



# Enhance Machine Learning Workflows: Snowflake Feature Store & Model Registry

Learn How to Enhance Machine Learning Workflows with Snowflake's Feature Store and Model Registry: Improve Data Consistency, Boost Model Deployment Efficiency, and Gain Insights Using Customer Churn Prediction

evolv Consulting · [Follow](#)

8 min read · Oct 7, 2024



51



1



...

The evolv logo is located in the top right corner of the header. It consists of the word "evolv" in a lowercase, sans-serif font, with the "e" and "o" in orange, and the "v" and "l" in light blue.

# Enhance Your ML Workflows with Snowflake

## with Feature Store & Model Registry



blog written by

A stylized signature of the author's name, "Rose Ellison", in red and orange.

Senior ML Consultant, evolv

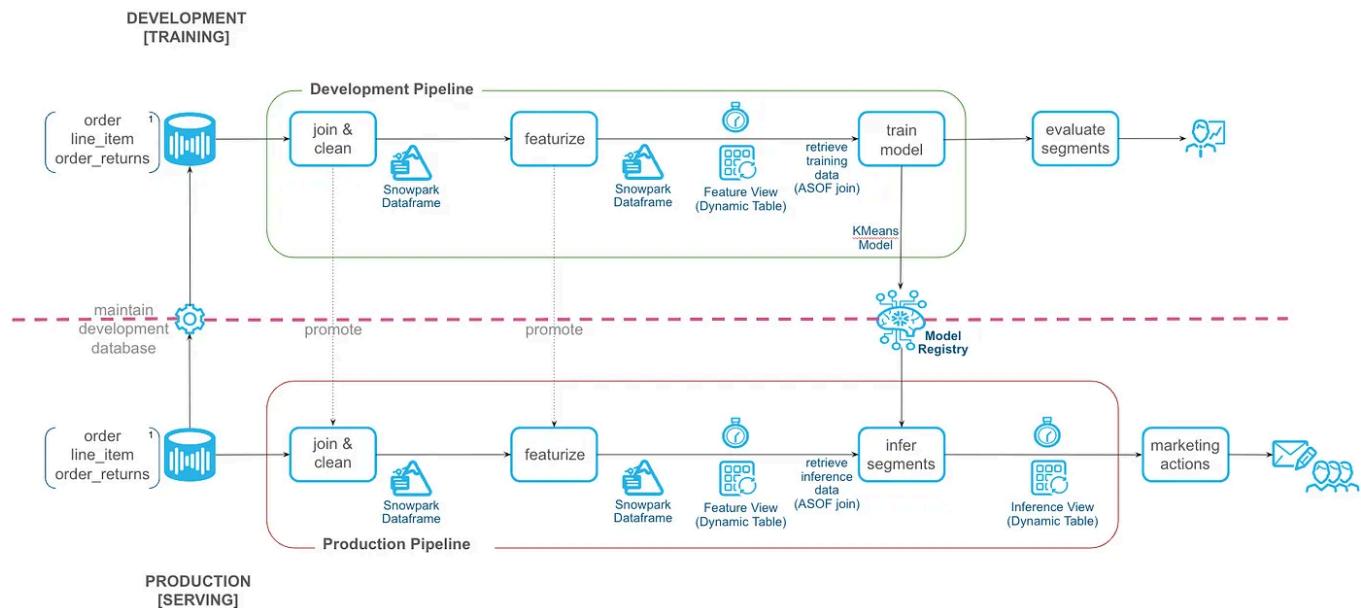
## Introduction

Predicting customer churn — the probability of customers ending their service — is a crucial challenge for many businesses. In financial services, churn can lead to substantial revenue loss and diminish customer trust. In healthcare, it impacts patient retention and continuity of care, while manufacturing firms face heightened operational costs from losing repeat customers. Accurate churn prediction empowers companies to proactively mitigate customer dissatisfaction, customize retention strategies, and ultimately boost profitability. However, reliable predictions require robust machine learning workflows that ensure data consistency, efficient model deployment, and insightful business analytics.

Snowflake's Feature Store and Model Registry offer essential tools to streamline workflows. By centralizing feature management and simplifying model tracking, they empower businesses to effectively address common

ML challenges. This article explores how leveraging these features can enhance your ability to predict customer churn, ensuring your models remain accurate and scalable as your business expands.

