From Data to Decisions

Modernizing Analytics at FSU with Snowflake, AI, and Power BI

ITS - Data & Analytics
Dinesh Kumar Sundararajan & Tazmeen Alam

PRESENTERS

Dinesh Kumar Sundararajan

Sr. Data Integration Architect
ETL Pipelines, Data Replication
Masters in Computer Engineering
Graduate Certificate in Information Leadership & Management
Experience working in Robotics, IoT & Data Analytics

Tazmeen Alam

Analytics Architect
Data Analytics & Business Intelligence
Masters in Information Technology
(with graduate certificate in Information Architecture)
Experience working in BI solutions, Database management & Data Analytics

“We are drowning in information but starved for knowledge. The alchemy of the modern age is not turning lead into gold, but turning frozen data into fluid intelligence”

—Adapted from E.O. Wilson

AGENDA

01 Foundations

- The Evolution of Data & Traditional Challenges
- Why Modern Data Architecture Matters in Academia
- Introduction to Snowflake & How it became popular
- Snowsight User Interface

02 Infrastructure

- Types of data supported in Snowflake
- Data Ingestion & Loading
- Working with Semi-Structured Data
- Connecting Snowflake to Power BI

03 Superpower

- Evolution toward an AI platform
- Cortex AI Umbrella
- Live Demo: Raw Data to AI-Driven Power BI Dashboard
- Key Takeaways

FOUNDATIONS

Why modern data platforms became necessary

WHY THIS TOPIC MATTERS?

1 The Data Maze

Silos are holding us back. Our data lives in “locked rooms”—Different systems don’t talk to each other.

2 The “Wait” Problem

Information has an expiration date. In a world of instant answers, waiting days for a manual report feels like waiting for snail mail.

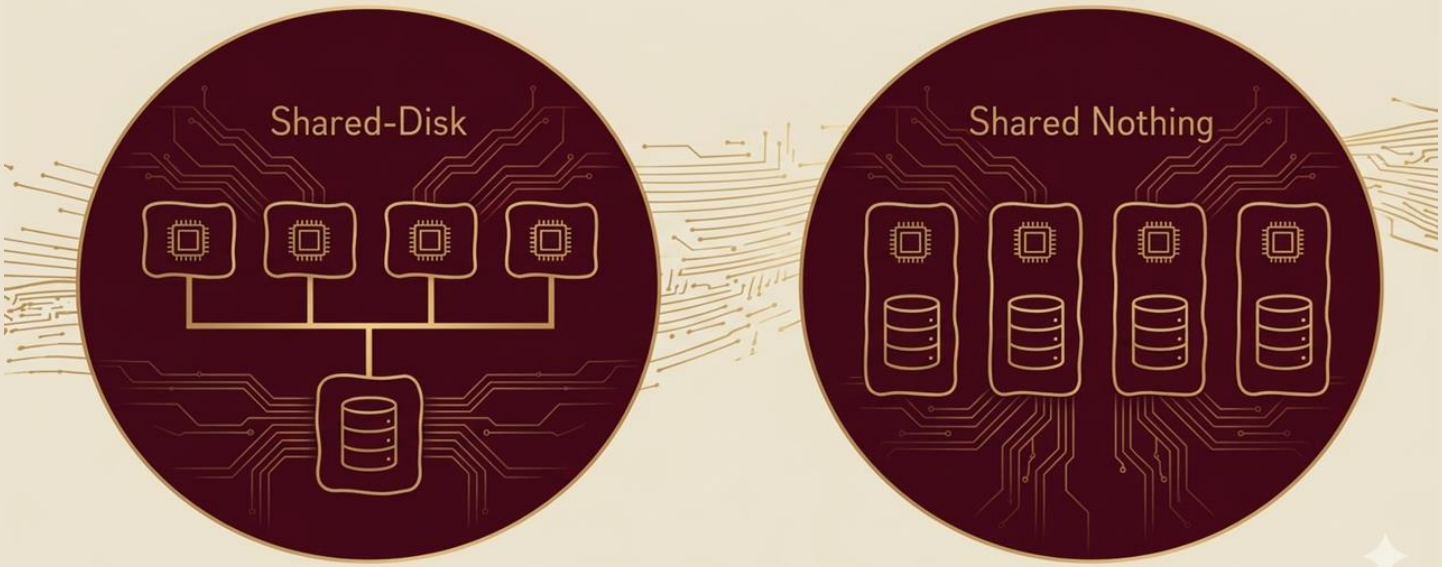
3 The Variety Gap

Data isn't just spreadsheets. Modern research uses PDFs, sensor data, and notes. We need a system that reads it all, not just rows and columns.

