OSA CON 23

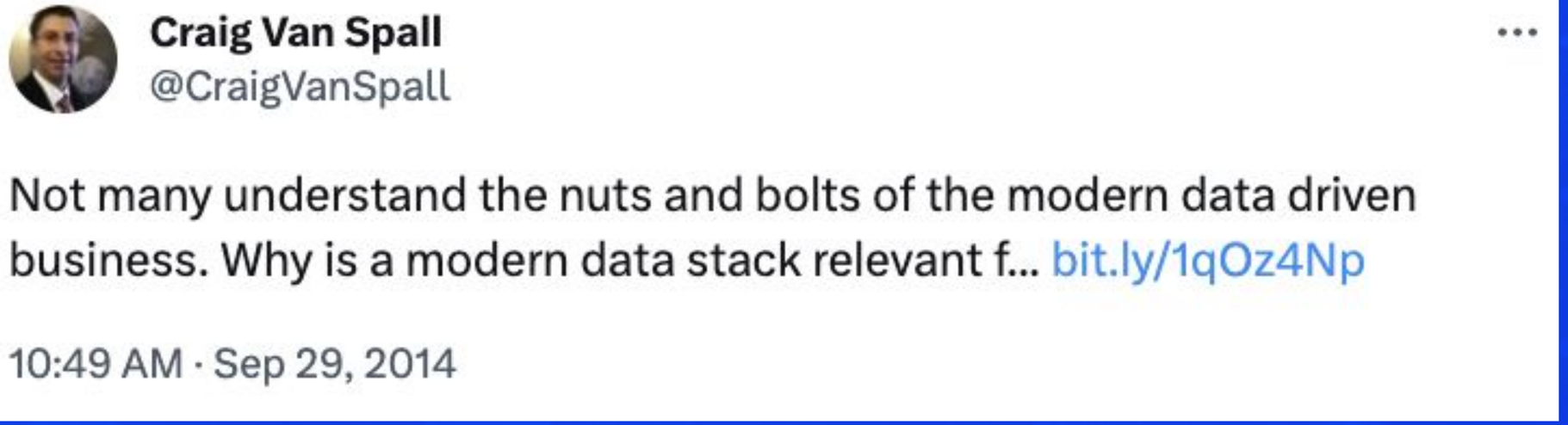
MAYBE THE REAL MODERN DATA STACK WAS THE OSS WE GOT ALONG THE WAY

Pedram Navid
Head of Data Engineering, Dagster Labs
@pdrmnvd / linkedin.com/in/pedramnavid

December 12-14 2023

Modern Data Stack: Origins

MDS: Origins



Craig Van Spall
@CraigVanSpall

Not many understand the nuts and bolts of the modern data driven business. Why is a modern data stack relevant f... bit.ly/1qOz4Np

