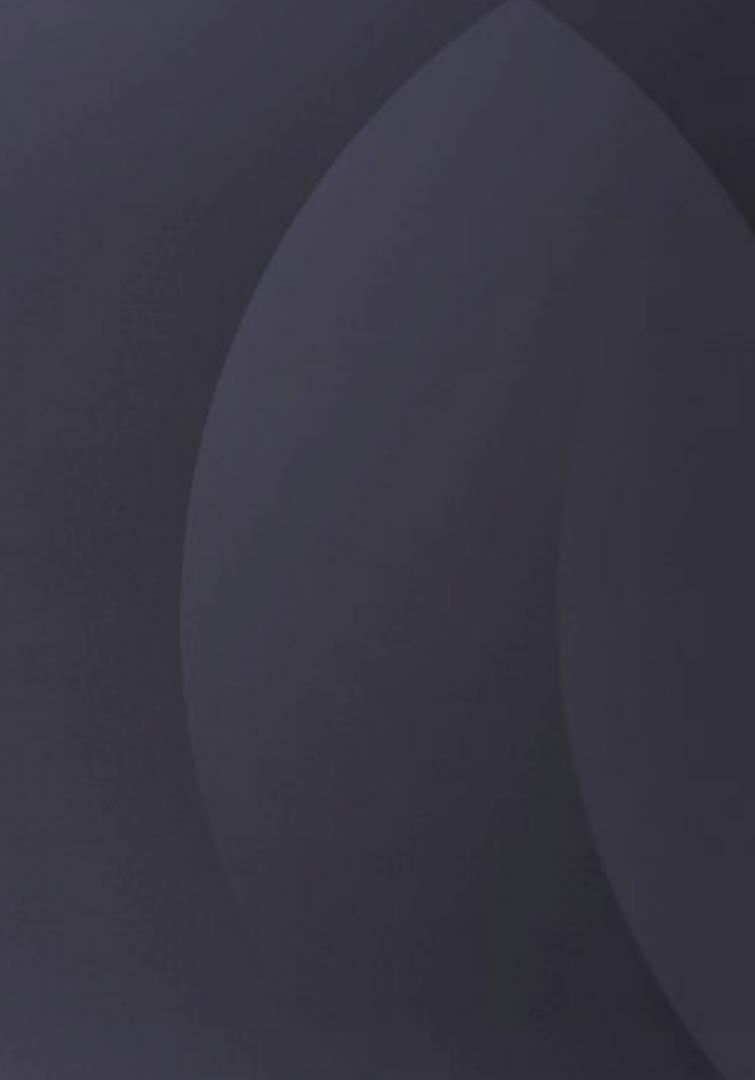


Data Engineering Rises Again

Escaping The Modern Data Trap With Composable, Embeddable Software

Nick Schrock Founder, Dagster Labs



What is the modern data stack?

What is the modern data stack?



The modern data stack (MDS) is a suite of tools used for data integration. These tools include, in order of how the data flows:

- a fully managed ELT data pipeline
- a cloud-based columnar warehouse or data lake as a destination
- a data transformation tool
- a business intelligence or data visualization platform.

What is the modern data stack?

Fivetran

Ultimately, the modern data stack lowers the technical barrier to entry for data integration. The components of the modern data stack are built with analysts and business users in mind, meaning that users of all backgrounds can not only easily use these tools, but also administer them without in-depth technical knowledge.

The Modern Data Stack has reshaped the data ecosystem with extremely positive benefits.

The Modern Data Stack has reshaped the data ecosystem with extremely positive benefits.

However, one goal was to eliminate data engineering.

