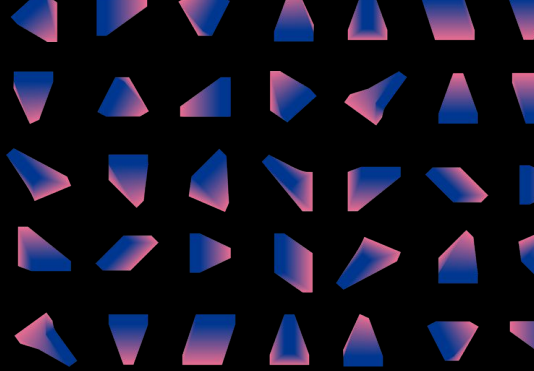


MODE

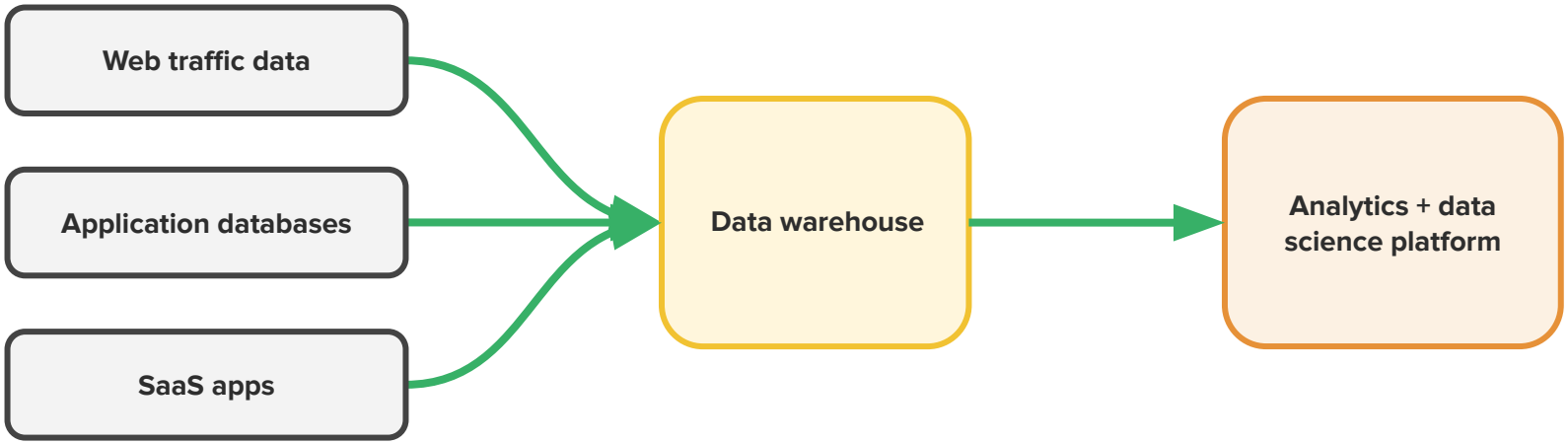
Building an end-to-end data stack in 30 minutes

Benn Stancil
Chief analytics officer, Mode

MODE

A modern data stack

A modern data stack



Part 1

The data warehouse

What to consider

What to consider

- Easy to set up and maintain

Vicki Boykis
@vboykis

Follow

Is there "Hadoop for Complete Morons?" Hadoop for Idiots is just not cutting it for me.

What to consider

- Easy to set up and maintain
- Uses plain SQL

Compare

SQL
Programming language

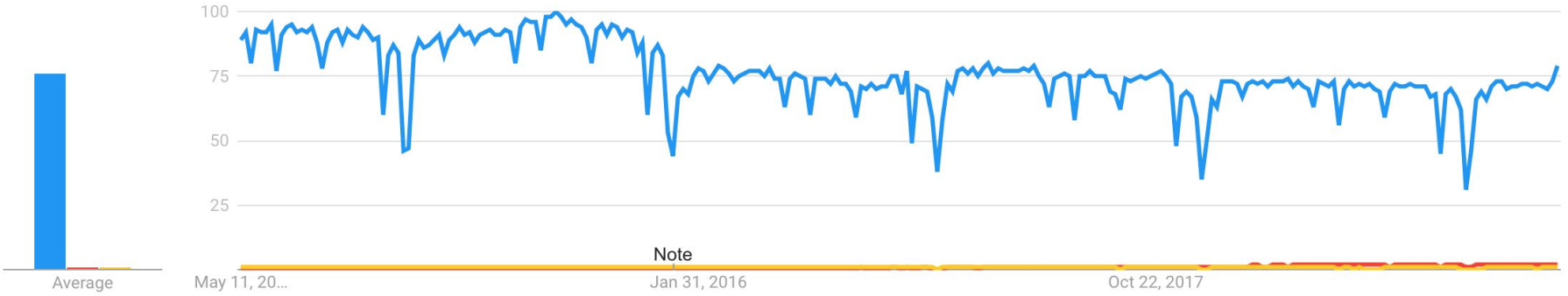
GraphQL
Topic

MapReduce
Topic

+ Add comparison

United States | Past 5 years | All categories | Web Search

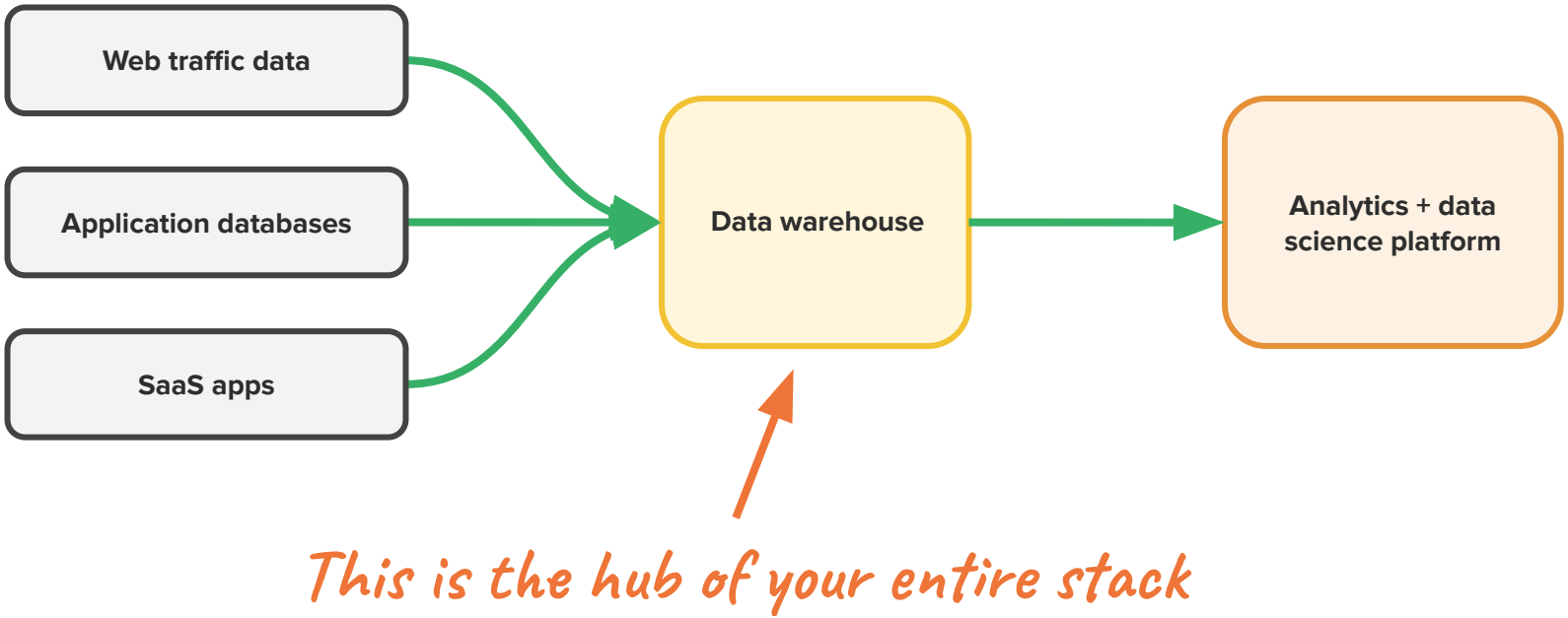
Interest over time



What to consider

- Easy to set up and maintain
- Uses plain SQL
- A strong ecosystem

A modern data stack



amazon
REDSHIFT

Launch cluster Learn more

With a few clicks, you can create your first Amazon Redshift cluster in minutes.

Quick launch cluster

PostgreSQL

AWS Partner Network - Amazon Redshift Partners

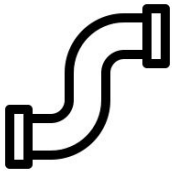
Enhance your Amazon Redshift data warehouse by working with industry-leading tools and experts for loading, transforming and visualizing data. The partners below have certified their solutions to work with Amazon Redshift. To search all AWS technology partners, visit the Partner Solutions Finder. Or use the AWS Marketplace to find and immediately start using third party software.

- Load and transform your data with Data Integration Partners »
- Analyze data and share insights across your organization with Business Intelligence Partners »
- Architect and implement your analytics platform with System Integration and Consulting Partners »
- Query, explore and model your data using tools and utilities from Query and Data Modeling Partners »