10:49 AM · Sep 29, 2014

MDS: Origins

Product Announcements

What you need to know about the sunset of consumer Google+ on April 2

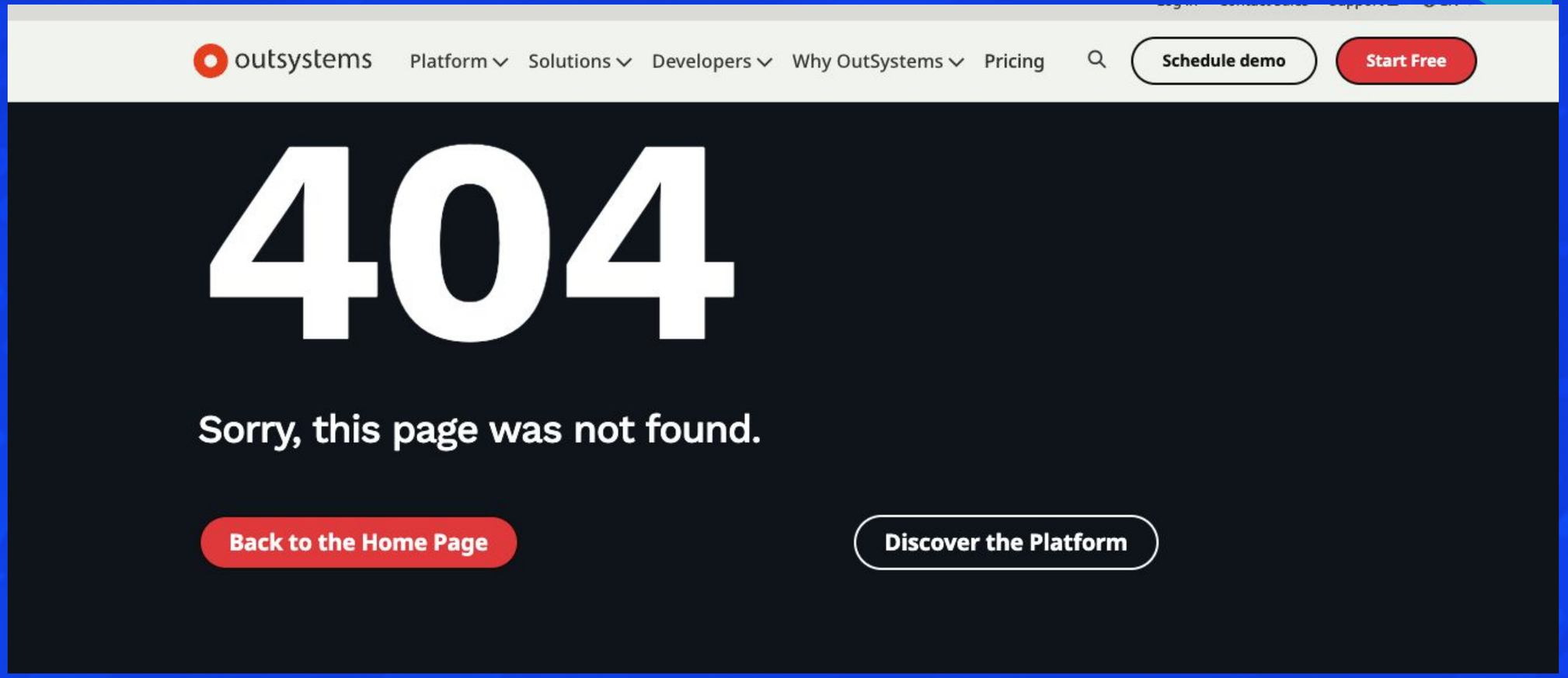
January 30, 2019

David Conway
Product Manager

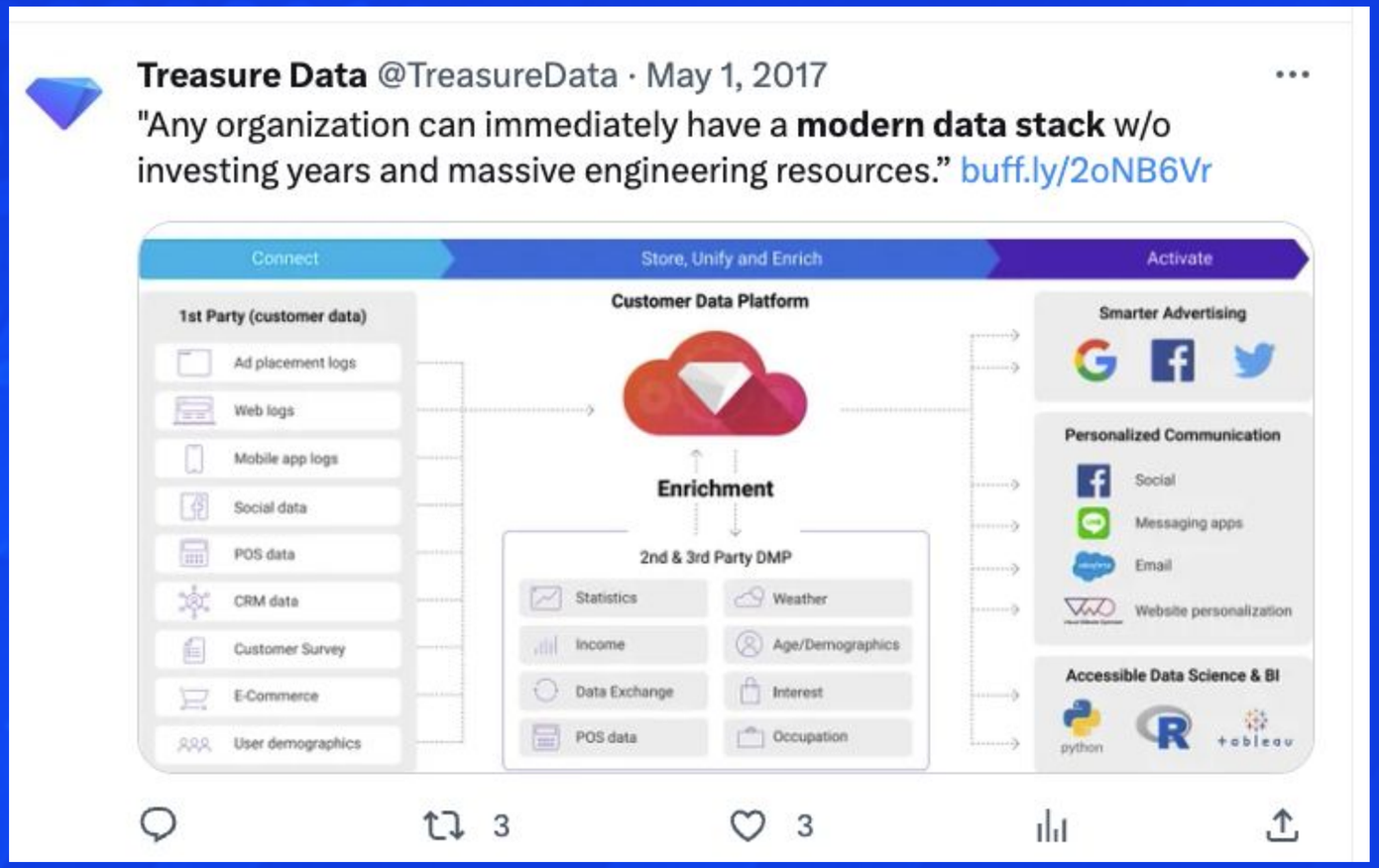
MDS: Origins

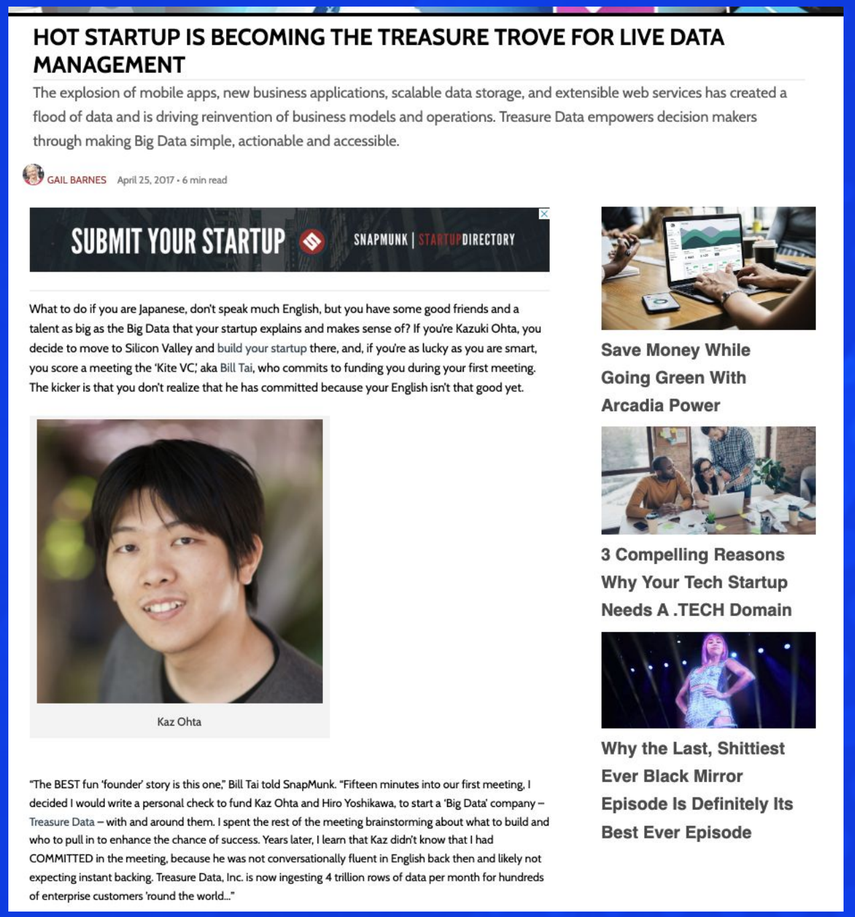
Gonçalo Gaiolas
@ggaiolas
Come build a modern #data stack to completely transform how a company utilizes to serve its customers. #datascience lnkd.in/dB5_Kx8
7:56 AM · Oct 3, 2016

