# SNOWFLAKE FOR
# DATA ANALYSIS

**Vino Duraisamy | Nov 2023**

# PRE-REQUISITES FOR CODE-ALONG

# quickstarts.snowflake.com

# quickstarts.snowflake.com

## Prerequisites

- [Git](#) installed

- [Python 3.9](#) installed

  - Note that you will be creating a Python environment with 3.9 in the **Get Started** step

- A Snowflake account with [Anaconda Packages enabled by ORGADMIN](#). If you do not have a Snowflake account, you can register for a [free trial account](#).

- A Snowflake account login with ACCOUNTADMIN role. If you have this role in your environment, you may choose to use it. If not, you will need to 1) Register for a free trial, 2) Use a different role that has the ability to create database, schema, tables, stages, tasks, user-defined functions, and stored procedures OR 3) Use an existing database and schema in which you are able to create the mentioned objects.

> IMPORTANT: Before proceeding, make sure you have a Snowflake account with Anaconda packages enabled by ORGADMIN as described [here](#).

Next

# SNOWFLAKE SQL
# +
# SNOWPARK PYTHON

**BEFORE SNOWPARK**

**AFTER SNOWPARK**

Two compute clusters to manage

**External Compute**

Python | Java | Scala
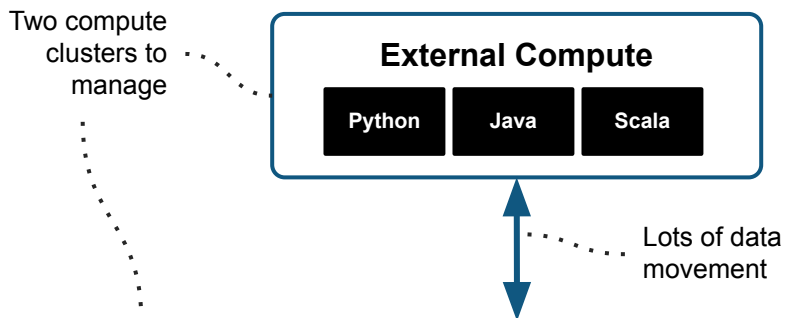
Lots of data movement

No data movement - processing comes to where the data is

SQL

SQL | Python | Java | Scala

**Snowpark**

❄️ snowflake®
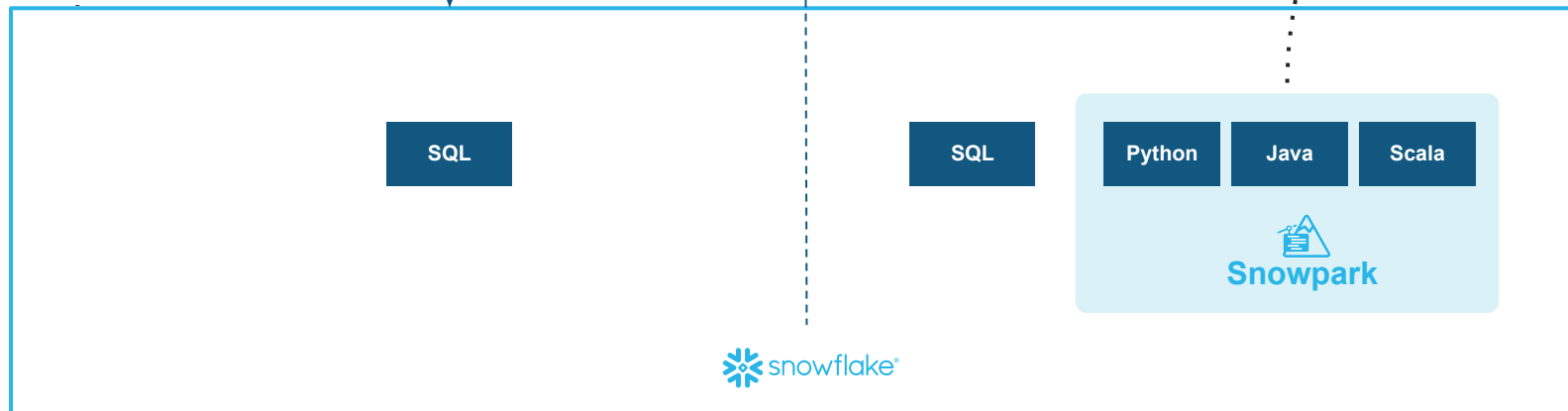
# Snowpark: Securely Deploy & Process Non-SQL Code