Find an Amazon Redshift Partner

snowflake

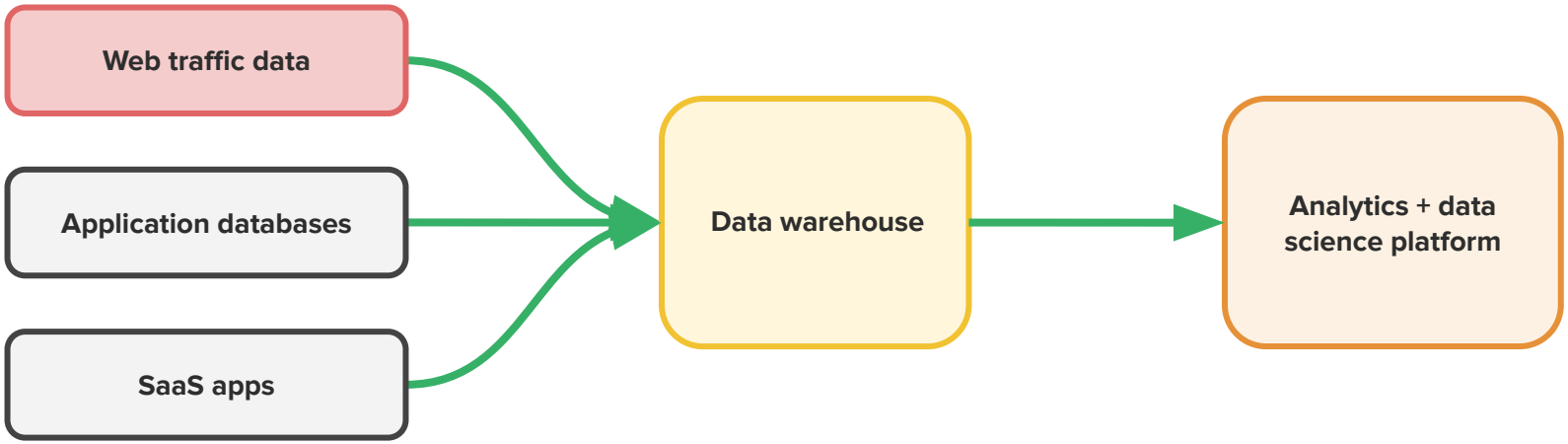
Google
BigQuery



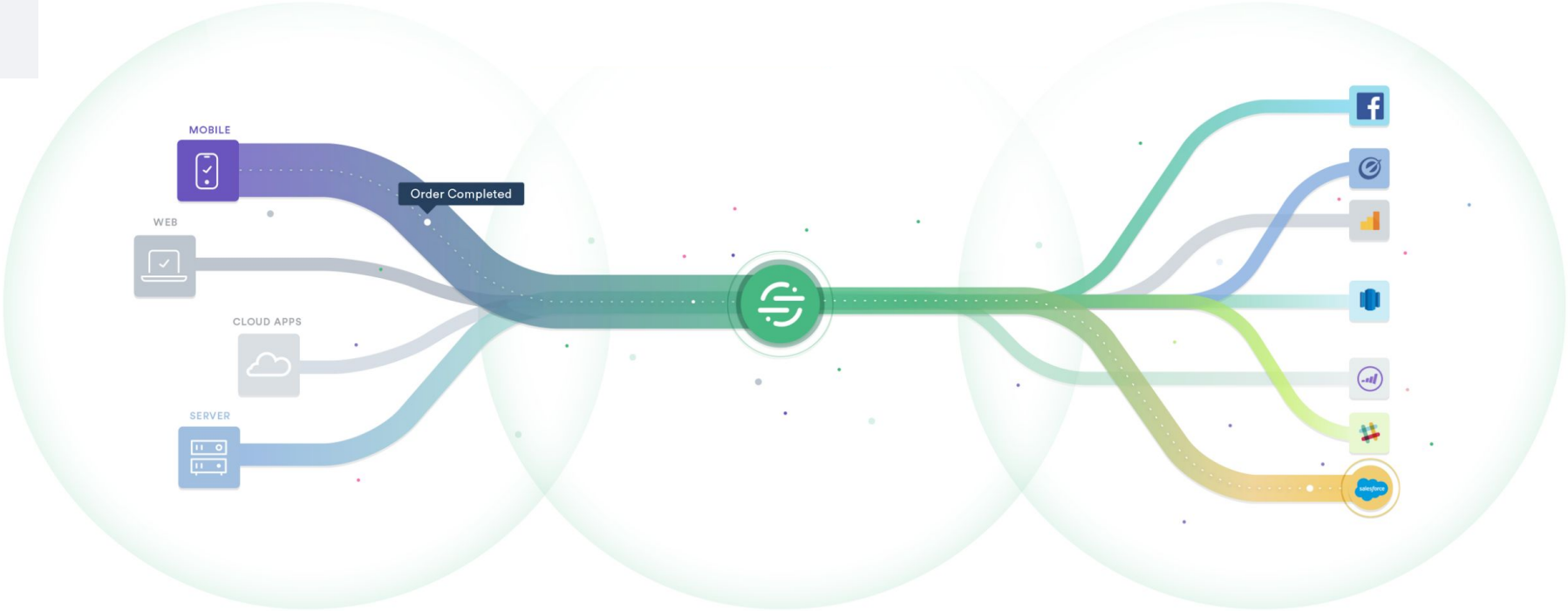
Part 2

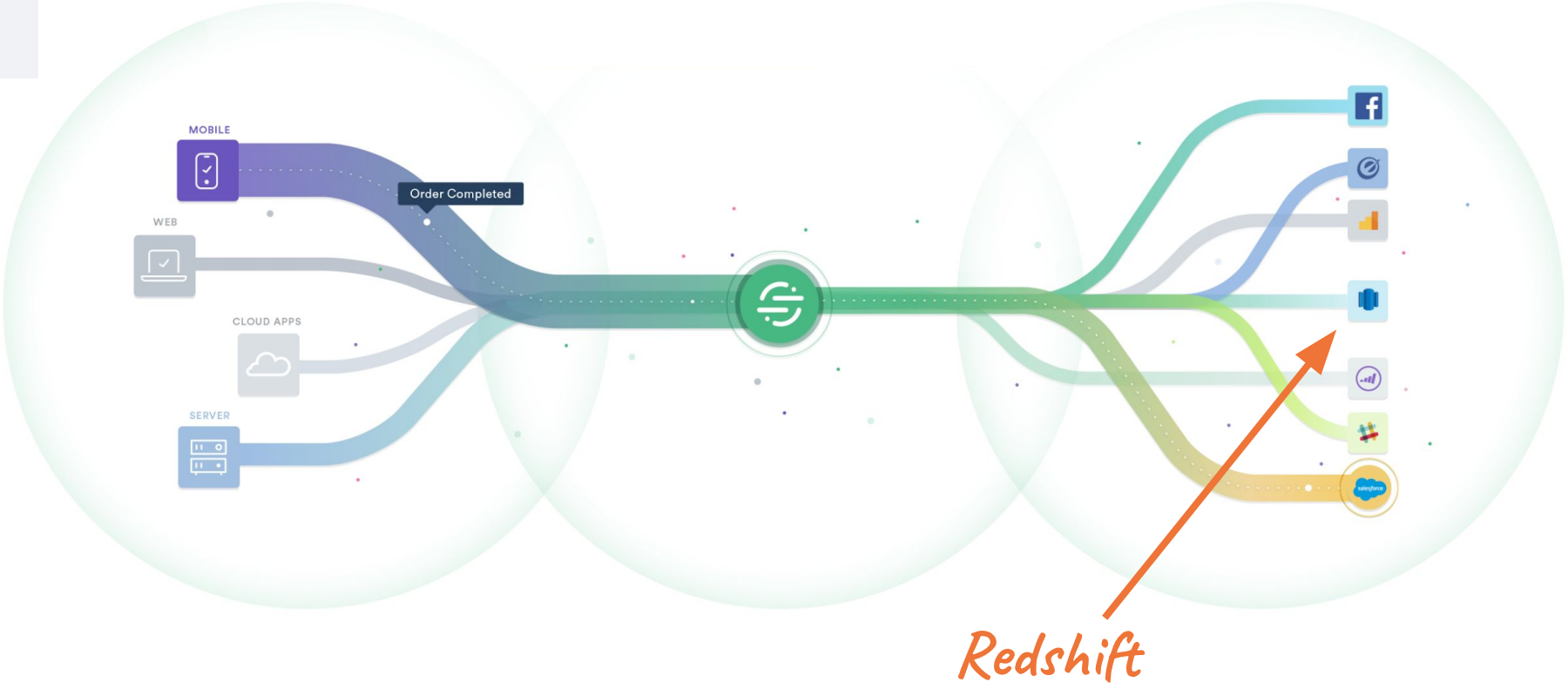
Collect data

Three major categories



Segment


MOBILE
Order Completed
WEB
CLOUD APPS
SERVER

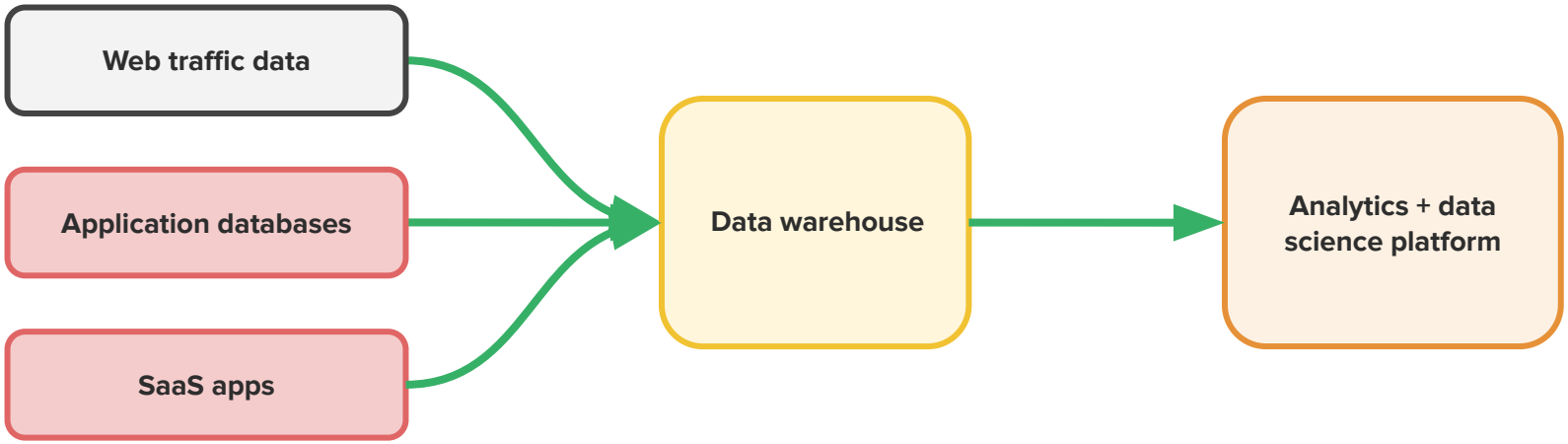

MOBILE
Order Completed
WEB
CLOUD APPS
SERVER
Redshift

```
analytics.track("photo_viewed", {
  height: "480",
  width: "640"
});
```

SNOWPLOW

Google Analytics

Three major categories



ELT > ETL

Stitch

Configure your Intercom Integration

Check out the documentation if you need some help setting up your Intercom integration. If you're not sure where to find this information you can invite a member of your team to help.

Integration Name

Intercom Integration Name

Intercom App Id

Intercom App Id

Sync Historical Data

This will determine the start date for your Intercom integration. Data equal to or newer than this date will be replicated from your Intercom integration to your Stitch data warehouse. Learn more about Intercom integration here.

Configure your PostgreSQL Integration

This integration supports PostgreSQL version 9.3 and above (9.4 for log-based data replication).

Check out the documentation if you need some help setting up your PostgreSQL integration. If you're not sure where to find this information you can invite a member of your team to help.

Integration Name

Integration Name

Host (Endpoint)

Your IP address or host name

Port

3306

Enter your IP address or host name.

