Does this metrics workflow look familiar?

Create Table → Schedule Airflow Job → Backfill Table → A/B Testing Config → Add table to Superset → Create Dashboard → Anomaly Detection Config → Tell consumers to update queries → Monitor Pipelines
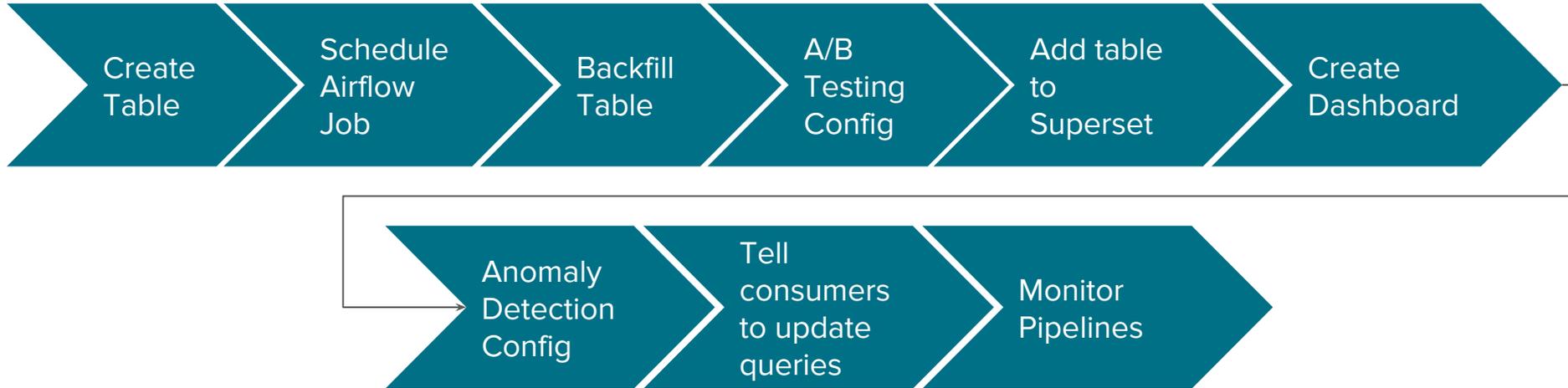
Lauren Chircus

Company: Airbnb
Role: Product Manager
Previous Role: Data Scientist
Twitter: @lchircus
Fun Fact: This Airbnb near
Salinas was my favorite

You can change the paradigm!

Global Metric Config

Create Dashboard

Changing the metric management paradigm

1. Airbnb's journey

2. Why you should make dimensions first class citizens

3. Why prioritize bonus features early

Airbnb's Journey

Plethora of tools for building & accessing data

Strong, open source-based compute environment

Consuming metrics was painful, too

Metrics weren't reusable across tools -> discrepancies

Consuming metrics was painful, too

Metrics weren't reusable across tools -> discrepancies

Metrics were hard to find

Consuming metrics was painful, too

Metrics weren't reusable across tools -> discrepancies

Metrics were hard to find

Required SQL knowledge or prepared dashboards

Global Metrics Framework

Superset

A/B testing

Airflow

Anomaly Detection

Dataportal

HIVE

presto

druid

What is Global Metrics?

"Global Metrics" is the concept that metrics should be **defined in one place**, **have strong metadata**, and available **wherever you need them**.

Can we reuse existing infra?

