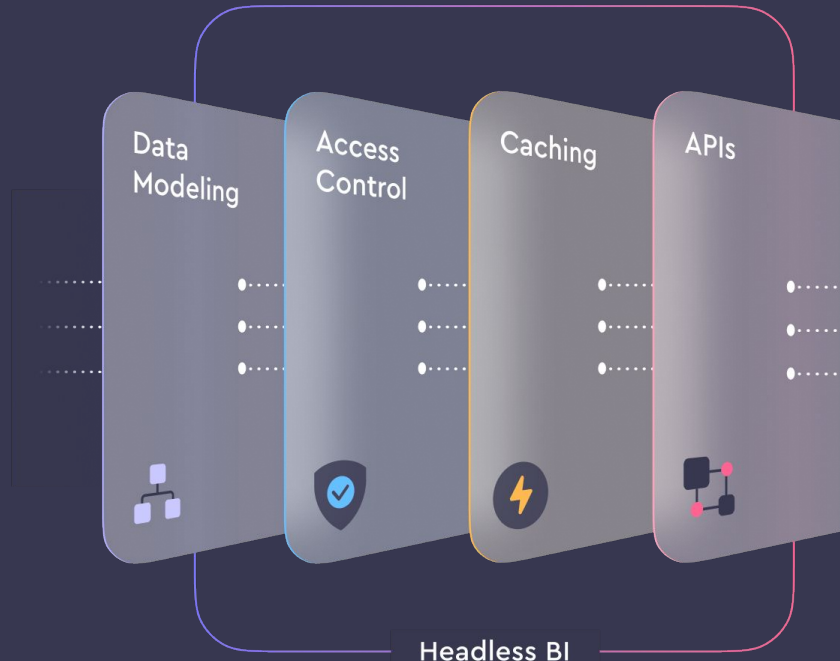
cube

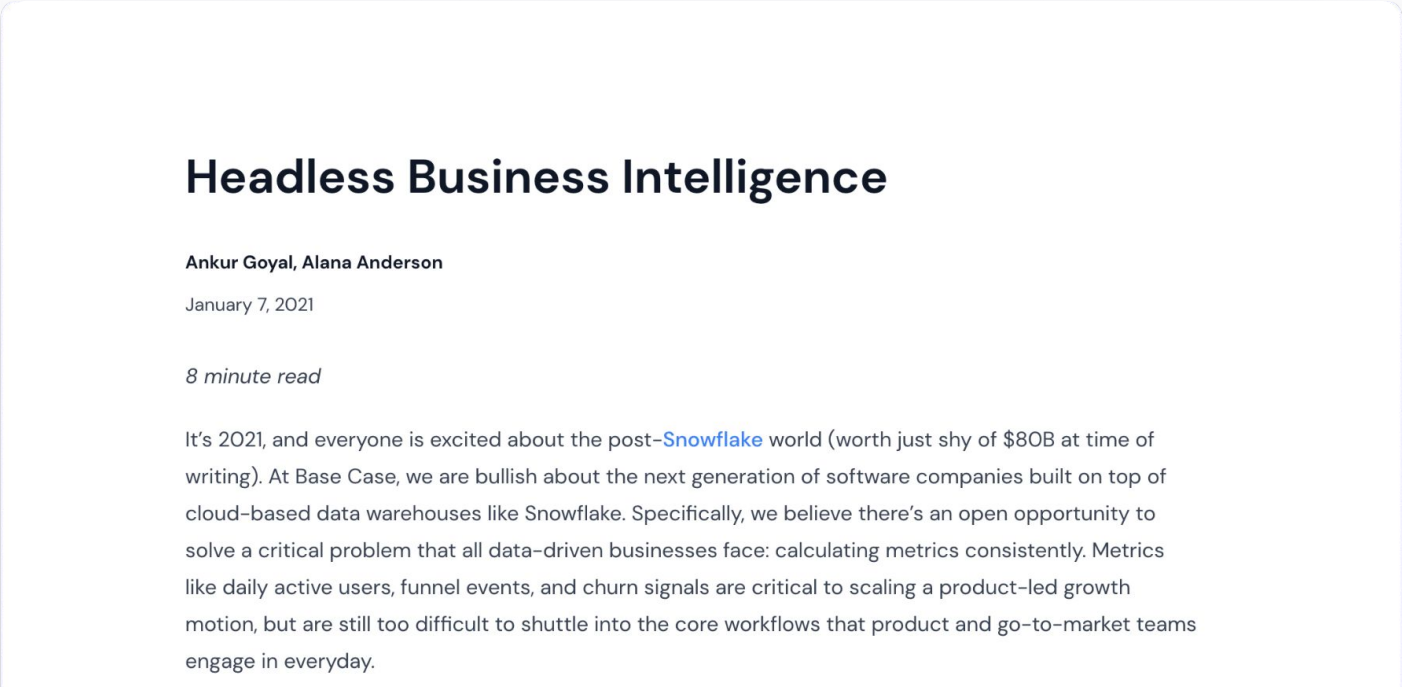
Navigating the Modern Data Stack with Open-Source Headless BI

Artyom Keydunov
Co-founder & CEO at Cube



What is headless BI?

Basecase VC was the first to publicly talk about ‘Headless BI’ in early 2021



Headless Business Intelligence

Ankur Goyal, Alana Anderson

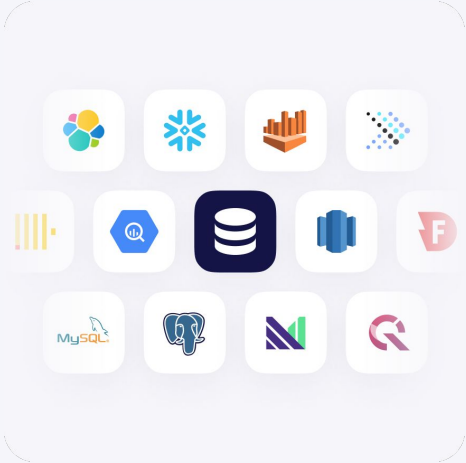
January 7, 2021

8 minute read

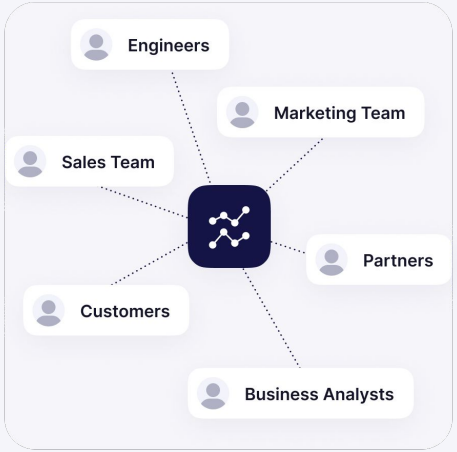
It’s 2021, and everyone is excited about the post-Snowflake world (worth just shy of $80B at time of writing). At Base Case, we are bullish about the next generation of software companies built on top of cloud-based data warehouses like Snowflake. Specifically, we believe there’s an open opportunity to solve a critical problem that all data-driven businesses face: calculating metrics consistently. Metrics like daily active users, funnel events, and churn signals are critical to scaling a product-led growth motion, but are still too difficult to shuttle into the core workflows that product and go-to-market teams engage in everyday.

