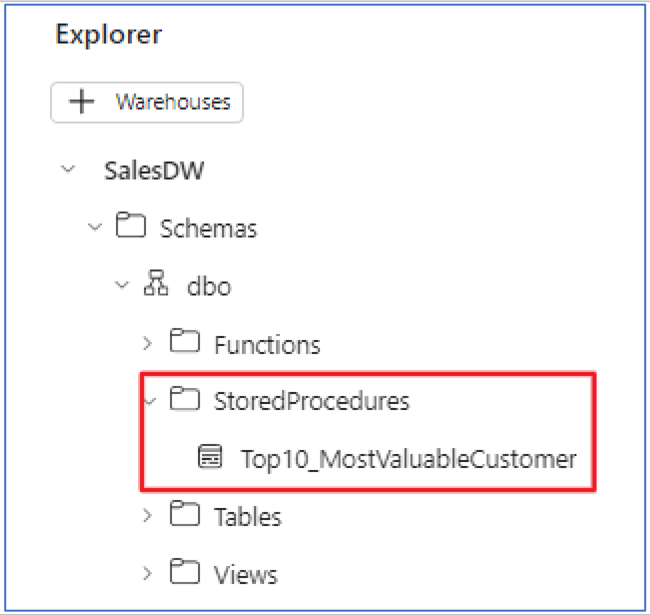
Narrative	Steps	Screenshot
Just like any enterprise warehouse, the Microsoft Fabric warehouse also provides the most commonly used ways for table creation and data ingestion. We have already experienced one of them with the help of data pipelines. Let's take a look at more conventional ways of doing this.	1. In the Explorer panel, **scroll down** to Queries and **expand** Shared Queries. 2. Under Shared Queries folder, **select** "00 Ingest Data In DW using COPY INTO Command".	
Serena, the data analyst, has ingested the data in table quickly using "Copy Into" command. The COPY statement is the primary way to ingest data into Warehouse tables. COPY performs high-throughput data ingestion from an external Azure storage account, with the flexibility to configure source file format options, a location to store rejected rows, skipping header rows, and other options.	3. **Select** the query from line number 2 to 9. 4. **Click** on the run to get data ingested in table.	

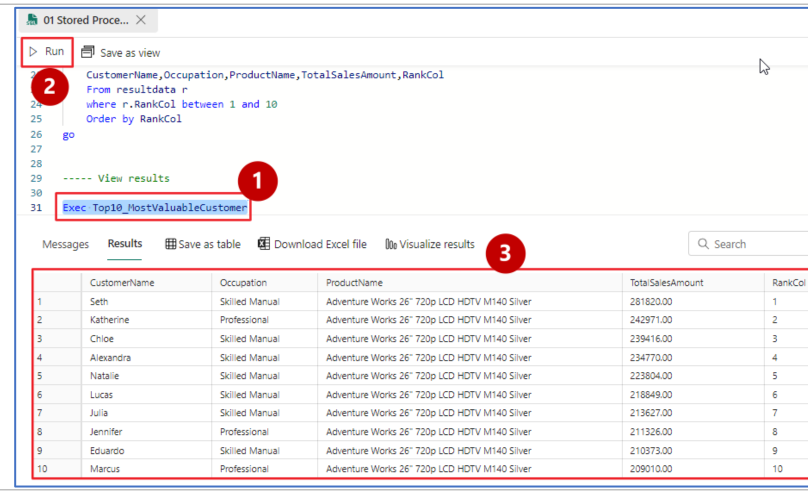
<table>
<tr>
<td>Serena can confirm the operation ran successfully by running a query.</td>
<td>5. Select the query from line number 13.
6. Click on the run.
7. View the table result.</td>
<td>
</td>
</tr>
<tr>
<td>Now, she can create a new table as well based on SELECT statement using CTAS.
The CREATE TABLE AS SELECT (CTAS) statement is one of the most important T-SQL features available. CTAS is a parallel operation that creates a new table based on the output of a SELECT statement.</td>
<td>8. Under Shared Queries folder, select “02 CREATE TABLE AS SELECT (CTAS)”.</td>
<td>
</td>
</tr>
</table>

CTAS is the simplest and fastest way to create and insert data into a table with a single command.	9. **Select** the query from line number 8 to 11. 10. **Click** on Run to execute the selected query. 11. **View** the message ‘table is created & row affected in table’.	

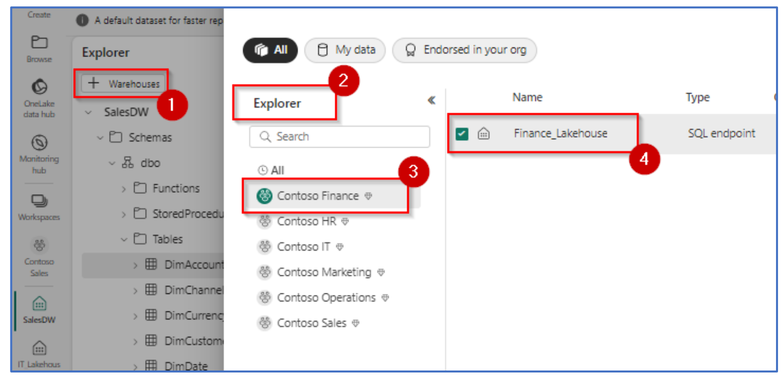
3.5.6 Create a table in Data Warehouse using Stored Procedure

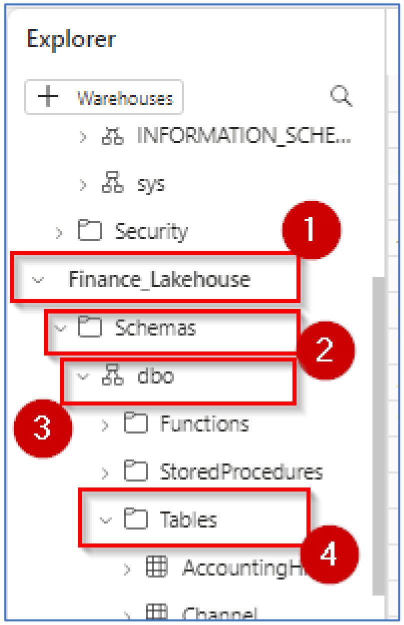
Narrative	Steps	Screenshot
As Data Analyst, Serena can create Stored procedures too.	1. In the Explorer panel, **scroll down** to Queries. 2. Under Shared Queries folder, **select** “01 Stored Procedure: Top 10 Most Valuable Customer”.	

<table>
<tr>
<td>A stored procedure is used to retrieve data, modify data, delete data in database table, perform operations in the database and return a status value to a calling procedure or batch.</td>
<td>3. Select query from line number 3 to 29.
4. Click on Run.</td>
<td>
</td>
</tr>
<tr>
<td>In the Object explorer, verify that she can see the newly created stored procedure by expanding the StoredProcedures node under the dbo schema.</td>
<td>5. In explorer, expand StoredProcedures folder.</td>
<td>
</td>
</tr>
</table>

