



SUMMIT 2023



SUMMIT 2023

Powering Centene's real-time enterprise data analytics with Upsolver

Santona Tuli | Director of Data
Alexander Adam | Manager, Cloud Platform Engineering

February 23, 2023



Santana Tuli
Director of Data
Upsolver



Alexander Adam
Manager, Cloud Platform Engineering
Centene Corporation



AGENDA

- **Intro to Upsolver**
High quality, observable data ingestion at scale
- **Intro to Centene**
Largest Medicaid managed care organization
- **Centene data architecture**
A perfect match for Upsolver and Snowflake
- **Why Centene chose Upsolver**
Benefits of managing data quality at ingestion



About Upsolver



Data quality at ingestion



A cloud data ingestion platform, purpose-built to overcome the quality challenges posed by complex data (streams, files, CDC).

Certifications



Partnerships



High quality, observable data at work

Over **10,000** production pipelines

Processing **petabytes per day**

Across multiple industries in **enterprise, midsize, and startups**

Cox
AUTOMOTIVE™

ConocoPhillips

CENTENE®
Corporation

proofpoint.

AppsFlyer

SISENSE

is ironSource
FROM UNITY

yotpo.

Gem

flywheel

PEER39

avantis

SimilarWeb

herdwatch
simplifying farming.

