# H2O.ai and Snowflake
# **Powering Innovation with Data and AI**

# Accelerating Data Science Impact with Snowflake and H2O.ai

Snowflake's Cloud Data Platform enables organizations to drive more business value from their data by easily unifying, integrating, analyzing and sharing their data within the Data Cloud. H2O.ai enables easy access to the entire data science lifecycle from inside the Snowflake Cloud Data Platform. Through multiple pre-built integrations, users can easily leverage H2O.ai machine learning capabilities for real-time analysis from within the Snowflake toolset. This shortens learning cycles, significantly reduces processing time, ensures predictions are based on the most recent data and makes those predictions available to any application built on top of Snowflake.

## Make.

Build machine learning models and AI applications with accuracy, speed and transparency.
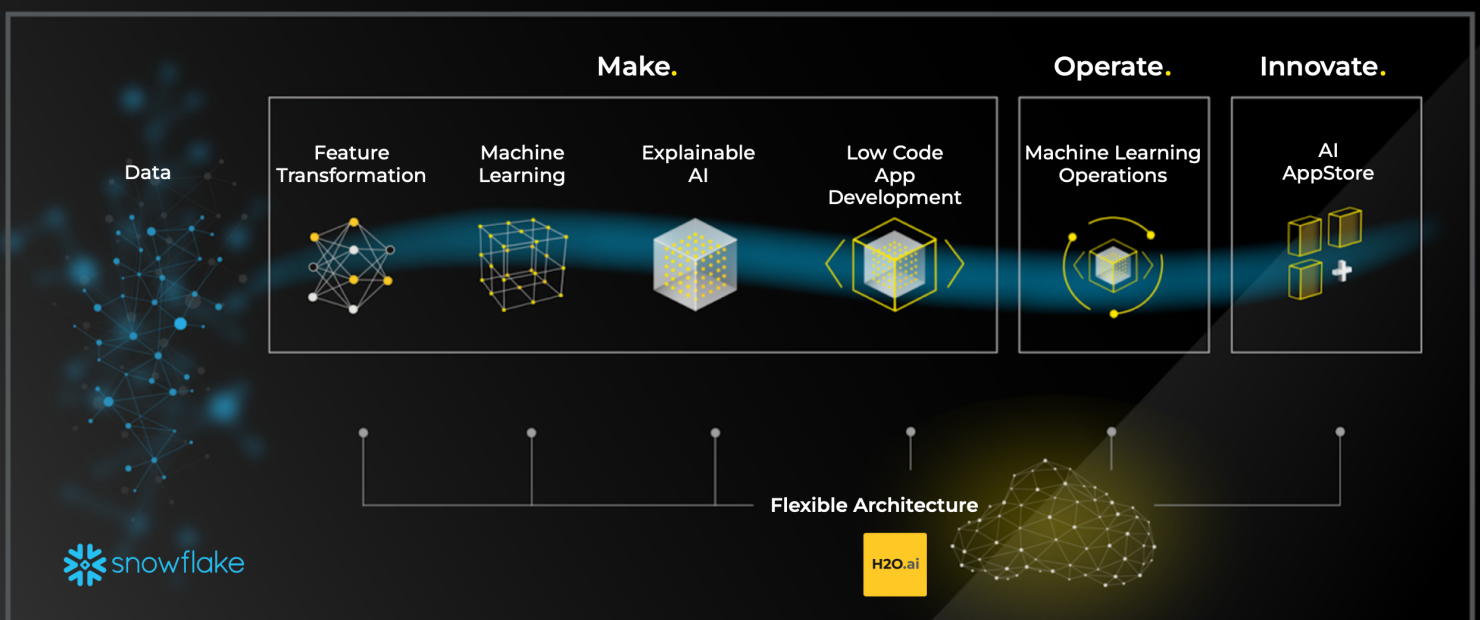
## Operate.

Streamline performance monitoring and rapidly adapt to changing conditions.

## Innovate.

Easily deliver innovative solutions to end users with an intuitive AI AppStore.