MDS: Origins

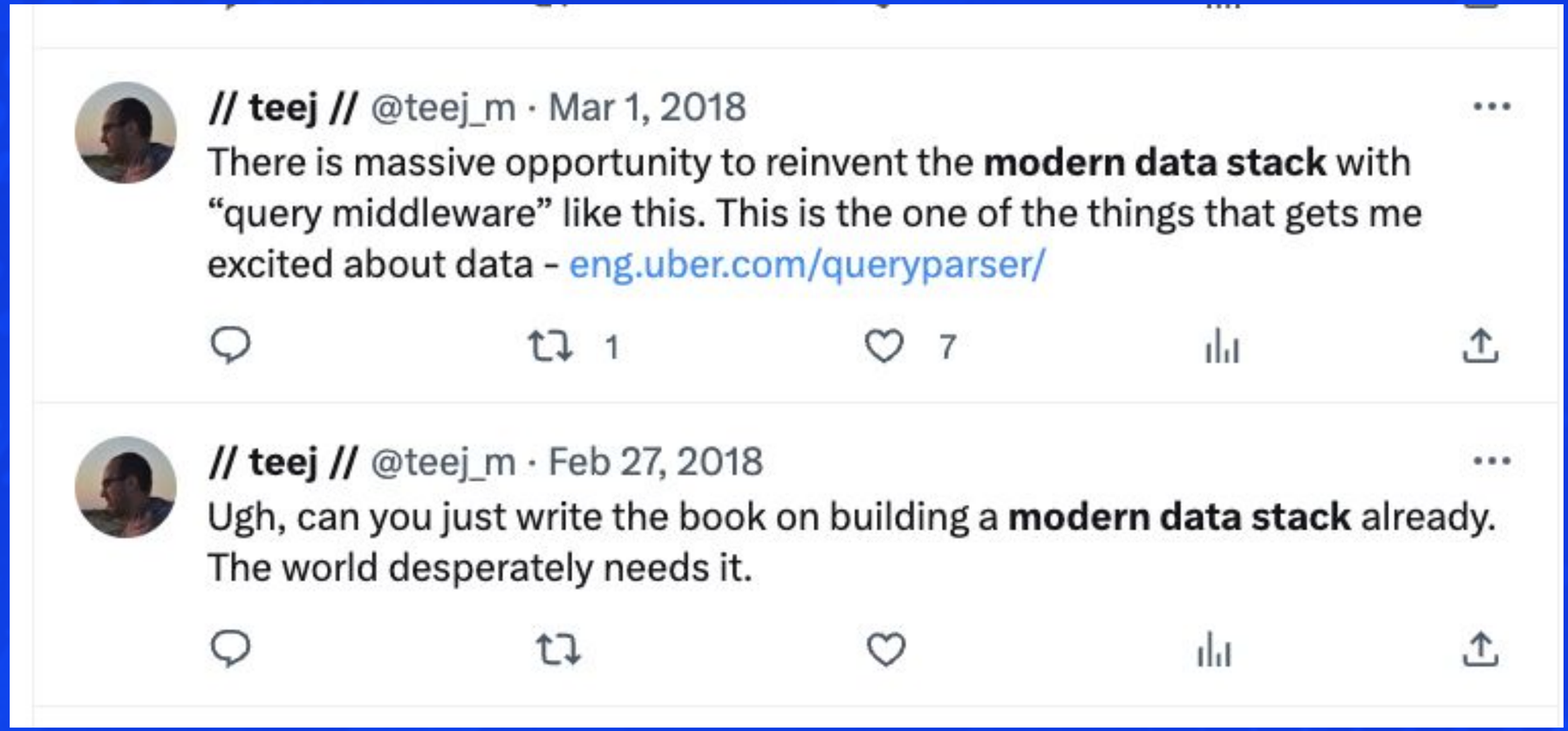

outsystems
Platform
Solutions
Developers
Why OutSystems
Pricing
Schedule demo
Start Free
404
Sorry, this page was not found.
Back to the Home Page
Discover the Platform

MDS: Origins


Treasure Data @TreasureData · May 1, 2017
"Any organization can immediately have a modern data stack w/o investing years and massive engineering resources." buff.ly/2oNB6Vr
Connect
Store, Unify and Enrich
Activate
1st Party (customer data)
Ad placement logs
Web logs
Mobile app logs
Social data
POS data
CRM data
Customer Survey
E-Commerce
User demographics
Customer Data Platform
Enrichment
2nd & 3rd Party DMP
Statistics
Weather
Income
Age/Demographics
Data Exchange
Interest
POS data
Occupation
Smarter Advertising
Personalized Communication
Social
Messaging apps
Email
Website personalization
Accessible Data Science & BI
python
tableau
3
3


HOT STARTUP IS BECOMING THE TREASURE TROVE FOR LIVE DATA MANAGEMENT
The explosion of mobile apps, new business applications, scalable data storage, and extensible web services has created a flood of data and is driving reinvention of business models and operations. Treasure Data empowers decision makers through making Big Data simple, actionable and accessible.
GAIL BARNES April 25, 2017 • 6 min read
SUBMIT YOUR STARTUP
SNAPMUNK | STARTUPDIRECTORY
What to do if you are Japanese, don't speak much English, but you have some good friends and a talent as big as the Big Data that your startup explains and makes sense of? If you're Kazuki Ohta, you decide to move to Silicon Valley and build your startup there, and, if you're as lucky as you are smart, you score a meeting the 'Kite VC,' aka Bill Tai, who commits to funding you during your first meeting. The kicker is that you don't realize that he has committed because your English isn't that good yet.
Kaz Ohta
"The BEST fun 'founder' story is this one," Bill Tai told SnapMunk. "Fifteen minutes into our first meeting, I decided I would write a personal check to fund Kaz Ohta and Hiro Yoshikawa, to start a 'Big Data' company – Treasure Data – with and around them. I spent the rest of the meeting brainstorming about what to build and who to pull in to enhance the chance of success. Years later, I learn that Kaz didn't know that I had COMMITTED in the meeting, because he was not conversationally fluent in English back then and likely not expecting instant backing. Treasure Data, Inc. is now ingesting 4 trillion rows of data per month for hundreds of enterprise customers 'round the world..."
Save Money While Going Green With Arcadia Power
3 Compelling Reasons Why Your Tech Startup Needs A .TECH Domain
Why the Last, Shittiest Ever Black Mirror Episode Is Definitely Its Best Ever Episode

MDS: Origins



// teej // @teej_m · Mar 1, 2018

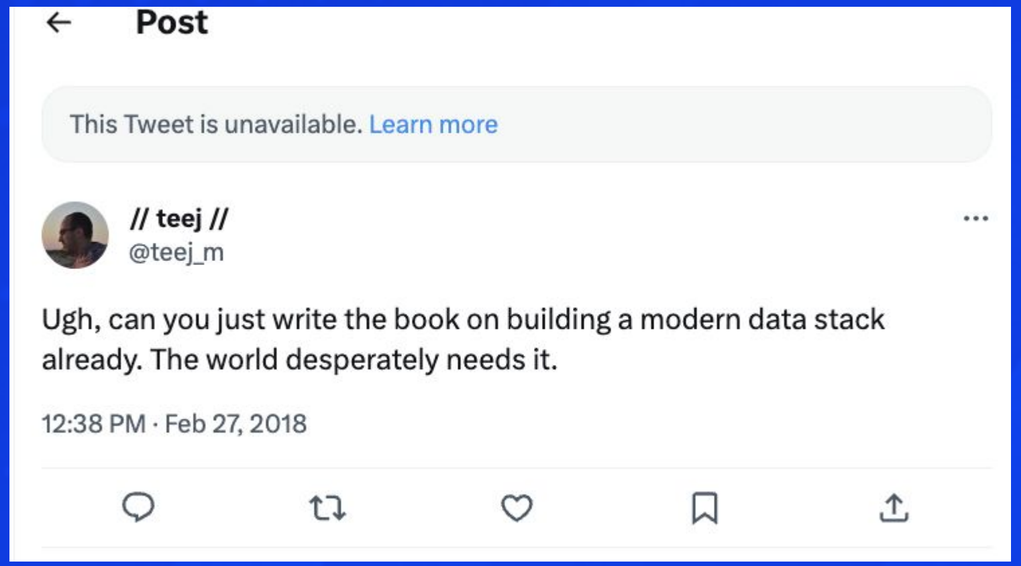
There is massive opportunity to reinvent the **modern data stack** with "query middleware" like this. This is the one of the things that gets me excited about data - eng.uber.com/queryparser/

1 7

// teej // @teej_m · Feb 27, 2018

Ugh, can you just write the book on building a **modern data stack** already. The world desperately needs it.

