INTRODUCING A SQL-BASED METRICS LAYER POWERED BY DUCKDB

Mike Driscoll
Co-Founder, CEO at Rill Data

Themes

- Why are metrics the core building blocks of analytics?
- Why should they be defined in SQL?
- Why is DuckDB the ideal engine for a metrics layer?
- How might AI accelerate metrics modeling and exploration?

Rill is a BI tool designed for DuckDB's unique speed

Optimized for OLAP engines

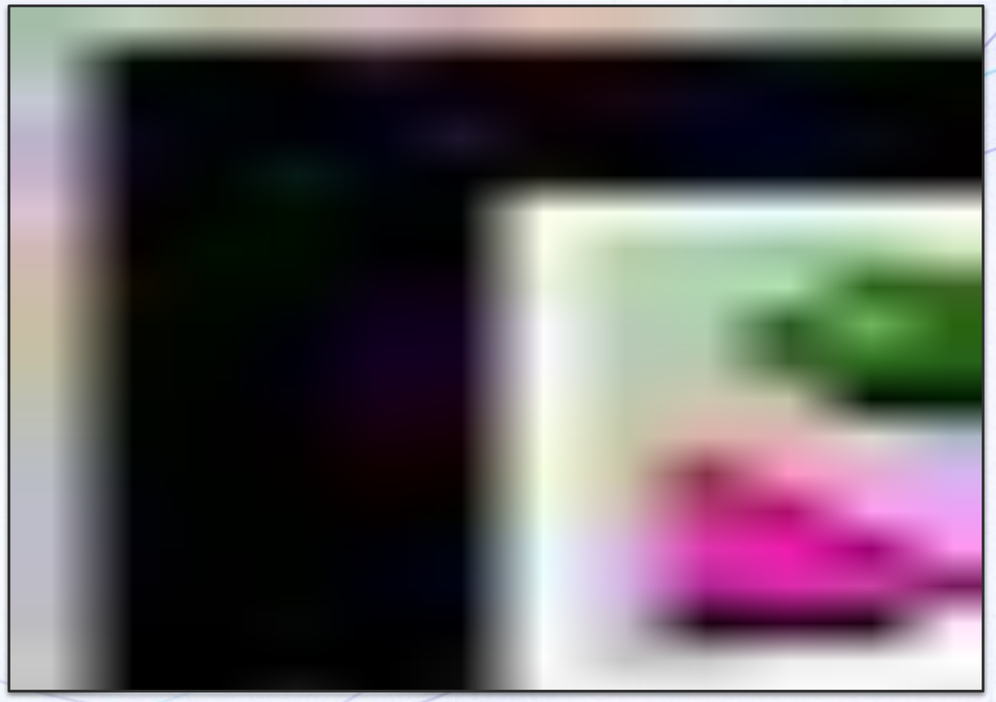
- Drilling, slicing, and dicing is instant, meant to feel tactile

BI-as-Code

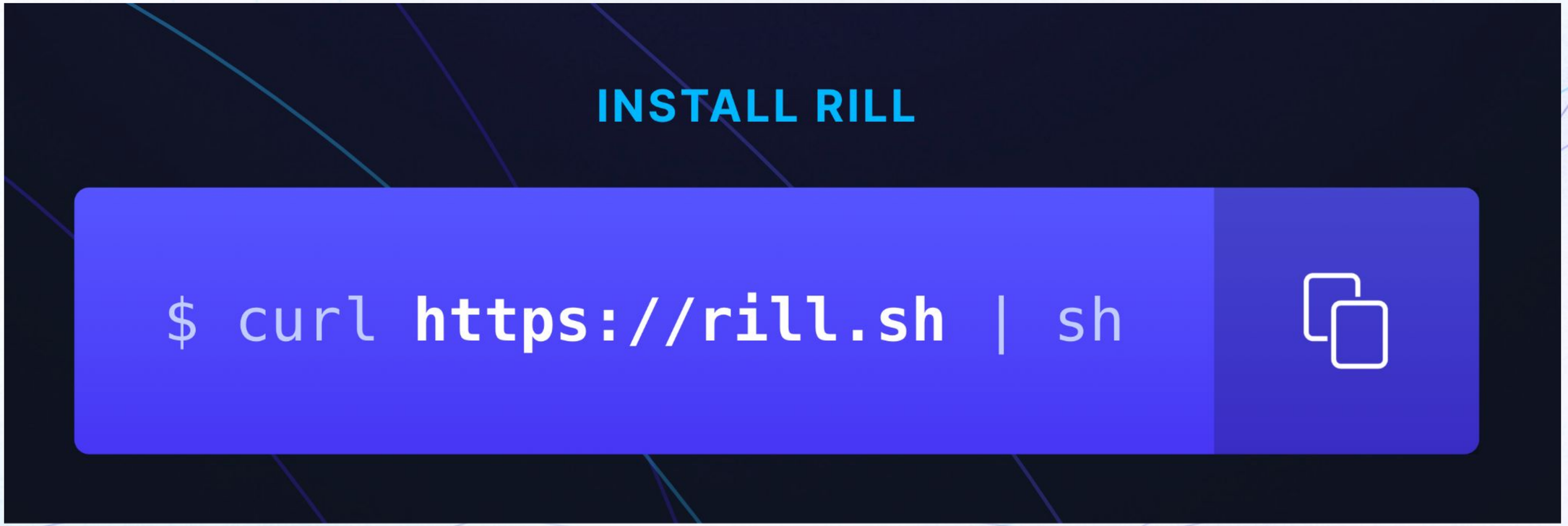
- Develop dashboards locally, then publish globally with Git workflows