The category of products that should be addressing this is business intelligence (BI) tools; however, by

Proliferation of data applications



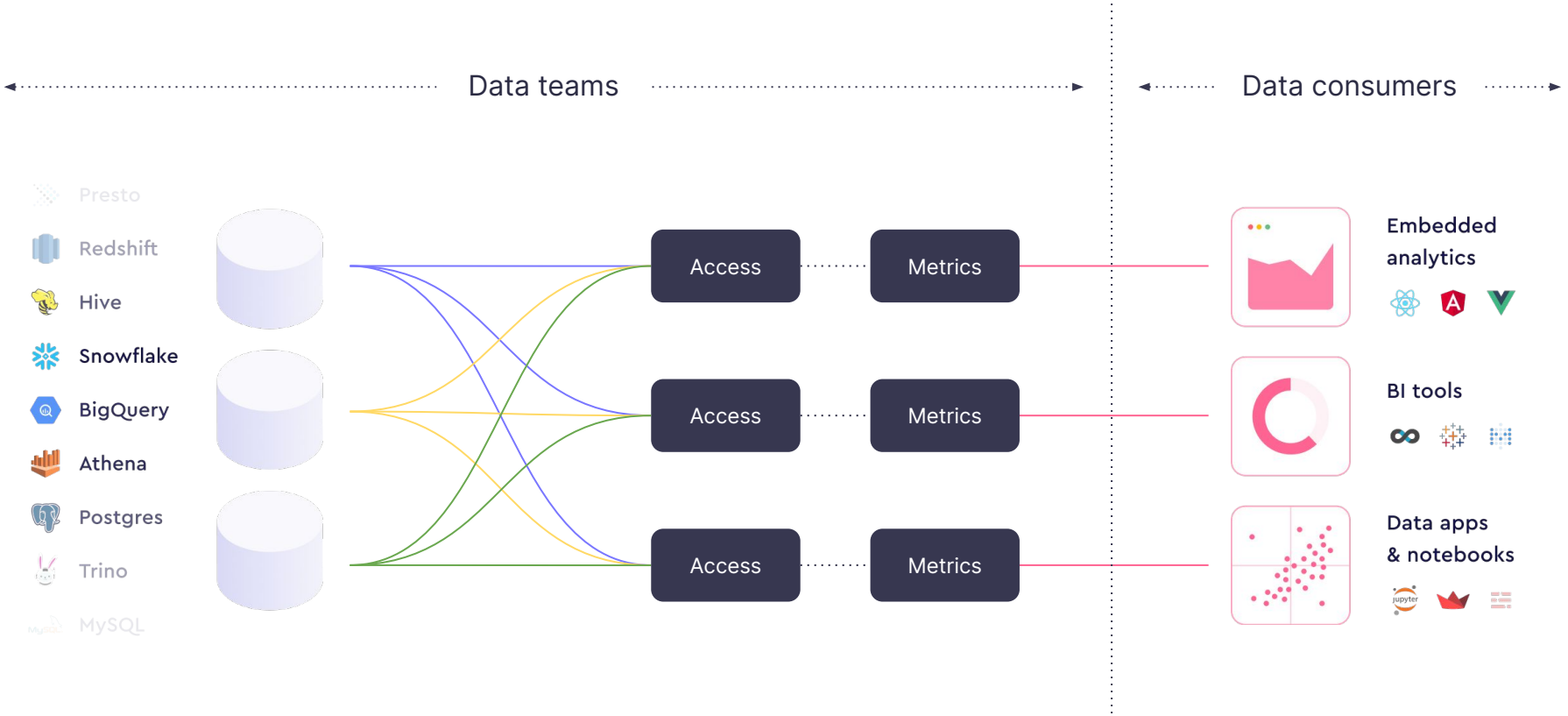
More data source



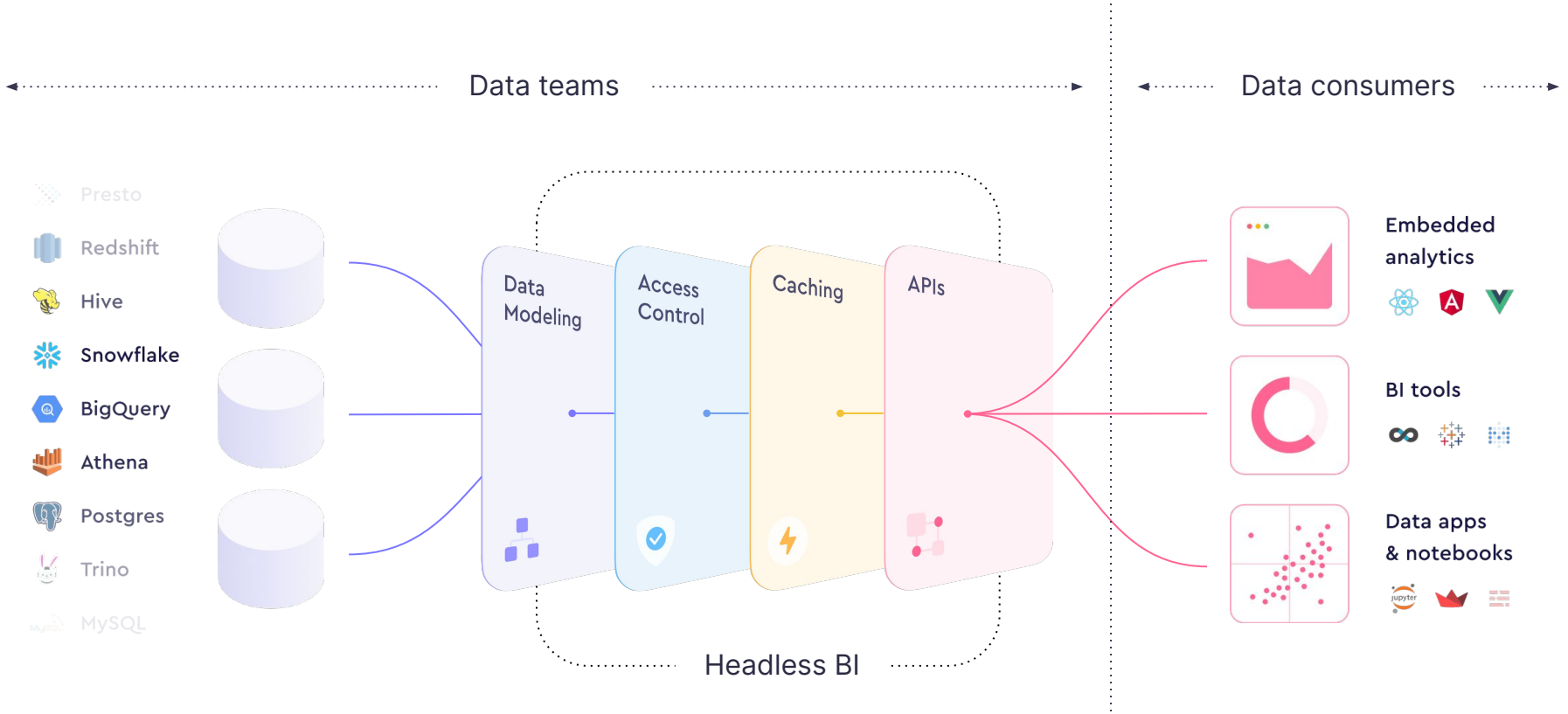
More data consumers

