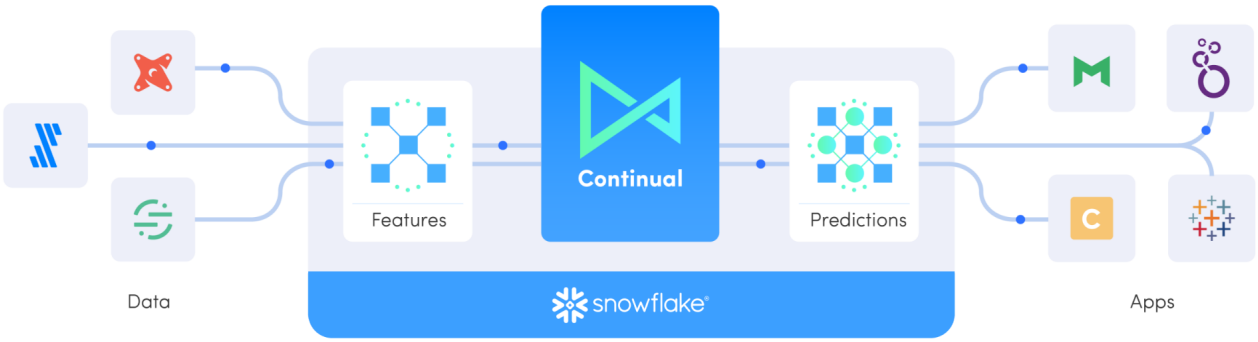
ENTERPRISE AI FOR THE MODERN DATA STACK

Overview

This guide will show you how to easily add Continual as the AI layer to your modern data stack with Snowflake at the core. The intention is to provide an introduction to using Continual on Snowflake. After completing this tutorial, users are invited to try more advanced examples.

We are going to demonstrate connecting Continual to Snowflake, building feature sets and models from data stored in Snowflake, and analyzing and maintaining the predictive model continuously over time.



To keep things simple at the start, we'll use a nicely manicured, fictitious dataset to illustrate how Snowflake and Continual combine to enable modern data teams to effectively build, deploy, and utilize production grade models. The dataset consists of customer information such as account data, demography, geographic area, and phone activity of a fictional telecommunications business. It also conveniently contains a boolean value per customer defining whether or not the person ended their contract and "churned". While this dataset will suffice the purposes of quickly trying Continual + Snowflake, we don't believe the telco churn dataset is the most realistic example of customer churn, which is why we created a more comprehensive example you can try next!

WHAT YOU'LL LEARN

- How to connect Continual to Snowflake and do machine learning on your data warehouse data cloud
- Create feature sets and models in Continual
- Evaluate and maintain production machine learning models
- Analyze model performance, input data, and features to iteratively improve performance
- Writing back prediction to Snowflake

PREREQUISITES

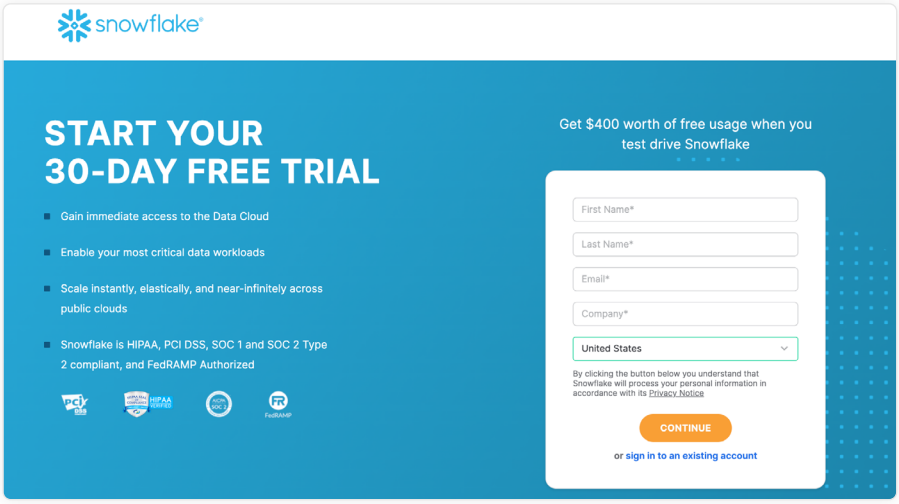
- Basic experience with Snowflake and SQL
- Basic knowledge of machine learning and data science problems

Prepare your lab environment

SETTING UP SNOWFLAKE

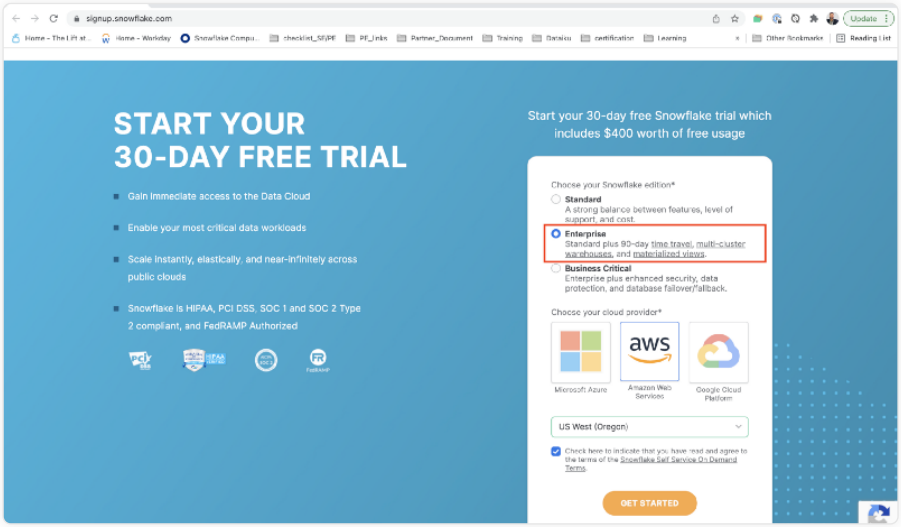
If you have a Snowflake account, then login using your unique credentials.

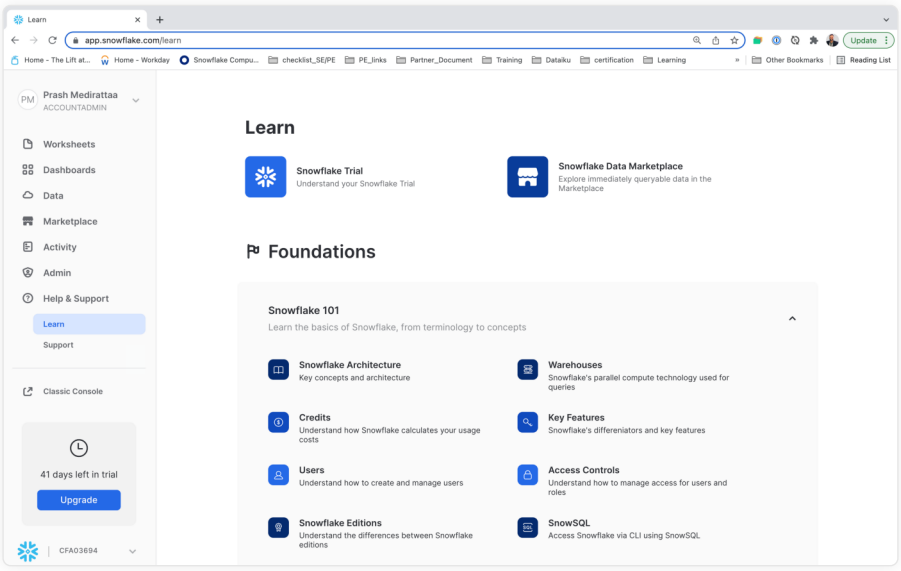
If you don't have a Snowflake account, visit https://signup.snowflake.com/ and sign up for a free 30-day trial environment.



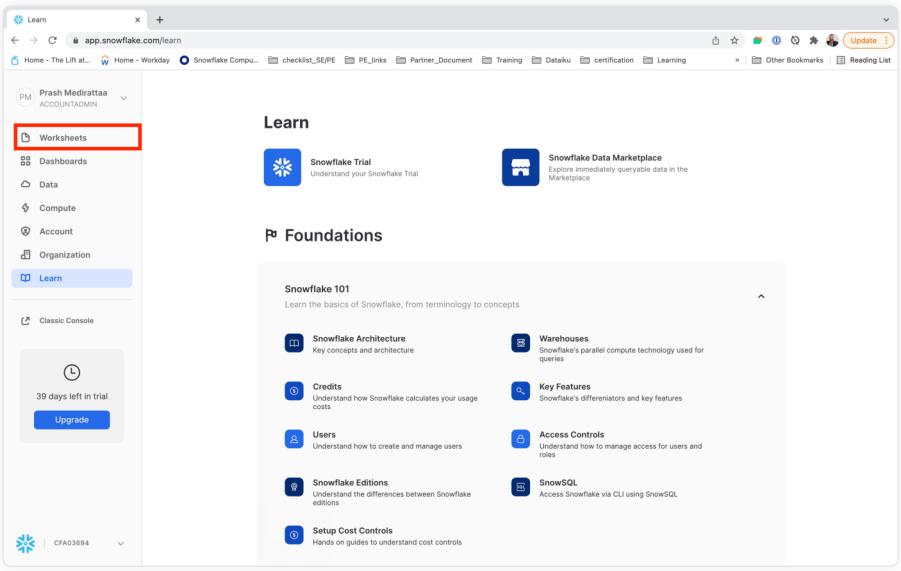
For this example, you will only need the Standard edition on AWS. But you may want to select Enterprise to try out rad features like time travel, materialized views, or database failover.

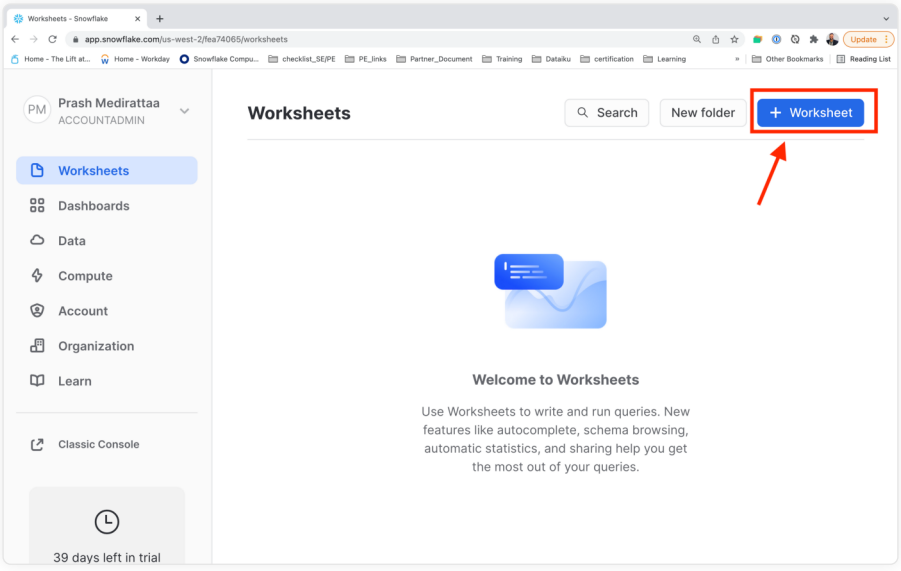
Choose US West (Oregon) for the AWS region.