4 The Modern Standard

Growth without the headache. We need a platform that scales instantly, stays flexible for any project, and keeps our data under "lock and key."

LIMITS OF LEGACY ARCHITECTURE



SNOWFLAKE ORIGIN STORY

2012: The Visionaries

Benoit Dageville
Thierry Cruanes
Marcin Zukowski

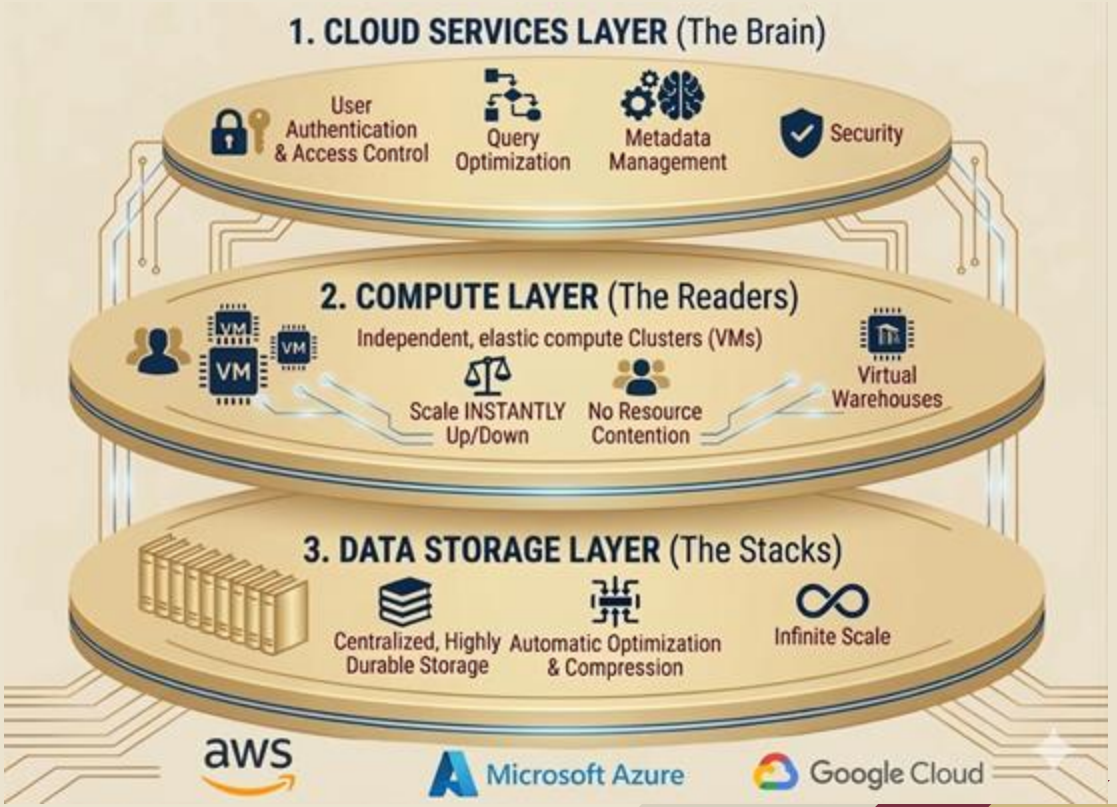
The “Cloud-Native” Philosophy

Born in the Cloud,
Not Moved to It.
- blank slate design

2014: The Disruption

Public Launch,
Rapid Adoption

SNOWFLAKE ARCHITECTURE OVERVIEW



HOW SNOWFLAKE BECAME POPULAR

1. True Freedom

Complete separation of storage & compute. Grow your library without having to hire more readers. Pay only for what you use.

