

# Data Products - Hype vs Reality

Gwen Shapira

# Data Products

- Where it started
- Where we are now
- What you really need
- What it takes

**Where it all started**

# Strata Conference - 2011

## The Evolution of Data Products

The data that drives products is shifting from overt to covert

Mike Loukides



O'REILLY®

O'REILLY®  
**Strata**  
Making Data Work

**Data products**  
→ **infrastructure innovation**  
and **adoption**

10:01 / 39:42

"1M. 10M. 100M. Data!" -- LinkedIn Data Scientist Monica Rogati @ O'Reilly Strata 2011

2012

# Data Jujitsu

The Art of Turning Data Into Product

DJ Patil



O'REILLY®

O'REILLY®

**Strata**  
Making Data Work

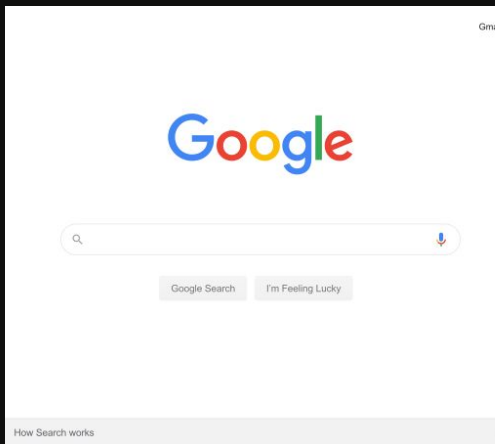
# LinkedIn's data products

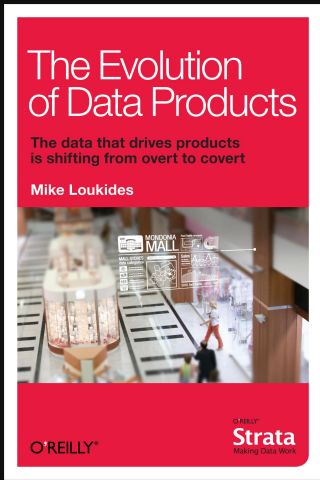
The screenshot shows the LinkedIn profile of Gwen (Chen) Shapira. The profile header includes her name, tagline "Building Something Amazing", and a "Start a post" button. Below the header, there are statistics for "Who's viewed your profile" (2,185) and "Impressions of your posts" (1,314). The main content area features a post from ZIR AI Inc. with a photo of a city skyline and the text "View from ZIR AI's first office at WeWork in San Jose (photo taken last Thursday). We are hiring anywhere in the US or Canada right now, you can see available jobs and apply here: <https://lnkd.in/gKawWym>". The post has 4 comments and 2 shares. The right sidebar shows "LinkedIn News" with articles about Zelensky, company loyalty, student debt, plastic waste, and quantum internet. The bottom of the page shows the LinkedIn logo and copyright information.

The screenshot displays the "People you may know" section of LinkedIn, showing two categories: "People you may know from Tel Aviv University" and "People you may know in San Francisco Bay Area". Each category lists several individuals with their profile pictures, names, titles, and mutual connections. The "Tel Aviv University" category includes Dan Shappir, Yaniv Levin, Ran Ziv, Eden Gal, Shahar Tarshish, Itay Perry, Tal Yalon, and Sagit Bitran. The "San Francisco Bay Area" category includes Harendra Chawla, Sophie Blee-Gold..., Alyssa Huang, and Ahmad Nawaz. Each person's card includes a "Connect" button. The "See all" link is visible at the top right of each category.



Many products  
use data for a great UX





” [data] products aren’t about the data; they’re about enabling their users to do whatever *they* want, which most often has little to do with data ”



**Where we are today**

# Many faces of data products

	Dashboard	Decision Support
Internal	Leaders and data analysts use DWH + Metabase.	Sales reps use Salesforce Einstein.
External	Your product shows charts or enforces rules.	Your product gives good advice to users

# How are we doing in each?

	Dashboard	Decision Support
Internal	We had this for decades. Still non-trivial.	Rare in B2B
External	Exists and growing fast. Most innovation is here.	Rare in B2B

How are we doing in each?

	Dashboard	Decision Support
Internal	<b>This is what "data products" mostly means today</b>	Rare in B2B
External		<b>This is what "data products" meant at LinkedIn, Netflix, etc</b>