The reality

The reality

Data engineers are more important than ever

The reality

Data engineers are more important than ever

They are poorly served by the current modern data stack



Spin-up to get started/prototype



Spin-up to get started/prototype



Can get lots of stakeholders involved



Spin-up to get started/prototype

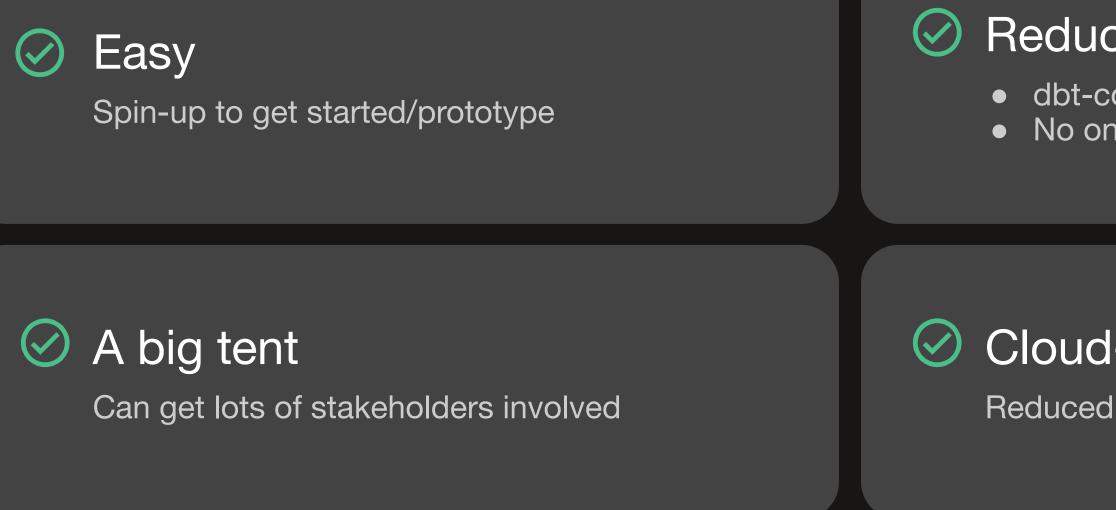




Can get lots of stakeholders involved

Reduced incidental complexity

• dbt-core, not custom templated SQL No one wants to write data movement code



Reduced incidental complexity

dbt-core, not custom templated SQL
No one wants to write data movement code

Cloud-native

Reduced fixed cost and better DX



Presumes homogenous storage and compute



Presumes homogenous storage and compute



Too many operational silos

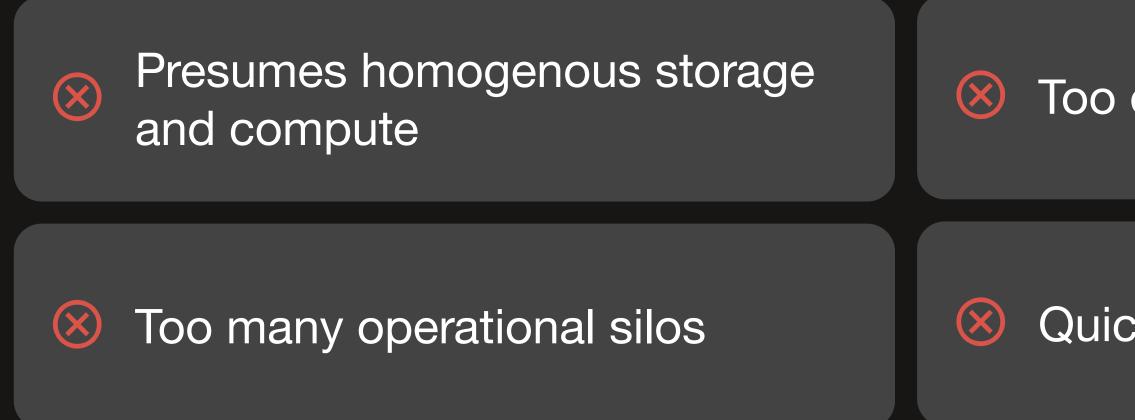


Presumes homogenous storage and compute



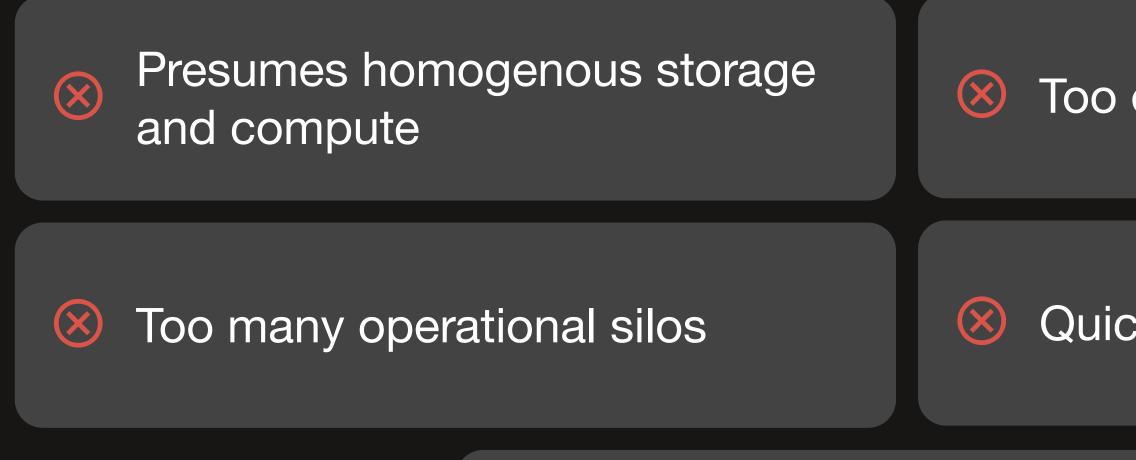
Too many operational silos

Too expensive



Too expensive

Quickly descends into chaos





Poor automation, flexibility, and composability

Too expensive

Quickly descends into chaos



Presumes homogenous storage and compute

Presumes homogenous storage and compute



Matthew Mullins @mullinsms

The lie of every MDS architecture diagram is that there is a single warehouse at the center of the business. I've almost never seen such an architecture.