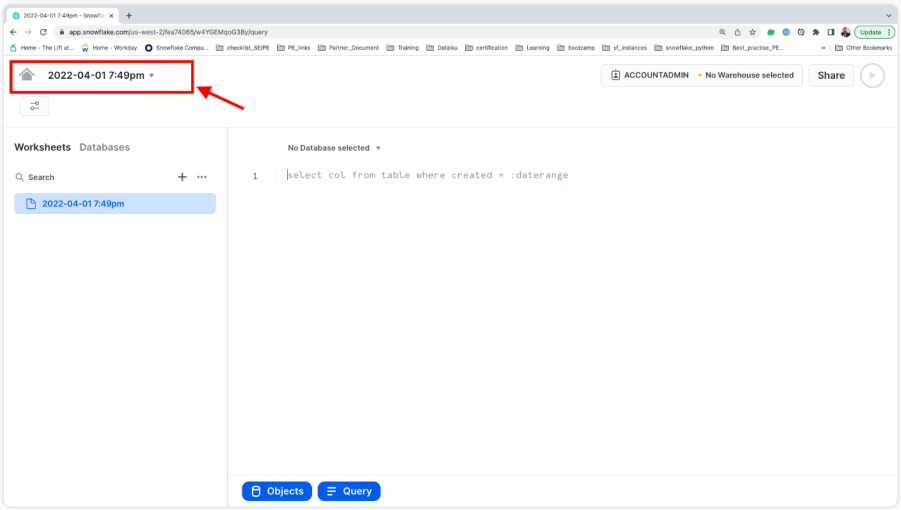


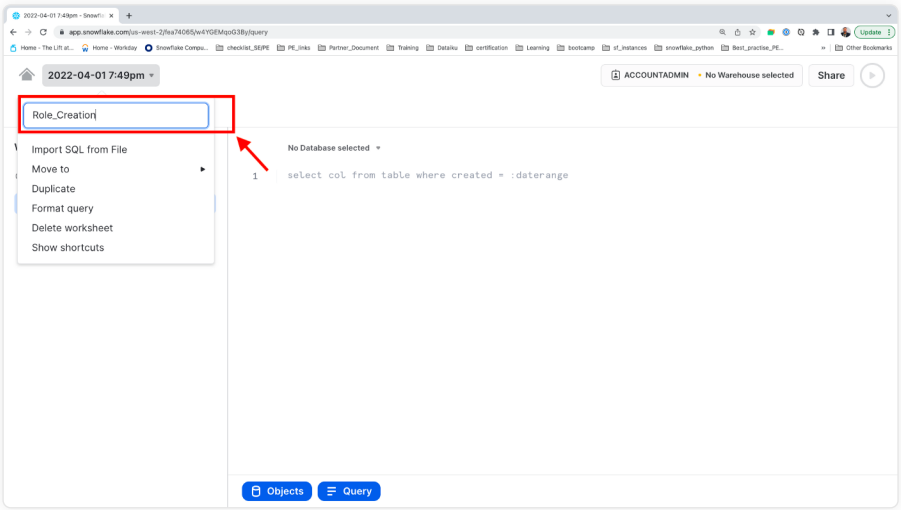


Once you've logged in, open a new Worksheet.









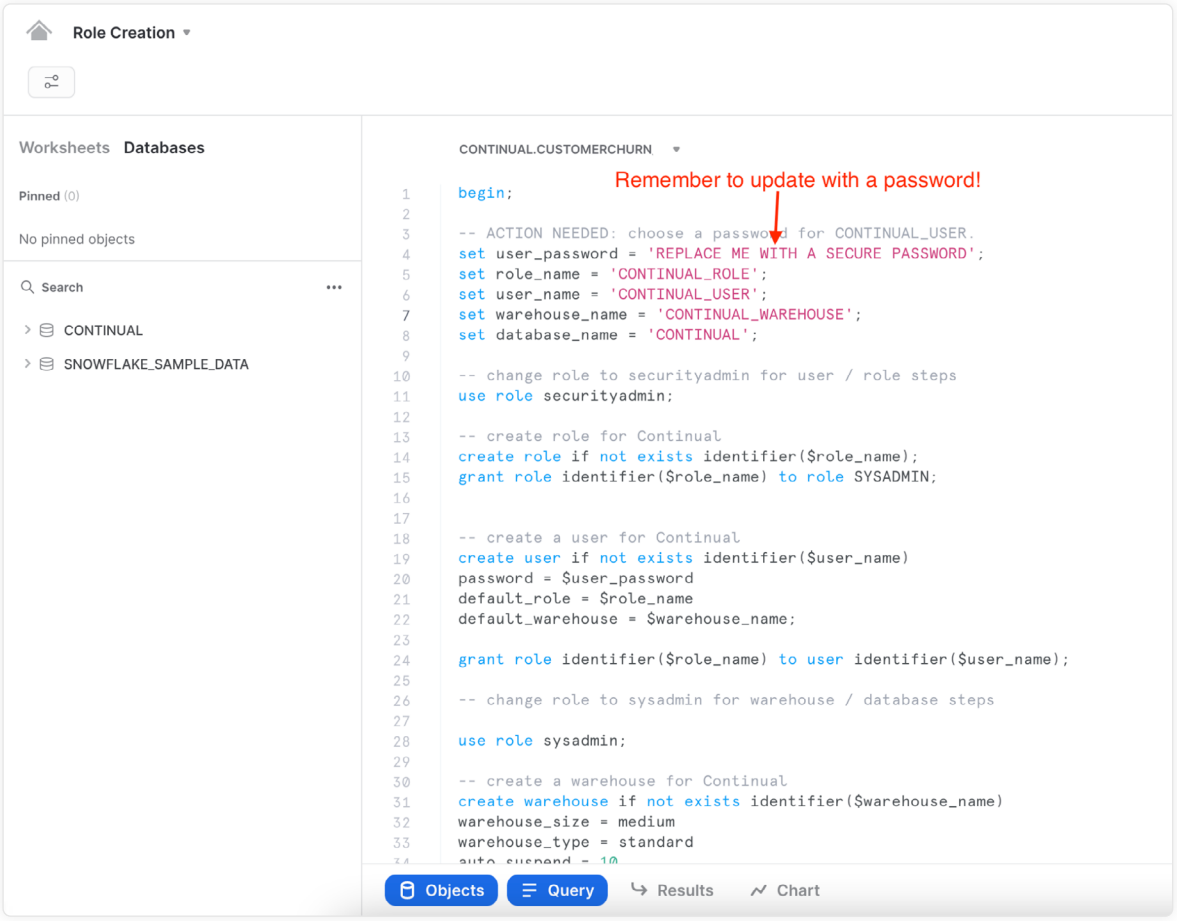
Let's create a user role, role, user, warehouse and database for use by Continual.

In Worksheets, copy and paste the following SQL into your worksheet. Make sure to update the user_password.

```
begin;

-- ACTION NEEDED: choose a password for CONTINUAL_USER.

set user_password = 'REPLACE ME WITH A SECURE PASSWORD';

set role_name = 'CONTINUAL_ROLE';

set user_name = 'CONTINUAL_USER';

set warehouse_name = 'CONTINUAL_WAREHOUSE';

set database_name = 'CONTINUAL';


-- change role to securityadmin for user / role steps
```

```
use role securityadmin;

-- create role for Continual
create role if not exists identifier($role_name);
grant role identifier($role_name) to role SYSADMIN;

-- create a user for Continual
create user if not exists identifier($user_name)
password = $user_password
default_role = $role_name
default_warehouse = $warehouse_name;

grant role identifier($role_name) to user identifier($user_name);

-- change role to sysadmin for warehouse / database steps

use role sysadmin;

-- create a warehouse for Continual
create warehouse if not exists identifier($warehouse_name)
warehouse_size = medium
warehouse_type = standard
auto_suspend = 10
auto_resume = true
initially_suspended = true;

-- create database for Continual
create database if not exists identifier($database_name);

-- grant Continual role access to warehouse
grant USAGE
```

```
on warehouse identifier($warehouse_name)

to role identifier($role_name);


-- grant Continual access to database

grant CREATE SCHEMA, MONITOR, USAGE

on database identifier($database_name)

to role identifier($role_name);


Commit;
```



In this tutorial, we will not use other databases/schemas/tables as source tables for feature sets or models. But for an actual use case, you will need to grant the continual user created above USAGE permission on any such resources. See our docs for more information.

Setting up Continual

SIGNUP FOR TRIAL ACCOUNT

To get started, navigate to Continual and fill in your user details to register an account. Continual has a free 30-day trial and no credit card is required.

You'll need to verify your email address. If you don't receive a verification email within a few minutes, check your spam folder and email support@continual.ai. If your link expires, you can log back into your account to send a new verification email.

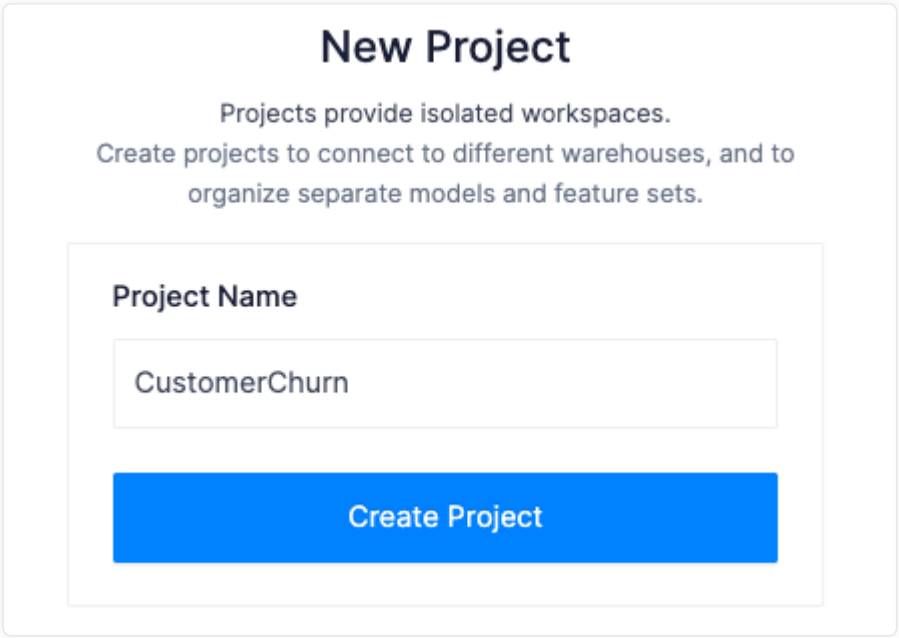
CREATE AN ORGANIZATION

Organizations allow you to share projects within a company and collaborate with team members under a shared billing account.

CREATE PROJECT

After creating your organization you will see your organization's project dashboard with the option to create a project. Projects are isolated workspaces for feature sets and models and connect bi-directionally with Snowflake.

