
FIVETRAN

1.
INTRODUCTION

FIVETRAN helps you to centralize the data from various sources

Example, you're trying to create a common central repository of data where you trying to pull the data from different, different sources.

Like, SQL DB, Microsoft Azure storage account, Google Analytics ,Maybe different email marketing system and you want to pull the data for that as well, and you want to combine everything in probably you want to do some analytics.

You may want to pull out the reader from all those different sources. So in that case, FIVETRAN could be useful.

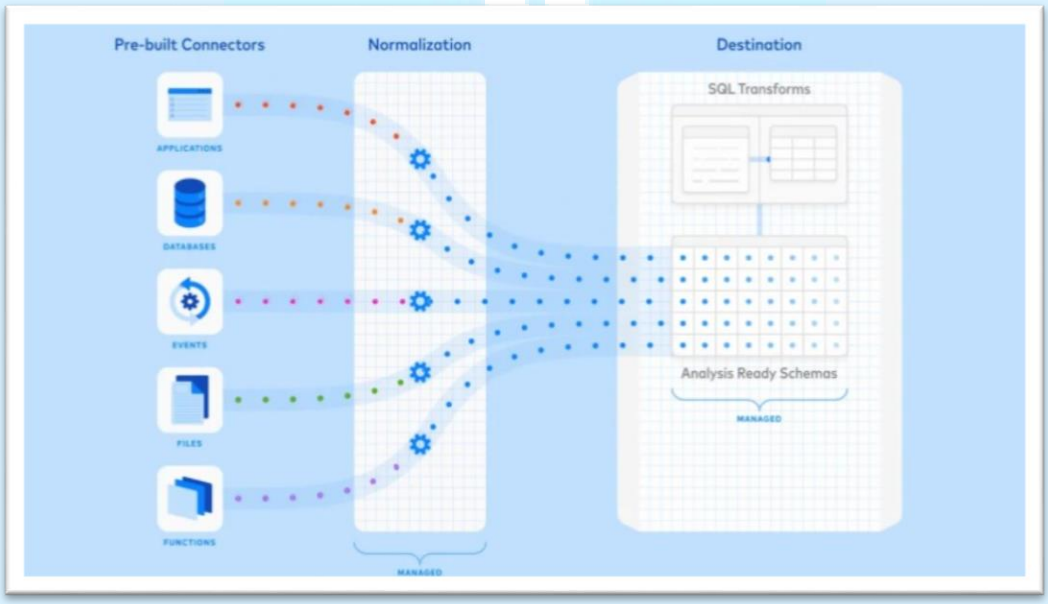
You wanted to extract the data and load into your data destination. But you are loading it from multiple places and you have to store it in a common one area.

Maybe.

For example, you bring up all this data in entire into a snowflake or maybe you pulling out all these it maybe into the ADLS gen2. So it depends or maybe into the data bricks.

Overall the idea is you have multiple sources and you want to move the data from all these multiple sources into a one common location, and that is your destination.

ADF Pipeline for SQL-DB to Blob Storage



Multiple sources of data, you want to move into a one common location that is a destination, then all this can be managed and moved to using the FIVETRAN.

You have multiple sources through which you are bringing in the data and on the Right hand side you have a destination where you pull all this data from these different sources and stored it in your destination. It could be your data warehouse, it could be your snowflake, and it could be your any data destination.

It is a kind of an ELT tool and rather than an ETL Tool.

ETL will stands for extract, transform and load. So you want to pull the data. So that means you want to extract the detail. You've done some transformation, and then you load this data into some destination.

For example, you have a data in the SQL server. You pull this data that lets you extract this data. You done some transformation, removing all knowledge column or a bad records. And then you are you push all this good data into some snowflake data warehouse or somewhere else.

ELT is like, you extract the data and as it is, you load the data into the destination. And once you load it after that, you are doing the transformation.

So that's the basic difference between ETL and ELT.

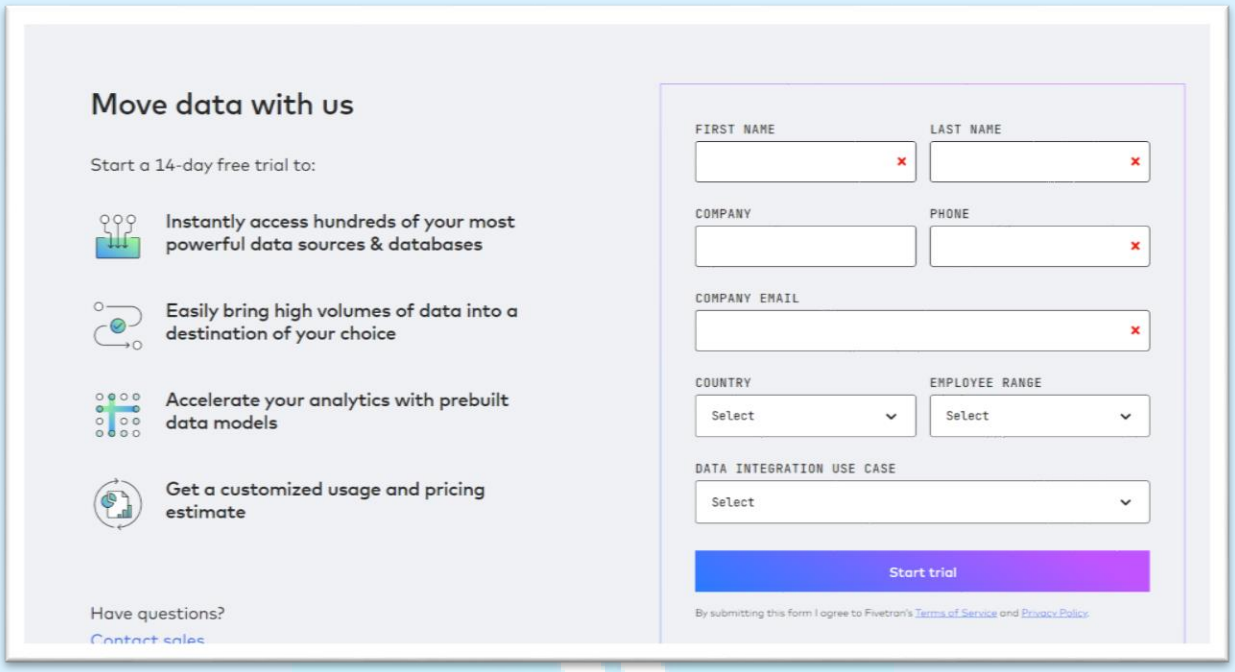
Why fire train is becoming more popular?

- There can be n a number of analytics can be done.
- No installation needed.
- Really quick to set up.
- It’s a very high able to move quickly and easily.