Proliferation of role-specific data applications

Very *un*DRY

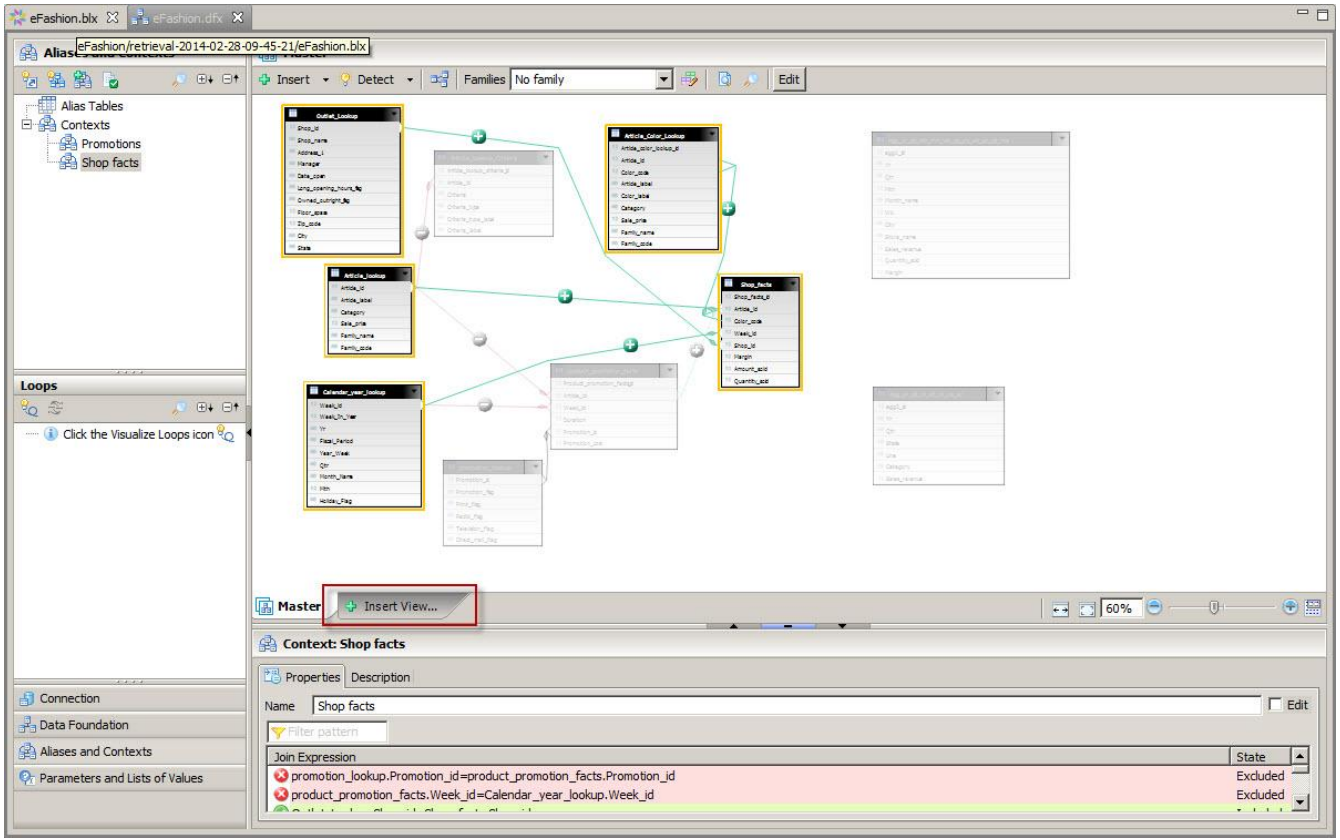

Data teams
Data consumers
Presto
Redshift
Hive
Snowflake
BigQuery
Athena
Postgres
Trino
MySQL
Access
Access
Access
Metrics
Metrics
Metrics
Embedded analytics
BI tools
Data apps & notebooks