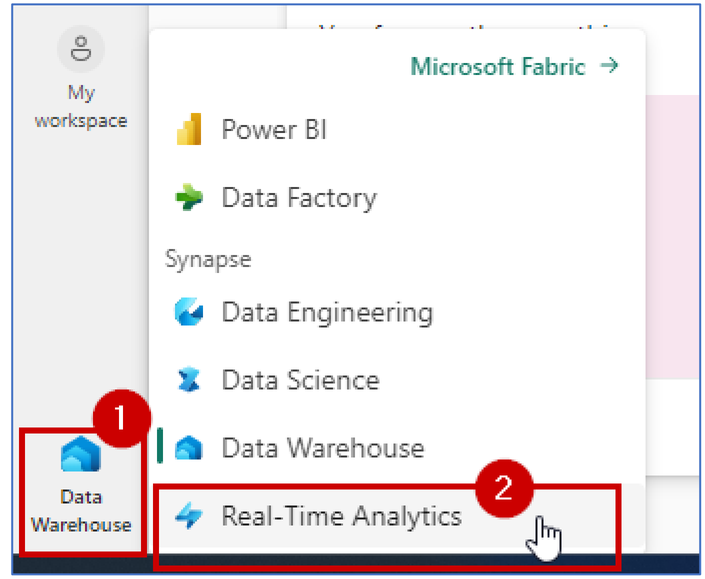
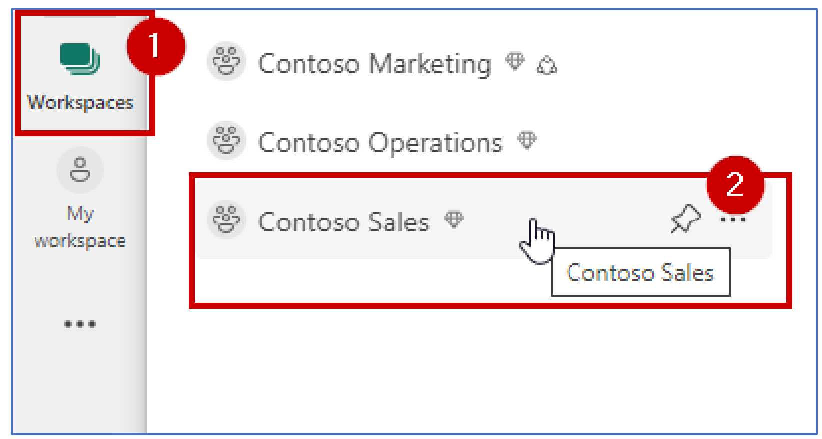
Let’s take a look at the ‘Top 10 Valuable customer’ query results.	6. **View** the result ‘Top 10 Valuable customer’.	

3.5.7 Virtual Warehouse

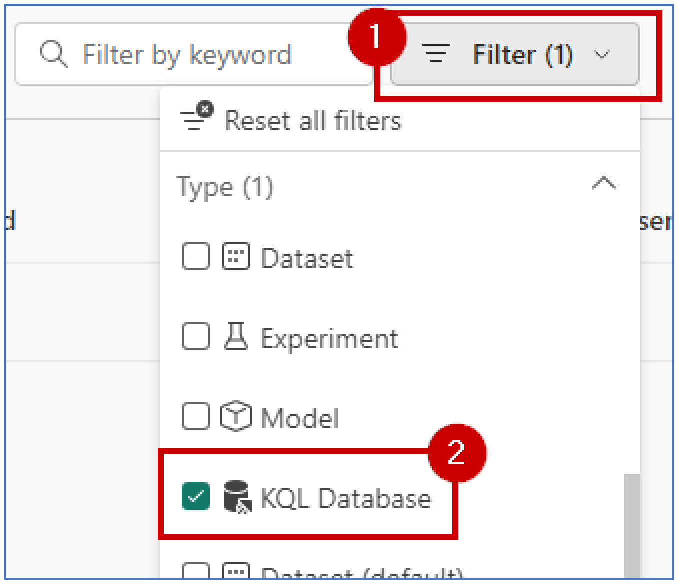
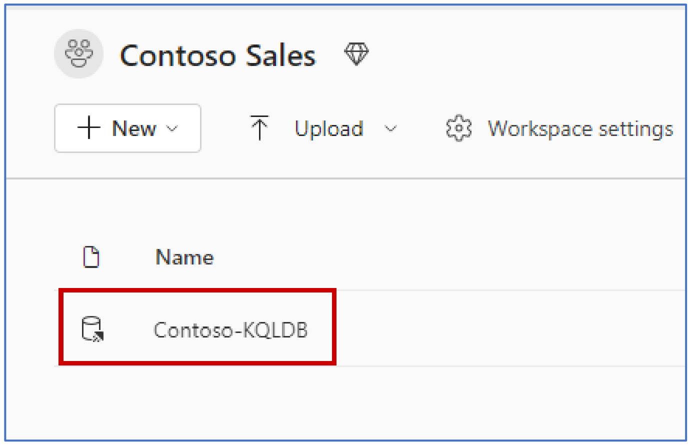
Narrative	Steps	Screenshot
Now that was all about ingestion and table creation. But that was nothing new. We have done all this with conventional warehouses, so what’s special? Introducing virtual warehouses. Serena can not only analyze data from her department, but she can also query any data from other warehouse or a lakehouse SQL end point across the organization from any department.	1. **Click** on the warehouse button. 2. **Click** on explorer. 3. **Select** finance workspace. 4. **Select** Finance_Lakehouse (SQL Endpoint)**.** 5. **Click** on Confirm. **Note:** It will take a few seconds for the new warehouse to appear.	

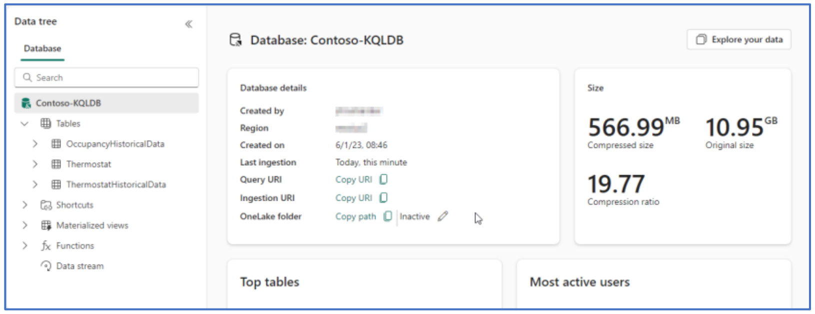
Serena wishes to pull some data from Finance Lakehouse into the sales warehouse. Guess what? Without having to copy data from anywhere she can simply create a virtual warehouse pointing to the Finance workspace. Any of these tables can be dragged into queries and joined together to derive insights and collaborate with different departments.	6. **Show** the Finance lakehouse and tables appearing.	

3.6 Microsoft Fabric for Real-Time Analytics experience

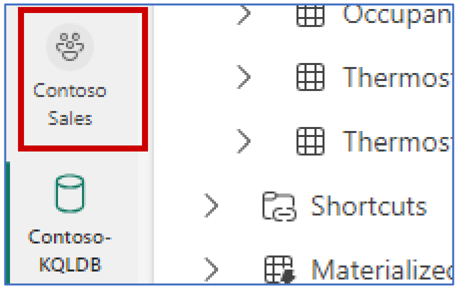
Narrative	Steps	Screenshot
To ensure a pleasant in-store shopping experience for their customers, Contoso set up IoT sensors and thermostats to monitor the store's temperature every minute. As a data analyst, Serena keeps an eye on data coming from the thermostats to ensure they are running smoothly. Serena switches to Real-Time Analytics to work with this data.	From the experience switcher: 1. **Select** Data Warehouse. 2. **Select** Real-Time Analytics.	
Serena opens the in-store real-time analytics, created in the Contoso Sales workspace.	3. **Select** Contoso Sales workspace.	

