07/20/22

Generate Data Value Faster using ELT

Google Cloud

Jason Davenport
Tech Lead,
Google Cloud

Closing the Data Value Gap

Data

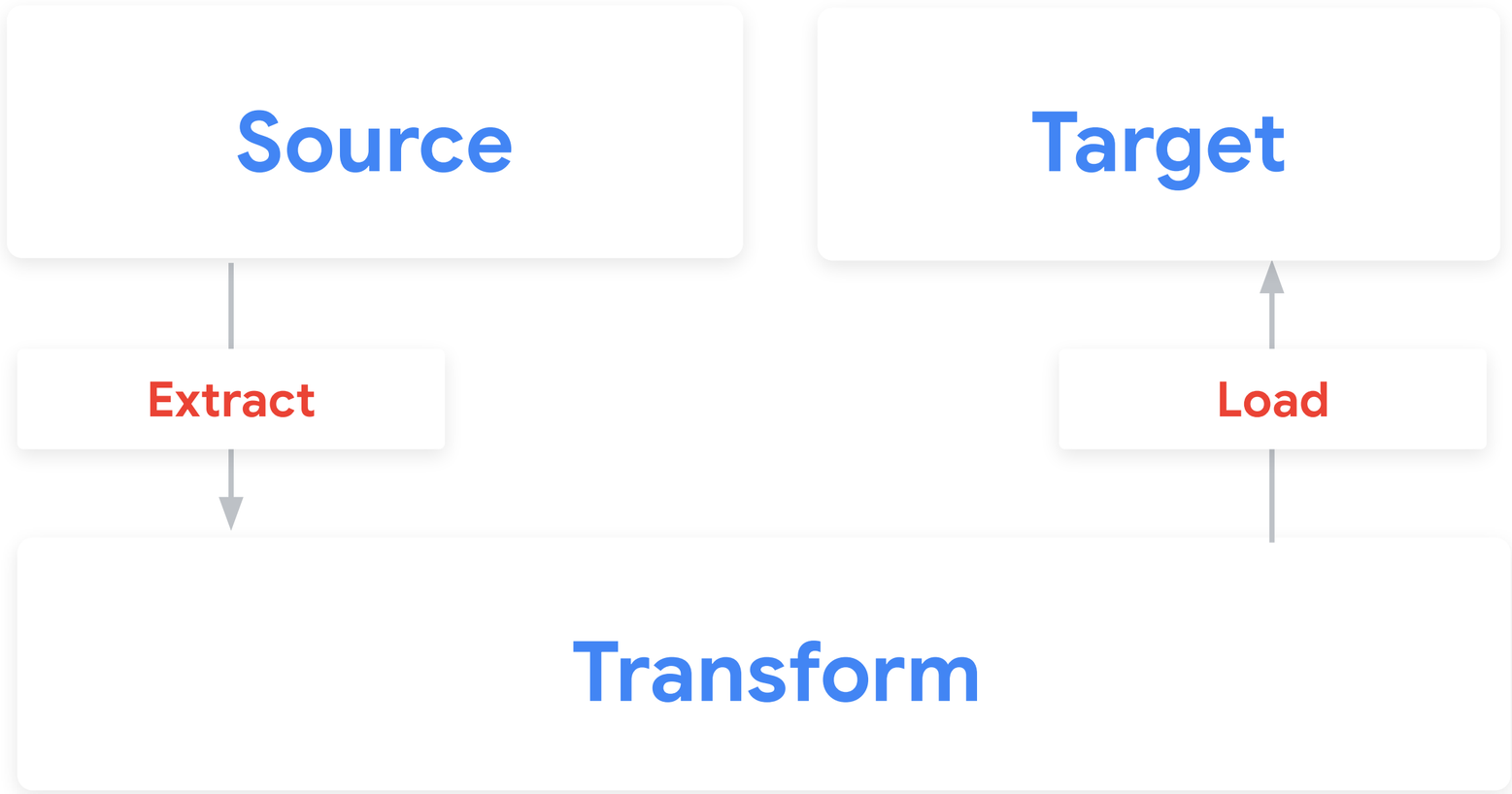
68%