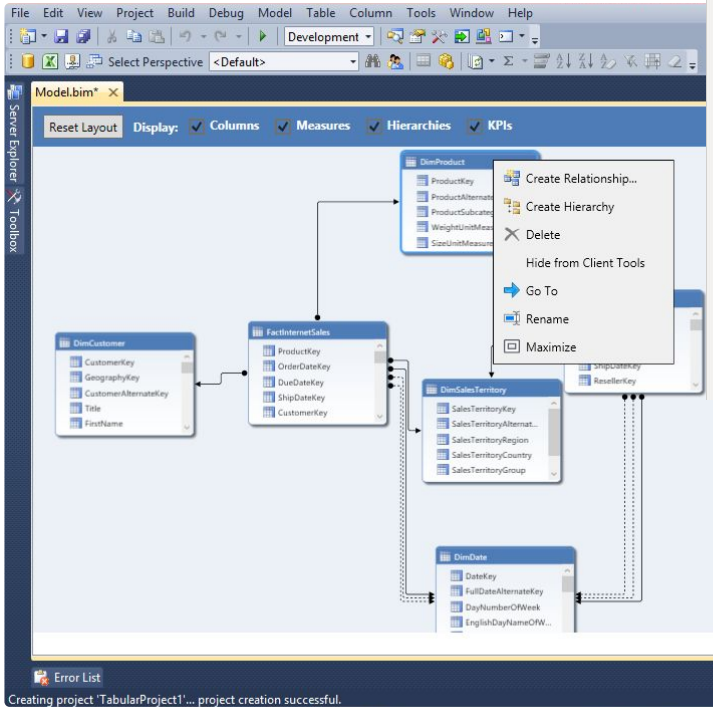
Instead,extract common logic upstream

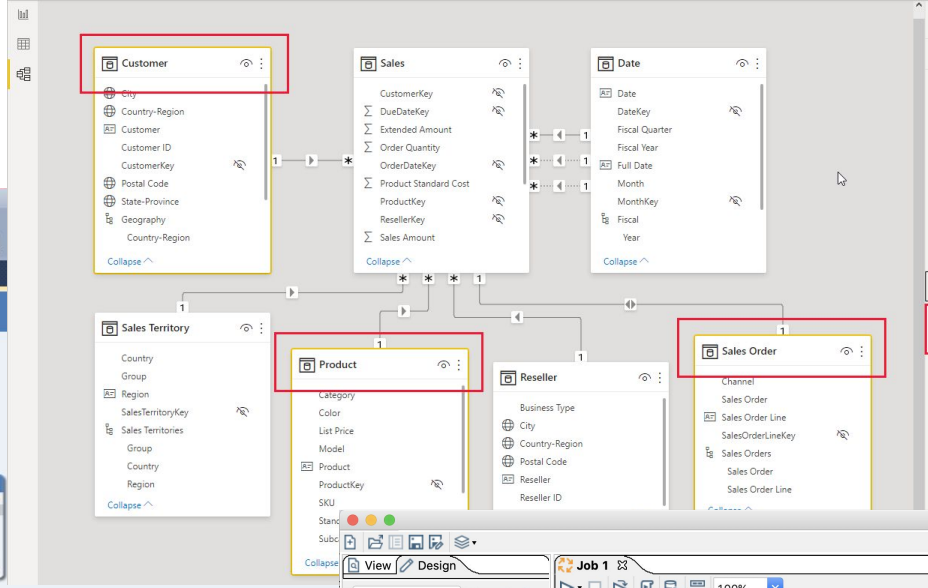


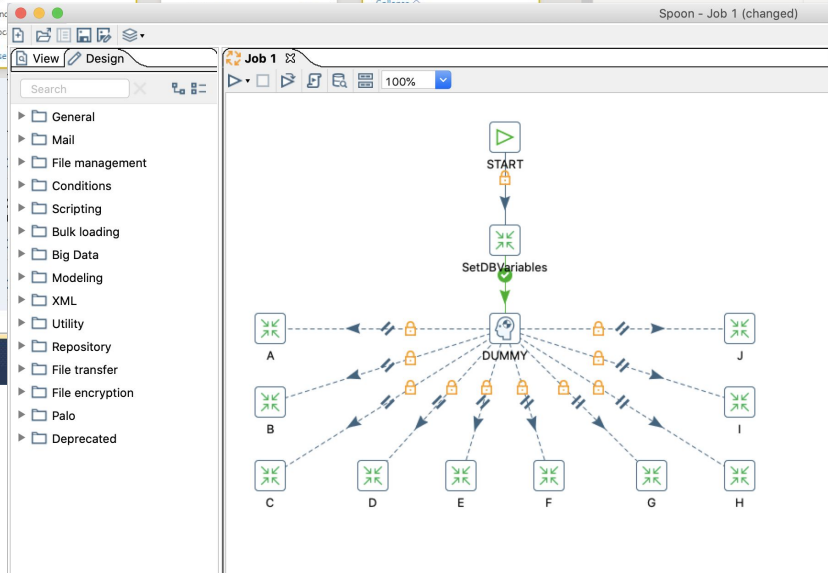
The core of headless BI is the semantic layer—

—a concept that has existed for decades.


eFashion.blx
eFashion.dfx
eFashion/retrieval-2014-02-28-09-45-21/eFashion.blx
Insert
Detect
Families
No family
Edit
Alias Tables
Contexts
Promotions
Shop facts
Loops
Click the Visualize Loops icon
Connection
Data Foundation
Aliases and Contexts
Parameters and Lists of Values
Master
Insert View...
60%
Context: Shop facts
Properties
Description
Name
Shop facts
Edit
Filter pattern
Join Expression
State
promotion_lookup.Promotion_id=product_promotion_facts.Promotion_id
Excluded
product_promotion_facts.Week_id=Calendar_year_lookup.Week_id
Excluded