Metrics-first Philosophy

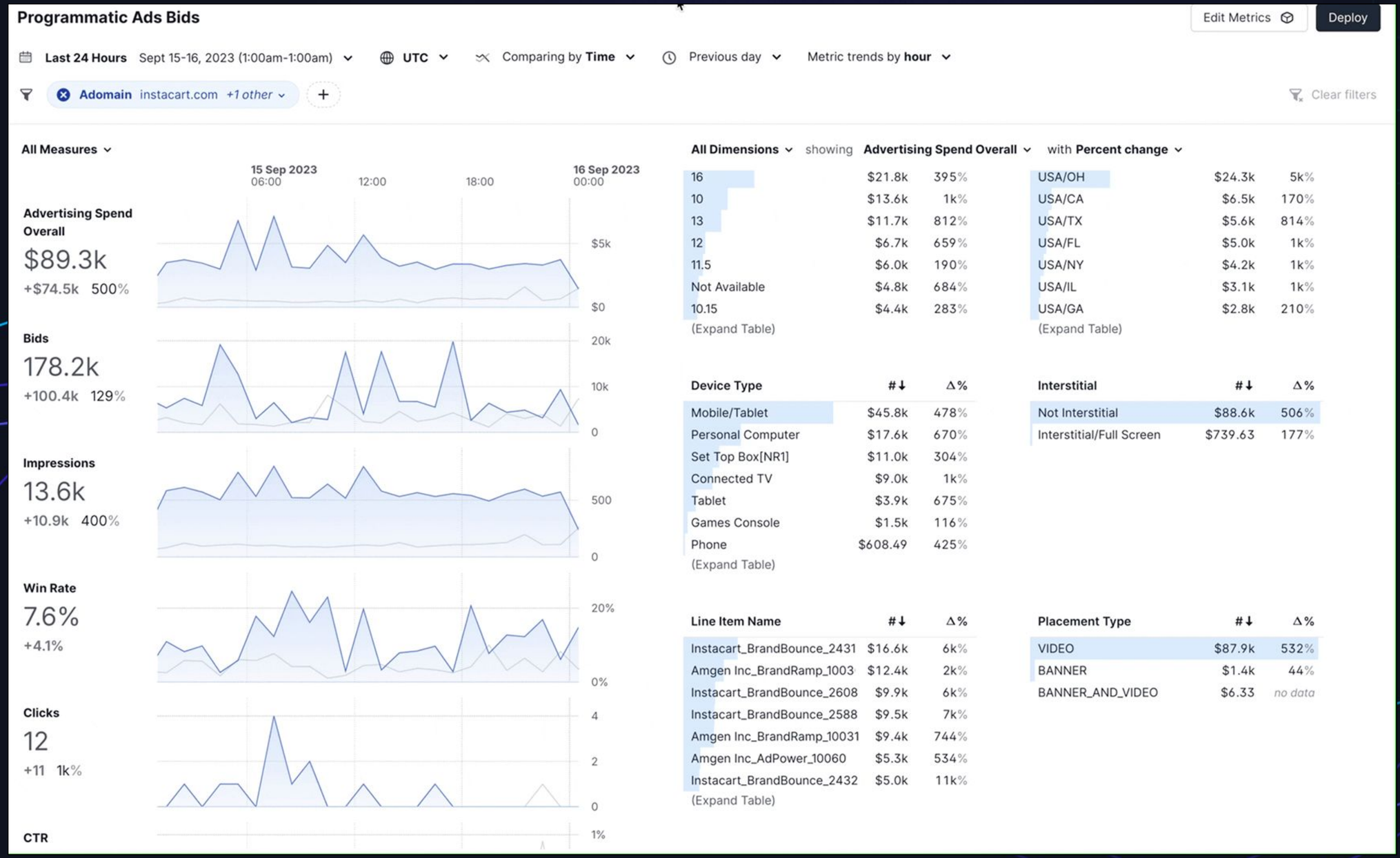
- Declare metrics with SQL expressions, and Rill *auto-generates* dashboards



Live Demo of a DuckDB-Powered Metrics Layer



Visit www.rilldata.com to copy command


Programmatic Ads Bids
Edit Metrics
Deploy
Last 24 Hours Sept 15-16, 2023 (1:00am-1:00am)
UTC
Comparing by Time
Previous day
Metric trends by hour
Adomain instacart.com +1 other
Clear filters
All Measures
15 Sep 2023
06:00
12:00
18:00
16 Sep 2023
00:00
Advertising Spend Overall
$89.3k
+$74.5k 500%
$5k
$0
Bids
178.2k
+100.4k 129%
20k
10k
0
Impressions
13.6k
+10.9k 400%
500
0
Win Rate
7.6%
+4.1%
20%
0%
Clicks
12
+11 1k%
4
2
0
CTR
1%
All Dimensions showing Advertising Spend Overall with Percent change
16 $21.8k 395%
10 $13.6k 1k%
13 $11.7k 812%
12 $6.7k 659%
11.5 $6.0k 190%
Not Available $4.8k 684%
10.15 $4.4k 283%
(Expand Table)
USA/OH $24.3k 5k%
USA/CA $6.5k 170%
USA/TX $5.6k 814%
USA/FL $5.0k 1k%
USA/NY $4.2k 1k%
USA/IL $3.1k 1k%
USA/GA $2.8k 210%
(Expand Table)
Device Type # ↓ Δ%
Mobile/Tablet $45.8k 478%
Personal Computer $17.6k 670%
Set Top Box[NR1] $11.0k 304%
Connected TV $9.0k 1k%
Tablet $3.9k 675%
Games Console $1.5k 116%
Phone $608.49 425%
(Expand Table)
Interstitial # ↓ Δ%
Not Interstitial $88.6k 506%
Interstitial/Full Screen $739.63 177%
Line Item Name # ↓ Δ%
Instacart_BrandBounce_2431 $16.6k 6k%
Amgen Inc_BrandRamp_1003 $12.4k 2k%
Instacart_BrandBounce_2608 $9.9k 6k%
Instacart_BrandBounce_2588 $9.5k 7k%
Amgen Inc_BrandRamp_10031 $9.4k 744%
Amgen Inc_AdPower_10060 $5.3k 534%
Instacart_BrandBounce_2432 $5.0k 11k%
(Expand Table)
Placement Type # ↓ Δ%
VIDEO $87.9k 532%
BANNER $1.4k 44%
BANNER_AND_VIDEO $6.33 no data