Global Metrics Framework

?
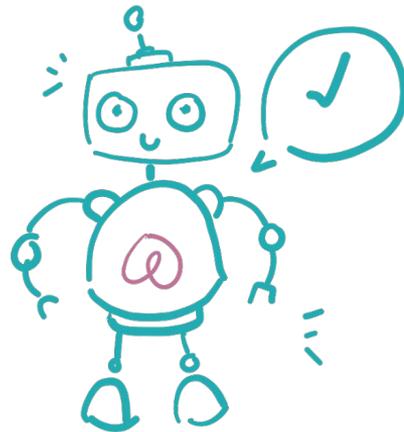
ML Feature Framework

The basic frameworks look similar

Logic & metadata store

Compute data

Consuming Apps

ML: serve data to models

Logic & metadata store

Compute data

Consuming Apps

Search	Pricing	Fraud	....
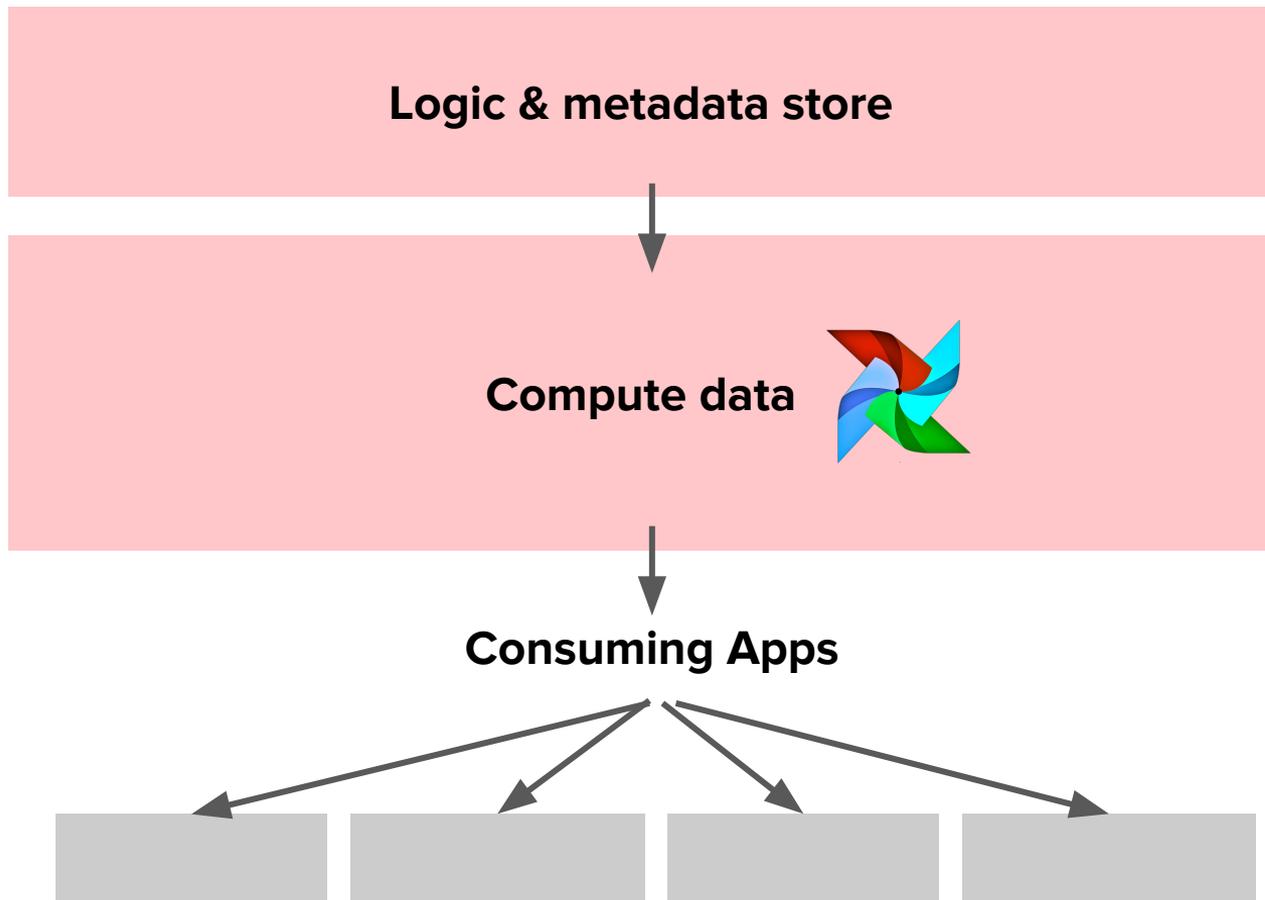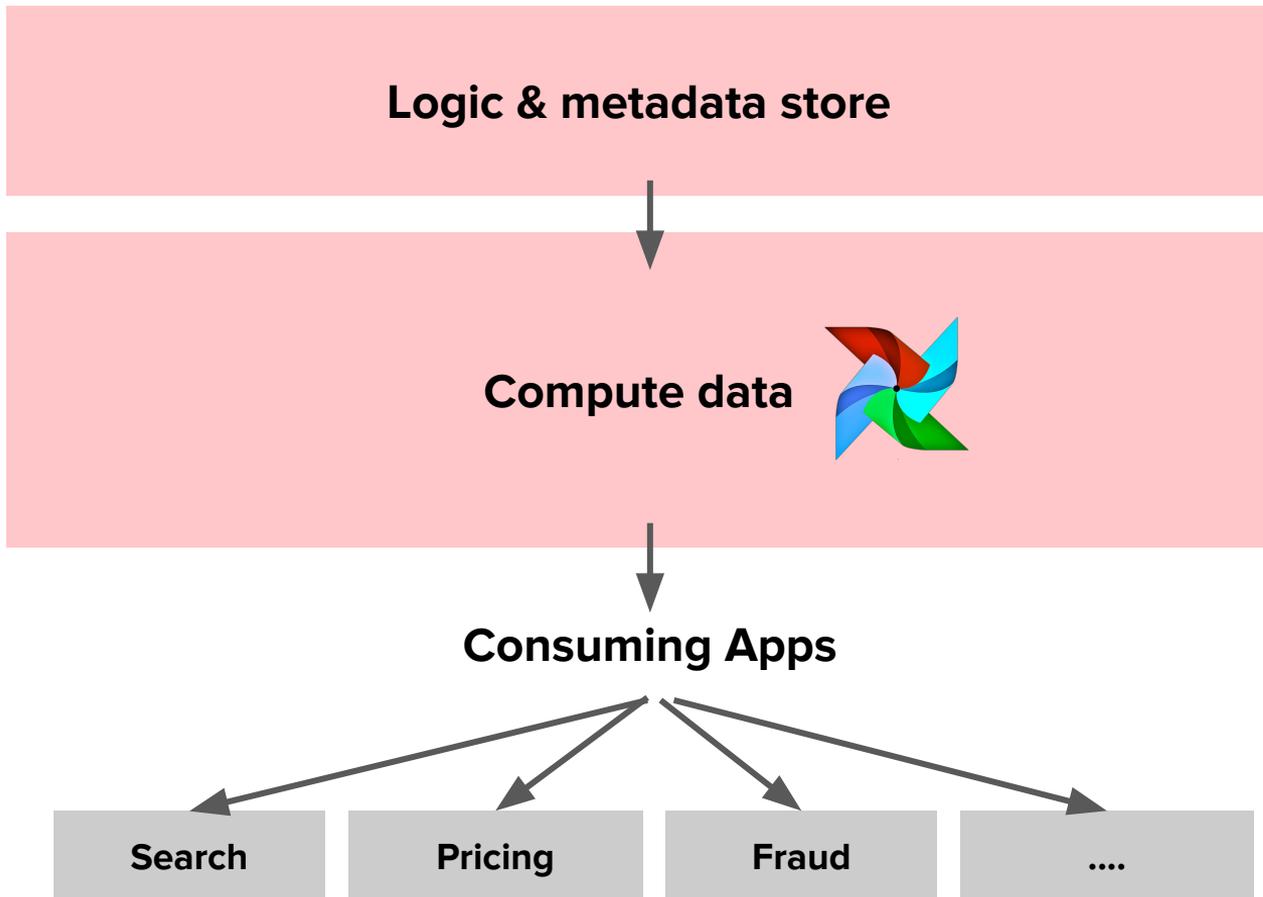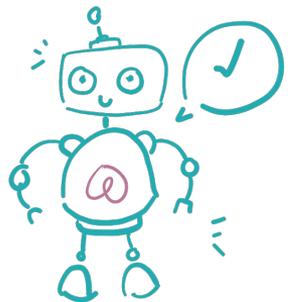
Metrics: serve data to apps

Logic & metadata store

Compute data

Consuming Apps

Superset

Anomaly
Detection

A/B testing

Dataportal
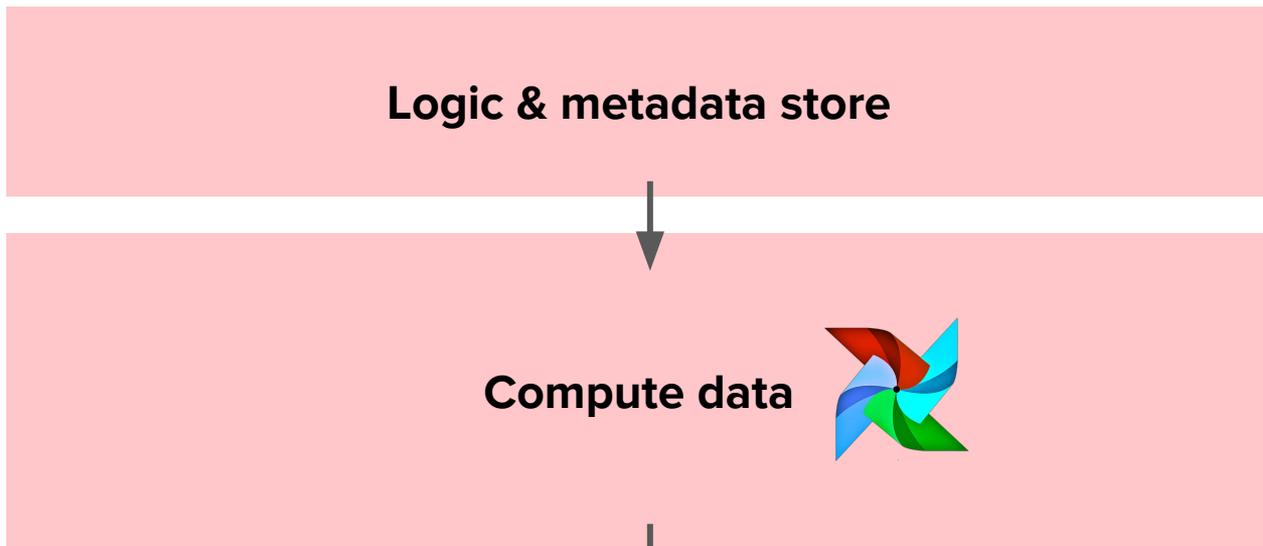
Metrics are different than ML features

Metrics

ML Features
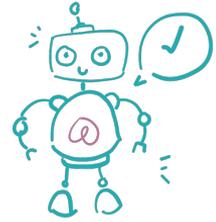
- Leverage as much information as possible

- Entirely offline

- Diverse metric types

Metrics are different than ML features

Metrics

- Leverage as much information as possible
- Entirely offline
- Diverse metric types

ML Features
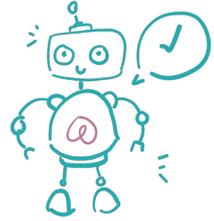
- Prevent data leakage to keep models clean
- Available online and offline
- Windowing functions