# Data powers AI.

| | Make. | | | | Operate. | Innovate. |
|---|---|---|---|---|---|---|
| Data | Feature Transformation | Machine Learning | Explainable AI | Low Code App Development | Machine Learning Operations | AI AppStore |

Flexible Architecture

snowflake

H2O.ai

# Embed Machine Learning Directly into Your Snowflake Environment

An estimated 90% of machine learning models fail to get into production for many reasons, including process complexity and incompatibility across technologies. H2O.ai and Snowflake have eliminated those friction points to seamlessly deliver predictive models inside of the Snowflake environment. This allows for machine learning capabilities to scale across your organization while maintaining speed in performance and accuracy of results.

## Train.

H2O.ai's prebuilt connector streamlines data access for the training of machine learning models.

## Deploy.

Completed models are deployed inside of Snowflake through H2O.ai's point-and-click interface.

## Integrate.

H2O.ai auto-generates code in multiple programming languages to simplify Snowflake integration.

### Train.

| Snowflake Data | Data Prep & Enrichment | Auto Generated Code | Feature Transformation | Machine Learning |
|---|---|---|---|---|
| | Batch | SQL Scala Python | Data Visualization and Insights | Model Training and Fitting |

### Deploy.

| Machine Learning Operations | Model Repository | Model Deployment | Model Monitoring | Machine Learning Scoring Pipeline |
|---|---|---|---|---|
| | | | | Batch In Stream |

### Integrate.

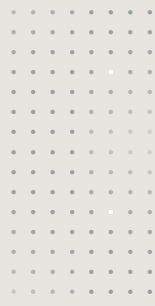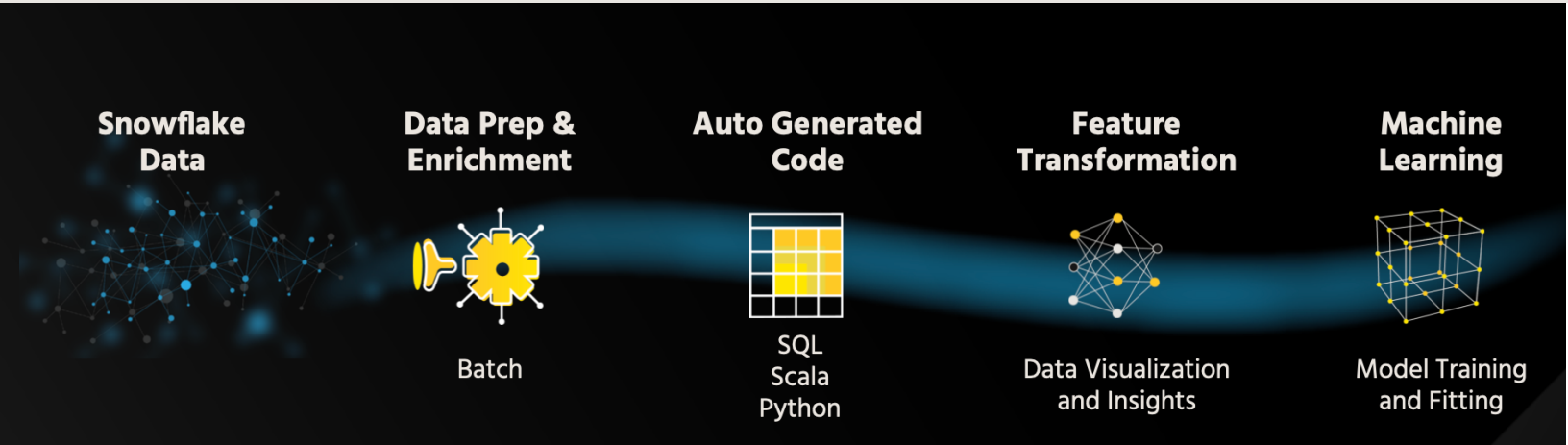| User Defined Function | AI Powered Business Solutions | Explainable Insights |
|---|---|---|
| Java Python | H2O Wave Third Party Solutions | |

# Train
# Machine Learning Models

Accurate machine learning models need high quality data and sophisticated data science tooling. Snowflake provides the data sets, as well as the tools, needed to prepare that data for ingestion. H2O.ai provides the comprehensive automated machine learning platform to make and operate those models at scale.