MDS: Origins


Post
This Tweet is unavailable. Learn more
// teej //
@teej_m
Ugh, can you just write the book on building a modern data stack already. The world desperately needs it.
12:38 PM · Feb 27, 2018

Mode Analytics @ModeAnalytics · Apr 17, 20

Mo
be
da

Jake Klein (Taylor's Version) @JakeKlein94 · Mar 13, 2019

Wanna do big data but not really sure where to get started? Check out this

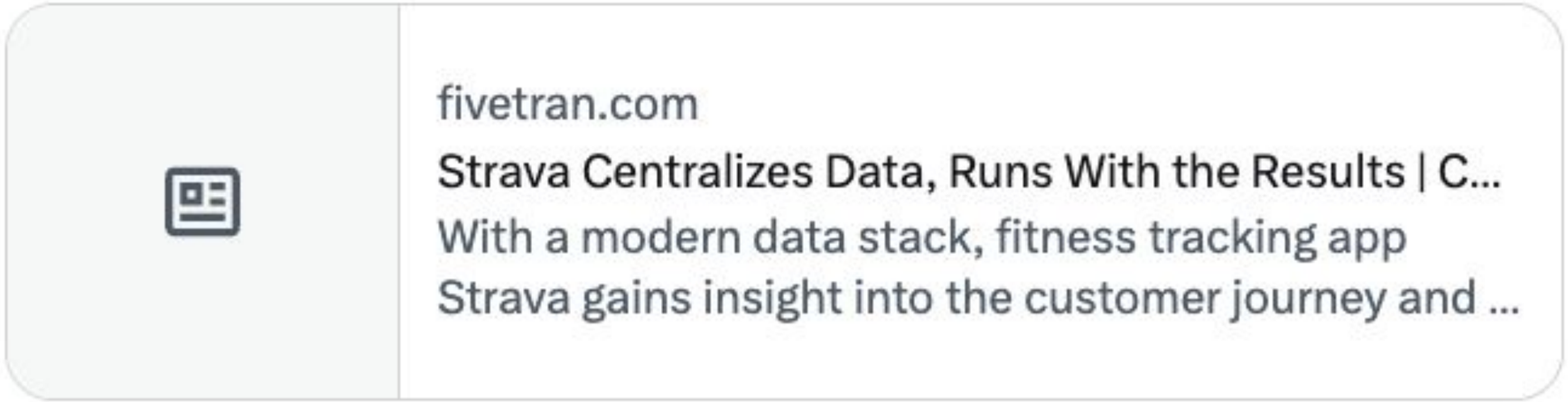
he

SPEt

Snowflake @SnowflakeDB · Aug 9, 2019

With a modern #data stack, fitness tracking app @Strava gains insight into the customer journey and makes more informed marketing decisions. #SnowflakeDB❄

Learn how Snowflake + @fivetran + @tableau teamed to make this a reality:



OpsMat
The lates
Data Sta
Accelera
ow.ly/xQ

Softwar
How to E

Aditya A
From Da
Data Sta

Daryl Ki
From Da
Data Sta

Big Data
The **Mod**
bigdataldn.com/intelligence/t...

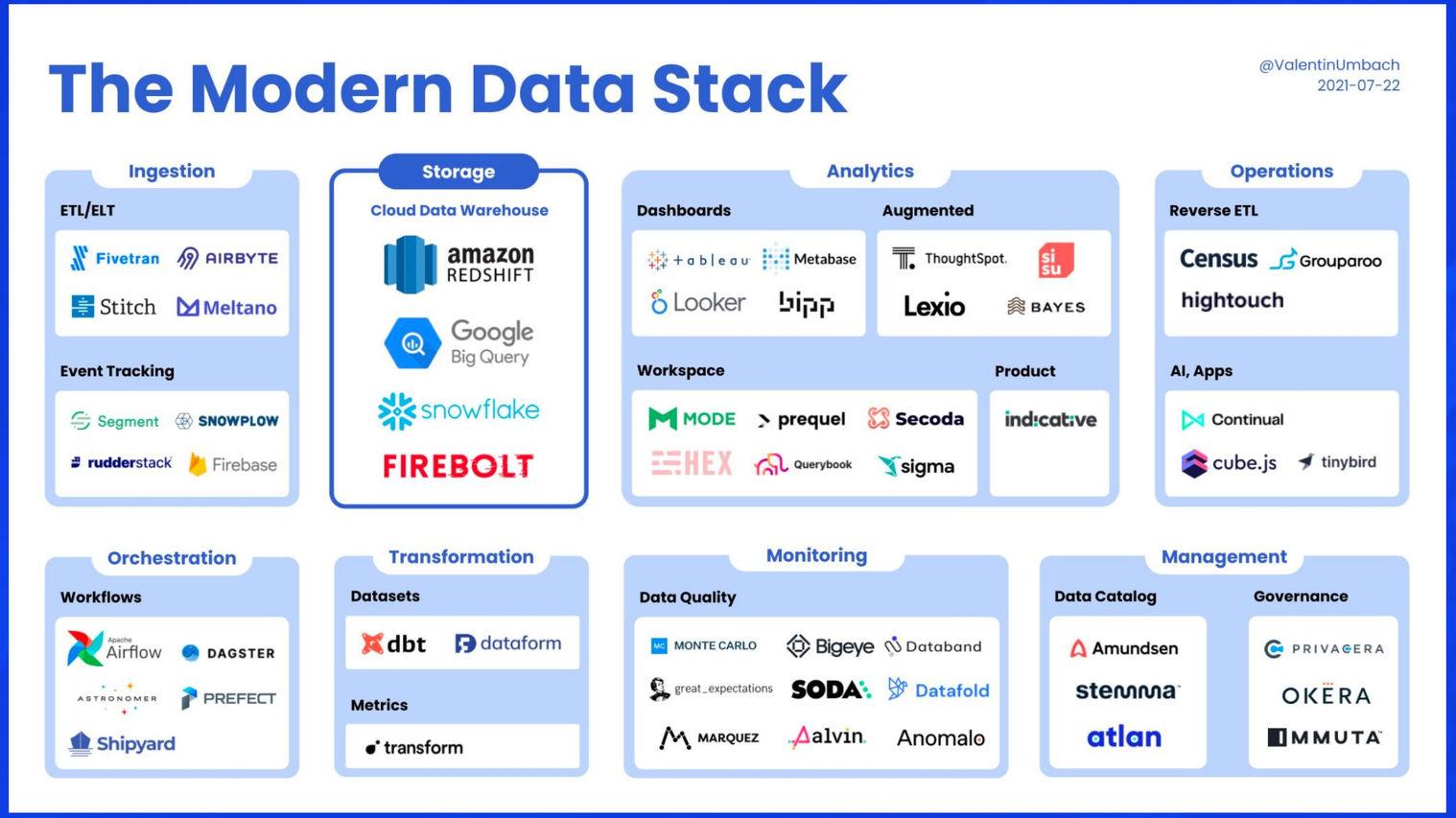
modern

ng, led
the

requirement of the **modern data stack**"