Metrics are the right-sized building blocks of analytics

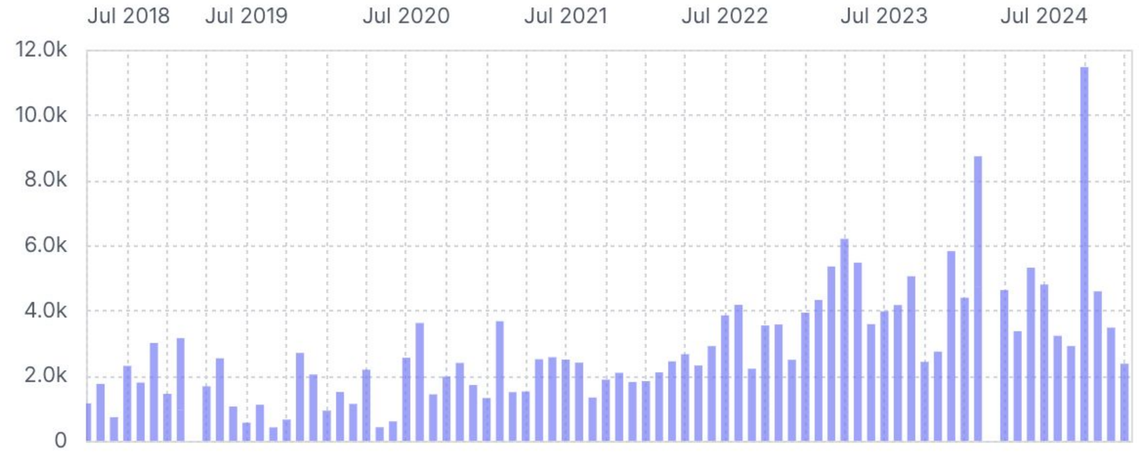
Fact tables are too raw, and…

date	commit_hash	commit_message	username
2024-10-02 07:53:52 Z	00a605270719941ca0412ad5d0a14b1bdfbf9eb5	compiling, GetExportOrder needs to be porte...	Tishj
2022-07-08 05:58:04 Z	4504caa679f5dc5e4f40e93af8da5ca81d0bd...	More parameter fixes	Mark Raasveldt
2022-01-11 03:30:15 Z	a342e1176485d05d3283dd78dfcdb928b6b3...	No need to protect	Kirill Müller
2023-04-24 11:20:16 Z	d475e5809051df6325bae74bafb87bbb0a8af...	refactor sink api for interrupts	Sam Ansmink
2023-11-24 11:17:20 Z	78115aad658c4e2a9b9921b786933862516b1...	test	
2024-10-02 07:53:52 Z	00a605270719941ca0412ad5d0a14b1bdfbf9eb5	com	
2022-10-19 14:01:16 Z	28eed524788dfa4caa3dbf7026b93cdd8db0...	Opt	
2022-06-06 12:46:45 Z	6e099b510cf72f5fa62b9e5a2c14fdc39e89d4...	Adc	
2022-10-21 07:56:57 Z	e5be393df64b5bfe0a33764445e7a9170aeca...	fixi	

Reports are too baked

Total Commits

220k



Metrics are flexible

A metric is an aggregate function whose value depends on the filters (predicates) placed on its dimensions.