Similar basics, different details
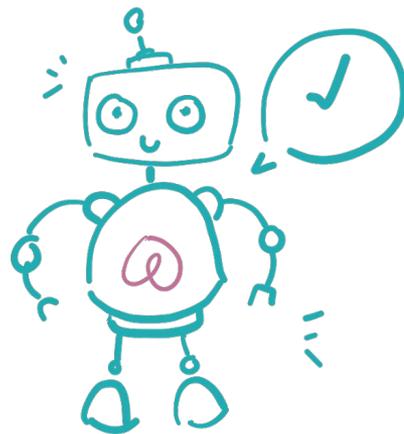
Global Metrics Framework ≠ **ML Feature Framework**

Why dimensions are 1st class citizens

Denormalization makes analytics speedy
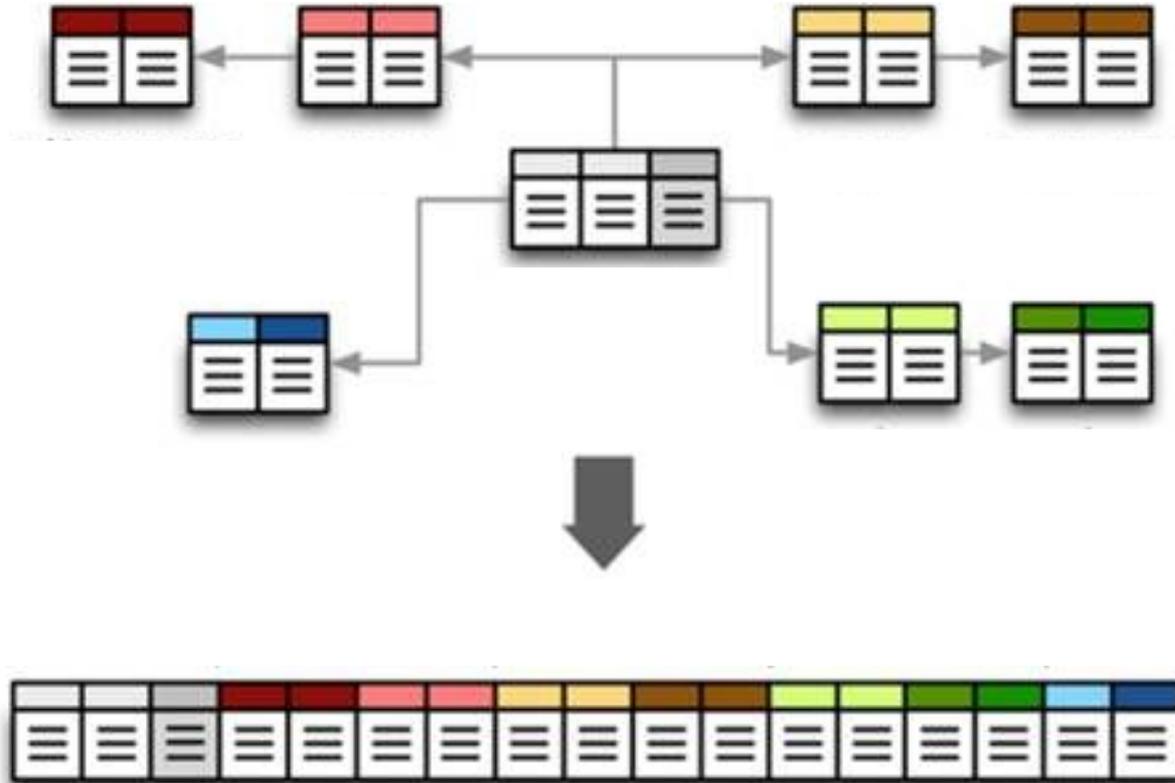
druid **doesn't allow joins**

timestamp	shape	color	count
12:00	square	yellow	23
12:00	circle	yellow	2
12:00	square	red	57
12:00	circle	red	188

Many metrics are dimensional cuts

Company

Bookings

Many metrics are dimensional cuts

Growth

First Time Bookings

Company

Bookings

Many metrics are dimensional cuts

Growth

First Time Bookings

Company

Bookings

China

Bookings in China

Many metrics are dimensional cuts

Company

Bookings

Growth

First Time Bookings

China

Bookings in China

Airbnb for Work
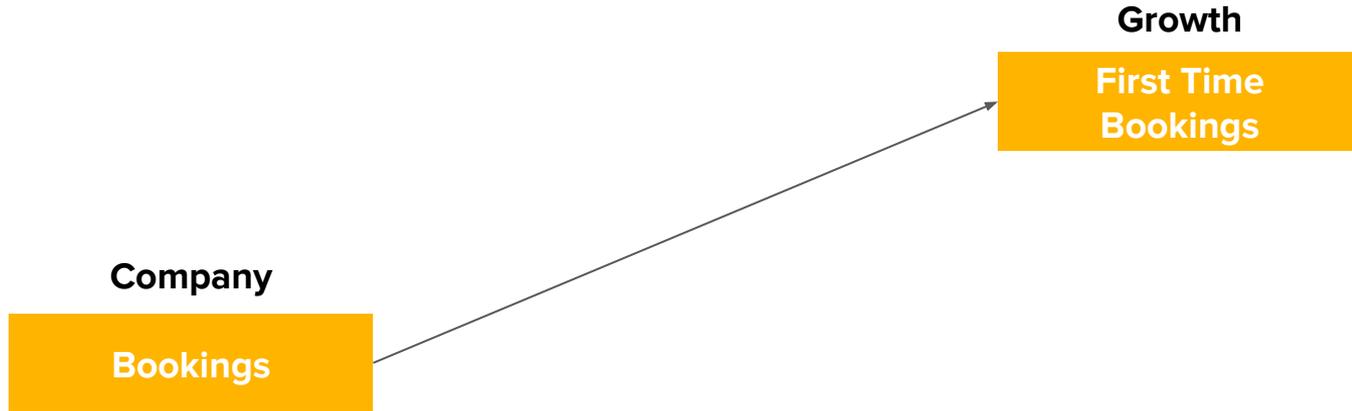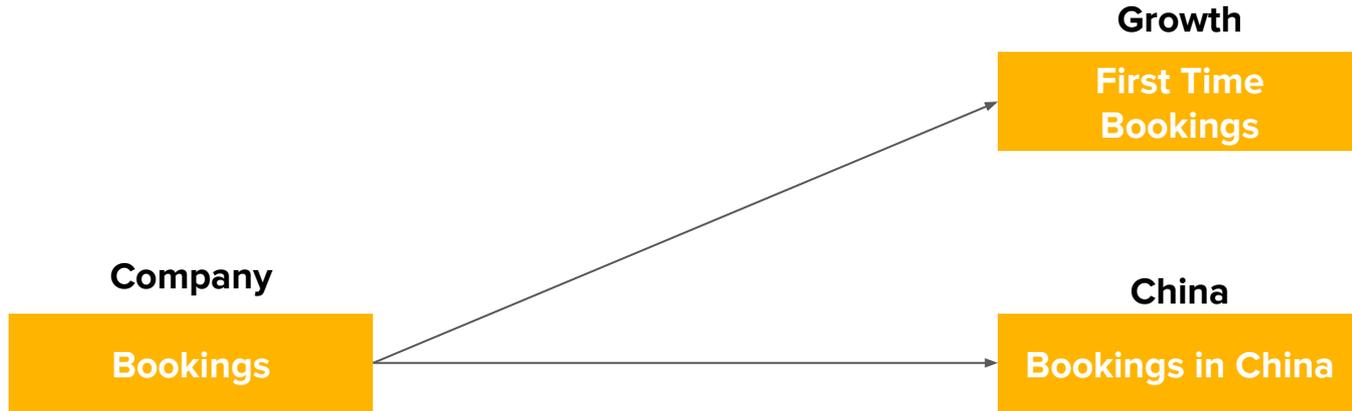
Business Trip Bookings
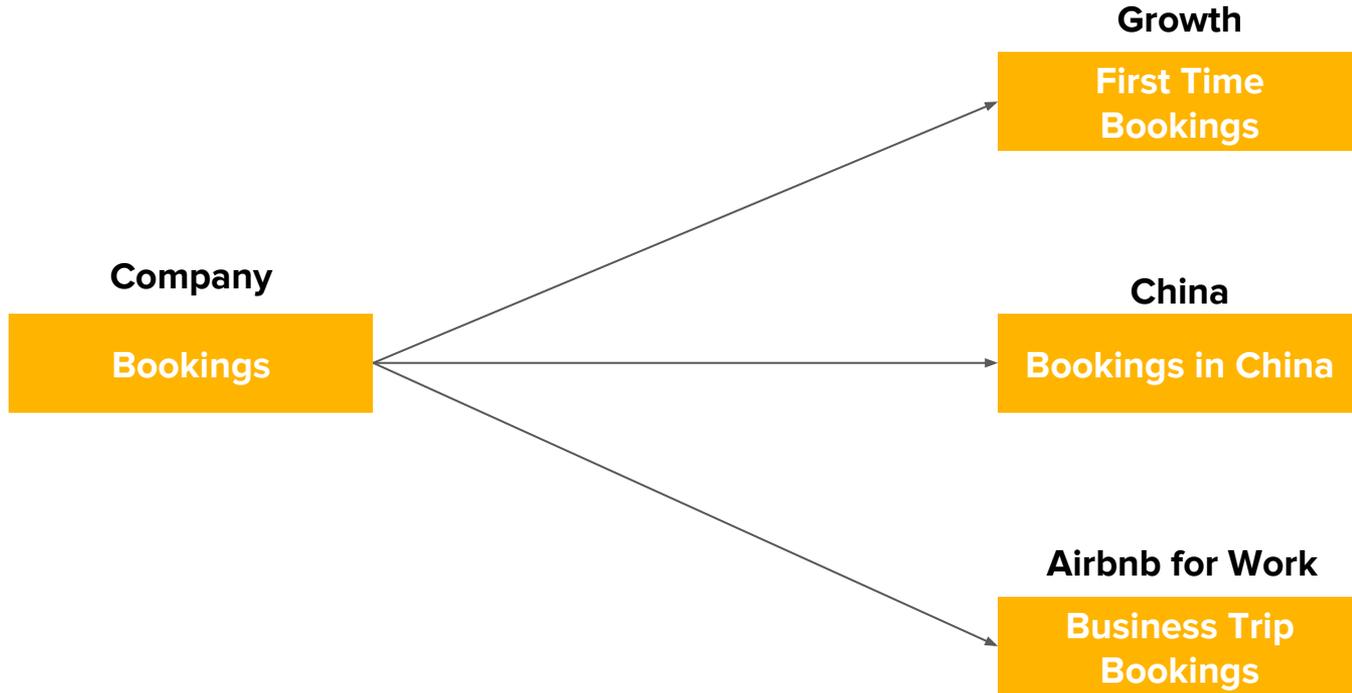
Exploratory analysis across many dimensional cuts