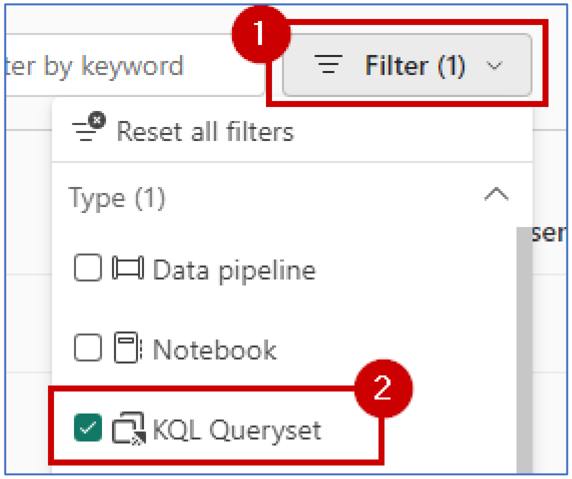
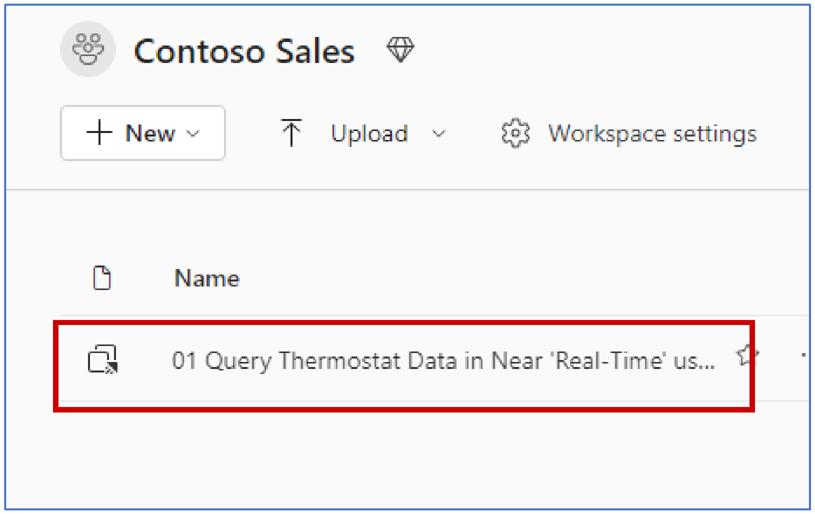
3.6.1 KQL Database

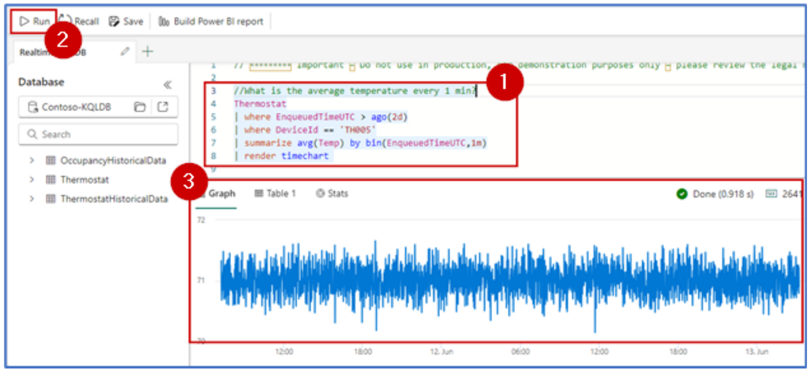
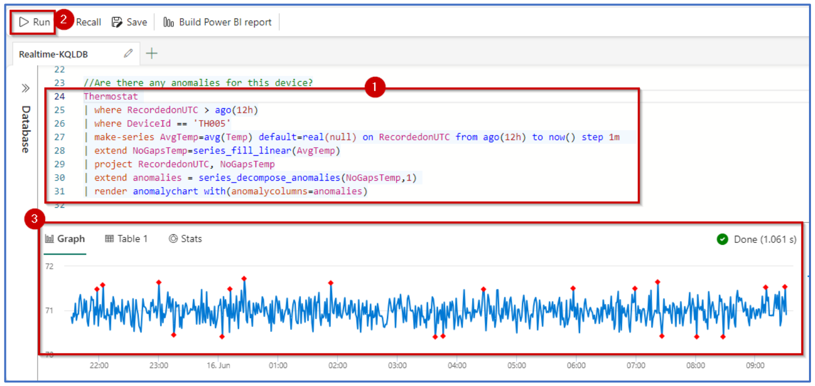
Narrative	Steps	Screenshot
For Serena to interact with real-time data, it must be in the context of databases. So, she looks at the database properties where the real-time analytics data is stored.	1. **Select** Filter. 2. **Select** KQL Database.	
She selects the Contoso_KQLDB database.	3. **Click** Contoso-KQLDB.	

Here, Serena looks at the tables, columns, and other shortcuts where data is added in this database. Once the KQL database has data, she can query the data using Kusto Query Language in a KQL Queryset.	4. **Talk** about the Database properties.	

3.6.2 KQL Query 'For Thermostat Data'

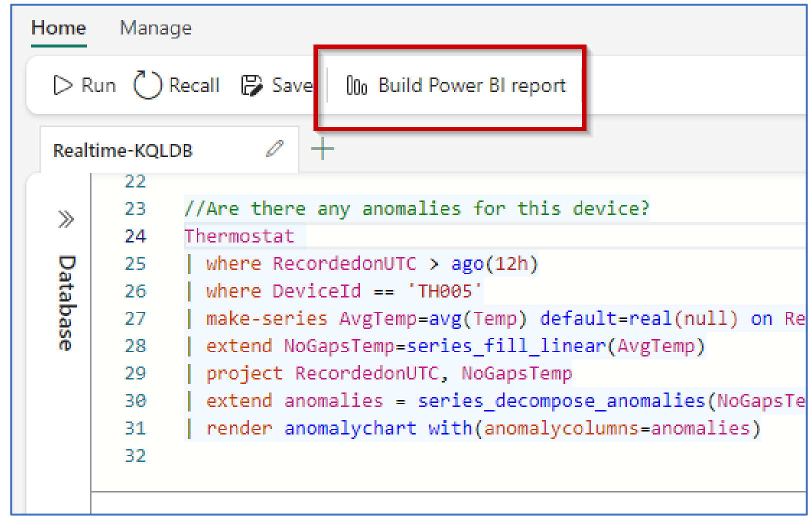
Narrative	Steps	Screenshot
Kusto Query Language is a powerful tool Serena uses to explore Contoso's data, discover patterns, identify anomalies and outliers, create statistical modeling, and more. Serena goes back to the workspace to open the Queryset.	1. **Select** Contoso Sales.	

Here, she sets the filter for KQL Queryset.	2. **Select** Filter. 3. **Select** KQL Queryset.	
Serena can see there's a Query set present to query the thermostat data that's streaming in near real-time from the devices installed in Contoso's stores.	4. **Click** "01 Query Thermostat data in Near Real-Time' using KQL script.	

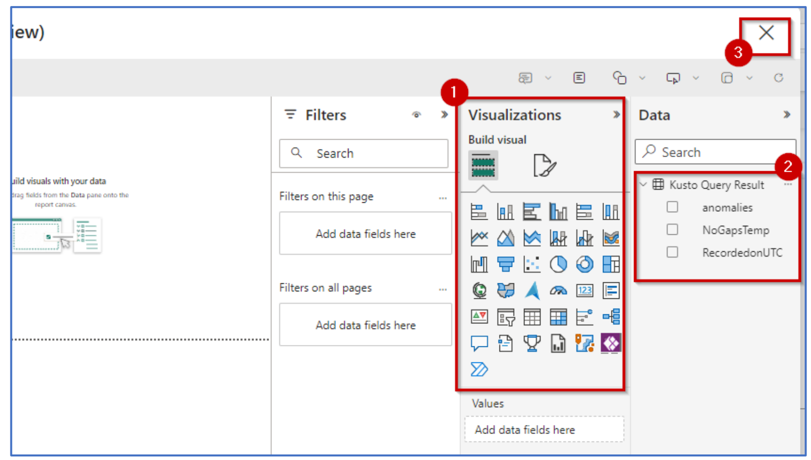
First, Serena wants to know the average temperature per minute. So, she runs the query and visualize the data in a line chart. She can see that it looks like the temperature is currently quite pleasant. Now, she wants to know the average temperature in the next 15 minutes, in anticipation of heavy foot traffic due to the ongoing sale.	5. **Select** the query. (Line 4-8) 6. **Click** Run. 7. **Point** to the graph/result.	
She sees that the temperature is going to remain pleasant for a while.	8. **Select** the query. (Line 12-19) 9. **Click** Run. 10. **Point** to the graph/result.	
She uses a third query to keep an eye on the temperature and detect any anomalies. A sudden rise or drop in temperature triggers an alert for the Contoso staff to check the situation and take necessary action to bring the temperature back to an optimal level. Now that she has real-time data coming to KQL DB, she wants to connect to Power BI to visualize this data.	11. **Select** the query. (Line 24-31) 12. **Click** Run. 13. **Point** to the graph/result.	