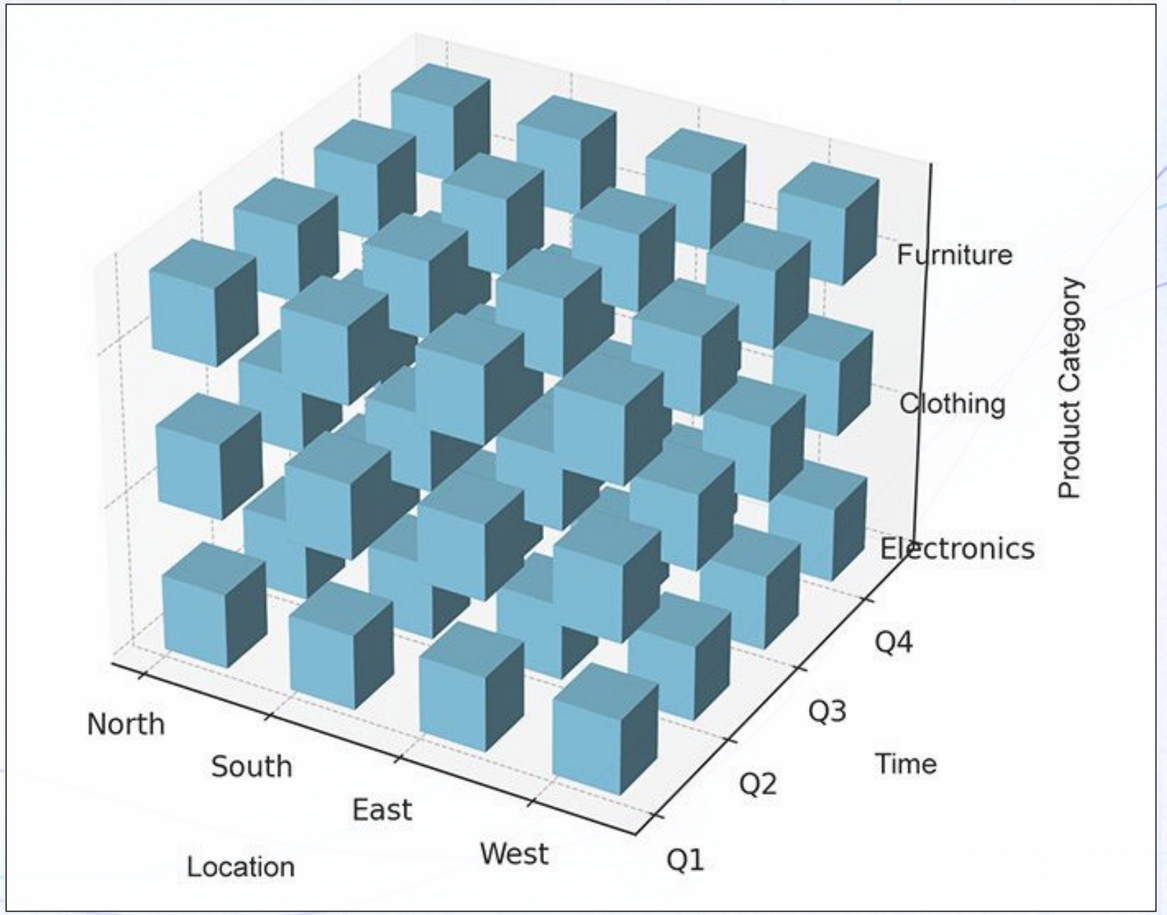
Analytics applications can flexibly define the predicate context(s).

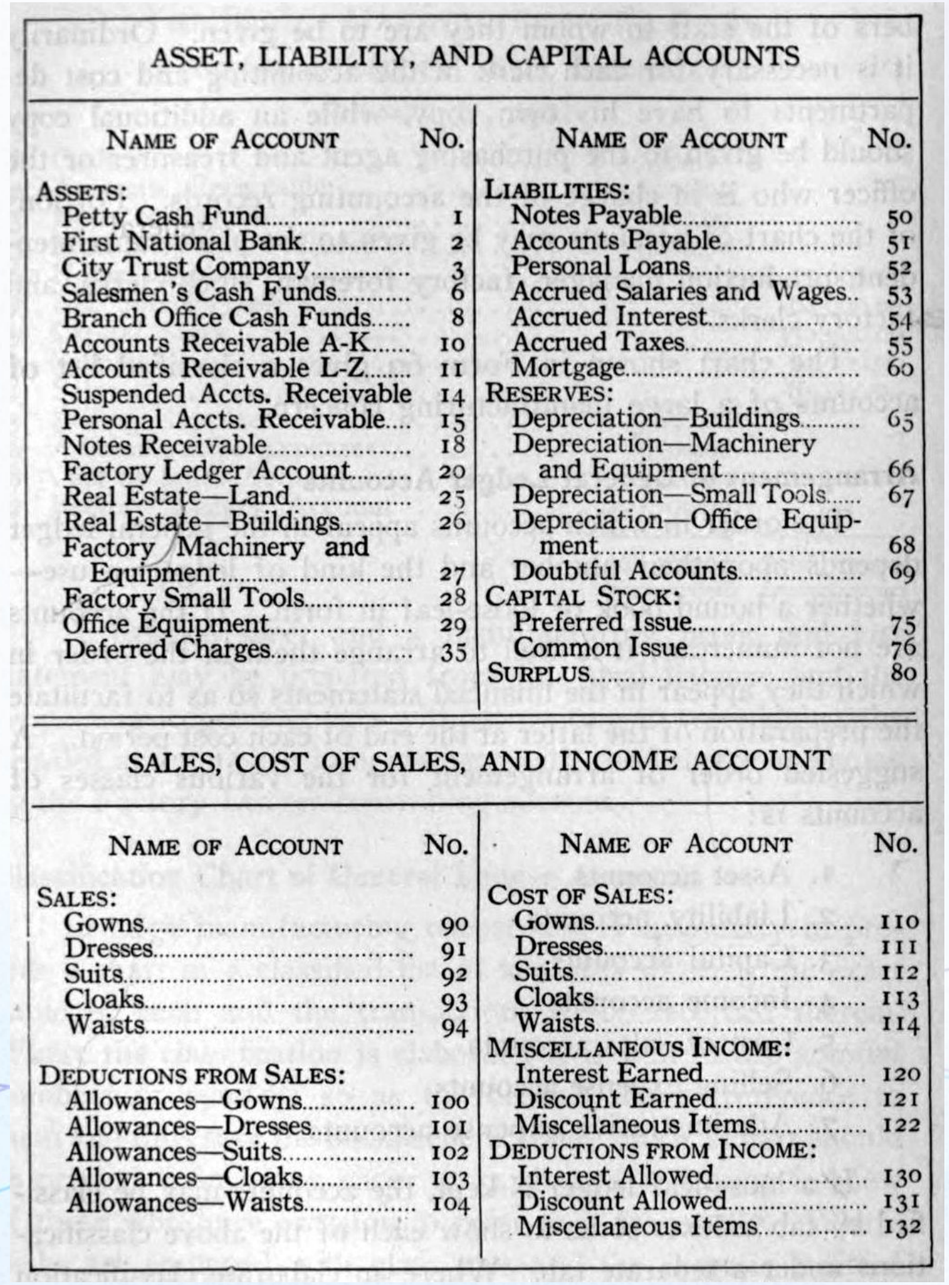
```
## revenue metric
SUM(sales) FROM transactions
WHERE <??>

SUM(sales) FROM transactions
WHERE country = 'NL'
AND date >= '2025-01-01'
```

