

# Quick-Start Guide - Zetaris NDP Enterprise (on-prem)

## Introduction

Welcome to the Networked Data Platform SaaS Quick-Start Guide!

This Quick-Start Guide provides a fundamental understanding of the Zetaris platform. The Quick-Start Guide helps you to explore the potential business benefits of the platform by saving the manual time involved in performing Extract-Translate-Load (ETL) or Extract-Load-Translate (ELT) processes.

ETL or ELT processes represent one of the five data delivery styles:

1. Bulk or batch data movement
2. Data replication or synchronisation
3. Message-oriented data movement
4. Data virtualisation
5. Stream data integration

The traditional approach to data integration using data movement methods like ETL or ELT is no longer sustainable. This traditional approach cannot meet the business and market demands as data comes in various volumes and frequencies. Structures and platforms used to manage this activity cannot cope with modern business demands.

The Networked Data Platform is an enterprise-wide harmonisation of the five data delivery styles, with a bias towards data virtualisation. Zetaris can ingest data when necessary, but the speed and scalability of the Networked Data Platform derive from virtualisation over metadata, which allows you views of any combination of data from disparate sources without having to move the data itself.

Instead of bringing the data to your query, Zetaris takes the query to the data.

## Contents

- [Introduction](#)
- [How to access the Networked Data Platform](#)
- [The Networked Data Platform UI Overview](#)
- [Data Fabric Builder](#)
  - [Register Physical Datasources](#)
  - [Register Logical Datasources](#)
  - [Register Streaming Datasources](#)
- [Schema Store](#)
- [Data Catalog](#)
- [Query Builder](#)
- [Virtual Data Mart](#)
  - [Create a Virtual Data Mart](#)
- [Virtual Data Warehouse Builder](#)
  - [Create a Virtual Data Warehouse:](#)
  - [How to access the Virtual Data Warehouse](#)
    - [Web Interface](#)
    - [JDBC and ODBC](#)
- [Data Lineage View](#)
- [Data Fabric File System](#)
- [User Management](#)
  - [Create a new user](#)
  - [Reset a user's password](#)

## How to access the Networked Data Platform

To access the Zetaris Networked Data platform, you need to log in to the platform. If you are the first user from your organisation, register as a user with your information.

### To register as the first user from an organisation:

- Visit <https://www.zetaris.com/> and click **Get started**.
- Enter your information in the online form for a 60-day free trial of the Zetaris Networked Data Platform.
- Click **Get Started**. You will receive your access credentials through email.



This registered user can add all other users for their organisation.

Existing users can access the Networked Data Platform through <http://ui.datafabric.zetaris.com/lightning-gui/login%20>.