## Snowflake Data

Data Prep & Enrichment

Auto Generated Code

Feature Transformation

Machine Learning

Batch

SQL
Scala
Python

Data Visualization and Insights

Model Training and Fitting

### Snowflake Data

Enables the rich Snowflake data ecosystem to be used for training models. For example, users can produce augmented datasets to further improve accuracy by combining customer data with marketplace data.
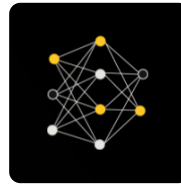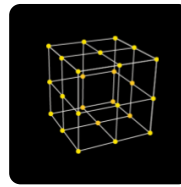
### Feature Transformation

Automatically visualize and address data quality issues with advanced feature engineering that transforms your data into an optimal modeling dataset.

### Data Prep & Enrichment

Leverage the scale of the data cloud to perform comprehensive data preparation and enrichment.

### Machine Learning

Quickly create and test highly accurate and robust models with state-of-the-art automated machine learning that spans the entire data science lifecycle and can process a variety of data types within a single dataset.

### Auto-generated Code

Autogeneration of artifacts allows all the team members to use the model in their preferred programming language, reduce the time required to pass models between groups and results in faster consumption within the environment.

The H2O AI Cloud is built for speed and agility for experimentation, deployment and feedback. By reducing time-to-insight through automation, speed, accuracy and explainability, users are able iterate faster, innovate continuously and deliver more value to the business.
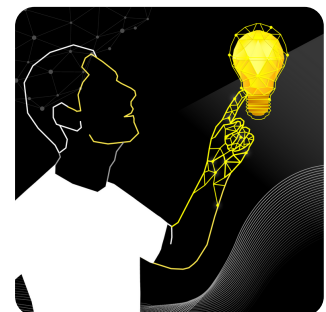
# Train Models



**Auto Generated Code**

Select * from myTable;

**Raw Data** (Table)

**Data Prep & Enrichment**

**Feature Engineering & Transformation**

**Machine Learning**

snowflake



## Data Engineers

Use machine learning models as part of the data preparation and execution pipeline to both enrich data and check for anomalies.
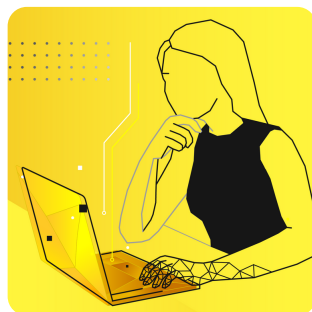
## Data Scientists

Train machine learning models with familiar tooling, and easily interpret, monitor and refine model output.

## Developers

Initiate training via a rich API to create highly accurate and explainable models.

# Deploy
# Machine Learning
# Pipelines

H2O.ai provides a comprehensive suite of capabilities surrounding machine learning operations that support data scientists, machine learning engineers and IT professionals in the management, deployment, and monitoring of their models in production. Additionally, the H2O AI Cloud provides a highly open and flexible architecture with distributed processing, optimized compute efficiency and the ability to deploy in the environment of your choice.
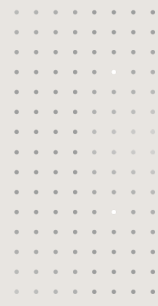


| Machine Learning Operations | Model Repository | Model Deployment | Model Monitoring | Machine Learning Scoring Pipeline |

Batch
In Stream

### Machine Learning Operations

Monitor models in real-time and set custom thresholds to receive alerts on prediction accuracy and data drift and guarantee deployed models are operating as intended.

### Model Monitoring

Maintain model oversight and know when data drift occurs. Feature importance delivers local explanations as to which features are contributing the most or least to prediction values.
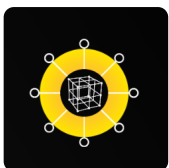
### Model Repository

Create a central place to host and manage all experiments. Maintain a view of all deployed versions with complete, integrated model management capabilities.