Too many disconnected tools

Too many disconnected tools

THE 2023 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE

THE 2023 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE									
INFRASTRUCTURE		ANALYTICS	MACHINE LEARNING & ARTIFICIAL INTELLIGENCE						
STORAGE STO	PGStrem > byen	BI PLATFORMS Supervised and a supervised at the supervised at th	MACHINE LEARNING & ARTIFICIAL INTELLIGENCE DATA SCIENCE NOTEBOOKS DATA SCIENCE Numpyter COMPANY Company						
Progress in MarkLogic Countine DATASTAIC Countine DATAS	VECTOR DATABASES @Pinecone	anod ¹ ************************************	Sama Landov Contex a Second WHYLABS Contex Market Contex a Second WHYLABS Contex Market Predibase Market Contex a Second Predibase M						
DATA TRANSFORMATION Model talend alteryx Image: talend alteryx Im	ATALOG Wormstax AsaiPoint Collibre IERA Alassa Collibre metapher Derry Collibre MANUTA SternAna SternAna Securiti Statut Statu Securiti Statut Statu Securiti Statut Statu Collibre Securiti Statut Statu Securiti Statut Statu Securiti Minderri Derry Statut Statu Statut Statut Statut Statut Securiti Minderri Derry Statut Statut Statut Statut Statut Statut Statut Statut Securiti Minderri Statut Statut Statut Statut Statut Statut Securiti Minderri Statut Statut Sta	ANALYTICS Splunk> Splunk> Spl	VISION Siri Construction Siri						
ORCHESTRATION ASTRONUMER ASTRONUMER ASTRONUMER Immerser Constitution Without Astronumer Astronumer Constitution Without Astronumer Astronumer Constitution Without Astronumer Astronumer Astronumer Constitution Without Astronumer Astronumer Constitution Without Astronumer Astronumer Astronumer Astronumer Constitution Without Astronumer Astronumer Constitution Without Astronumer Astronumer Constitution Without Astronumer Astronumer Constitution Without Astronumer Constitution Without Astronumer Constitution Without Astronumer Constitution Without Constitution Constitution Constitution Constitution Constitution Constitution Constitution Constitution Constitution Constitution Constitution	With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 Mith Secure 10 With Secure 10 With Secure 10 With Secure 10 With Secure 10 Mith Secure 10 With Secure 10 With Secure 10 With Secure 10	QUERY ENGINE ENTERPRISE SEARCH Image: Constraint of the second	AI HARDWARE GPU CLOUD GPU CLOUD						
FRAMEWORKS FORMAT QUERY / DATA FLOW DATA ACCESS	MongoDB @reds // Influedoto	STRUCTURE MALAQUEZ MESSAGING LAN	OPEN SOURCE INFRASTRUCTURE TOOLS & MLOPS & AI INFRA AI FRAMEWORKS & LIBRARIES GUAGES AI MONT Bastani						

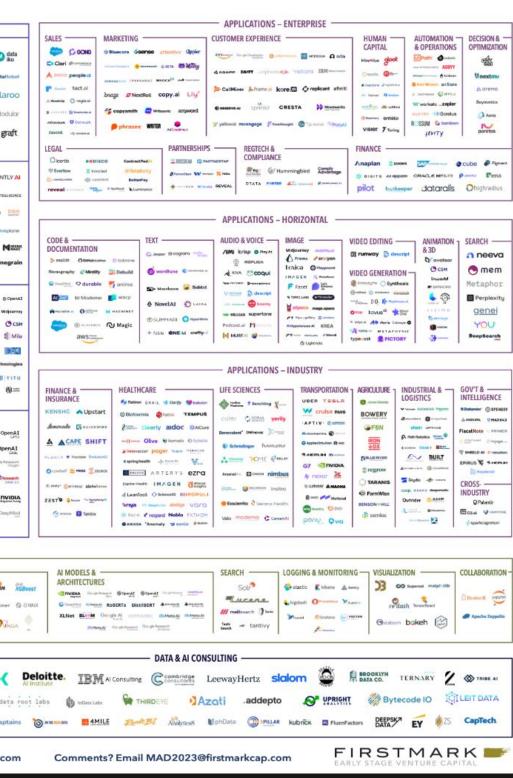
DATA SOURCES & APIs									
FRAMEWORKS FRAMEW	5世業 ①xnet 谷 Chainer								

		DATA SOURCES &	APIS			
DATA MARKETPLACES	FINANCIAL & MARKET DATA	AIR / SPACE / SEA	PEOPLE / ENTITIES	LOCATION INTELLIGENCE	ESG	QuantumBlack BCG
& DISCOVERY	Bloomberg C THENSON NEUTERS D DOW JONES Management	O constant Aspire PRELIGE		rounsouwer 🕐 mapbox 💥 weighter 🕜 sensitie	N'INNEN SUSTAINALYTICS MSCI . TRUVALUE LABE	Al by McKinsey
aws @ soundar @ DAWI	X BALLE CONCERNENT AND THE WORLD SAVES		Epsilon klue Standwatch			
≫ explorer IIM seconce № MAMO	xignite Qinteneo Educatento earnest erigne MX tink?	LIST SHARES.	Quantcast Cognism			
Denne Lante. O Seterate Summer 23	M CANDE	Synspective E/ eagleview ICEYE	The same mellissi zignali destruction	A Radar O Placer.al & Cuebiq Denseventing 200"	CCLARITY AL Arcadia novisto 😳	MERU BIG 💑 Data Capta
			1			

Version 1.0 - Feb 2023 © Matt Turck (@mattturck), Kevin Zhang (@ykevinzhang) & FirstMark (@firstmarkcap)