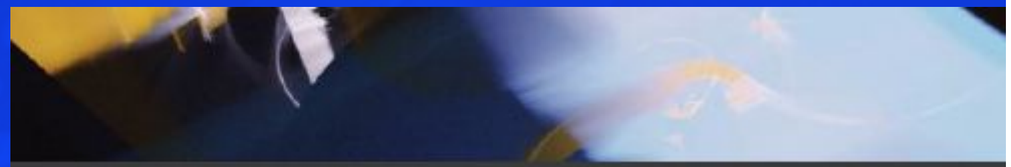
What is it?

- Composable pieces
- Centered around the Cloud Warehouse
- Easy to implement and plug together



Hype Cycle

Goodbye, World



data tech

Is the Modern Data Stack dead already ?

Huge hype, an ocean of critics. Are we done with the modern data stack ?

Jul 17, 2022

Apple Podcasts Preview

Ep.99 'Is the Modern Data Stack Dead' with Special Guest Chris Tabb

Drill to Detail

Technology

Listen on Apple Podcasts ↗

Episode Website ↗

More Episodes

48 min PLAY ▶

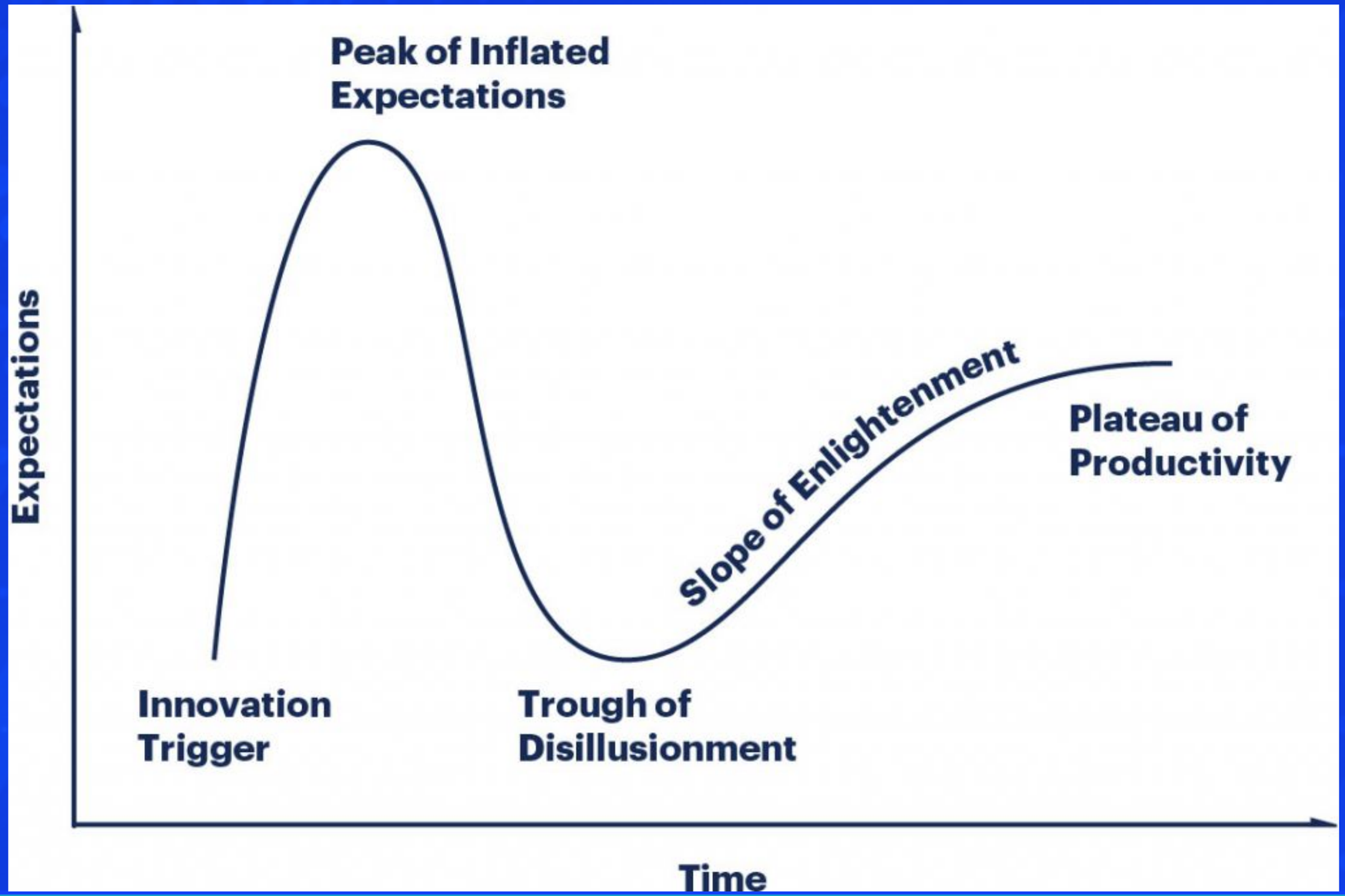
Rittman Analytics 2023

The Modern Data Stack is Dead... Long Live the Modern Data Stack - Part 1

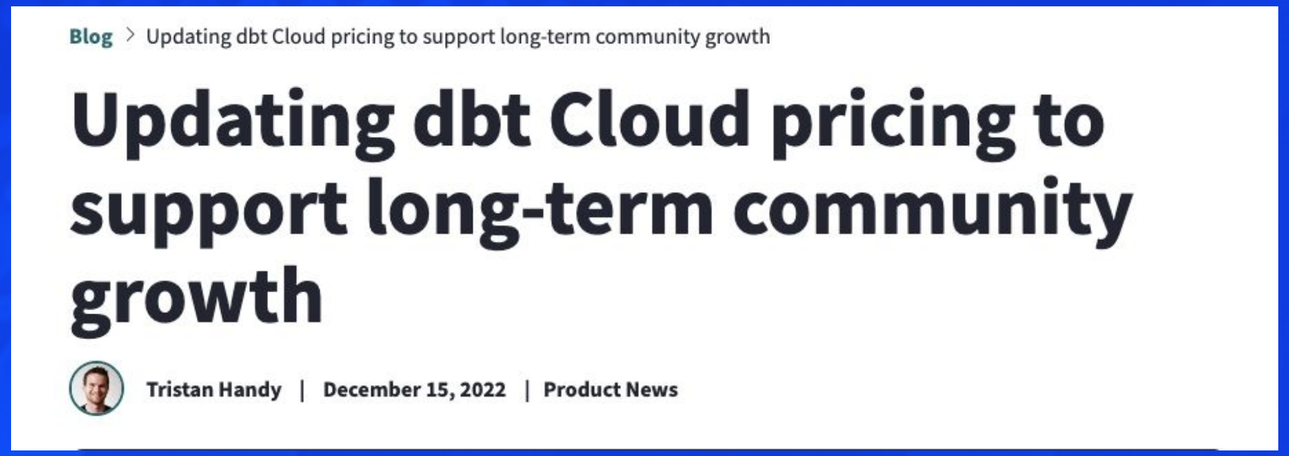
Where we've come from, lest we forget

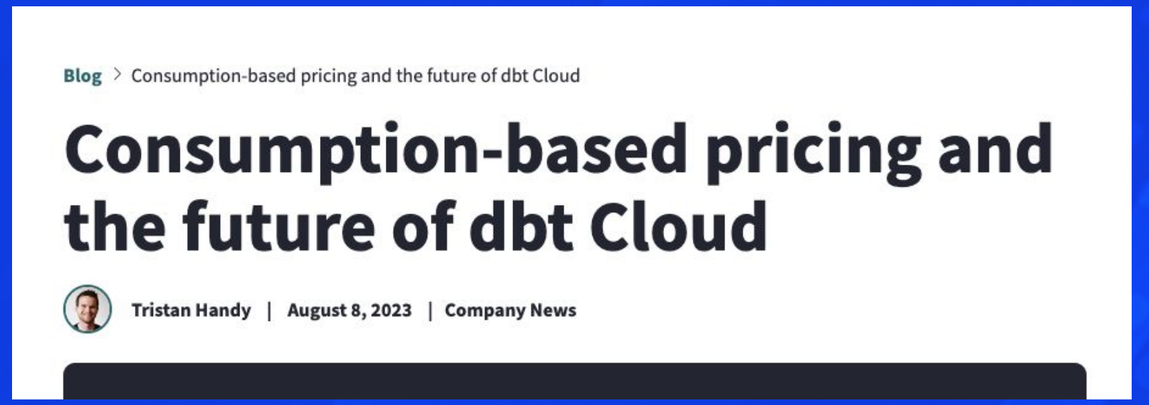
DAVID JAYATILLAKE

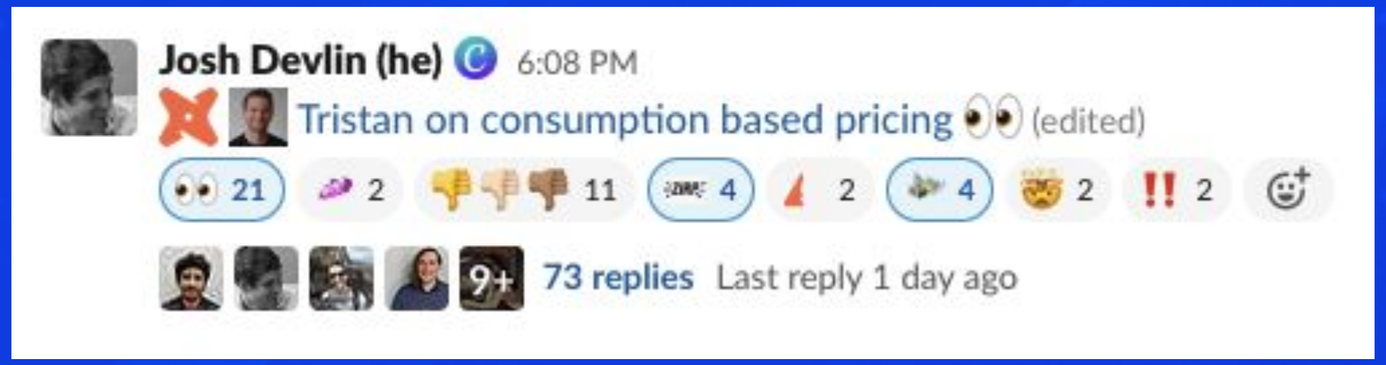
JUL 19, 2023


Peak of Inflated Expectations
Expectations
Slope of Enlightenment
Plateau of Productivity
Innovation Trigger
Trough of Disillusionment
Time

