



aws SUMMIT

INDIA | MAY 25, 2023

AIML005

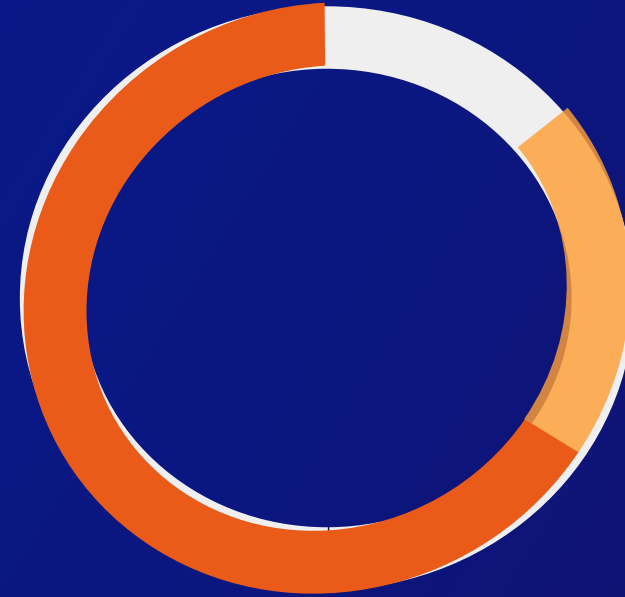
Unlock Data Value with Snowpark and Streamlit

Sumeet Tandure
SE Manager, India, Snowflake



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Companies
strategically
scaling AI
are achieving 3x
return
yet...



- 62% Struggle to operationalize or still in proof of concept
- 24% Some models in prod but formalizing process
- 14% Defined scalable and repeatable process

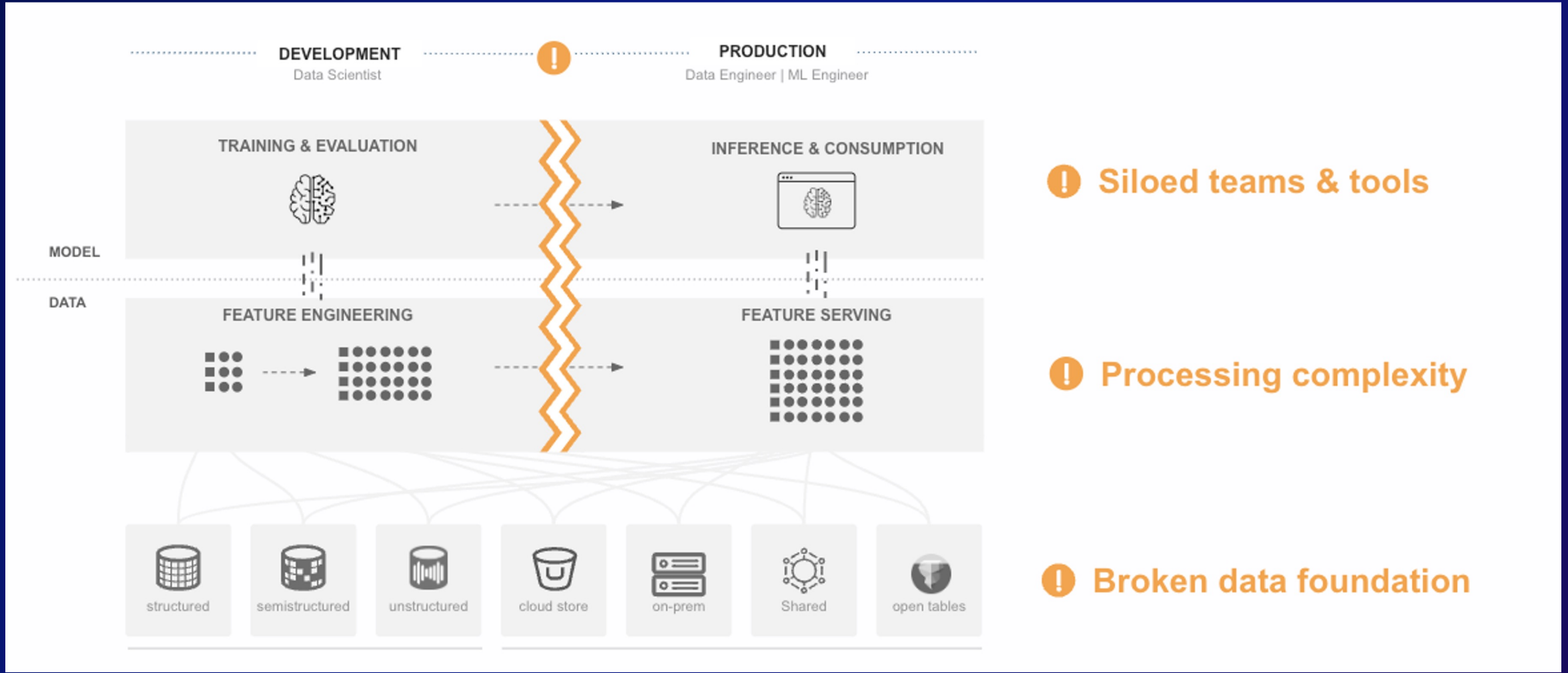
Source: Accenture. AI: Built to Scale. 2019