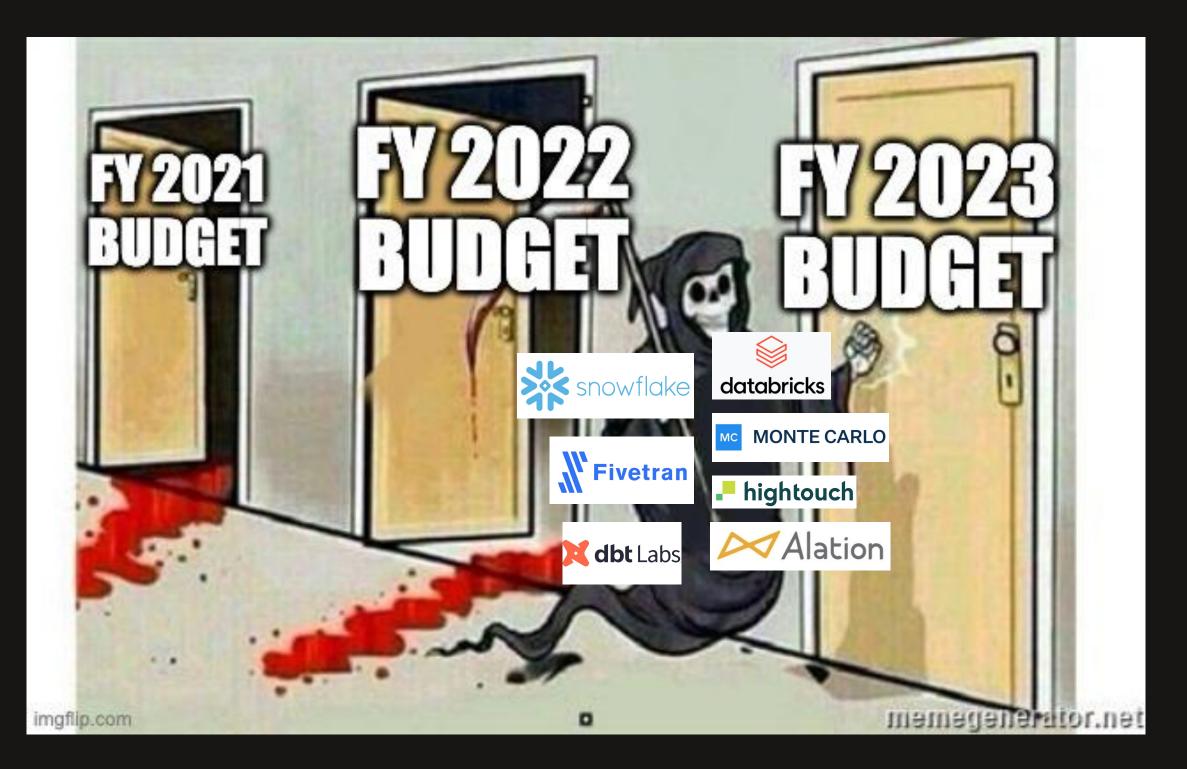
Blog post: mattturck.com/MAD2023

Interactive version: MAD.firstmarkcap.com



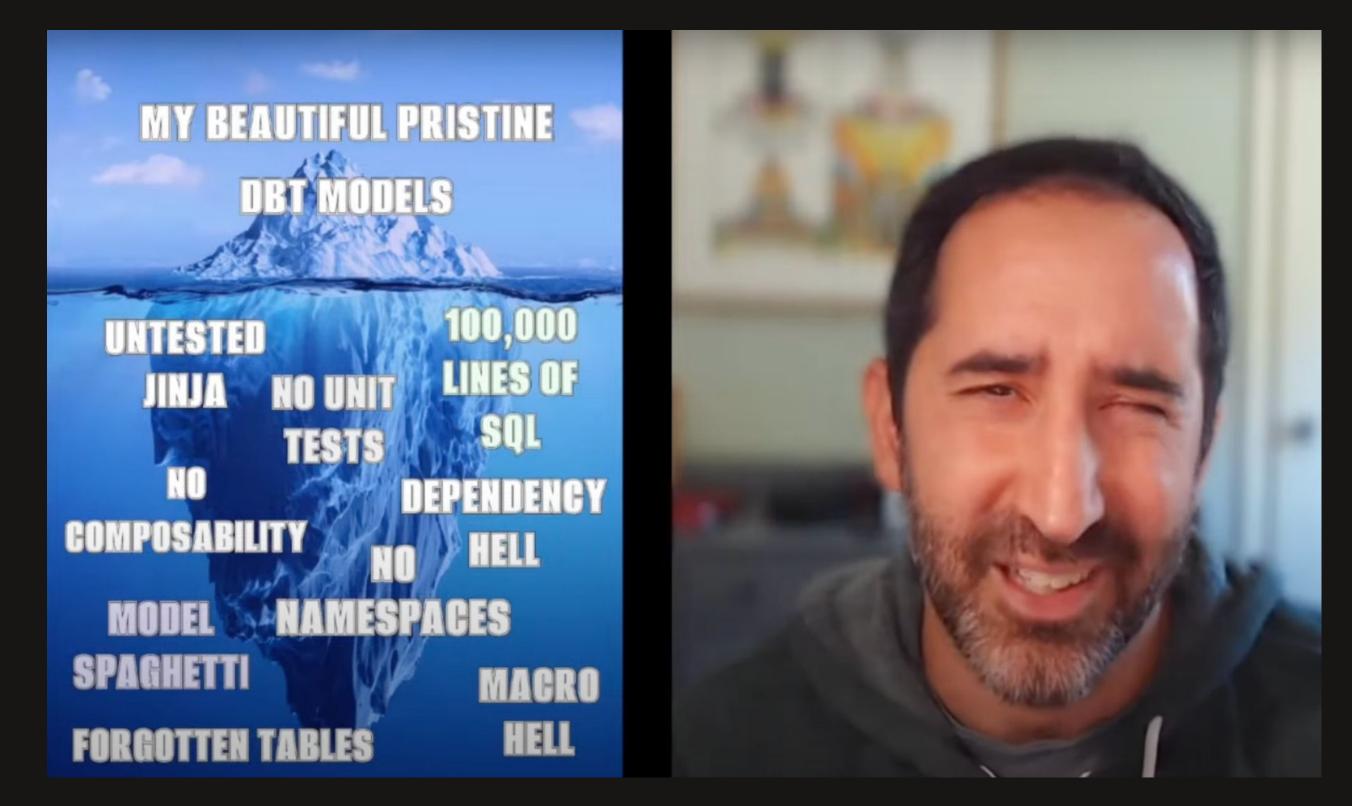
Too expensive

Too expensive



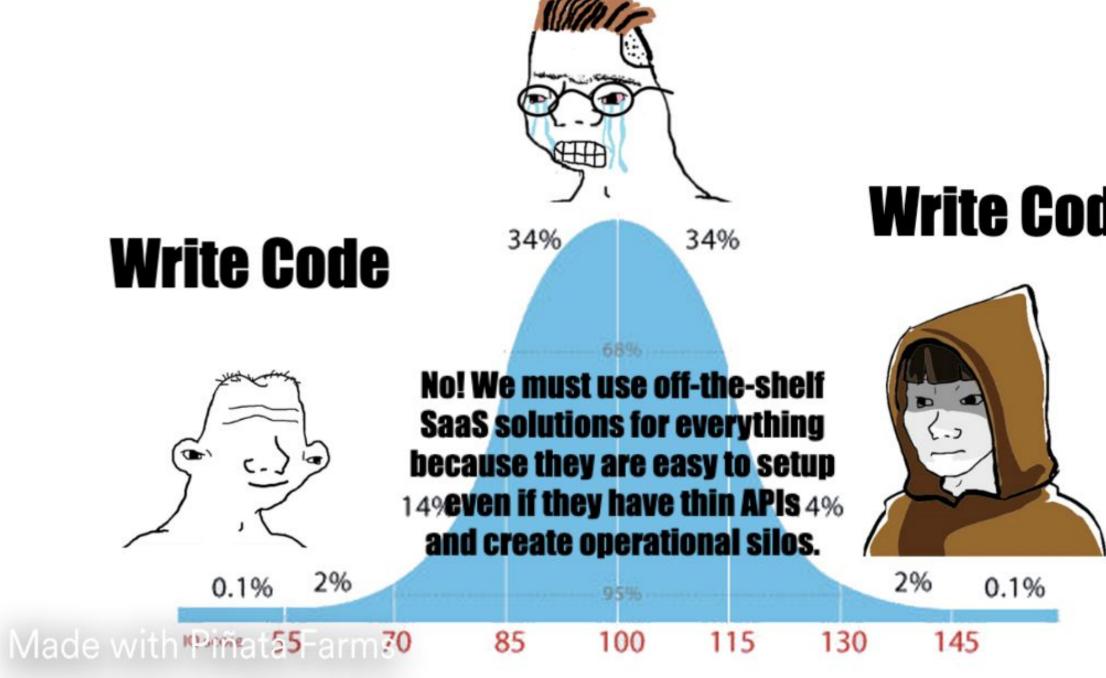
Descends into chaos

