SNOWFLAKE & ACADEMIC MEDICAL RESEARCH

Accelerating research with the AI Data Cloud

Greater Plains Collaborative Learning Engagement Conference
August 2024
Paul Boal, Snowflake Healthcare and Life Sciences

Safe Harbor and Disclaimers

Other than statements of historical fact, all information contained in these materials and any accompanying oral commentary (collectively, the “Materials”), including statements regarding (i) Snowflake’s business strategy, plans or priorities, (ii) Snowflake’s new or enhanced products, services, and technology offerings, including those that are under development or not generally available, (iii) market growth, trends, and competitive considerations, (iv) our vision for Snowpark, the AI Data Cloud, and industry-specific AI Data Clouds, including the expected benefits and network effects of the AI Data Cloud; and (v) the integration, interoperability, and availability of Snowflake’s products, services, or technology offerings with or on third-party platforms or products, are forward-looking statements. These forward-looking statements are subject to a number of risks, uncertainties and assumptions, including those described under the heading “Risk Factors” and elsewhere in the Annual Reports on Form 10-K and the Quarterly Reports on Form 10-Q that Snowflake files with the Securities and Exchange Commission. In light of these risks, uncertainties, and assumptions, the future events and trends discussed in the Materials may not occur, and actual results could differ materially and adversely from those anticipated or implied in the forward-looking statements. As a result, you should not rely on any forwarding-looking statements as predictions of future events. Forward-looking statements speak only as of the date the statements are first made and are based on information available to us at the time those statements are made and/or management's good faith belief as of that time. Except as required by law, we undertake no obligation, and do not intend, to update the forward-looking statements in these Materials.

Any future product or roadmap information (collectively, the “Roadmap”) is intended to outline general product direction. The Roadmap is not a commitment, promise, or legal obligation for Snowflake to deliver any future products, features, or functionality; and is not intended to be, and shall not be deemed to be, incorporated into any contract. The actual timing of any product, feature, or functionality that is ultimately made available may be different from what is presented in the Roadmap. The Roadmap information should not be used when making a purchasing decision. In case of conflict between the information contained in the Materials and official Snowflake documentation, official Snowflake documentation should take precedence over these Materials. Further, note that Snowflake has made no determination as to whether separate fees will be charged for any future products, features, and/or functionality which may ultimately be made available. Snowflake may, in its own discretion, choose to charge separate fees for the delivery of any future products, features, and/or functionality which are ultimately made available.

The Materials may contain information provided by third-parties. Snowflake has not independently verified this information, and usage of this information does not mean or imply that Snowflake has adopted this information as its own or independently verified its accuracy.

Barriers to faster progress in academic medical research

Access to Data
69%

Skills / Expertise
74%

Scalable Technology

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8417604/

STRENGTHEN YOUR DATA FOUNDATION

Unify access to all data types and analytics in a single, scalable, secure, collaborative platform.

ACCELERATE ENTERPRISE AI

Leverage existing and readily accessible expertise like SQL, Python, R, and open LLMs for advanced analytics.

BUILD & DISTRIBUTE APPLICATIONS

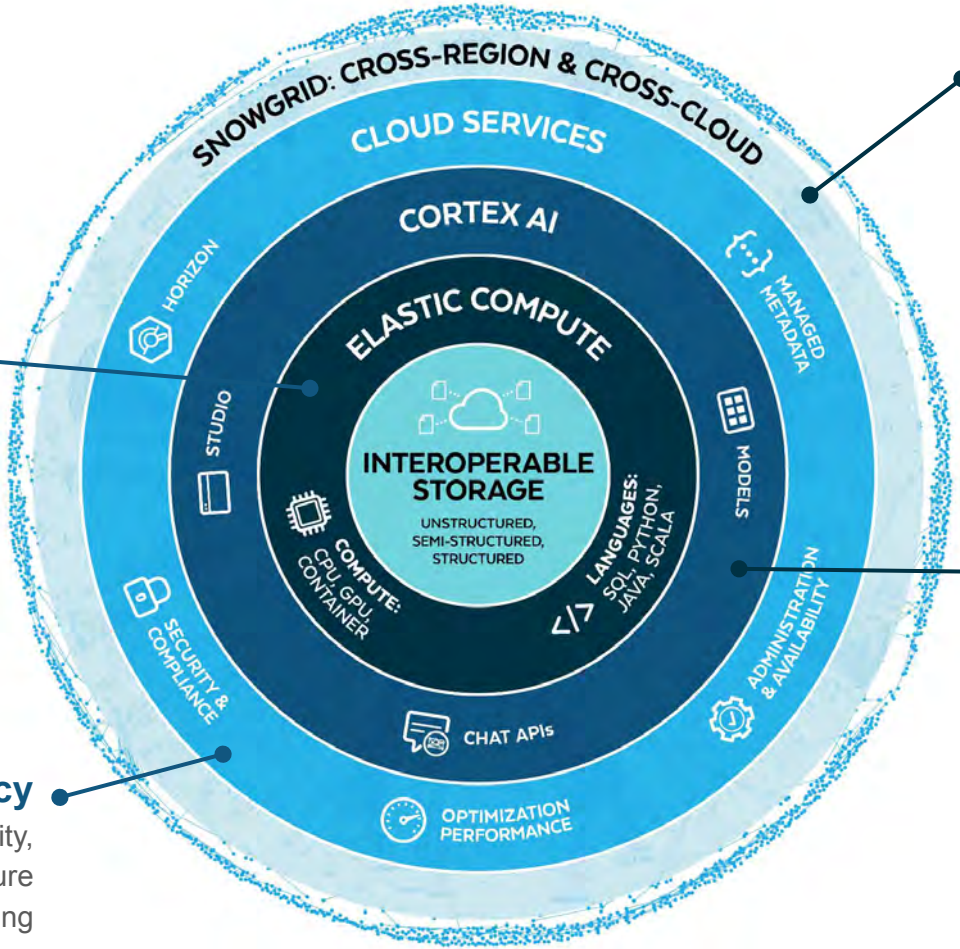
Share data and analytical algorithms while maintaining data privacy and protecting intellectual property rights.