Metrics are fast

Metrics can be calculated from partially aggregated data tables (OLAP cubes) which are often 10-100x smaller than fact tables.





ASSET, LIABILITY, AND CAPITAL ACCOUNTS

Name of Account	No.
Assets:	
Petty Cash Fund	1
First National Bank	2
City Trust Company	3
Salesmen's Cash Funds	6
Branch Office Cash Funds	8
Accounts Receivable A-K	10
Accounts Receivable L-Z	11
Suspended Accts. Receivable	14
Personal Accts. Receivable	15
Notes Receivable	18
Factory Ledger Account	20
Real Estate—Land	25
Real Estate—Buildings	26
Factory Machinery and Equipment	27
Factory Small Tools	28
Office Equipment	29
Deferred Charges	35
Liabilities:	
Notes Payable	50
Accounts Payable	51
Personal Loans	52
Accrued Salaries and Wages	53
Accrued Interest	54
Accrued Taxes	55
Mortgage	60
Reserves:	
Depreciation—Buildings	65
Depreciation—Machinery and Equipment	66
Depreciation—Small Tools	67
Depreciation—Office Equipment	68
Doubtful Accounts	69
Capital Stock:	
Preferred Issue	75
Common Issue	76
Surplus	80

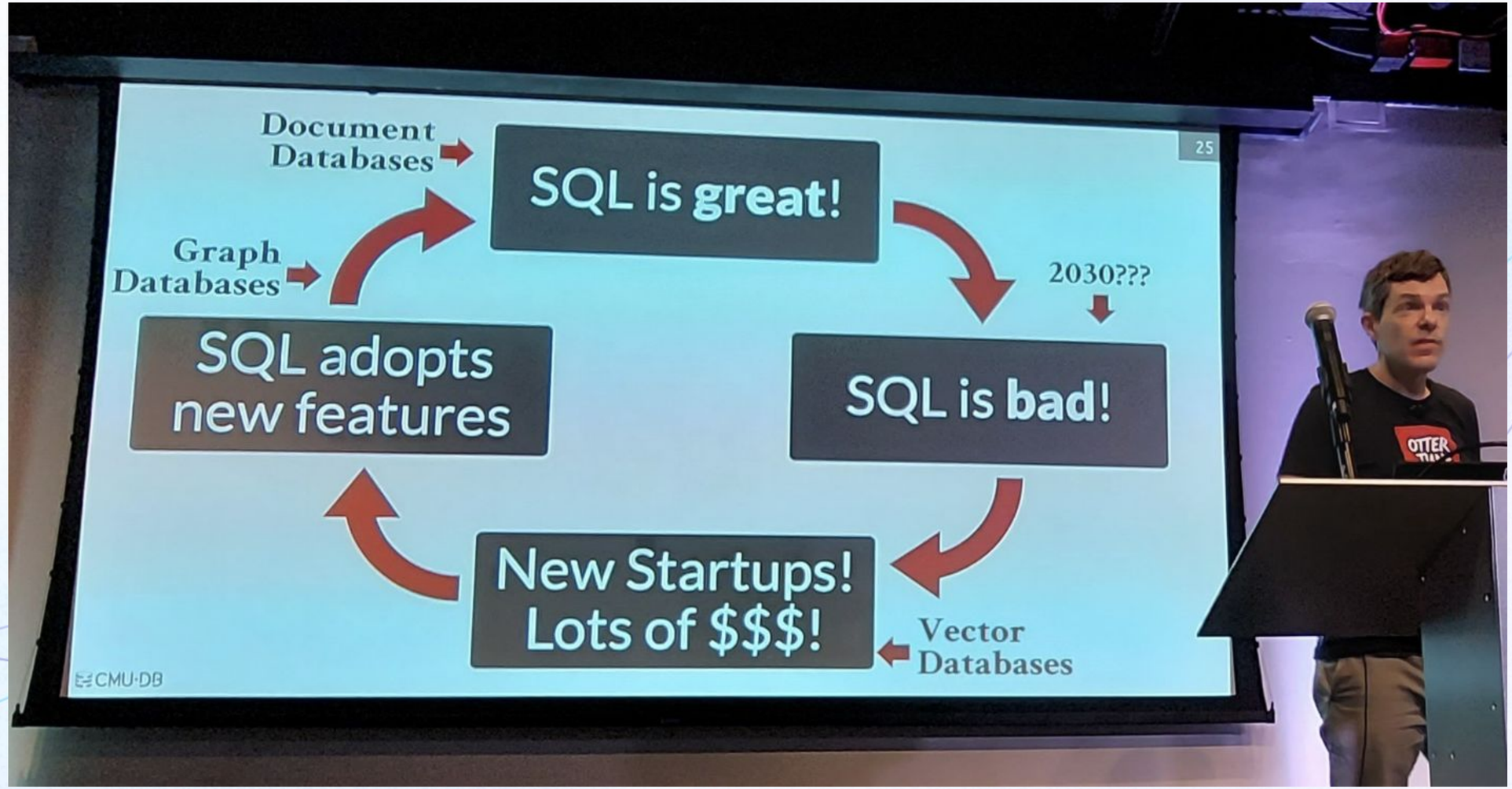
SALES, COST OF SALES, AND INCOME ACCOUNT