Descends into chaos

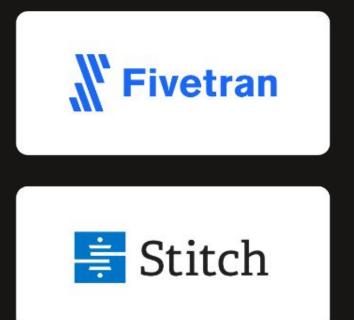


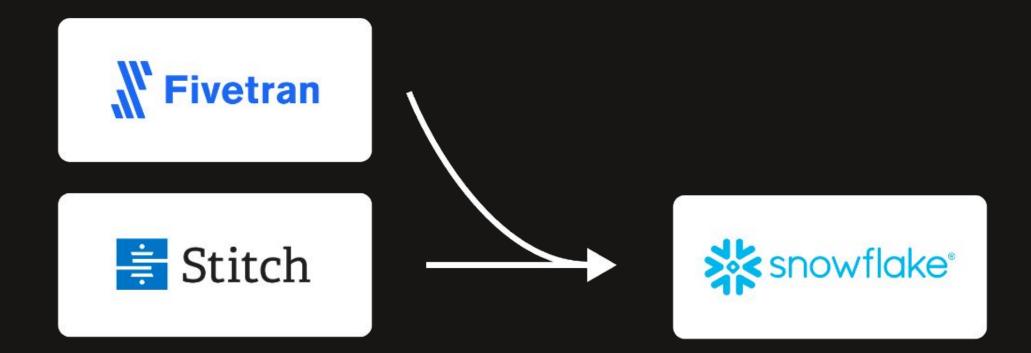
Poor automation, flexibility, composability

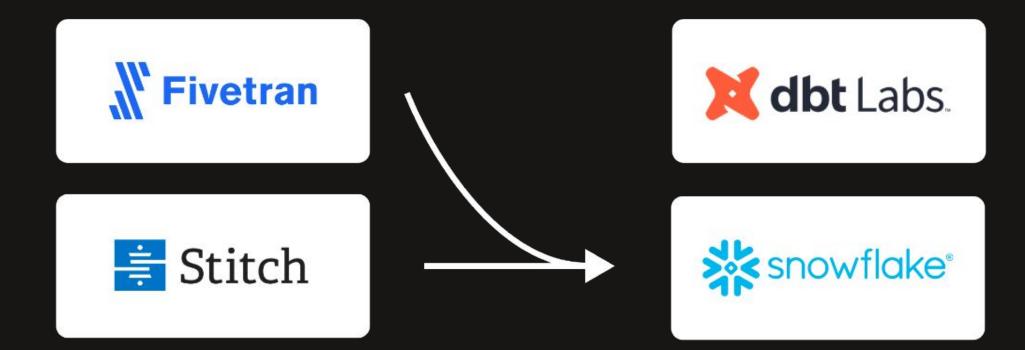
Poor automation, flexibility, composability