2.
QUICK SETUP

2.1 Create Free Trial Account for Fivetran

https://FIVETRAN.com/signup



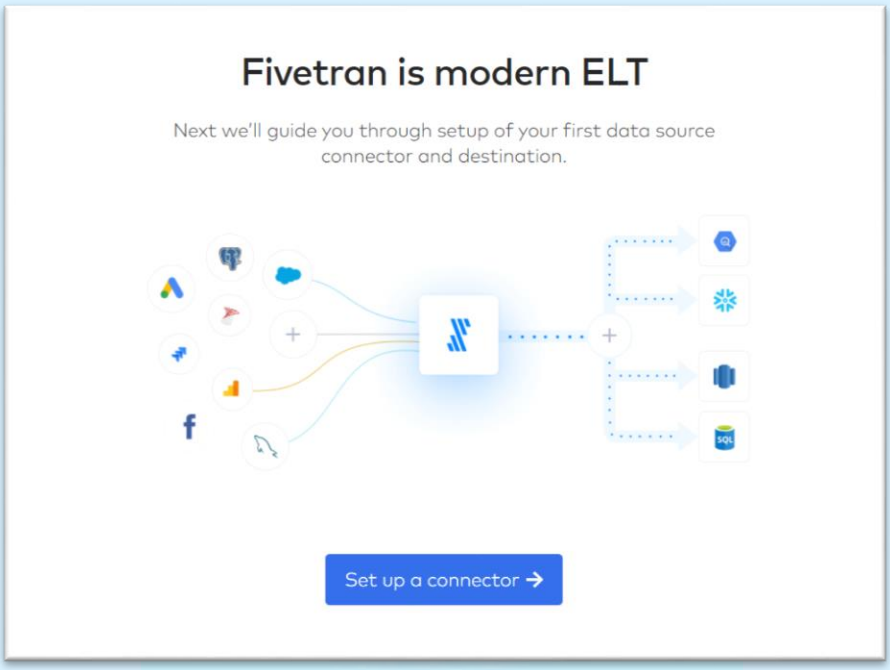
2.2 Add source in Fivetran

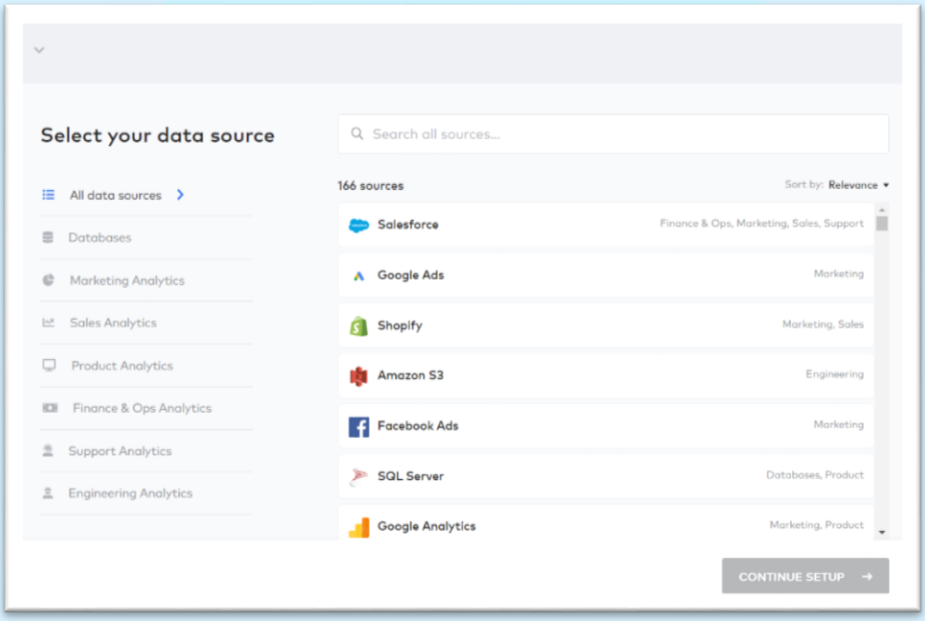
Add First Connector

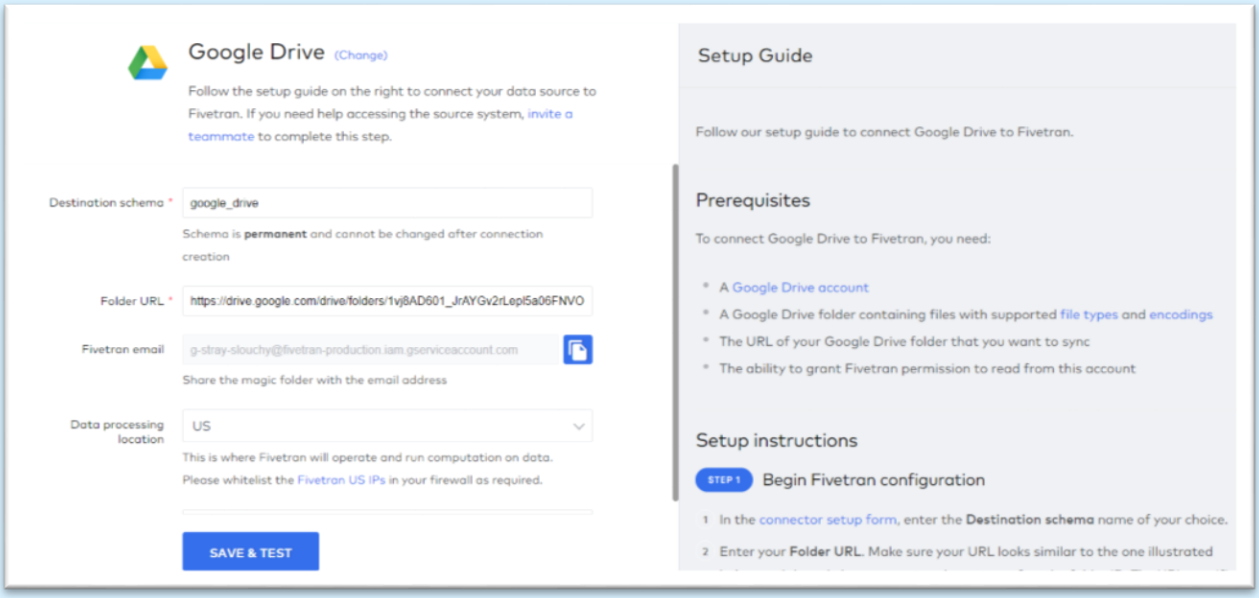
e.g. to synch data from Azure SQL DB to destination warehouse.

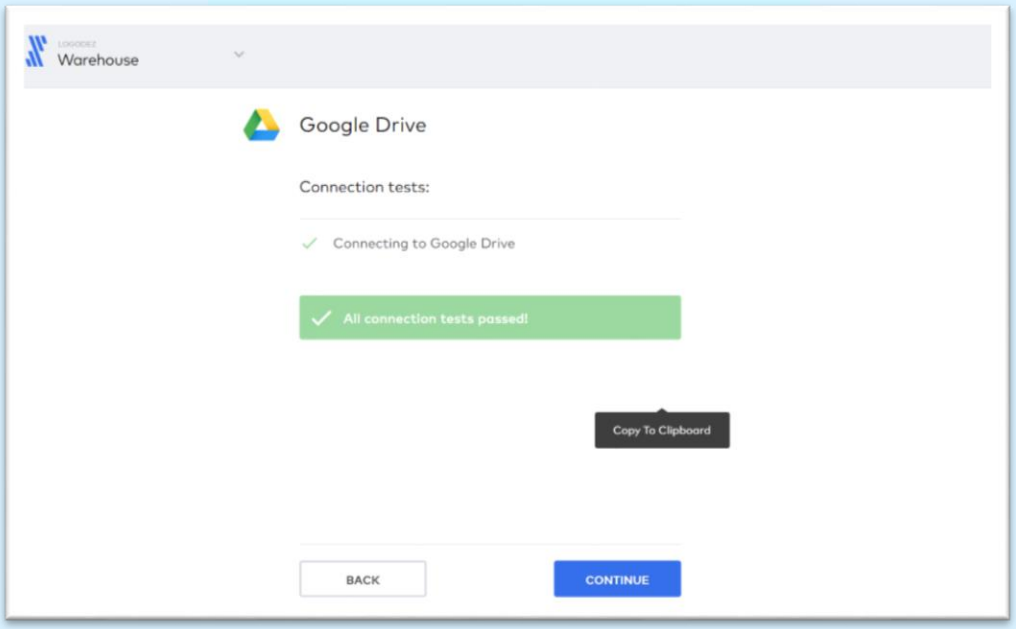
We will going to see, how we can add our first source connector to pull the data from our source to the destination.

We will going to do is we will try to pull the data from this Azure SQL database into our destination warehouse.






Google Drive (Change)
Follow the setup guide on the right to connect your data source to Fivetran. If you need help accessing the source system, invite a teammate to complete this step.
Destination schema *
google_drive
Schema is permanent and cannot be changed after connection creation
Folder URL *
https://drive.google.com/drive/folders/1vj8AD601_JrAYGv2rLepI5a06FNVO
Fivetran email
g-stray-slouchy@fivetran-production.iam.gserviceaccount.com
Share the magic folder with the email address
Data processing location
US
This is where Fivetran will operate and run computation on data.
Please whitelist the Fivetran US IPs in your firewall as required.
SAVE & TEST
Setup Guide
Follow our setup guide to connect Google Drive to Fivetran.
Prerequisites
To connect Google Drive to Fivetran, you need:
A Google Drive account
A Google Drive folder containing files with supported file types and encodings
The URL of your Google Drive folder that you want to sync
The ability to grant Fivetran permission to read from this account
Setup instructions
STEP 1
Begin Fivetran configuration
1 In the connector setup form, enter the Destination schema name of your choice.
2 Enter your Folder URL. Make sure your URL looks similar to the one illustrated


Warehouse
Google Drive
Connection tests:
Connecting to Google Drive
All connection tests passed!
Copy To Clipboard
BACK
CONTINUE

2.3 Fivetran types of destinations

In the FIVETRAN there are two types of destination is available. First is you can use your own destination. It means that you want to move all your data into your choice.

Choice of data warehouse, for example: You might have a snowflake, you might have a SQL Data warehouse, you might have some another data warehouse which you wanted to choose as a destination.

So if you have your own destination is ready, you can use that.

Otherwise, you have a second option in which your destination is nothing but a top FIVETRAN managed destination.

So that destination is managed by the FIVETRAN and FIVETRAN automatically give you a default destination. And this destination is basically a big data warehouse.

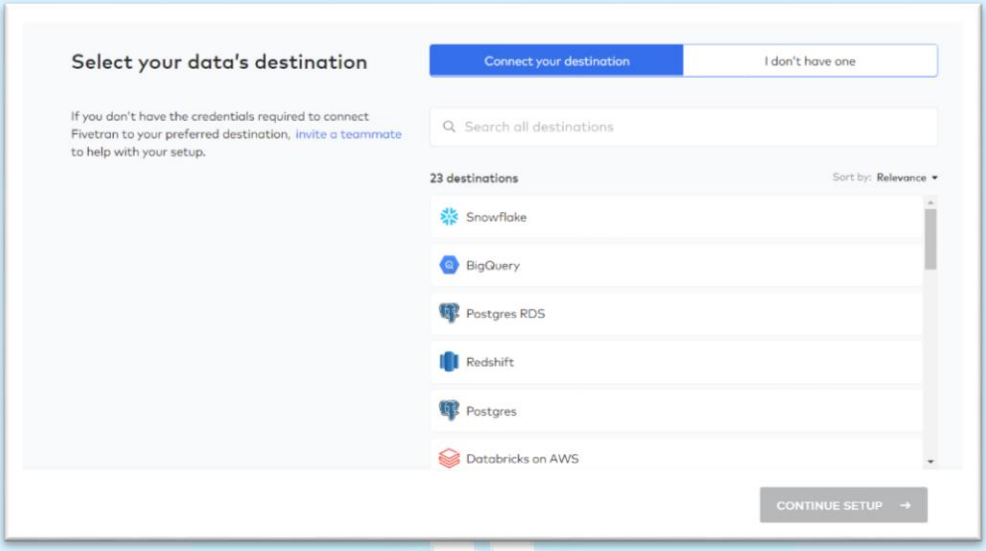
So you will get 14 days of free trial for a big data warehouse, also indirectly. And you can go and use that. So it's up to you how you want to set your destination. Either you want to go ahead and use your own destination or you just for this practice purpose.