### Machine Learning Scoring Pipeline

By enabling the model to execute within the Snowflake environment, any workflow that uses Snowflake can now consume the model as well.
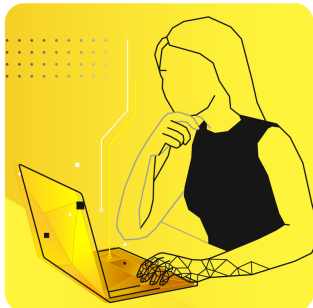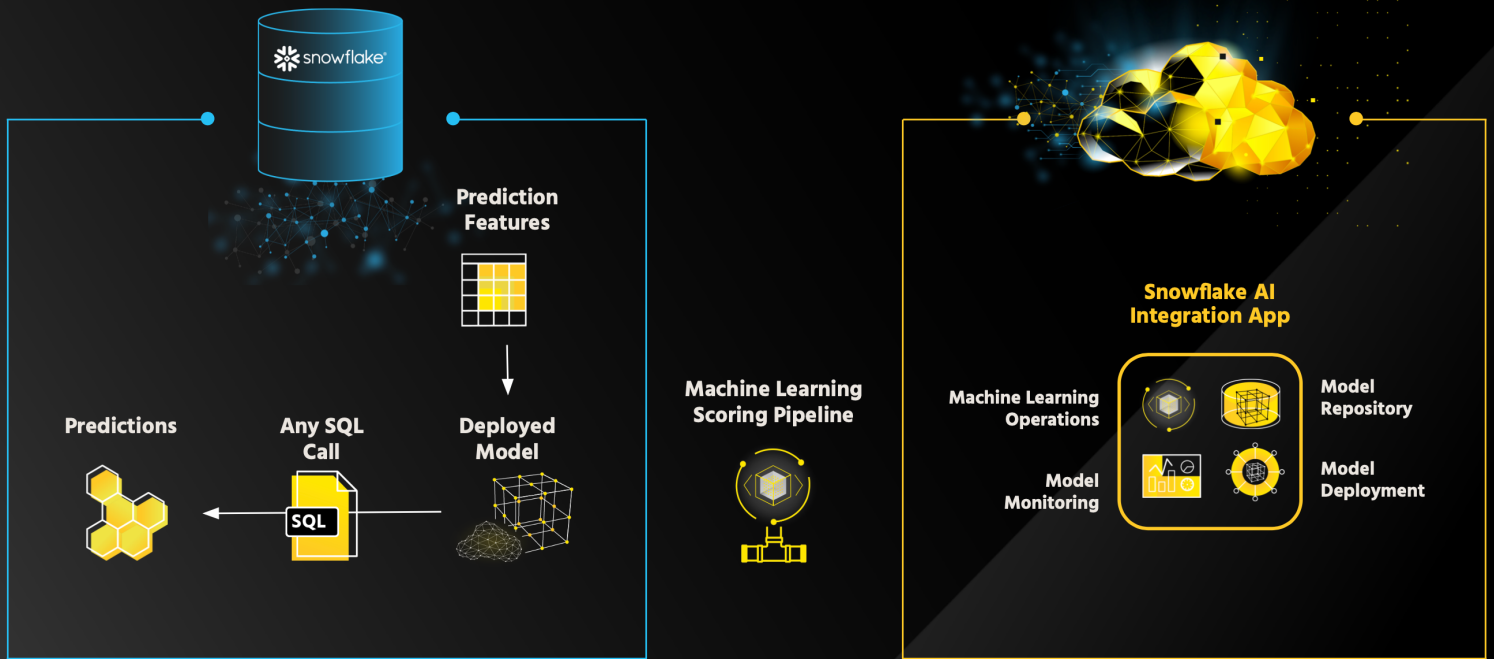
### Model Deployment

You can build once and deploy to any scoring environment with target deployments. Models can be scored in real-time, in batch, asynchronously or as streaming data.

The H2O AI Cloud is platform agnostic with clients for Python, R and Java. Users benefit from the latest versions of all major open source packages and gain control over them with our built-in custom recipe architecture. Distributed machine learning backends can handle any data size by scaling out to multiple worker nodes, with model training occurring across multiple CPUs and GPUs.