**What you really need  
to build**

Internal Decisions

External Decisions

External Dashboards

Internal Dashboards

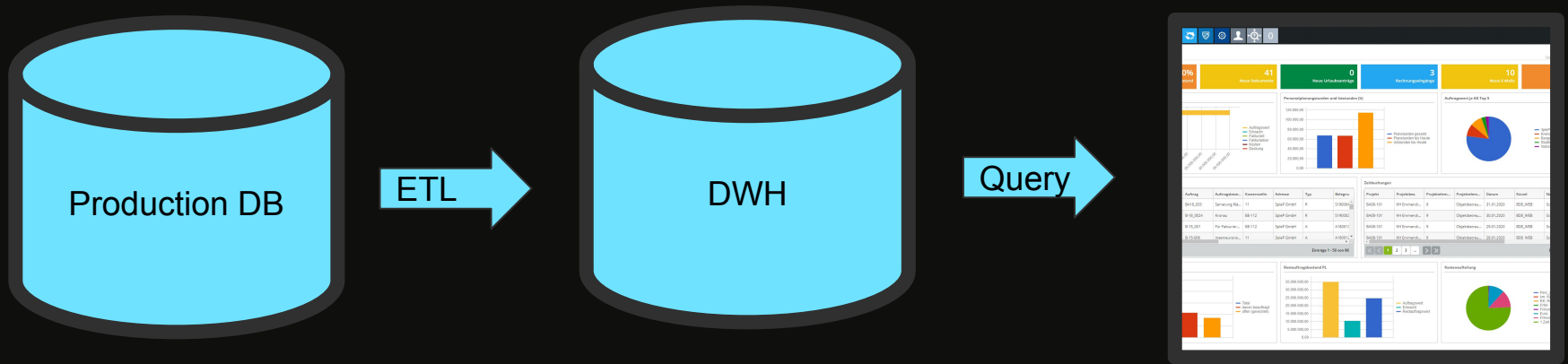
Vendors point to top of pyramid  
But we are often still at the bottom