To learn more about what's new with Snowflake's ML, [check out this video](#), recorded at SUMMIT 2024.



An end-to-end customer segmentation ML use case diagram using Snowflake Feature Store & Model Registry.

## Leveraging Snowflake's Integrated ML Capabilities

Snowflake ML offers a unified platform for end-to-end machine learning, enabling data scientists and ML engineers to develop and deploy scalable features and models securely. By directly operating on governed data within Snowflake, teams can eliminate data silos and avoid the pitfalls of data movement and governance issues. This integration ensures that your machine learning workflows are both efficient and compliant with data governance standards.

Imagine your company wants to predict which customers are likely to churn. With Snowflake's integrated ML capabilities, you can seamlessly manage the data required for this prediction, ensuring your features are consistent, and your models are deployed efficiently.

**evolv Consulting Expert Tip:** *Utilize Snowflake's secure data sharing features to collaborate with different departments without the need to duplicate data, ensuring data integrity and security across your organization.*

## Key Features:

- **Feature Store:** The Feature Store centralizes feature creation and management, ensuring consistent feature use across models and projects. It acts as a repository where features are stored, versioned, and made accessible for model training and inference. This consistency is crucial for maintaining model accuracy and reliability.
- **EX:** By using the Feature Store, you can define features such as customer tenure, monthly charges, and payment methods once, and reuse them across different churn prediction models. This eliminates discrepancies and ensures that every model is trained on the same feature set.

**evolv Consulting Expert Tip:** *Implement feature versioning within the Feature Store to track changes over time, allowing you to revert to previous feature definitions if needed, and ensuring consistency across different model versions.*

- **Model Registry:** The Model Registry simplifies model tracking and deployment by providing a centralized repository for storing, versioning, and managing machine learning models. It ensures that models are easily accessible, auditable, and seamlessly promoted from development to production environments.

- EX: After training your churn prediction model, the Model Registry allows you to log and version the model. If you improve your model's accuracy, you can register the new version without disrupting the existing deployment, ensuring smooth transitions and maintaining model integrity.

**evolv Consulting Expert Tip:** *Use the Model Registry to enforce model governance by defining life-cycle stages (e.g., development, staging, production) and implementing approval workflows for promoting models to higher stages.*

- **Snowpark ML Python Library:** The Snowpark ML Python Library offers robust APIs for developing and deploying ML pipelines within Snowflake. This library facilitates seamless integration of machine learning tasks into existing workflows, enabling data scientists to leverage Snowflake's powerful data processing capabilities without moving data across platforms.
- EX: Using Snowpark ML, you can build your churn prediction pipeline directly within Snowflake, leveraging its data processing power to handle large datasets efficiently. This reduces latency and improves the performance of your ML workflows.

**evolv Consulting Expert Tip:** *Optimize your Snowpark ML scripts by utilizing Snowflake's scalable compute resources. Adjust your warehouse size based on workload demands to ensure efficient processing without unnecessary costs.*

## Benefits of Adopting Snowflake's Tools

1. **Improved Data Consistency:** A centralized Feature Store ensures all models are built and trained using uniform data definitions, minimizing discrepancies and errors.

**2. Efficient Model Deployment:** The Model Registry accelerates the transition of models from development to production, enabling faster deployment and iteration.

**3. Enhanced Collaboration:** Integrated tools allow teams to collaborate more effectively, sharing features and models across projects and departments.