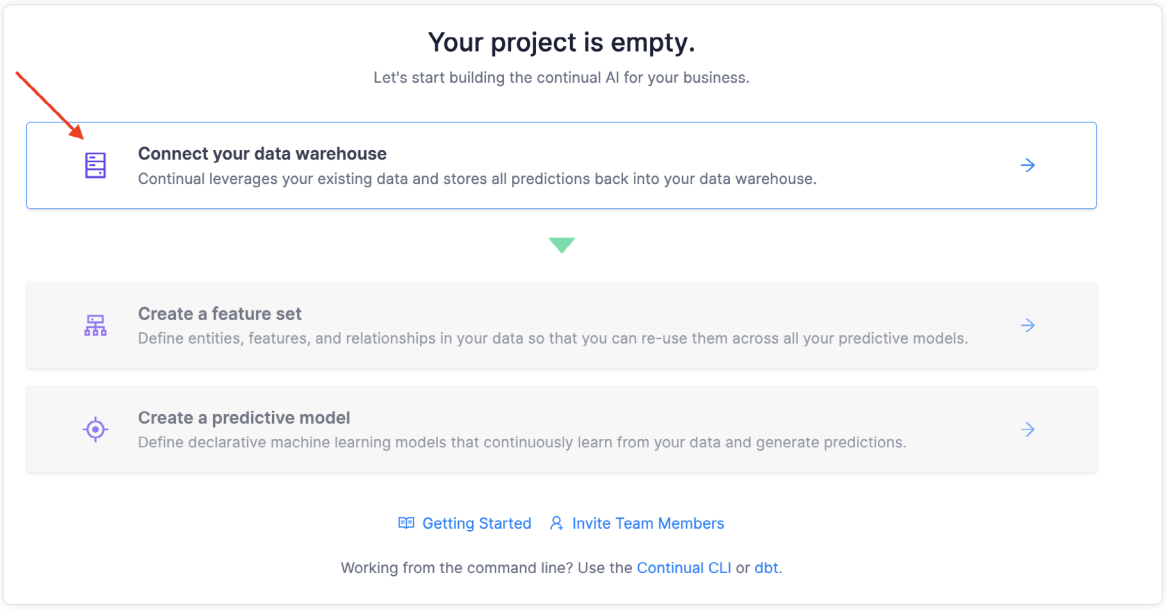
Go ahead and create a new project and name it CustomerChurn

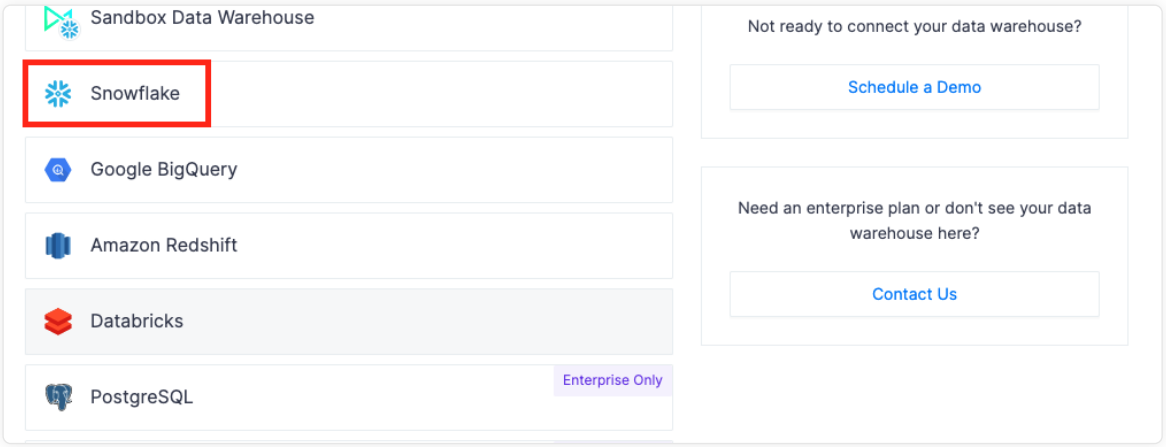


CONNECT TO SNOWFLAKE

Continual was designed for cloud data warehouses and, consequently, connectivity is simple. Each Continual project connects bi-directionally to one Snowflake Database. Continual maintains tables and views for all your feature sets and models, as well as all predictions made by your models, inside a schema. This makes it easy to build models from your existing data and consume the predictions Continual maintains using your existing tools in Snowflake!

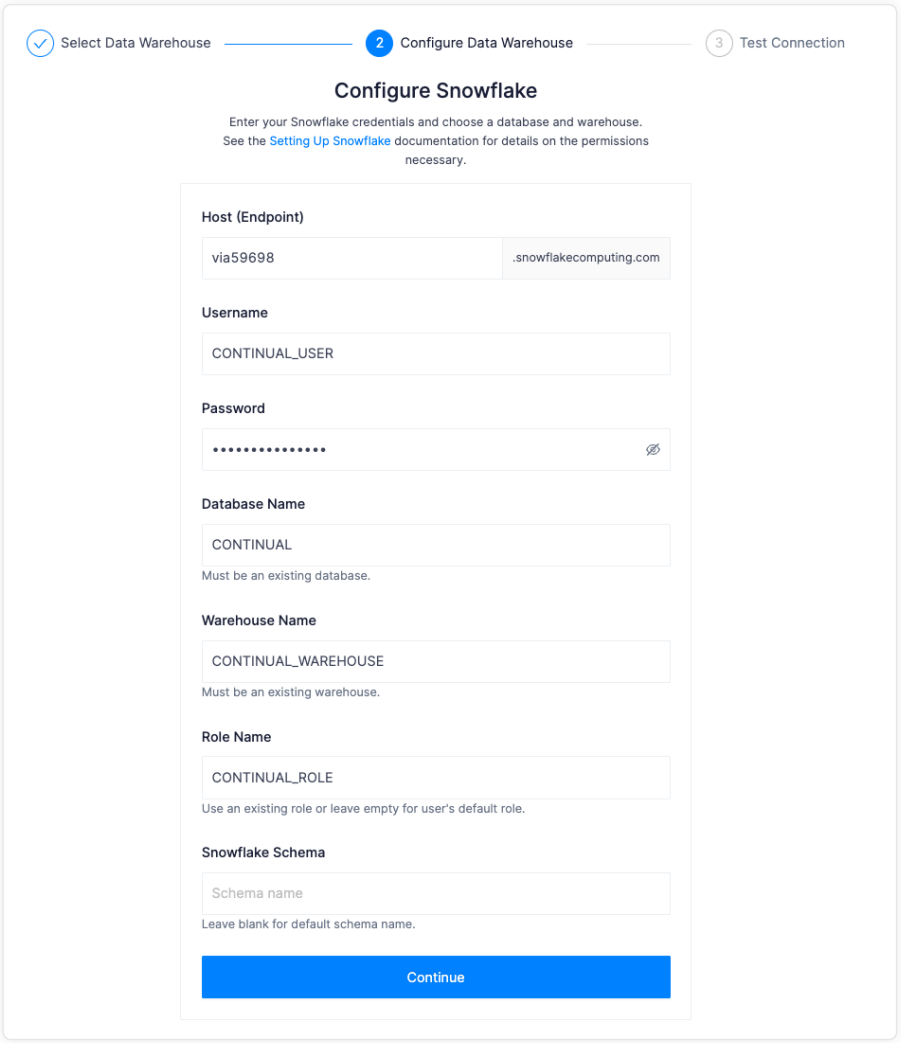
Click “Connect your data warehouse” and then select “Snowflake”



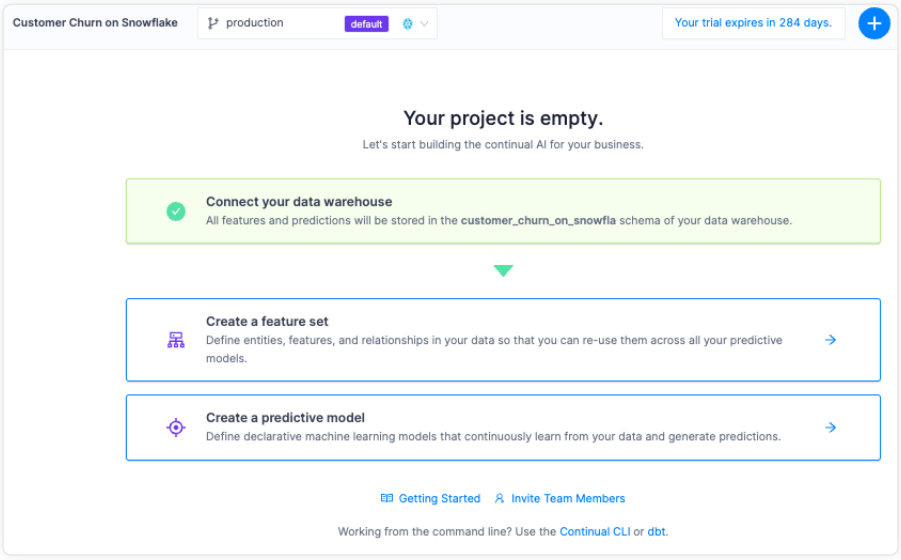


Enter your snowflake account identifier, username, password, database name, warehouse name, and role. Leave the schema field blank.

NOTE: The Host (Endpoint) is the Snowflake account identifier. If you selected a region other than US West (Oregon) you need additional segments depending on the region.



Test the connection and then create it. And there we have it: Continual and Snowflake are connected!

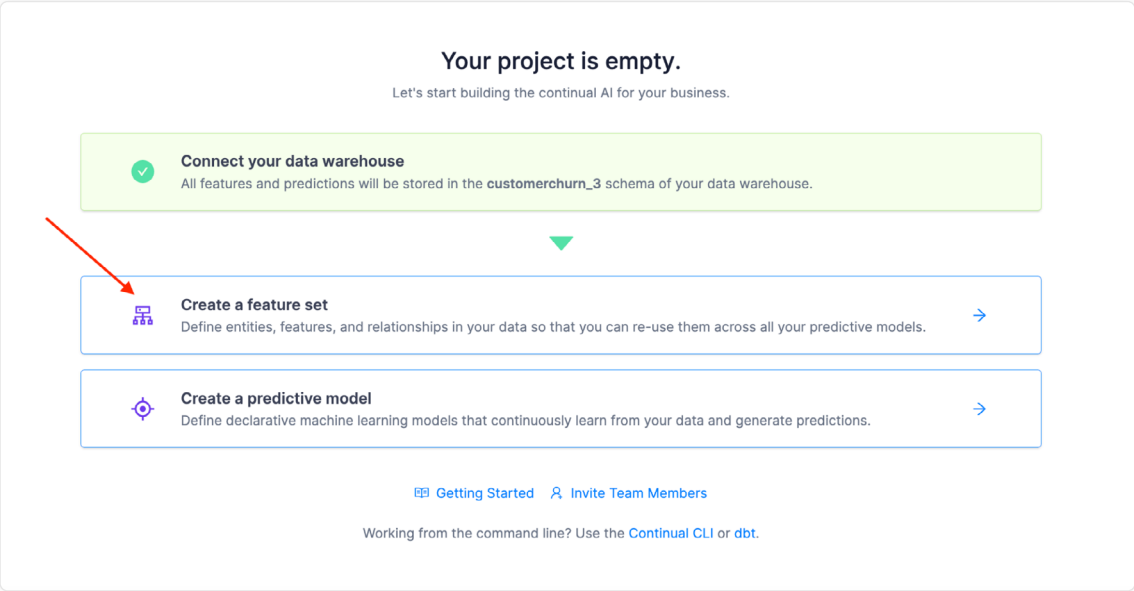


Create a feature set with SQL

Now that we've established our connection and can access our data in Snowflake, it's time to prepare features for a model.