User

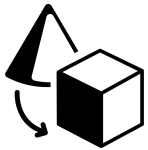
Your PostgreSQL username

Password

Your PostgreSQL password

etleap

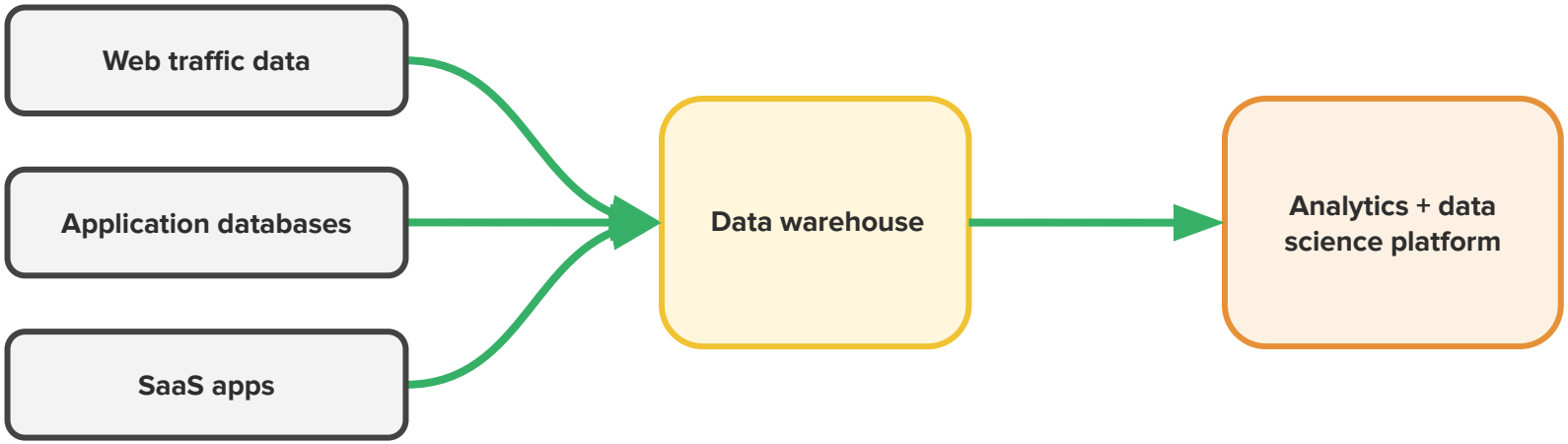
Fivetran



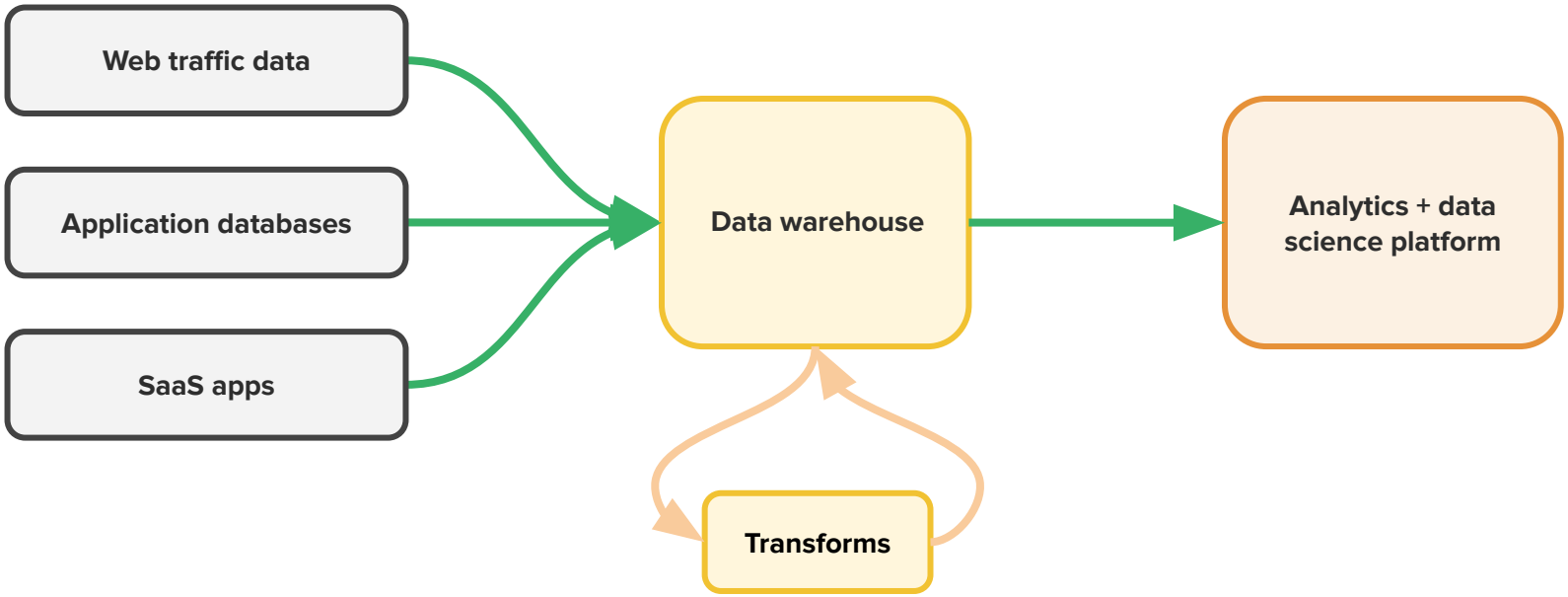
Step 3

Clean and model data

A modern data stack



A modern data stack



What to consider

- Uses plain SQL

Engineers Shouldn't Write ETL: A Guide to Building a High Functioning Data Science Department

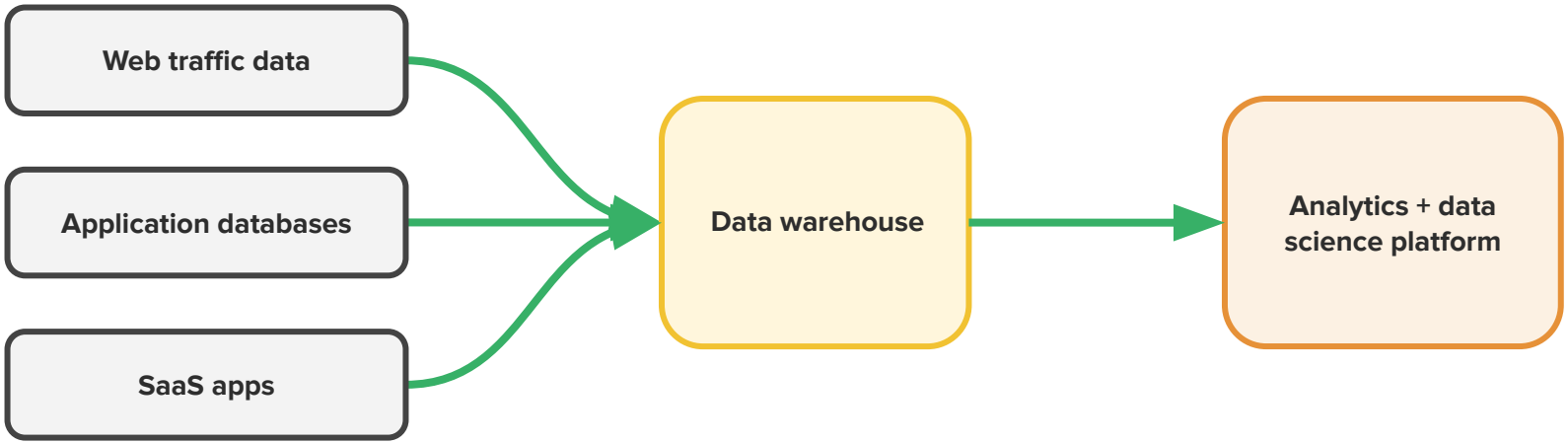
JEFF MAGNUSSON

March 16, 2016 - San Francisco, CA

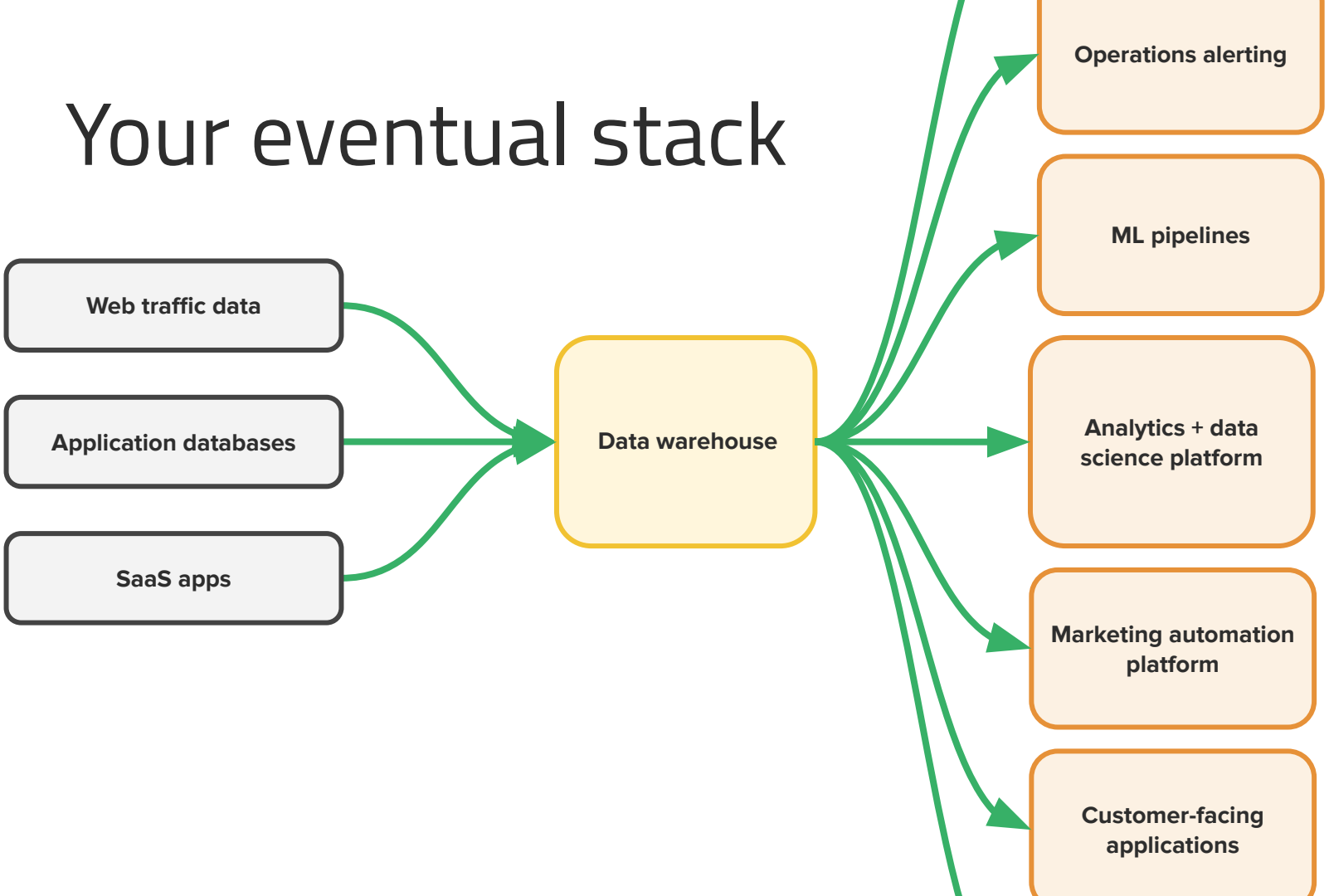
What to consider

- Uses plain SQL
- Agnostic to downstream applications

Your initial stack



Your eventual stack



dbt_

dataform

Step 4

Analyze and share results

What to consider

- Deliver results quickly

Speed is [an analyst's] highest virtue.