# Deploy Models



**Prediction Features**

**Snowflake AI Integration App**

Machine Learning Operations

Model Repository

Model Monitoring

Model Deployment

**Machine Learning Scoring Pipeline**

**Predictions**

**Any SQL Call**

SQL

**Deployed Model**

### Data Scientists

The time to move a model into production is reduced, so models that are sensitive to decay are able to be used for a extended period of time before retraining.
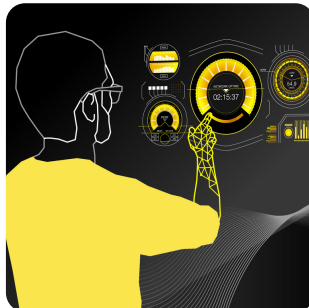
### Machine Learning Engineers

Providing a flexible number of options for deployment makes it simpler for the engineer to use the models in new and creative ways.
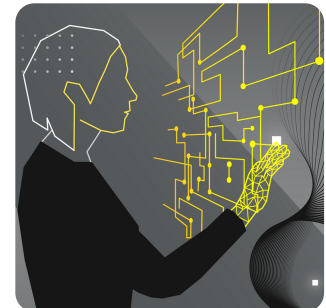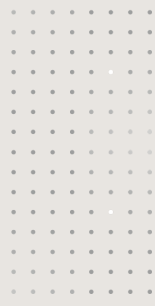
### DevOps/IT Professionals

The return on investment can be realized quickly by reducing the time to deploy a model into a production environment.

# Integrate
# Machine Learning
# Solutions

Contact Us

Snowpark enables the data in Snowflake to be available as a dataframe using Scala or Python. The Java UDF feature allows Java code to be executed as a user defined function (UDF) within the Snowflake environment. That means once a model is trained, it can be imported into where the data lives for an easier path to deployment. This opens up an amazing opportunity to use familiar H2O.ai tools and models alongside the power and scaling of the Snowflake platform.

## User Defined Function

Java
Python

## AI Powered Business Solutions

H2O Wave
Third Party Solutions

## Explainable Insights

### User Defined Function and External Functions

Increase the use cases that can quickly be addressed with machine learning models using native SQL within the Snowflake environment.

### AI Powered Business Solutions

Advanced machine learning models can be used to easily power existing and new applications.

### Explainable Insights

Provide trust and transparency in machine learning predictions with explainable results.

The H2O AI Cloud supports rapid prototyping and solution development, while also fostering collaboration between technical teams and business users. You can interact with your Snowflake instance and manage the end-to-end machine learning pipeline with seamless integration between your Snowflake and H2O.ai platforms. Easily discover your data, train new models, score with existing models and integrate models into your business applications.
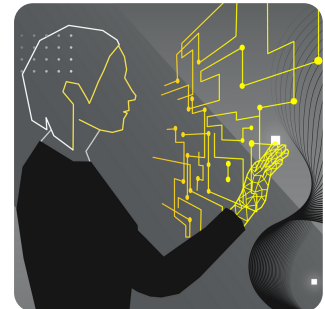
# Integrate Models
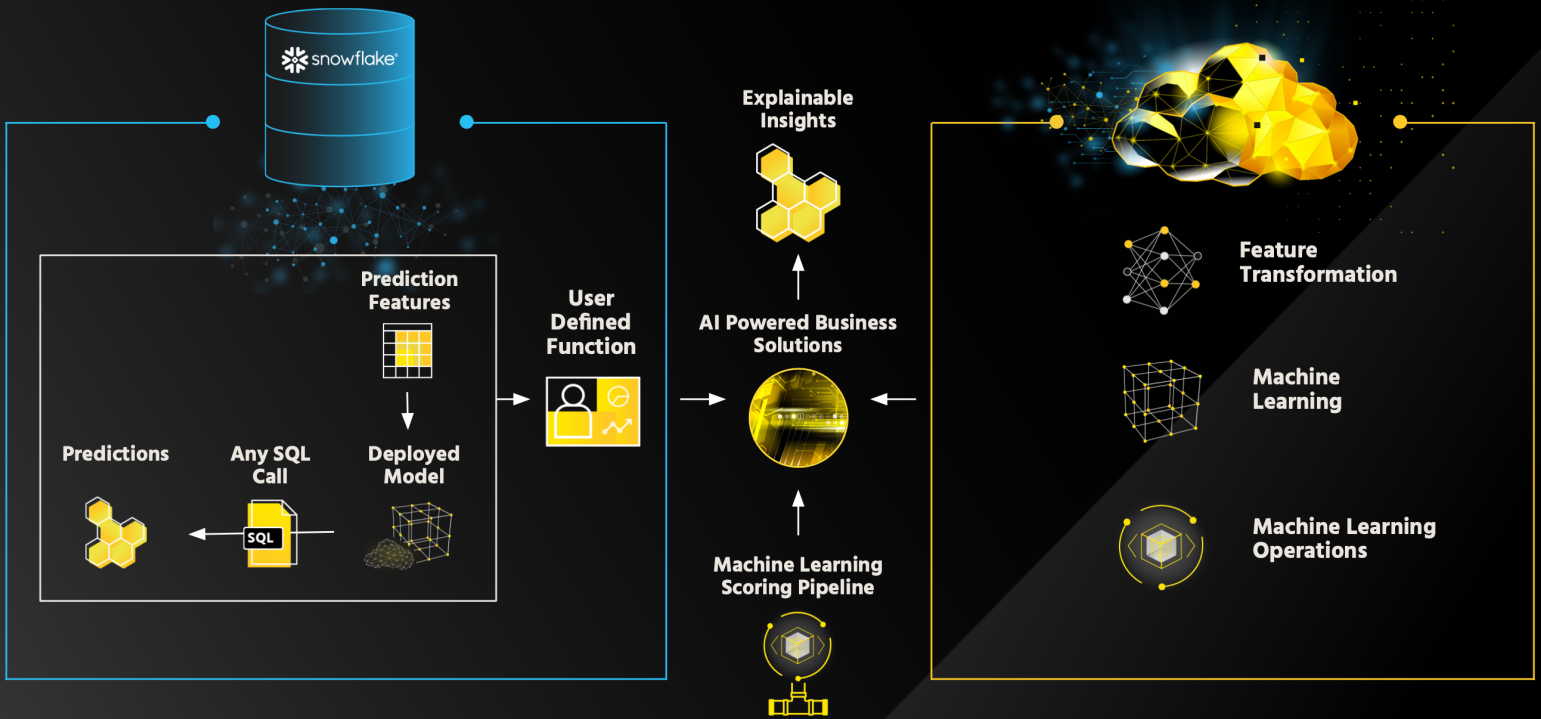