3.6.3 Real-Time Power BI report using KQL DB

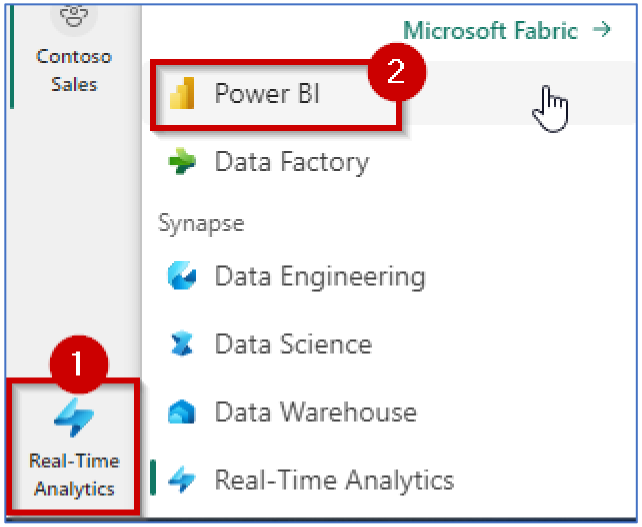
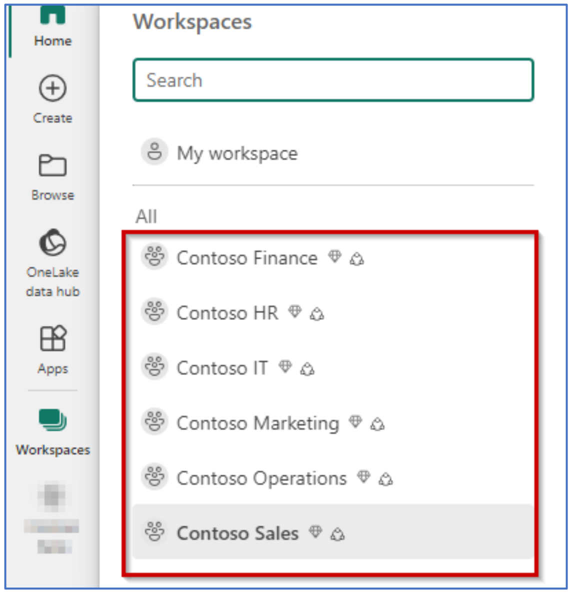
<table>
<tr><th>Narrative</th><th>Steps</th><th>Screenshot</th></tr>
<tr>
<td>Power BI report creation is an integral part of Microsoft Fabric. Power BI reports can be easily created without connecting to any external data sources or Kusto connector.

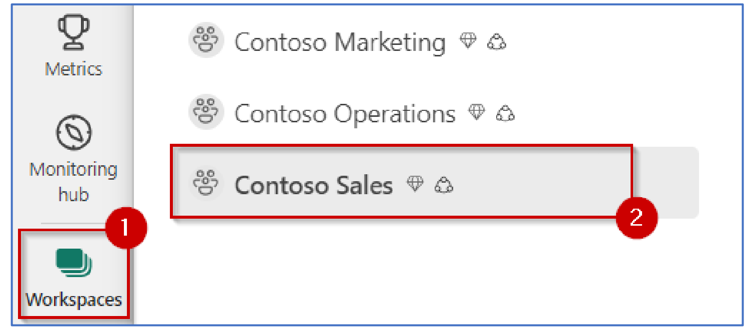
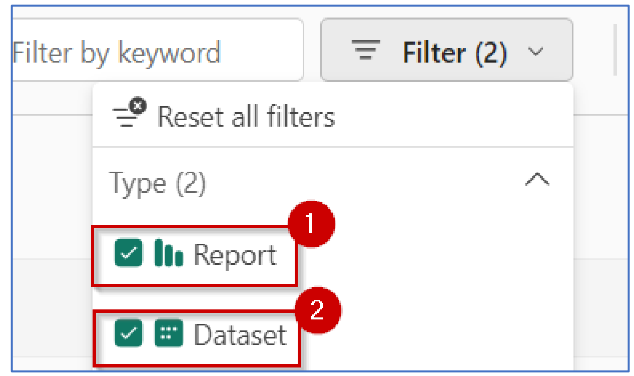
The connector is built into the fabric.</td>
<td>1. Click on “Build Power BI report”.</td>
<td>
</td>
</tr>
<tr>
<td>This action generates a connection to the data set created by Kusto Query running on the Kusto DB.

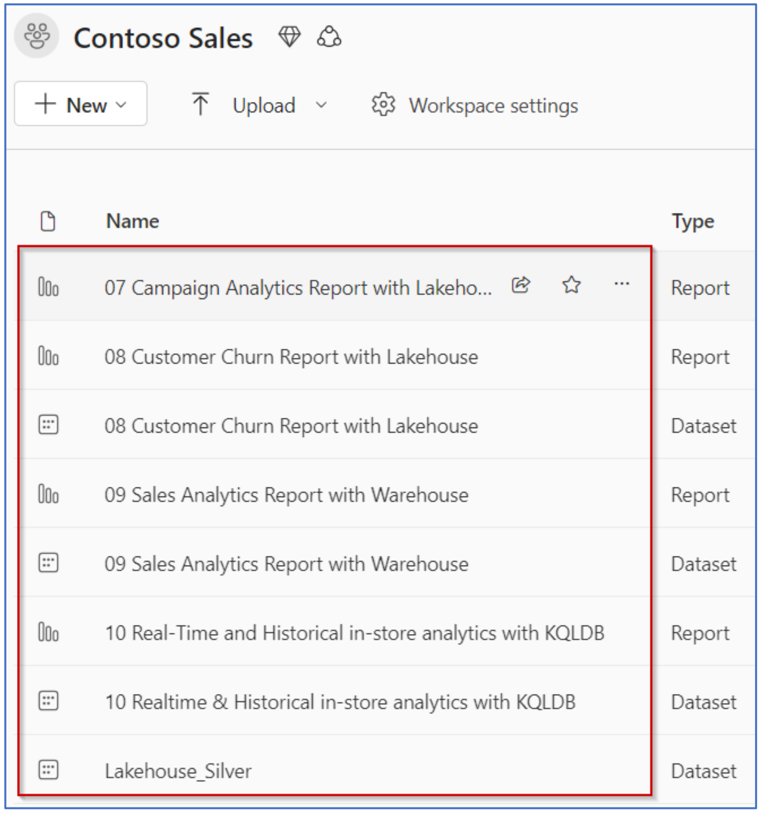
Serena can see that she can create a Power BI report within this plane, with a full fledge feature set available.