Cassie Kozyrkov
Harvard Business Review

Data Teams

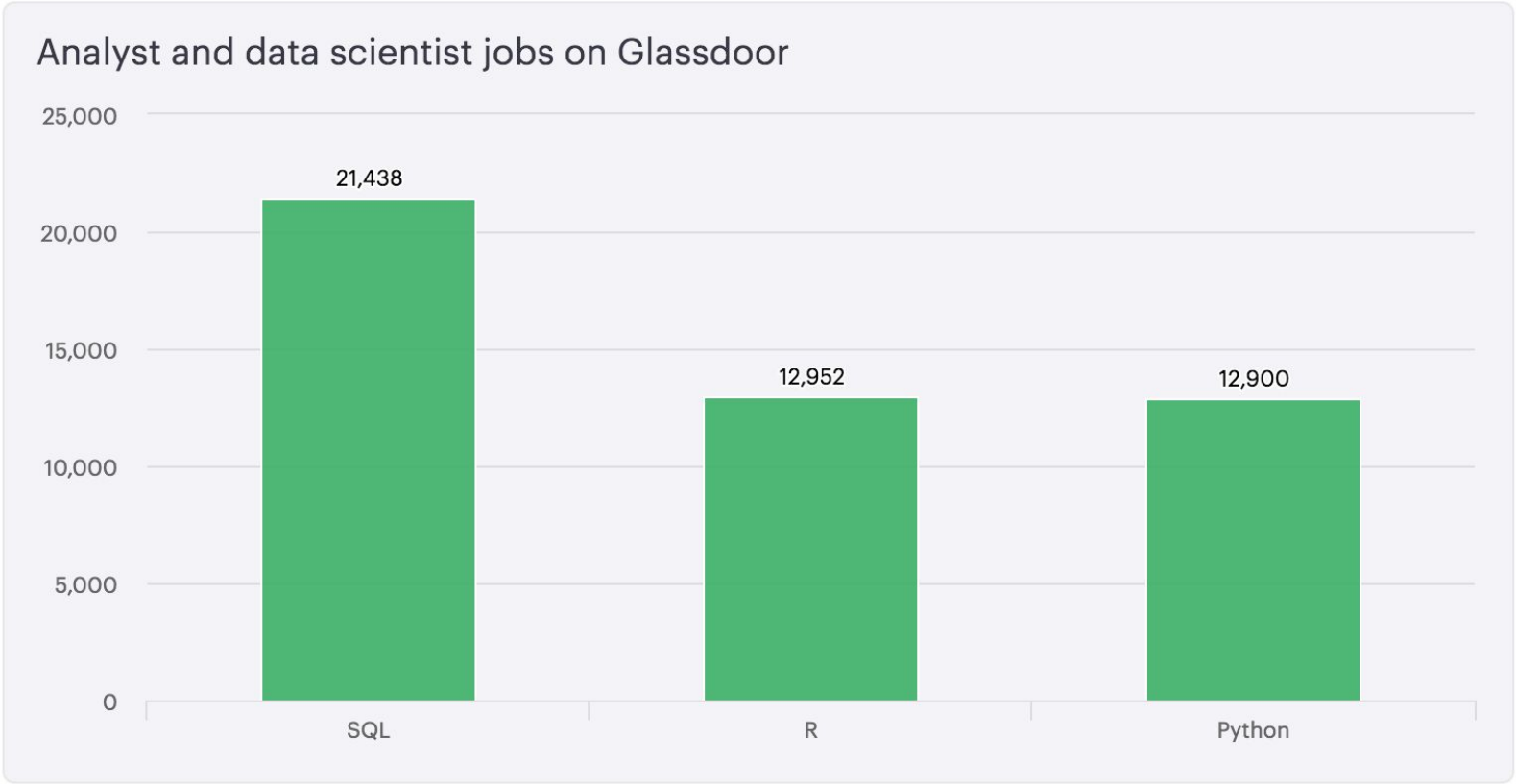
January 14, 2019 • 5 minute read

Overplanned Analytics Initiatives Are Doomed to Fail

BENN STANCIL
CHIEF ANALYST

What to consider

- Deliver results quickly
- Uses plain SQL


Analyst and data scientist jobs on Glassdoor
25,000
20,000
15,000
10,000
5,000
0
21,438
12,952
12,900
SQL
R
Python

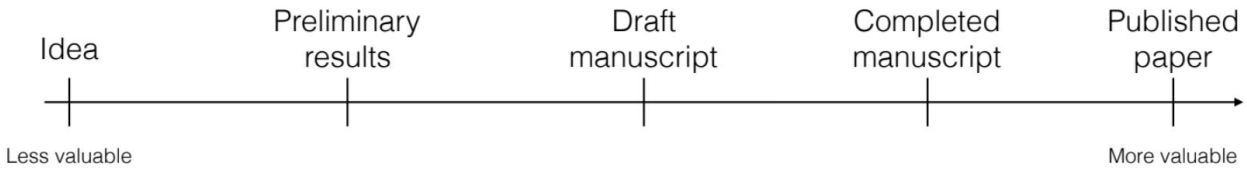
What to consider

- Deliver results quickly
- Uses plain SQL
- Easy to share

The Unreasonable Effectiveness of Public Work

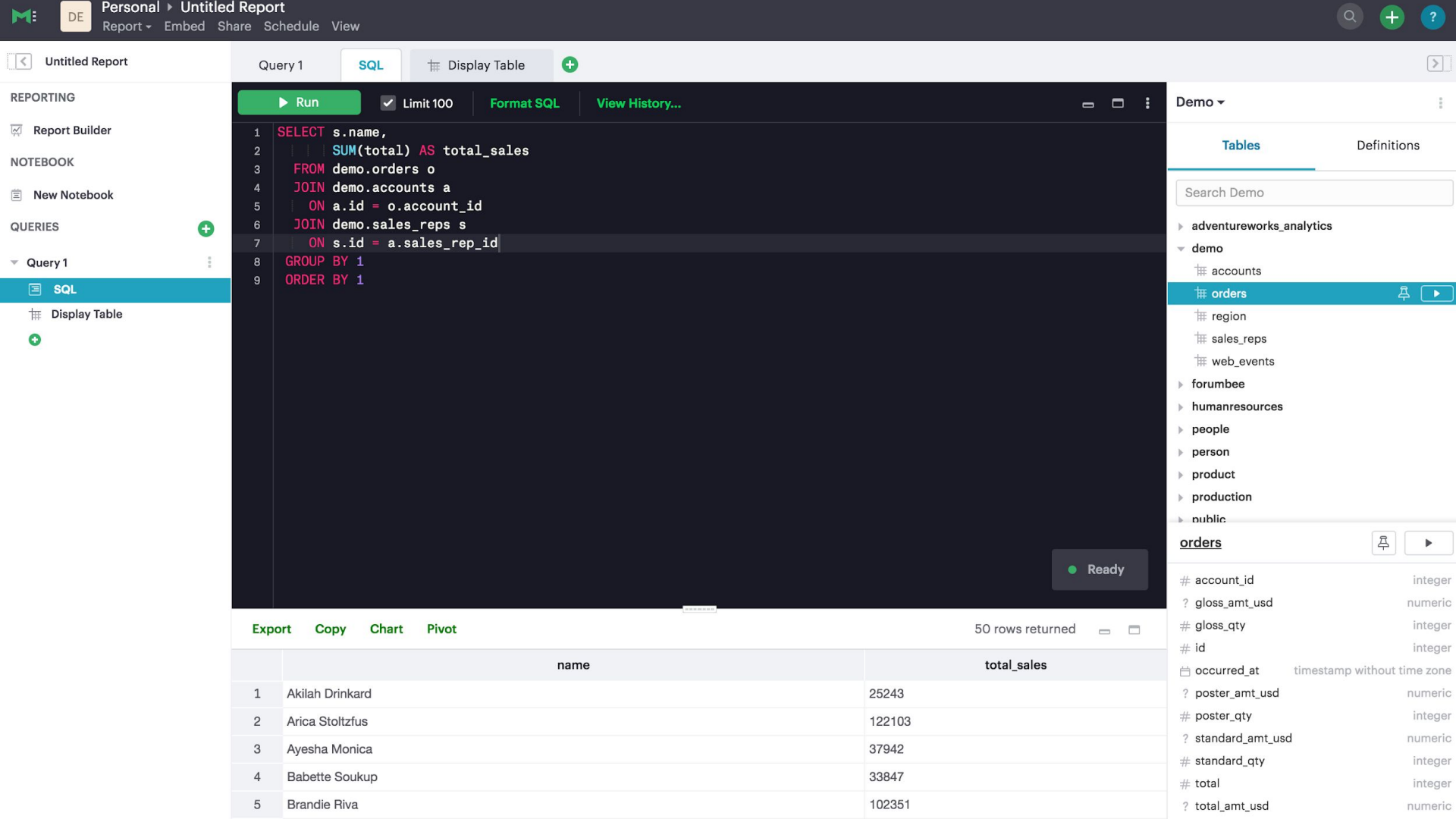
David Robinson
DataCamp
2019-01-18

How I used to think of my goals:



How I should have been thinking of them:

MODE


DE
Personal ▸ Untitled Report
Report Embed Share Schedule View
Untitled Report
REPORTING
Report Builder
NOTEBOOK
New Notebook
QUERIES
Query 1
SQL
Display Table
Query 1
SQL
Display Table
Run
Limit 100
Format SQL
View History...
1 SELECT s.name,
2 SUM(total) AS total_sales
3 FROM demo.orders o
4 JOIN demo.accounts a
5 ON a.id = o.account_id
6 JOIN demo.sales_reps s
7 ON s.id = a.sales_rep_id
8 GROUP BY 1
9 ORDER BY 1
Ready
Export Copy Chart Pivot
50 rows returned
name
total_sales
1 Akilah Drinkard 25243
2 Arica Stoltzfus 122103
3 Ayesha Monica 37942
4 Babette Soukup 33847
5 Brandie Riva 102351
Demo
Tables
Definitions
Search Demo
adventureworks_analytics
demo
accounts
orders
region
sales_reps
web_events
forumbee
humanresources
people
person
product
production
public
orders
account_id integer
gloss_amt_usd numeric
gloss_qty integer
id integer
occurred_at timestamp without time zone
poster_amt_usd numeric
poster_qty integer
standard_amt_usd numeric
standard_qty integer
total integer
total_amt_usd numeric

Query 1 | SQL | Display Table | Orders by Regi...