## Data Analysts

The Snowflake and H2O.ai integration enable new data insights using advanced machine learning models that can unlock value for the business.

## Business Users

Access accurate and trustworthy models easily with Snowflake tools you are already familiar with.

## DevOps/IT Professionals

Reduce the time and friction for machine learning models to produce value by accessing models through native Snowflake tooling.
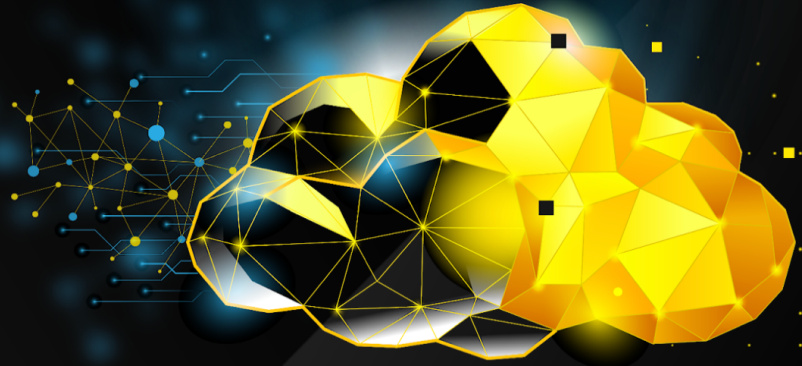
# Innovate with
# Data and AI

**Snowflake Data**

**H2O.ai Machine Learning**

## Scale AI initiatives across your business.

**End-to-End Integration**

**1**

Support any business use case with unified data and AI operations.

**Best-in-Class Product**

**2**

AutoML platform delivering accuracy, speed and trust across the entire data science lifecycle.

**Winning Team**

**3**

Built by a community of data scientists, business leaders, nonprofits and academics

Recognized as a global visionary and thought leader in automated machine learning (autoML), time series forecasting and responsible AI, H2O.ai is the trusted AI partner to more than 20,000 organizations around the world. Our platform, the H2O AI Cloud, enables businesses, government entities, nonprofits and academic institutions to make, operate and innovate with AI to accelerate responsible innovation and push the boundaries of what is possible with artificial intelligence.

# Email the H2O.ai team to learn more.

**snowflake@h2o.ai**