2. Infinite Performance

No more resource contention. The Registrar can run huge reports *at the same time* as Finance, and neither slows the other down.

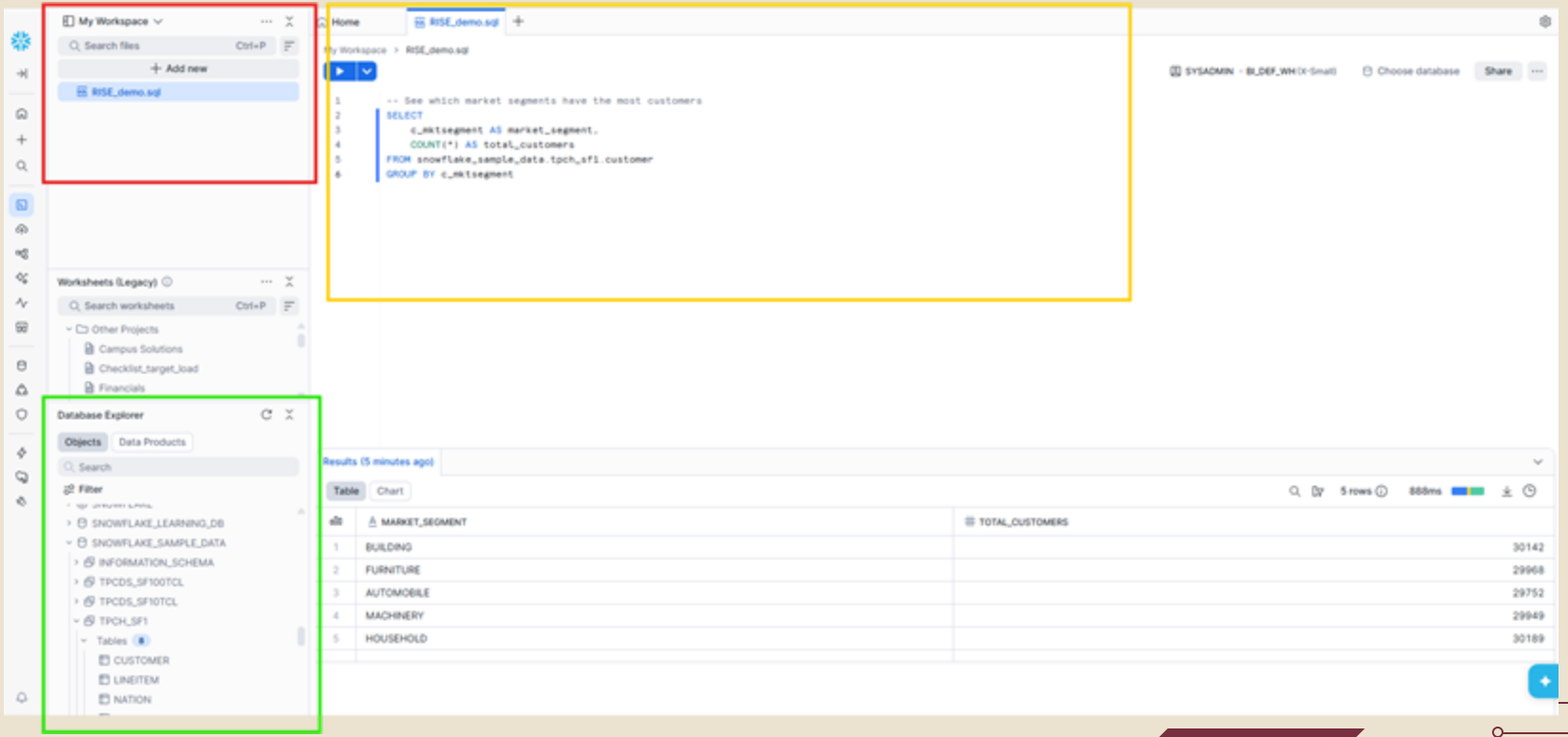
3. Zero Overhead

A **fully managed platform.** It's 'self-driving' data architecture. No hardware to maintain, no knobs to tune.