Name of Account	No.
Sales:	
Gowns	90
Dresses	91
Suits	92
Cloaks	93
Waists	94
Deductions from Sales:	
Allowances—Gowns	100
Allowances—Dresses	101
Allowances—Suits	102
Allowances—Cloaks	103
Allowances—Waists	104
Cost of Sales:	
Gowns	110
Dresses	111
Suits	112
Cloaks	113
Waists	114
Miscellaneous Income:	
Interest Earned	120
Discount Earned	121
Miscellaneous Items	122
Deductions from Income:	
Interest Allowed	130
Discount Allowed	131
Miscellaneous Items	132

Metrics are intuitive

Humans have been tabulating in ledgers, spreadsheets, and pivot tables for centuries.

Metrics should be expressed in SQL



Metrics should be expressed in SQL

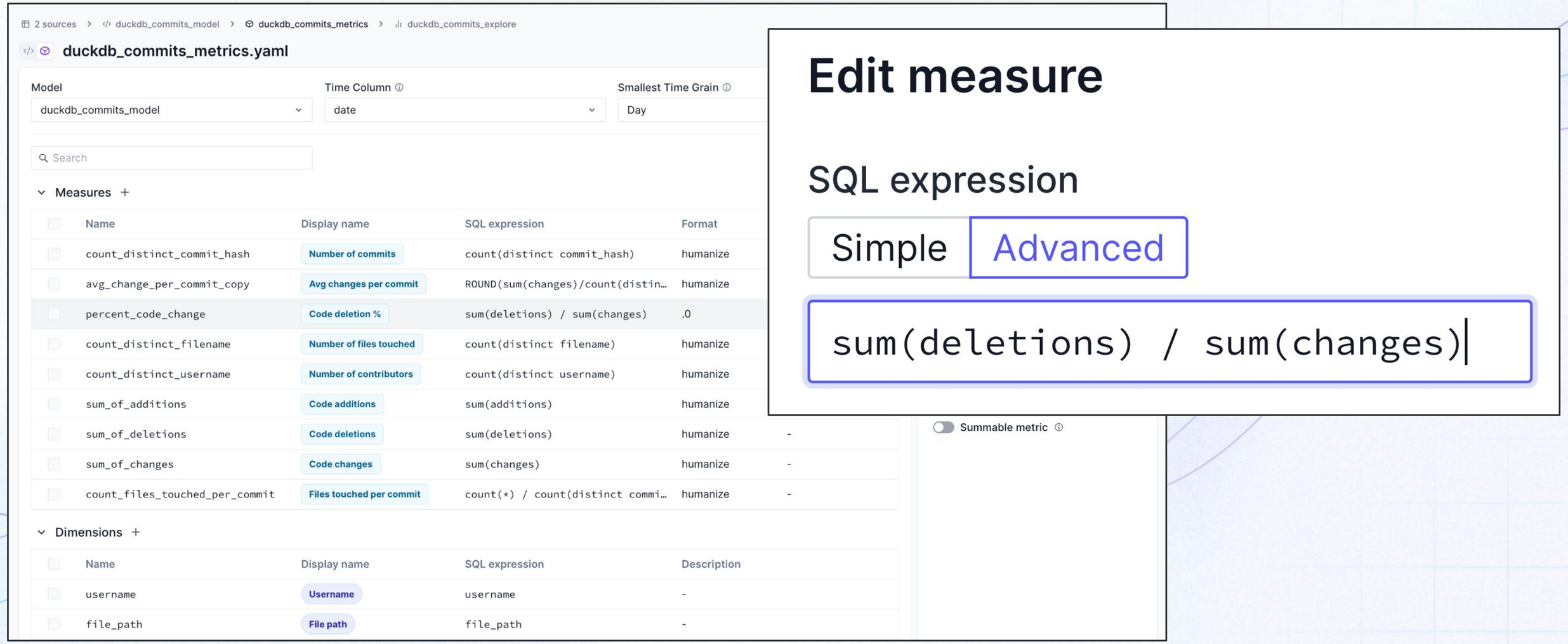
I don't want to learn your garbage query language

2018-08-30

This is a bit of a rant but I really don't like software that invents its own query language. There's a trillion different ORMs out there. Another trillion databases with their own query language. Another trillion SaaS products where the only way to query is to learn some random query DSL they made up.

I just want my SQL back. It's a language *everyone* understands, it's been around since the

Metrics should be expressed in SQL



Measures in SQL

Julian Hyde
Google Inc.
San Francisco, CA, USA
julianhyde@google.com

John Fremlin
Google Inc.
New York, NY, USA
fremlin@google.com

ABSTRACT

SQL has attained widespread adoption, but Business Intelligence tools still use their own higher level languages based upon a multidimensional paradigm. Composable calculations are what is missing from SQL, and we propose a new kind of column, called a measure, that attaches a calculation to a table. Like regular tables, tables with measures are composable and closed when used in queries.