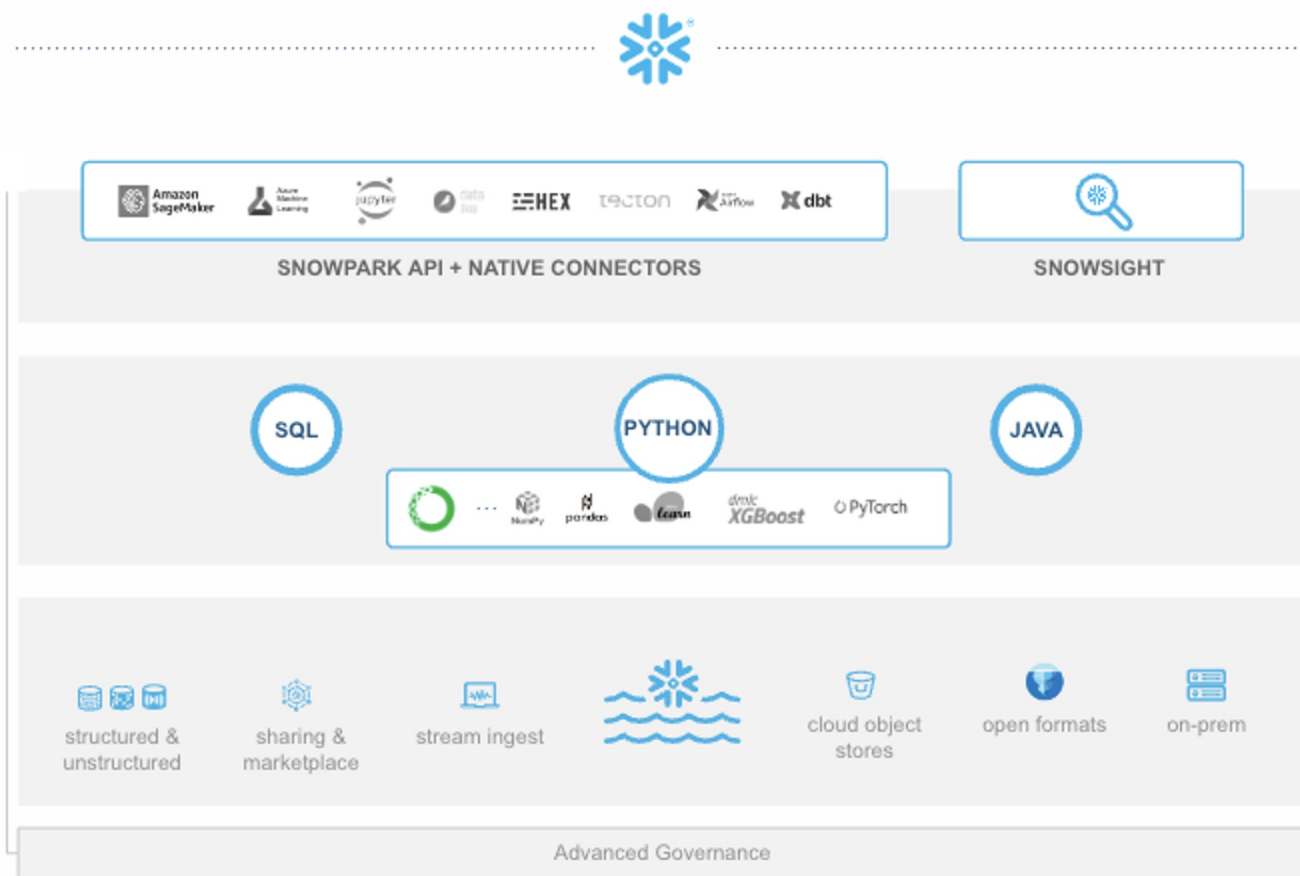
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Forrester. Operationalize Machine Learning. 2020