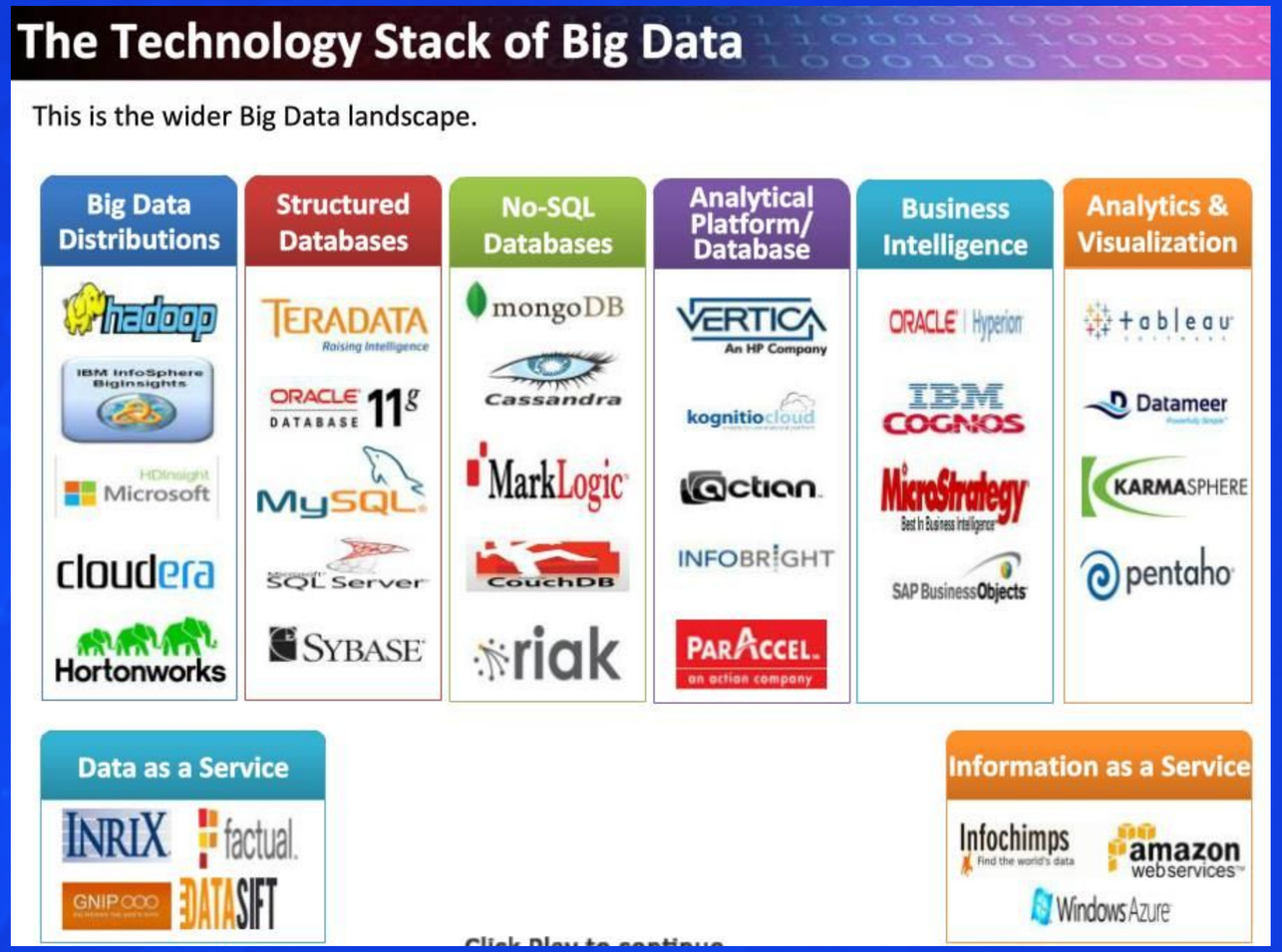
Nail in the Coffin

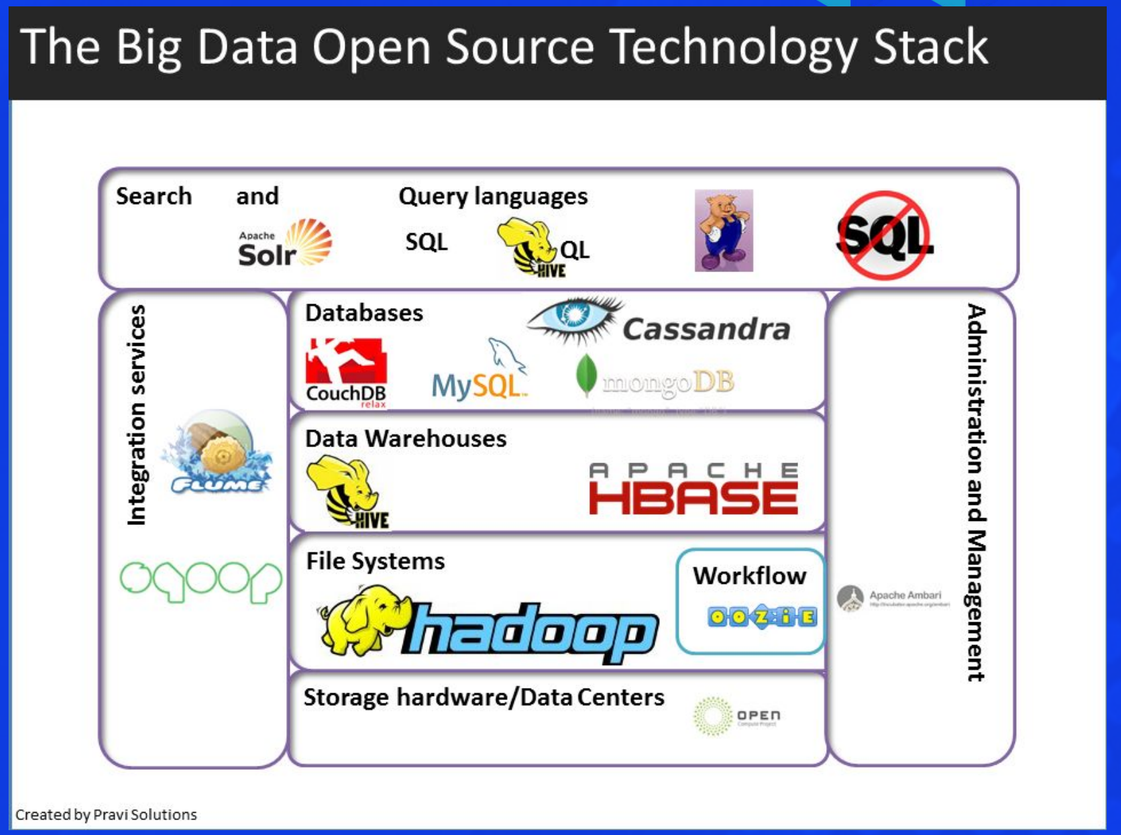




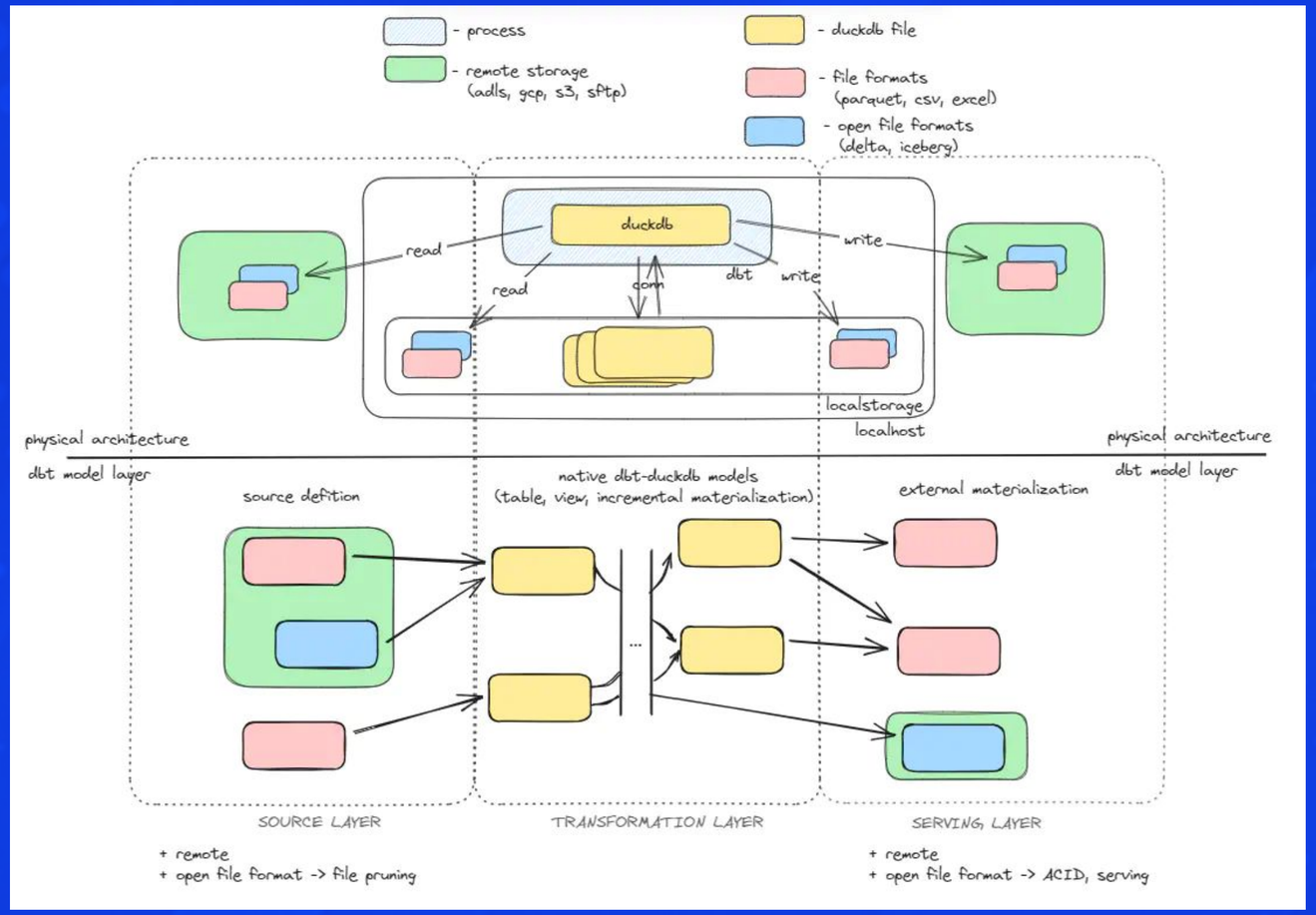


Nostalgia Time


The Technology Stack of Big Data
This is the wider Big Data landscape.
Big Data Distributions
Structured Databases
No-SQL Databases
Analytical Platform/ Database
Business Intelligence
Analytics & Visualization
Data as a Service
Information as a Service


The Big Data Open Source Technology Stack
Search and
Query languages
SQL
Integration services
Databases
Data Warehouses
File Systems
Workflow
Storage hardware/Data Centers
Administration and Management
Created by Pravi Solutions

The Open Source Data Stack Architecture



https://georgheiler.com/2023/12/11/dagster-dbt-duckdb-as-new-local-mds/

Data Pipelines

- Ingestion
- Storage
- Transformation
- Orchestration
- Activation
- Visualization

Ingestion

- Plain Ol’ Python Scripts
- Meltano / Alto
- Airbyte
- dlt
- Steampipe
- Sling
- pg2parquet
- And many many others

Storage

- Postgres
- Clickhouse
- DuckDB
- Parquet
- Arrow
- Deltalake

Transformation

- dbt
- dbt-duckdb
- polars
