Today's Speakers

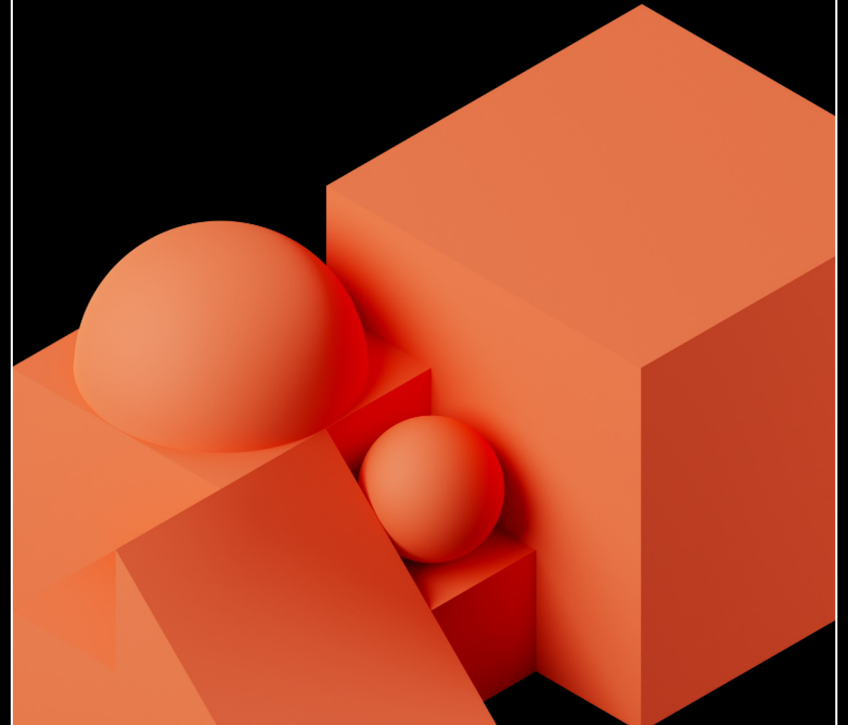**JOE WALSH**
Director, Measurement & Attribution

DEREK SUGDEN
Ad Measurement Lead, Measurement & Attribution

Effectv's Custom Ad Solutions Help Businesses Reach Potential Customers

effectv
A COMCAST COMPANY

Our Approach

Effectv is an
audience delivery company that combines
the **best of digital** with the **power of TV**

Find
We identify the right audiences for an advertiser's business and use data insights to find matching households

Target
We plan effective campaigns to reach audiences during any content they watch

Report
We show transparent, easy-to-read performance results that help optimize future campaigns

Prove
We correlate campaigns to business results, so advertisers can be confident their campaign is working

Our Teams' Focus...

From Impressions to Attribution

Impressions

- Household level data
- Number of exposures
- Timestamp of exposure
- Dimensions: Product, Geography, Network, Creative, etc.