of companies are **unable to realize tangible & measurable Value from Data.**

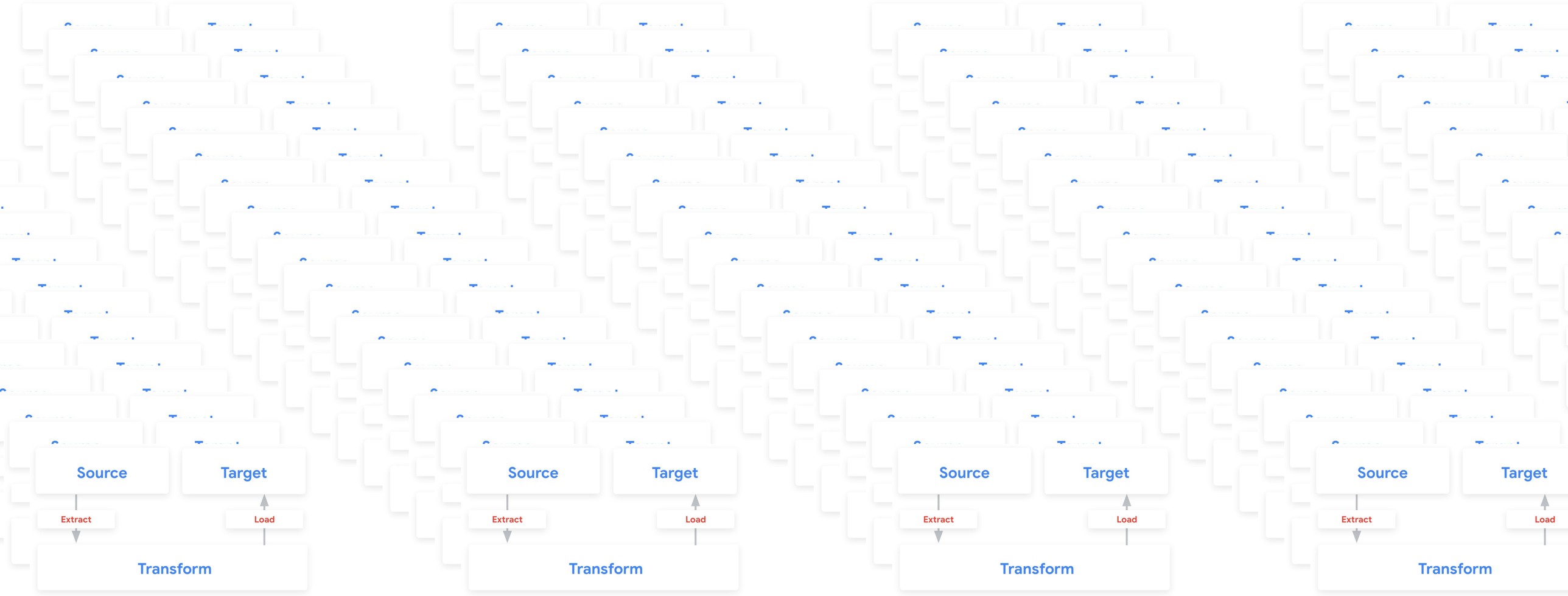
Value

We need to re-imagine how we effectively perform data engineering at scale.