Bookings

in China

from North America

by host status

Standard Star Schema

Foreign Key

Dim_Date
- Id
- Date
- Day
- Day_of_Week
- Month
- Month_Name
- Quarter
- Quarter_Name
- Year

Fact_Sales
- Date_Id
- Store_Id
- Product_Id
- Units_Sold

Foreign Key

Dim_Store
- Id
- Store_Number
- State_Province
- Country

Foreign Key

Dim_Product
- Id
- EAN_Code
- Product_Name
- Brand
- Product_Category
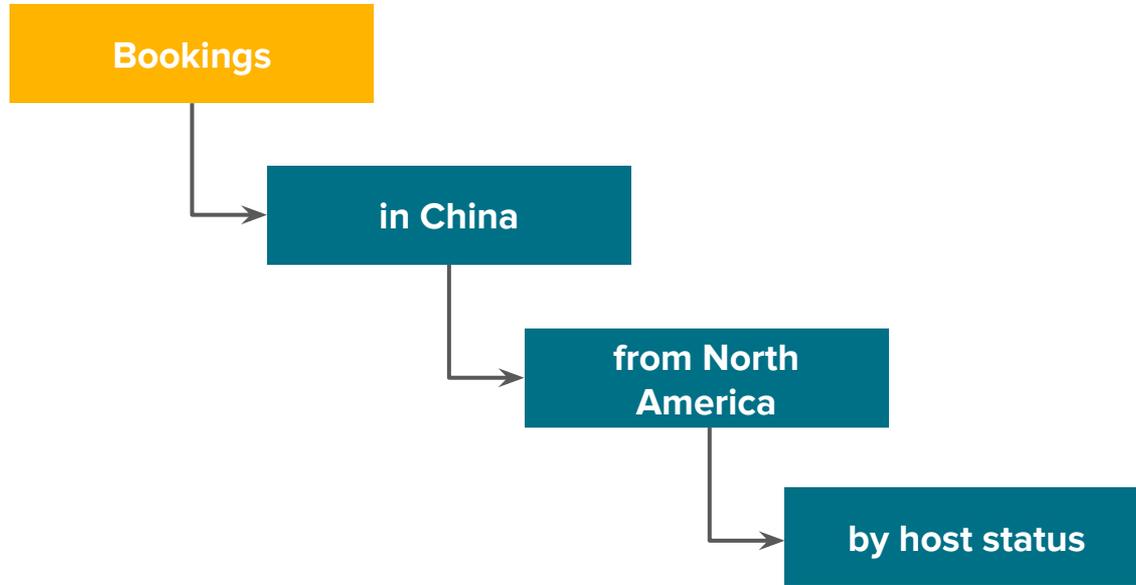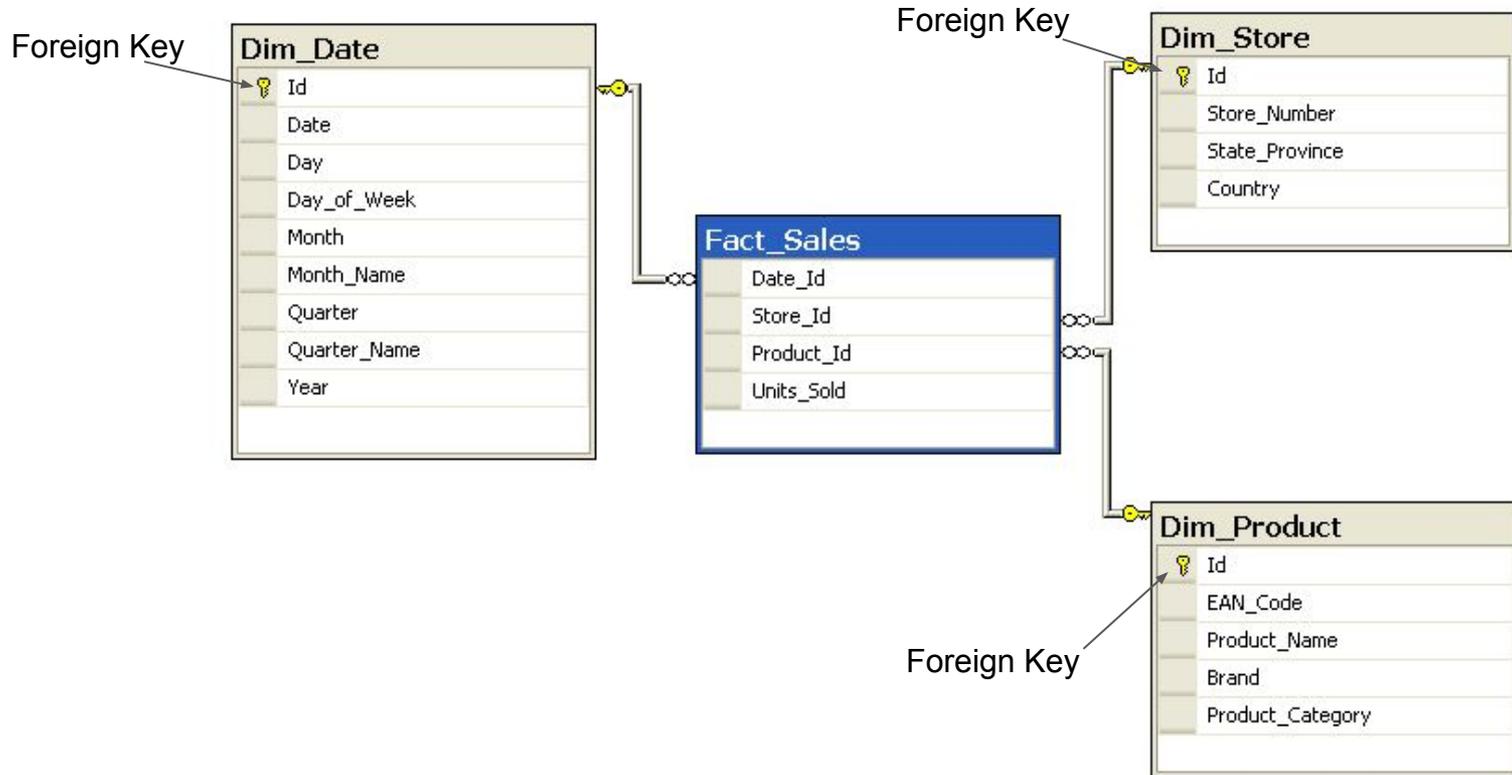
Global Metrics Framework Naming