3.
Implementation

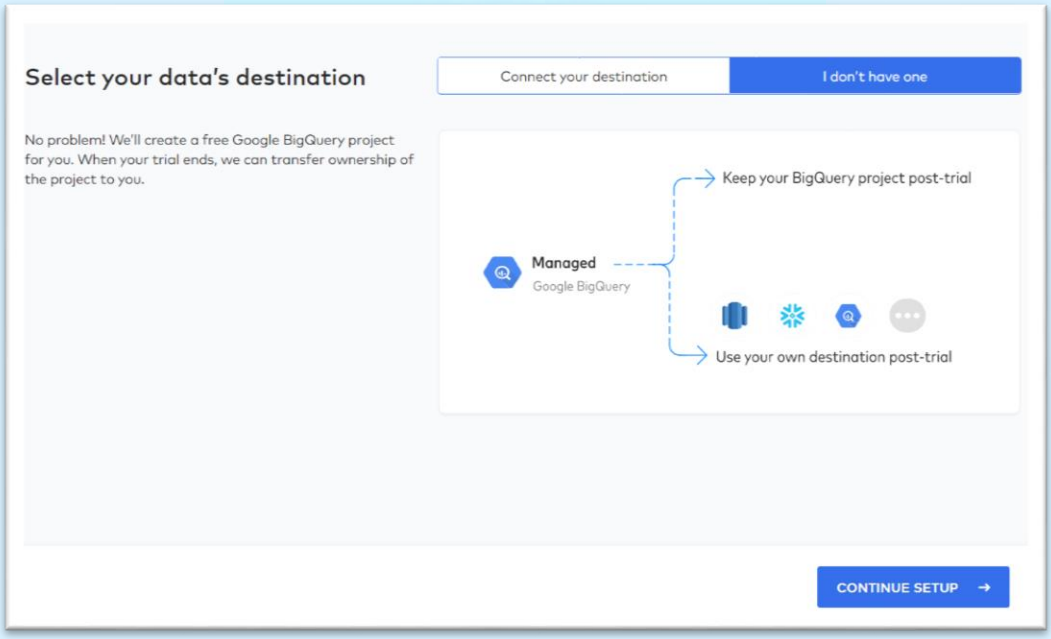
3.1 Add the Fivetran Managed destination

Source was our G DRIVE and now you can see that it is asking us to connect your destination.so you can see you have an option. Either you can select anything from here, from all these available destinations. So all these 23 destinations are possible. Either it could be a SQL server, it could be a Databricks..... etc.

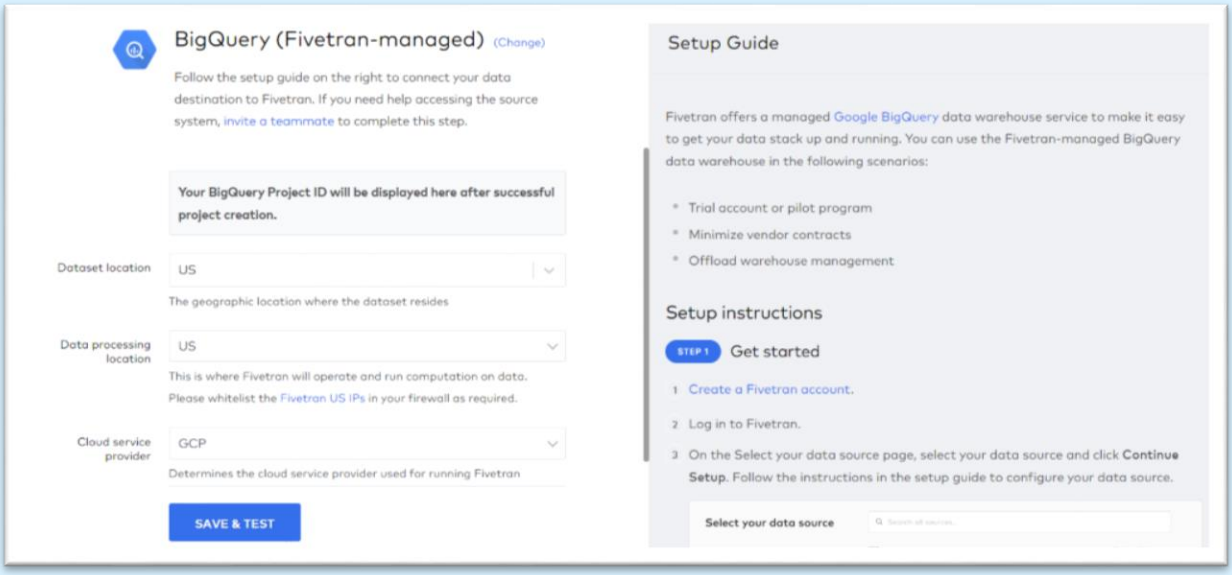
So you can choose the destination based on your choice, or else what you can do is you can click on this, I don't have one, and then I will create a one destination for you And that will be not your own, but that will be the FIVETRAN and manage one.

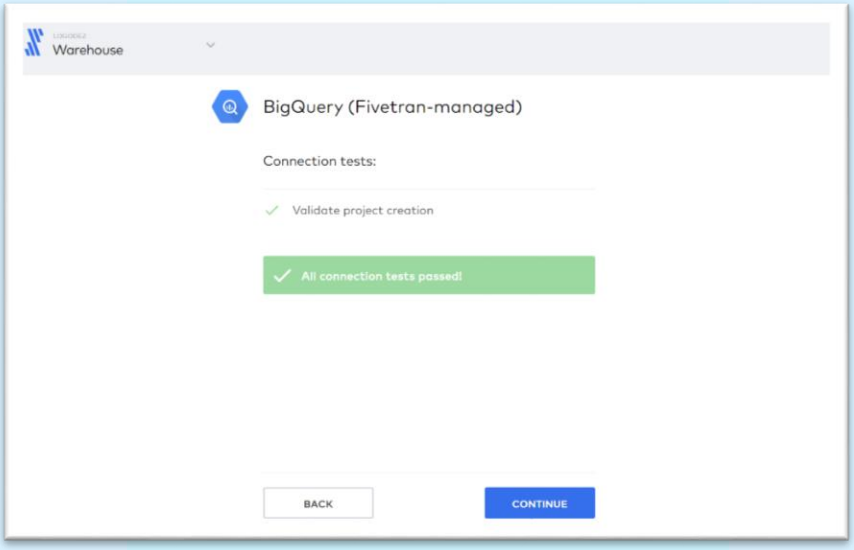


Another option



So let just click on I don't have one and you can see that it saying that it will go and keep all your data now into the query.

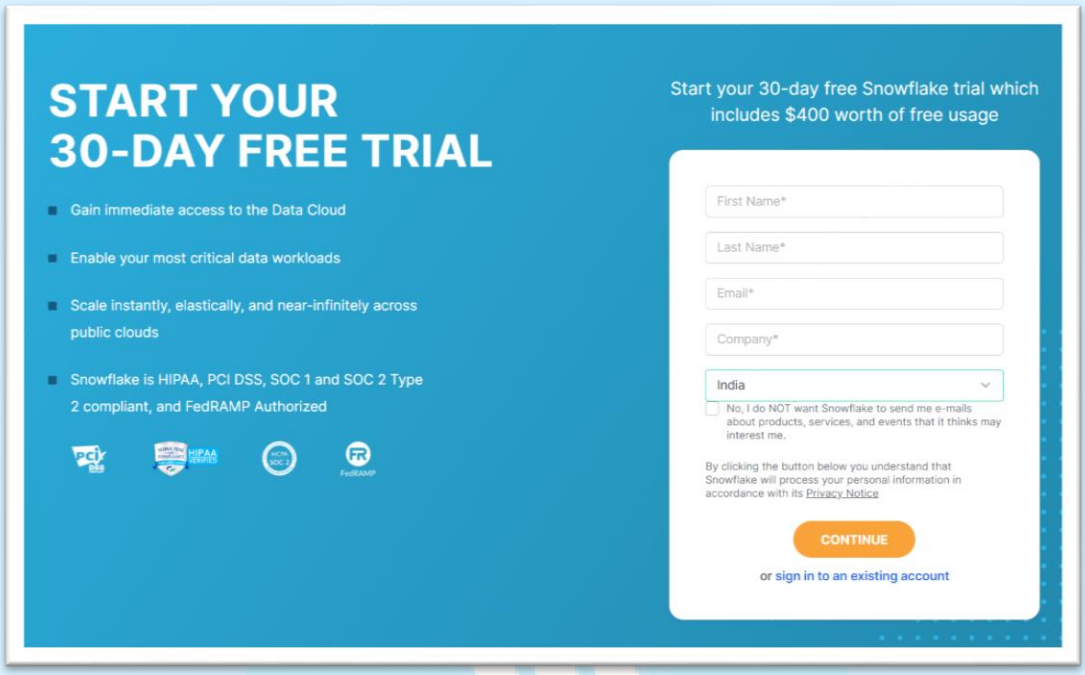




So your connection is really your source was really source you have already selected that was your single database and we have created a destination that as the fight train managed Big Query.