The ML development to production gap



Snowflake platform for data science and ML



- Natively integrated ecosystem
- Multi-language, elastic engine
- Unified data access



Snowpark for Python



Snowpark
Client API

Snowflake

DataFrame Query

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

Python Functions & SProcs

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

Query Translator

Object Serializer

Snowflake Connector for Python

{...} SQL Query

Python Bytecode

Processing Engine

SQL Engine

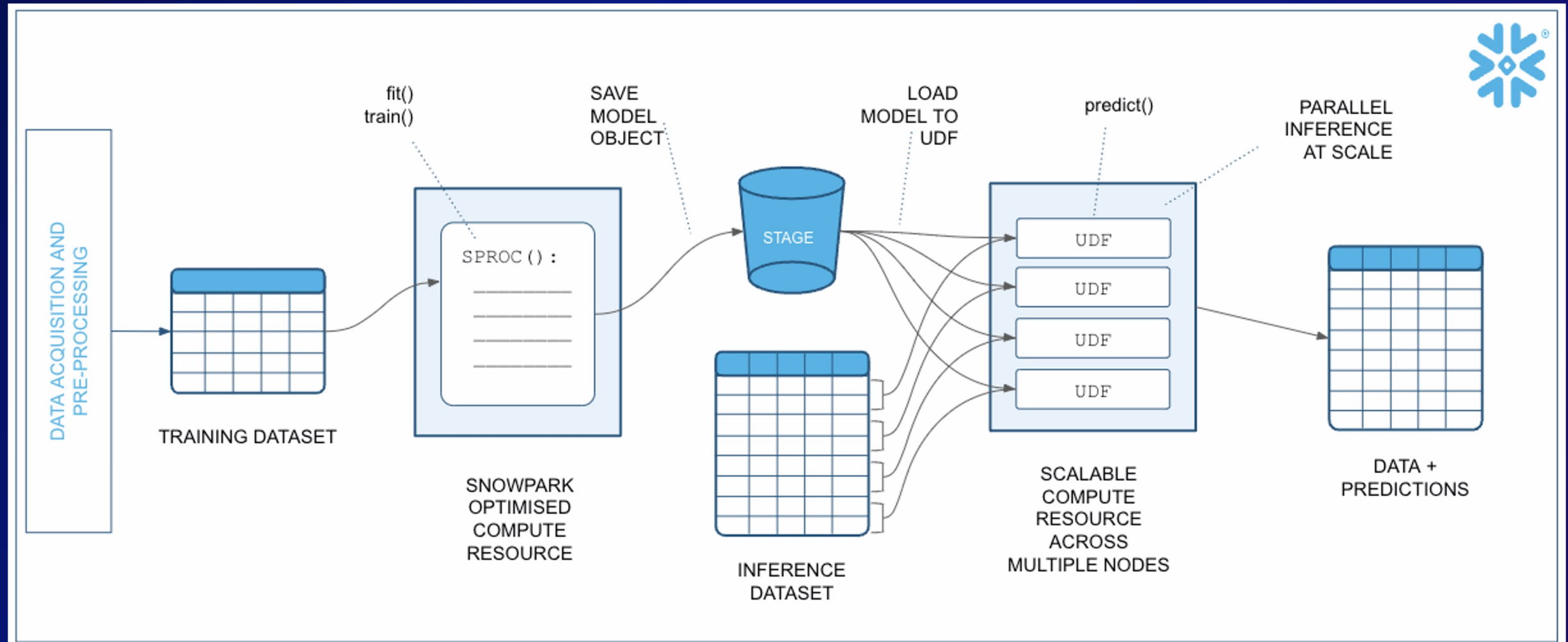
Python Secure
Sandbox

Built-in Anaconda
Packages



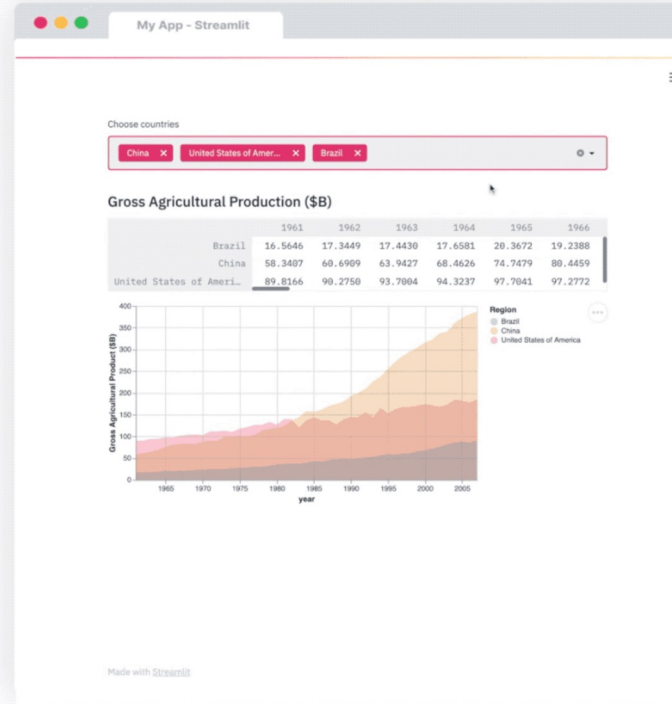
& more

End-to-End Machine Learning with Snowpark



Streamlit open source Python library

```
1 import streamlit as st
2 import pandas as pd
3 import altair as alt
4
5 @st.cache
6 def get_UN_data():
7     AWS_BUCKET_URL = "https://streamlit-demo-data.s3-us-west-2.amazonaws.com"
8     df = pd.read_csv(AWS_BUCKET_URL + "/agri.csv.gz")
9     return df.set_index("Region")
10
11 df = get_UN_data()
12
13 countries = st.multiselect(