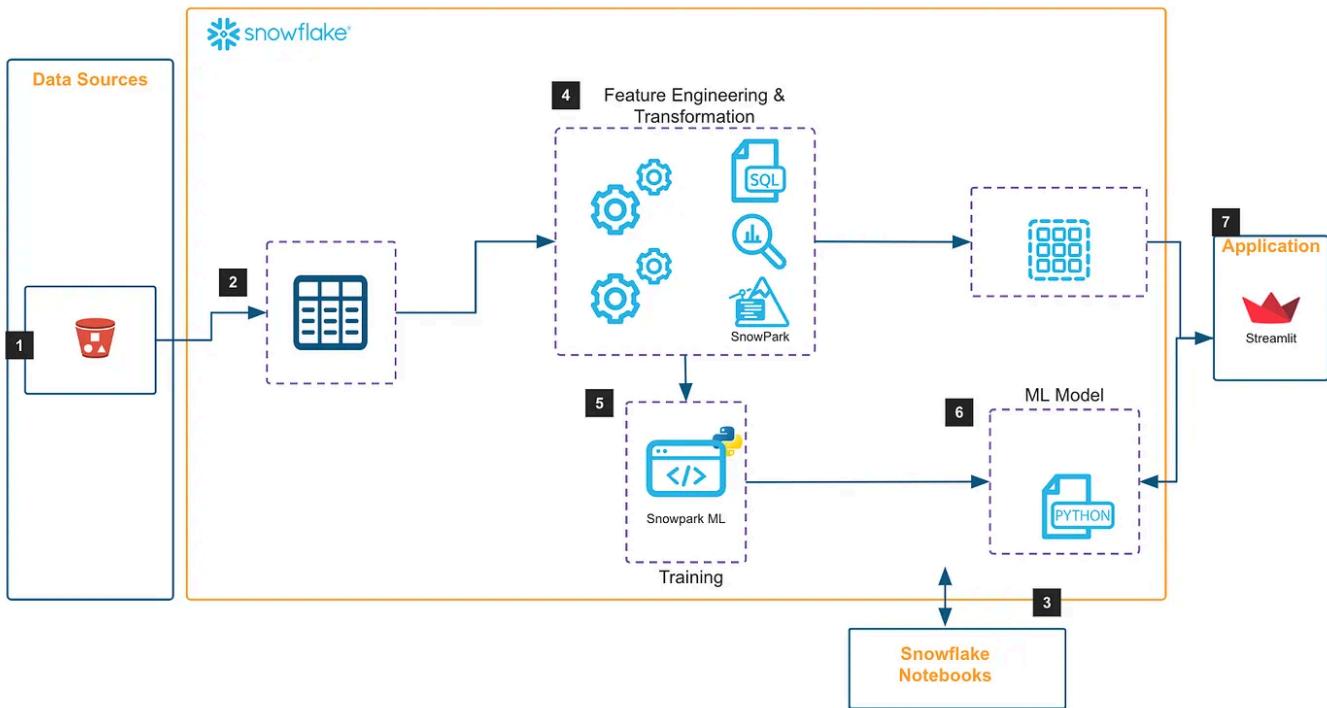
**4. Governance and Security:** Operating within Snowflake's governed environment maintains data compliance and security standards throughout the ML life-cycle.

**5. Scalability:** Snowflake's scalable architecture supports large datasets and complex models, ensuring that your machine learning workflows can grow with your business needs.

**EX:** By ensuring data consistency and efficient model deployment, your churn prediction models remain reliable and can be scaled as your customer base grows. Enhanced collaboration allows different departments, such as marketing and customer service, to access and utilize the churn predictions effectively.

**evolv Consulting Expert Tip:** *Leverage Snowflake's data masking and encryption features to protect sensitive customer information, which ensures compliance with data protection regulations and enables effective churn prediction.*

**Getting Started: Predicting Customer Churn with Snowflake ML**



<https://developers.snowflake.com/solution/data-analysis-and-churn-prediction-using-snowflake-notebooks/>

To demonstrate the power of Snowflake's Feature Store and Model Registry, let's walk through a common business problem: predicting customer churn. We'll use the widely recognized **Telco Customer Churn** dataset, familiar to many data scientists, to illustrate the implementation of Snowflake's tools effectively.

## 1. Initialize the Feature Store: Create a Feature Store client and register entities and feature views to manage your features effectively.

```
from snowflake.snowpark import Session
from snowflake.ml.feature_store import FeatureStore, CreationMode

# Define your connection parameters
connection_parameters = {
    "account": "<YOUR_ACCOUNT>",
    "user": "<YOUR_USER>",
    "password": "<YOUR_PASSWORD>",
    "role": "ML_MODEL_ROLE",
    "warehouse": "ML_MODEL_WH",
```

```
"database": "ML_MODEL_DATABASE",
"schema": "ML_MODEL_SCHEMA"
}

# Create a Snowflake session
session = Session.builder.configs(connection_parameters).create()

# Initialize the Feature Store
fs = FeatureStore(
    session=session,
    database=session.get_current_database(),
    name=session.get_current_schema(),
    default_warehouse=session.get_current_warehouse(),
    creation_mode=CreationMode.CREATE_IF_NOT_EXIST,
)
```

**evolv Consulting Expert Tip:** Utilize environment variables or secure secret management tools to handle your Snowflake connection parameters securely, avoiding hard-coding sensitive information in your scripts.