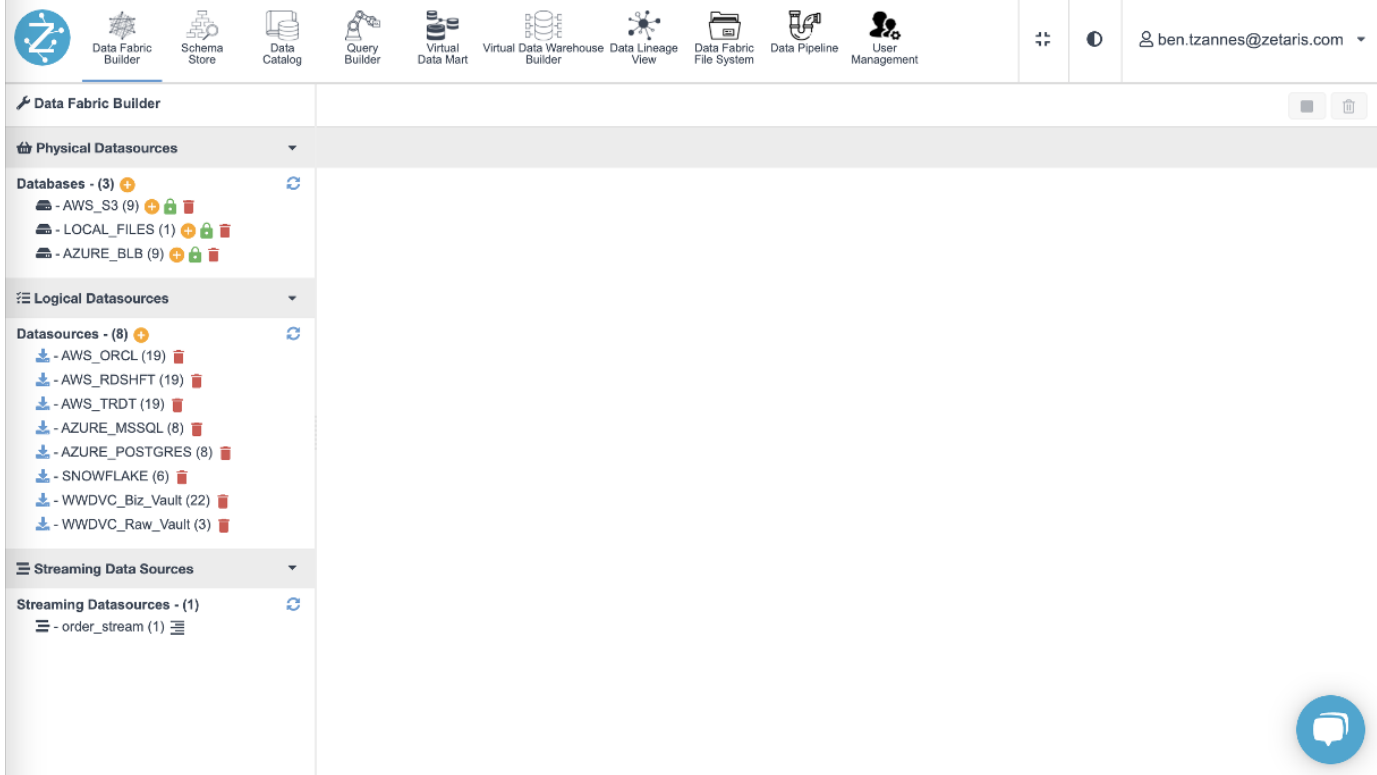
**To log in:**

1. Enter your **user email** and **password**.
2. Select the **terms and conditions** checkbox and click **Log in**.






**Note:** It is mandatory to check the box **"I agree Terms and Conditions"** to log in.





**The Networked Data Platform UI Overview**

On login, you see nine widgets next to the Zetaris logo on the top panel.






These widgets are:

Widget	Name and Action
	<p><b>Data Fabric Builder</b> - enables you to combine data from different sources as if it were physically consolidated through virtualisation over metadata.</p>
	<p><b>Schema Store View</b> - enables logical configuration of entities across disparate data sources so that you can query them as if they were physically consolidated.</p>
	<p><b>Data Catalog</b> - enables you to curate your data assets for better collaboration, stronger regulatory compliance, and reduced overhead.</p>
	<p><b>Query Builder</b> - enables you to query your disparate data sources in any combination through a simple drag and drop process.</p>
	<p><b>Virtual Data Mart</b> - enables you to combine disparate data sources in any combination through a simple drag and drop process.</p>


	<b>Virtual Data Warehouse Builder</b> - creates small, medium, or large data warehouses flexibly.
	<b>Data Lineage View</b> - shows the full context of your data sources, how you have been building and aggregating your datasets and what transformations have occurred.
	<b>Data Fabric File System</b> - enables you to create views of local files (e.g. CSV).
	<b>User Management</b> - manages user roles and permissions.

The user panel (top right) includes the following tools:

Buttons	Button Name
	<b>Expand Navigation</b> - view the nine widget icons with their names
	<b>Colour Adjustment</b> - view the interface in either dark or light mode.
 admin@zetaris.com ▾	<b>User Profile</b> - change your password, sign out securely, and discover information about the Zetaris Lightning Interface. <ul style="list-style-type: none"> <li>• Click <b>About</b> to get information about Zetaris Lightning.</li> <li>• Click <b>Change Password</b> to reset your password. <ul style="list-style-type: none"> <li>To change the password: <ul style="list-style-type: none"> <li>• Enter the old password in the <b>Change Password</b> dialog box.</li> <li>• Enter the <b>new password</b>.</li> <li>• Confirm the new password, and click <b>Update</b>.</li> </ul> </li> </ul> </li> <li>• <b>Sign Out</b> to end the session and log out securely.</li> </ul>