Affordable Scalability

Built in the cloud for elastic scalability and response to change without large capital purchases

Security and Privacy

Are Snowflake's top priority, regardless of the feature you're using



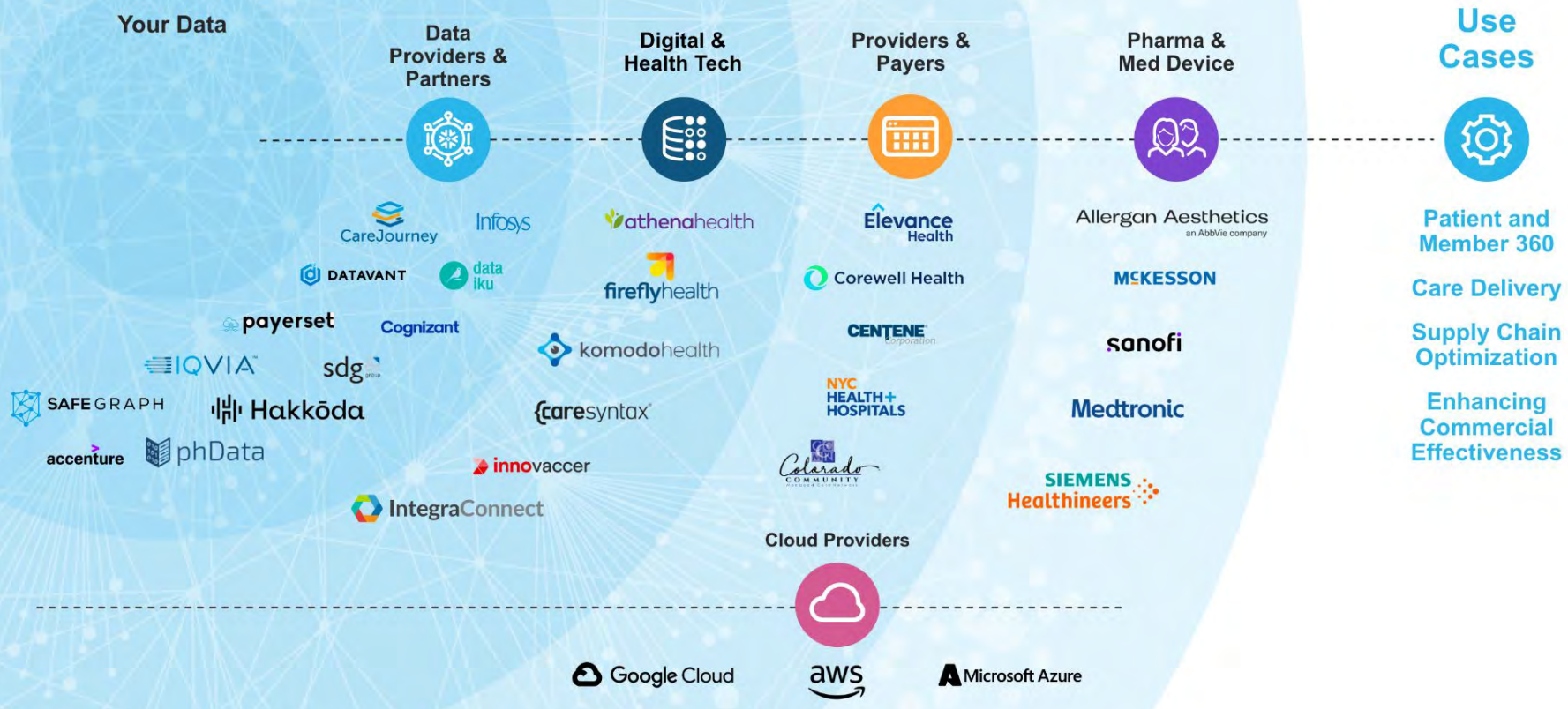
Collaboration Built-In

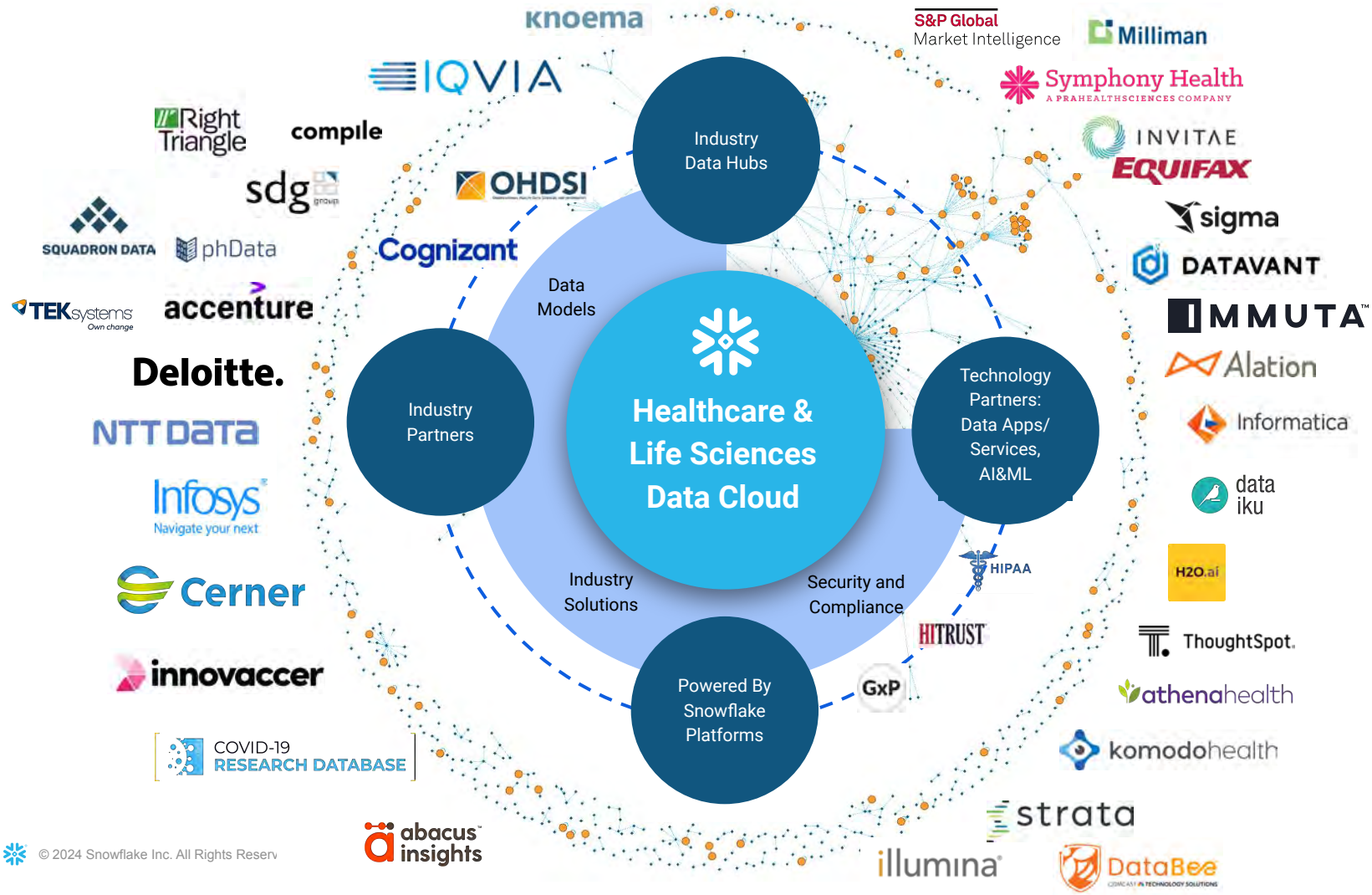
Snowflake customers can seamlessly and securely share data without time-consuming export/import processes

Advanced Analytics

Built into Snowflake are statistics and traditional machine learning via SQL and Python to hosted LLMs for secure GenAI applications

Data Available in Snowflake





Academic Medical Centers & Research Entities

python R OHDSI i2b2 sas Streamlit

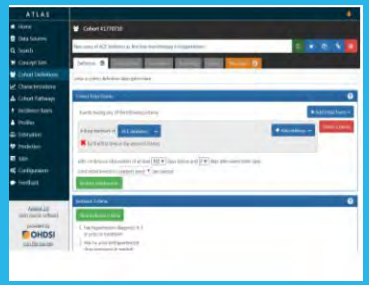
Research Objectives

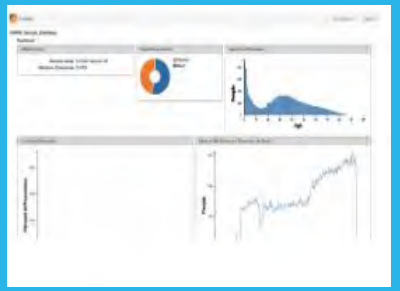
- Deepen our understanding of diseases
- Enhance our ability to predict clinical outcomes
- Accelerate development of next generation therapeutics
- Pioneer new approaches to engagement in the clinic and beyond