Starting with ETL

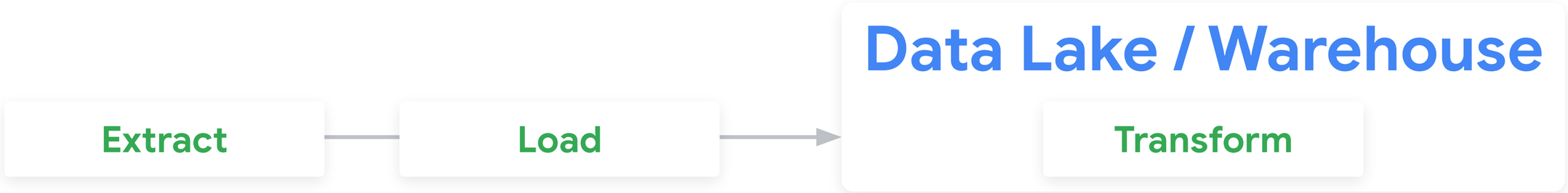


Starting with ETL



In reality, this pattern runs 100’s or 1000’s of times, and doesn’t leverage the power of purpose built platforms

Shifting to ELT



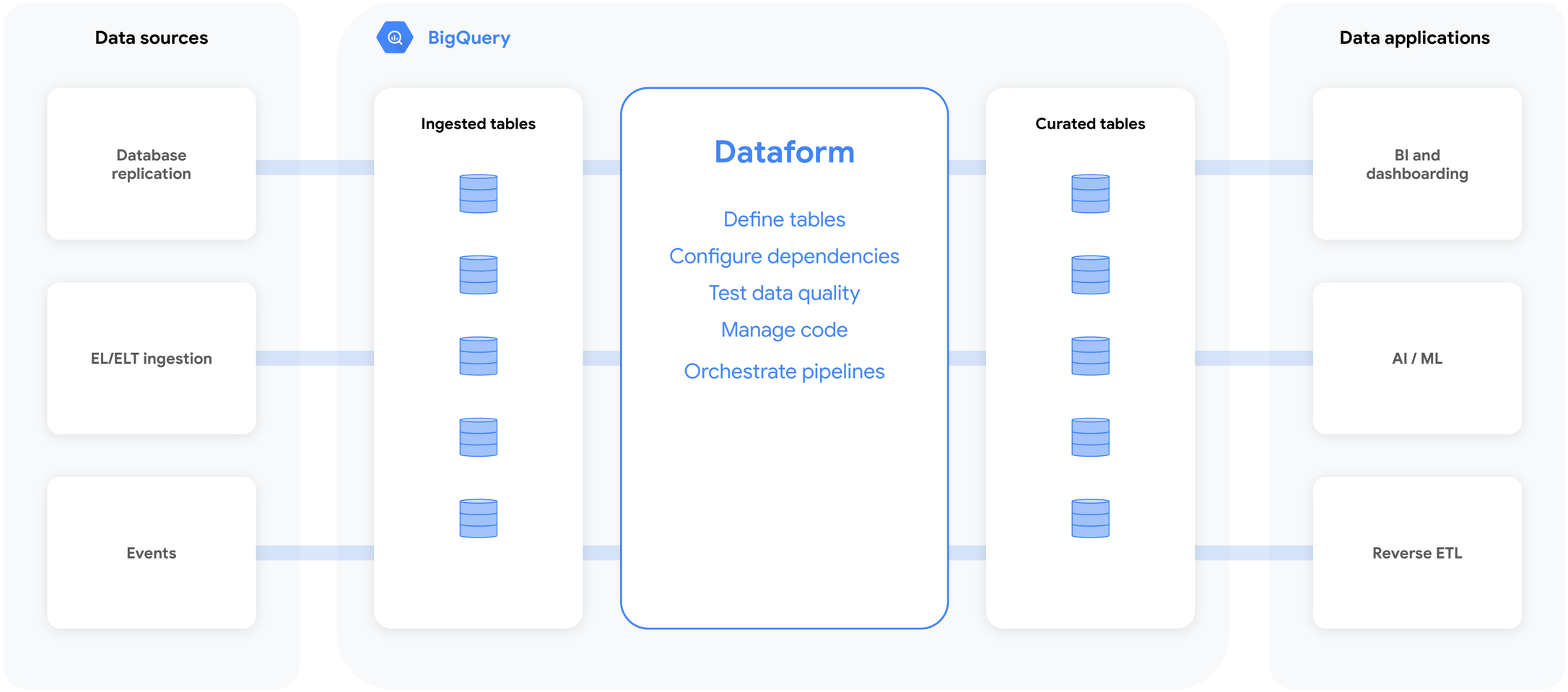
Our data is extracted and loaded once, and transforms get performed on the best platform for the workload

How can we create value building Transforms quickly and effectively?