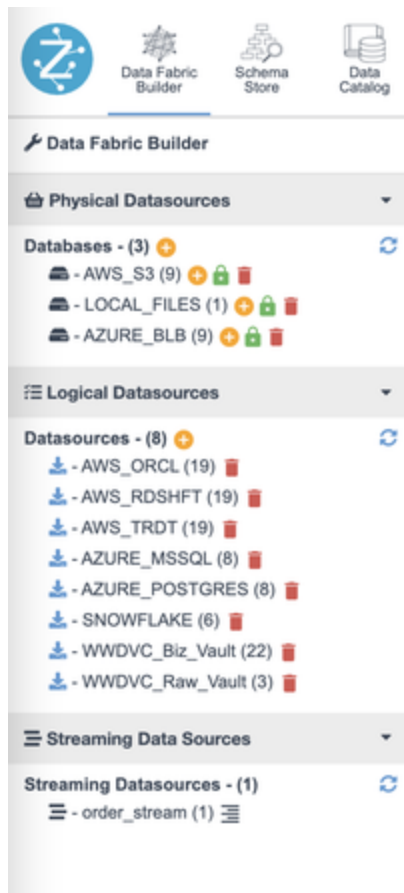
## Data Fabric Builder

Data Fabric Builder allows you to create Data Fabrics from data sources consisting of different databases, streams, or files. These data sources are registered in the left panel.

 The physical data sources and their data remain where they are. Registration allows Zetaris to locate and interrogate them and ingests only the metadata from these data sources.

There are three different types of registered data sources with Zetaris:

- **Physical Datasources** denote files such as CSV, XML, JSON, Parquet, ORC, and Delta. Examples of these sources include AWS S3, Azure Blob, local/network file drives, and RESTful API data sources.
- **Logical Datasources** denote databases (RDBMS and NoSQL).
- **Streaming Datasources** include Apache Kafka and AWS Kinesis.



## Register Physical Datasources


To register physical data sources into the Zetaris Networked Platform:

1. Create and name a virtual connection to your physical data source, local Files, or AWS S3, and so on, by clicking the plus button *plus*:
2. Once the virtual connection is added, click the plus button *plus*: next to the registered data source.

3. Select **File Store** to register data from known locations, or select **API** to retrieve data from a RESTful API. Click **Next**.

4. Complete fields to register data source tables.

The physical data sources should now be available.

 You can register these tables in bulk through the Schema Store View.

 For details on specific connections and the variables that need to be completed, see our Data Source Connectors list.

## Register Logical Datasources

To register Logical Datasources in the Zetaris Networked Platform:

1. Click the plus button *plus*: under Logical Datasources to initiate the connection wizard.

2. Select **Pull Type** and click **Next**.

3. Complete fields to register logical data source tables.

4. Click **Next**. A summary of all the virtual database and driver parameters with the selected virtual tables is displayed. To accept these changes and create the virtual data source, click **Register**.

Logical data source tables should now be available.

You can register these tables in bulk through the Schema Store View.

For details on specific connections and the variables that need to be completed, check out our Data Source Connectors list.

## Register Streaming Datasources

Streaming data sources can be set up through the **Schema Store**.

### Schema Store

Schema Store is the Zetaris SQL editor, which allows you to:

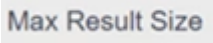



- Register metadata from physical, logical, and streaming data sources in bulk
- Manage metadata schema platform-wide
- Run ad-hoc queries across virtual data sources
- Materialise and cache virtual data sources
- Run virtual data sources statistics
- Specify virtual data sources partitions
- Create permanent views and data marts of our virtual data sources
- Configure security, LDAP integration, user management, and role-based access control

Result	History
04/06/2021, 12:31:59	DESCRIBE USER ADMIN@ZETARIS.COM
03/06/2021, 15:34:53	SELECT * FROM VIEW REGIONSALES
03/06/2021, 15:34:46	SELECT METADATA FROM VIEW REGIONSALES
03/06/2021, 12:35:33	SELECT * FROM AZURE_BLB.region
03/06/2021, 12:35:27	SELECT * FROM AZURE_BLB.supplier

In the left panel, you can view the **Schema Browser**. Here you can view all the virtual databases, data sources, streams, data marts, and views for use in any query.

The **SQL Editor** toolbar is near the top of your screen. It allows you to perform the following functions:

Buttons	Name and Action
	<b>Plus</b> – it adds a tab.
	<b>File Open</b> – opens any previously saved SQL queries.
	<b>Save</b> – allows save or save as of any new or edited SQL queries.