**2. Register Entities and Feature Views:** Define and register your entities and feature views using the Telco Customer Churn dataset. Entities represent the primary keys for joining data, while feature views define the features used for model training.

```
from snowflake.ml.feature_store import Entity, FeatureView

# Define an entity for customers
customer_entity = Entity(
    name="CUSTOMER",
    join_keys=["customerID"],
    description="Entity representing individual customers"
)

# Register the entity
fs.register_entity(customer_entity)

# Define a feature view for customer features
```

```

customer_feature_view = FeatureView(
    name="CUSTOMER_FEATURES",
    query="""
        SELECT
            customerID,
            tenure,
            MonthlyCharges,
            TotalCharges,
            CASE
                WHEN InternetService = 'DSL' THEN 1
                WHEN InternetService = 'Fiber optic' THEN 2
                ELSE 0
            END AS InternetService_Fiber,
            CASE
                WHEN Contract = 'Month-to-month' THEN 1
                WHEN Contract = 'One year' THEN 2
                WHEN Contract = 'Two year' THEN 3
                ELSE 0
            END AS Contract_Type,
            CASE
                WHEN PaymentMethod = 'Electronic check' THEN 1
                WHEN PaymentMethod = 'Mailed check' THEN 2
                WHEN PaymentMethod = 'Bank transfer (automatic)' THEN 3
                WHEN PaymentMethod = 'Credit card (automatic)' THEN 4
                ELSE 0
            END AS Payment_Method,
            churn
        FROM TELCO_CUSTOMER_CHURN
        """
    ,
    entities=[customer_entity],
    version="1.0",
    description="Features derived from Telco customer churn data",
    refresh_frequency="1 day"
)

# Register the feature view
fs.register_feature_view(customer_feature_view)

```

**3. Generate Training Data:** Create a spine DataFrame from the target data and generate a training dataset using the registered feature views.

```
# Create a spine DataFrame from the target data
spine_df = session.table("TELCO_CUSTOMER_CHURN").select("customerID", "churn")
```

```
# Generate the dataset
training_dataset = fs.generate_dataset(
    name="CUSTOMER_CHURN_PREDICTION_DATASET",
    spine_df=spine_df,
    features=[customer_feature_view],
    version="1.0",
    spine_label_cols=["churn"],
    desc="Dataset for predicting customer churn"
)
```

**evolv Consulting Expert Tip:** Regularly update your spine DataFrame to include the latest customer data. This ensures that your training datasets reflect the most recent customer behaviors and trends, enhancing the accuracy of your churn predictions.

**4. Train and Register Models:** Train a machine learning model using Snowpark ML and log it into the Model Registry for easy management and deployment.

```
from snowflake.ml.modeling.tree import RandomForestClassifier
from snowflake.ml.registry import Registry

# Convert the dataset to a Snowpark DataFrame
training_data = training_dataset.read.to_snowpark_dataframe()

# Define feature columns and label column
feature_cols = ["tenure", "MonthlyCharges", "TotalCharges", "InternetService_Fi
label_col = "churn"

# Split the data into training and testing sets
train_df, test_df = training_data.random_split([0.8, 0.2], seed=42)

# Initialize and train the model
rf = RandomForestClassifier(
    input_cols=feature_cols,
    label_cols=label_col,
```

```

n_estimators=100,
max_depth=5,
random_state=42
)
rf.fit(train_df)

# Evaluate the model
predictions = rf.predict(test_df)
predictions.select("customerID", "churn", "OUTPUT_churn").show()

# Initialize the Model Registry
registry = Registry(
session=session,
database_name=session.get_current_database(),
schema_name=session.get_current_schema()
)

# Log the model
model_name = "CUSTOMER_CHURN_RF_MODEL"

registry.log_model(
model_name=model_name,
version_name="v1",
model=rf,
comment="Random Forest model for predicting customer churn"
)