View

Orders by Region

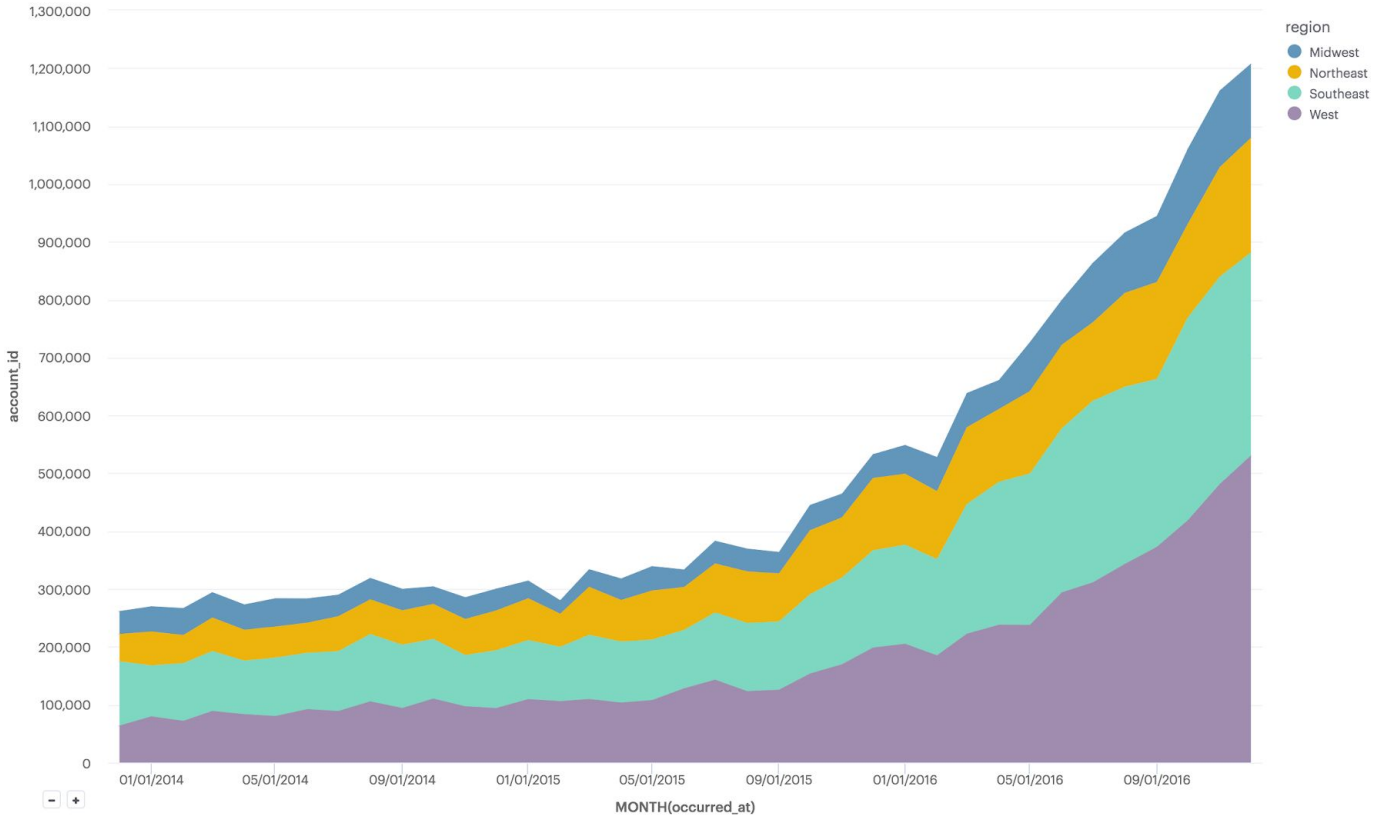


Chart Data

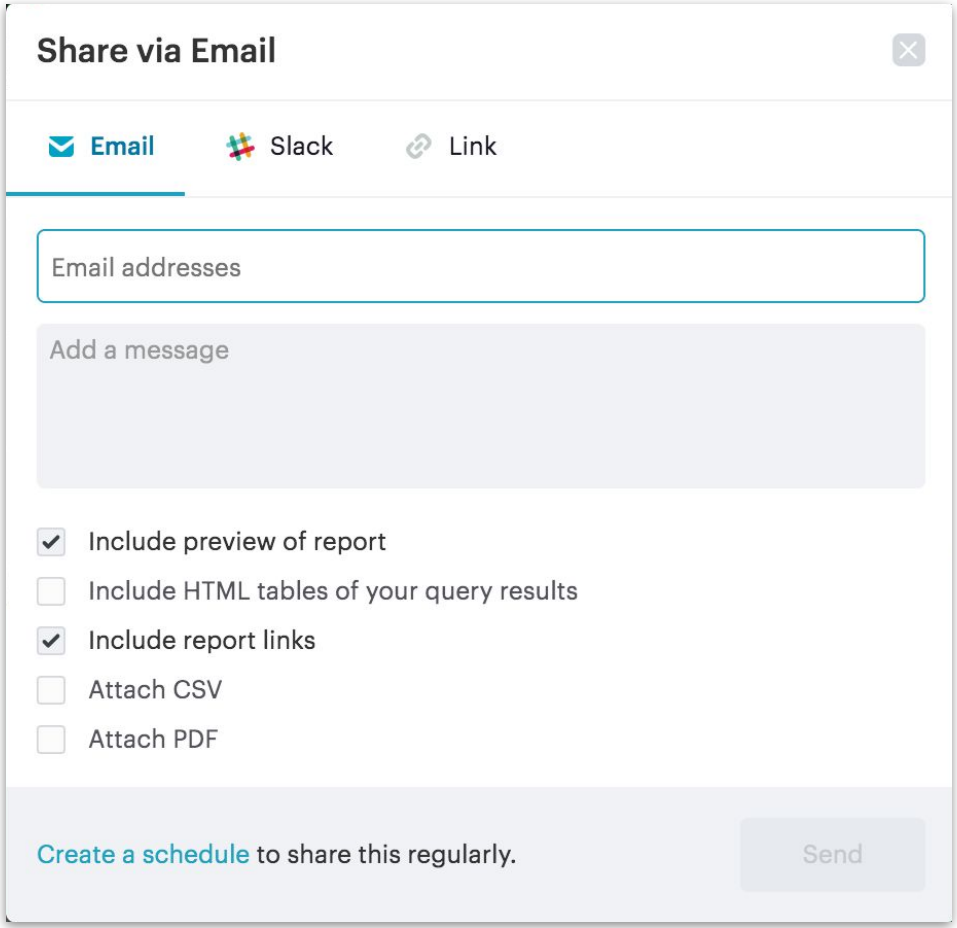
SETTINGS
Area Chart
X-Axis
MONTH(occurre...
Y-Axis
SUM(account_id)
Color
region

FILTERS
DAY(occurred_at)

Fields | Format

DIMENSIONS
occurred_at
region

MEASURES
account_id
gloss_amt_usd
gloss_qty
id
poster_amt_usd
poster_qty
standard_amt_usd
standard_qty
total
total_amt_usd


Share via Email
Email
Slack
Link
Email addresses
Add a message
Include preview of report
Include HTML tables of your query results
Include report links
Attach CSV
Attach PDF
Create a schedule to share this regularly.
Send

