



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

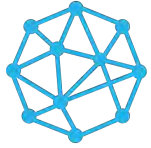
Other than statements of historical fact, all information contained in these materials and any accompanying oral commentary (collectively, the “Materials”), including statements regarding (i) Snowflake’s business strategy, plans or priorities, (ii) Snowflake’s new or enhanced products, services, and 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

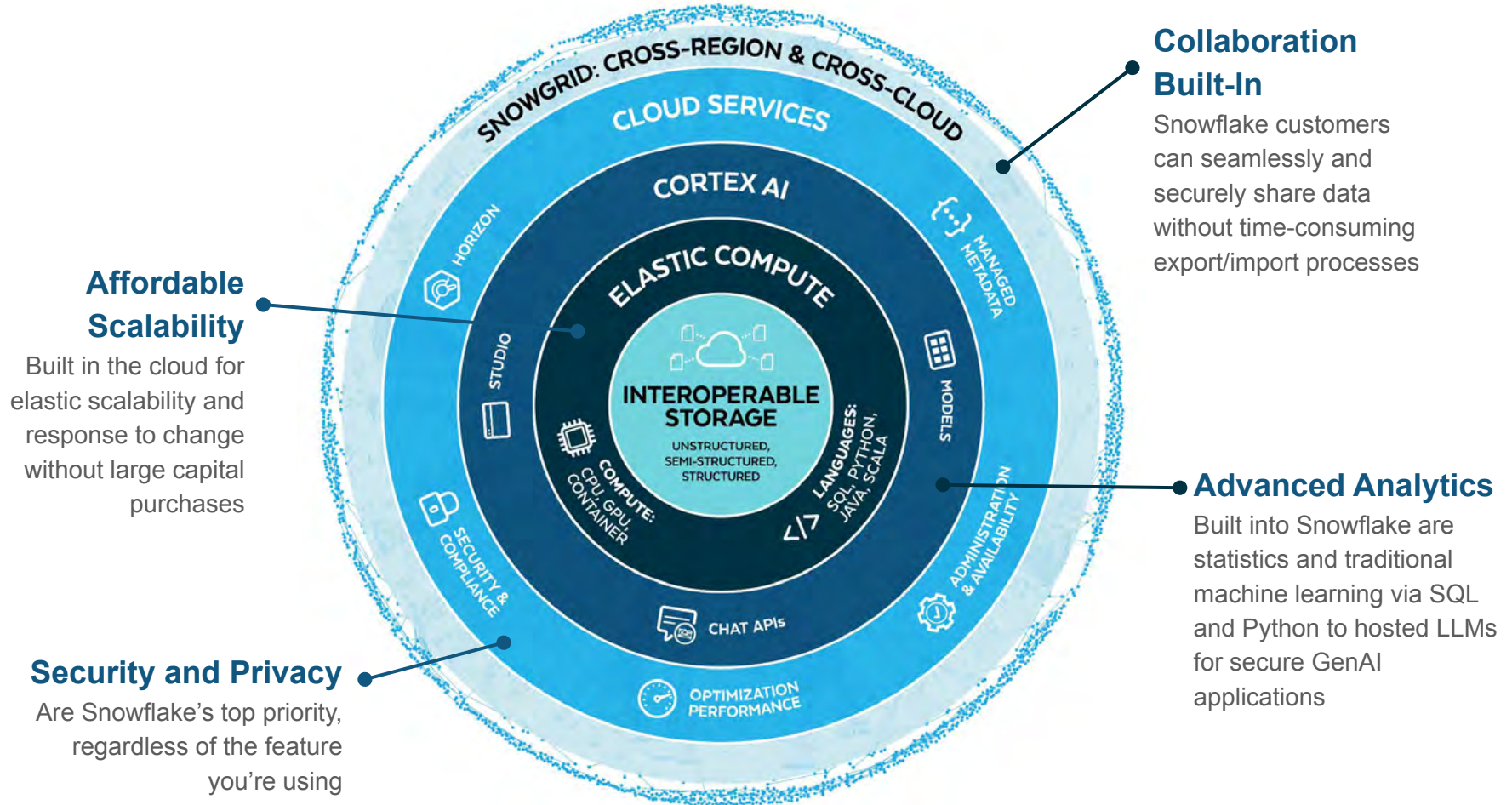
Leverage existing and readily accessible expertise like SQL, Python, R, and open LLMs for advanced analytics.