4. Total Flexibility

Multi-cloud and multi-data. Run on AWS, Azure, or Google Cloud. Handle structured Excel and unstructured PDFs in one place.

SNOWSIGHT USER INTERFACE



INFRASTRUCTURE

From raw data to actionable insights

Snowflake Data Types

Structured

Highly organized data that fits into predefined tables and rows (CSVs)

DATA TYPES

- Numeric
- String & binary
- Logical
- Date & time
- Geospatial
- UUID

Semi-structured

Data that does not reside in a rational wireframe but contains tags or markers to separate semantic elements (JSON, XML, Parquet)

DATA TYPES

- Variant
- Object
- Array

Unstructured

Data that doesn't have any predefined data mode (document, image, or audio file)

DATA TYPES

- PDF
- Image
- Audio files
- Word docs
- Vector
- User-defined

How to handle semi-structured/unstructured data?

Storage

Files stay where they are

- Files live in **cloud storage** - Snowflake points to them, no copying
- **Internal stage** (within Snowflake) or **External stage** (S3, Azure, GCS)
- Query file metadata - name, size, date - via a **directory table**

Semi-structured

Query flexible formats natively

- Stored as **VARIANT** - holds JSON, XML, Parquet as-is, no rigid schema needed
- Navigate nested fields with **dot notation** - e.g. `col:customer.name`
- **FLATTEN** - splits arrays into rows for easy analysis in SQL or Power BI

Processing Images

Classify, detect, and describe with AI

- **Cortex AI** - describe or classify images from a stage using a vision model
- **Snowpark (Python)** - run custom ML models like OCR inside Snowflake
- **External functions** - call Azure Computer Vision or Amazon Textract from SQL

Processing PDFs

Extract structured data from documents

- **Document AI** - extracts fields (dates, names, amounts) from PDFs automatically
- Results load directly into a table with `COPY INTO`
- **Cortex AI** can summarize or answer questions from any staged PDF

Note: Snowflake does not support BLOB/CLOB - all file handling goes through stages and functions like Document AI or Cortex AI.

FSU Snowflake Architecture

HIPAA NON-HIPAA

S_FI → LANDING ZONE

"S_" Layer

The landing zone for raw data migrated from original SQL views and legacy systems.

ODS → INTERMEDIARY

ODS Layer

The intermediary operation data storage between S_ layer and O_ layer.

O_IR → PRIMARY ZONE

"O_" Layer

The primary zone for business users, analysts, and Power BI connectivity.

Snowflake FSU Access

User Accounts

STEP 1
Submit a Salesforce ticket mentioning the FSUID of the user to be added.

STEP 2
Ticket gets assigned to the D&A queue.

STEP 3
New user is created in Snowflake with a default role and warehouse.

STEP 4
User's FSUID is added to Azure Active Directory for SSO.