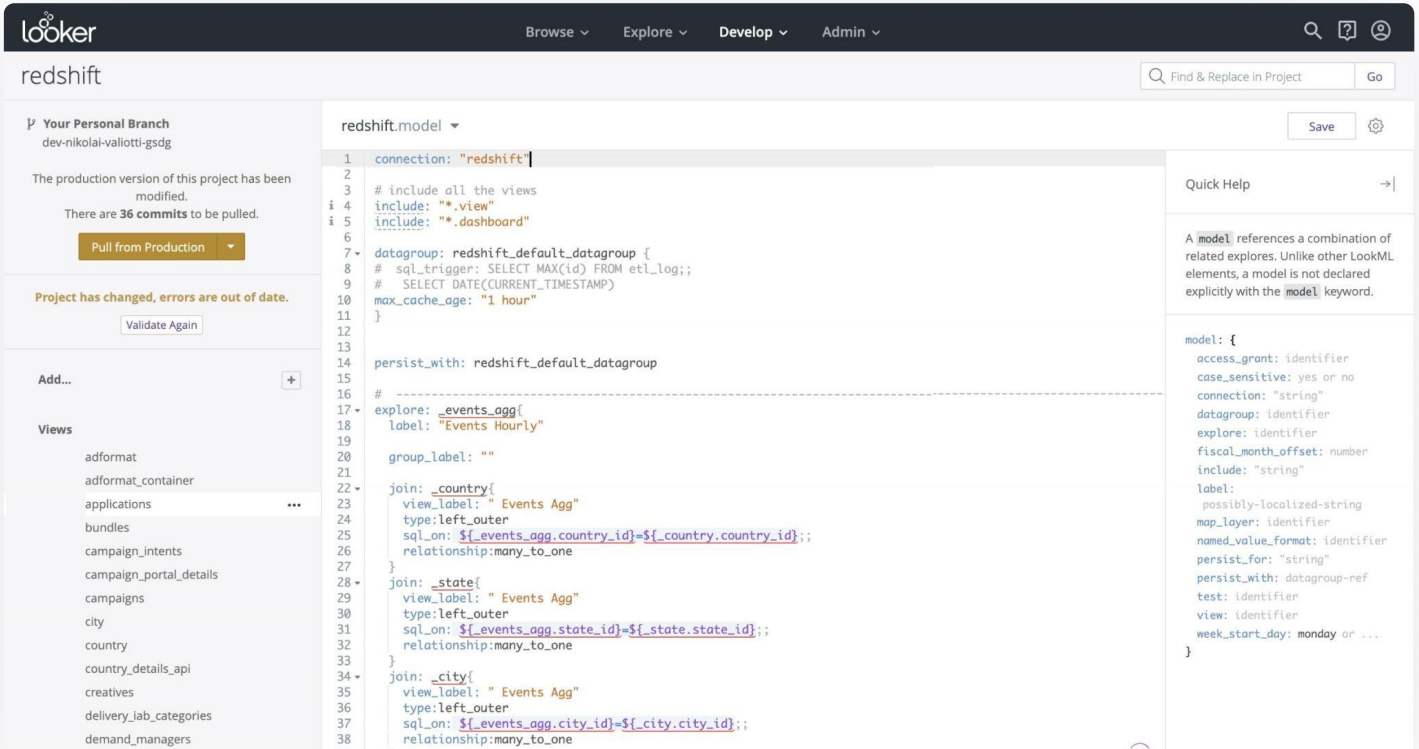

File Edit View Project Build Debug Model Table Column Tools Window Help
Development
Select Perspective
<Default>
Model.bim*
Reset Layout
Display:
Columns
Measures
Hierarchies
KPIs
DimProduct
ProductKey
Create Relationship...
Create Hierarchy
Delete
Hide from Client Tools
Go To
Rename
Maximize
DimCustomer
CustomerKey
GeographyKey
CustomerAlternateKey
Title
FirstName
FactInternetSales
ProductKey
OrderDateKey
DueDateKey
ShipDateKey
CustomerKey
DimSalesTerritory
SalesTerritoryKey
SalesTerritoryRegion
SalesTerritoryCountry
SalesTerritoryGroup
DimDate
DateKey
FullDateAlternateKey
DayNumberOfWeek
Error List
Creating project 'TabularProject1'... project creation successful.


Customer
City
Country-Region
Customer
Customer ID
CustomerKey
Postal Code
State-Province
Geography
Country-Region
Collapse
Sales
CustomerKey
DueDateKey
Extended Amount
Order Quantity
OrderDateKey
Product Standard Cost
ProductKey
ResellerKey
Sales Amount
Collapse
Date
Date
DateKey
Fiscal Quarter
Fiscal Year
Full Date
Month
MonthKey
Fiscal
Year
Collapse
Sales Territory
Country
Group
Region
SalesTerritoryKey
Sales Territories
Group
Country
Region
Collapse
Product
Category
Color
List Price
Model
Product
ProductKey
SKU
Collapse
Reseller
Business Type
City
Country-Region
Postal Code
Reseller
Reseller ID
Sales Order
Channel
Sales Order
Sales Order Line
SalesOrderLineKey
Sales Orders
Sales Order
Sales Order Line
Properties
General
Description
Enter a description
Synonyms
Enter a comma-separated list of synonyms for Q&A
Is hidden
No
Advanced
Storage mode
DirectQuery
Fields

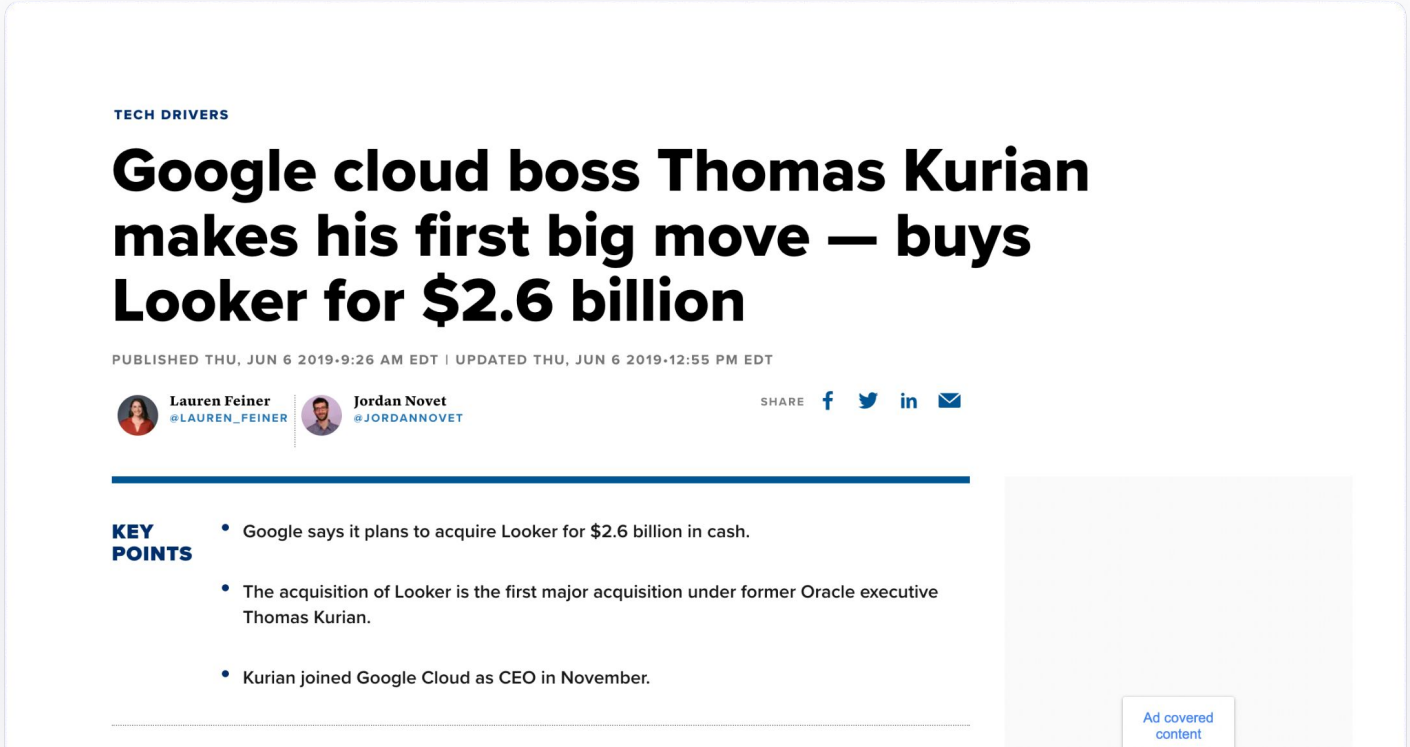

Spoon - Job 1 (changed)
View
Design
Search
General
Mail
File management
Conditions
Scripting
Bulk loading
Big Data
Modeling
XML
Utility
Repository
File transfer
File encryption
Palo
Deprecated
Job 1
100%
START
SetDBVariables
DUMMY
A
B
C
D
E
F
G
H
I
J

The semantic layer was and still is a part of many, many BI tools.