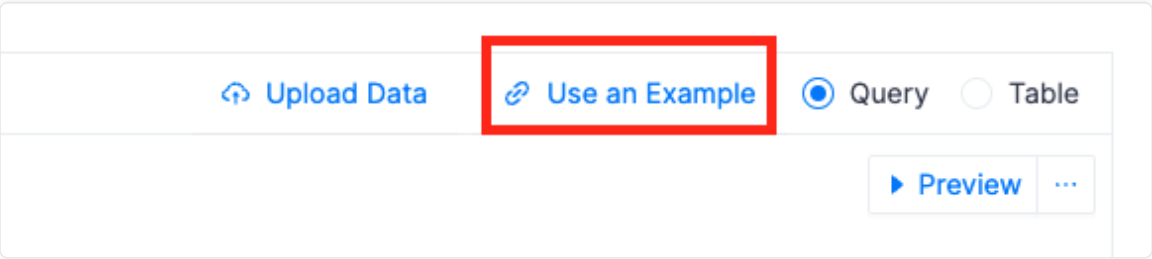
A feature set is one of the main objects in Continual. It describes a collection of related features and the data underlying those features. You can think about it as a view or table of your data warehouse that organizes the data in a way that is easiest for the machine learning model to understand. Just as we'll do when creating a model, we use SQL to query the data and a YAML file to define metadata.

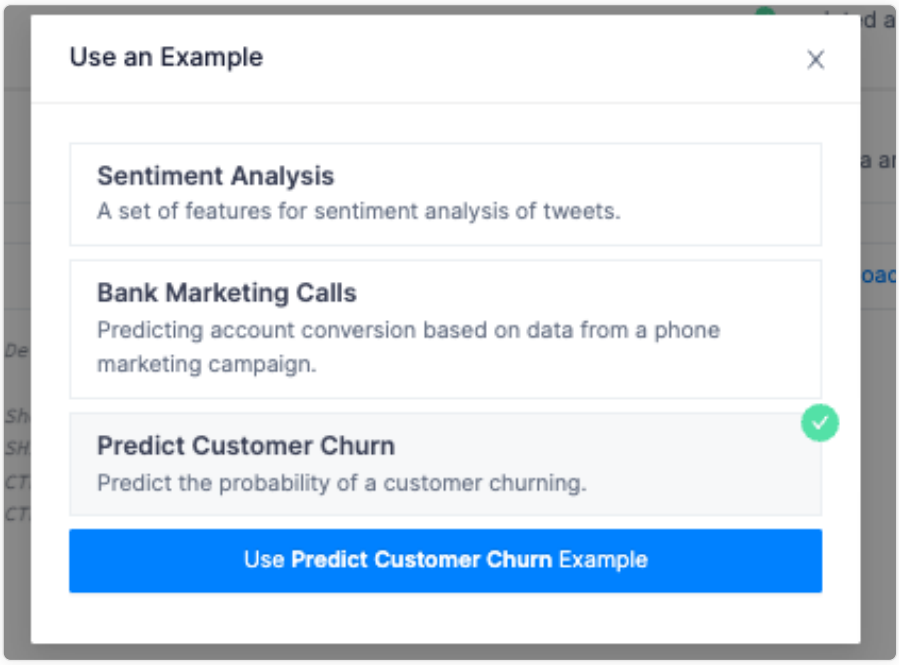
Click "Create a feature set"



The Query Data step is where we use SQL to select the data for our feature set. To make it easy, we have an example ready to go that will copy a csv from an object store into your Snowflake database and pre-populate the query editor, configurations and metadata, and schema. You are living the good life!

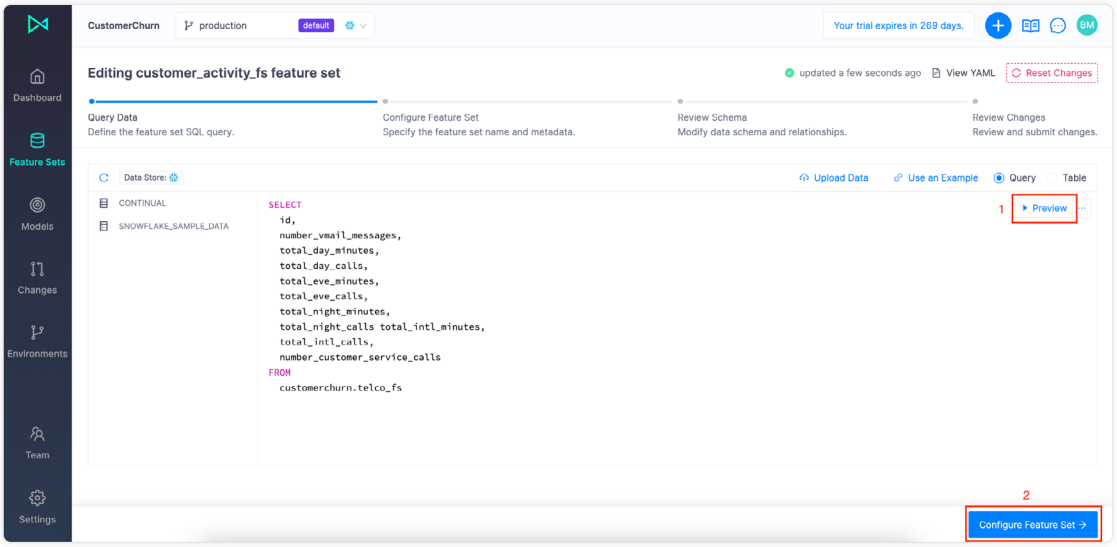
Click "Use an Example" on the right-hand side and select "Predict Customer Churn"





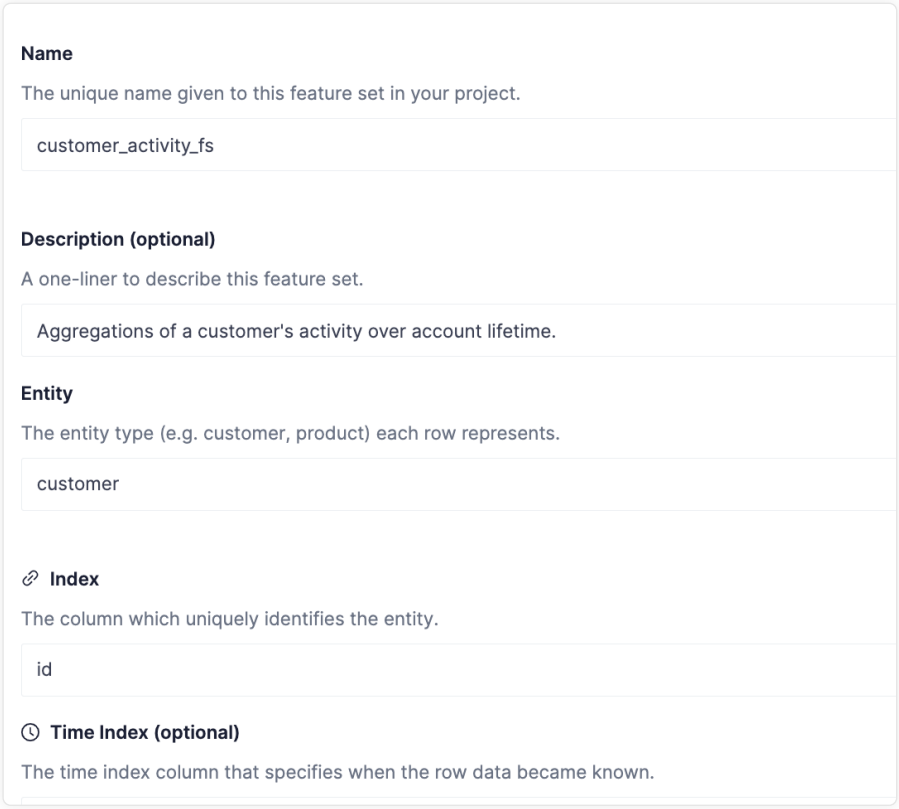
Preview the data to verify the query is selecting the data required for the feature set.

Then select Configure Feature Set on the bottom right to advance to the next step.



The Configure Feature Set step is where you add all the metadata to the featureset: name, description, entity, and index. An entity is a higher level object that combines feature sets that represent common business objects such as “customers”, “products”, and “sales”. The index is what uniquely identifies the feature set and connects it to an entity. All feature sets in an entity have the same index.

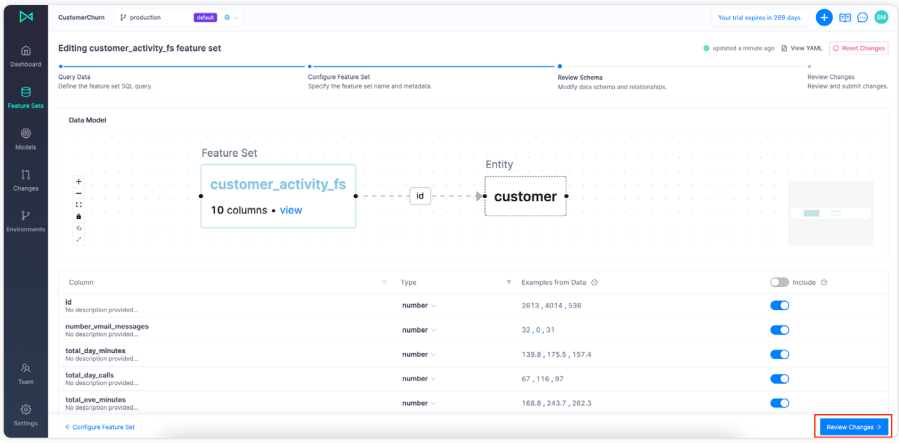
Populate the fields as shown below and create a new entity called "customer".



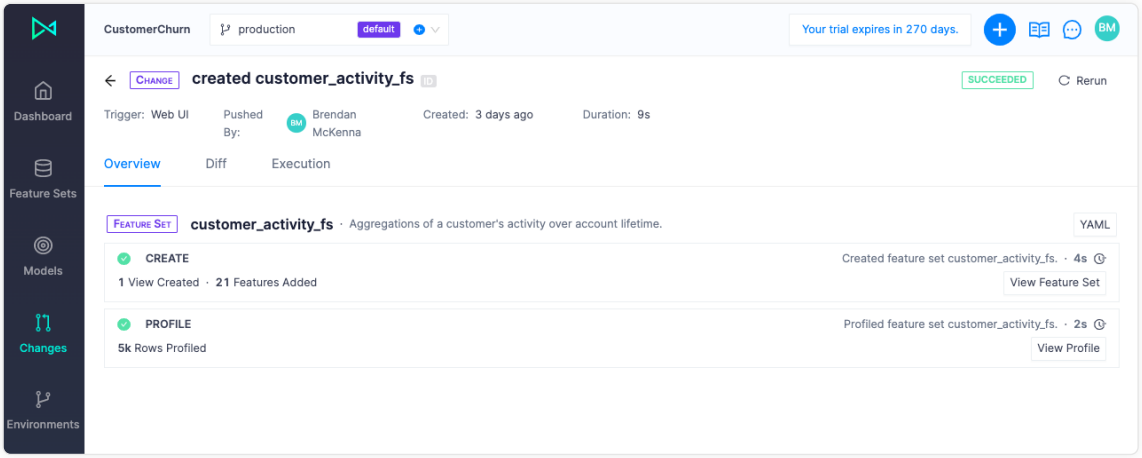
Click "Define Schema" to advance to the next step.