On the right side, she can see the real-time dataset in the form of a table.</td>
<td>2. Hover over the Power BI Report layout and navigation panel.
3. Show the table on the right side.
4. Close the Power BI report page.</td>
<td>
</td>
</tr>
</table>

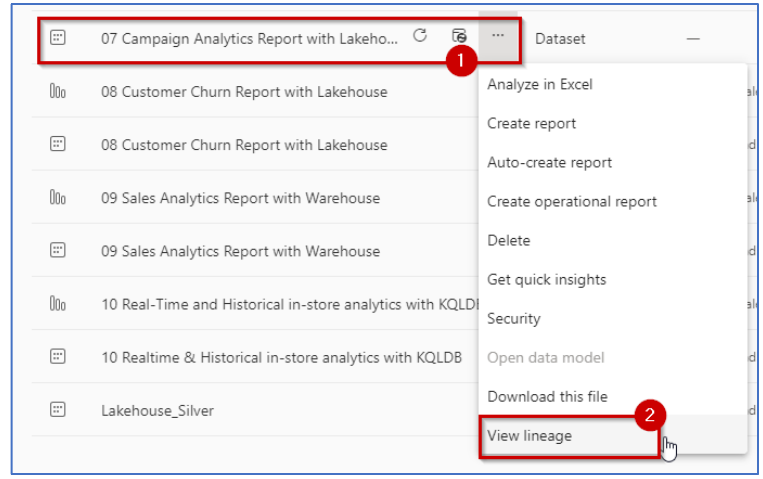
3.7 Microsoft Fabric for Power BI experience

Narrative	Steps	Screenshot
Let's navigate to the Power BI experience in Microsoft Fabric.	In the left navigation bar, 1. **Click** on the experience switcher icon. 2. **Select** Power BI.	
Here we can see all the other workspaces from Finance, HR, IT, Marketing, Operations, and Sales.	3. **Click** on the 'Workspaces' to see all the workspaces. 4. **View** the list of all the departmental workspaces.	

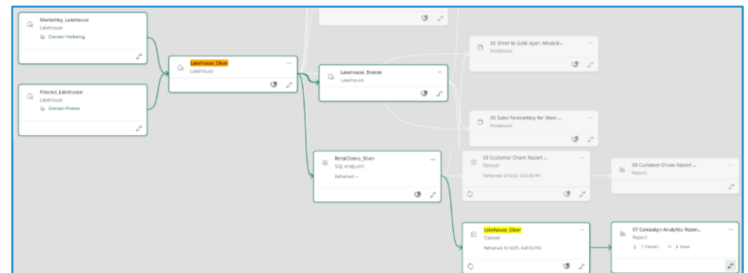
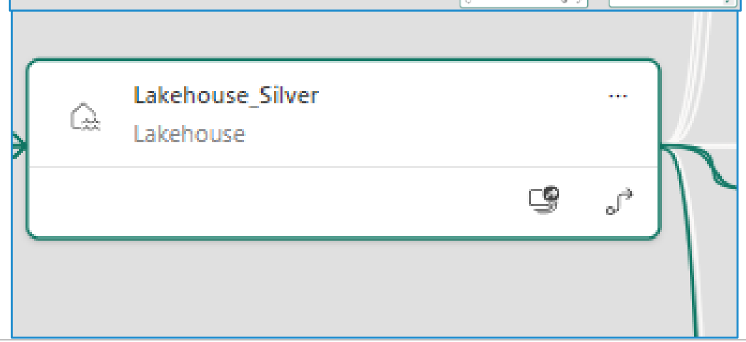
Let's look at Contoso Sales Workspace.	5. **Click** on Contoso Sales workspace.	
To open the datasets and reports, set the filter for these two assets.	6. **Click** on the filter in the top left corner. 7. **Select** Report and Dataset.	

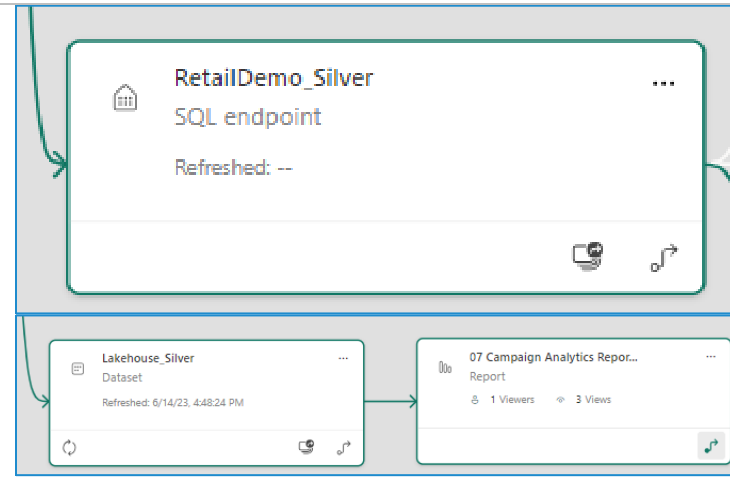
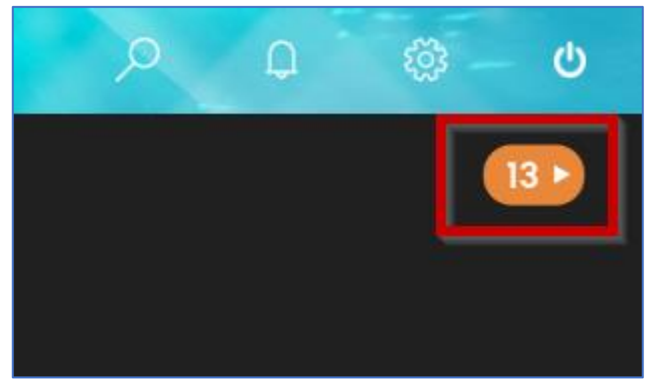
Here is the list of all the datasets and reports available in Contoso Sales workspace.	8. **View** the list of the Reports and their Datasets respectively.	

<table>
<tr>
<td>The Campaign Analytics report is created using the Direct Lake mode. Direct Lake mode is a groundbreaking new dataset capability for analyzing very large data volumes in Power BI. Direct Lake is based on loading parquet-formatted files directly from a data lake without having to query a Lakehouse endpoint, and without having to import or duplicate data into a Power BI dataset.</td>
<td>9. Click on the 3 ellipses for the drop down.

10. Click on the view lineage.</td>
<td>
</td>
</tr>
<tr>
<td>Here, in the lineage view, we see the report Campaign Analytics is coming from the Lakehouse named ‘Lakehouse_Silver’.</td>
<td>11. View the Lineage of the Campaign Analytics Report.

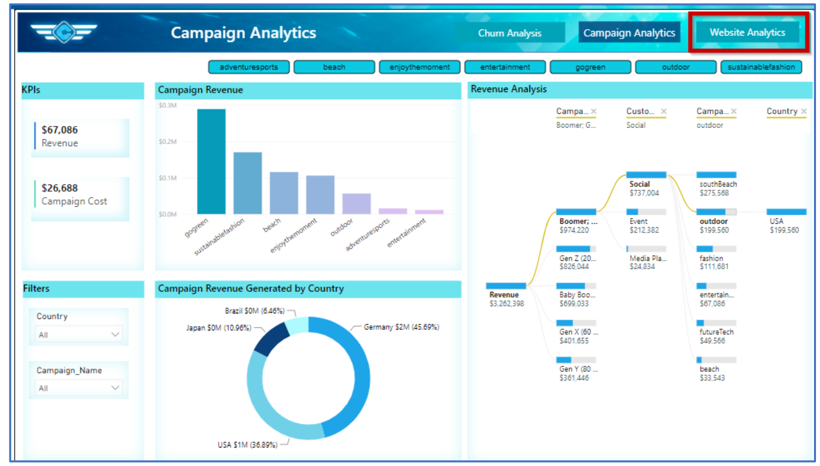
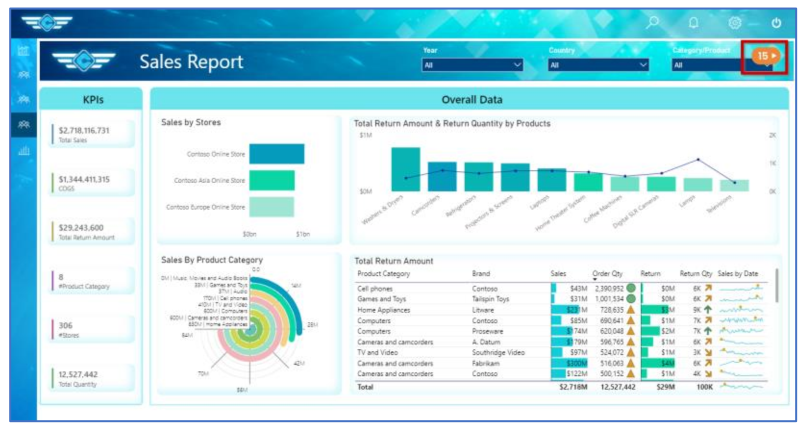
Note:

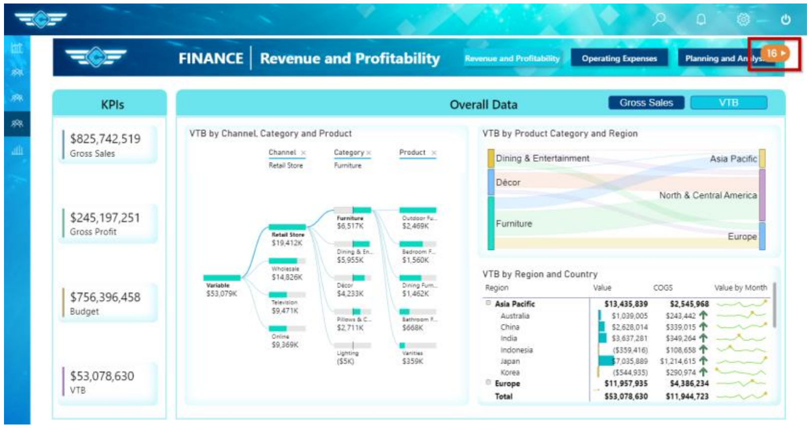
Lineage will appear different for you as reports have not been connected to lakehouse by default. You can make connection updates to the dataset to come from lakehouse.</td>
<td>

</td>
</tr>
</table>

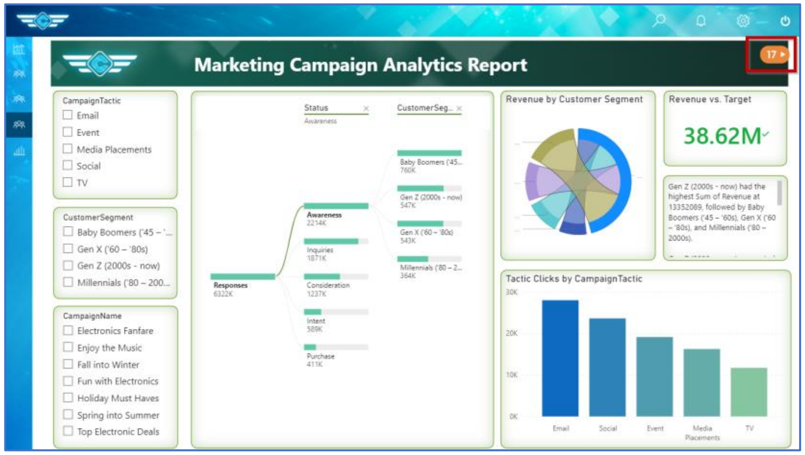
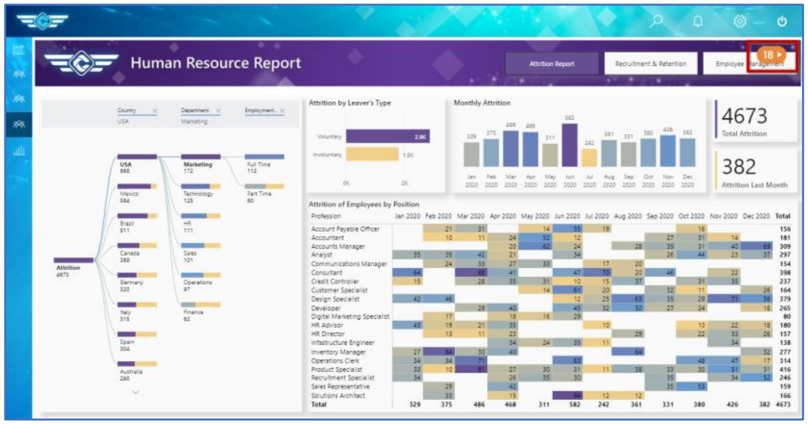
		
Contoso has several other departments, including finance, marketing, operations, and IT. Let’s see the reports that data analysts created using Direct Lake in these individual workspaces.	20. **Click** on arrow 13.	

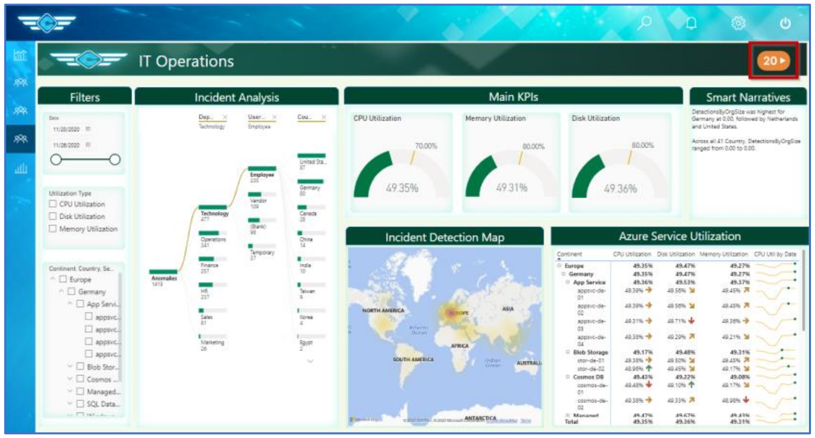
3.7.1 Departmental reports [Direct Lake]

Narrative	Steps	Screenshot
Direct Lake mode eliminates the import requirement by loading the data directly from OneLake. Because there is no explicit import process, any changes can be picked up as they occur at the data source. This combines the advantages of both DirectQuery and import modes while avoiding their disadvantages. Direct Lake mode can be the ideal choice for analyzing very large datasets and datasets with frequent updates at the data source. ***Note:*** *Direct Lake is supported on Power BI Premium P and Microsoft Fabric F SKUs only. It's not supported on Power BI Pro, Premium Per User, or Power BI Embedded A/EM SKUs.* This Report shows the probability of Customer Churn due to various reasons, which can be seen in the Bar chart in the bottom right corner. The scatter plot on the top left shows that Customers with lesser tenures at Contoso and lesser spend amounts are most likely to churn.	**<<The following steps are performed in the demo web app>>** 1. **Talk** about the Customer Churn Report. 2. **Click** on Campaign Analytics**.**	