Conversions

- Household level data
- Type of conversion
- Timestamp of conversion
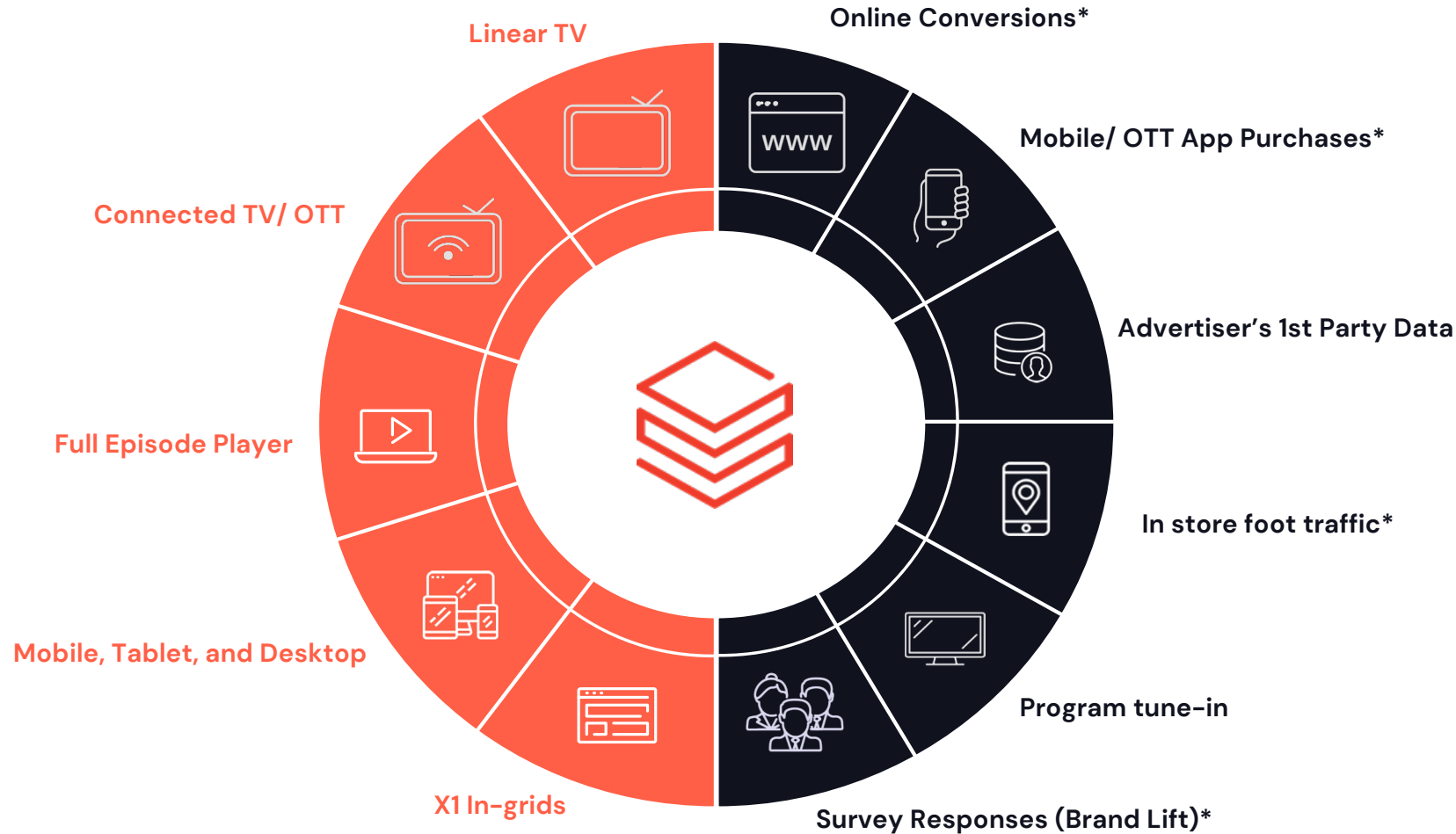- Dimensions (data source specific)

Attribution

- Multi-touch modeling/ customer journey
- Total attributed conversions for a given lookback window
- Number of converting households
- Conversion rates
- Lift metrics
- Cost per acquisition
- ROI

Tying Ad Exposures to Business Related Outcomes

effectv
A COMCAST COMPANY

AD EXPOSURES

BUSINESS OUTCOMES

Linear TV

Connected TV/ OTT

Full Episode Player

Mobile, Tablet, and Desktop

X1 In-grids

Online Conversions*

Mobile/ OTT App Purchases*

Advertiser's 1st Party Data

In store foot traffic*

Program tune-in

Survey Responses (Brand Lift)*

*Datasets licensed from third parties.