Notice our feature set is displayed in the Data Model graph, with all the columns, their data types, and whether they are included in this feature set.

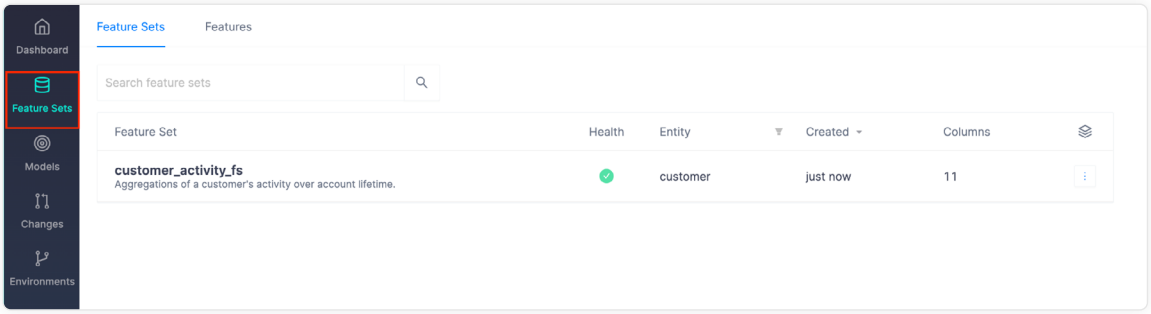
Okay, time to review and create! Click "Review Changes" and then "Submit Changes"



Now, click on the "Changes" tab on the left hand side to see the action added to the activity feed.



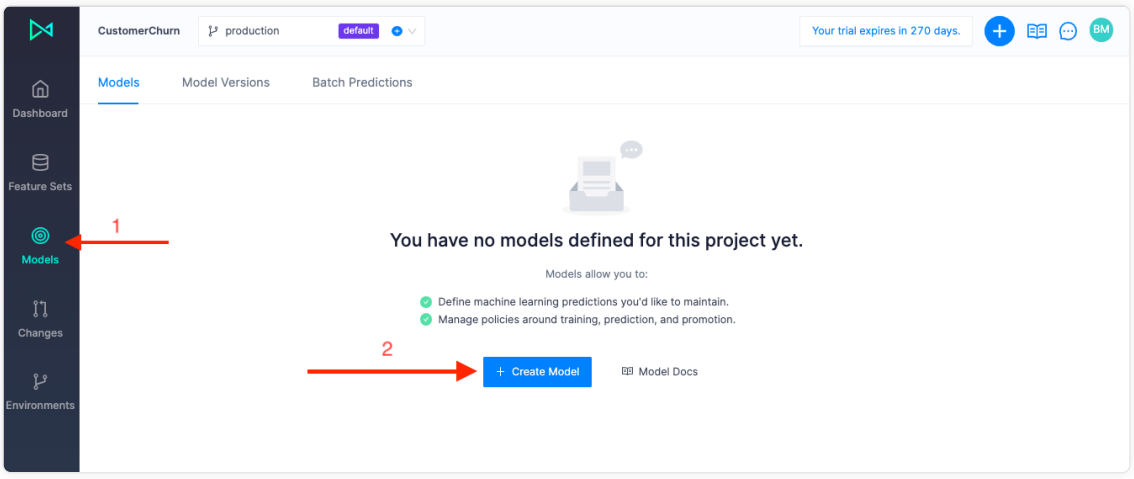
Once the Feature Set has been created, we can see it listed on "Feature Sets" on the left vertical menu:



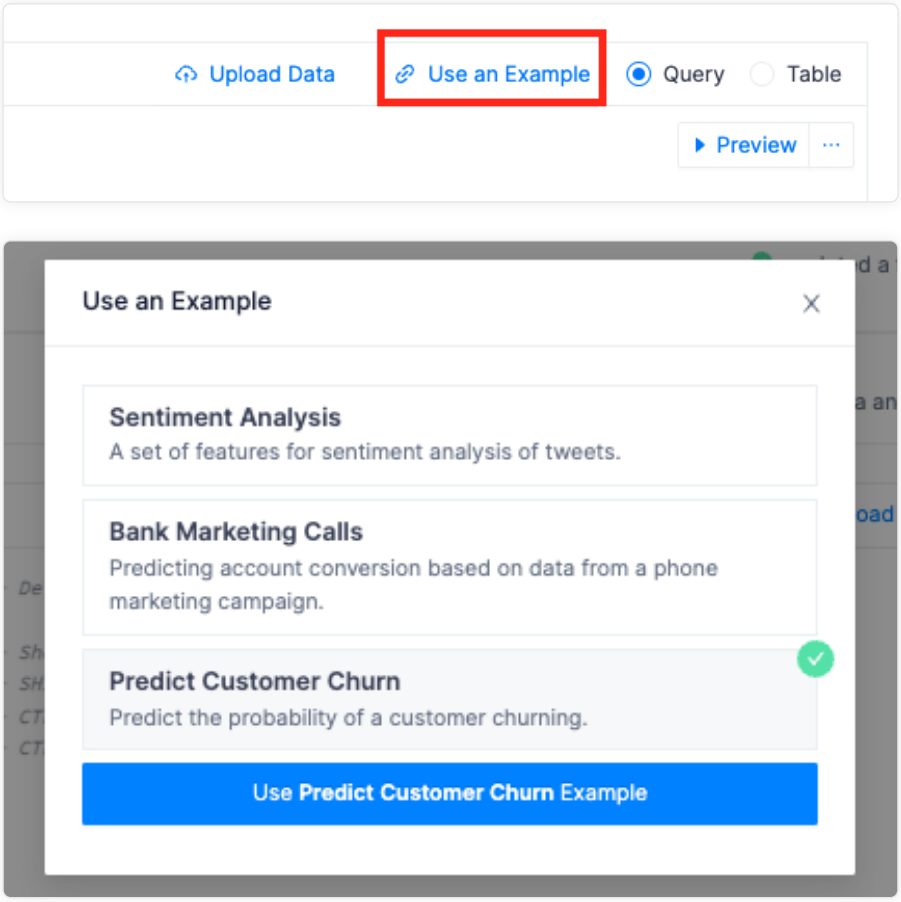
Create a model

We've connected to Snowflake and created a feature set for a model. Now it's time to create a model that we will use our feature set and some additional data to predict the probability of a customer churning. The flow is very similar to creating a feature set except with some key additions.

At the configuration step, we'll need to provide a target column to train our model against. Then we need to set policies for re-training, promotion, and running predictions. Click on "Models" on the left hand side, then "Create Model"



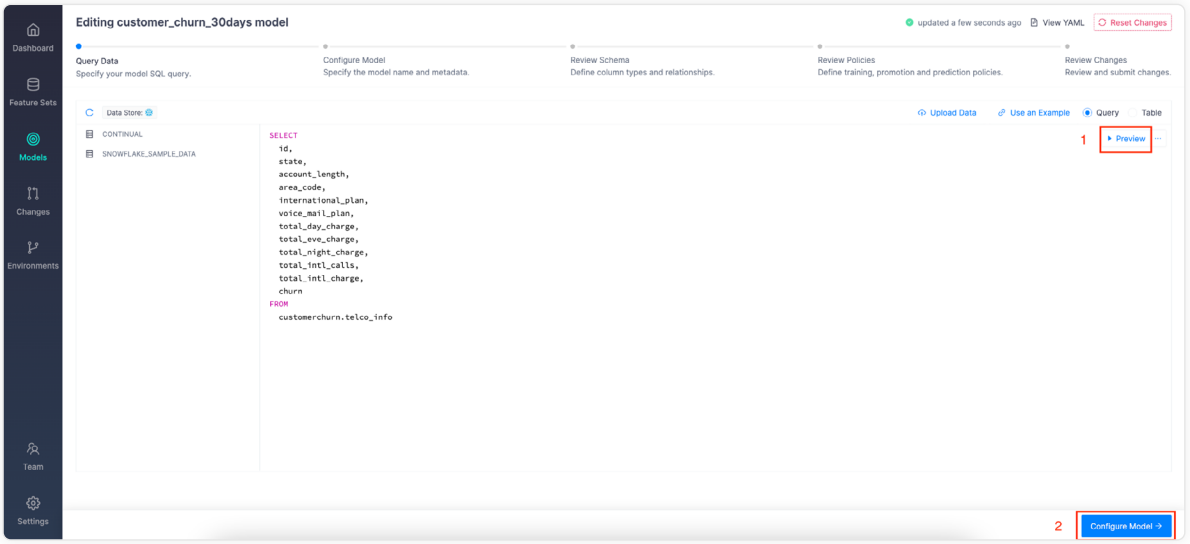
Click "Use an example" and then select "Predict Customer Churn".



We need to make sure our SQL query contains a unique index, features, and a target. In addition to new features we'll define in our model spine, we want to include the feature set we previously built. We do this by including the index column of our feature set in our query and then linking it to our "customers" entity in the "Review Schema" step. Then, at model training time, Continual will join the feature set with the model to create the training data set.

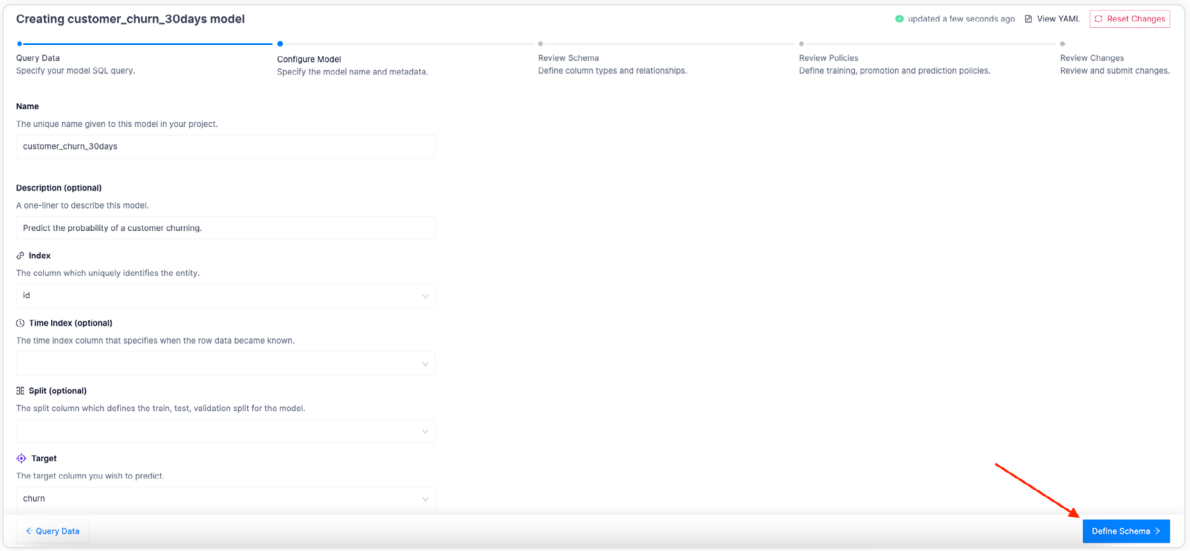
We typically recommend storing your features in feature sets and connecting your models to them via entity linking, but it's also possible to specify a list of columns in your model that represent additional features to bring into the model.