```

**Enhanced Evaluation with Metrics:** To better understand the model's performance, you can calculate additional metrics:

```

from snowflake.ml.modeling import metrics

# Calculate accuracy
accuracy = metrics.accuracy(predictions, "churn", "OUTPUT_churn")
print(f"Accuracy: {accuracy}")

# Calculate precision and recall
precision = metrics.precision(predictions, "churn", "OUTPUT_churn")

```

```
recall = metrics.recall(predictions, "churn", "OUTPUT_churn")
print(f"Precision: {precision}, Recall: {recall}")
```

**evolv Consulting Expert Tip:** Incorporate cross-validation techniques during model training to ensure that your churn prediction model generalizes well to unseen data, thereby improving its reliability and performance.

**5. Deploy and Predict:** Use the registered model to perform predictions on new data, ensuring that predictions are based on the most recent feature values.

```
# Retrieve the model from the registry
model_version = registry.get_model(model_name).version("v1")

# Prepare new data for prediction
new_customers_df = session.table("NEW_TELCO_CUSTOMERS").select("customerID")

# Retrieve the latest feature values
features_df = fs.retrieve_feature_values(
    entity_df=new_customers_df,
    features=[customer_feature_view]
)

# Perform predictions
predictions = model_version.run(features_df, function_name="predict")

# Show the predictions
predictions.select("customerID", "OUTPUT_churn").show()
```

## Best Practices for Using Feature Store and Model Registry

**1. Feature Versioning:** Maintain different versions of features to track changes and ensure that models are trained on consistent data.

- 2. Data Quality Checks:** Implement automated checks to validate the quality and integrity of data before it enters the Feature Store.
- 3. Model Monitoring:** Continuously monitor model performance in production to detect and address any degradation over time.
- 4. Documentation:** Keep comprehensive documentation of feature definitions, model architectures, and training processes to facilitate collaboration and knowledge sharing.
- 5. Access Controls:** Utilize Snowflake's role-based access controls to secure sensitive data and restrict access to authorized personnel only.
- 6. Automated Pipelines:** Develop automated pipelines for feature updates and model retraining to ensure your machine learning workflows remain current and efficient.

**EX:** By implementing these best practices, your churn prediction models remain reliable and up-to-date. For instance, regularly updating feature definitions and retraining models ensures that changes in customer behavior are accurately reflected in your predictions.

## **Conclusion**

Snowflake's Feature Store and Model Registry are transformative tools that can significantly enhance your machine learning workflows. By adopting these integrated capabilities and utilizing Snowpark scripts as demonstrated, businesses can achieve greater efficiency, consistency, and insight from their data. These tools not only streamline the development and deployment of machine learning models, but also ensure that data governance and security standards are upheld throughout the ML life-cycle.

Now is the time to elevate your machine learning projects! Embrace these tools to transform your data into actionable insights and stay ahead in the competitive landscape.

## Connect with evolv Consulting

Transform your machine learning workflows with [evolv Consulting](#).

Schedule a consultation today to explore how Snowflake's Feature Store and Model Registry can deliver unparalleled business insights. Our expert team is ready to seamlessly integrate these solutions, maximizing your data-driven success.

*Senior Machine Learning Consultant*  
**Rose Ellison**

Rose excels in leveraging Snowflake's cloud data platform to enhance AI-driven projects combining advanced data science techniques with a deep understanding of Snowflake to address complex data challenges. Her approach centers on practical applications of AI, aiming to transform data into a powerful tool for business optimization. In addition, Rose is known for her forward-thinking strategies, constantly exploring new frontiers in AI to keep her clients at the cutting edge of technology and market trends.



**evolv**

Machine Learning

Snowflake Data Cloud

AI

Customer Churn

ML Model Deployment



## Written by evolv Consulting

Follow

14 Followers · 34 Following

We are cloud-native, business consultants who bring a fresh perspective to help clients overcome #management and #technology challenges.

## Responses (1)



Elliott Botwick

What are your thoughts?