Core Databricks Capabilities Used In Our Process

effectv
A COMCAST COMPANY

Scheduled Jobs

Several codependent workflows can operate in tandem to speed up processing and automate outputs

AUTOMATION

Auto-Scaling Clusters

Auto-scaling allows for the efficient processing of both large and small datasets while keeping costs at a minimum

SCALE

Confidential Cluster Policies

Customer data is handled with security and privacy in mind by establishing confidential cluster policies
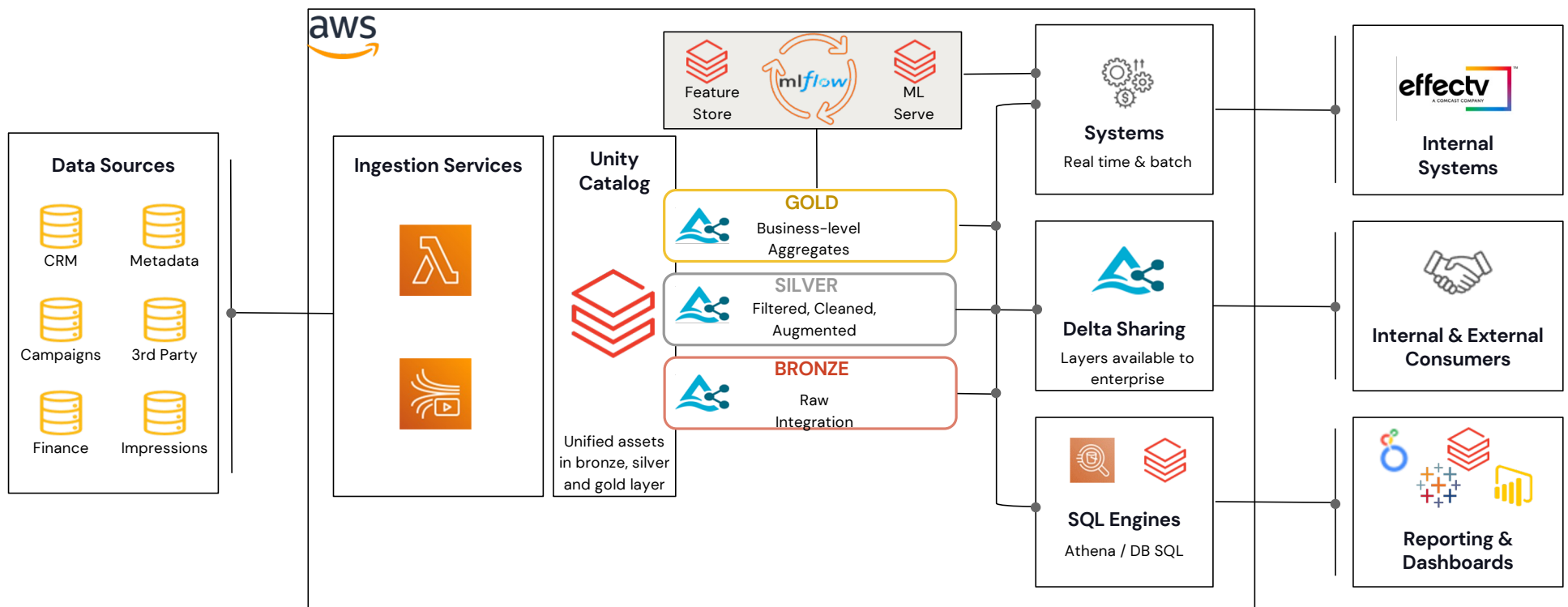
SECURITY