Clicktripz

Bigabid

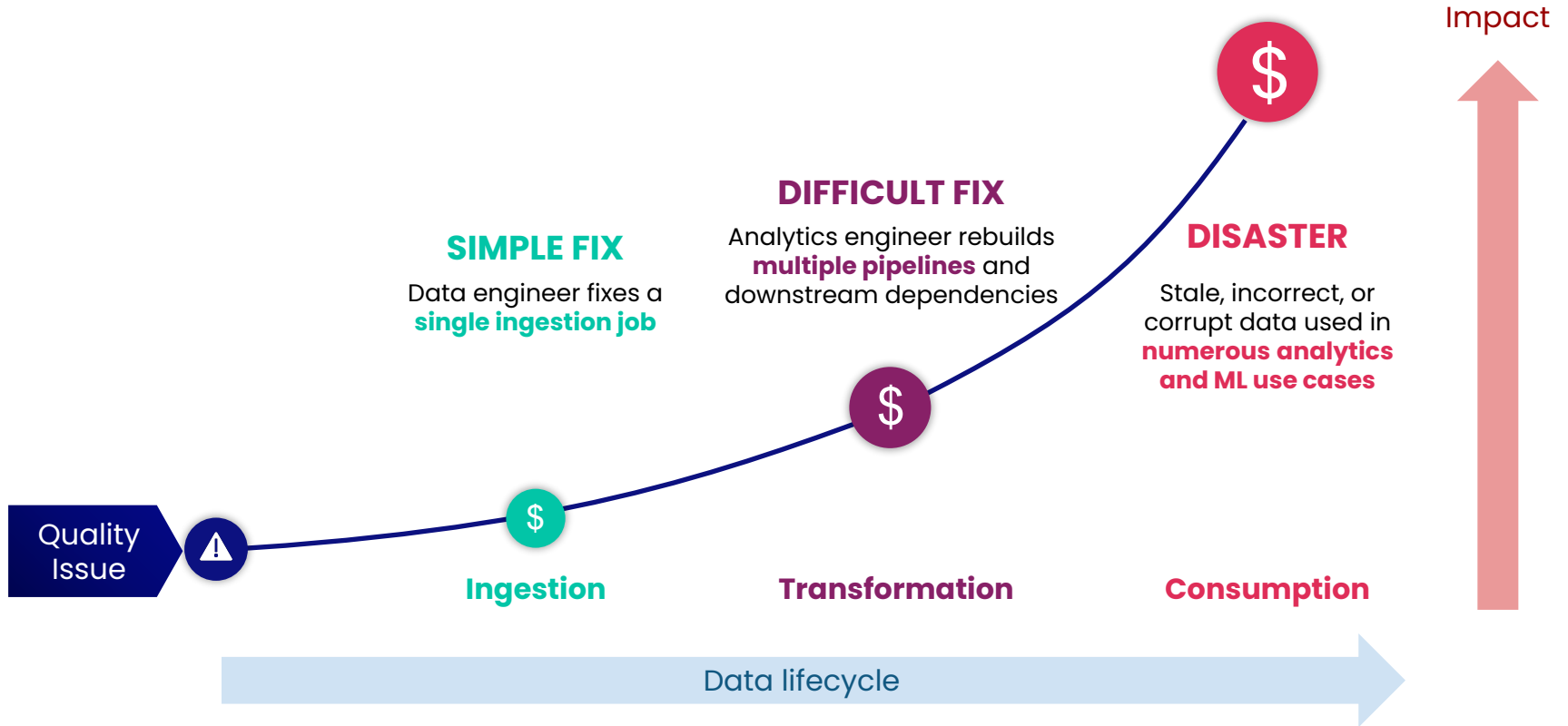
Zazzle

eMed™

VEGAS.COM

HORIZON
POWER

The compounding cost of quality issues



Issues at ingestion wreck havoc later

Root cause

Source data may arrive out-of-order and include duplicates

Sources & targets use different data types and naming conventions

Source systems' outputs change without warning

Downstream impact

Transform models need to be really complex to handle different cases

Dependent models break, columns disappear or are repeated, data are cast incorrectly

Unexpected or empty values in important fields ruin analytics



How can we prevent it?

Upsolver solution

ExactSync

- exactly-once delivery
- strong data ordering

Auto schema evolution

Resolves type conversion and column naming violations

Real-time data observability

Continuous statistics per field for monitoring and retrospection

Benefits

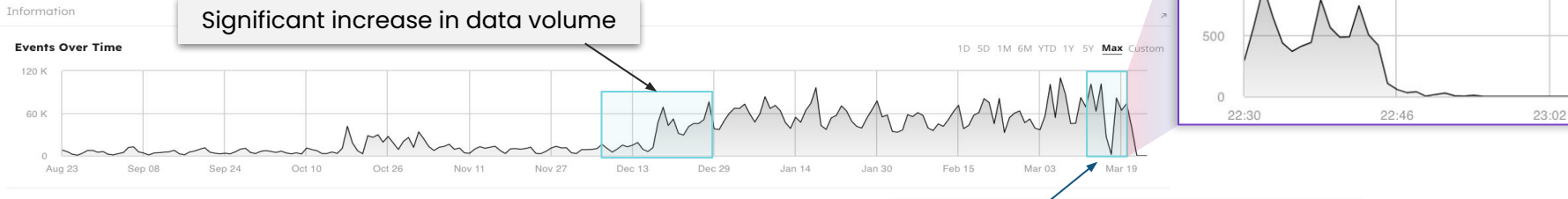
Transformations can focus on data, rather than error-handling

No schema-related breakage or manual evolution work

Quick detection dramatically reduces data issues and cost to fix



Real-time data observability



Significant number of NULLS in important fields

Fields Statistics

Field	Type	Density	Top Values
action	string	3.28%	all, EXPAND, VIEW
anonymous_id	string	28.86%	c240747c-18f9-49f9-ba3c-cefec
api_info.build.branch	string	100%	develop, hotfix/189-with-upsolv
api_info.build.revision	string	100%	293, 380, 4290401484
api_info.build.tag	string	100%	integ-e2e, sqlake, prod-global-a
api_info.environmenttype	string	100%	prod, integ-e2e, integ
api_info.trackingversion	string	100%	2
app			dbux, Sqlake
cluster_id			4be0753f-aeb0-4e8c-9a64-772e
cluster_name	string	<0.01%	Default Compute, testcompute2
completed_item_text	string	0.52%	(, ;, JOB
completed_item_type	string	0.52%	Unit, Constant, Field
completion_options[]	string	0.81%	NOT, (-
connection_id	string	0.69%	6f7458ed-e476-4858-87c8-ccf0

Not what I expected

Number of Arrays

Value Distribution

(_____
;	_____
JOB	_____
TABLE	_____
)	_____
CONNECTION	_____
SELECT	_____
DELETE_DATA	_____
DROP	_____
FROM	_____
CREATE	_____
MAP_COLUMNS_BY_NAME	_____