Surendra Reddy

Oct 11, 2024

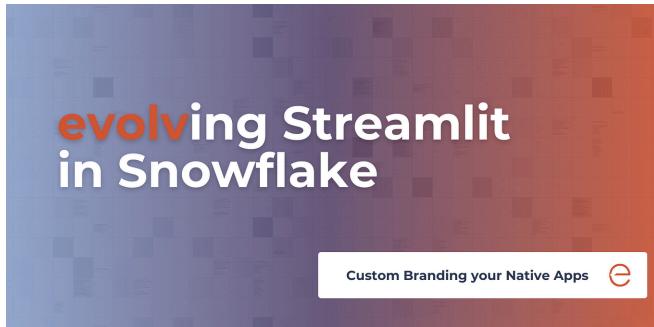
...

Your blog provides a clear and concise explanation, which I found very helpful. However, I encountered a challenge regarding the dataset TELCO\_CUSTOMER\_CHURN. It appears there are no details on how to create or load this table. Could you provide some guidance on that aspect?



[Reply](#)

## More from evolv Consulting



 In evolv Consulting by evolv Consulting

### evolving Streamlit in Snowflake

Branding your Native Apps

Feb 15, 2024  2



...



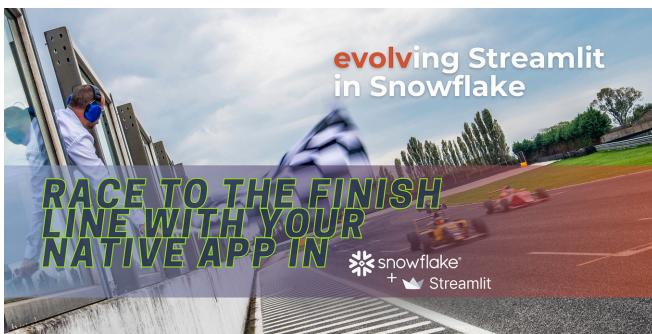
 evolv Consulting

### Solving Vendor Lock in Using a Composable Data Infrastructure

Jul 9, 2024  3



...



 evolv Consulting

## evolving Streamlit in Snowflake: Race to the Finish Line with your...

Part 3, by Chris Schneider

Mar 8, 2024  1



Jul 10, 2024  1

evolv

## The Unsung Hero of AI Transformation

Why LLMs for Extractive AI is the Powerful, Yet Safer, Enterprise AI Use Case

blog by Geoff Clark  
Cloud Data Architect

 evolv Consulting

## The Unsung Hero of AI Transformation

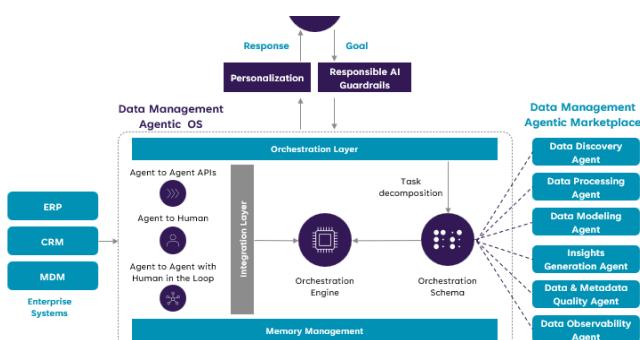
Why LLMs for Extractive AI is the Powerful, Yet Safer, Enterprise AI Use Case



...

See all from evolv Consulting

## Recommended from Medium





In AI Advances by Debmalya Biswas

## Agentic AI for Data Engineering

Reimagining Enterprise Data Management leveraging AI Agents

Mar 23

625

16



...



In Coding Beauty by Tari Ibaba

## This new IDE from Google is an absolute game changer

This new IDE from Google is seriously revolutionary.

Mar 11

3.2K

190



...



In Snowflake Builders Blog: Data Engin... by Piers ...

## 5 Key Drivers Behind Snowflake Cortex LLM Speed & Cost

If you're like me, you've probably wondered which large language model (LLM) can spit...

6d ago



...

Lewis Gavin

## Choosing the BEST Data Architecture in 2025

Lakehouse vs Data Mesh vs ...

Mar 11

187

6



...



Chinonso Nnaji

## Step-by-Step: Solving a Customer Churn Prediction for Energy...

A Comprehensive Guide to Data Science Techniques for Predicting Customer Churn i...



Warda Rahim

## A Simplified Guide to ML Model Deployment Using MLflow on...

Azure Databricks has emerged as a powerful cloud-based platform for big data analytics...

Nov 12, 2024  16

  Mar 19

---

See more recommendations