3.2 Create a Snowflake Free trail account

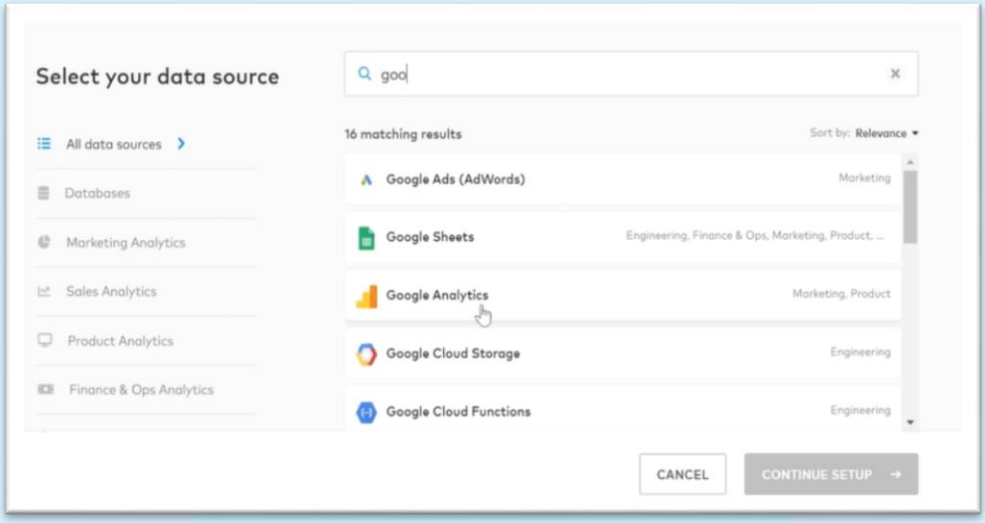
https://signup.snowflake.com/

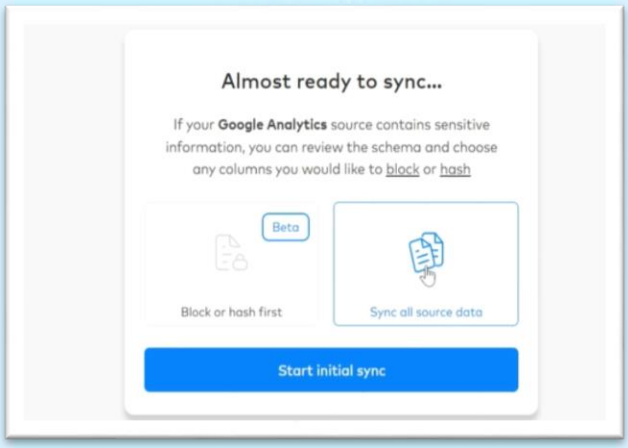


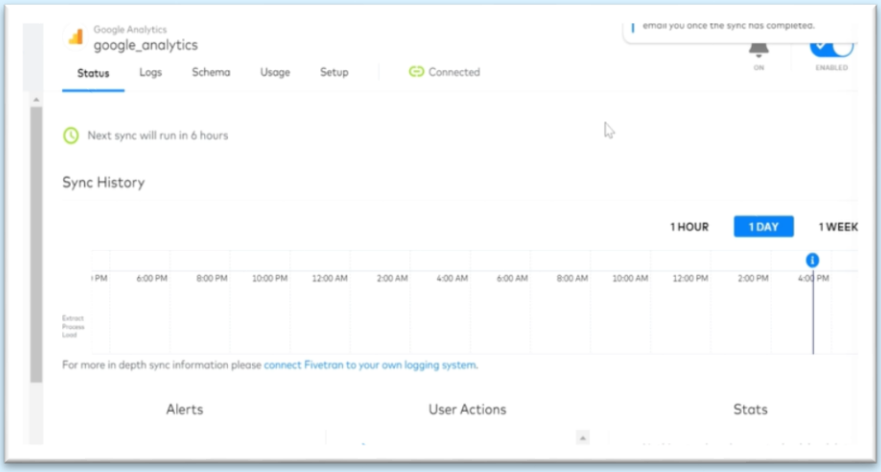
You will see lots of sample Data here.

3.3 Add Google Analytics as Source

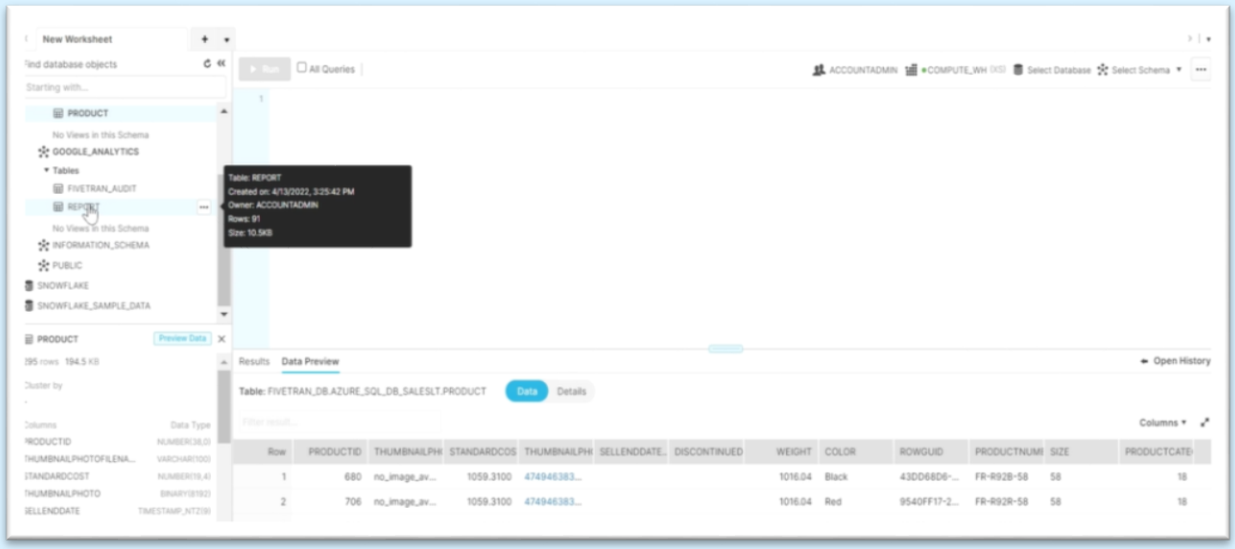
Add connector -







Also you can see this in snowflake -



3.4 Add Transformations

Transformations are SQL scripts that are executed on your data based on specific events or conditions.

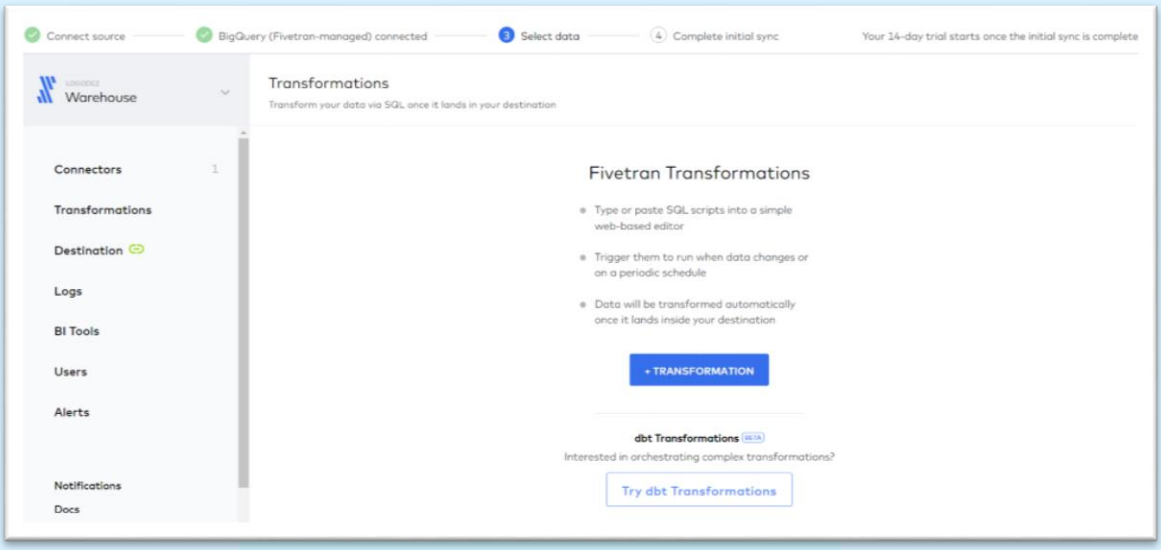
Transformations run in your destination after FIVETRAN load your data in destination.

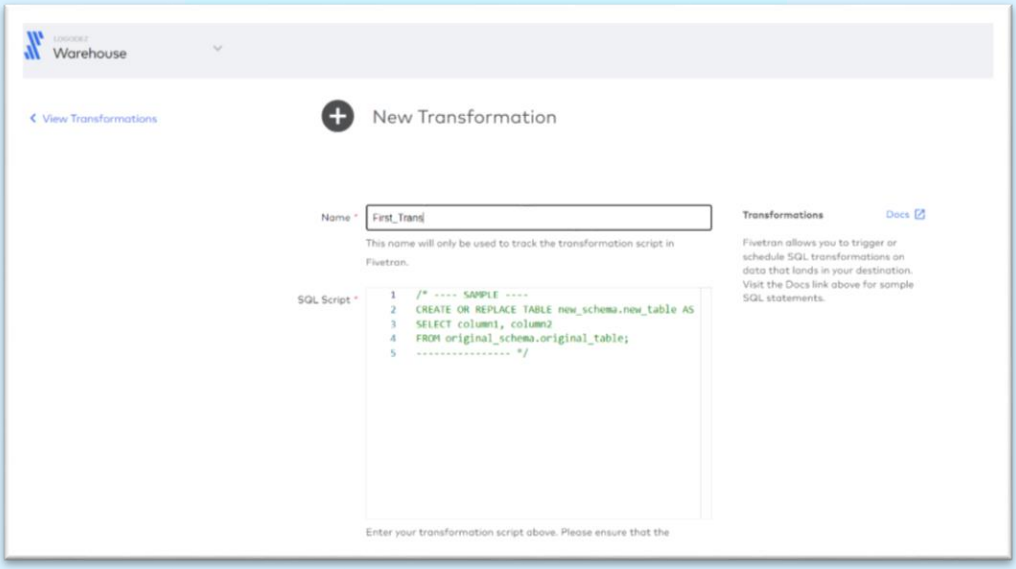
If a transformation fails, you do not lose data.