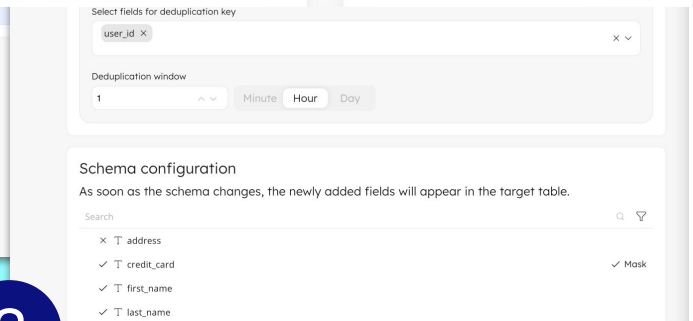
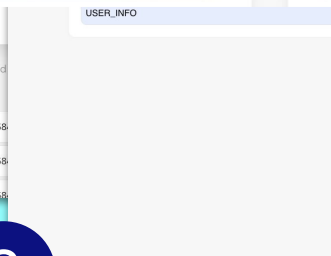
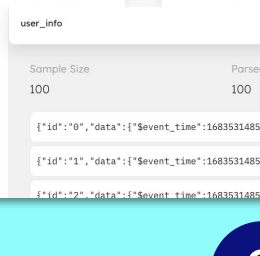
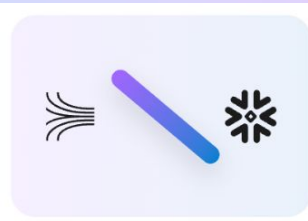
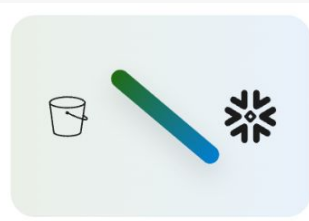
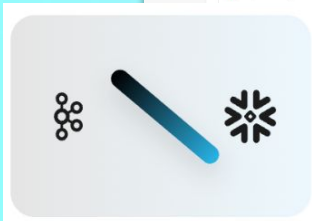
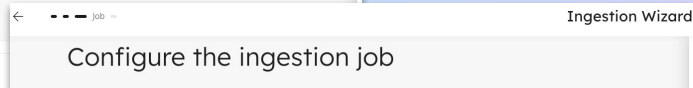
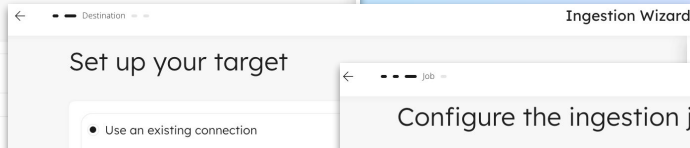
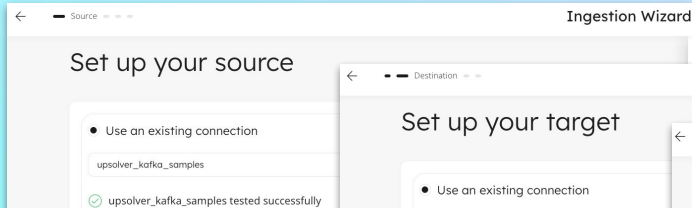
Outage!! - unexpected drop in events

Values	Array	First Seen	Last Seen
(false	2022-08-24 07:12	2023-03-22 03:46
;	false	2022-08-24 07:12	2023-03-22 03:46
JOB	false	2022-08-24 07:12	2023-03-22 03:46
TABLE	false	2022-08-24 07:12	2023-03-22 03:46
)	false	2022-08-24 07:12	2023-03-22 03:46
CONNECTION	false	2022-08-24 07:12	2023-03-22 03:46
SELECT	false	2022-08-24 07:12	2023-03-22 03:46
DELETE_DATA	false	2022-08-24 07:12	2023-03-22 03:46
DROP	false	2022-08-24 07:12	2023-03-22 03:46
FROM	false	2023-01-23 07:09	2023-03-16 09:12
CREATE	false	2023-01-23 06:57	2023-03-16 07:26
MAP_COLUMNS_BY_NAME	false	2023-01-05 08:35	2023-03-22 03:41
	false	2023-01-05 08:35	2023-03-22 03:41
	true	2023-01-05 08:35	2023-03-22 03:45
	false	2022-08-24 07:21	2023-03-22 03:41