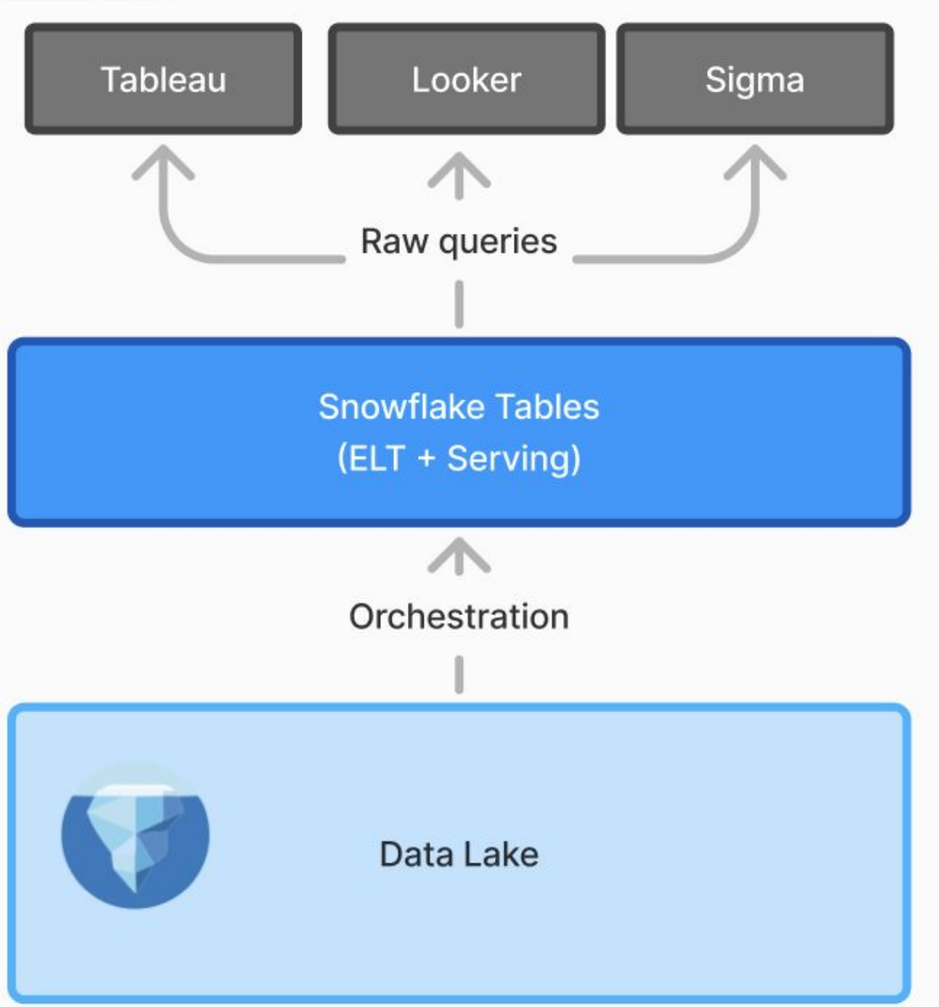
SQL-with-measures has the power, conciseness and reusability of multidimensional languages but retains SQL semantics. Measure invocations can be expanded in place to simple, clear SQL.

perhaps graphical, perhaps textual). They guide users in the construction of queries, and aid creation in visualizations and reports. But we believe that their core strength is the ability to express calculations in a concise manner, and to compose and reuse those calculations.

In this paper, we show that the relational model imposes repetition of filter expressions: changing the date range of a query requires updating many `WHERE` clauses. Therefore the challenge is how to extend the data model offered by SQL, in ways that do not change the semantics of currently valid SQL expressions or