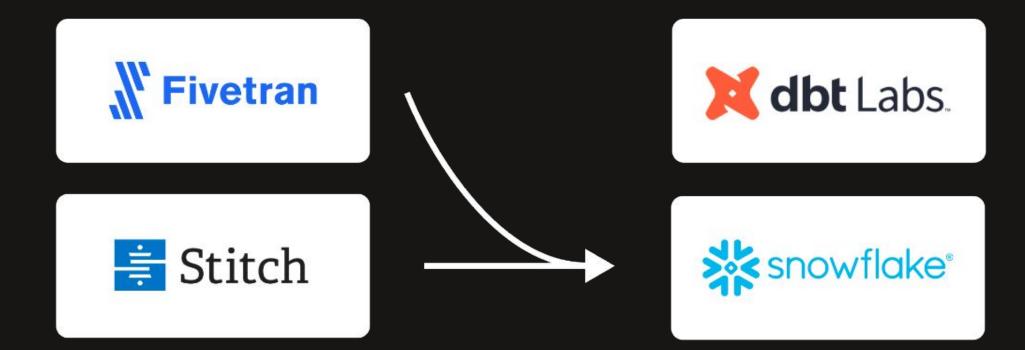


Write Code

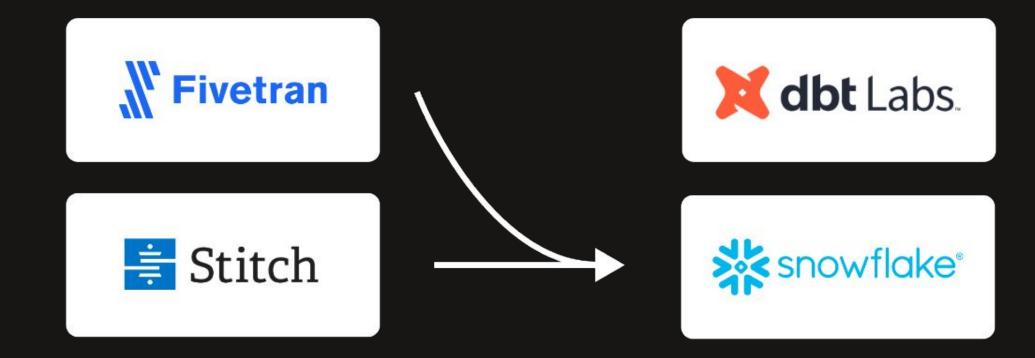






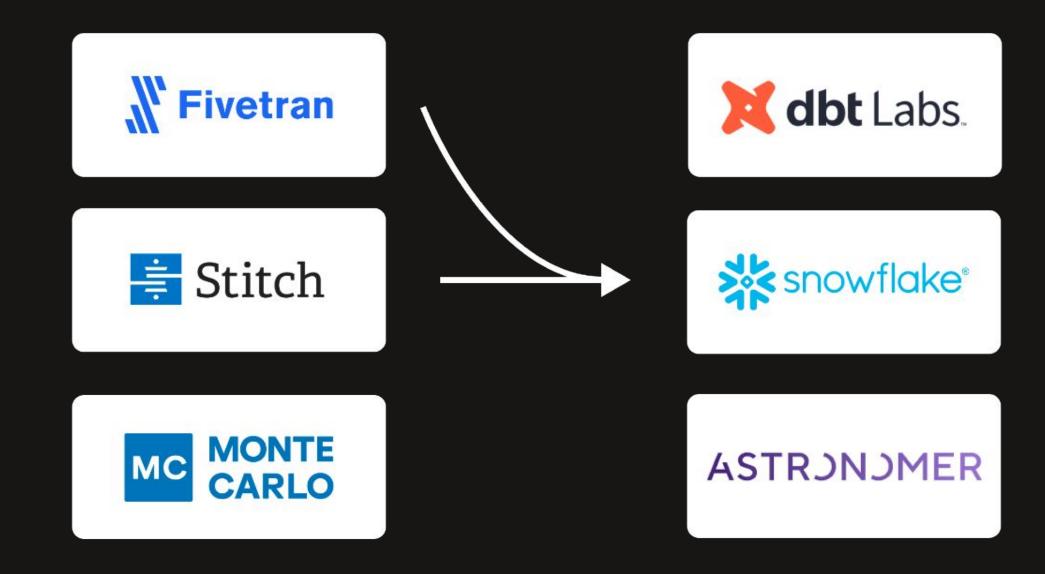




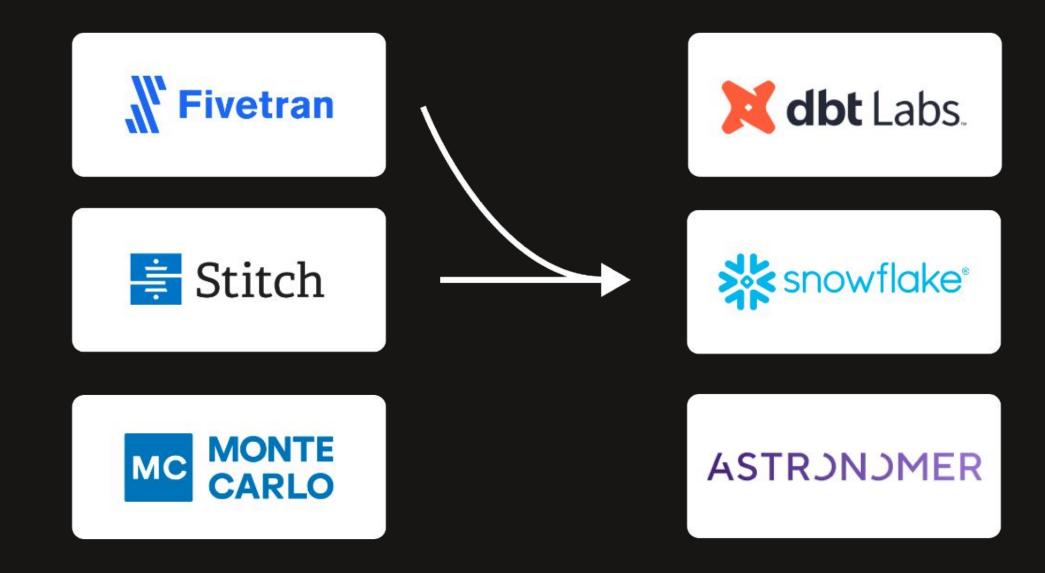


ASTRONOMER



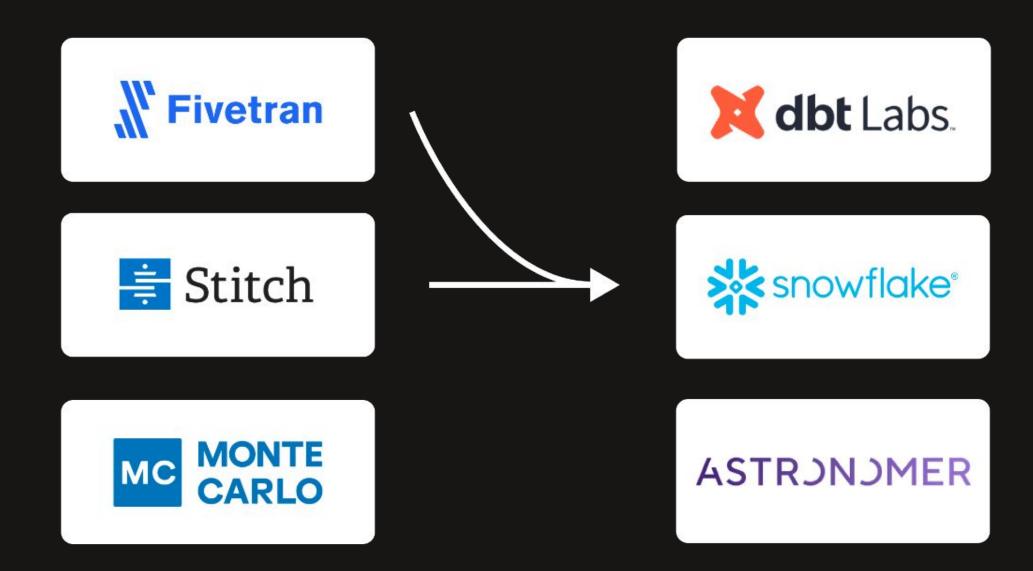








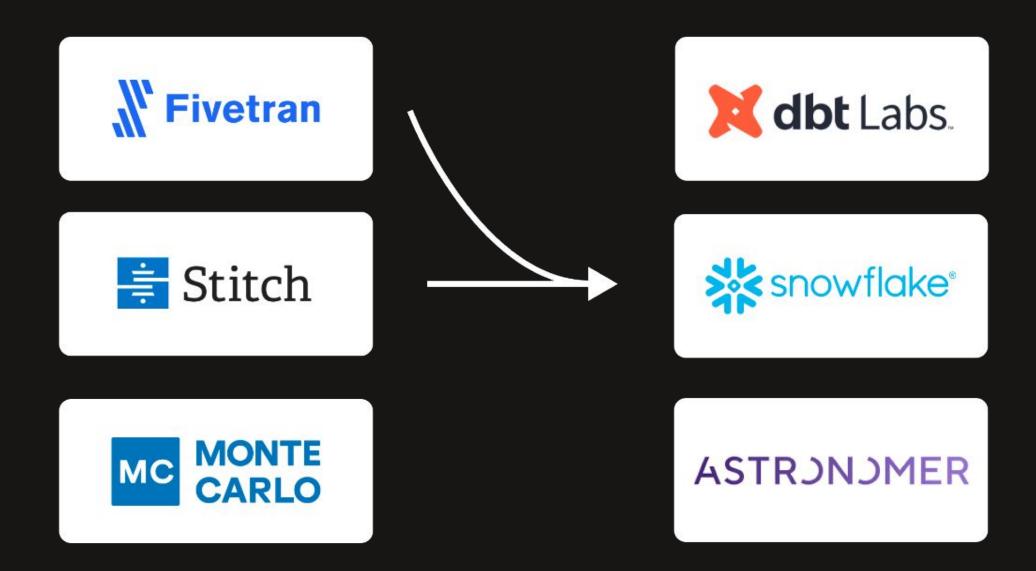
stemma



All this to build a single quality end-to-end data pipeline with searchable metadata



stemma



All this to build a single quality end-to-end data pipeline with searchable metadata



• Seven vendors to negotiate Four schedulers to babysit Six UIs to operate

Standalone tools or embedded libraries?

Standalone tools or embedded libraries?



Standalone tools or embedded libraries?



Standalone tools or embedded libraries?



Depends on context!