- Low Code
- Automation
- Data Quality
- Collaboration
- Documentation

DataOps

Dataform helps analytics teams manage data in BigQuery using ELT with SQL

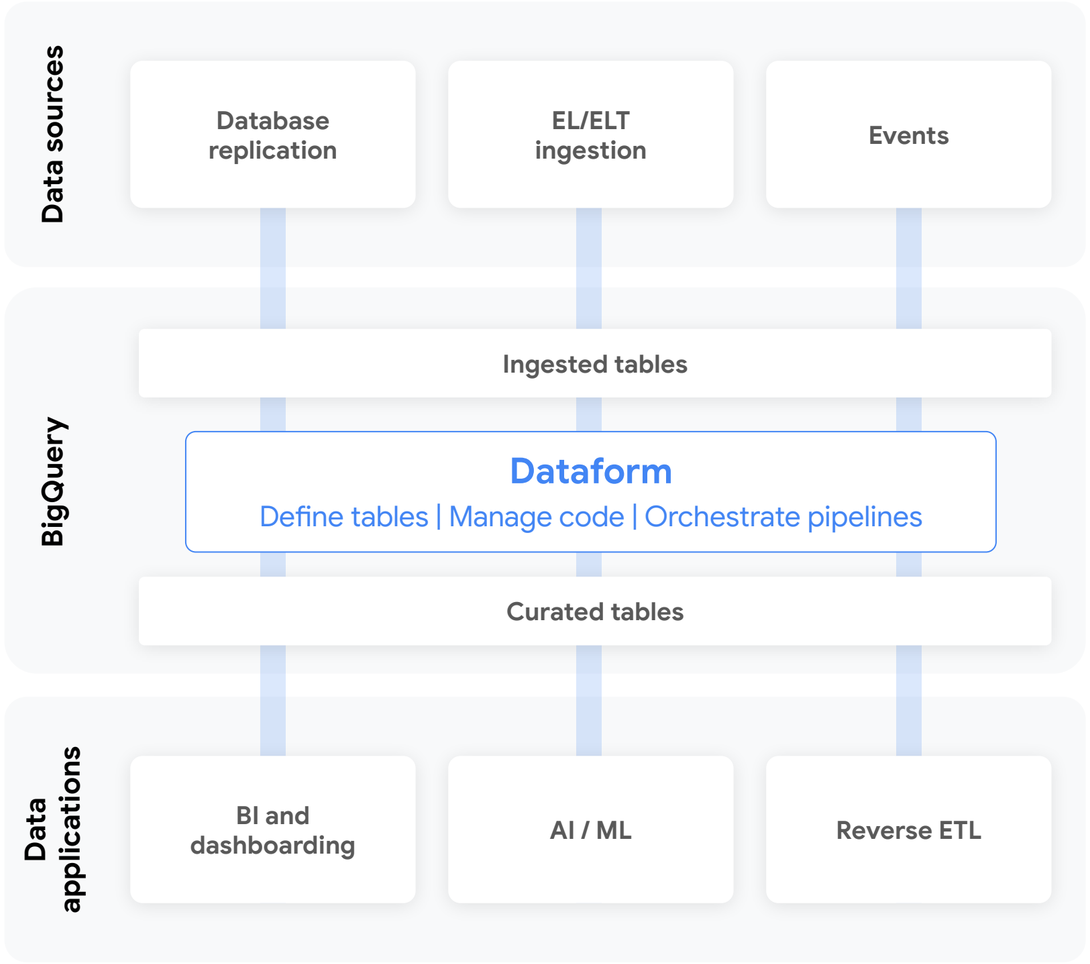


Develop SQL pipelines without the complexity

Build scalable data transformation pipelines in BigQuery using SQL from a single environment and without additional dependencies.