 <b>Max Result Size</b> 50	<b>Max Result Size</b> - sets the maximum number of results to be returned by a query.
	<b>Play</b> - executes any SQL code that has been highlighted in the SQL Editor.
	<b>Stop</b> - stops SQL query execution.
	<b>Format</b> - displays SQL Code in the Editor window into a human-readable format.

The log panel below the Schema Browser canvas has two tabs: Result and History.

- The **Result** tab allows you to view the output of your query.
- The **History** tab allows you to view the history of the queries you have run.

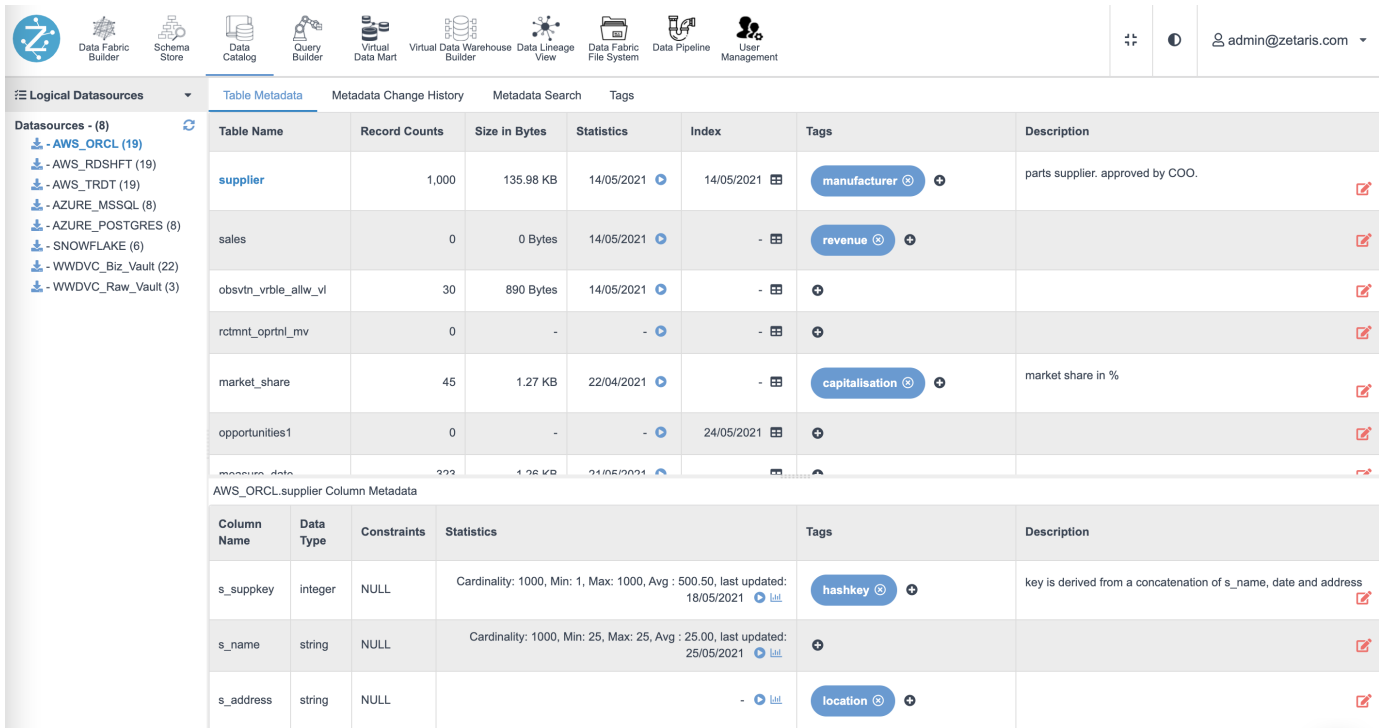
## Data Catalog

The Zetaris Data Catalog captures information on table and column metadata and tracks metadata changes.

This metadata initially only includes tables and columns; however, tags and descriptions can be added manually or uploaded.

Statistics can be run against tables and columns to retrieve quantitative metrics.

Table	<ul style="list-style-type: none"> <li>• Record Counts</li> <li>• Size in Bytes</li> <li>• Data statistics were last run</li> </ul>
Column	<ul style="list-style-type: none"> <li>• Cardinality (uniqueness of data values)</li> <li>• Min and Max (string length, or integer / float values)</li> <li>• Average (string length, or integer / float values)</li> <li>• Date statistics were last run</li> </ul>



The screenshot shows the Zetaris Data Catalog interface. At the top, there are navigation icons for Data Fabric Builder, Schema Store, Data Catalog, Query Builder, Virtual Data Mart, Virtual Data Warehouse, Data Lineage View, Data Fabric File System, Data Pipeline, and User Management. The user is logged in as admin@zetaris.com.