14     "Choose countries", list(df.index), ["China", "United States of America"]
15 )
16
17 data = df.loc[countries]
18 data /= 1000000.0
19 st.write("## Gross Agricultural Production ($B)", data.sort_index())
20
21 data = data.T.reset_index()
22 data = pd.melt(data, id_vars=["index"]).rename(
23     columns={"index": "year", "value": "Gross Agricultural Product ($B)"}
24 )
25
26 chart = {
27     alt.Chart(data)
28     .mark_area(opacity=0.3)
29     .encode(
30         x="year:T",
31         y=alt.Y("Gross Agricultural Product ($B)", stack=None),
32         color="Region:N",
33     )
34 }
35 st.altair_chart(chart, use_container_width=True)
```



Bring your data to life in minutes by building an interactive app – all in Python

Get started today

- Install Streamlit
- Connect to Snowflake
- Create!



DEMO



MODERNIZE YOUR
PIPELINES: SPARK TO
SNOWPARK MIGRATION
GUIDE

RUN END TO END MACHINE
LEARNING WITH SNOWPARK
& SNOWPARK-OPTIMIZED
WAREHOUSES

HANDS-ON LAB: Getting
Started with Data
Engineering and ML using
Snowpark for Python

Thank you!

Sumeet Tandure
SE Manager, India, Snowflake



Please complete the
session survey