Subject

Dimension Source

🔑	Id
	Date
	Day
	Day_of_Week
	Month
	Month_Name
	Quarter
	Quarter_Name
	Year

Metric Source

Date_Id
Store_Id
Product_Id
Units_Sold

Subject

Dimension Source

🔑	Id
	Store_Number
	State_Province
	Country

Dimension Source

🔑	Id
	EAN_Code
	Product_Name
	Brand
	Product_Category
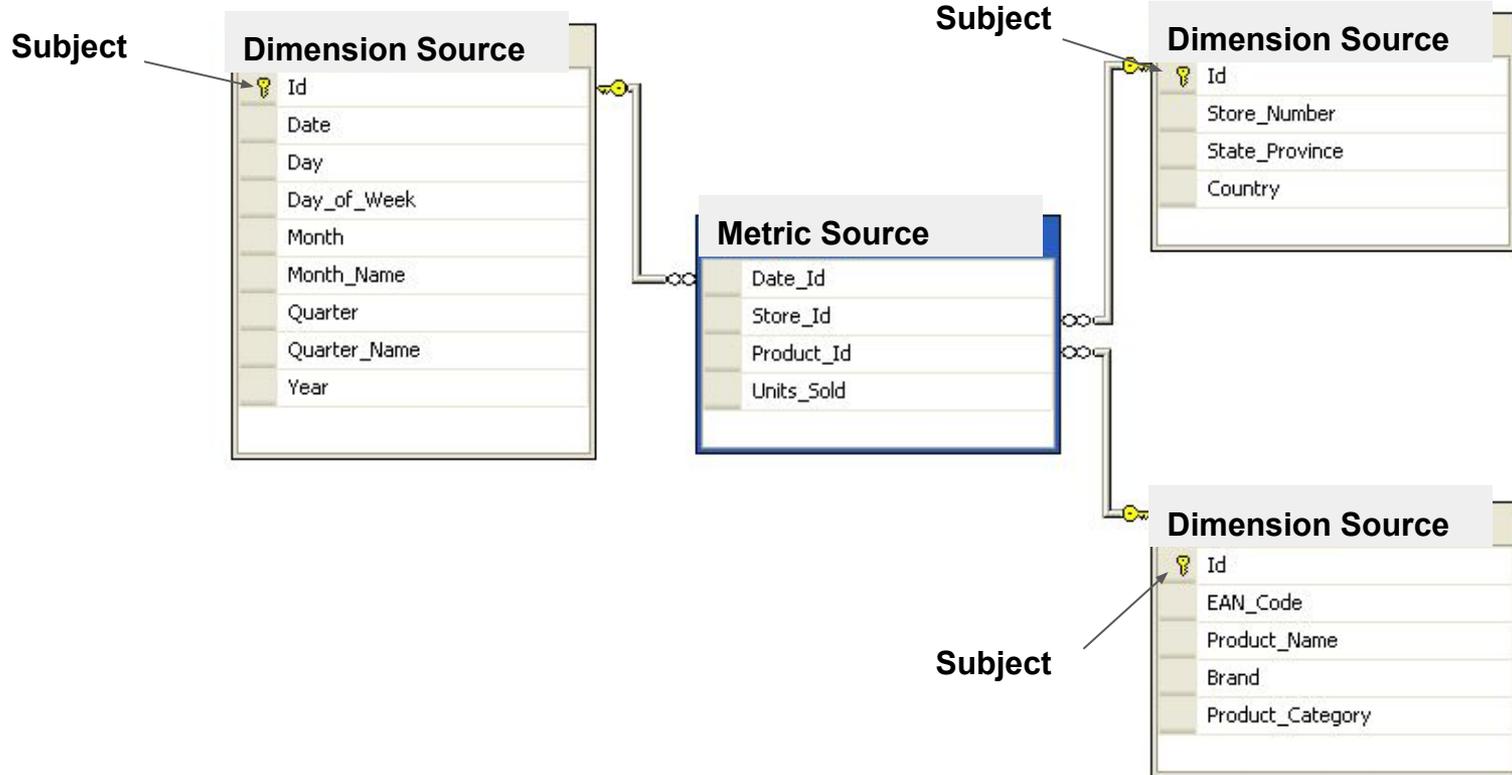
Subject