- Arrow DataFusion

Katie Bauer @imightbemary · Aug 21

Software engineers hate no code until someone asks them to write a SQL query

11 9 104 9,446

pedram 🐤

@pdrmnvd

software engineers will literally invent another dataframe framework before writing one line of sql

9:43 PM · Aug 21, 2023 · **1,348** Views

Orchestration

- Dagster
- Prefect
- Airflow

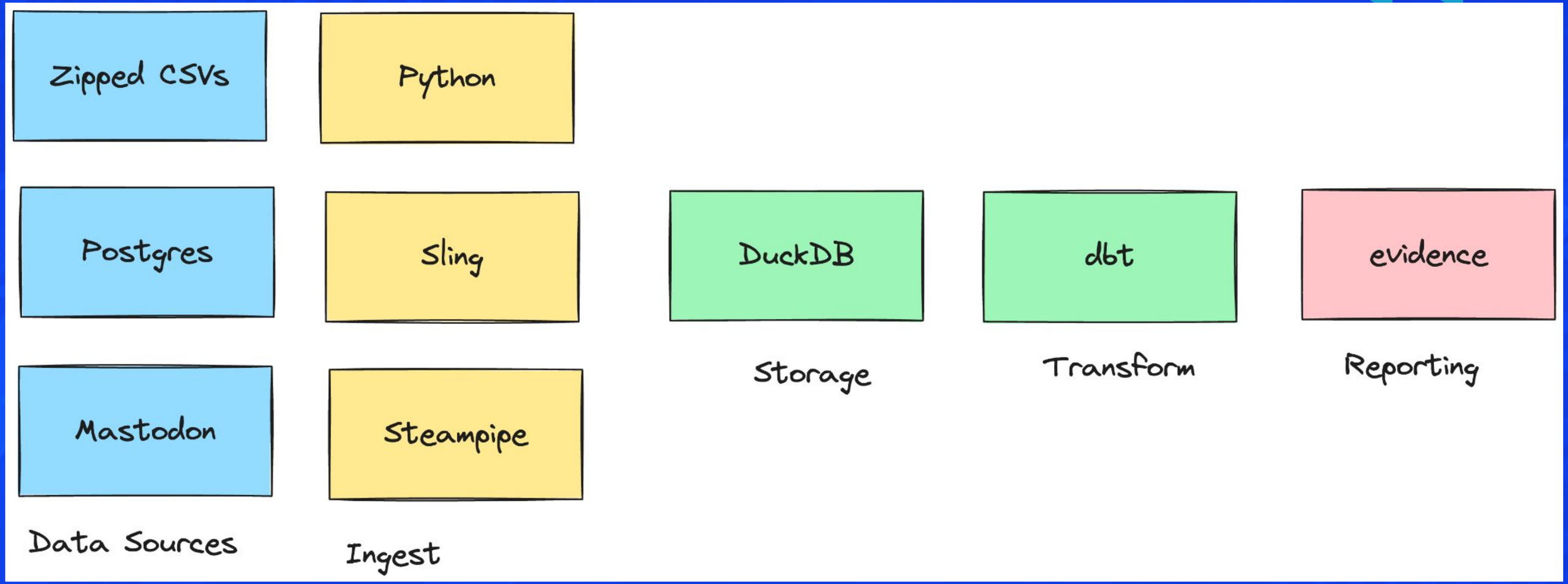
Conclusion

Regardless of whether you're an Airflow stan or one of the 500 data influencers who are being held at gunpoint to promote Mage (sorry Tommy, I still love you), I think we can all agree that if you can avoid Airflow, you'll be better off doing so.

Visualization

- Superset
- Evidence.dev
- Lightdash
- Metabase

Let’s Build a Stack


Zipped CSVs
Python
Postgres
Sling
DuckDB
dbt
evidence
Mastodon
Steampipe
Storage
Transform
Reporting
Data Sources
Ingest

Demo Time

Github Repo:
https://github.com/dagster-io/mdsfest-opensource-mds

Source Data:
https://feederwatch.org/explore/raw-dataset-requests/

Q&A

Follow Me

- ex-Twitter: @pdrmnvd
- LinkedIn: Pedram Navid
- Dagster Slack: dagster.io/slack