**What it takes to get there?**

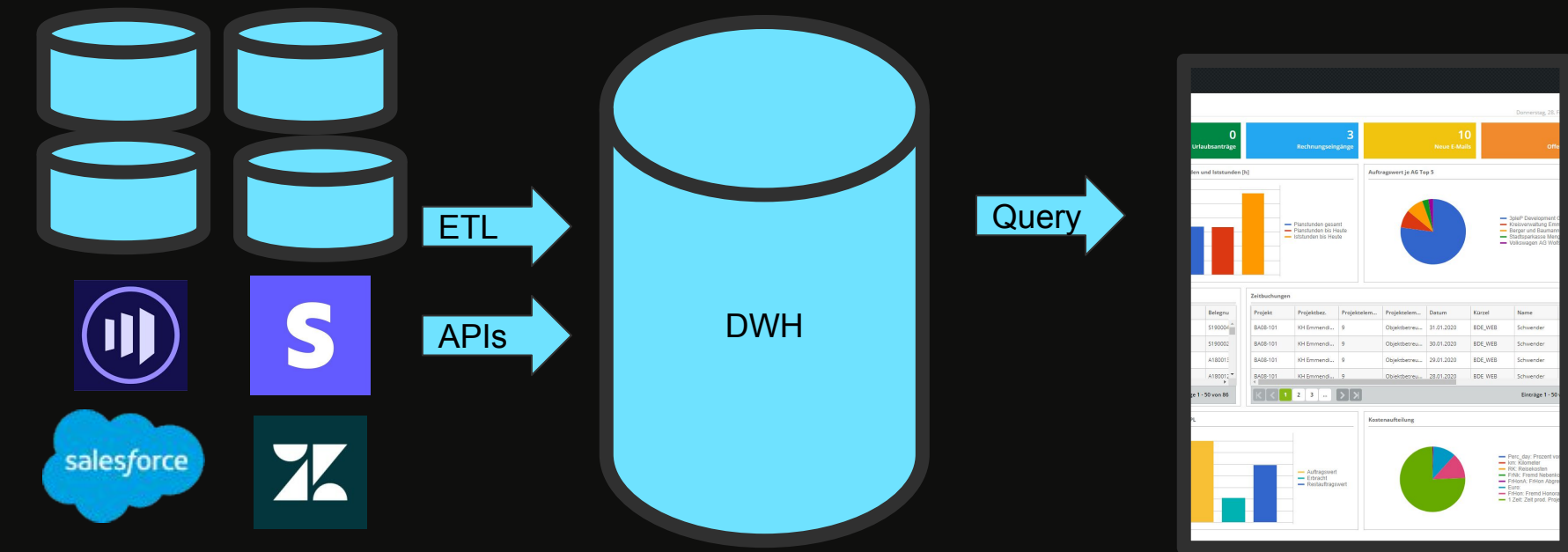


# Internal Dashboards

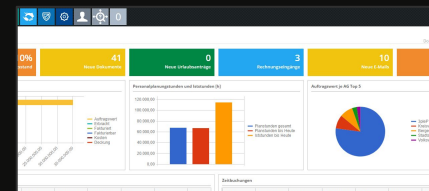
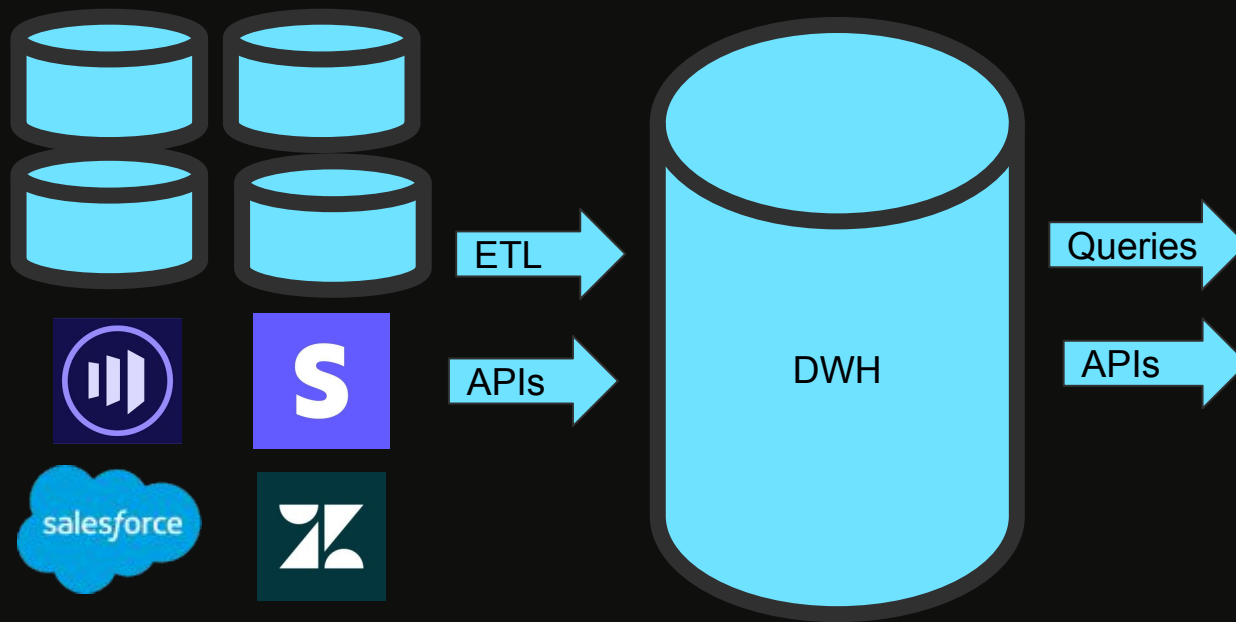
# Internal Dashboards are Deceptively Simple...



# You forgot the other sources



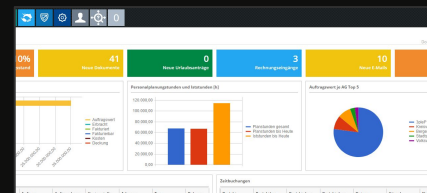
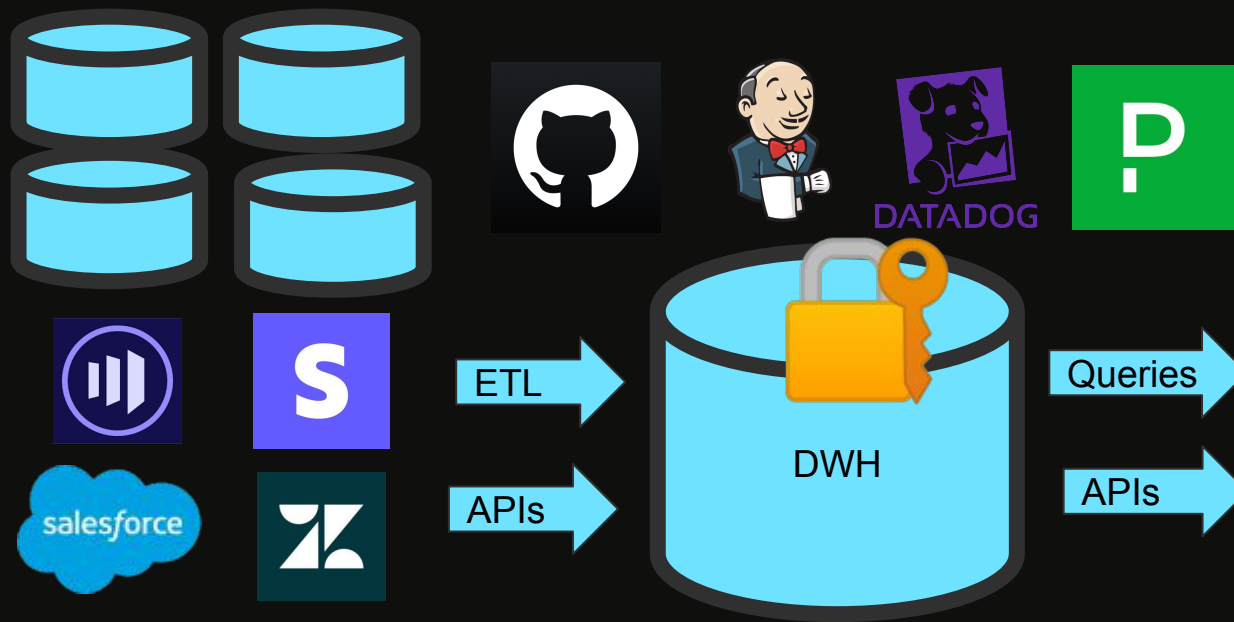
And the other users...



```
1 SELECT TO_CHAR("Activity"."created_date",  
2         "Activity"."activity_name" AS "Activ  
3         COUNT(DISTINCT "Activity"."activity_  
4 FROM "public"."activity" AS "Activity"  
5 WHERE ("Activity"."created_date" >= DATE '2  
6       AND "Activity"."created_date" < DATE  
7 DAY')  
8 GROUP BY TO_CHAR("Activity"."created_date",  
9         "Activity"."activity_name"  
10 ORDER BY "Created Date" ASC,  
        "Activity Name" ASC
```