Looker innovated the semantic layer by making it code-based



LookML was a major factor in Looker's success



TECH DRIVERS

Google cloud boss Thomas Kurian makes his first big move — buys Looker for $2.6 billion

PUBLISHED THU, JUN 6 2019•9:26 AM EDT | UPDATED THU, JUN 6 2019•12:55 PM EDT

Lauren Feiner @LAUREN_FEINER

Jordan Novet @JORDANNOVET

SHARE

KEY POINTS

- Google says it plans to acquire Looker for $2.6 billion in cash.
- The acquisition of Looker is the first major acquisition under former Oracle executive Thomas Kurian.
- Kurian joined Google Cloud as CEO in November.

Ad covered content

What if a semantic layer could be accessed outside of a BI tool?

Enterprise

Google's Looker partners with Tableau

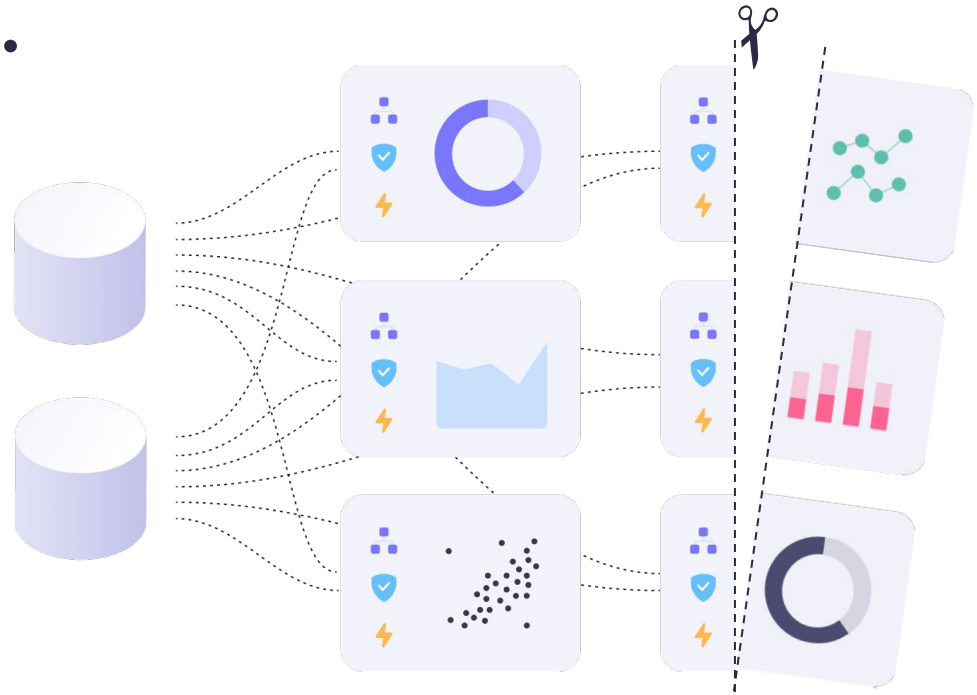
Frederic Lardinois @fredericl / 5:00 AM PDT • October 12, 2021

Comment

In 2019, Google acquired the business intelligence service Looker for $2.6 billion and Salesforce picked up Tableau for $15.7 billion. But today, the two competitors are growing closer together thanks to a new integration between the two. Specifically, Tableau users will soon be able to access Looker's semantic layer, while Google's Looker users will soon be able to use Tableau's visualization layer on top of the Looker platform.

This may seem like an unlikely partnership at first, but in a way, this allows both services to play to their strengths. Tableau's advanced visualization capabilities have always been a draw for its users, but unlike Looker, it was born on the desktop and despite all of the company's efforts, that

Headless BI: keep the semantic layer, caching, and access control—but decouple visualizations from them.



2020–2022: The rise of Headless BI tools in the modern data stack

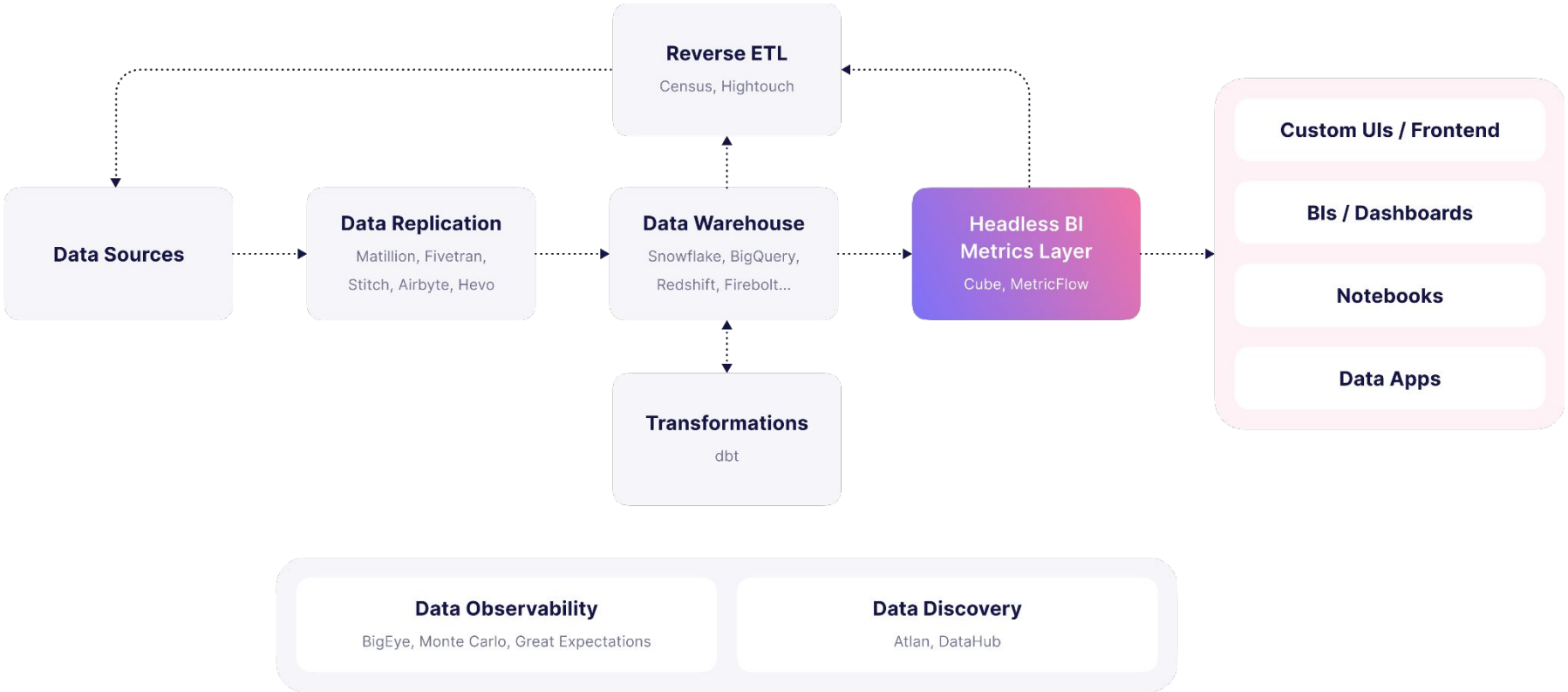
Why open-source?