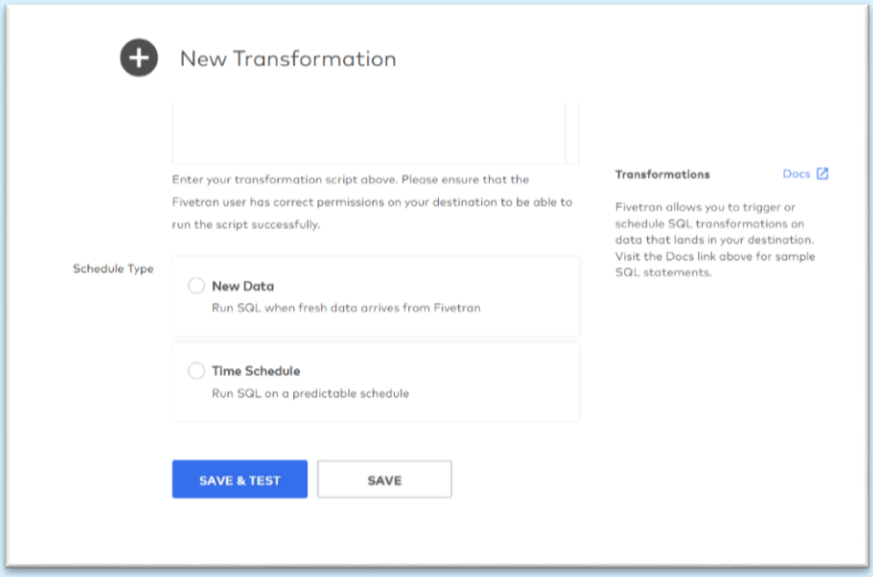
Transformation options -

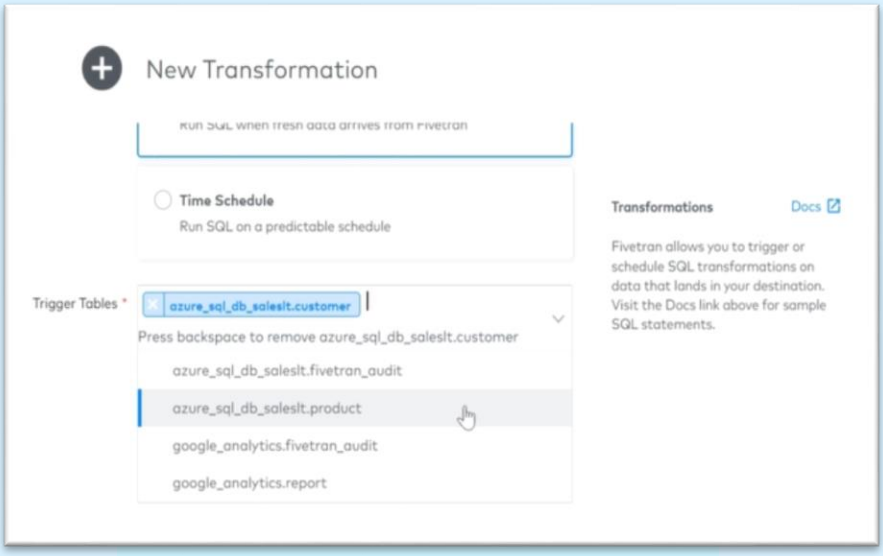
- Basic SQL Transformations
- Transformations for dbt Core

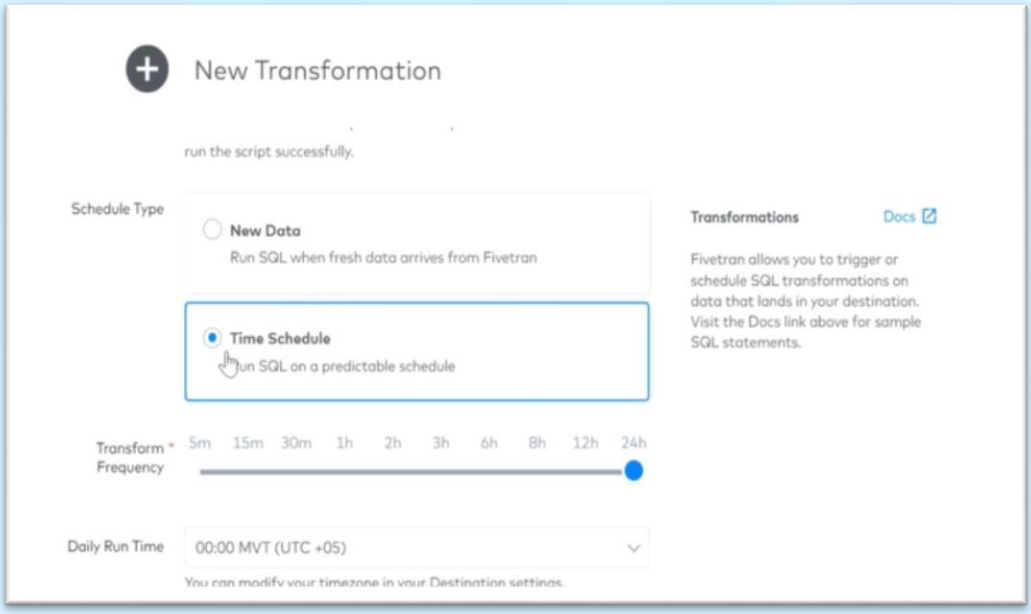
dbt has its own learning curve and whatever you can do there as far as the business is concerned, same thing you can achieve through your SQL as well.






New Transformation
Enter your transformation script above. Please ensure that the Fivetran user has correct permissions on your destination to be able to run the script successfully.
Schedule Type
New Data
Run SQL when fresh data arrives from Fivetran
Time Schedule
Run SQL on a predictable schedule
SAVE & TEST
SAVE
Transformations
Docs
Fivetran allows you to trigger or schedule SQL transformations on data that lands in your destination. Visit the Docs link above for sample SQL statements.

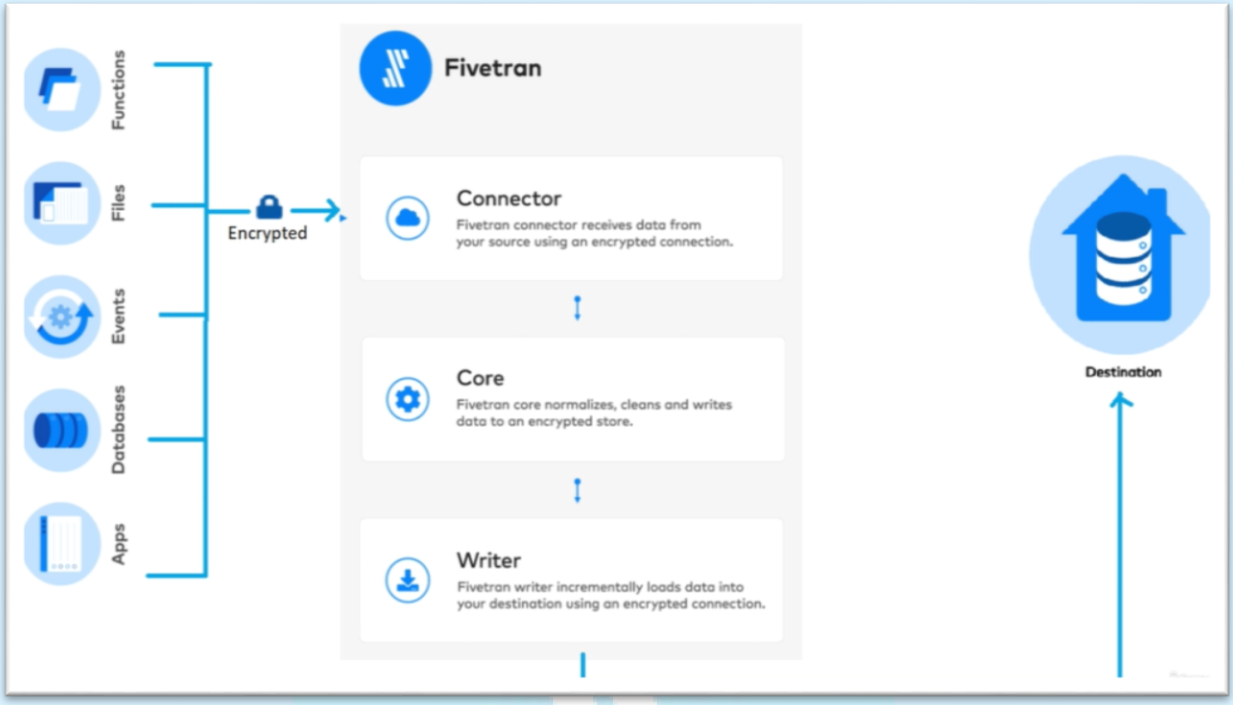

New Transformation
Time Schedule
Run SQL on a predictable schedule
Trigger Tables *
azure_sql_db_saleslt.customer
Press backspace to remove azure_sql_db_saleslt.customer
azure_sql_db_saleslt.fivetran_audit
azure_sql_db_saleslt.product
google_analytics.fivetran_audit
google_analytics.report
Transformations
Docs
Fivetran allows you to trigger or schedule SQL transformations on data that lands in your destination. Visit the Docs link above for sample SQL statements.


New Transformation
run the script successfully.
Schedule Type
New Data
Run SQL when fresh data arrives from Fivetran
Time Schedule
Run SQL on a predictable schedule
Transform Frequency *
5m 15m 30m 1h 2h 3h 6h 8h 12h 24h
Daily Run Time
00:00 MVT (UTC +05)
You can modify your timezone in your Destination settings.
Transformations
Docs
Fivetran allows you to trigger or schedule SQL transformations on data that lands in your destination. Visit the Docs link above for sample SQL statements.

4.
Architecture

4.1 Architecture

https://FIVETRAN.com/docs/getting-started/core-concepts



FIVETRAN core is having a temporary storage basically and this temporary storage, it's the place where FIVETRAN normalized that it does clean and right data. So you can see that on the left hand side we have different types of data sources like ABS, database events, files and functions. So all these data sources are these different types of data sources sends the data to the FIVETRAN.

Now to receive the data from the data sources, we first create the connectors. So what this FIVETRAN connector do is FIVETRAN connector receives the data from these data sources through the encrypted connection so that all your data transfer happens in a secure manner.

Once this data get passed through this connector FIVETRAN do the normalization and FIVETRAN cleans this data and do deduplication also and writes this data to the internal temporary storage which is in the encrypted form. So internally FIVETRAN holds your data in the temporary storage. And once all this process has been done, now the FIVETRAN writer actually going to write incrementally.

Your data into the destination that is your destination data source and that happens again through the encrypted connection so that you can see that all the data which flows from the source to the FIVETRAN and from FIVETRAN to the destination, all this data are processed through the in the encrypted manner Only and FIVETRAN internally does the normalization clean and a do the deduplication, etc. in the Temporary storage location and hold your data in that

temporary storage for almost seven days. And once everything get done after that, these data get deleted. And also the Data has been stored temporarily within the FIVETRAN when that data also store and then you get paid for them.

So your data is encrypted addressed. Basically there are two types of connector

First one is the pull connector.

Pull connector - FIVETRAN pull connectors actively retrieve, or pull, data from a source FIVETRAN connects to and downloads data from a source system at a fixed frequency.

Is it something where you go and you retrieve the data from the source based on the the predefined frequencies

For example, you're Azure SQL DB, SQL server.