Haven't been updated recently

Newly added columns

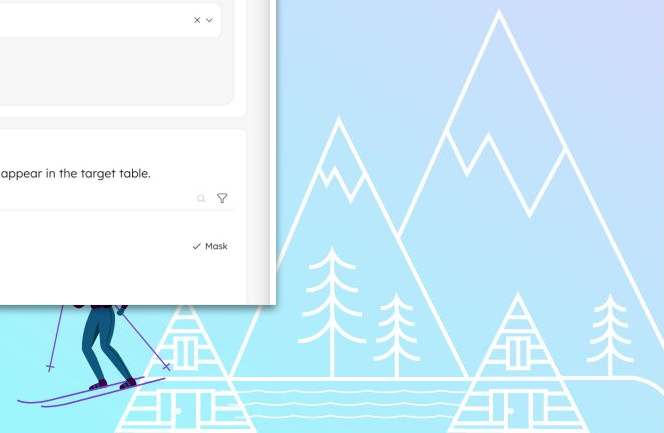
Getting started with upsolver



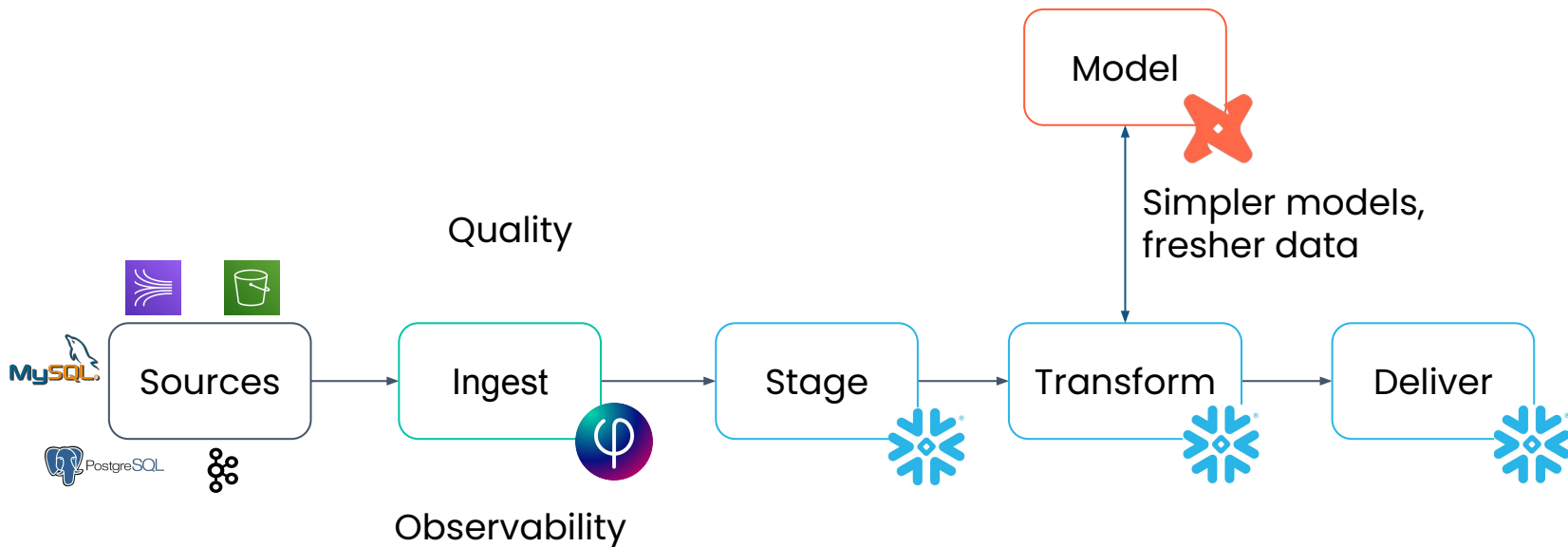
1

2

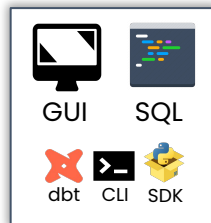
3




Reference architecture




Upsolver in the stack




Streams
Object Stores
Operational Databases



ExactSync

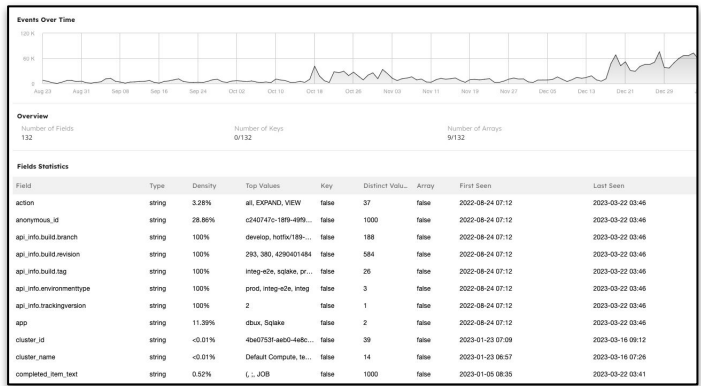


Auto Schema Evolution

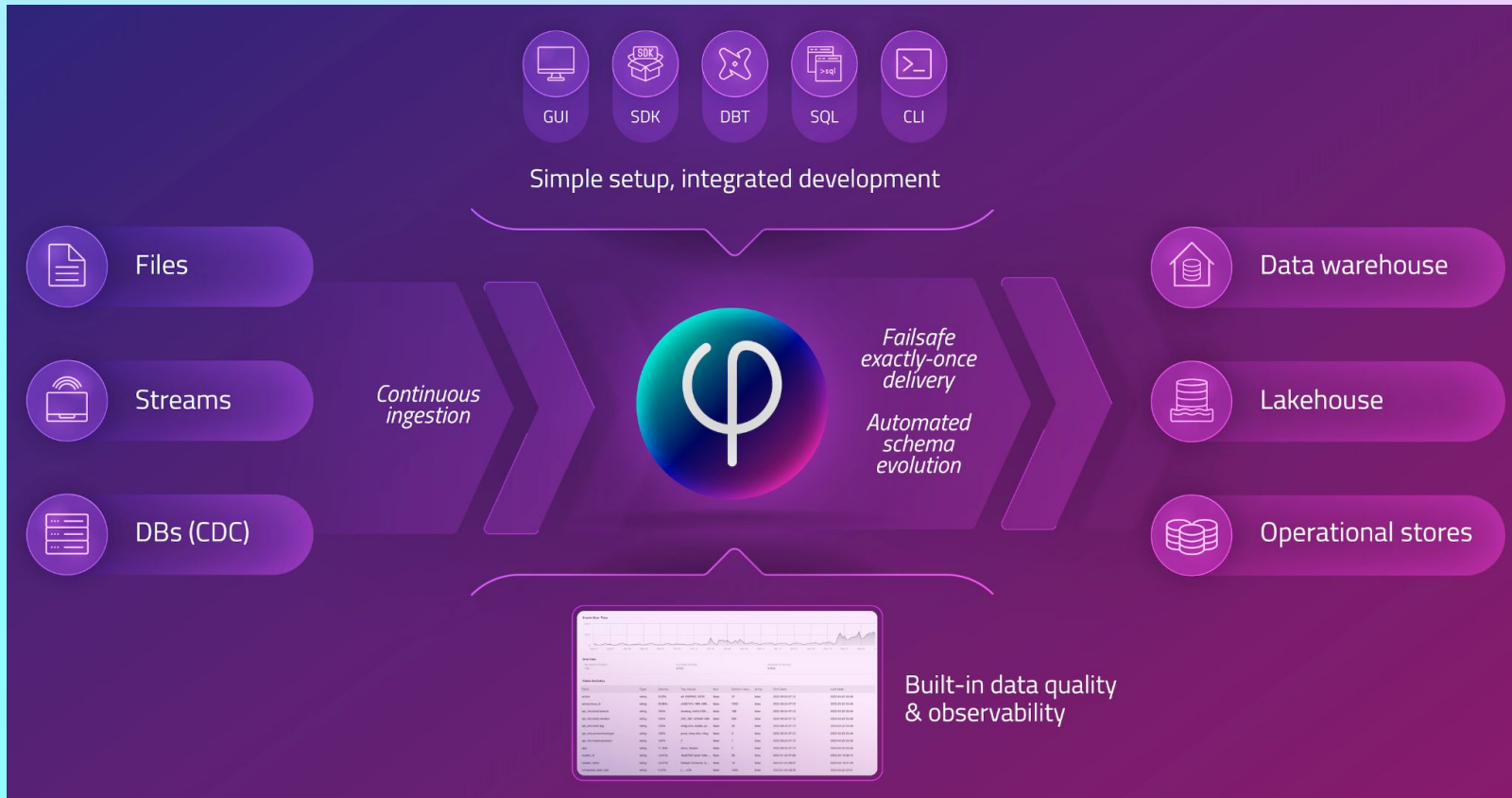


Real-time Data Observability

Data warehouse
Lakehouse
Operational databases



Upsolver in any stack



TESTIMONIALS

Upsolver plays a crucial part in our core data infrastructure, and the team has proven to be a reliable partner that's been **committed to our success from day one**.

Amit Attias | CTO, Bigabid

With Upsolver, I could handle **massive amounts of streaming data** and see real results, in a fraction of the time I thought it would take.

Guy Levy-Yurista | ex CSO, Sisense

Upsolver has saved thousands of engineering hours and significantly reduced total cost of ownership, **enabling us to invest in our hypergrowth rather than data pipelines**.

Seva Feldman | VP of R&D, Ironsource (Unity)