Snowflake Marketplace Data Providers

Snowflake Data Clean Rooms

OMOP and OHDSI Ecosystem Supported by the Snowflake Platform





ATLAS Version 2.11

ACHILLES

ATLAS/WebAPI

DQD - Data Quality Dashboard

HADES (OHDSI Analytical Methods)

Snowflake export as a SQL dialect in ATLAS

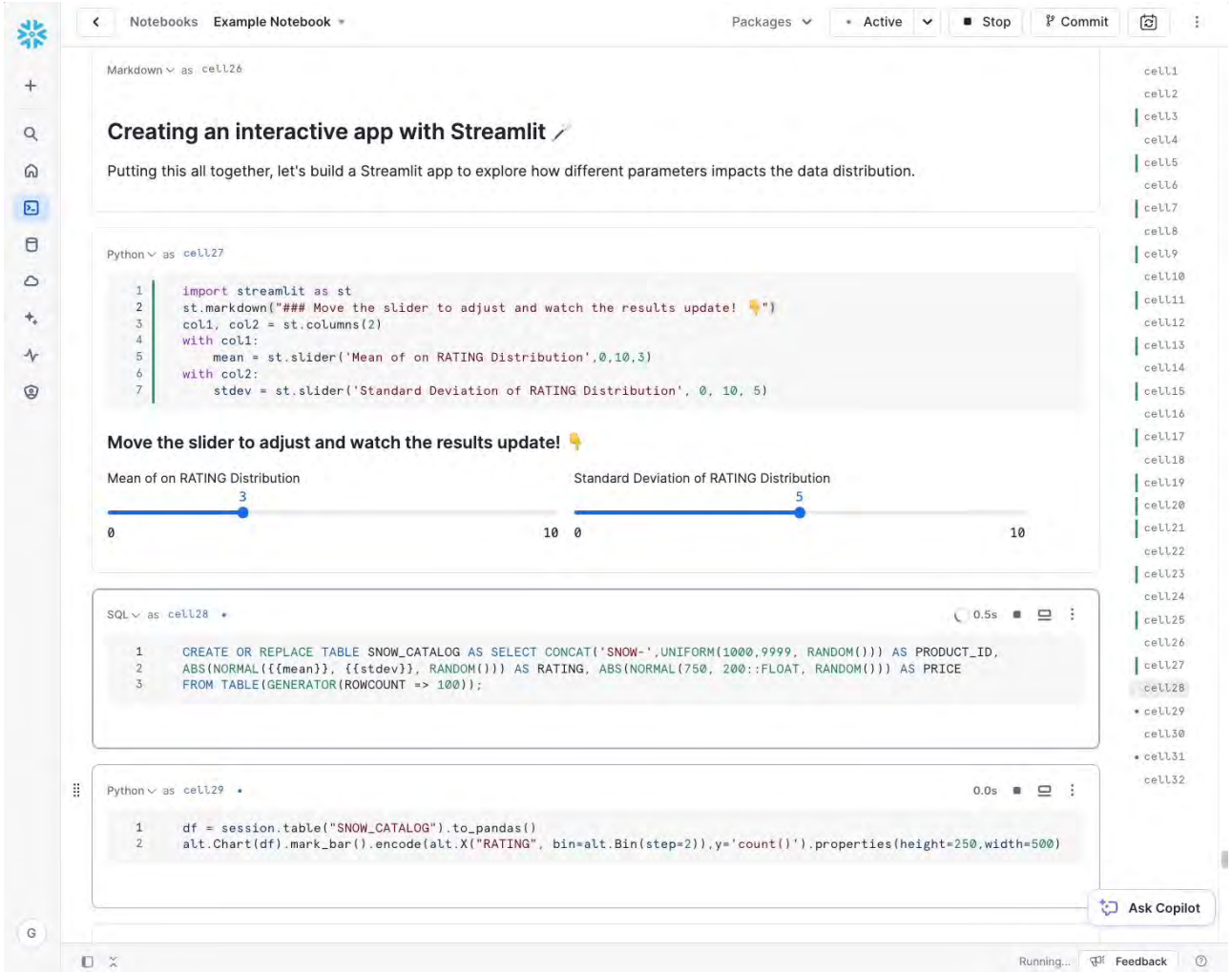
GPC Running i2b2 on Snowflake

Abu Mosa, AMIA 2024 Informatics Summit

- 32 million patients
- 1.3 billion Encounters
- 25 billion clinical observations
- ENACT Ontology v4.1
- 13 sites from GPC
  - Data is distributed in 13 SnowFlake databases
  - Shared in PCORnet CDM format
  - i2b2 executes queries against PCORnet CDM data from 13 sites over SnowFlake secure views and shares, and through harmonization scripts (secure views)