YAML configs instead of tables

Dimension Source

Destination_geo.yaml

Dimension Source

Origin_geo.yaml

Metric Source

Bookings.yaml

Dimension Source

Host_status.yaml
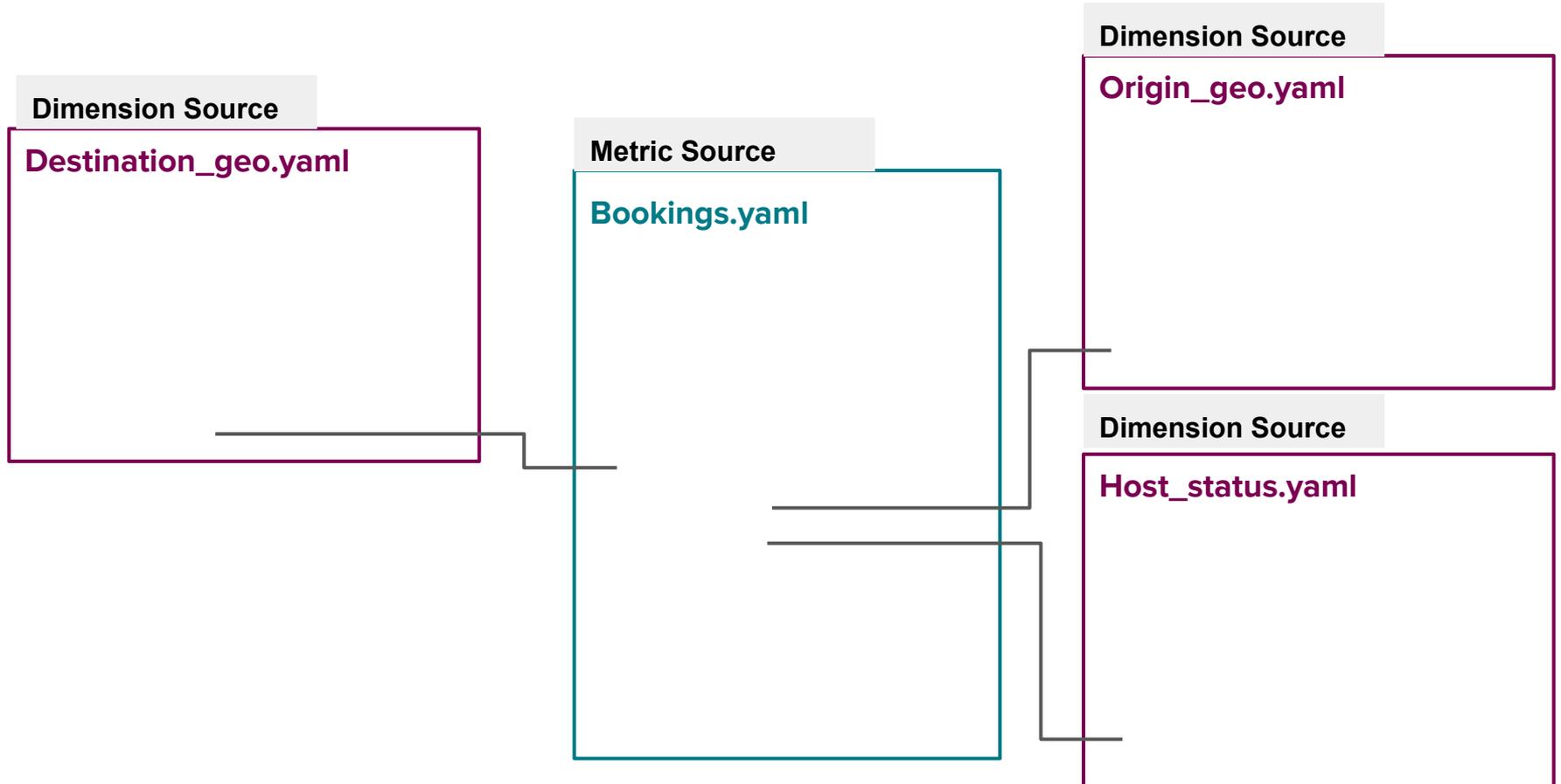
Data scientists list which dimensions to include

_metric_source: bookings_

metrics:
    - bookings
    - nights

subjects:
    -   listing
    -   guest
    -   host

**dimensions:**
    - dim_destination_china
    - dim_origin_region
    - dim_new_host

Automatically joins to the relevant dimension sources

_dim_source: destination_geo_

dimensions:
- dim_destination_region
- dim_destination_china

subject:
-     listing

_metric_source: bookings_

metrics:
- bookings
- nights

subjects:
-     listing
-     guest
-     host

dimensions:
- dim_destination_china
- dim_origin_region
- dim_new_host

_dim_source: origin_geo_

dimensions:
- dim_origin_region

subject:
-     guest

_dim_source: host_status_

dimensions:
- dim_new_host

subject:
-     host

Bookings has hundreds of dimensions
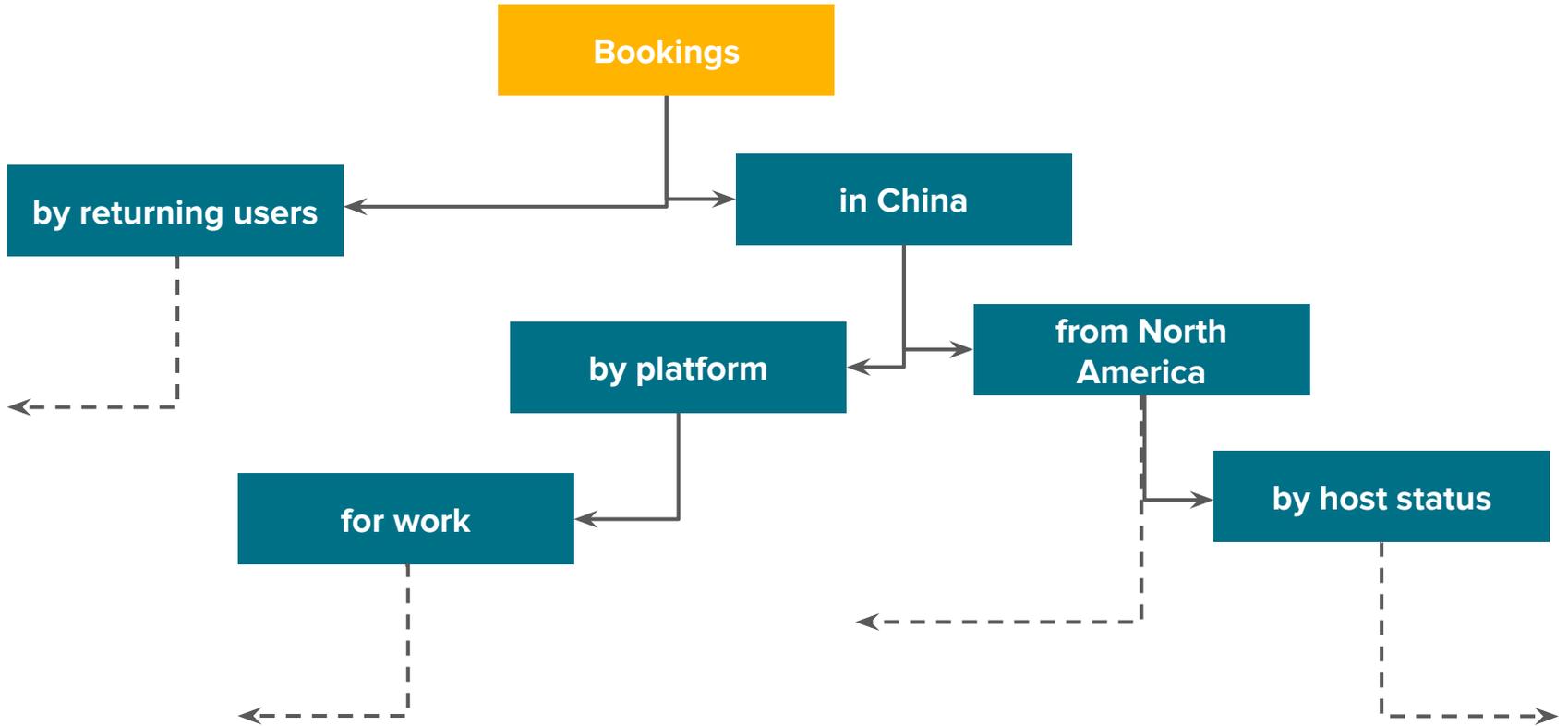
Expensive dimensions
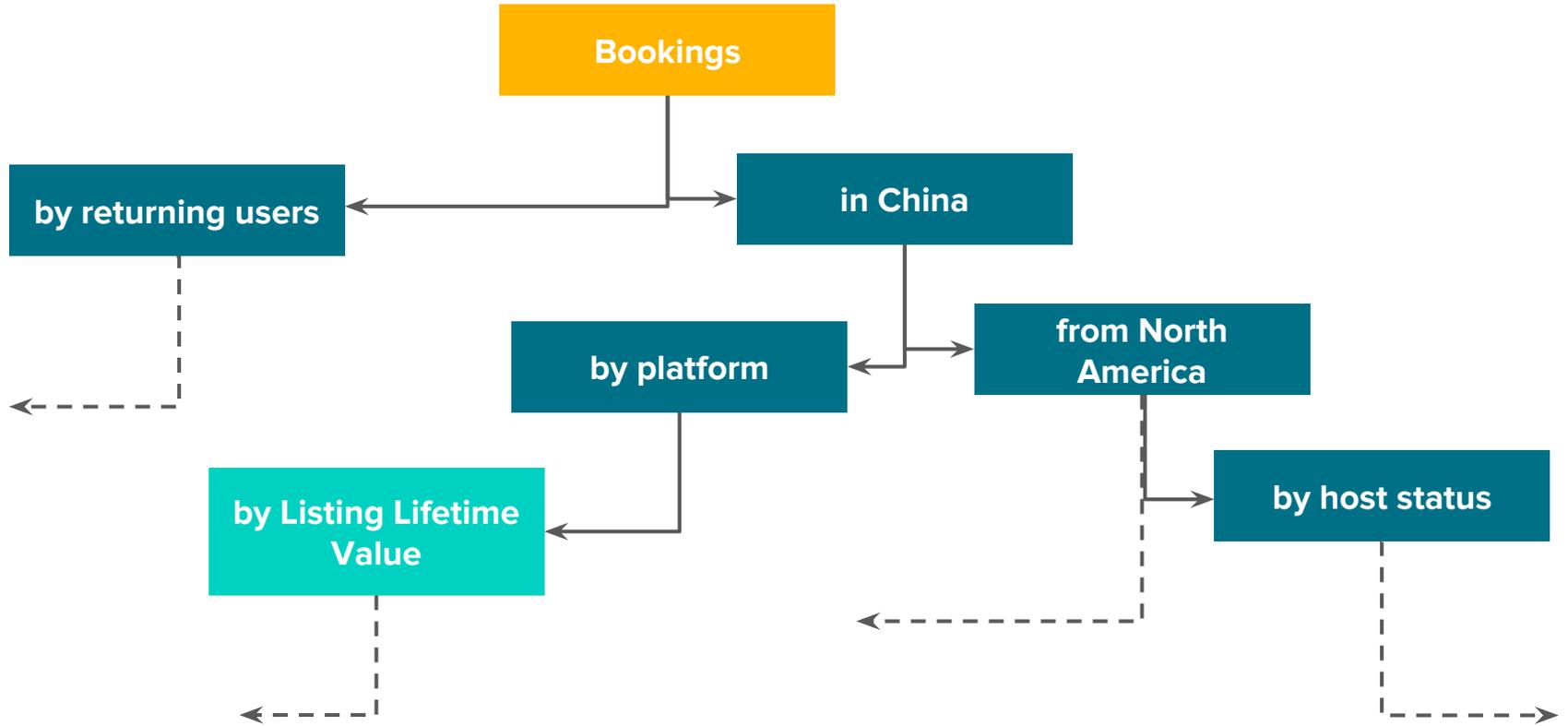
Dimension sets give DS control over SLAs