Upsolver is solving a problem at scale that I don't want to think about. The fact that **it just works** and people don't complain is a win.

Yuji Xie | Analytics Lead, Gem



“

Upsolver is like the “easy button” for Snowflake.
We ingest data from our Kafka streams, process it as necessary for different use cases, and deliver it, all while *observing how our schema and data are changing in real time.*

Alexander Adam

Cloud Platform Engineering Manager, Centene



Meet our client Centene Corporation



Centene Corporation in a snapshot

- Purpose: Transforming the health of the community, one person at a time.
- Mission: **Better health outcomes at lower costs.** Centene provides access to high-quality healthcare, innovative programs and a wide range of health solutions that help families and individuals get well, stay well, and be well.
- By the numbers:
 - 74,300 employees
 - Covering all 50 States
 - 27.1 Million Members
 - Fortune #26



How Centene uses data

We use data to help provide care and solutions for our members.

Our data lake receives data from many different source systems related to: Claims, Member, Provider, Authorization, Corporate Operations, Health Plan, etc.

Different business domain teams leverage data to improve an array of solutions, such as:

- Processing authorizations on the latest information about a member or their care.
- Validating prescriptions by combining pharmacy system data and member data.

We have a strict cloud security model so we can protect our member data.



Data platform team

Our group's goal is to ingest data from source systems and ship relevant data to domain-driven data stores, depending on different criteria and SLAs.

- Information in the source systems get updated throughout the day.
- Data get streamed through Kafka, processed by Upsolver, and pushed into Snowflake.
- Data from different sources are transformed and combined into larger, domain-informed data sets for downstream operations by business partners.



Platform enablement

Moving from on-prem to cloud, simply adding headcount to the platform team didn't scale:

- Our domain-driven use cases are highly specialized.
- With our previous setup, it was taking too long to onboard and train data engineers.