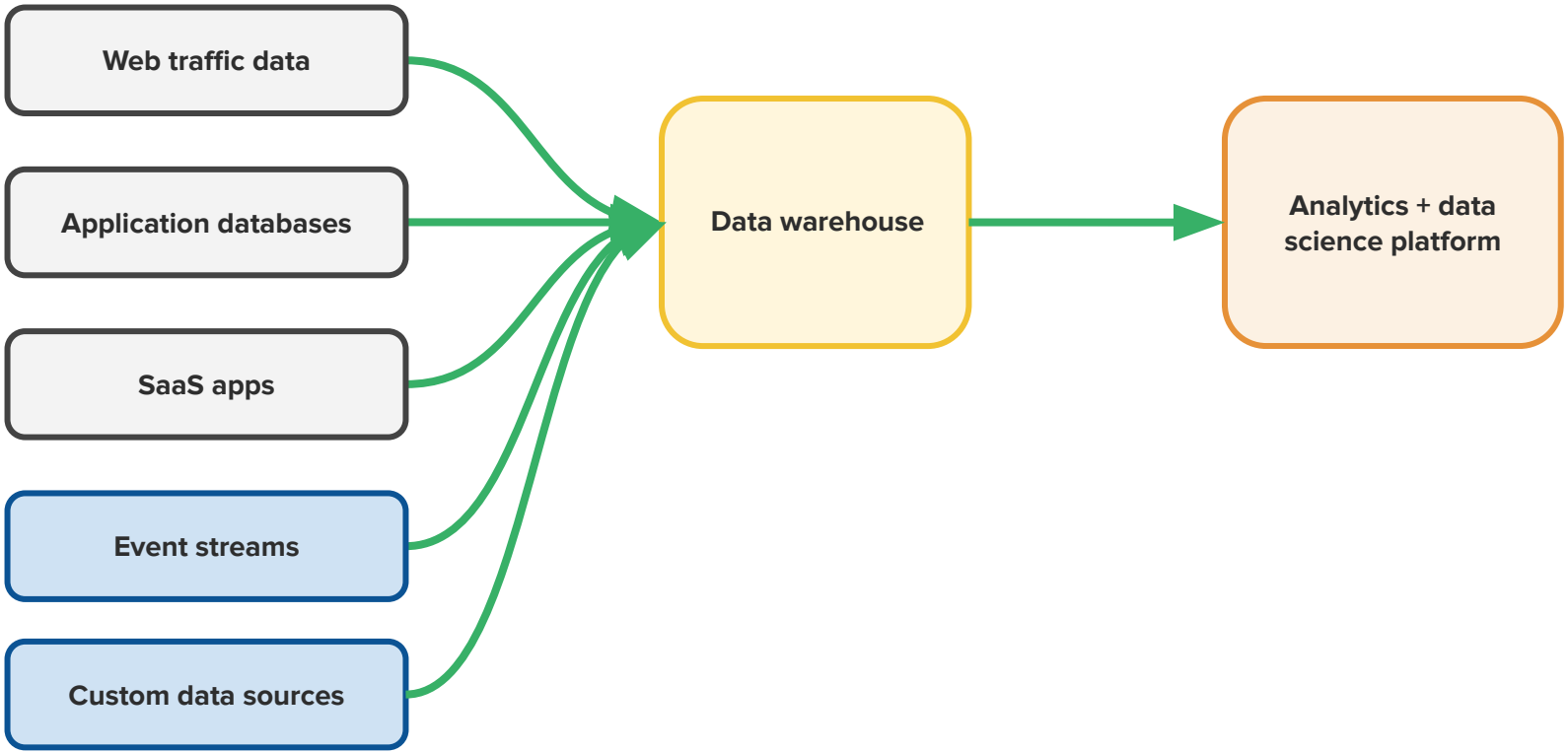
redash

Metabase

Step 5

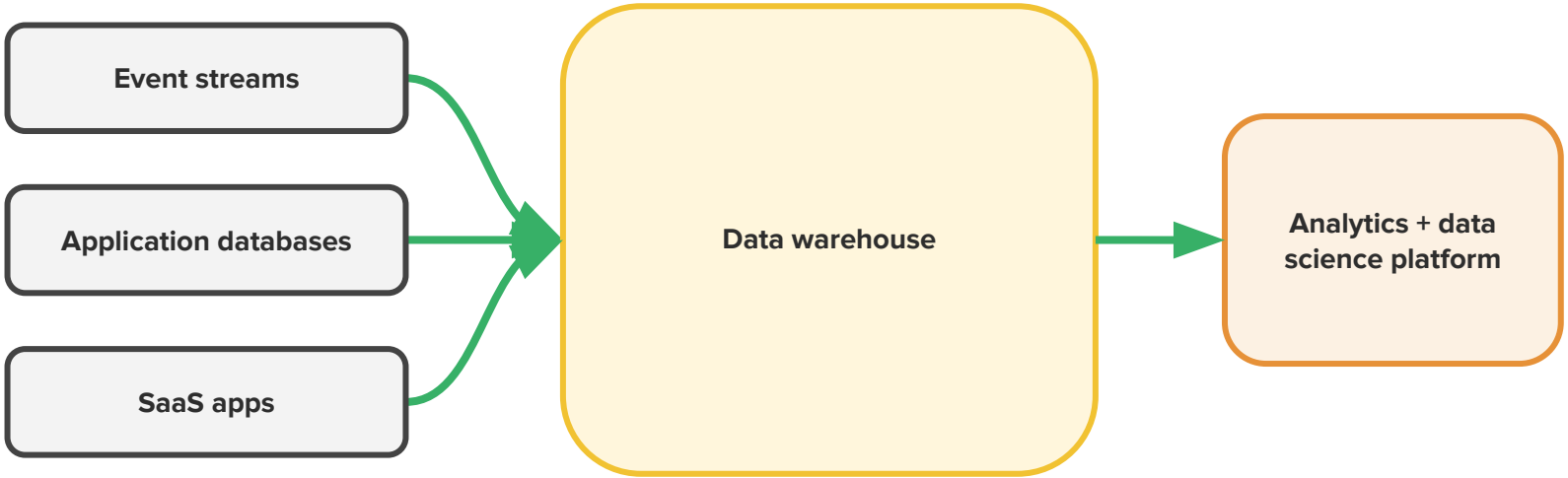
What comes next

Integrating more data sources

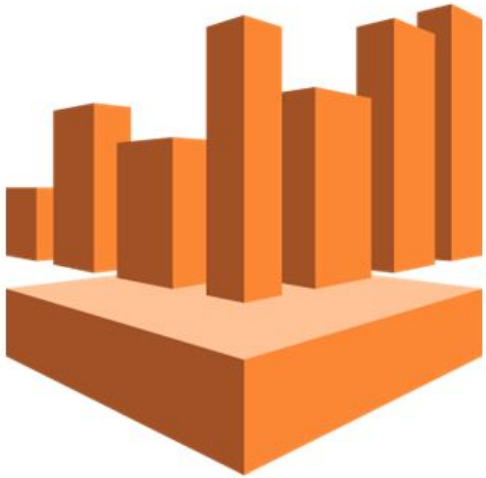


Kinesis
Firehose

Dealing with enormous scale

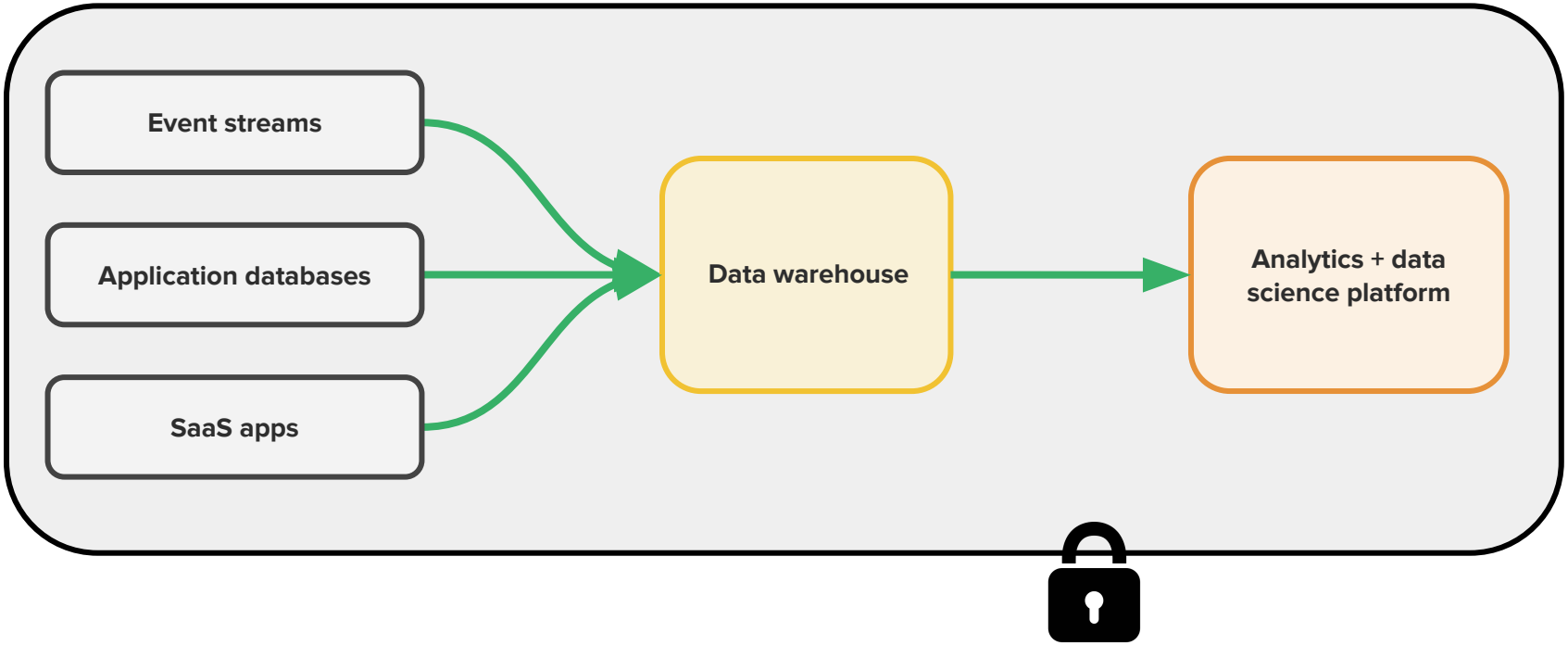


Cluster: resize-example
Cluster
Database
Backup
Query cluster
Modify cluster
Resize cluster
Delete cluster
Reboot cluster
Upgrade release version
hzjsmnbrsd.us-east-1-qa.redshift-dev.
resize-example
Multi Node
dc2.large
8
us-east-1d



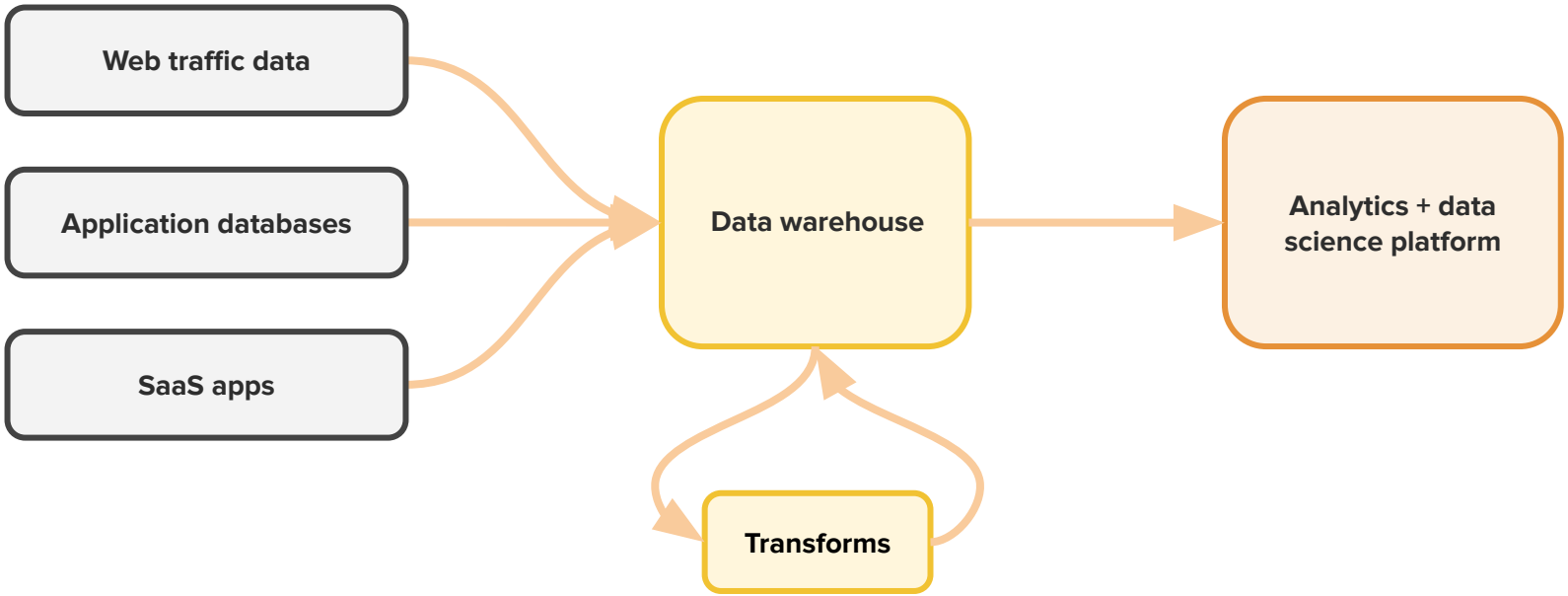
Amazon Athena

Data governance

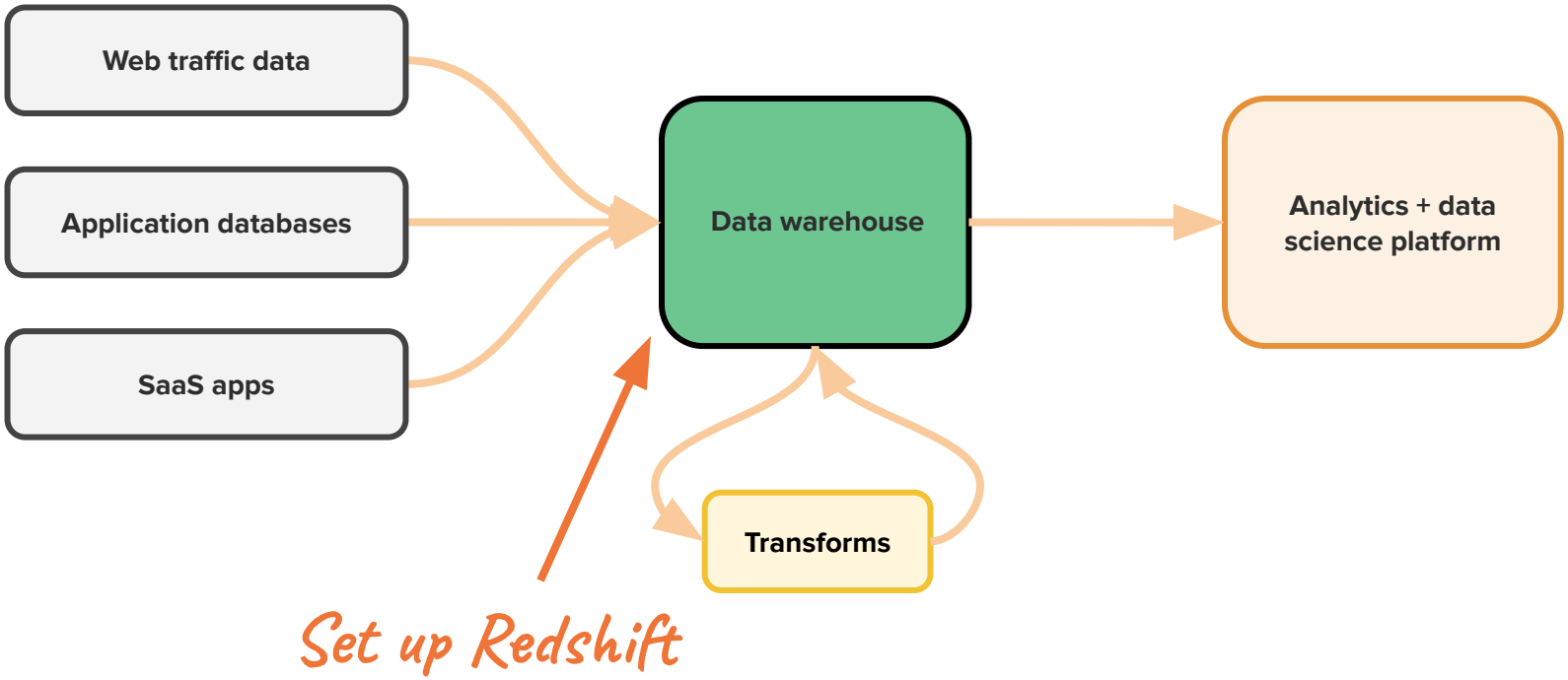


A live demo!