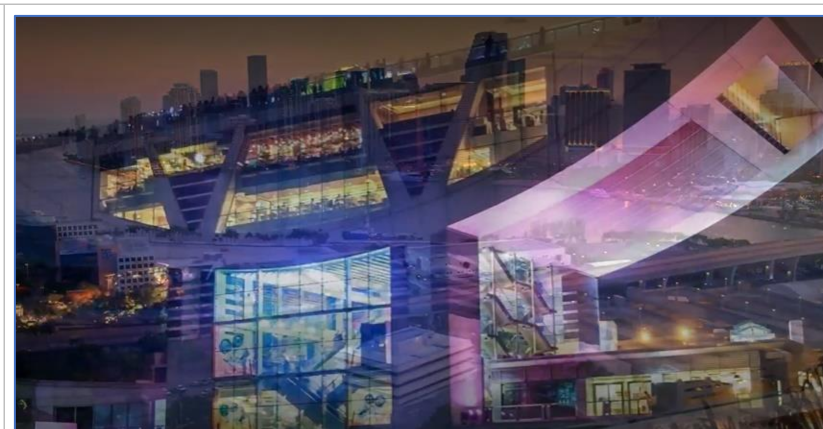
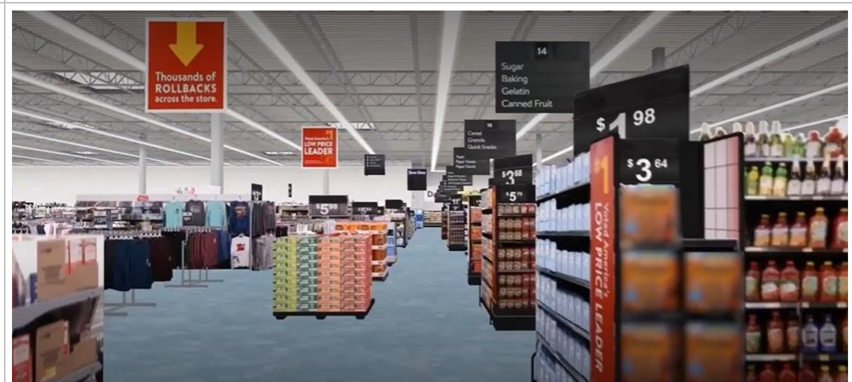
In this report, we can see the campaign cost and the revenue generated through various campaigns. Contoso's Marketing leadership were expecting the #EnjoyTheMoment to be most popular. But look, they find that #go green has generated the highest revenue.	3. **Talk** about the Campaign Analytics Report. 4. **Click** on Website Analytics**.**	
The website analytics report shows that customers in the 31-40 age group, the millennials, are the most unhappy customers because they can't find their favorite products while searching for them on their mobile devices. This is what's causing the high Bounce Rate.	5. **Talk** about the Website Analytics Report. 6. **Click** on arrow 14**.**	
Now let's see the departmental reports. If you remember, in the past the data in departments at Contoso was in silos. This caused their inter-departmental data integration challenges resulting in higher costs, slow time to market and decreased data quality. But now, with Microsoft Fabric, they have one copy of data and with shortcuts they can easily refer to data from	7. **View** the Sales Report. 8. **Click** on arrow 15.	

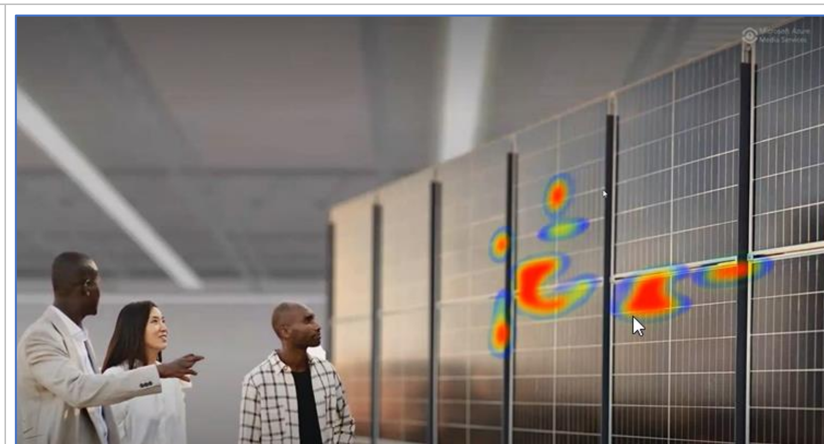
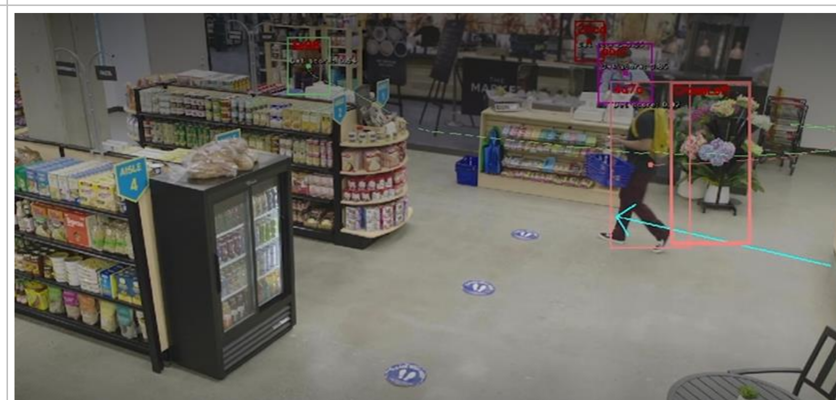
other departments without ever needing to actually move the data, how cool is that? Here we see the Sales report directly from the Lakehouse. The visualization shows us sales details from various stores, departments, product category, and online as well as offline stores. It also shows us the quantity of products which are returned most vs quantity of the products sold. Thanks to the advantages of direct query mode, now available to Contoso, they no longer have to choose between performance and data latency. They get BOTH! In fact, this is true for all the departmental reports that follow.		
Let's look at the finance report which shows Revenue and Profitability, Operating Expenses, Planning, and Analysis. The Revenue and Profitability section shows the gross sales and variance to budget departments wise, channel wise, and product wise. We see the same two benefits here: fast performance and no data latency. What about the Marketing department? Let's see!	6. **View** the Revenue and Profitability Report. 7. **Click** on arrow 16.	

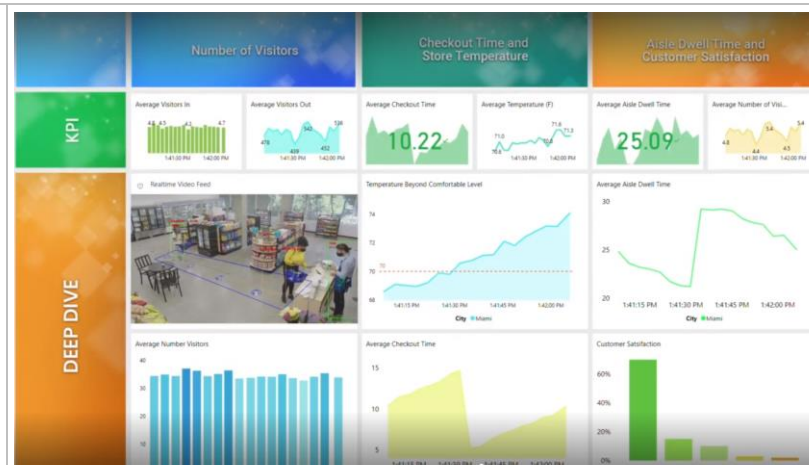
The Marketing Campaign Analytics report shows the responses from customers with different age groups/segments through various campaigns. It also displays the different methods of clicks in the Tactic Clicks by Campaign Tactic chart. We can see that the customer segment GenZ has contributed the most to the revenue, while millennials have contributed the least.	8. **View** the Marketing Campaign Analytics Report. 9. **Click** on arrow 17.	
The Human Resource department wants to make sure that their current and potential employees are happy with the company and benefits. This report has been built on top of the data in the SQL endpoint of the HR workspace. The Human Resources Report here helps them monitor the attrition rate, recruitment, and retention rate – the reasons people are joining or leaving Contoso.	10. **View** the Human Resource Report. 11. **Click** on arrow 18.	

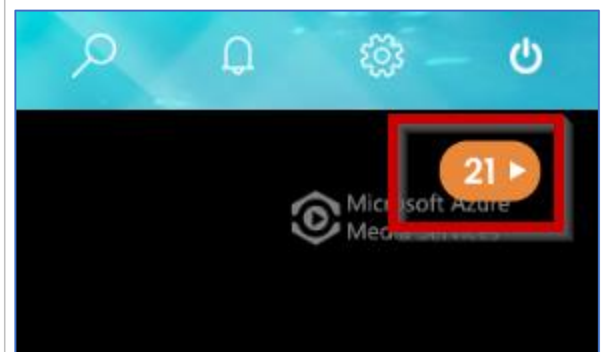
For the Operations department, one of the most critical parts of keeping the organization running is to understand the warehouse expenses, orders vs returns patterns, and the channels that are most popular among their customers. This data helps them strategize their inventory management and warehouse operations.	12. **View** the Operations Report. 13. **Click** on arrow 19.	
And last but not the least, here is the report from the IT department. They are capturing all the ongoing information related to IT needs and consumption by Contoso employees like computation and disk utilization required for their data storage. They are also keeping track of IT incidents, anomalies, etc. being reported in their offices across the world. They obviously need data from pretty much all other Contoso locations. Again, zero data movement and the CDO's integration challenges are now a thing of the past!	14. **View** the Operations Report. 15. **Click** on arrow 20.	