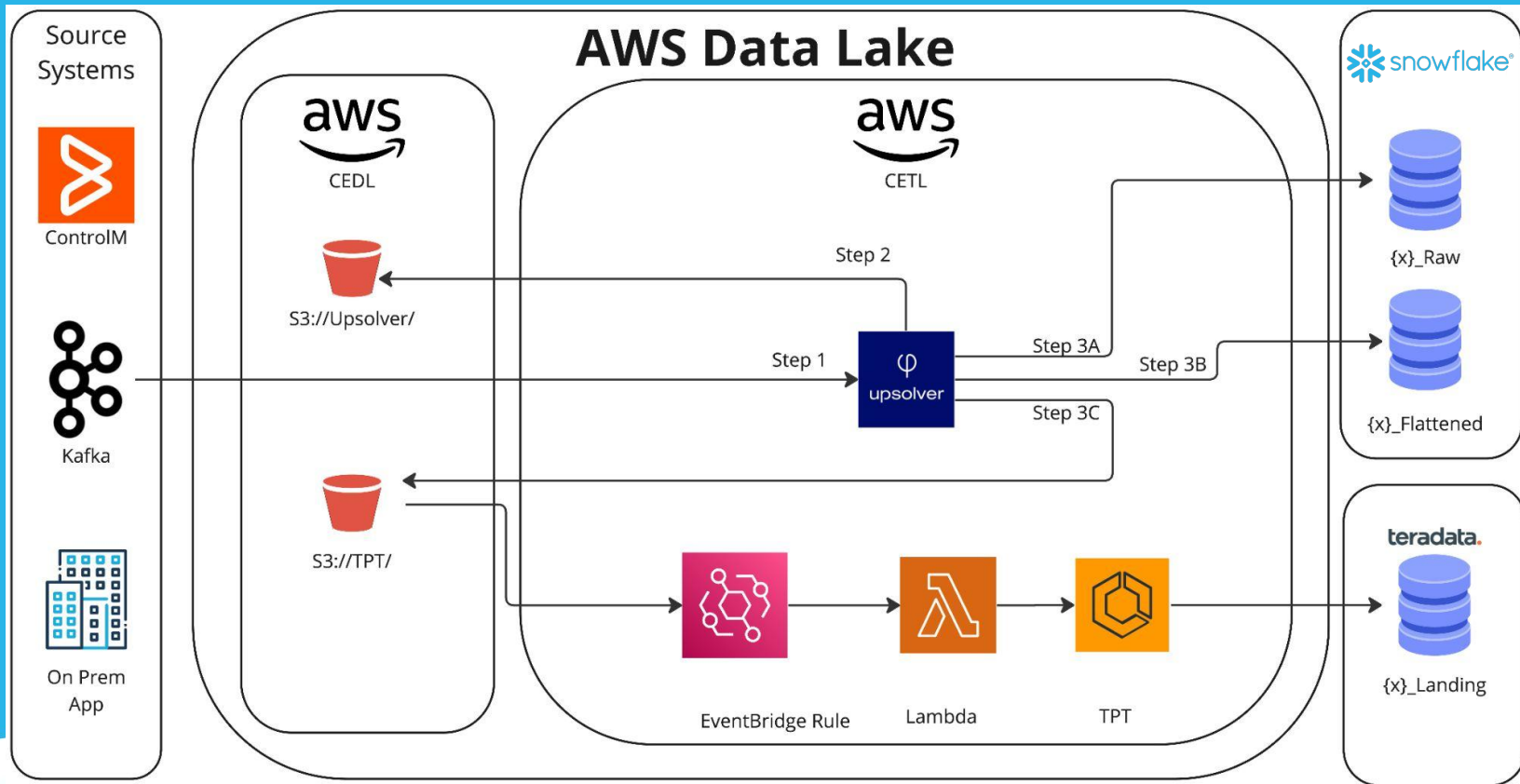
We needed:

- A platform that allows different teams to self-serve their relevant data.
- An enablement team for setting up processes and establishing usage patterns, providing templates, documentation, and how-to guides on data flow patterns.
- A solution for data platform engineers, domain data engineers, and data analysts.

We found a simple solution in Upsolver → “the easy button”



Data ingestion for Snowflake at Centene



Why Centene chose Upsolver

Simplification of workflows



- Quick way to ingest and flatten JSON data, with metrics on schema drift.



- Fan out source feed to multiple Snowflake databases with different transformation requirements.



- Eliminated the development complexities of hand-coding Glue/PySpark jobs.



- Avoid pitfalls such as rogue Glue jobs that may unpredictably drive up costs.



- Require less access provisioning to AWS account.