Oh, and it is critical, right?



```
1 SELECT TO_CHAR("Activity"."created_date",  
2         "Activity"."activity_name" AS "Activ  
3 COUNT(DISTINCT "Activity"."activity_  
4 FROM "public"."activity" AS "Activity"  
5 WHERE ("Activity"."created_date" >= DATE '2  
6       AND "Activity"."created_date" < DATE  
7 DAY')  
8 GROUP BY TO_CHAR("Activity"."created_date",  
9         "Activity"."activity_name"  
10 ORDER BY "Created Date" ASC,  
        "Activity Name" ASC
```



# Data as a Product



Every product has a  
data product

Data products have  
many users / personas

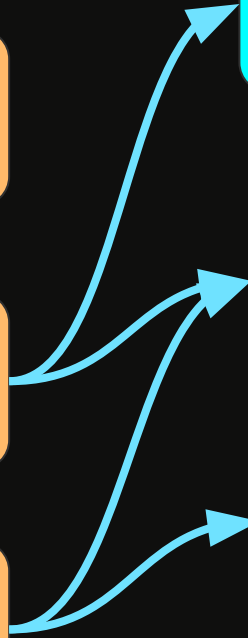
Product teams own  
“source data products”

Some users create new  
data products

Cleaning, standardization,  
discovery, security,  
governance, documented

Architecture, code,  
infrastructure

Aggregation, correctness,  
discovery, security,  
governance, documented

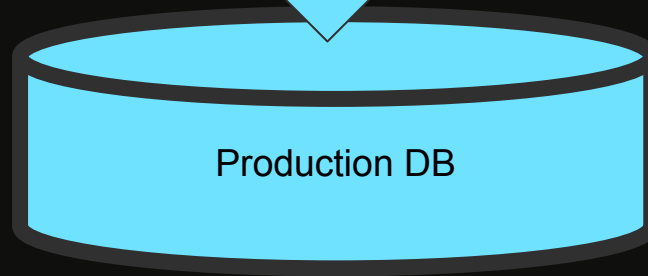
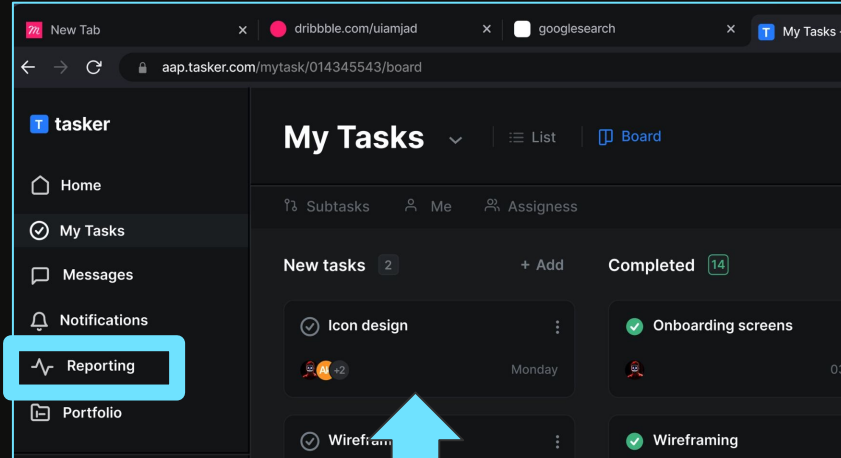


Every data product  
internal or external  
needs owners

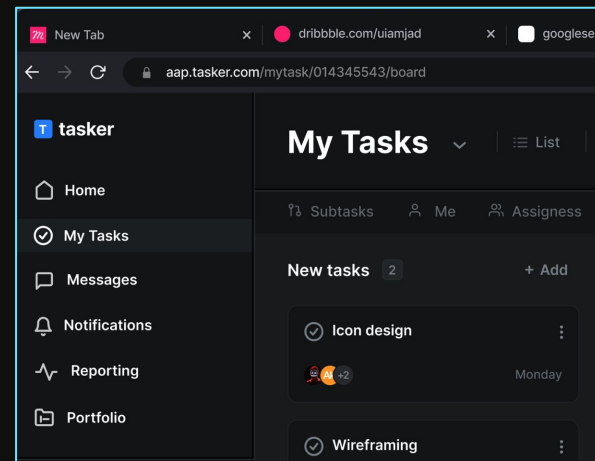
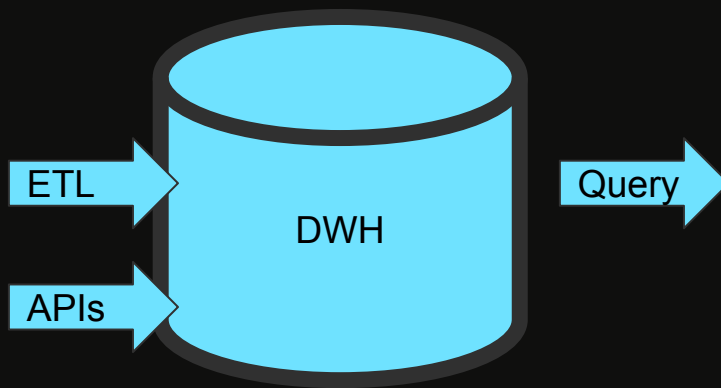
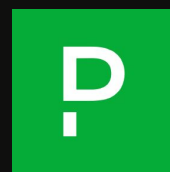
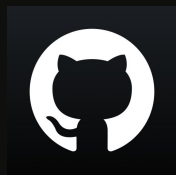
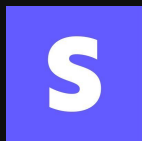
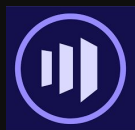