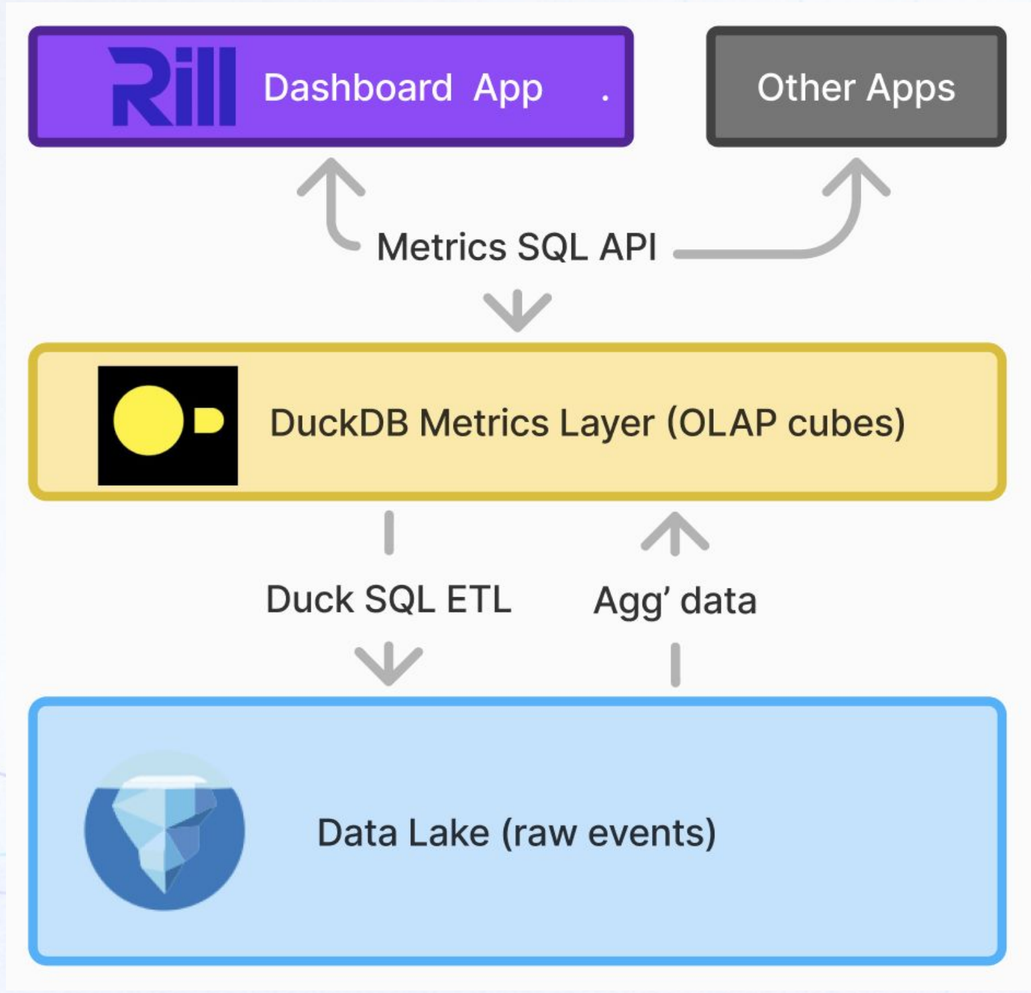
Architecture of a legacy metrics stack

- Inconsistent metrics definitions
- Slow table queries
- Many coding languages



Architecture of a DuckDB-powered metrics stack

- Consistent metrics definitions
- Sub-second OLAP queries
- SQL-everywhere stack



Challenges of a DuckDB-powered metrics stack

- Data modeling is required *up front* - no pain, no gain
- Metric changes can be expensive - rebuild OLAP
- Single-node scale still has its limits - << 1TB

DeepDuck: An AI Agent for DuckDB-Powered Metrics?

- AI can assist in the challenges of metrics modeling
  - Today 80%+ of Rill's metrics views and dashboards are first created with GPT-4o
  - More opportunity for basic optimizations like casting of data types (enums), column-ordering
- We predict conversational AI will prefer metrics over tables
  - Richer metadata of metrics provides more context to an agent
  - Faster queries mean more queries, faster time to autonomous insights

Inspiration & Further Reading

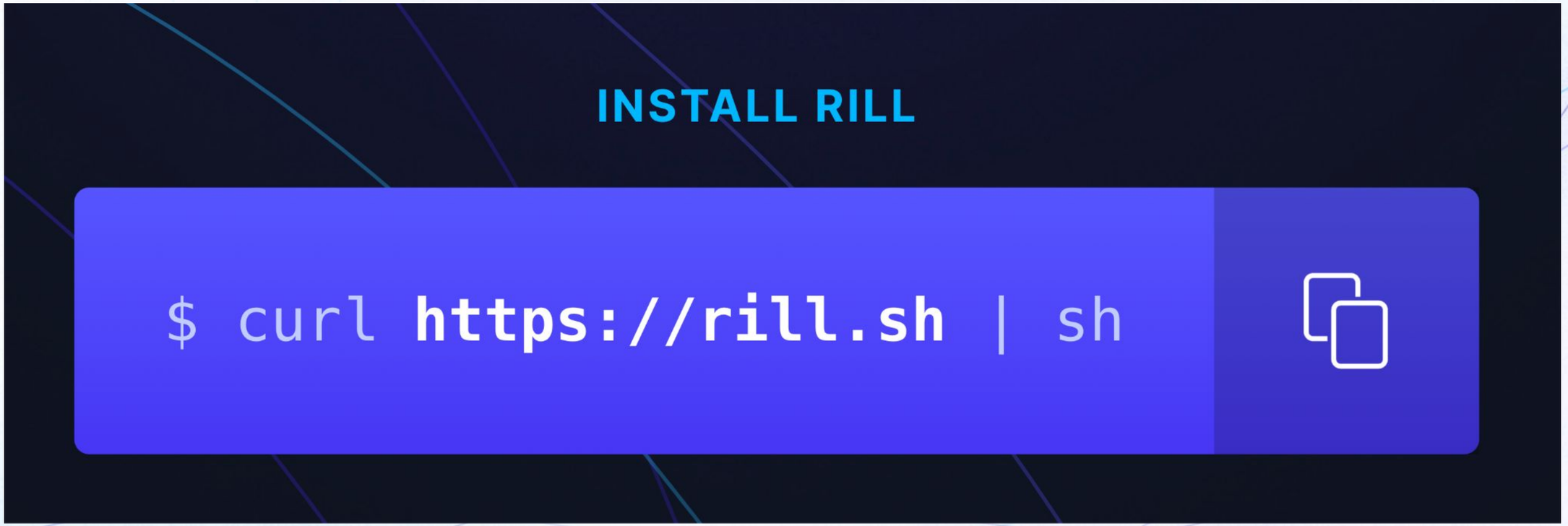
Hyde, J., & Fremlin, J. (2024). Measures in SQL. In Companion of the 2024 International Conference on Management of Data (pp. 31–40). Santiago, Chile: ACM. https://doi.org/10.1145/3626246.3653374

Handel, N. (2021, December 15). A Brief History of the Metrics Store. Towards Data Science. Retrieved from https://towardsdatascience.com/a-brief-history-of-the-metrics-store-28208ec8f6f1

Fowler, D. (2020, December 7). Kimball in the Context of the Modern Data Warehouse: What's Worth Keeping, and What's Not. Coalesce Conference 2020. YouTube. https://www.youtube.com/watch?v=3OcS2TMXELU

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling (3rd ed.). Wiley.

Thank you! Questions & Answers.



Visit www.rilldata.com to copy command