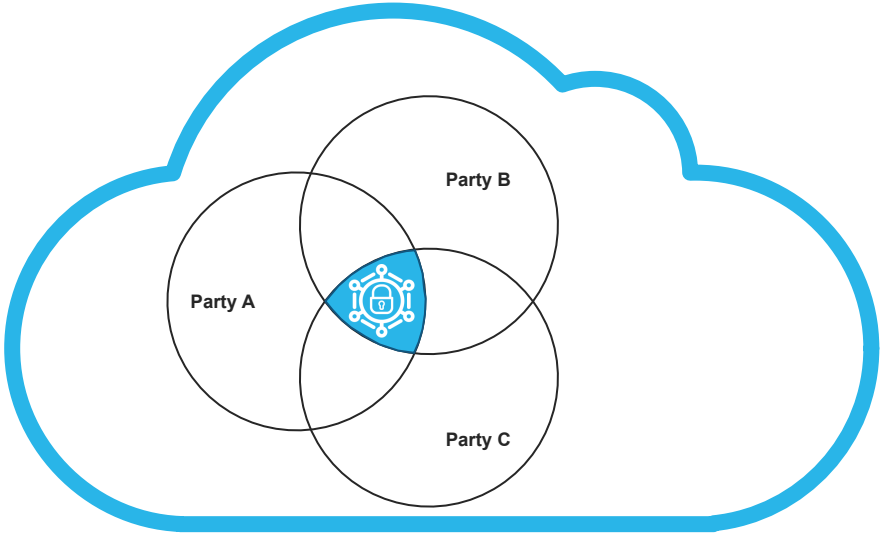
Snowflake Notebooks

- SQL + Python
- Collaborative
- Git Integration
- Pandas
- Stats
- ML
- AI



Snowflake Data Clean Rooms

A secure environment where multiple parties can collaborate on sensitive or regulated data without exposing or moving the underlying data



Snowflake Data Clean Rooms

Accelerating data clean room innovations for Snowflake Customers

SNOWFLAKE NATIVE APP FRAMEWORK

Never Leave the AI Data Cloud

The native app framework allows applications to run directly in your Snowflake account, without data movement or copy.

SNOWFLAKE HORIZON

Compliance | Security | Access | Privacy | Interoperability

Built-in governance solution with a unified set of compliance, security, privacy, interoperability and access capabilities in the AI Data Cloud.

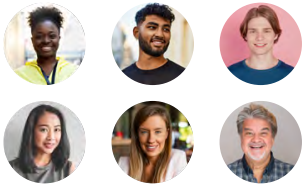
CROSS-CLOUD SNOWGRID

Cross-Cloud Collaboration

Interconnect your business' ecosystems across regions and clouds with a cross-cloud technology layer that lets you operate at global scale

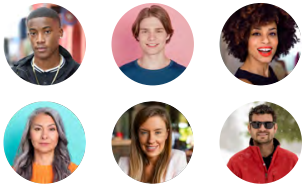
Cohort Builder

Cohort 01



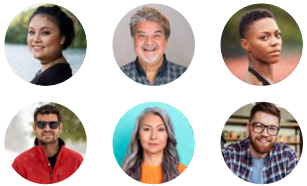
- ☒ Diabetes
- ☒ # PCP Visit > 4
- ☒ Open Care Gap: failure to follow prescription guidelines