# External Dashboards

# External Dashboards are Deceptively Simple...

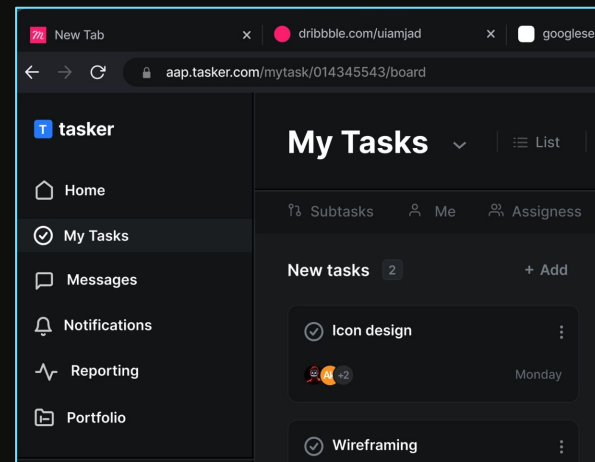
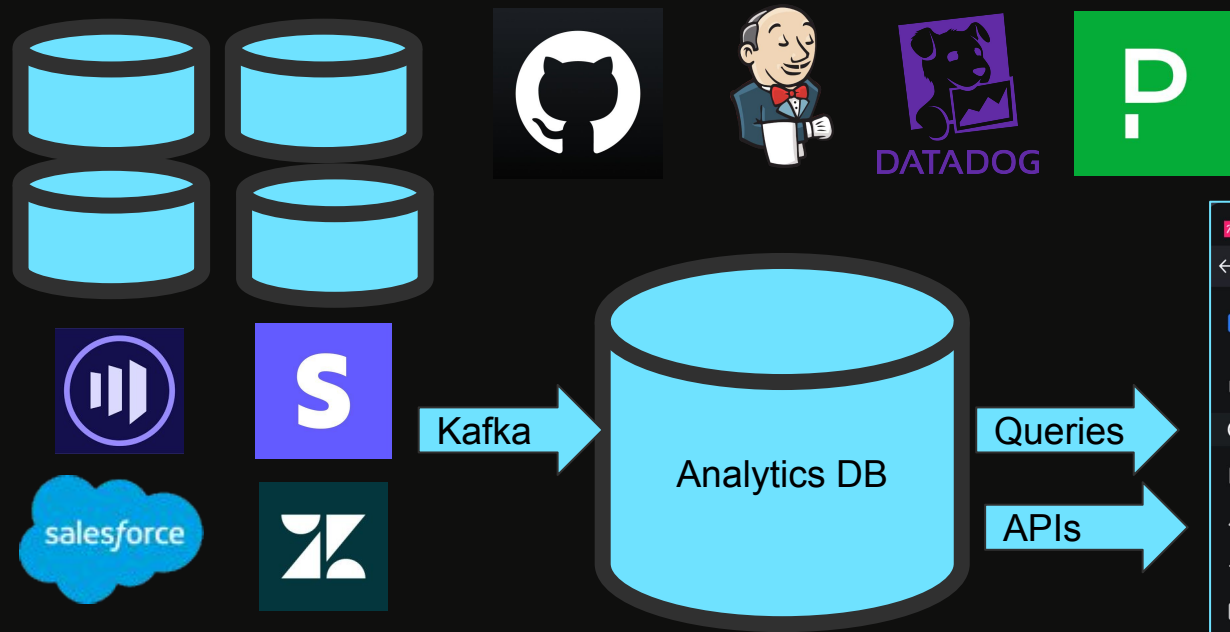


# Is this it?

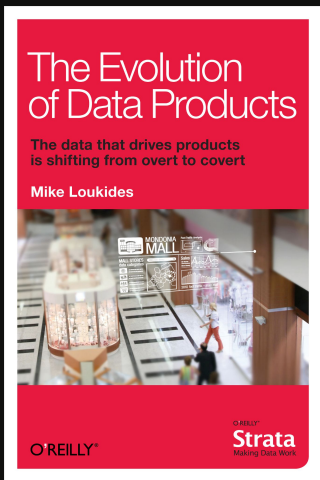


Your customers expect consistent low latency  
20-200ms for anything IOT  
200-2000ms for web apps

This is pretty much it.