Overserve engineers

Could be just embedded libraries

Overserve engineers

Could be just embedded libraries

Underserve engineers

Thin APIs, inflexible, hard to automate

Overserve engineers

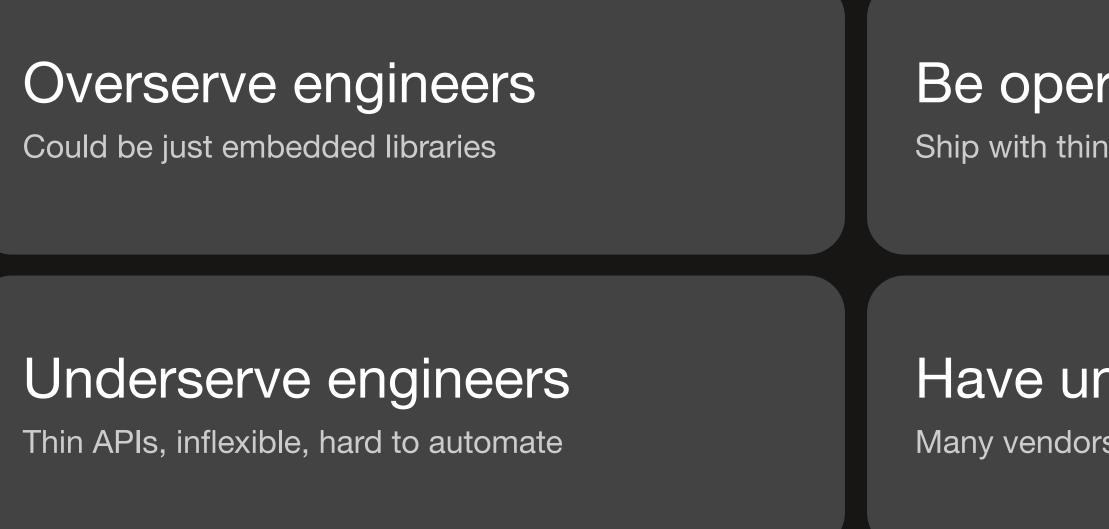
Could be just embedded libraries

Be operationally fragile

Ship with thin, underpowered orchestrators

Underserve engineers

Thin APIs, inflexible, hard to automate



Be operationally fragile

Ship with thin, underpowered orchestrators

Have unpredictable costs

Many vendors, many pricing models, poor visibility

Are standalone services the only option?

Adopting off-the-shelf SaaS tools without a unified control plane

Adopting off-the-shelf SaaS tools without a unified control Dane

"Easy" but disconnected tools often end up quickly creating more complexity



Builds on the good parts of the MDS

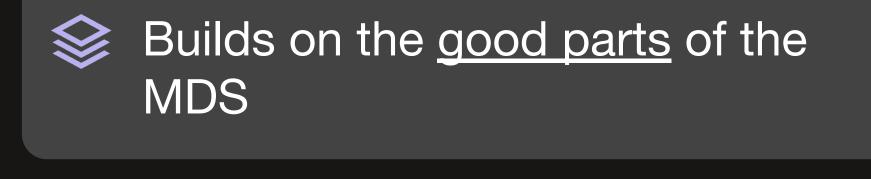
Builds on the good parts of the MDS

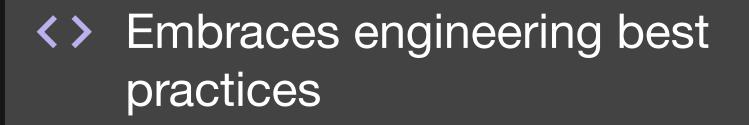
Embraces engineering best practices

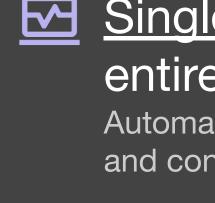
Builds on the good parts of the MDS

Embraces engineering best practices

A <u>composable</u> toolkit for data engineers





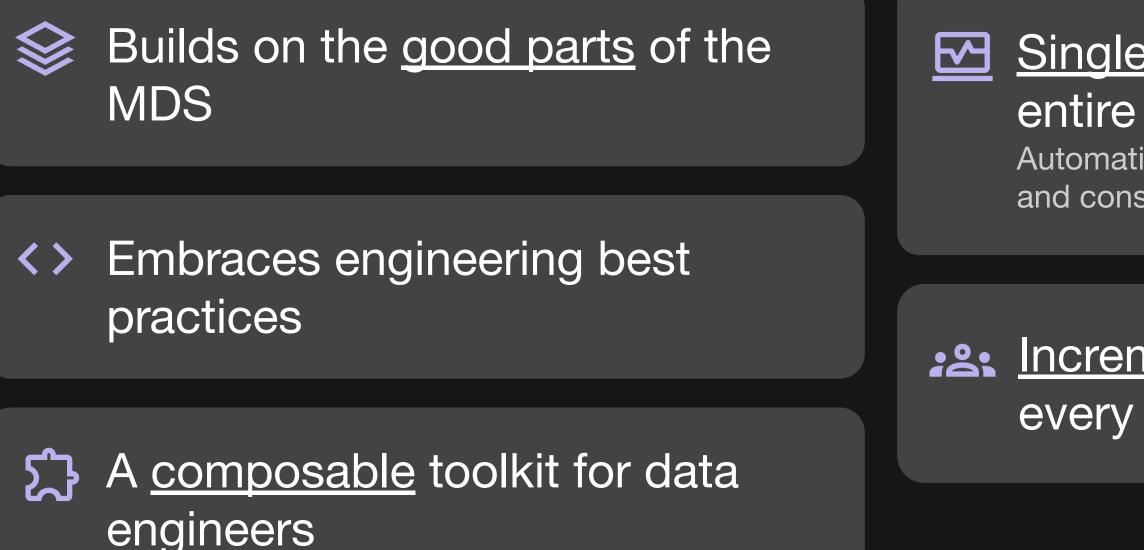


A <u>composable</u> toolkit for data engineers

Single pane of glass for the

entire data platform

Automation, observability, lineage, data quality, and consumption management



Single pane of glass for the entire data platform

Automation, observability, lineage, data quality, and consumption management

Incremental adoption across every stakeholder team





Thank you