Cohort 02



- ☒ Vascular
- ☒ Smoking
- ☒ Open Care Gap: missed vaccination screening

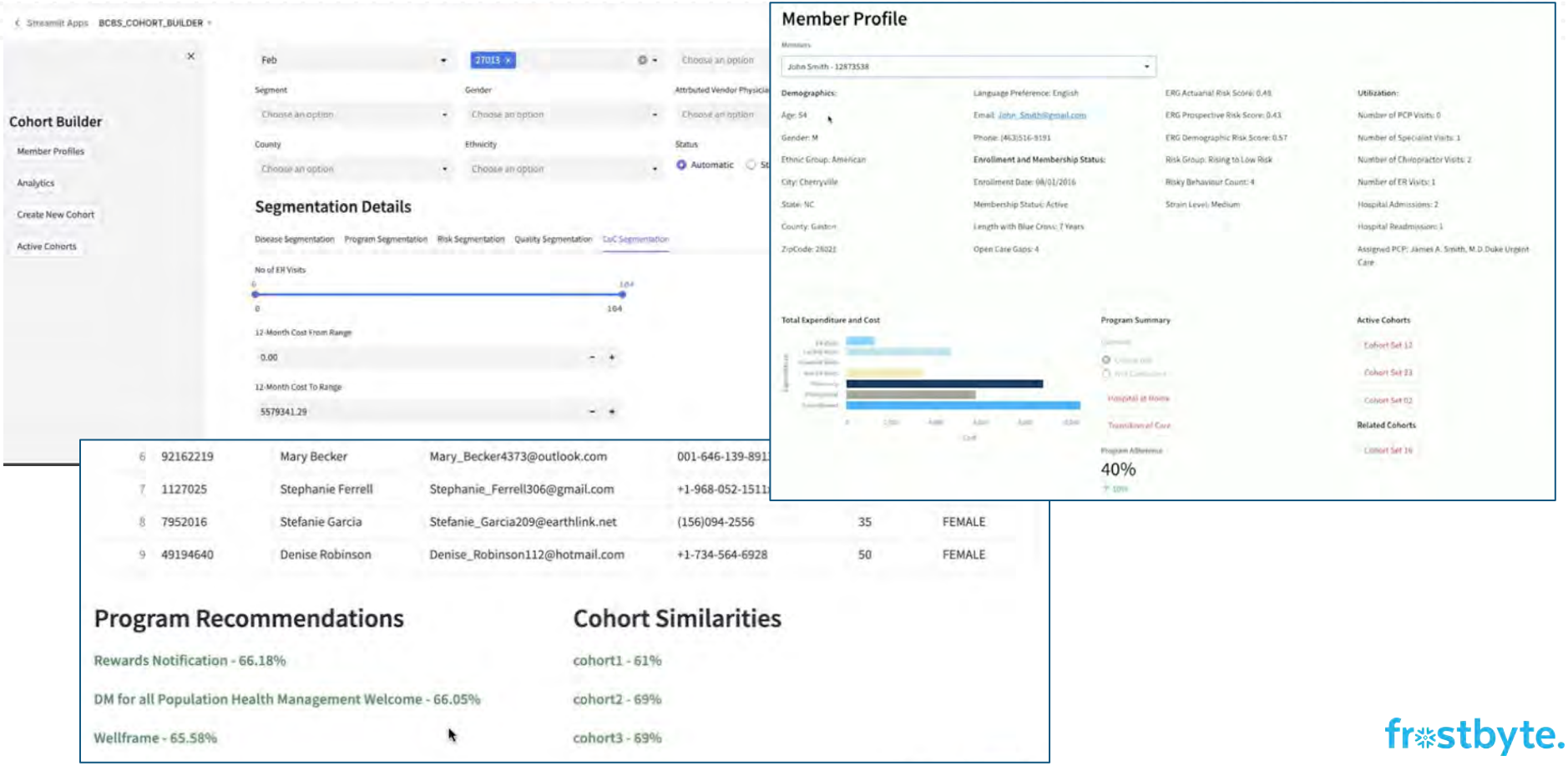
Cohort 03



- ☒ Cardiac
- ☒ # ER Visit > 4
- ☒ Open Care Gap: can't access the care they need in their area

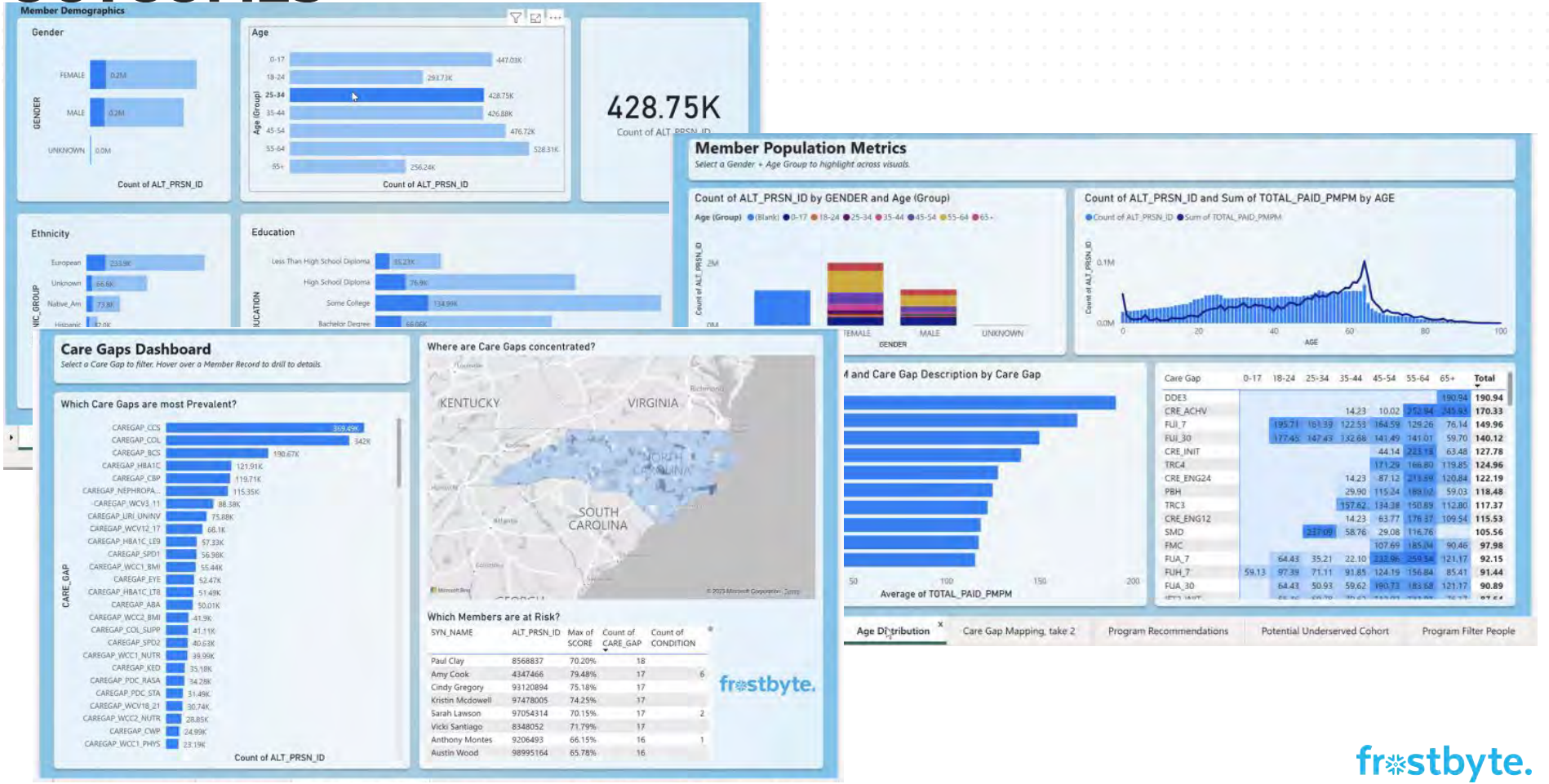
frostbyte.