Service Accounts

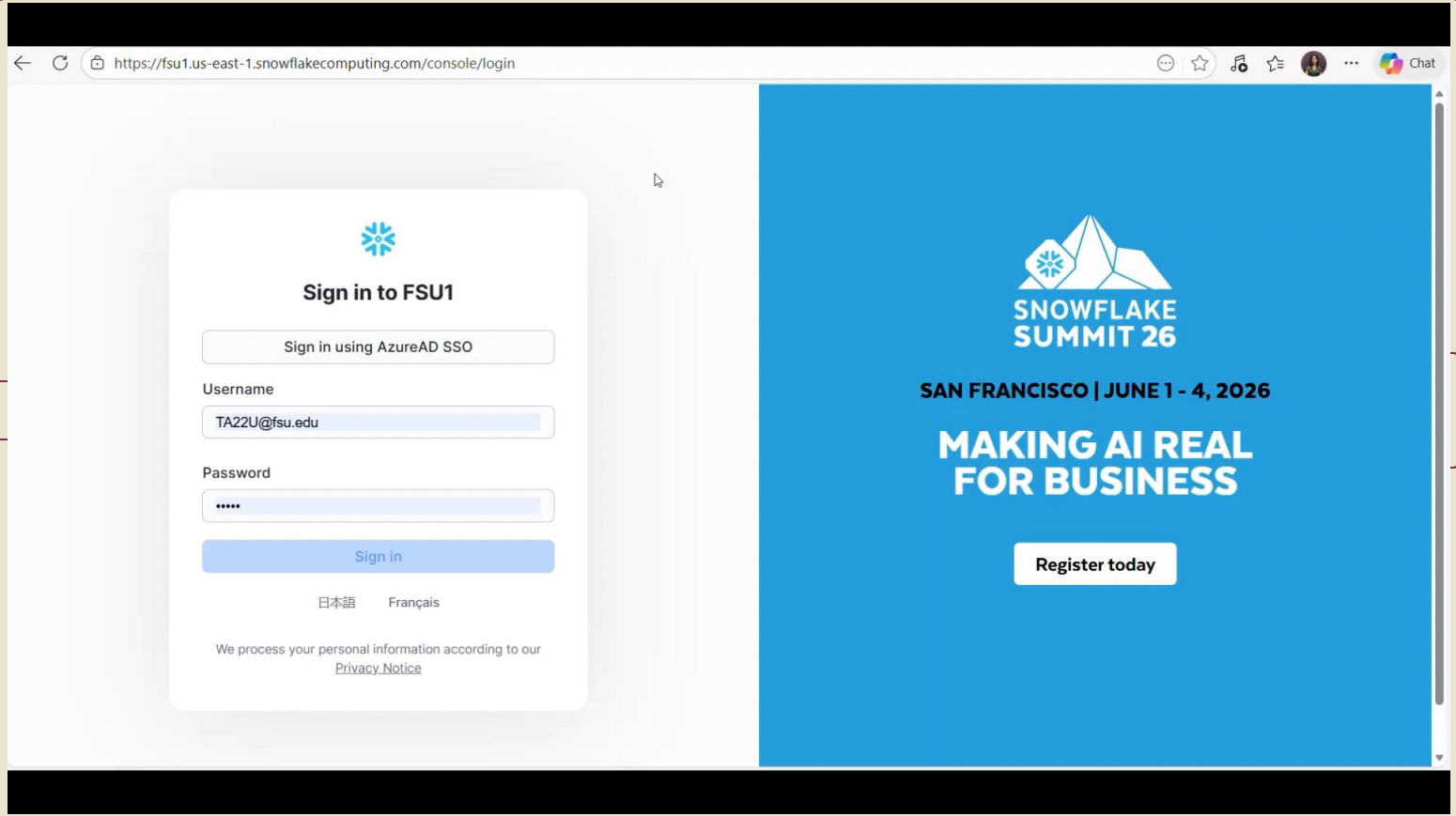
STEP 1
Submit a Salesforce ticket mentioning the FSUID and public key to be added.

STEP 2
Ticket gets assigned to the D&A queue.

STEP 3
User is updated in Snowflake using the `RSA_PUBLIC_KEY` property.

STEP 4
Application authenticates via **JSON web token** using the local Private Key (SSO not used).

Connect Snowflake and Power BI



SUPERPOWER

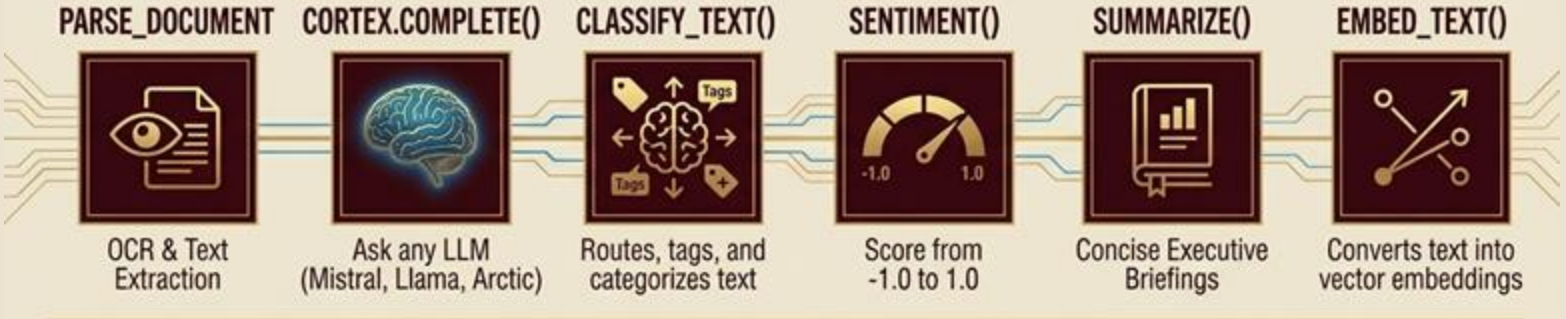
Turning data into intelligent decisions

Why AI inside Snowflake?