Click "Configure Model"

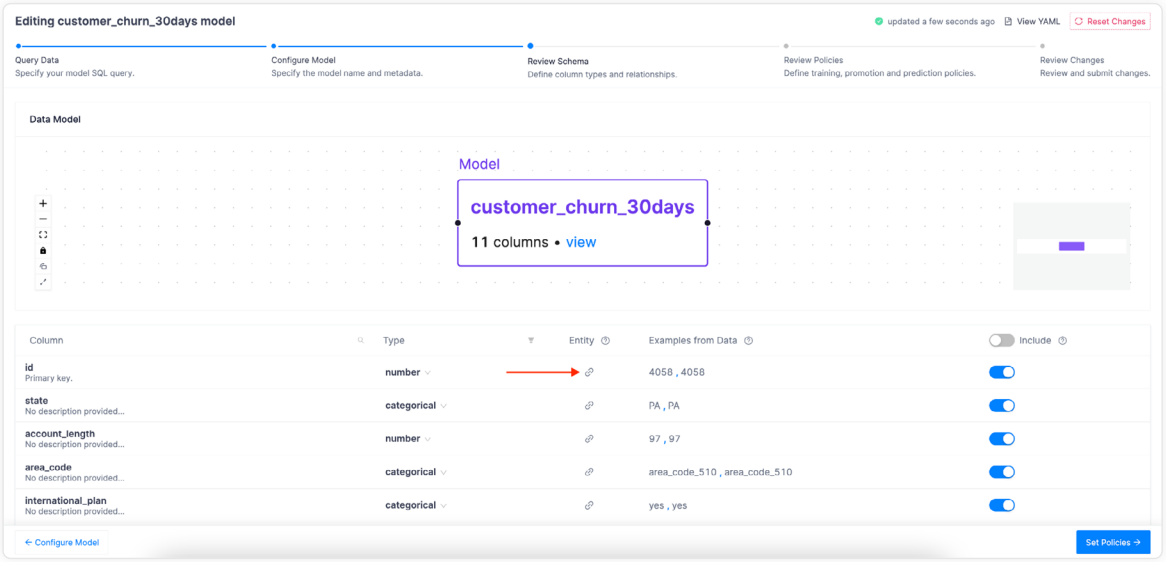


Cool, so let's give our model a name and description and define our model index and target column. These attributes, along with a sql query that generates the data and linked entities, forms the core of a model definition, and this is sometimes referred to as the model spine.

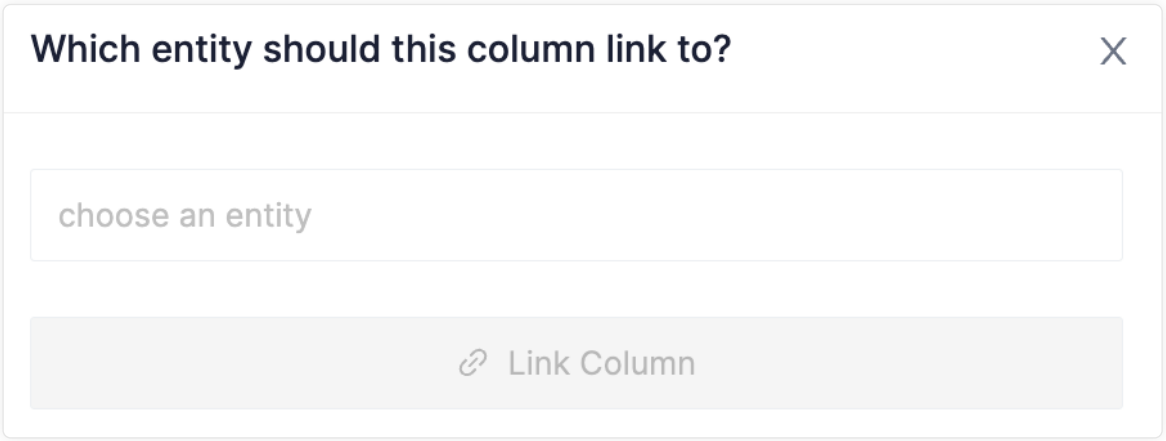
Click "Define Schema"



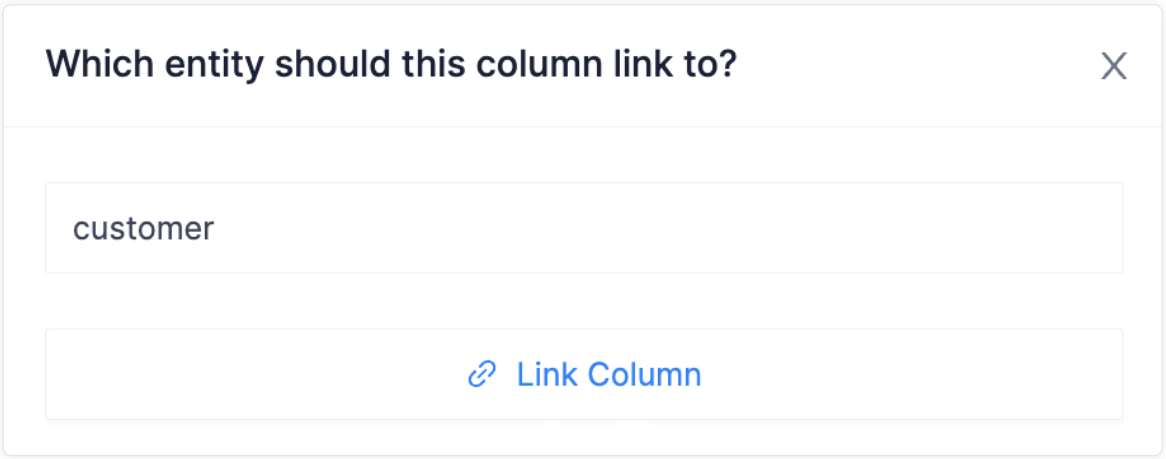
Now it's time to link our feature set index to our "customer" entity. Click the chain icon on the "id" row and then select "customer".



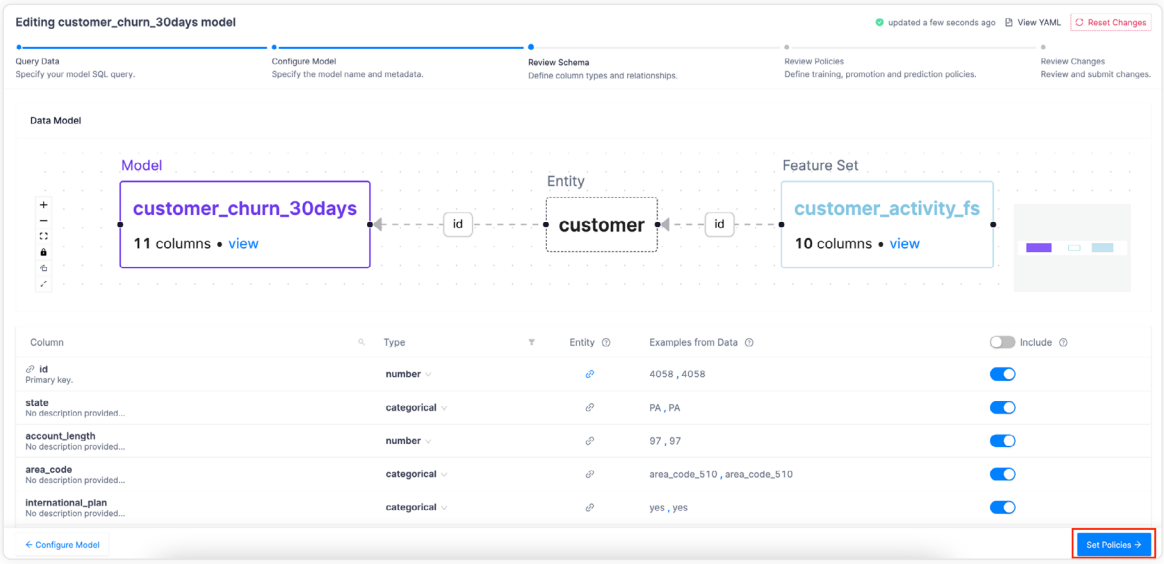
Type "customer" into the pop up box:



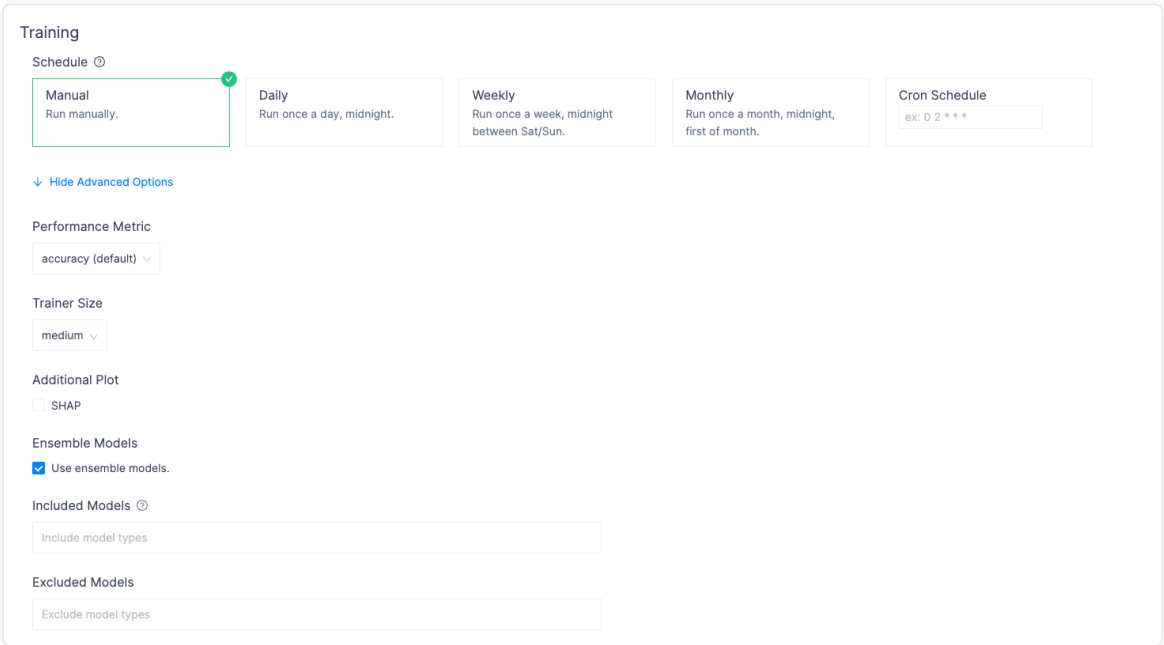
Then click "Link Column"