So these are your normal up pull connectors. So you just go there and fetch the data.

second one is the push connector

Push connector - In push connector, such as webhooks or snowplow, source systems send data to FIVETRAN as events.

So like in the push connectors, there are different types of sources, right? You have about webhooks where you have a callback URL, which is like even driven. So like Snowplow.

So these kinds of source system sends the data to the FIVETRAN in terms of the event and FIVETRAN all those events in terms of the JSON to fetch the data, etc.

4.2 Manage sync

The first historical sync that FIVETRAN does for a connector is called the initial sync.

Whenever you make or add over the first connector and attach it to the destination, the first type of sync, which happens up it's called is an initial sink.

Initial sink is nothing but basically an historical sink.

So like when first time you make us think you will try to pull all the data, whatever is at your source site. So your initial sink in general will take time. And by default, by train will see that as soon as your initial is complete, you will get notified.

So that's the first initial thing, which you can also call it as an historical sink to pull out entire data available at the source site.

After a successful initial sync, the connector runs in incremental sync mode. In this mode, only data that has been modified or added - also known as incremental changes.

Once your initial sink done, then you might probably wanted to run the sink on the periodic basis. So that does nothing but called as your incremental sink. So in that case, most of the time you wanted to move the data or pull the data which has been modified, added or got deleted.

So you try to get know about that, those information and you will pull that kind of data only. So that comes up as an incremental change data.

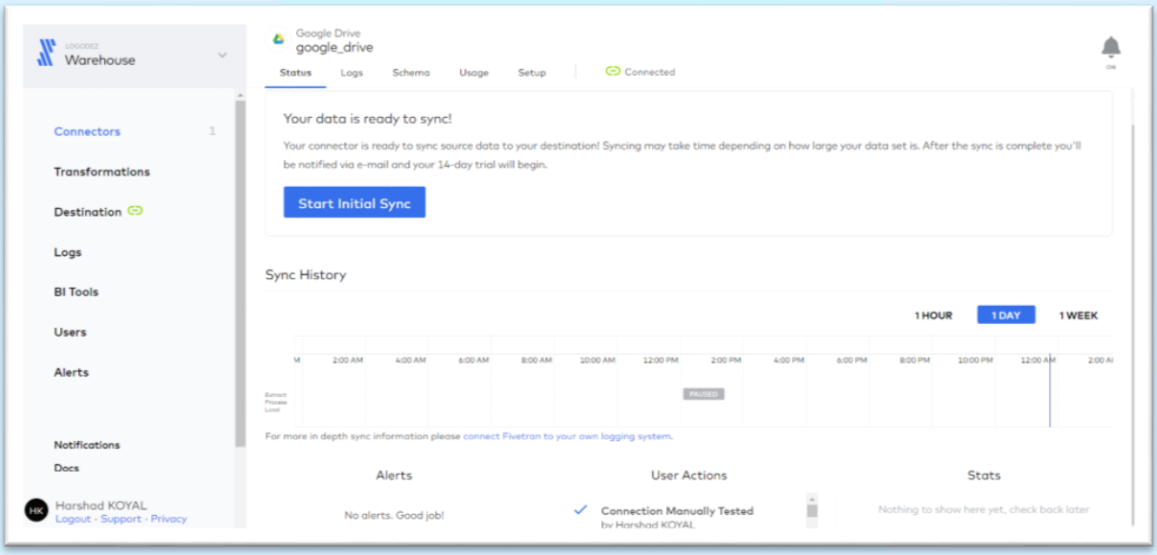
In some cases, you may need to re-run a historical sync to fix a data integrity error. This is called a re-sync.

In some cases, you may want to read an end and start SQL thing because something bad happened at the destination site.

You're getting some data issues, some integrity issues. Something got deleted accidentally at the destination maybe. So in those cases, you might want to do the racing. So Resync is nothing. It's again a kind of an initial thing that you upload the entire data and get overwrite at the destination.

Sync History Chart

COLOR	STATUS	NOTES
Green	Completed	Your sync finished running with no errors.
Blue	Incomplete	Your sync is currently running.
Purple	Rescheduled	Fivetran reschedules syncs in two situations: • Fivetran has reached the request limit for your application's API. • Fivetran has exceeded the set number of API calls we have provisioned for your connector. In either rescheduling scenario, the sync automatically pauses and resumes the following day, once the daily API limit has been restored.
Gray	Unknown	Your sync has exhibited atypical behavior and has not completed for unknown reasons.
Red	Failed	Your sync could not finish running. The stage your sync failed in appears in light red.



Sync Frequency Overlapping

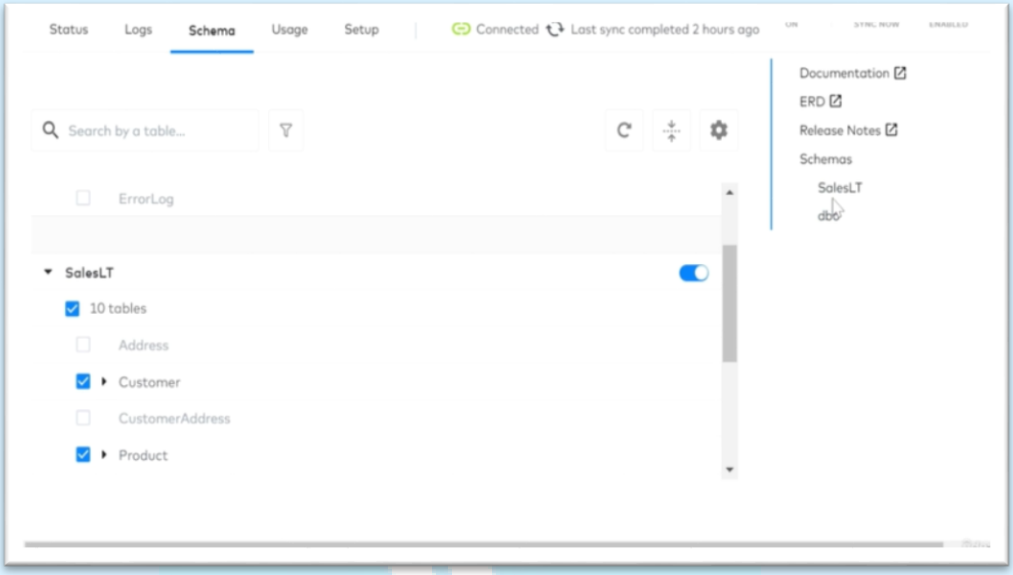
SYNC NUMBER	START TIME	END TIME	NOTES
1	08:00	08:10	
2	08:15	08:35	
3	~~08:30~~ 08:45	08:55	Sync 3 is postponed because Sync 2 is still running. Sync 3 then starts at 08:45.

4.3 Connector schema and the logs

In the dashboard, on the left hand side you can see that we have a connectors. Select one of the connector. So, whenever you select the connector here, you see the multiple things like status, logs, schemas. Click on this schema. This schema is nothing but the schema of this connector.

For example, our SQL DB has two schemas. One, what "sells LT" and another one was "DBO".

So on the right hand side, you can see that we have a self LP and we have a double.



In some cases you might want to add a few more tables which you wanted to sync on. You want to remove the existing table from the sync. So what you can do is you can make the changes from here. Or maybe you want to resync [keep the cursor on schema] the entire table.

Unhashed- Data values at the source will be replicated identically to the destination. [Keep the cursor on column]

Hashed - Hash data values in this column before syncing to your destination. [Keep the cursor on column]

So like a going to hash your values before sinking into the destination or it will not going to sync based upon whether you have selected and unhashed or hashed.

So that can also you can set for the data privacy, although this schema tab little bit changes when you have a different type of a connector. So this is what we are looking for, the SQL.

You can see that automatically you are getting an option search by executor. So that's how you can see the schema.

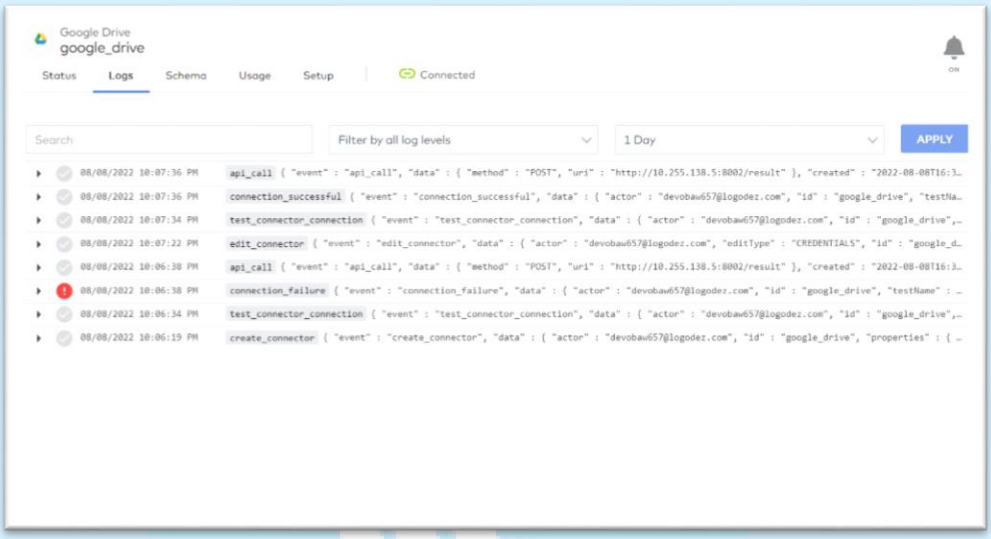
Remember one thing if I turn automatically manages the changes in the schema.

So let's say for example, you already created a connection and after that do more columns got added up into your source.

So automatically when the next thing going to happen, FIVETRAN will pull these to extra column and it will go and create those extra columns into the destination. So that's how it can manage the schema changes as well.

You can click on this logs, so you will see a separate logs for every connector. So that defines what happens to that connector when they try to sync.

So you can see you can search all those log by different types like filter, filter by warnings, filter by schema change, and events like that. You can select the duration from here and based on that you can apply and all the logs will be automatically displayed to you here at.