The main view is titled "Table Metadata" and shows a list of tables with columns for Table Name, Record Counts, Size in Bytes, Statistics, Index, Tags, and Description. The tables listed are:

Table Name	Record Counts	Size in Bytes	Statistics	Index	Tags	Description
supplier	1,000	135.98 KB	14/05/2021	14/05/2021	manufacturer	parts supplier. approved by COO.
sales	0	0 Bytes	14/05/2021	-	revenue	
obsvtn_vrble_allw_vl	30	890 Bytes	14/05/2021	-		
rctmnt_oprtnl_mv	0	-	-	-		
market_share	45	1.27 KB	22/04/2021	-	capitalisation	market share in %
opportunities1	0	-	-	24/05/2021		
monocure_data	222	1.28 MB	21/05/2021			

Below the table list, there is a section for "AWS\_ORCL.supplier Column Metadata" with columns for Column Name, Data Type, Constraints, Statistics, Tags, and Description:

Column Name	Data Type	Constraints	Statistics	Tags	Description
s_supkey	integer	NULL	Cardinality: 1000, Min: 1, Max: 1000, Avg: 500.50, last updated: 18/05/2021	hashkey	key is derived from a concatenation of s_name, date and address
s_name	string	NULL	Cardinality: 1000, Min: 25, Max: 25, Avg: 25.00, last updated: 25/05/2021		
s_address	string	NULL		location	

The **Metadata Change History** picks up changes in tables and columns, with an associated timestamp.

Logical Datasources

- Datasources - (8)
  - AWS\_ORCL (19)
  - AWS\_RDSHFT (19)
  - AWS\_TRDT (19)
  - AZURE\_MSSQL (8)
  - AZURE\_POSTGRES (8)
  - SNOWFLAKE (6)
  - WWDVC\_Biz\_Vault (22)
  - WWDVC\_Raw\_Vault (3)

Table Metadata | **Metadata Change History** | Metadata Search | Tags

Update Cycle

MONTHLY  WEEKLY  DAILY  HOURLY

Change History

Change On	Entity	Operation	Description
2021-05-18 19:09:07.181	DATASOURCE	INITIALIZE	SNOWFLAKE is created
2021-05-18 19:09:18.192	DATASOURCE	INITIALIZE	4 tables ingested
2021-05-20 13:26:29.475	TABLE	ADD	schema : PUBLIC, table : ORDERS was added
2021-05-21 09:58:55.953	TABLE	ADD	schema : PUBLIC, table : PART was added

**Metadata Search** allows you to search for tags, descriptions, column names, table names, and values.

Table Metadata | Metadata Change History | **Metadata Search** | Tags

customer

Metadata 60 found

Showing 1-20 of 60 items

Rows per page:

Data Source	Table	Column	Tags	Description
AZURE_MSSQL	customer	na	customer	this is customer
AZURE_MSSQL	customer	c_custkey	customer,key	
AZURE_MSSQL	customer	c_name	customer,name	
SLRT	customer	na		
SLRT	customer	c_custkey		
SLRT	customer	c_name		
SLRT	customer	c_address		

First < **1** 2 3 > Last

Contents 1,501 found

Showing 1-20 of 1,501 items

Rows per page:

Data Source	Table	Column	Value
AZURE_MSSQL	customer	c_name	Customer#000000001
AZURE_MSSQL	customer	c_name	Customer#000000002
AZURE_MSSQL	customer	c_name	Customer#000000003

**Tags** are a fast way to access the tag repository and view all associated columns and tables.

Total 27 tags found

Total 12 entities found

Data Source	Entity	Name	Tags	Description
AWS_ORCL	table	customer	customer, customer_oracle, customer_in_orcl	
AWS_TRDT	table	customer	customer	
AZURE_MSSQL	table	customer	customer	this is customer
AZURE_MSSQL	column	customer.c_custkey	customer, key	
AZURE_MSSQL	column	customer.c_name	customer, name	
AWS_TRDT	column	customer.c_custkey	customer, customer_key	
AWS_TRDT	column	customer.c_name	customer, customer_name	

## Query Builder

You can use this widget to **build a query** and save **permanent views** interactively through a drag and drop method, auto-generating the SQL statements.