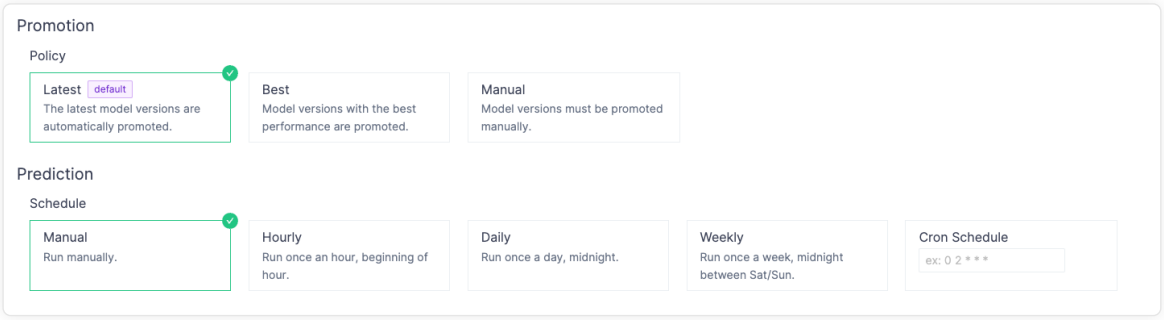
Click "Set Policies"



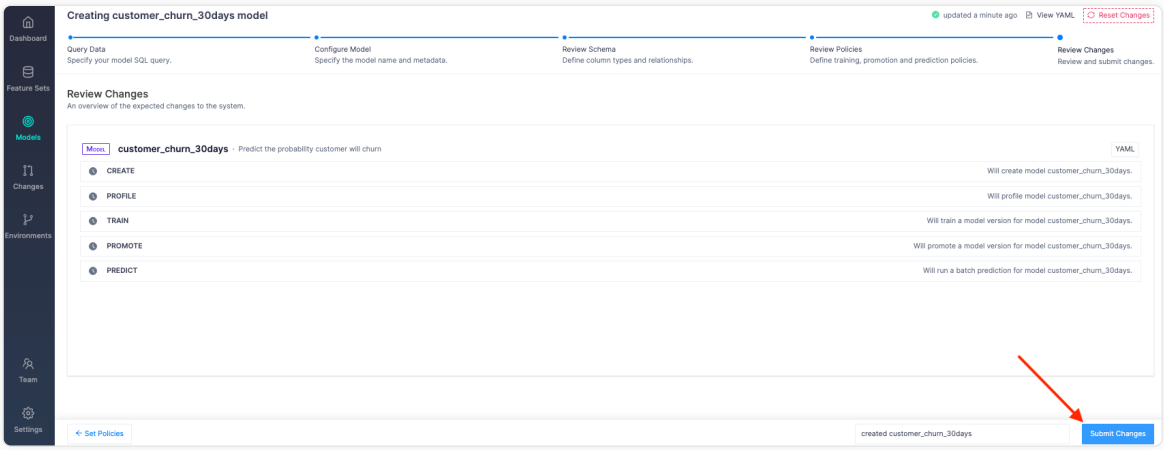
In Continual, you can configure recurring training schedules to ensure your model is updating as frequently as it needs to. You can also set advanced settings such as which performance metric to optimize for, the size of the container, and even which models to include or exclude in the experiment. While automated, Continual allows you to have control over how your model is created, optimized, deployed, and managed.



You can also set how the system chooses which model to promote to production and when new predictions should be made.

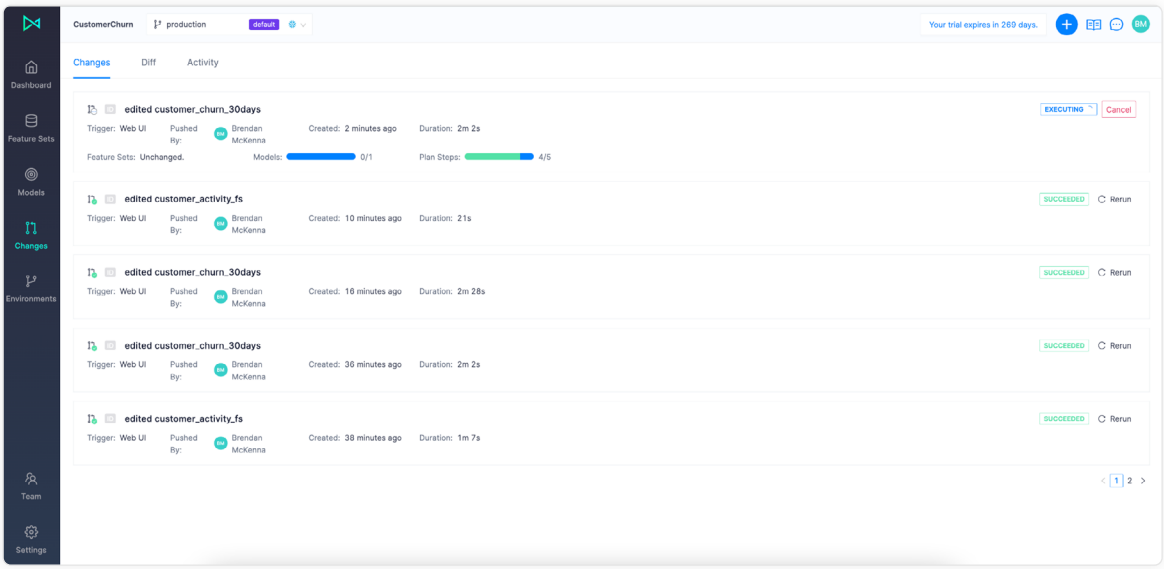


Go ahead and create the model by clicking "Submit Changes"



Well done! How easy was that?

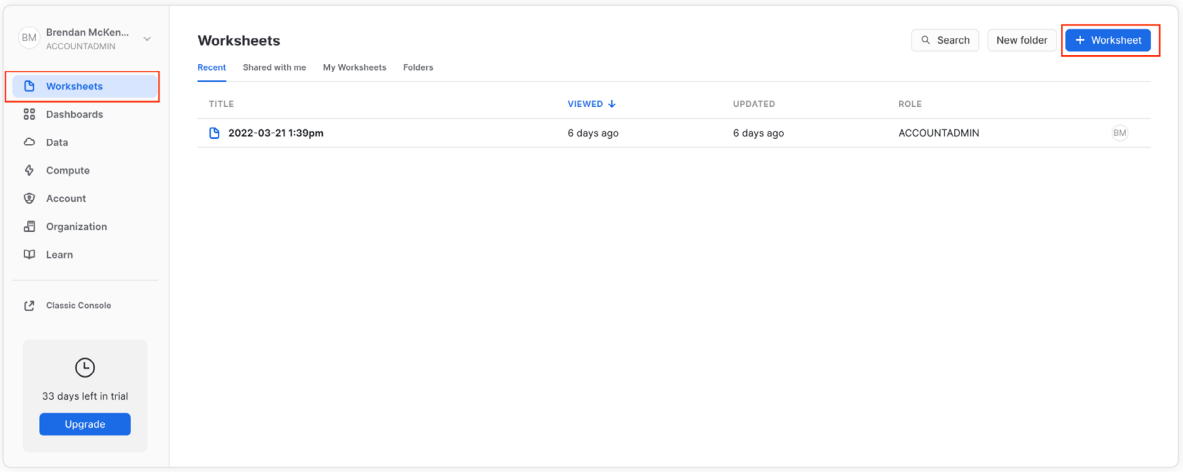
All changes you make in Continual, such as creating a new feature set or editing/updating an existing model, is listed in the "Changes" tab. This gives you a lineage of your team's work you can reference at any time.

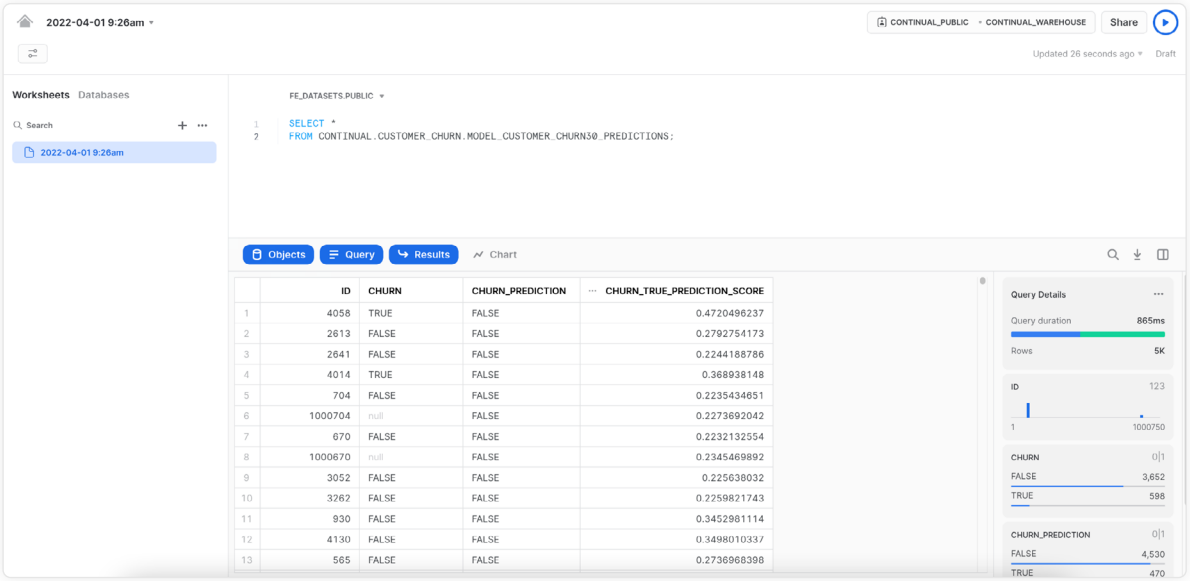


Once your model has been created and promoted it will write predictions directly back to Snowflake. Continual creates a table in your feature store for every model you create in the system that tracks all predictions made by model versions in that model over time. This table lives under `<feature_store>.<project_id>.model_<model_id>_predictions_history`. Continual additionally builds a view under `<feature_store>.<project_id>.model_<model_id>_predictions` which represents the latest prediction made for each record in your model spine.