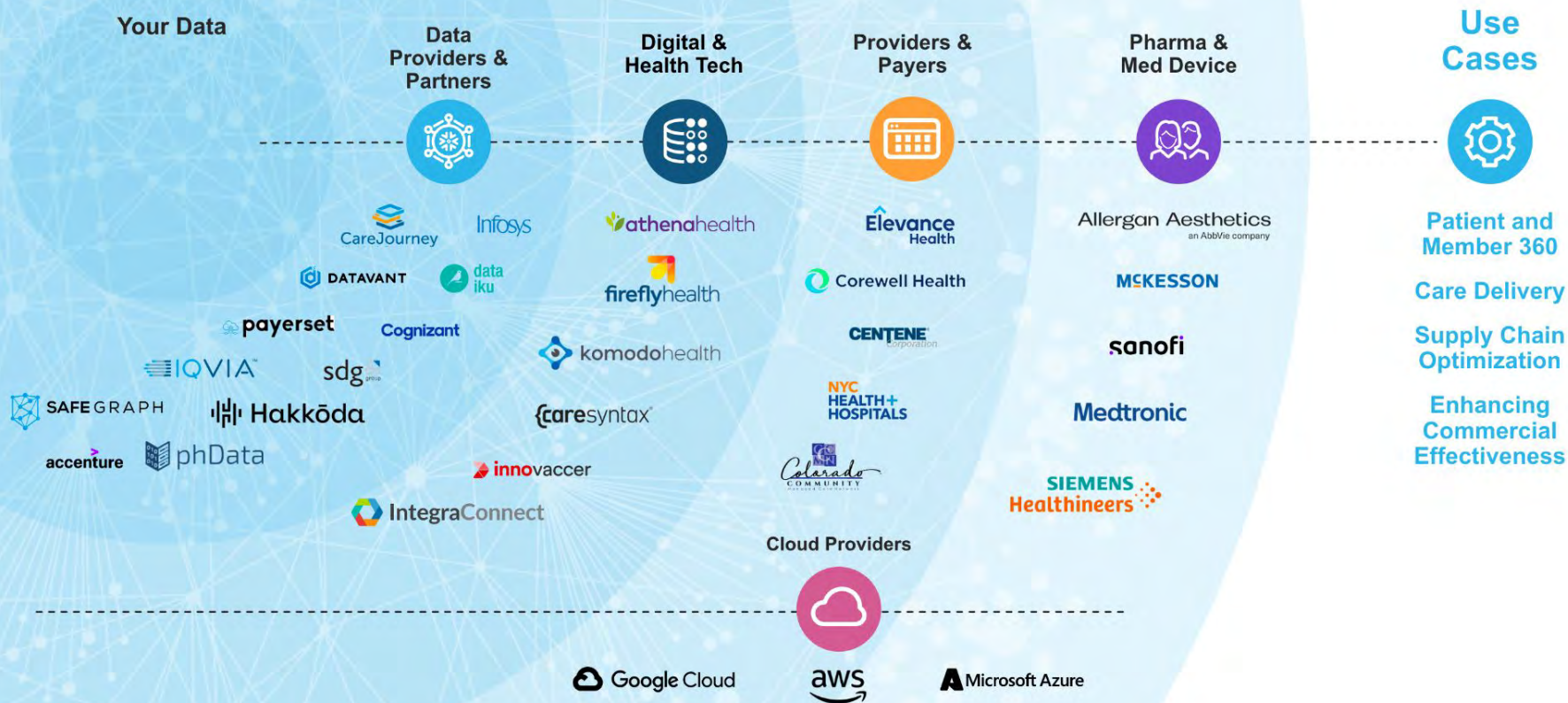
BUILD & DISTRIBUTE APPLICATIONS

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





Data Available in Snowflake



**Academic Medical Centers
& Research Entities**

Provide
Direct Share

BYOBI -
Analyze with
Tools of
Choice

TRUSTED RESEARCH ENVIRONMENT

Snowflake
Marketplace
Resources

Privacy
Preservation

**Snowflake Marketplace
Data Providers**



i2b2



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



i2b2

**Snowflake Data
Clean Rooms**

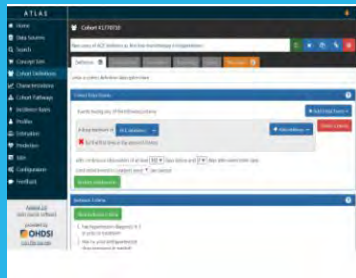


OMOP and OHDSI Ecosystem Supported by the Snowflake Platform



ATLAS Version 2.11

**DQD - Data Quality
Dashboard**



ACHILLES



**HADES (OHDSI
Analytical Methods)**



ATLAS/WebAPI

**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

The screenshot displays the Snowflake Notebooks interface with a sidebar on the left containing navigation icons. The main workspace is titled "Example Notebook" and contains three cells:

- Cell 26 (Markdown):** Titled "Creating an interactive app with Streamlit". It contains the text: "Putting this all together, let's build a Streamlit app to explore how different parameters impacts the data distribution."
- Cell 27 (Python):** Contains Python code for a Streamlit app:

```
1 import streamlit as st
2 st.markdown("### Move the slider to adjust and watch the results update! 🚀")
3 col1, col2 = st.columns(2)
4 with col1:
5     mean = st.slider('Mean of on RATING Distribution', 0, 10, 3)
6 with col2:
7     stdev = st.slider('Standard Deviation of RATING Distribution', 0, 10, 5)
```

Below the code, the app's UI is shown with two sliders: "Mean of on RATING Distribution" (set to 3) and "Standard Deviation of RATING Distribution" (set to 5).
- Cell 28 (SQL):** Contains SQL code to create a table:

```
1 CREATE OR REPLACE TABLE SNOW_CATALOG AS SELECT CONCAT('SNOW-', UNIFORM(1000, 9999, RANDOM())) AS PRODUCT_ID,
2 ABS(NORMAL([mean], [stdev], RANDOM())) AS RATING, ABS(NORMAL(750, 200::FLOAT, RANDOM())) AS PRICE
3 FROM TABLE(GENERATOR(ROWCOUNT => 100));
```
- Cell 29 (Python):** Contains Python code to load the data and create a chart:

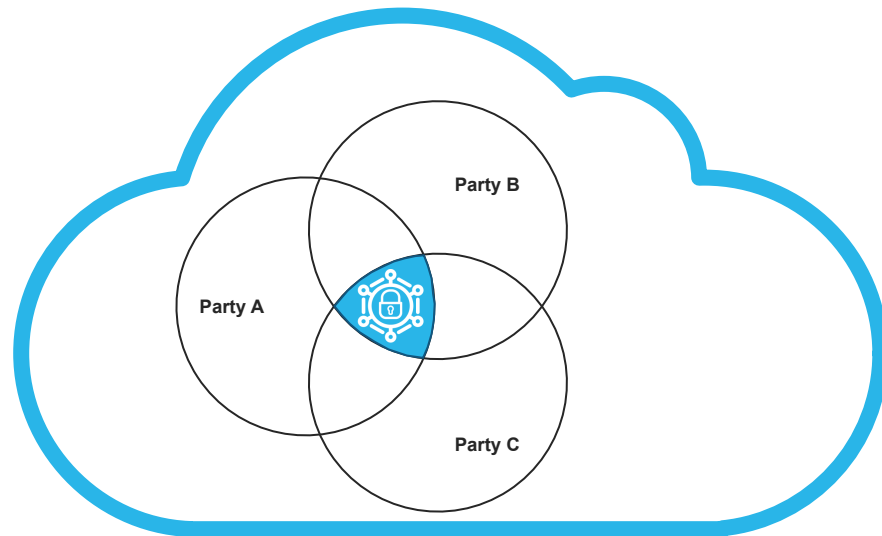
```
1 df = session.table("SNOW_CATALOG").to_pandas()
2 alt.Chart(df).mark_bar().encode(alt.X("RATING", bin=alt.Bin(step=2)), y='count()', properties(height=250, width=500))
```

At the bottom right, there is an "Ask Copilot" button. The status bar at the very bottom indicates "Running..." and includes a "Feedback" link.