# External Decisions



[data] products aren't about the data; they're about enabling their users to do whatever *they* want, which most often has little to do with data

Smart products lead to successful users lead to successful products.

“Features you  
may like“

In-product  
usage guide

Looks like this isn't  
working. Try...

Auto-scale

Personalized  
experience

“You really  
shouldn't“

Usage  
based billing

“Successful  
users often  
do Z“

“You started using  
X...”

Self-heal

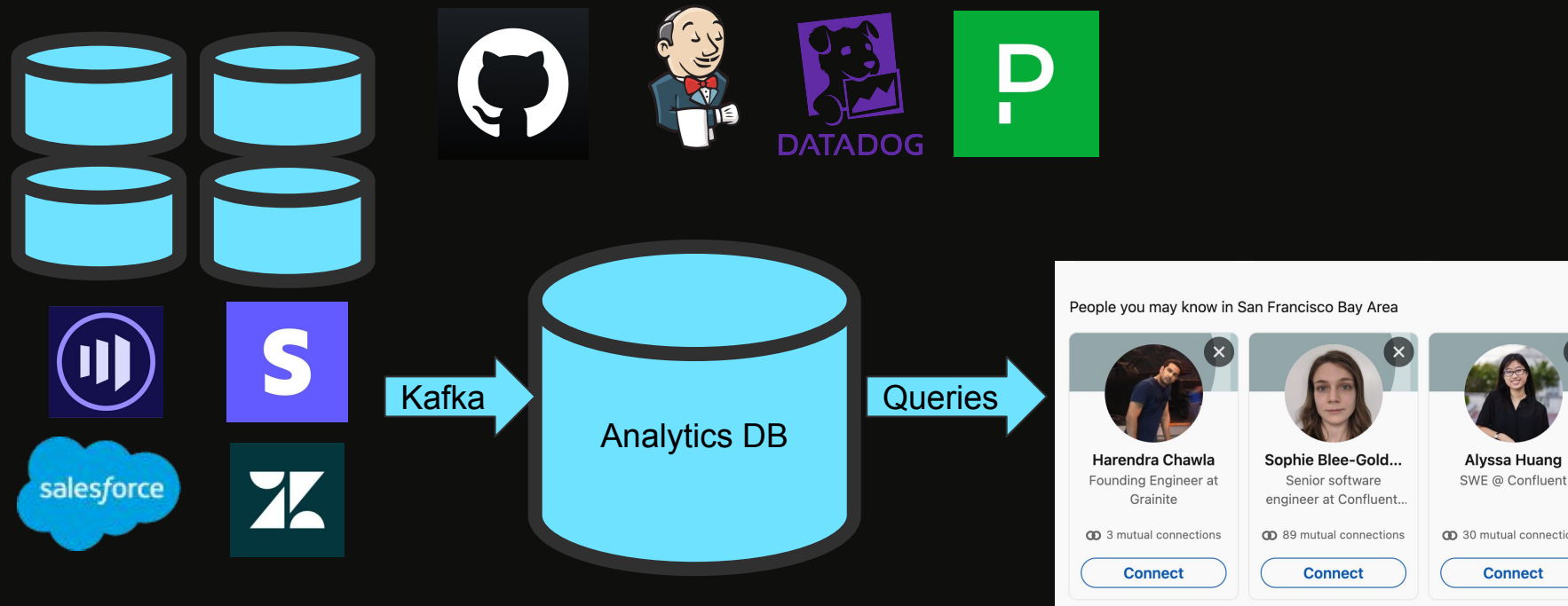
Smart  
search  
(cmd-K)