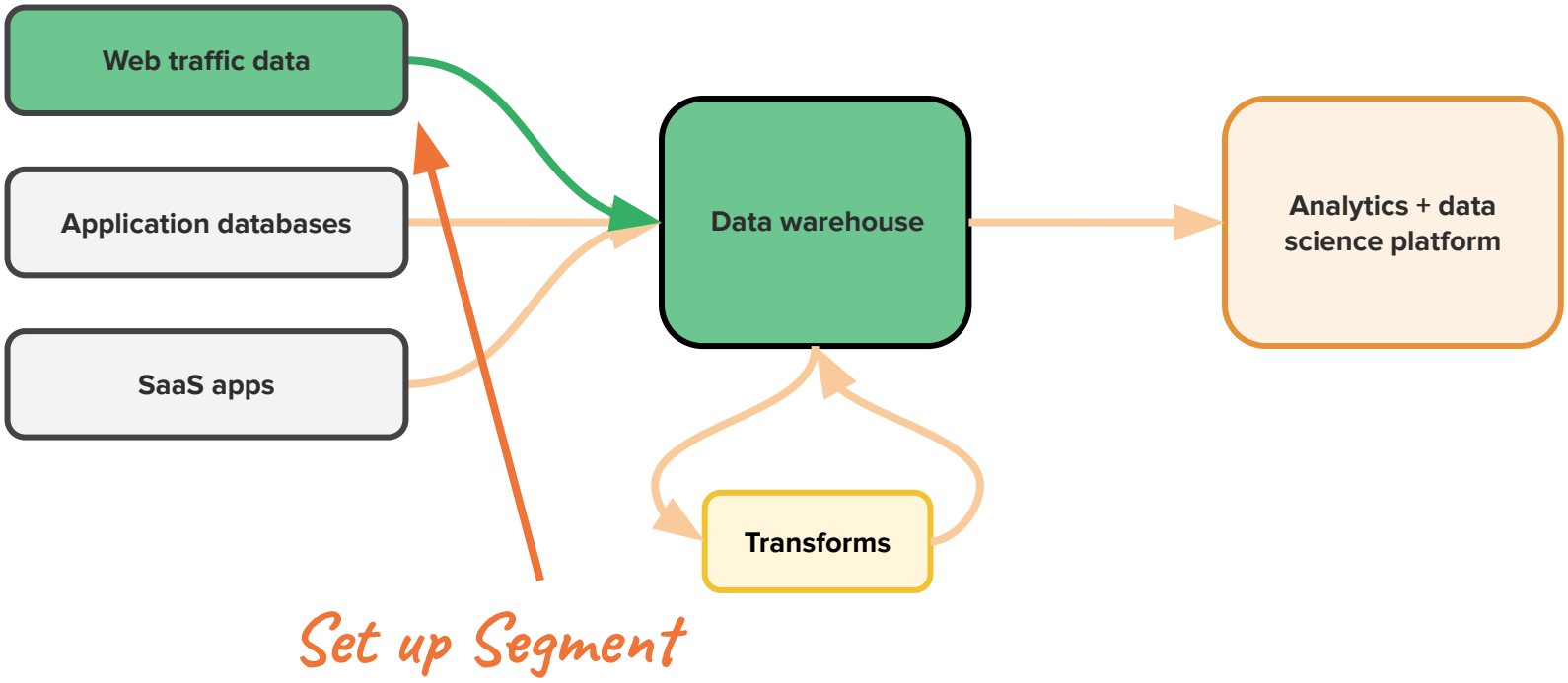
How we'll set this up



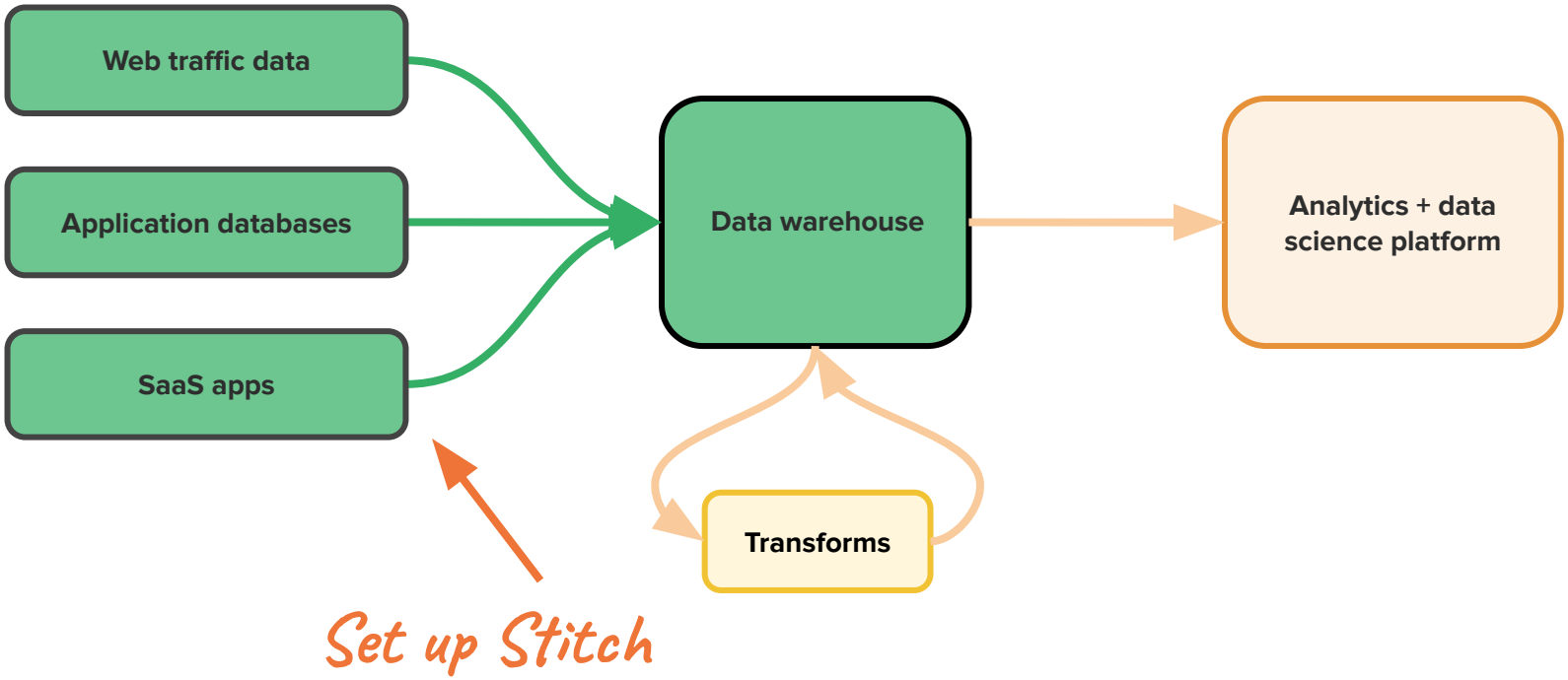
How we'll set this up



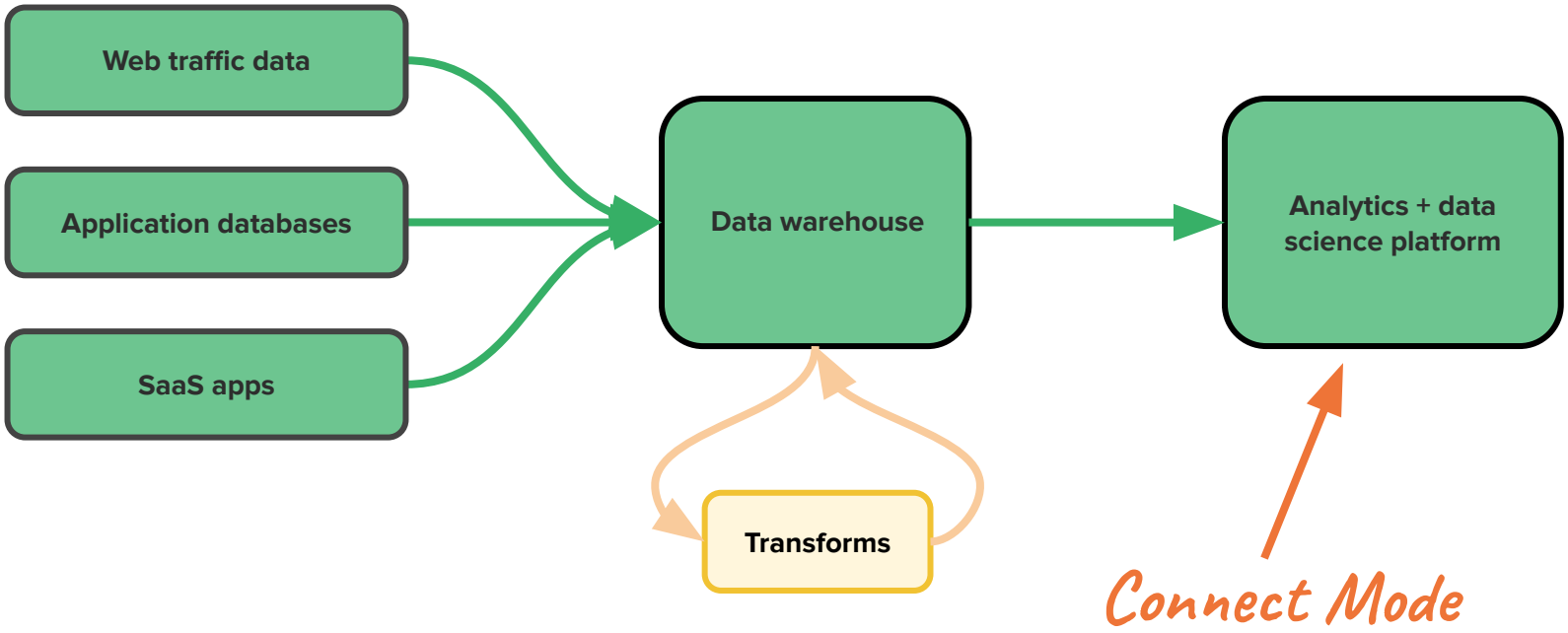
How we'll set this up



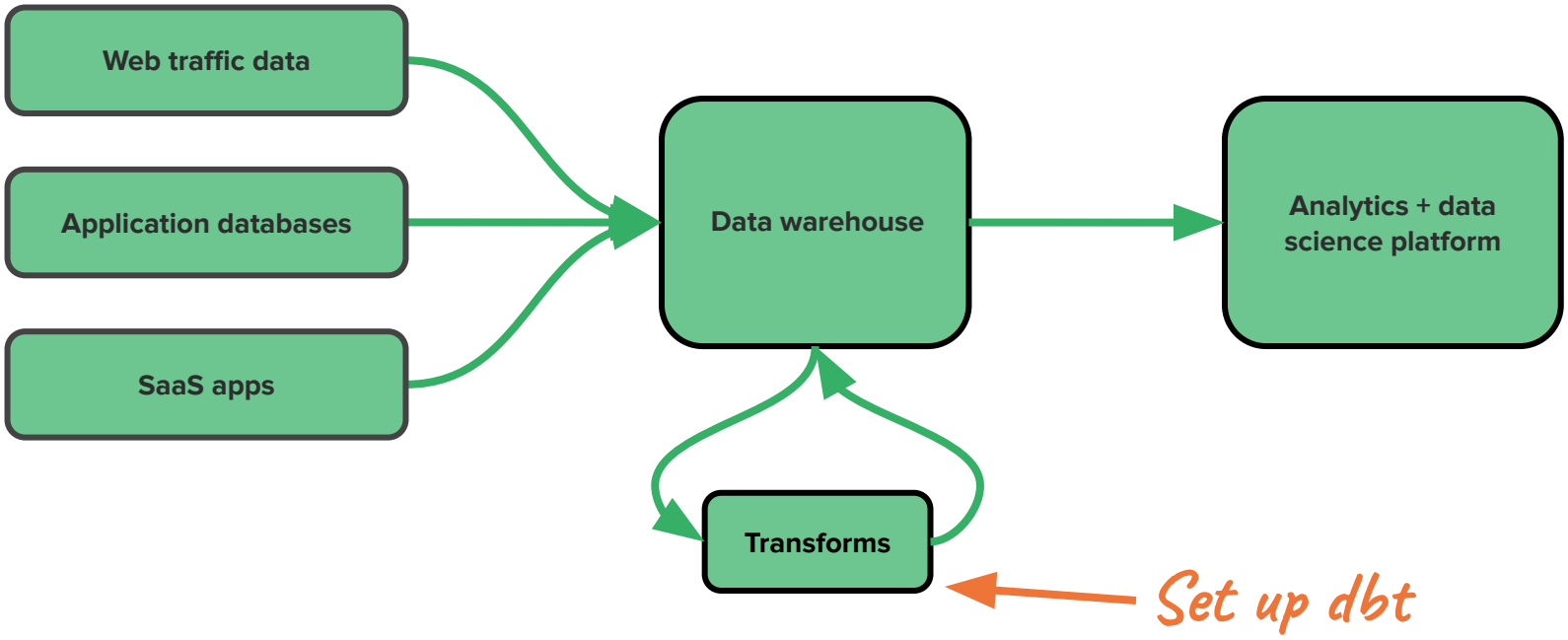
How we'll set this up



How we'll set this up



How we'll set this up



Existing services

- AWS account
- An application database
- Intercom (or any other 3rd party SaaS tool)

New services

- Amazon Redshift
- Segment
- Stitch
- Mode
- dbt

Questions?

Backup

Why invest in data?

Analytics and data science are now must-haves, not afterthoughts.

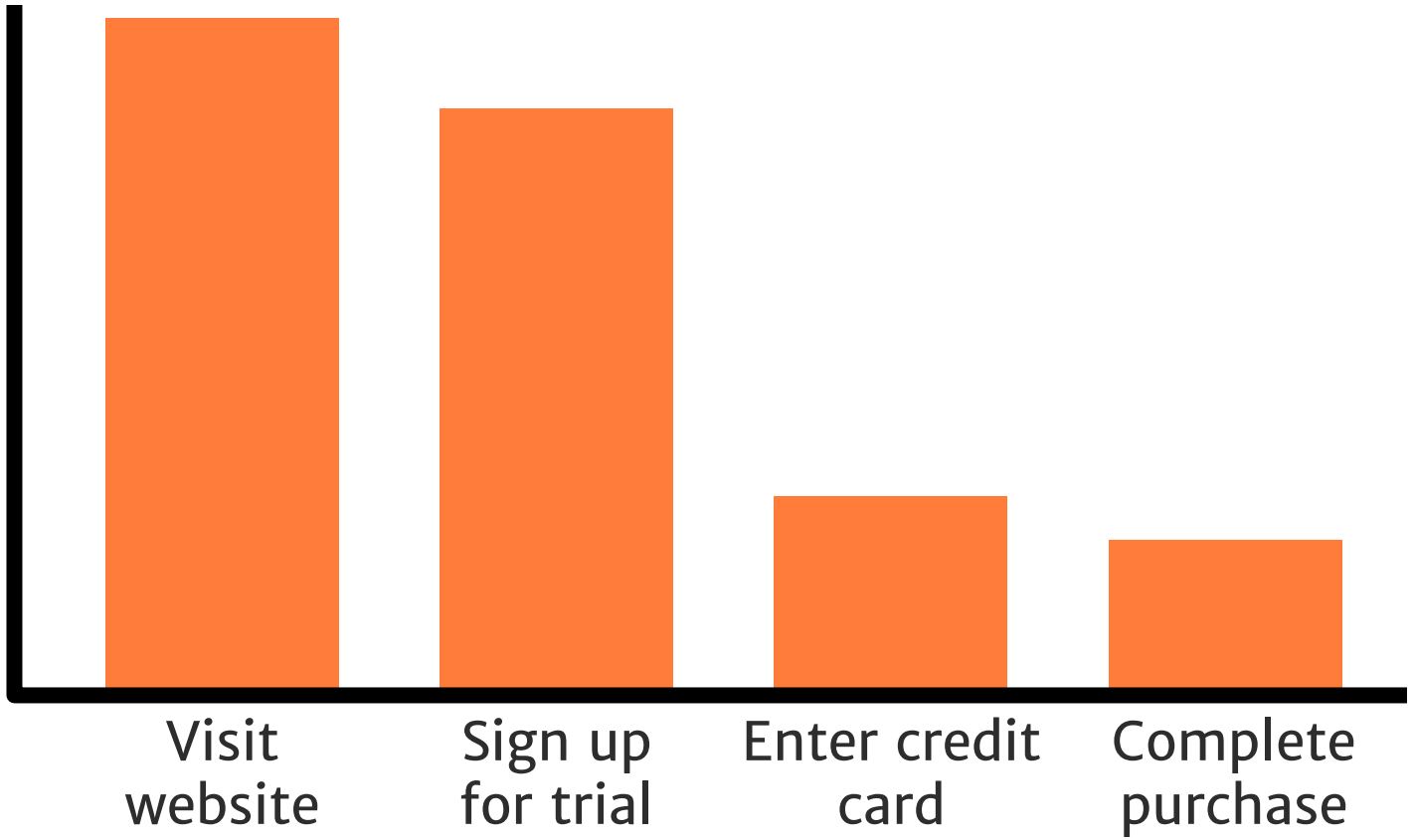
Sequoia Capital

Source: "Data-Informed Product Building," Sequoia Capital

Three other reasons to do this today

1. **It's the best way to prioritize**

0
5
10
15
20
25
30
35
40
45
50
55
TIME TIMER®


Visit website
Sign up for trial
Enter credit card
Complete purchase

Wedding Planner by The Knot

4+

All-in-one Wedding Planner

The Knot Inc.

#133 in Lifestyle

★★★★★ 4.8, 23.9K Ratings

Free

Source: Segment.com, “The Rise of the Chief Data Officer.”

Three other reasons to do this today

1. **It's the best way to prioritize**