Help teams collaborate following software development best practices including version control, environments, CI/CD, testing, and inline documentation.

Empower data analysts to build production-grade SQL pipelines to manage the data they need - without requiring data engineers.



An end-to-end data transformation experience in BigQuery

Open source, SQL-based language to configure tables. **Available now**

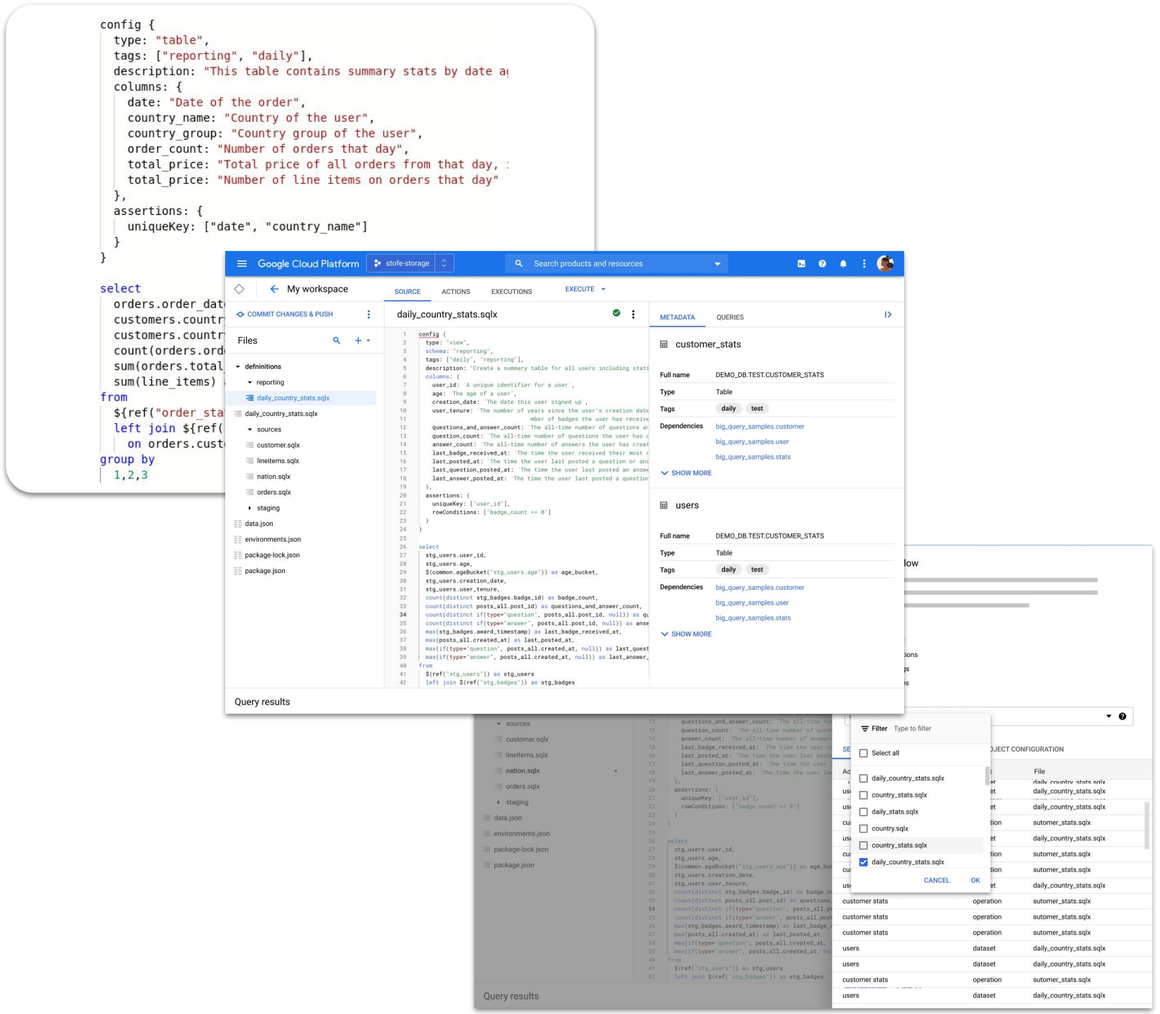
Define tables, configure dependencies, assert data quality, and document data tables