**PYTHON • JAVA • SCALA**

**CLIENT SIDE LIBRARIES**

**Snowpark DataFrame API**

**Snowpark ML API**

ML Modeling API (PuPr)

ML Operations API (PrPr)

**SERVER SIDE RUNTIMES**

UDFs

Stored Procedures

Services

Jobs

**Built-in Anaconda Packages**  NumPy  learn  pandas  Keras  & more

**Warehouses (Standard & Snowpark-Optimized)**

**Snowpark Container Services (PrPr)**

> **Develop in IDE of Choice**
Use Python worksheets, VSCode, Jupyter or any client with a python kernel

> **Process with security & scalability**
Bring processing to governed data for DataFrame operations and custom Python/Java code

> **Automate code execution**
Orchestrate your code with DAG representations using SProcs with Snowflake Tasks

**DataFrame Query**

**Python Functions & Stored Procedures**

```
df.filter(df.state == 'WA')
```

```
@udf def detect_fraud()
```

**IDE with Snowpark Library**

| Query Translator | Object Serializer |

| Snowflake Connector for Python |
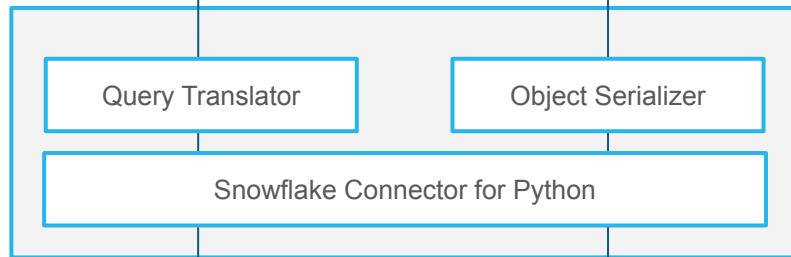
`{...} SQL Query`

`Python Bytecode`

**Server Side Runtimes**

Processing Engine

| SQL Engine | ⟷ | Python Secure Sandbox | ⟷ |

**Built-in Anaconda Packages**

NumPy    pandas

learn    K Keras

TF    & more

| Snowflake Tasks |
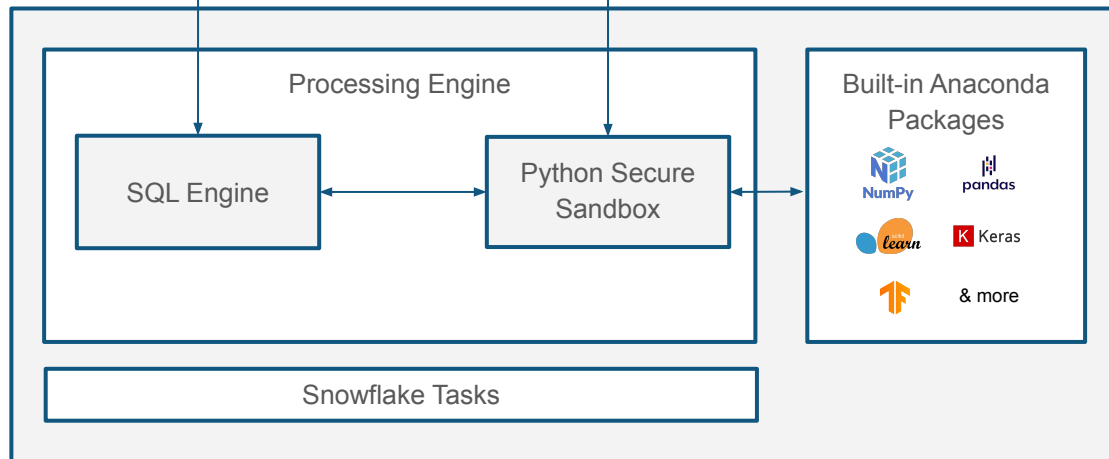
# End-to-End Data Science Workflow