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



Encrypted Snowpark
for AI/ML



Cryptographic Compute
Support



Differential Privacy



Attestation Security
Guarantees



Strong Public Key
Cryptosystem Integration



Data & Activation
Integrations

ANCHORED BY CORE SNOWFLAKE FEATURES

Snowflake Native App Framework
Snowflake Horizon
Cross-Cloud Snowgrid

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

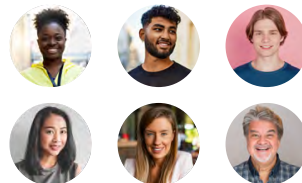


Cohort Builder

Total Member Population



Cohort 01



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



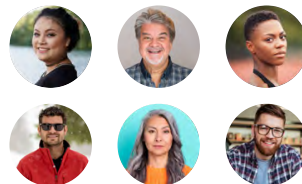
Cohort 02



- ❑ 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

Cohort Self-Service Design for the Business

Streamlit Apps

BCBS_COHORT_BUILDER

Cohort Builder

Member Profiles

Analytics

Create New Cohort

Active Cohorts

Feb

27013

Choose an option

Segment

Gender

Choose an option

Choose an option

County

Ethnicity

Choose an option

Choose an option

Status

Automatic

Segmentation Details

Disease Segmentation

Program Segmentation

Risk Segmentation

Quality Segmentation

CAC Segmentation

No of ER Visits

12-Month Cost From Range

0.00

12-Month Cost To Range

5579341.29

6	92162219	Mary Becker	Mary_Becker4373@outlook.com	001-646-139-891
7	1127025	Stephanie Ferrell	Stephanie_Ferrell306@gmail.com	+1-968-052-1511
8	7952016	Stefanie Garcia	Stefanie_Garcia209@earthlink.net	(156)094-2556
9	49194640	Denise Robinson	Denise_Robinson112@hotmail.com	+1-734-564-6928

Member Profile

Members

John Smith - 12873538

Demographics:

Age: 54

Gender: M

Ethnic Group: American

City: Cherryville

State: NC

County: Gaston

ZipCode: 28021

Language Preference: English

Email: John_Smith@gmail.com

Phone: (463)316-9191

Enrollment and Membership Status:

Enrollment Date: 06/01/2016

Membership Status: Active

Length with Blue Cross: 7 Years

Open Care Gaps: 4

ERG Actuarial Risk Score: 0.48

ERG Prospective Risk Score: 0.43

ERG Demographic Risk Score: 0.57

Risk Group: Rising to Low Risk

Risky Behaviour Count: 4

Strain Level: Medium

Utilization:

Number of PCP Visits: 0

Number of Specialist Visits: 1

Number of Chiropractor Visits: 2

Number of ER Visits: 1

Hospital Admissions: 2

Hospital Readmissions: 1

Assigned PCP: James A. Smith, M.D. Duke Urgent Care

Total Expenditure and Cost

Program Summary

Program Adherence

40%

Program Recommendations

- Rewards Notification - 66.18%
- DM for all Population Health Management Welcome - 66.05%
- Wellframe - 65.58%

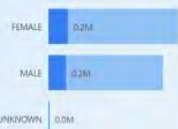
Cohort Similarities

- cohort1 - 61%
- cohort2 - 69%
- cohort3 - 69%

COHORT INSIGHTS | AI-PROGRAM DRIVEN OUTCOMES

Member Demographics

GENDER



Count of ALT_PRSN_ID

Age



Count of ALT_PRSN_ID

428.75K

Count of ALT_PRSN_ID

Member Population Metrics

Select a Gender + Age Group to highlight across visuals.