Why Centene chose Upsolver

Putting quality, governance, and security first



- Data never leaves Centene's cloud infrastructure.
- We can build monitoring based on the metadata found within the system tables in Upsolver.
- Clusters auto-scaling significantly reduces costs.
- Enables cost attribution to teams based on usage.



How Upsolver stood out

In one word, simplicity.

- Ease of setting up ingestion from Kafka to Snowflake.
- Creates a responsive staging table.
- Live metrics and metadata on fields as data flow.

Added more than 100 data pipelines

Onboarded 45 active users

... in two months



Business impact in a short time

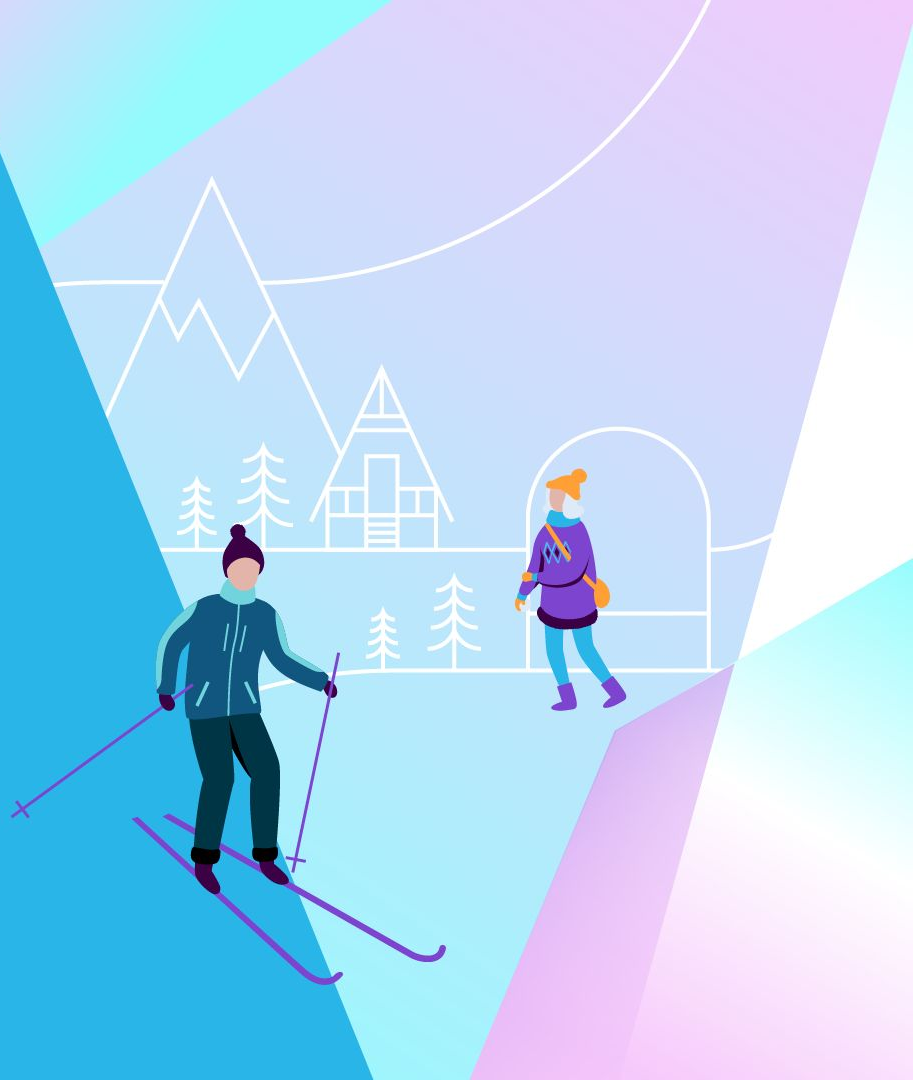
- With Upsolver, cloud data pipeline creation went **from quarters to a *single sprint***.
- A platform for subject matter expert data engineers to self-serve their data through the lake to EDWs that abstracts the complexities of creating an ETL pipeline.
- A lot less overhead on infrastructure management compared to other tools, while still providing flexibility and customization of infrastructure.



Time
Cost
Ops



Q&A





SUMMIT 2023

THANK YOU!