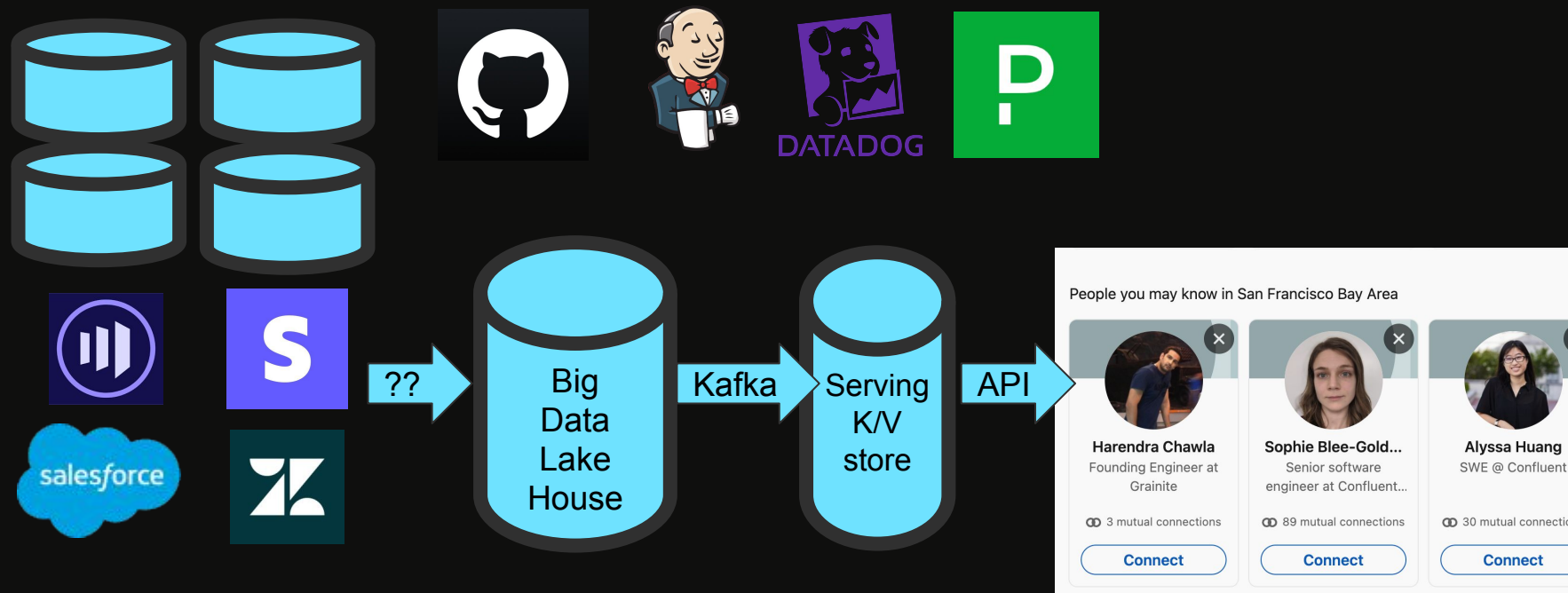
Smart products are about delivering  
delightful user experience.

Design experiences where users  
and algorithms share control

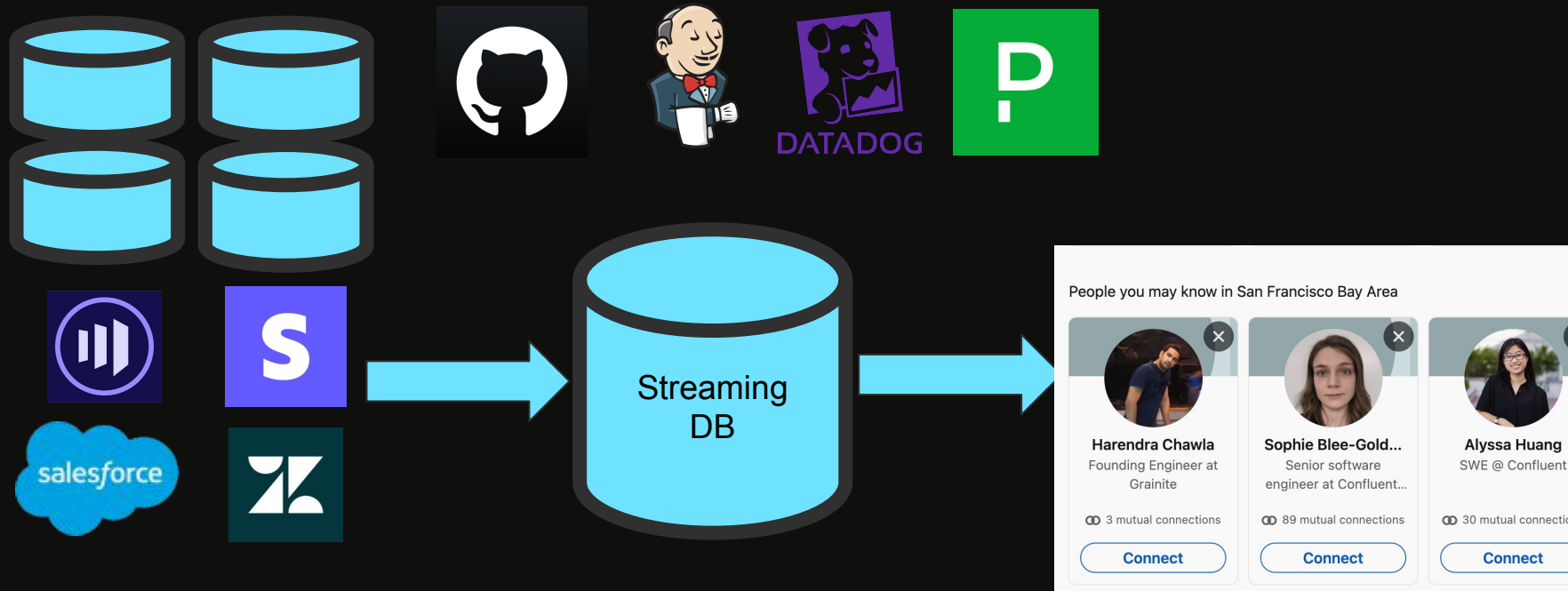
This may be a misfit



This architecture is challenging but common:



# Maybe a better future?



Sophisticated  
analysis

Low latency



Data stores are placed on this continuum

Great user experience requires both

We end up running multiple DBs and ETL pipes

”A Database can’t do everything  
but a data platform could“

--Sriram Subraminian,  
CEO of Stealth Startup

**Wrap it up**



## Key points:

Different problems  
require  
different databases

Good data products  
make data usable

Great data products  
make data disappear

# Thank you for attending!

If you are building infrastructure SaaS (DB, ETL, ML, etc as a service), or if you are interested in control-plane architectures and their unique data challenges.