The modern data stack is inherently more open

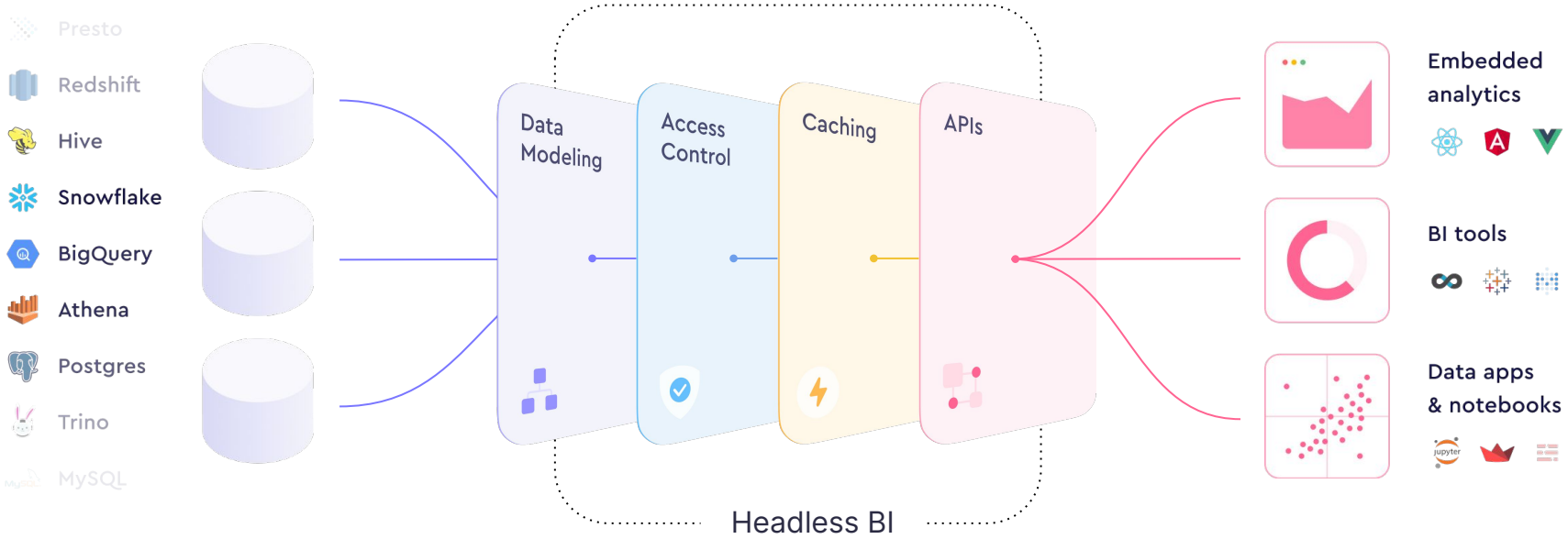
Many core components of the modern data stack are open-sourced

ETL, CDP	OLAP DBMS	Transformations	Orchestration	Headless BI
meltano	ClickHouse	dbt	Apache Airflow	cube
SNOWPLOW			dagster	MetricFlow
Airbyte			PREFECT	
rudderstack				

Headless BI in the modern data stack

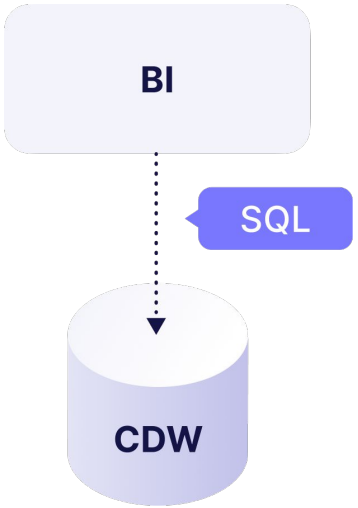


Architecture of Headless BI



Push down compute to CDW

Build on top of the innovation of previous generation BI.
BI tools generate and execute SQL queries in CDWs.



- Power BI direct query
- Tableau live connections
- Thoughtspot
- Looker

Data modeling (semantic) layer in code

```yaml
cubes:
  name: active_users
  sql: SELECT * from events

  measures:
    - name: weekly_active
      type: count_distinct
      rolling_window:
        trailing: 7 day
        offset: start

  dimensions:
    - name: time
      type: time
      sql: timestamp
```

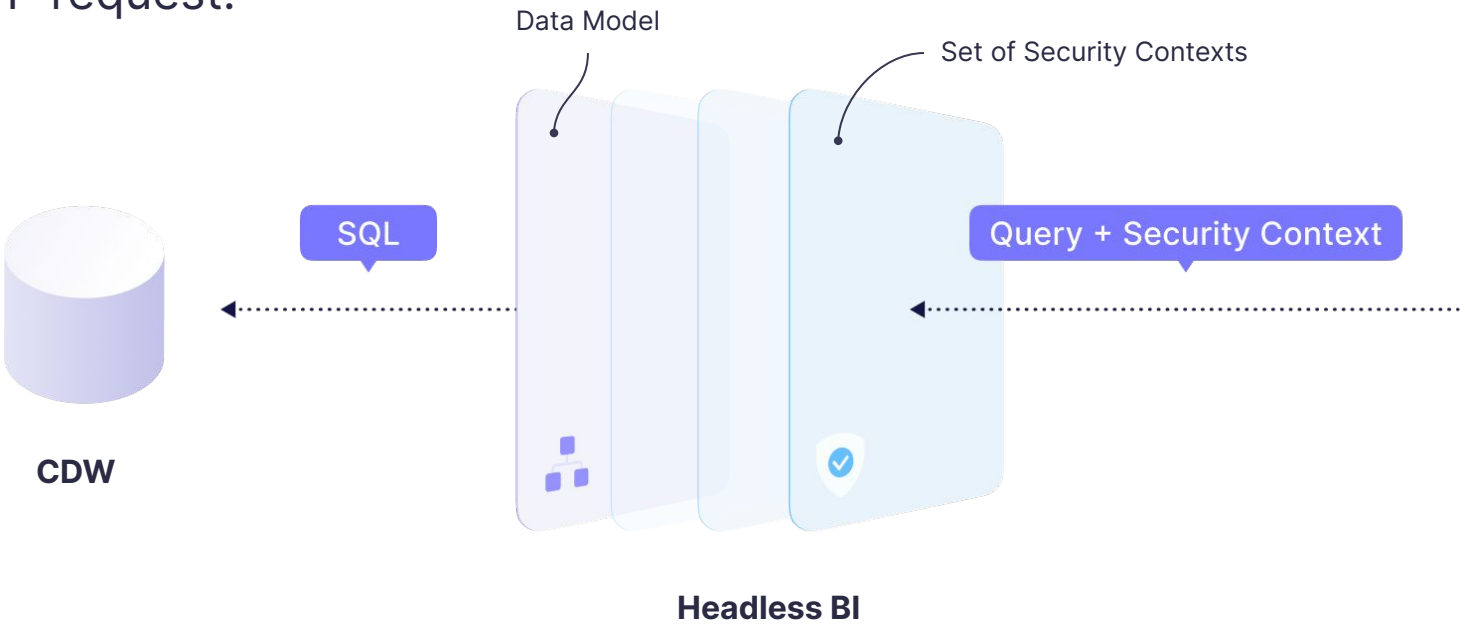
cube

```yaml
data_source:
  name: transactions
  sql_table: schema.transactions

  identifiers:
    - name: transaction_id
      type: primary

  measures:
    - name: revenue
      description:
      expr: price * quantity
      agg: sum
```