- **Bring the Brain to the Data**
  Eliminates risky file transfers; AI runs securely within the FSU data vault.
- **Instant Intelligence**
  Immediate access to 10+ industry-leading models (GPT-4o, Claude 3.5) with zero setup.
- **Academic-Grade Security**
  Data is never used for public training and remains fully governed by FSU security rules.
- **Research-Friendly Scaling**
  No expensive hardware to manage, pay only for what you use at a "per-token" rate.

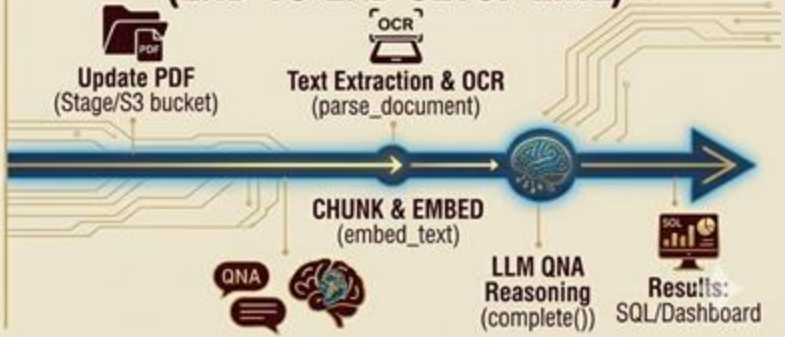
Snowflake Cortex AI Umbrella

AI Deep-Dive & Demo Setup



DOCUMENT AI SPECIALIZED WORKFLOW (BUILD PIPELINE)

DEMO SETUP ARCHITECTURE (END-TO-END SETUP LINE)



DEMO

KEY TAKEAWAYS

Elastic Scaling: Decoupling **storage and compute** allows you to scale resources instantly and only pay for what you use.

Maintenance-Free Design: As a **cloud-native** platform, Snowflake handles massive data volumes without the "noise" or bottlenecks of legacy systems.

Unified Data: Store and query both structured (tables) and semi-structured (JSON) data in one place without complex pre-processing..

Accessible Intelligence: The **Snowsight UI** democratizes data, allowing researchers to move from raw code to visual insights with a few clicks.

AI-Driven Outcomes: With **Cortex AI**, the platform evolves from simple storage into an intelligent engine for rapid, AI-powered decision-making.

REFERENCES

- **Snowflake Official Documentation:** Snowflake Inc. (2024). Snowflake Documentation: Overview and Guides. Link: **https://docs.snowflake.com/**
- **Snowflake Cortex AI Service:** Snowflake Inc. (2024). Snowflake Cortex AI: Large Language Model (LLM) and Machine Learning Functions. Link: **https://docs.snowflake.com/en/user-guide/snowflake-cortex**
- **Ultimate SnowPro Core Certification Course:** Bailey, T. (2024). Ultimate SnowPro Core Certification Course & Exam. Udemy. Link: **https://www.udemy.com/course/ultimate-snowpro-core-certification-course-exam/**
- **Foundational Research Paper (The Snowflake Origin):** Dageville, B., Cruanes, T., et al. (2016). The Snowflake Elastic Data Warehouse. Proceedings of the 2016 International Conference on Management of Data (SIGMOD).
- **U.S. Department of Education, Federal Student Aid. (2024)**. *Federal Student Aid Fiscal Year 2024 Annual Report*. Washington, D.C. Link: **https://studentaid.gov/strategic-planning-reporting**

Thank you!

Questions/Suggestions welcomed.