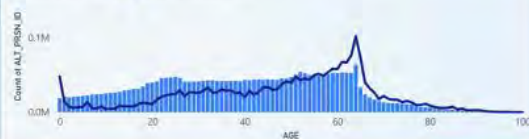
Count of ALT_PRSN_ID by GENDER and Age (Group)

Age (Group) (Blank) 0-17 18-24 25-34 35-44 45-54 55-64 65+



Count of ALT_PRSN_ID and Sum of TOTAL_PAID_PMPM by AGE

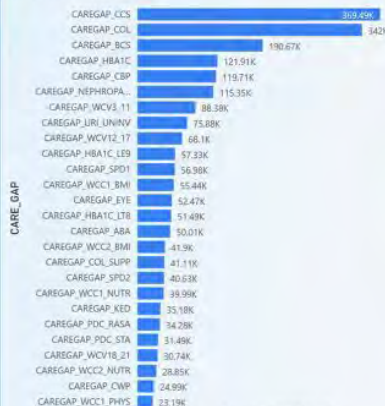
Count of ALT_PRSN_ID Sum of TOTAL_PAID_PMPM



Care Gaps Dashboard

Select a Care Gap to filter. Hover over a Member Record to drill to details.

Which Care Gaps are most Prevalent?



Count of ALT_PRSN_ID

Where are Care Gaps concentrated?

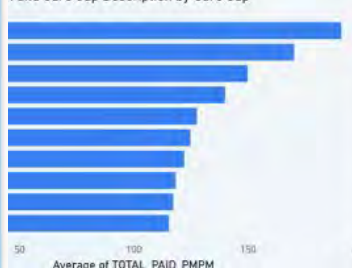


Which Members are at Risk?

SYN_NAME	ALT_PRSN_ID	Max of SCORE	Count of CARE_GAP	Count of CONDITION
Paul Clay	8568837	70.20%	18	
Amy Cook	4347466	79.48%	17	6
Cindy Gregory	93120894	75.18%	17	
Kristin McDowell	97478005	74.25%	17	
Sarah Lawson	97054314	70.15%	17	2
Vicki Santiago	8348052	71.79%	17	
Anthony Montes	9206493	66.15%	16	1
Austin Wood	98995164	65.78%	16	

frostbyte.

4 and Care Gap Description by Care Gap



Average of TOTAL_PAID_PMPM

Care Gap	0-17	18-24	25-34	35-44	45-54	55-64	65+	Total
DDE3							190.94	190.94
CRE_ACHV			14.23	10.02	252.34	245.83	170.33	170.33
FUL_7	195.71	161.39	122.53	164.59	129.26	76.14	149.96	149.96
FUL_30	177.45	147.43	132.68	141.49	141.01	59.70	140.12	140.12
CRE_INIT				44.14	233.18	63.48	127.78	127.78
TRC4				171.29	169.80	119.85	124.96	124.96
CRE_ENG24			14.23	87.12	213.94	126.84	122.19	122.19
PBH			29.90	115.34	189.06	59.03	118.48	118.48
TRC3				157.62	134.38	150.89	112.80	112.80
CRE_ENG12			14.23	63.77	176.37	109.54	115.53	115.53
SMD			217.09	58.76	29.08	116.76	105.56	105.56
FMC					107.69	185.04	90.46	97.98
FUA_7			64.43	35.21	22.10	232.96	259.36	121.17
FUL_7	59.13	97.39	71.11	91.85	124.19	156.84	85.41	91.44
FUA_30	64.43	50.93	59.62	190.73	183.68	121.17	90.89	90.89

Age Distribution

Care Gap Mapping, take 2

Program Recommendations

Potential Underserved Cohort

Program Filter People




frostbyte.