MetricFlow

Multi-tenancy and security control

Headless BI generates SQL from the data model within the specific security context that was supplied via incoming query, e.g. with JWT token in HTTP request.



Caching

Caching via aggregations inside CDW or purpose-built aggregation storage

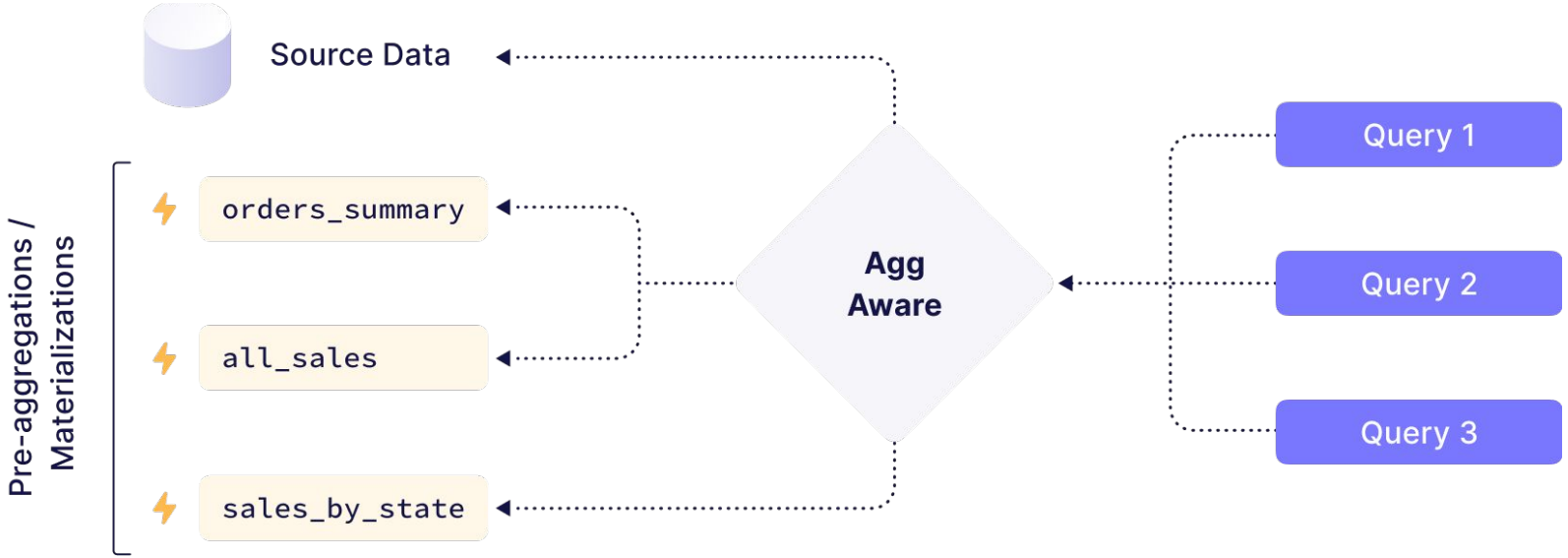
```yaml
cubes:
  pre_aggregations:
    - name: orders_summary
      measures:
        - count
        - total_value
      dimensions:
        - status
      time_dimension: created_at
      granularity: week
```

cube

```yaml
materialization:
  name: user_bookings_summary
  metrics:
    - bookings
    - booking_value
    - customer_service_tickets
    - guest_host_messages

  dimensions:
    - metric_time
    - guest__signup_channel
```

MetricFlow

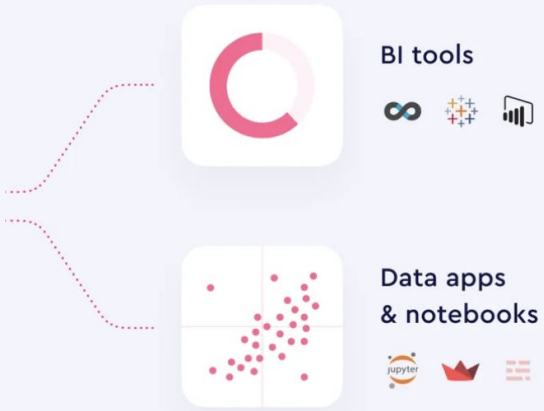
Querying with caching: aggregate awareness



Consumption level — APIs: SQL

Importance of SQL API; the SQL is lingua franca of data world

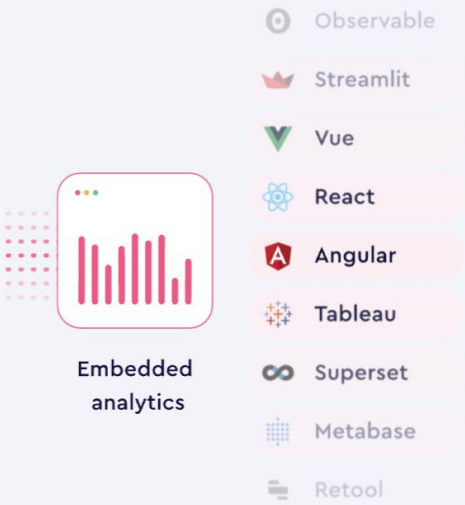
```
SELECT
  status,
  MEASURE(revenue)
FROM orders
WHERE STATUS != "completed"
GROUP BY 1
```



Consumption level — APIs: REST/GraphQL

Rest/GraphQL for web developers

```
query {
  cube {
    orders(where: {
      status: { notEquals: "completed" }
    }) {
      count
      status
      createdAt {
        month
      }
    }
  }
}
```



Thank you!


cube