## Query Builder Overview



- **Schema Browser** - This panel allows you to use the Query Builder for all the available virtual data sources, data marts, streams, pipeline tables, and views.
- **Main Panel** - In this panel, you can view or drag virtual tables to create queries and save permanent views.
- **Properties** - In this panel, you can group operators or set column properties.

Name	Purpose
GROUP BY	The columns with the GROUP BY clause are used in this section of the generated SQL query. Click the ROLLUP or CUBE radio buttons to apply those grouping operators to the GROUP BY.
CUBE	The columns with the GROUP BY clause are used in this section of the generated SQL query in a CUBE sub-clause. Drag a column to another column to change its order in the ROLLUP. The number next to the column is the rank order.
ROLLUP	The columns with the GROUP BY clause are used in this section of the generated SQL query in a ROLLUP sub-clause. Drag a column to another column to change its order in the ROLLUP. The number next to the column is the rank order.

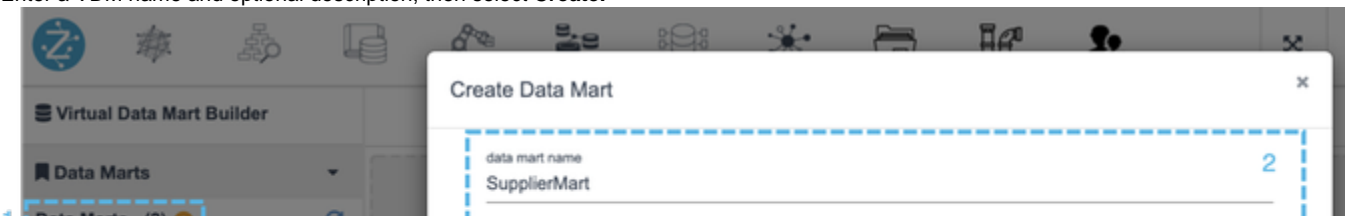
- **SQL** - Returns the mirrored SQL generated from the Query Builder actions.
- **Result** - Output from running the Query Builder.

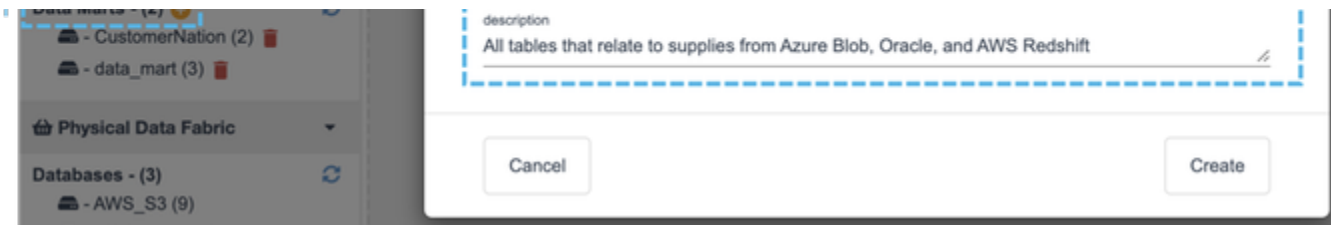
### Virtual Data Mart

Virtual Data Marts (VDMs) are logical collections of your virtual physical and logical sources that can be grouped and related together. These VDMs are assigned to a Virtual Data Warehouse (VDW), making it easy for the admin to assign end users access to data sources from a single, logical location. VDMs can be reused across VDWs to speed up the data allocation and management processes.

### Create a Virtual Data Mart

1. Click the plus button :plus: or :create: to add a new **Virtual Data Mart**.
2. Enter a VDM name and optional description, then select **Create**.





3. Click and drag tables from your virtual physical and logical data sources to the main grid.
4. Create relationships and rename columns and tables (optional).
5. Select **Save changes** in the top right.

## Virtual Data Warehouse Builder

A Virtual Data Warehouse (VDW) is a group of physically separate data sources analysed as a single, consolidated data source.