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*

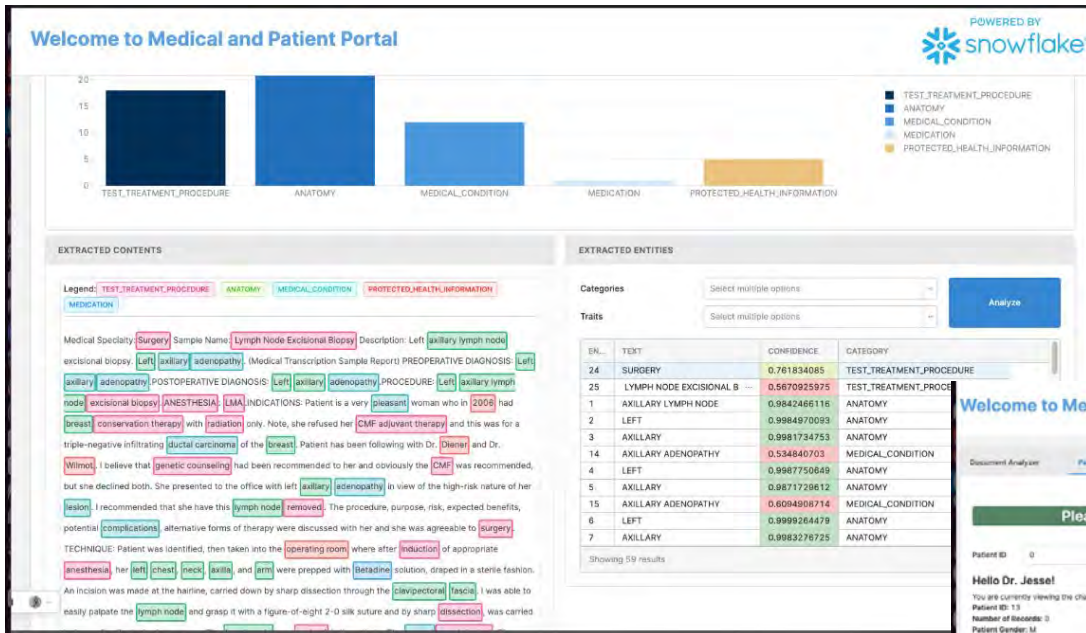
Cohort 03



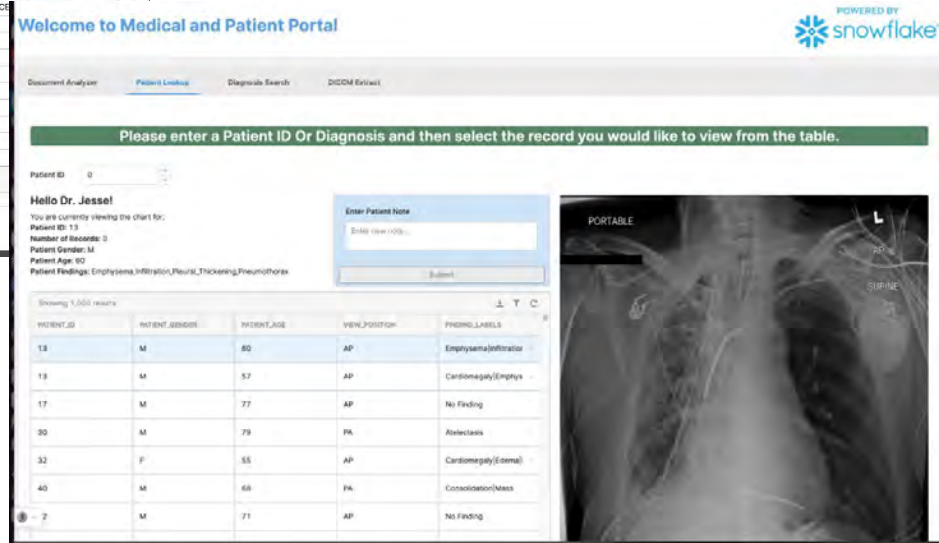
Program Recommendation:

-  *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



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 streamline the integration, deployment, maintenance and management of analytics across its members' data. Moving from an on-premises configuration that required significant deployment and maintenance of multiple servers to a time-consuming and costly endeavor, Snowflake's AWS deployments enabled serverless architectures of i2b2, enabling GPC's members to scale as needed and pay for only what they use. In addition to service, Snowflake also provides a managed environment for management, eliminating the time and effort of managing data warehouse infrastructure. This has benefited members and external national investigators by benefit from faster query processing and the ability for vertical and horizontal scaling of compute resources without disruption. Snowflake simplifies the data processing for GPC members, allowing for easier and cost-effective collaborative research projects. Snowflake also enables GPC to combine data sources from multiple institutions into one united view, making it easier for scientists to access data for their clinical research.

“

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 Missouri 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



© 2024 Snowflake Inc. All Rights Reserved