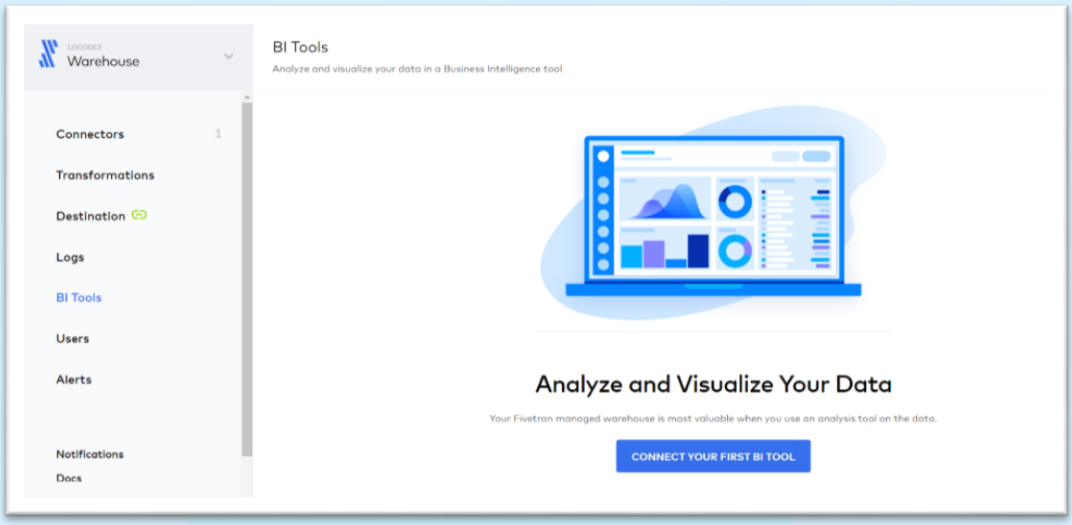


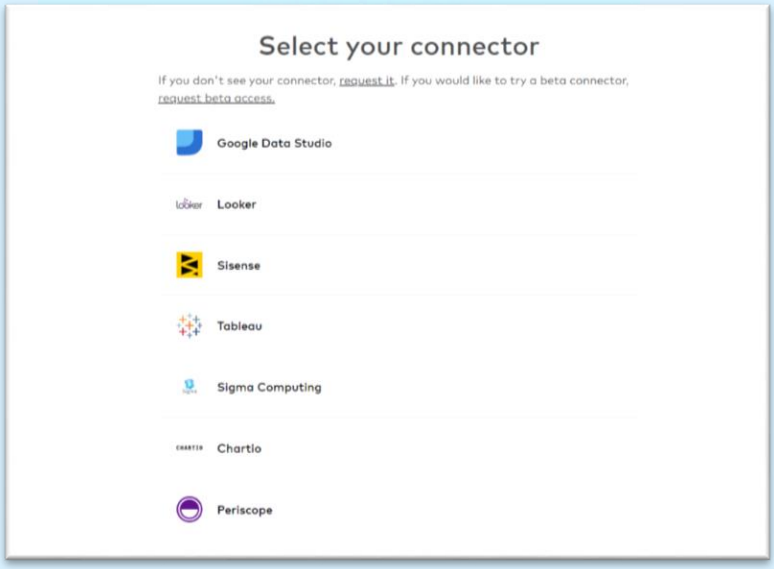
In some cases you might would like to go and check the different log types may come down here and you can see there are different connected event. So because of these events you can search for it in your logs and you will get the changes based on that. So if you have something specific you're searching for maybe this even they could help you.

So eventually every log is nothing but a kind of an adjacent object in itself like that. So you can see this name and you can build on such based on that even name as well. So you could do that searching and you can use this in debugging or monitoring purpose. Besides this, there is also an explicit way of whatever the logs are generating. You can go and externally you can see all these logs either in your AWG or maybe into your Azure monitor. So that can also be configured. So that's all about the connectors, schemas and the logs.

4.4 Connect With BI Tool

The FIVETRAN itself, you can connect the reporting tool which can fetch the data from this destination and can create the reports for you.





Come down to our FIVETRAN dashboard. And here on the left hand side, you can see that there is BI tools. So what you can do is you can come down to this bar tool and here on the right hand side, top corner you will see plus a tool. So here you can make a data connection to any bar tools.

So here you can make a data connection to any bar tools. So at the moment there are these many bar tools supported by the FIVETRAN.

EXAMPLE - So let's make a connection to Google Data Studio.

We come down to Google Data Studio, it will ask you to authorize. So in general, if you have a Google account, you will get a free Google data studio access. So there is nothing much you need to do.

Just click on Authorize. It will ask you to select the specific email ID to let me see. Like this is Gmail id. And you can see that the authorization has has been come out successful.

So you just know down this project ID, that this is your project. Go to the data studio. Now, here you can go down and you can click on this, plus create and create data source.

Now, in the data source, we know that our destination in the FIVETRAN was a big query. So we select our data source type as a big Query. And automatically you can see that we got this authentication for this data set.

And you can select the table and you can make the connect. So everything goes so smooth that you will find it so easy to do it. And now you can see that you get entire schema of your data source. And here you can go to create report and you can make your report as per your business need.

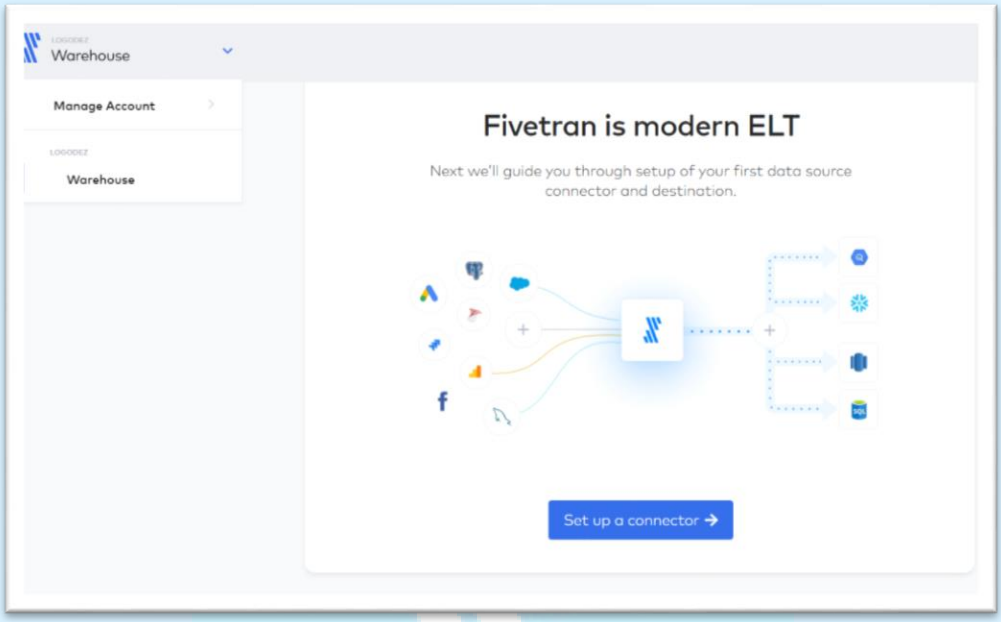
INSHORT,

I went down to The Fighter and I would select the tool from there, and that was I selected Google Data Studio. I Make authorization here. I select the my respective Gmail account and saved, later that's how I got to created a connection with the authorization. I come down to the Google data studio here. I come down to create ad data source and as they know, their destination was my big query in the FIVETRAN. So I select data sources, FIVETRAN

5.
ACCOUNT MANAGEMENT

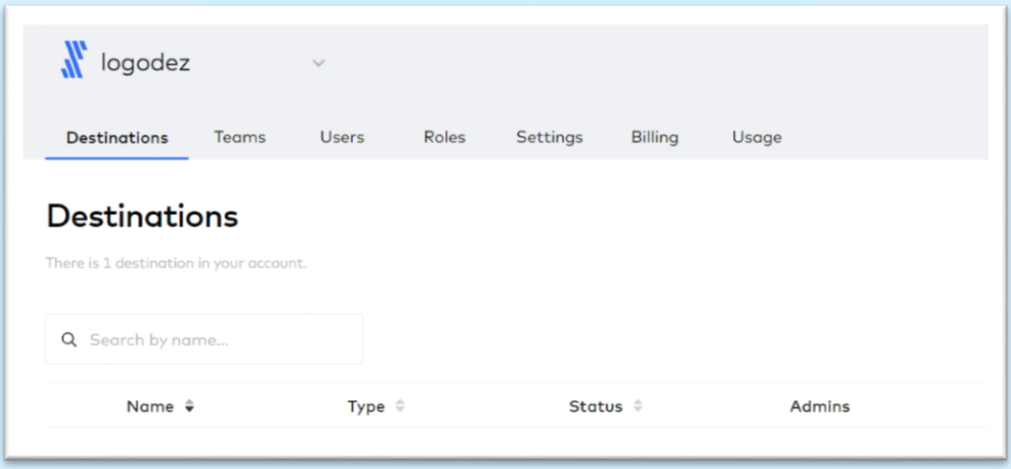
5.1 Account Management Overview

Our FIVETRAN portal where you can see that by default there is a one destination Got selected that you can see on left inside top corner. So if you can just click on that, you can see we have one more destination that was Big Query Warehouse. So this time for doing the account management.