Virtual Data Warehouse Name	Cluster Type	Size	Status	Jobs Running	Created At	Actions
vorderdw	Virtualised	Medium	AVAILABLE	0	02/06/2021, 17:00:21	[Stop] [Resume] [Remove]
FinanceVDW	Virtualised	Large	AVAILABLE	0	31/05/2021, 20:57:36	[Stop] [Resume] [Remove]
DetailsVDW	Virtualised	Medium	AUTO_SHUTDOWN	0	18/05/2021, 22:39:31	[Stop] [Resume] [Remove]
streams	Streaming	Large	STOPPED	0	01/06/2021, 20:36:10	[Stop] [Resume] [Remove]

This builder provides the flexibility to create small, medium, and large warehouses. Each warehouse is assigned different amounts of CPU and memory to enable the execution of queries. The type of warehouse used depends on the data to be processed and the execution response time required.

### Create a Virtual Data Warehouse:

1. Click **create** on the top-left side of the interface.
2. A pop-up window appears on your screen asking for the following details:
  - Enter the **Virtual Data Warehouse name**.
  - Choose the **type** of Virtual Data Warehouse from the drop-down menu.
  - Select the **size** of the Virtual Data Warehouse. The size will determine the number of resources allocated to that VDW. The sizes are:

Size	Resources (Resources per executor: 6 cores, 1.5 GB memory)
Small	Executors: 2 Cores: 12 Memory: 3 GB

Medium	Executors: 4 Cores: 24 Memory: 6 GB
Large	Executors: 6 Cores: 36 Memory: 9 GB

3. Select the appropriate **Users and Roles**. Click **Next**.
4. Select the appropriate **Lightning Databases** and/or **Data Sources**. Click **Next**.
5. Select the appropriate **Data Marts** and/or **Permanent Views**. Click **Next**.
6. A review of the Virtual Data Warehouse configuration appears on the screen. After reviewing all the information, click **Create**.

You have successfully created a new Virtual Data Warehouse. After the warehouse is created, initially it is in the pending state for a very brief time, but later changes to a running state.

To change **User access** or **Role access**, click :Assign: or :Revoke: buttons placed in the database panel.

### How to access the Virtual Data Warehouse

There are two ways to access the Virtual Data Warehouse:

#### Web Interface :VDW:

To access through Web Interface, click the Virtual Data Warehouse login icon on the right-hand side of the screen :VDW: . A new screen appears asking for your login credentials.

 You can log in only if you are provided with access to that warehouse.

The Zetaris VDW interface is a lightweight version of the platform, with access to **Schema Store** and **Query Builder**. End-users can build their own **ad-hoc queries**, and build their own **permanent views**.

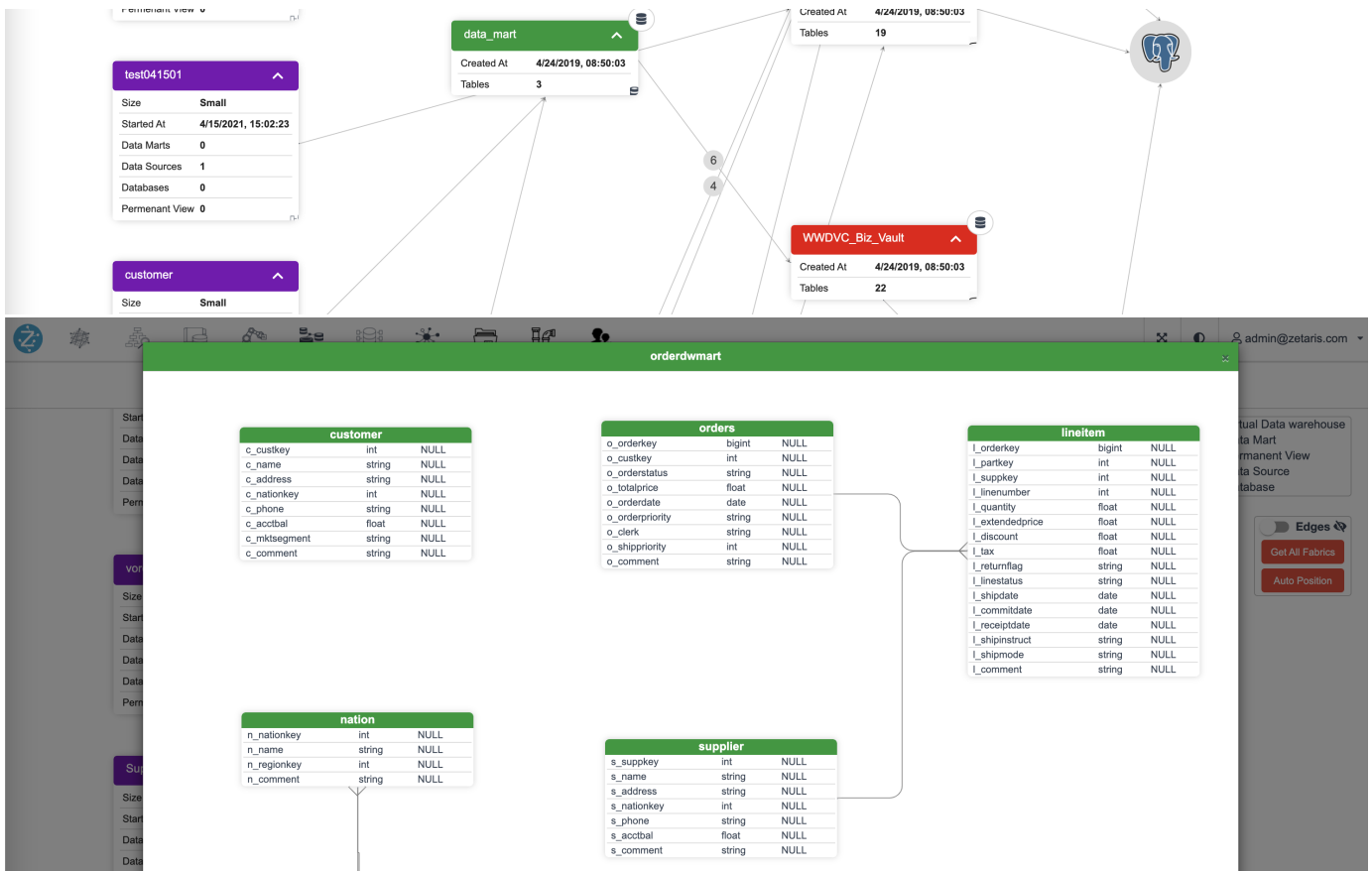
#### JDBC and ODBC :JDBC/ODBC:

To access Zetaris through third-party tools such as Tableau, PowerBI, Python and R, for example, you can use our JDBC or ODBC access points.

- For **JDBC**, use the JDBC string that can be copied from the URL icon on the right hand side of the page, in line with the VDW that you are trying to access :JDBC/ODBC:. You will also need the Zetaris JDBC driver. To obtain this, contact your Zetaris account manager.
- For **ODBC**, use the IP address from the JDBC string that can be copied from the URL icon the right hand side of the page, in line with the VDW that you are trying to access :JDBC/ODBC:. You will also need the Zetaris ODBC driver. To obtain this, contact your Zetaris account manager.

### Data Lineage View

The data lineage view provides the data that can be accessed through the Cloud Data Fabric. This data is queried by the customer administrator and allocated to the warehouse for the end user's query.



## Data Fabric File System

Data Fabric File System allows you to upload CSV, XML, JSON, Parquet, ORC, and Delta files.

The interface shows a navigation menu with icons for Data Fabric Builder, Schema Store, Data Catalog, Query Builder, Virtual Data Mart, Virtual Data Warehouse Builder, Data Lineage View, Data Fabric File System (active), Data Pipeline, and User Management. The main area displays the 'Data Fabric File System' view for the '/file-store' directory. A table lists the files in the directory:

File System Staging	File Name	Type	Created At	Owner	Size
file-store	annual-enterprise-survey-2019-financial-year-provisional-csv.csv	csv	22/04/2021, 11:05:25	Admin	4.9 MB
key-trust-store					

To upload a file:

- Click the **upload** tab, :upload:.
- Select **Browse Files** or **Choose from a folder**.


## User Management

User Management helps in managing different users and their roles.

### Create a new user

- Click the **plus** :plus: button.


- Enter the user's **email**, **name**, and **password**.

 Password must be a minimum of 8 characters. It should include at least one lowercase, one uppercase, one number, and one special character (!#\$%\$\*~@)

- **Confirm password** and click **Add**.

 If using local authentication, Administrators will initially set the User's password, though this can be changed by the user after login.

Similarly, you can create a **new role** by clicking the **plus** button next to Roles.

 The **assign and revoke** buttons at the bottom allow you to add or remove virtual data warehouses from the **User or Role**.

## Reset a user's password

Passwords can be reset by the user by selecting their name in the top right of the screen. Otherwise, the admin can reset through the SQL Editor in the Schema Browser with the following SQL command:

```
UPDATE USER user_id SET PASSWORD 'new_password'
```