2. **Data is a critical part of fundraising**

- Signups
- Daily active users
- Monthly revenue
- Retention
- DAU/MAU
- Quick ratio
- Payback period
- LTV/CAC
- Magic number

Three other reasons to do this today

1. **It's the best way to prioritize**
2. **Data is a critical part of fundraising**
3. **It will never get easier**

Google Analytics

Google Analytics

Google Analytics

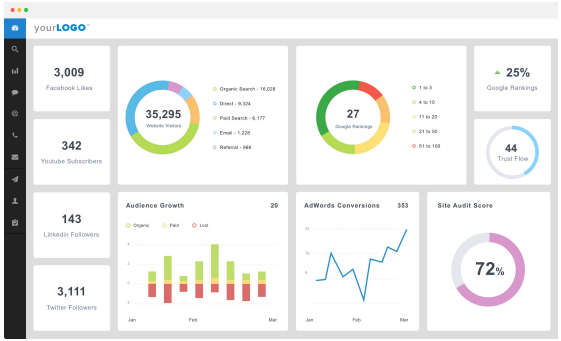
Google Analytics

Google Analytics

stripe

Google Analytics

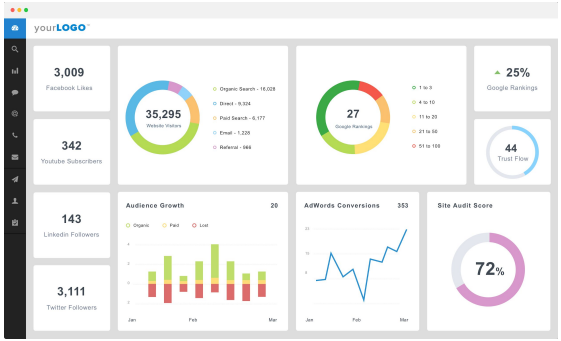
stripe



Google Analytics

stripe

salesforce



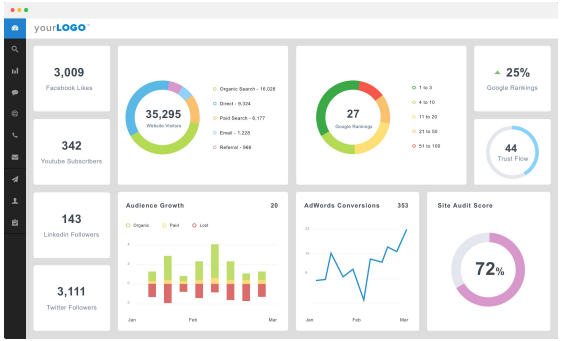
How much money did we make last quarter?

A board member

Google Analytics

stripe

salesforce



If you do it now, you can build the foundation *in an hour.*

Agenda

Agenda

- **Why startups should invest in data**

Agenda

- Why startups should invest in data
- An introduction to a modern data stack

Agenda

- Why startups should invest in data
- An introduction to a modern data stack
- A live demo

Agenda

- Why startups should invest in data
- An introduction to a modern data stack
- A live demo
- Q&A and hands-on help