3.8 Finale

So, after all that transformation with Microsoft Fabric at Contoso, let us see the results in action! It's Contoso's Black Friday Sale event and people are entering their stores by the numbers. They've deployed all these IoT sensors into the store. Let's see what happens next.	1. **Play** the store video.	
<<at the heat map scene>> All this data is being captured in real-time. We can see that these two aisles are more popular than other aisles. <<At the wine's aisle scene>> First is the wine aisle. And we can even see the sections which draw the most attention from customers here. Let us see the other aisle.	2. **Talk** about the popular aisles.	

<<At the solar panel’s scene>> And here we see environmentally conscious customers, mostly millennials, who are happy to see items like solar panels in the aisles for them to purchase. <<At the scene where customers are walking in aisles and getting into checkout lines>> Next, we see that as customers are shopping in their stores, the installed cameras and IoT sensors tell the staff in near real-time, the aisle dwell times as well as checkout times. Please note that there is NO facial recognition happening here, this is purely tracking aggregate customer shopping patterns. This way, if the checkout lines get too long, they are notified automatically, and more cashiers can be deployed.	3. **Talk** about the environmentally conscious customers.	
As the sales event proceeds, let’s jump ahead in time to 8:00 AM, the morning of the **Black Friday** sale.		

We can see on this Temperature chart that based on the historical sensor data Contoso collected in the past, on average, their thermostats are keeping temperatures at a comfortable range. As anticipated, when the number of customers increases, temperatures start to increase toward 75 degrees (Fahrenheit) – too hot to be comfortable. So naturally Aisle Dwell Times start to drop, and checkout times start to increase. Remember the anomaly detection in KQL Queryset that we talked about in the Real-Time Analytics Experience? This scenario is exactly what that was created for. Thanks to that anomaly detection, downstream systems are automatically activated when sensor batteries fall below 20% or thermostat sensors fail so they can auto-correct right away, ensuring a comfortable shopping experience for customers. Because of this quick correction, temperatures drop right back to comfortable values, restoring normal aisle dwell times as well as normal checkout times, and overall customer satisfaction remains high. What happens as the day progresses, though? Let's look at these charts in real-time.	4. **Point** to "Temperature Beyond Comfortable Level" chart.	

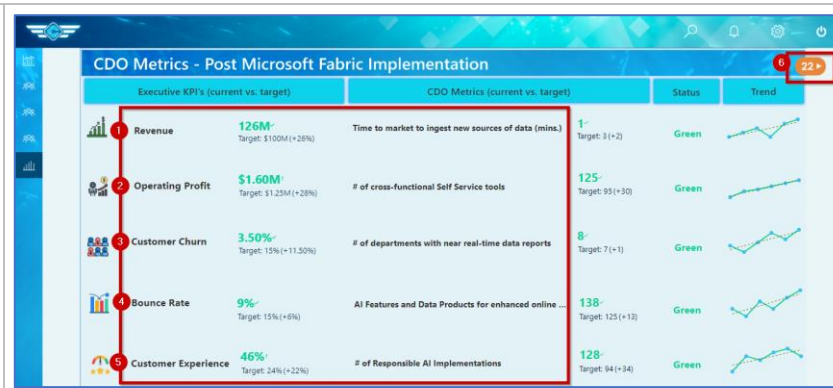
Of course, it's one thing to be able to have insights like this for one store, but with the power of Real-Time Analytics in Microsoft Fabric, Contoso now has insights into the correlations between temperature, aisle dwell time, average checkout times, and customer satisfaction across all their stores on the East Coast. When data belonging to various stores is shown on a single dashboard it becomes very easy to perform comparison analysis and apply the learnings from one store to improve customer satisfaction and aisle dwell time in other stores.	5. **Hover** over the charts.	
Now let's get a glimpse of the latest dashboards.	6. **Click** on arrow 21.	

Thanks to Microsoft Fabric, we can see that all the KPIs that matter most to Rupesh in the CDO Score card have started improving!

1. The LitWare Inc. acquisition is complete, which resolved the time to market issue that was affecting Revenue.
2. Operating Profit is up now that there are more self-service tools which has reduced Operating Expense.
3. Customer Churn is no longer a factor given the access to near real-time data maintaining comfortable temperatures and ensuring an optimal shopping experience.
4. And thanks to the implementation of AI features and data products for an enhanced online experience, the Bounce Rate is way down!

Contoso sees improved interdepartmental collaboration, elimination of data duplication and redundancy, and improvement in overall efficiency.

7. **Talk** about the green CDO metrics.
8. **Click** on arrow 22.



And as a result, the overall executive dashboard for Contoso also shows all green KPIs. April loved this green dashboard. The CDO, Rupesh, and his team knew the real heroes in this demo were Eva, the Data Engineer, Miguel, the Data Scientist, along with Serena and Wendy the Business Analyst. All of them worked so hard on these transformations.	9. **Talk** about the improved KPIs. 10. **Click** on arrow 23.	
Contoso had a great year and they celebrated with fireworks at all their locations! We saw how easy it was to transition to Microsoft Fabric from Contoso's current state and leverage the benefits to achieve their transformation vision! Imagine what Microsoft Fabric can do for your organization!!	**<<Demo End>>**	

DISCLAIMER

This presentation, demonstration, and demonstration model are for informational purposes only and (1) are not subject to SOC 1 and SOC 2 compliance audits, and (2) are not designed, intended or made available as a medical device(s) or as a substitute for professional medical advice, diagnosis, treatment or judgment. **Microsoft makes no warranties, express or implied, in this presentation, demonstration, and demonstration model.** Nothing in this presentation, demonstration, or demonstration model modifies any of the terms and conditions of Microsoft's written and signed agreements. This is not an offer and applicable terms and the information provided are subject to revision and may be changed at any time by Microsoft.

This presentation, demonstration, and demonstration model do not give you or your organization any license to any patents, trademarks, copyrights, or other intellectual property covering the subject matter in this presentation, demonstration, and demonstration model.

The information contained in this presentation, demonstration and demonstration model represents the current view of Microsoft on the issues discussed as of the date of presentation and/or demonstration, for the duration of your access to the demonstration model. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of presentation and/or demonstration and for the duration of your access to the demonstration model.

No Microsoft technology, nor any of its component technologies, including the demonstration model, is intended or made available as a substitute for the professional advice, opinion, or judgment of (1) a certified financial services professional, or (2) a certified medical professional. Partners or customers are responsible for ensuring the regulatory compliance of any solution they build using Microsoft technologies.

FEEDBACK

If you give feedback about the technology features, functionality and/or concepts described in this demo to Microsoft, you give to Microsoft, without charge, the right to use, share and commercialize your feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies, and services to use or interface with any specific parts of a Microsoft software or service that includes the feedback. You may not give feedback that is subject to a license that requires Microsoft to license its software or documentation to third parties because we include your feedback in them. These rights survive this agreement.

MICROSOFT CORPORATION HEREBY DISCLAIMS ALL WARRANTIES AND CONDITIONS WITH REGARD TO THE DEMO, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. MICROSOFT DOES NOT MAKE ANY ASSURANCES OR REPRESENTATIONS WITH REGARD TO THE ACCURACY OF THE RESULTS, OUTPUT THAT DERIVES FROM USE OF DEMO, OR SUITABILITY OF THE INFORMATION CONTAINED IN THE DEMO FOR ANY PURPOSE.