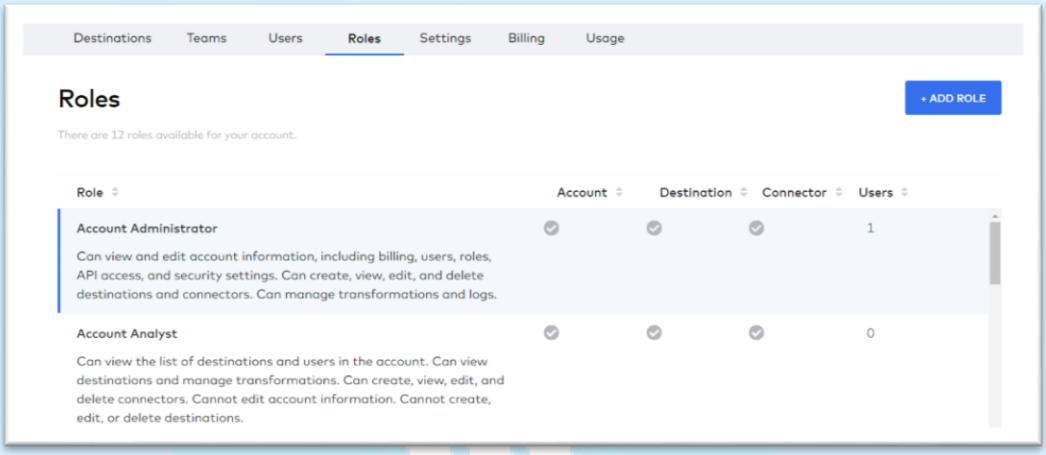
This will take me to the account management page of FIVETRAN.

Now here you can see you have a destination tab, you have a user tab, you have rules, you have settings, you have billing and you have a usage.



5.2 Add Custom Role | Manage Role

On the dashboard, you can see that the third tab is for the roles. So just click on the rules and you can see that by default there are multiple roles like account administrator. You can see that what kind of for different types of permission it has. It has all connectors related to account destination characters like that account analyst.

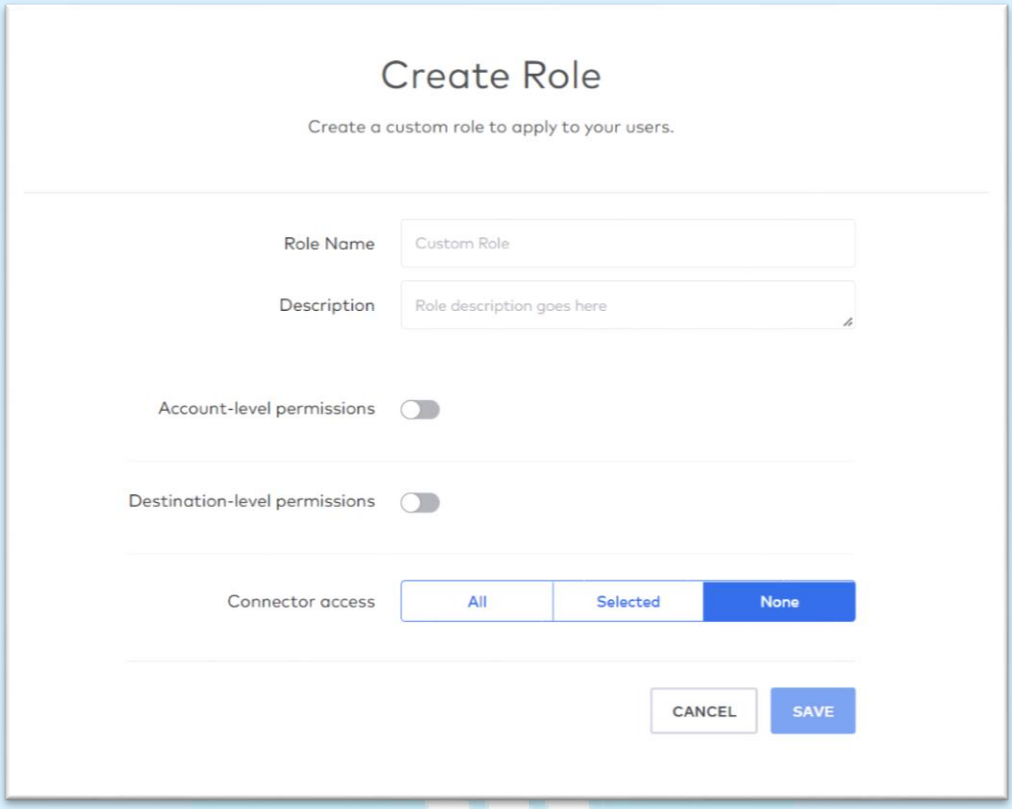


You can have account billing account reviewer. So all these different types of rules you already have. So there are total 14 by default rules out there.

In some cases, you might want to add your own rule. So that is called the custom rules. So when you want to define the rule based on your own needs, for that, you have to go and create the ADD rule. So just click on Add rule. You give the name of your rule, modify the rule. I just give that name. You want to give some description around it. You can give that description as well. Now it will ask you the account level permission. So like when you create this rule, what kind of account level permissions you wanted to give to this role.

So you click on this account level permission and here you can see that these are the different options.

You are getting in. So let's say settings. So what kind of settings he can do, he can edit, he can view or he having none access to the settings. So let's see. He can have a view access. You can have a billing.....etc.



So you can define a very fine grained access through these rules, even for the destination level.

If you want to give an access for a specific connector only, so you can select that.

Some cases you might want to delete this rule all together. So what you can do is you can click on this delete and that will going to delete that rule from your FIVETRAN account.

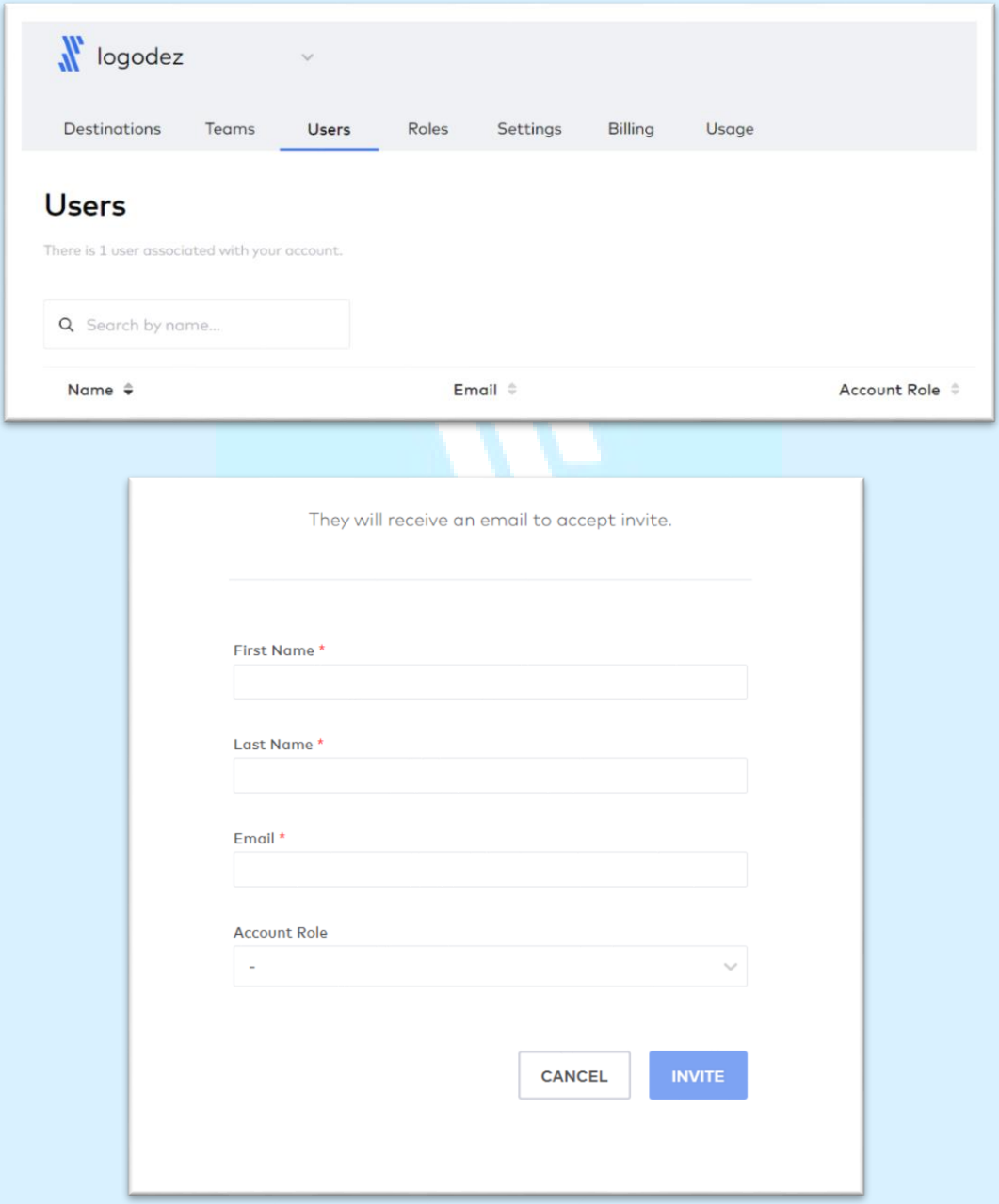
So basically in real world scenarios, most of the time as a developer, you might not have an access. To create multiple rules.

So like the network admin team or the DevOps team or maybe your manager or a very senior folk. So those kind of people generally have a limited access as an account to permissions to do all this thing. So they create these kind of rules and assigns these rules to the newly joined users, etc.

So most of the users, the user who just work as a developer, might not have an access to create a rule. But again, that will be defined by the kind of rule given to them by their administrator.

5.3 Add User

Let me give the name. I give them first name, last name and email ID and while giving the account role, you can see that there is a dropdown and in it by default the custom role which we have created is coming in. So it means that if I select this my FIVETRAN role, that is the customize role this user Mike will have and by default access permission as defined by this role.



So let me just click on invite. So this will send out an email to Mike on this email ID and once he join, he can have a specific role.

So let me just go to the email. This is my email only and let me just show you how an email comes like. So now you can see that I got an invite from FIVETRAN to connect and create an account.

So what I'll do is I'll just click on the sign up. So now you can see that automatically I get logged into the same account and this time you can see on the top, my user is Mike Hudson. And the good thing is I have a limited access.

So now you can see that when I click on Added Destination, it is not showing me I am not able to click it. So that's how we can add a new user with a predefined custom role.

If you want to delete any existing user, if you can just click on this delete and buy this, you can go and build that user right of it to just click on delete and this user will going to be deleted permanently.