Please contact me.

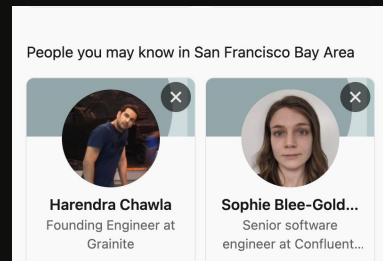
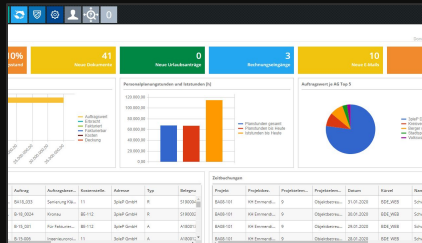
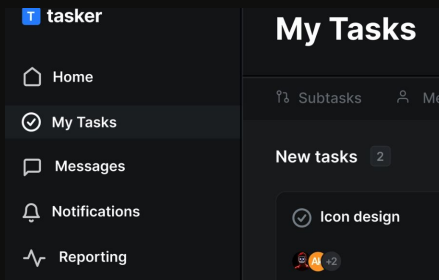


@gwenshap



gwen@thenile.dev

# Imagine: managed data platform



REST

Search

Queries

Handlers

Entities

Metrics / KPIs

Events

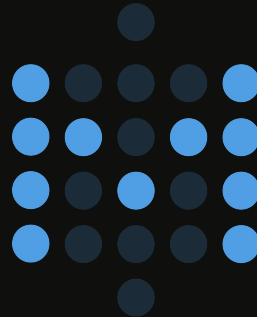
# Tons of OSS tools



druid



Apache  
Airflow



What went wrong?

Business

Conceptual Gaps

Technology Gaps

# 10 years later...

## STEPS TO GET STARTED WITH HUBSPOT

Progress: 100%



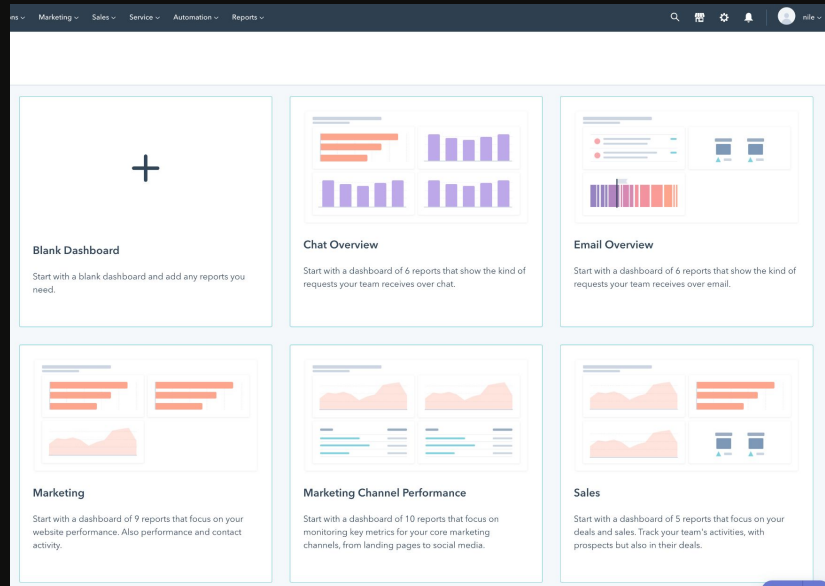
Sales Hub Free setup complete

**Recommended for scaling your business:**

Learn to attract and engage leads with your marketing tools.

[Go to Marketing Hub Free guide](#)

[Go to my dashboard](#)



# 10 years later...

The Data Product Platform is modular by design, and comprehensive in functionality.

It integrates with, and leverages, your existing data technology stack.



## Data Integration

Ingest and unify data from multiple sources, then pipeline it to target systems, always ensuring data integrity.



## Data Virtualization

Provide a logical abstraction layer to underlying systems, to make it easy to access trusted data.



## Data Preparation

Make data lakes and data warehouses instantly and always ready for analytics.



## Data Orchestration

Control data movement and data transformation, from source to target systems, code-free.



## Data Catalog

Discover and visualize metadata structure and lineage, from source to consuming microservices.



## Data Governance

Control data synchronization, access, integrity, and security – using configurable rules and processes.



## Data Masking

Protect data at rest, in use, and in transit – across production, testing, and analytics environments.



## Microservice Automation

Generate, debug, and deploy web services, in minutes, with an easy-to-use, no-code/low-code framework.