Effortless, scalable and secure processing without data movement across compute environments

**Prepare & Transform**

```
DF = DF.()
_____
_____
_____
```

Data lake

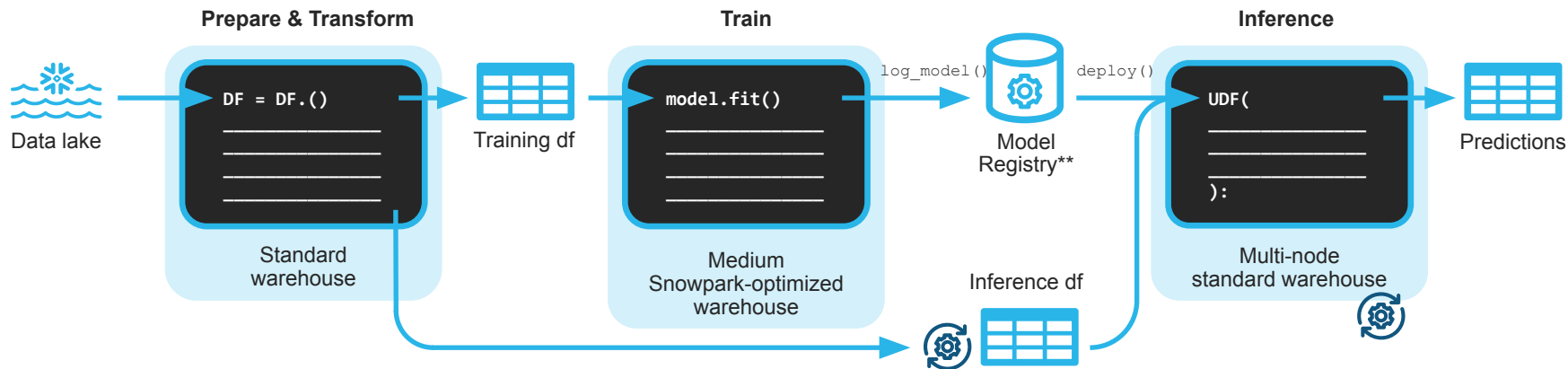Training df

Standard
warehouse

**Train**

```
model.fit()
_____
_____
_____
```

Medium
Snowpark-optimized
warehouse

log_model()

Model
Registry**

deploy()

**Inference**

```
UDF(
_____
_____
_____
):
```

Multi-node
standard warehouse

Predictions

Inference df

| Development | snowflake.ml.modeling.preprocessing* | snowflake.ml.modeling* | Deployment | snowflake.ml.registry** |
|---|---|---|---|---|
| Orchestration | snowflake.core.task** | | | |

LETS WRITE SOME CODE

# MARKETING SPEND AND REVENUE ANALYSIS

# STEP #1
# signup.snowflake.com

# signup.snowflake.com



## snowflake®

# START YOUR
# 30-DAY FREE TRIAL

- Gain immediate access to the Data Cloud

- Enable your most critical data workloads

- Scale instantly, elastically, and near-infinitely across public clouds

- Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized

Start your 30-day free Snowflake trial which includes $400 worth of free usage

First Name*

Last Name*

Company Email*

Company Name*

Role*

United States

By clicking the button below you understand that Snowflake will process your personal information in accordance with its Privacy Notice

**CONTINUE**

or sign in to an existing account

THANK YOU