metric_source: bookings

metrics:
   - bookings
   - nights

dimension_sets:
   china_dims:
      - dim_destination_china
      - dim_origin_region

   host_dims:
      - dim_new_host
      - dim_origin_region

Dimension sets give DS control over SLAs

_metric_source: bookings_

metrics:
- bookings
- nights

dimension_sets:
   china_dims:
     - dim_destination_china
     - dim_origin_region

   host_dims:
     - dim_new_host
     - dim_origin_region

_table: bookings__china_dims_

columns:
- bookings
- nights
- dim_destination_china
- dim_origin_region

_table: bookings__host_dims_

columns:
- bookings
- nights
- dim_new_host
- dim_origin_region

Global Metrics Framework = Denormalization Machine

Super powerful for ad hoc analysis
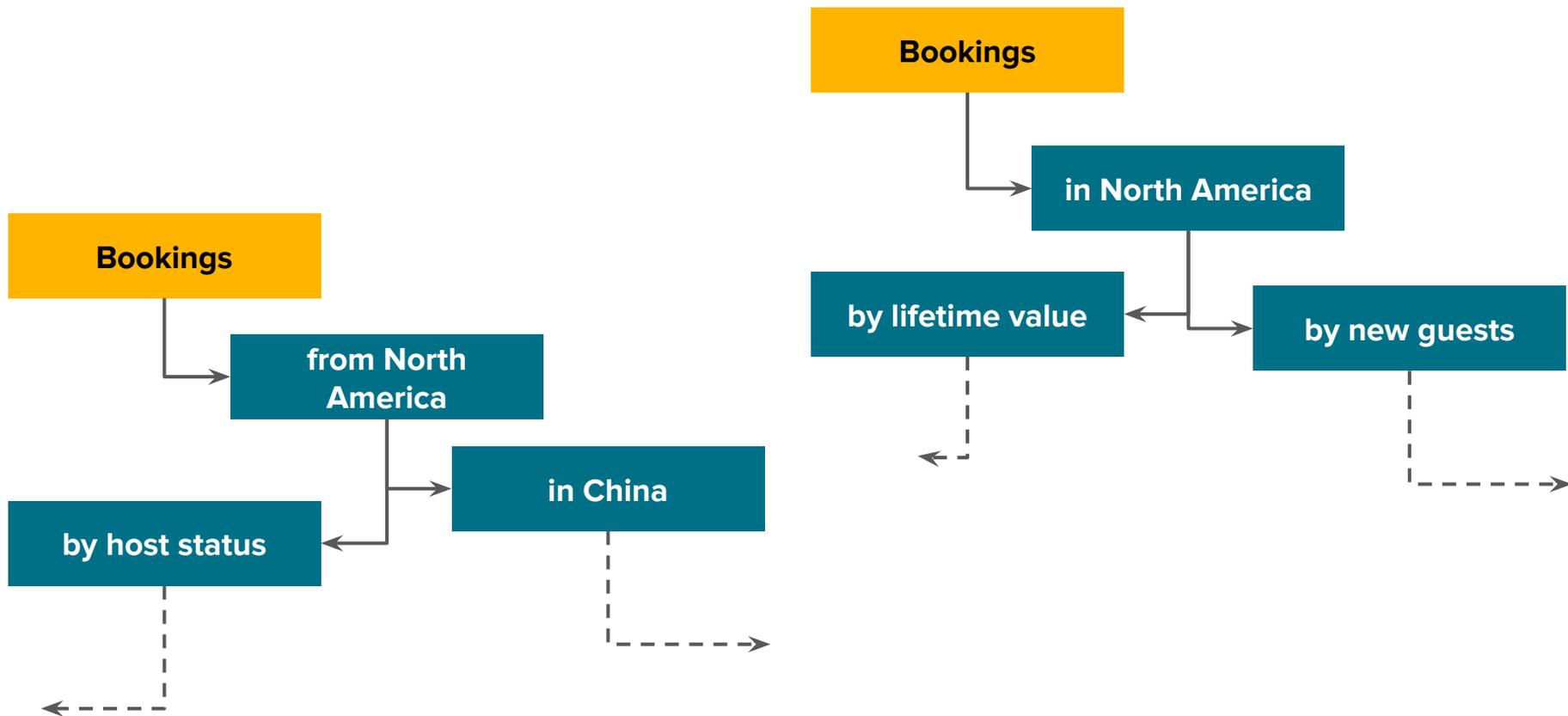
Config-driven pipeline generation eliminates 3 steps
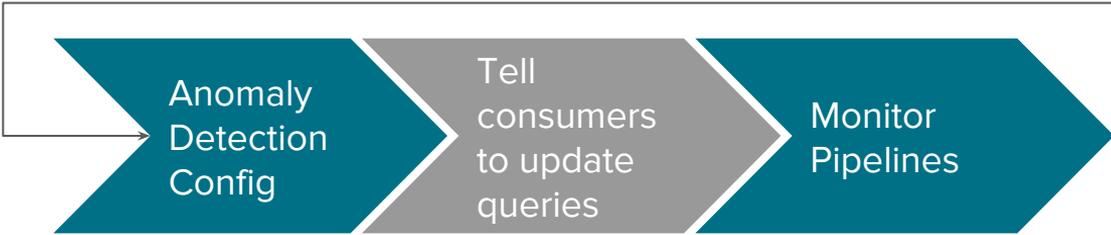
Serving data to apps eliminates 3 more steps
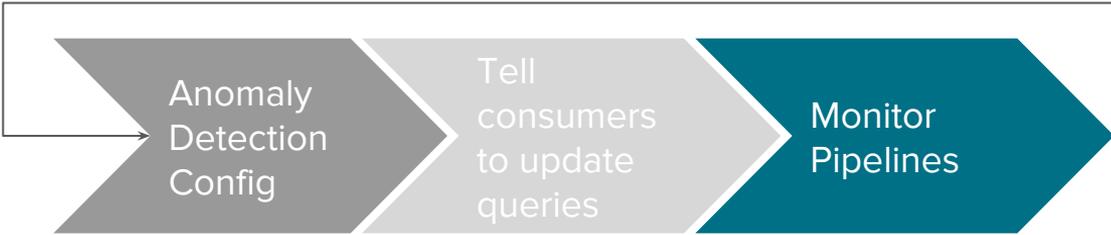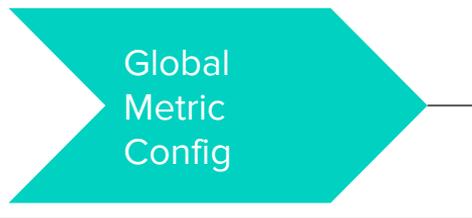
Global Metric Config