Cohort Self-Service Design for the Business



frostbyte.

COHORT INSIGHTS | AI-PROGRAM DRIVEN OUTCOMES



Program Recommendation

Cohort 01

Program Recommendation:

- *Eat smart move more prevent diabetes - 70%*
- *Care Gap Reminders Diabetes*
- *Diabetes Management Discharge Program*

Cohort 02

Program Recommendation:

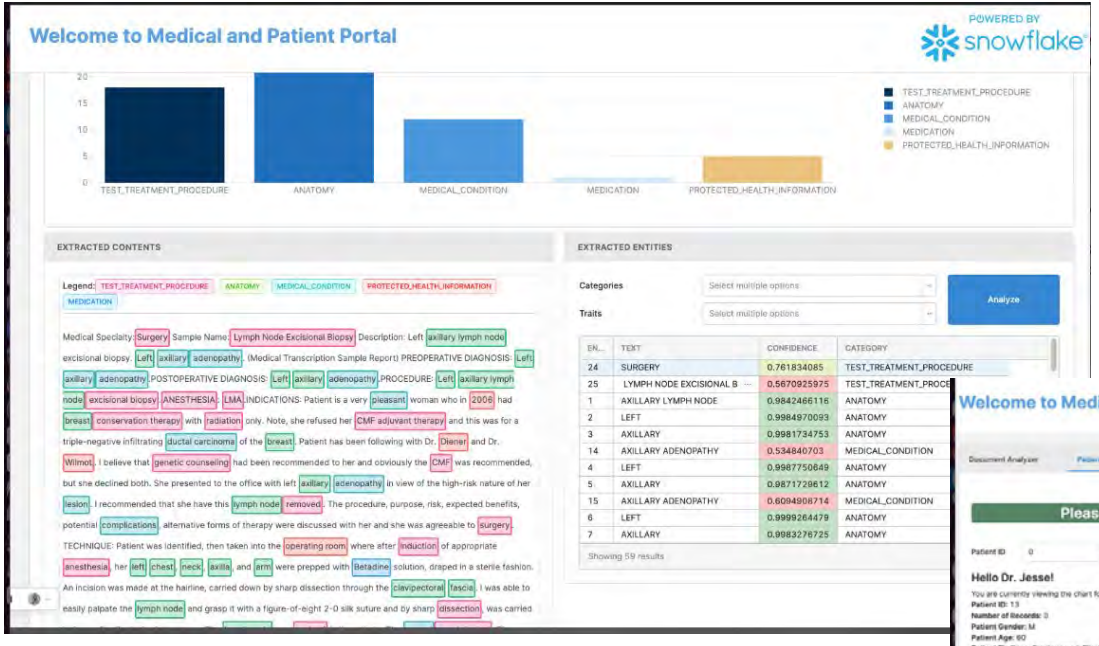
- *Telehealth Program*
- *Smoking Quitline*
- *Get Up and Get Moving*

Cohort 03

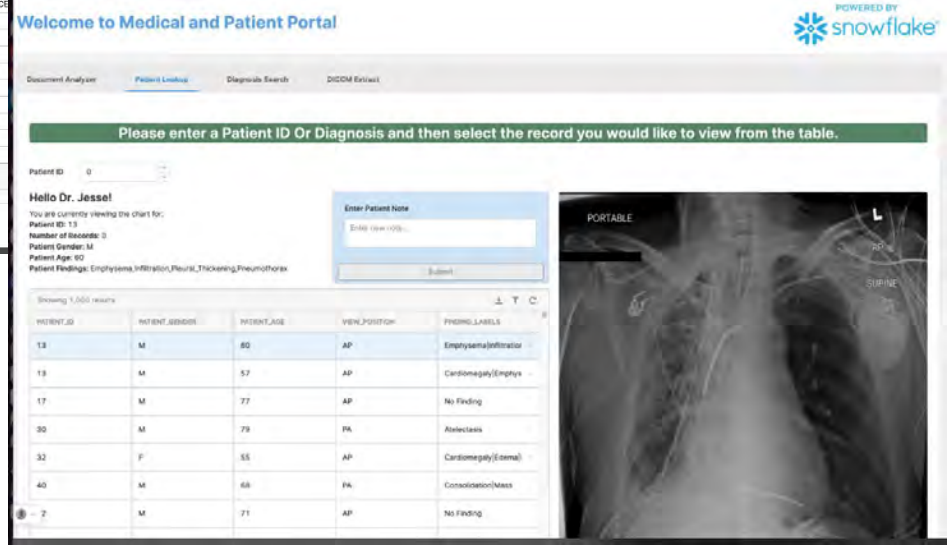
Program Recommendation:

- *Dental Health Integration*
- *Coronary Artery Disease Discharge Program*

Unstructured Data in Snowflake - unified platform data access



- Raw PDF notes and DICOM images imported to Snowflake
- Extract text and NLP/NER process
- Compare data to AWS Comprehend Medical API for insights/confidence
- Tables populated - data analysis ready; original objects still available



Using Snowflake to *Understand* Research

Ingest PubMed and other clinical knowledge using Snowflake LLM & RAG.

Query in natural language with a chatbot that shows its sources.

W&B Fully Connected > Articles > GenAI

How to create a biomedical RAG application using Snowflake Arctic for PubMed paper understanding

A tutorial about building a RAG application to better understand a large corpus of medical information

Anish Shah

Share Comment 2 stars

Created on August 13 | Last edited on August 15

You can check out the code for this project and see it in W&B Weave by following these links:
Code: https://github.com/ash0ts/snowflake-arctic-weave-demo
Weave: https://wandb.ai/a-sh0ts/bioasq_example/weave/traces

Introduction

Imagine a bustling hospital where clinicians need to make quick, accurate decisions for complex patient cases, like a pediatrician handling a rare genetic disorder in a newborn. This pediatrician needs to understand genetic factors, treatment options, and the latest research findings. However, sifting through the massive volume of biomedical literature on PubMed to find relevant information can be onerous.

In this context, a retrieval-augmented generation (RAG) system integrated with Snowflake Arctic can be invaluable. Here's how:

1. **Time-sensitive information retrieval:** The pediatrician inputs a clinical question like "Is Hirschsprung disease a Mendelian or

https://wandb.ai/wandb_fc/snowflake_arctic_rag/reports/How-to-create-a-biomedical-RAG-application-using-Snowflake-Arctic-for-PubMed-paper-understanding--Vmlldzo5MDIxNTk5

CUSTOMER SPOTLIGHT

The Greater Plains Collaborative (GPC) is a PCORnet®-funded clinical research network of 14 medical centers providing care in 14 states (CA, CO, IA, ID, IL, KS, MN, MT, MO, NE, TX, UT, WI, WY) committed to a shared vision of improving healthcare delivery through ongoing learning, adoption of evidence-based practices and active research dissemination. The GPC helps its members generalize their research to contribute to national studies using the PCORnet Common Data Model (CDM) to hold health records and tumor registries from its 35 million patients, securely linking statewide Medicare and Medicaid claims (25 million beneficiaries) to incorporate patients' healthcare experiences outside their medical centers, and the i2b2 platform for cohort analysis. This framework enables the secure sharing, integration, standardization and analysis of heterogeneous data from healthcare and research settings. However, the i2b2 platform has a complex architecture and can be challenging and costly to deploy and maintain.

GPC used Snowflake and AWS to s
integration, deployment, maintenar
analytics across its members' datab
an on-premises configuration that
deployment and maintenance of m
time-consuming and costly endeav
AWS deployments enabled serverl
of i2b2, enabling **GPC's members t
needed and pay for only what they
service, Snowflake also provides a
management, eliminating the time
managing data warehouse infrastr
members and external national inv
benefit from faster query processi
vertical and horizontal scaling of v
without disruption.** Snowflake sim
process for GPC members, allowing
easier and cost-effective collaborat
research projects. Snowflake also e
GPC to combine data sources from
institutions into one united view, m
scientists to access data for their c

Snowflake's been very supportive in helping us realize our vision for advancing patient-centered clinical research for the nation. Snowflake's high compression, dynamic ability to manage compute, secure views to protect patient privacy, price transparency, and nimbleness in supporting regulatory and contractual processes has allowed us to parsimoniously incorporate large data sources and scale analyses that wouldn't be possible previously. Snowflake's continued innovations and secure data sharing across regions and cloud platforms are now enabling i2b2 queries across multiple clinical research networks in PCORnet."

- RUSS WAITMAN,
Ph.D., FACMI, FAMIA, Associate Dean for Informatics at the University of MIssiouri School of Medicine
(Member of the Greater Plains Collaborative)

Snowflake - A Business & Technology Partner

Resources to Connect Your Business Objectives, Data Strategy, & Solutions

Snowflake for Academia

Enhanced Curriculum

Receive enhanced curriculum from Snowflake

Industry Readiness

Better industry readiness for students

Access

Opportunities for collaborations between Snowflake and Universities (research, career development)

Exposure

Access to leading-edge and industry-grade data platform and tools

Development

Better career development and job prospects

Networking

Community and networking opportunities

THANK YOU

paul.boal@snowflake.com

snowflake®