Let's use the latest predictions view. In Snowflake, paste the following sql statement in to view all your predictions:

```
SELECT *

FROM continual.customerchurn.model_customer_churn_30days_predictions;
```

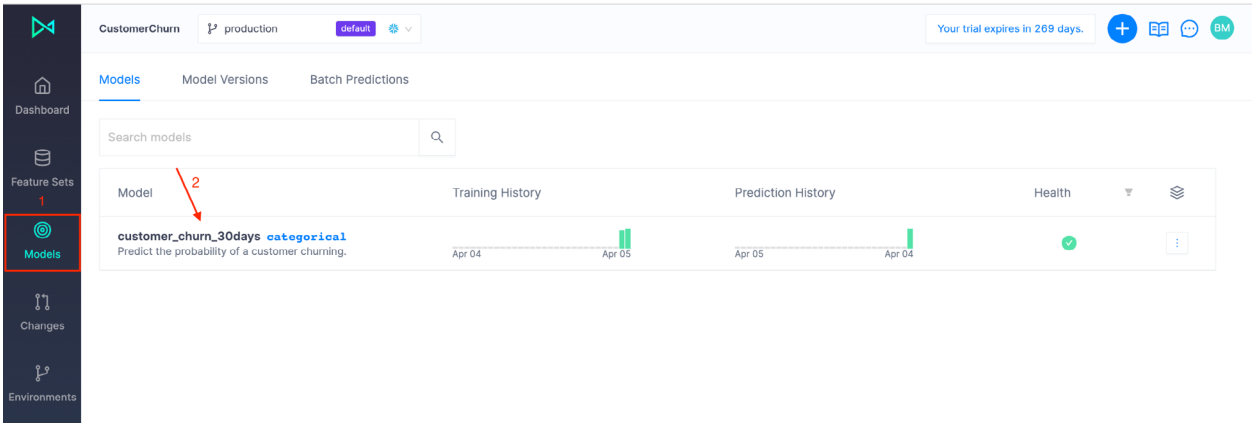




MLOps: Monitoring data and models

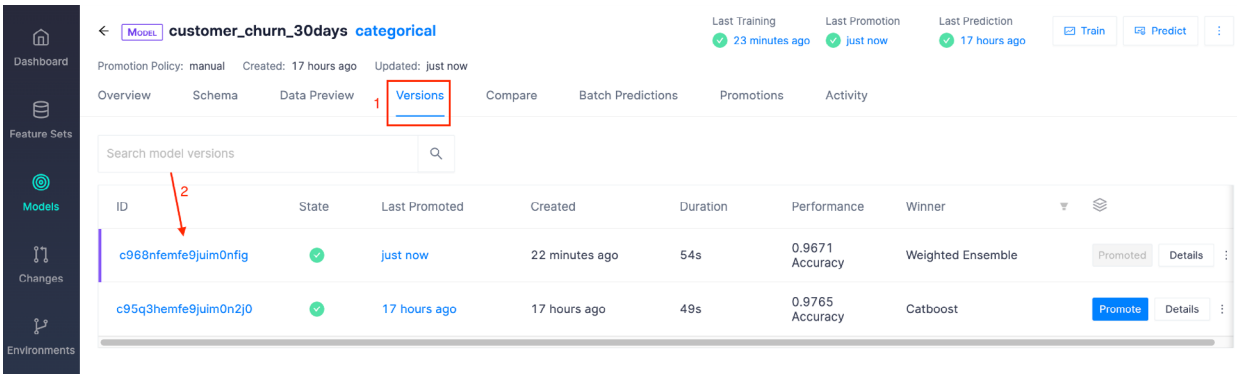
Back in Continual, there are many tools for monitoring your data, models, and prediction jobs.

Navigate to "Models" and select the customer_churn_30days model



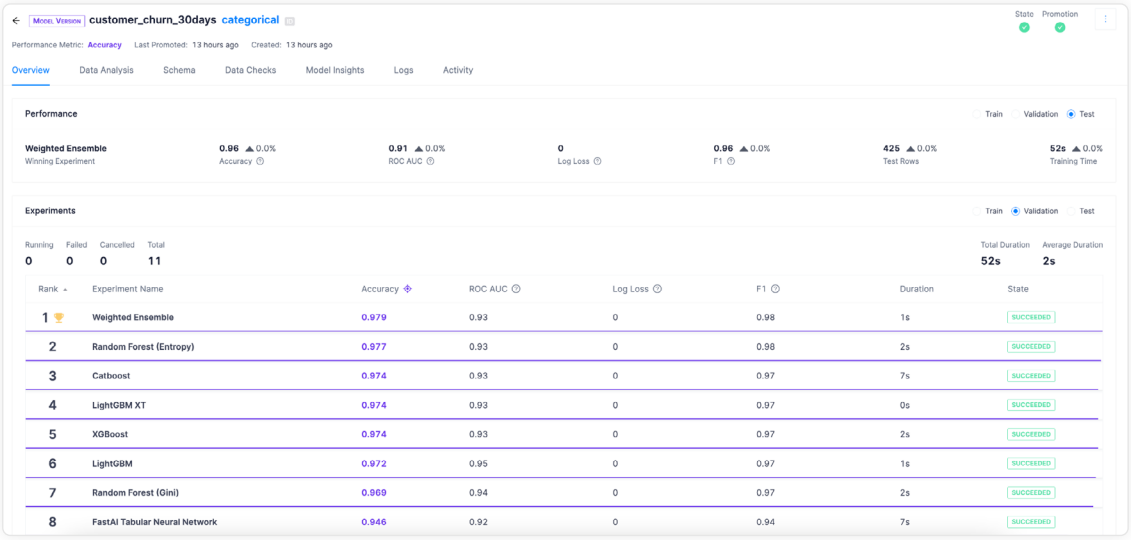
Each time you train a model, a new version is produced and managed under the "Model Version" view.

Click "Versions" and choose a Model Version to evaluate.



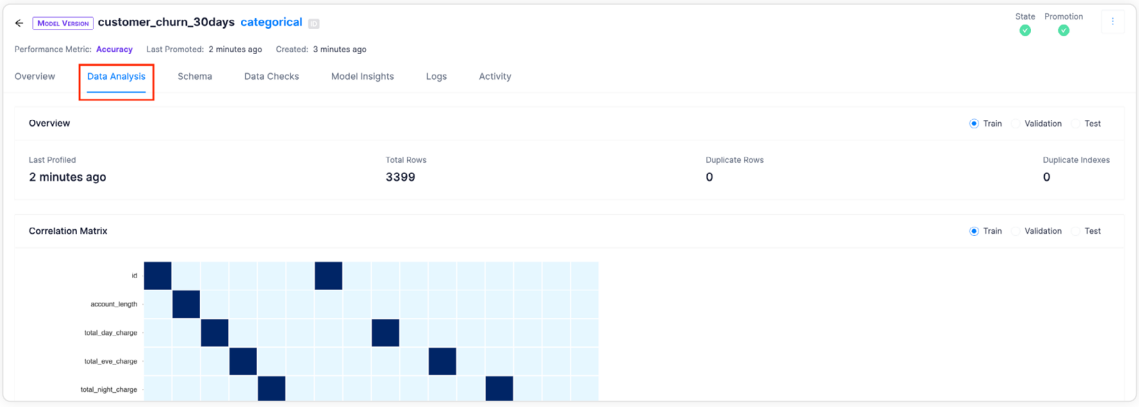
Performance Analysis

The "Overview" page shows the performance of the winning model, as well as each model that was tested. Continual runs a series of experiments across different model algorithms and optimizes performance across a specified performance metric.

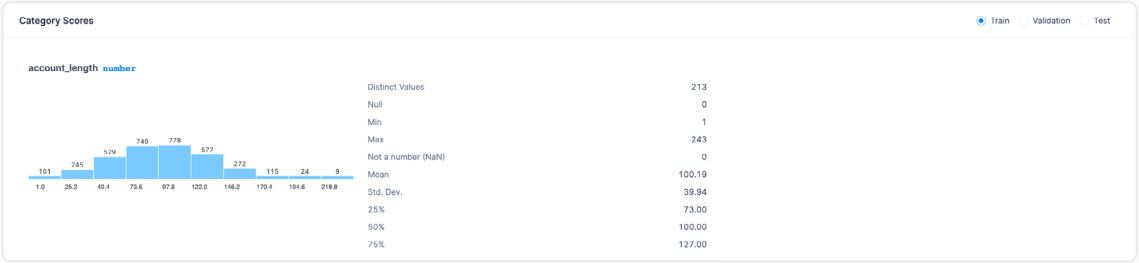


Monitoring data

Click on "Data Analysis" to look closer at the data used to train the model.

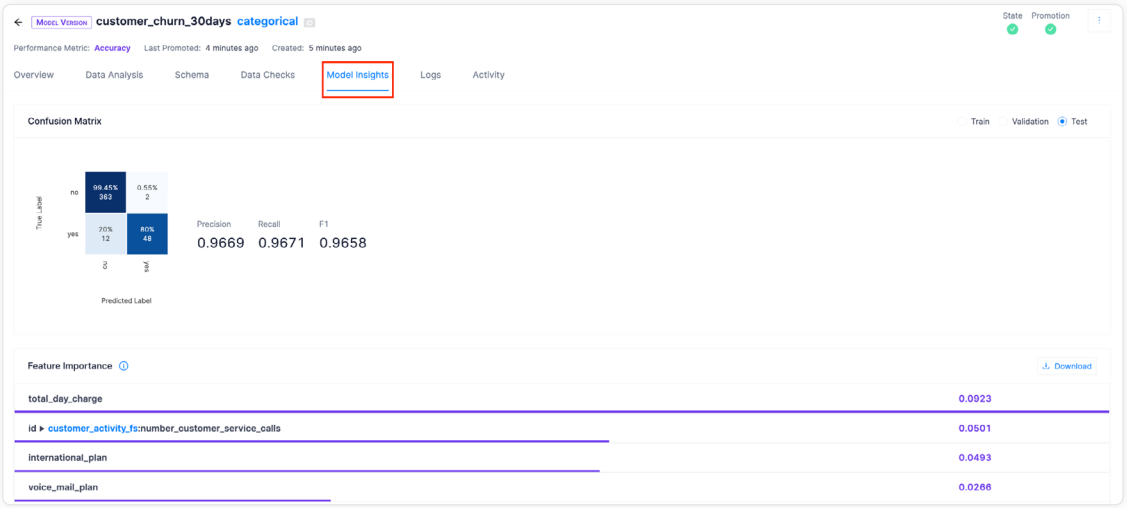


Here you can look at the correlation matrix to see which two variables are correlated and category scores to look at each feature's profile to check if there are features with many Null values, large outliers, or unexpected distributions.



Analyzing the model

Click on "Model Insights" and look at the confusion matrix to understand what your model is getting right and what types of errors it's making.



We can also reference Feature Importance to view which features were the most impactful.

Conclusion

Just like that you've enabled machine learning on Snowflake. Continual is the AI layer for the modern data stack and designed with the shared principles of simplicity, minimal management overhead, and elasticity.

In less than 15 minutes, we connected Continual to Snowflake, created a feature set, used it as input to experiment among more than 10 models and added other relevant features, promoted the best performing model to production, wrote prediction results back to Snowflake, analyzed our features and model performance to learn what improvements we can make. We did this all in the UI but could've used the CLI or SDK.

This concludes the guide to quickly getting started with Continual on Snowflake. Now you're ready for a more advanced example of predicting customer churn with Continual and dbt on Snowflake. We hope you'll dive in, and if you need a little help or have questions along the way, book some time with one of our AI experts.

ABOUT SNOWFLAKE

Snowflake delivers the Data Cloud—a global network where thousands of organizations mobilize data with near-unlimited scale, concurrency, and performance. Inside the Data Cloud, organizations unite their siloed data, easily discover and securely share governed data, and execute diverse analytic workloads. Wherever data or users live, Snowflake delivers a single and seamless experience across multiple public clouds. Snowflake's platform is the engine that powers and provides access to the Data Cloud, creating a solution for data warehousing, data lakes, data engineering, data science, data application development, and data sharing. Join Snowflake customers, partners, and data providers already taking their businesses to new frontiers in the Data Cloud

ABOUT CONTINUAL

Continual is the leading enterprise AI platform for the modern data stack, providing end-to-end automation for operational AI workflows on Snowflake. Customers use Continual to deploy continually improving predictive models to drive revenue, operate more efficiently, and build innovative products and services. Find out more at continual.ai.