Web development environment integrated with Git. **Preview in Q3**

Develop SQL and SQLX, commit and push to Git, sync with GitHub and GitLab, visualise dependencies

Fully managed, serverless orchestration for data pipelines embedded in GCP. **Preview in Q3**

Run SQL workflows manually or via API, view logs, configure execution environments, schedule pipelines, and get alerts



Use simple configuration to enable multiple development priorities

Configuration-as-code

Column descriptions

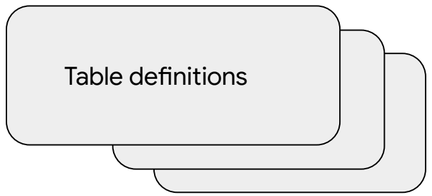
Data quality tests

Dependency management (`ref`)

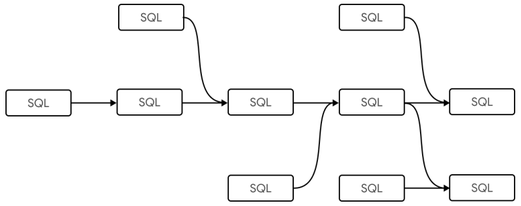
```
config {
  type: "table",
  tags: ["reporting", "daily"],
  description: "This table contains summary stats by date a
  columns: {
    date: "Date of the order",
    country_name: "Country of the user",
    country_group: "Country group of the user",
    order_count: "Number of orders that day",
    total_price: "Total price of all orders from that day,
    total_price: "Number of line items on orders that day"
  },
  assertions: {
    uniqueKey: ["date", "country_name"]
  }
}

select
  orders.order_date as date,
  customers.country_name as country_name,
  customers.country_group as country_group,
  count(orders.order_key) as order_count,
  sum(orders.total_price) as total_price,
  sum(line_items) as line_items
from
  ${ref("order_stats")} orders
  left join ${ref("customer_stats")} customers
    on orders.customer_key = customers.customer_key
group by
  1,2,3
```

Dataform then compiles your code into SQL statements to execute in BigQuery

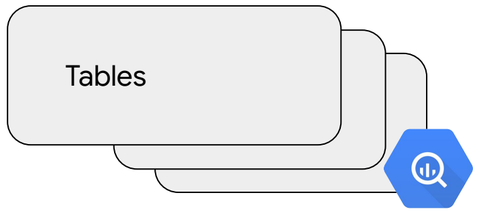
Develop table definitions with Dataform core and SQL



Dataform compiles into table definitions into SQL statements



Tables materialized in BigQuery



Go !

3 Ways to Get Started Today with Dataform CLI

Write Your First Pipeline

Test Your Data Quality

Deploy your Code to Production

Example 1: Let’s Build a Pipeline

We will:

1. Create a new set of sql objects
2. Check those objects into version control

In order to:

- Have a pipeline that can be executed to create value

Example 2: Test our data quality

We will:

1. Add standard data quality tests to our objects
2. Create our own quality tests

In order to:

- Ensure that our data is measured for quality during each pipeline run

Example 3: Deploy your pipelines to production

We will:

1. Create a deployment process
2. Deploy our code using our deployment process

In order to:

- Run our developed code in our environment

A bonus peek at our Preview launch

... coming soon to Google Cloud customers in Q3!

Get Started Today!

... at github.com/dataform-co/dataform

Thank you.

Google Cloud