Create Table

Schedule Airflow Job

Backfill Table

A/B Testing Config

Add table to Superset

Create Dashboard

Anomaly Detection Config

Tell consumers to update queries

Monitor Pipelines

Free stuff

Automatic backfills when metrics or dimensions change

Free stuff

Automatic backfills when metrics or dimensions change

Self-healing when days are missed

Free stuff

**Automatic backfills when metrics or dimensions change**

Self-healing when days are missed

Dashboard generation script

Bonus features eliminate 2 more steps

Global Metric Config

Create Table → Schedule Airflow Job → Backfill Table → A/B Testing Config → Add table to Superset → **Create Dashboard

Anomaly Detection Config → Tell consumers to update queries → Monitor Pipelines
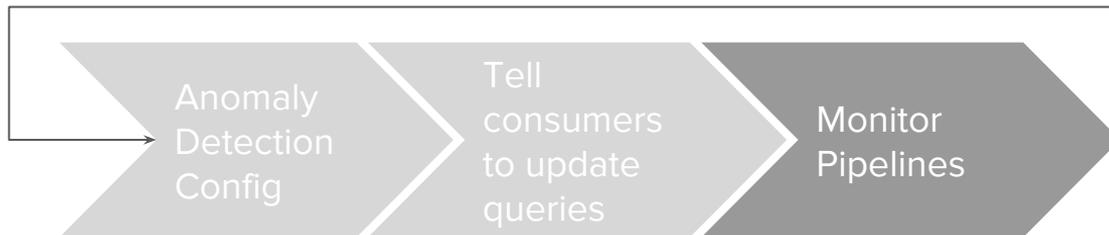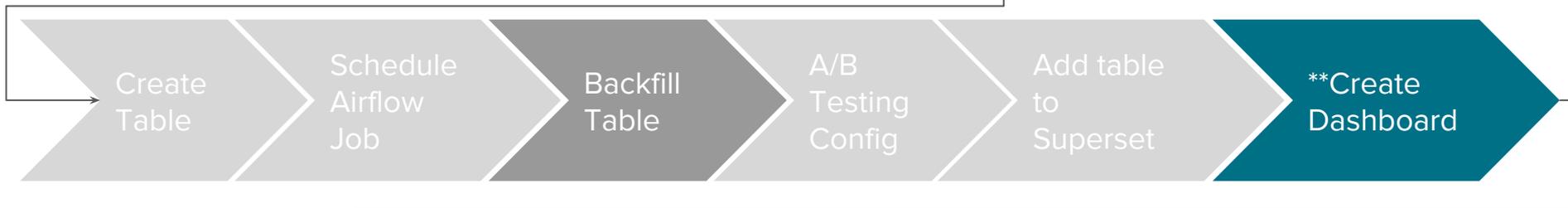
Old Data Science metric workflow took >2 weeks for simple changes

Create Table → Schedule Airflow Job → Backfill Table → A/B Testing Config → Add table to Superset → Create Dashboard → Anomaly Detection Config → Tell consumers to update queries → Monitor Pipelines
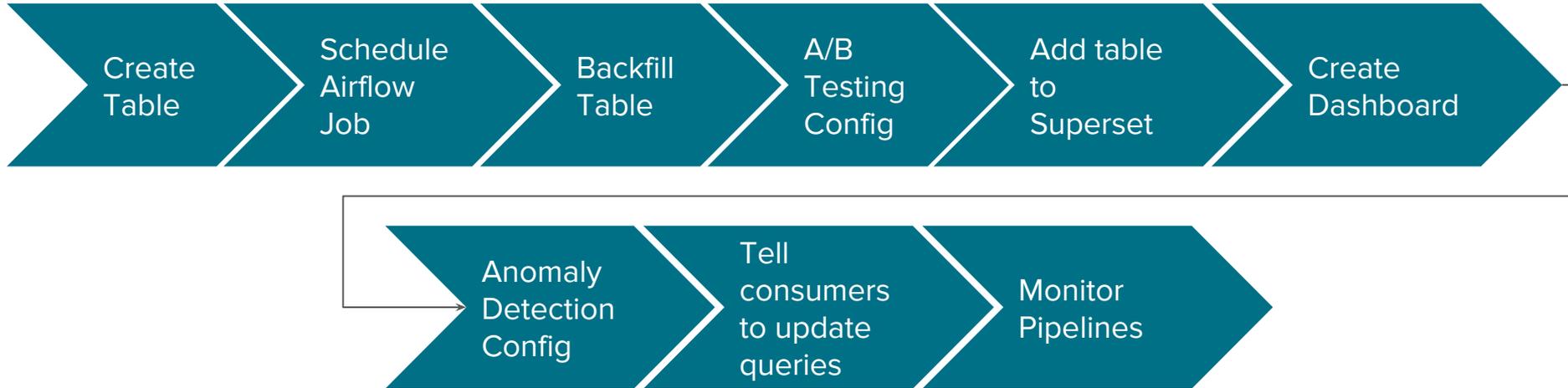
New Data Science metric workflow takes <2 days

Global Metric Config → **Create Dashboard

**semi-automated

Focusing on producers drives love

"It has dramatically reduced time to insight."

Focusing on producers drives love

"It has dramatically reduced time to insight."

"In our current world, even simple changes are painful. With Global Metrics, most of it becomes trivial."

Focusing on producers drives love

"It has dramatically reduced time to insight."

"In our current world, even simple changes are painful. With Global Metrics, most of it becomes trivial."

"You can put me in the satisfied customer quotes!"

At the time of official launch (last week)

Word-of-mouth adoption

- >20 teams contributing

- > 350 metrics added

- Less-technical contributors (Finance)

Changing the metric management paradigm

1. Airbnb's journey

2. Why you should make dimensions first class citizens

3. Why prioritize bonus features early

Where to go from here?

More features for metric consumers

Superset

Leverage metadata in
Superset integration

More features for metric consumers

Leverage metadata in Superset integration

Make metrics more discoverable

More features for metric consumers

Leverage metadata in Superset integration

Make metrics more discoverable

Metric certification process

Questions?

Twitter: @lchircus
LinkedIn: linkedin.com/in/lchircus
Email: lauren.chircus@airbnb.com