Simplified Data Stack

Use Unity Catalog and Delta Sharing to consolidate taxonomy, access controls, and the data itself into a single Data Stack with common tooling to service all the needs of our business.
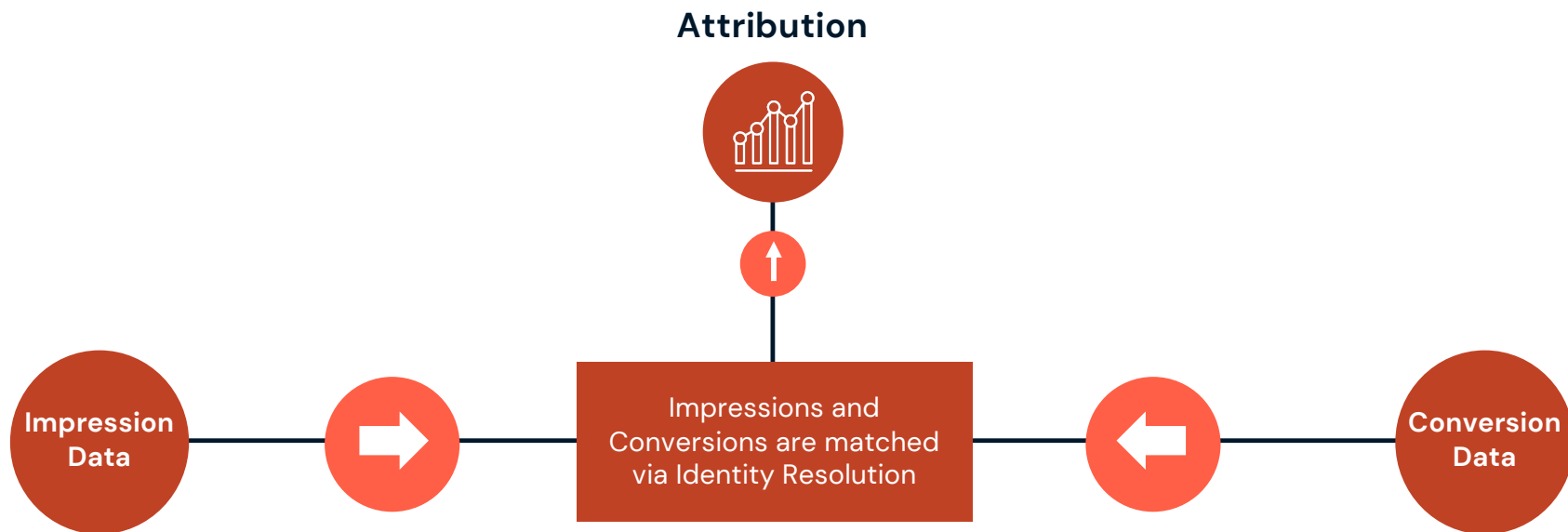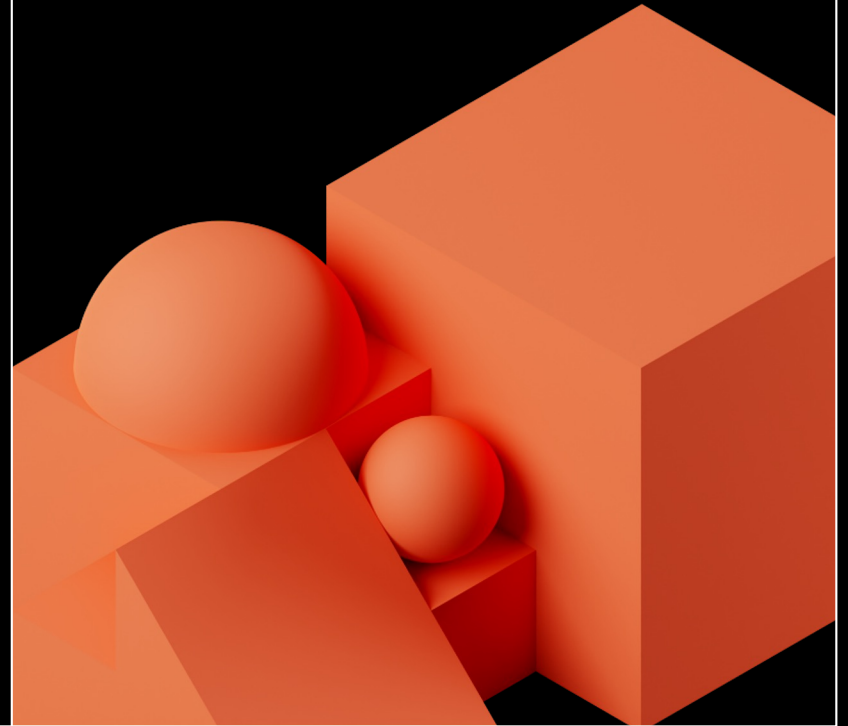
SIMPLIFICATION

Effectv Data Stack
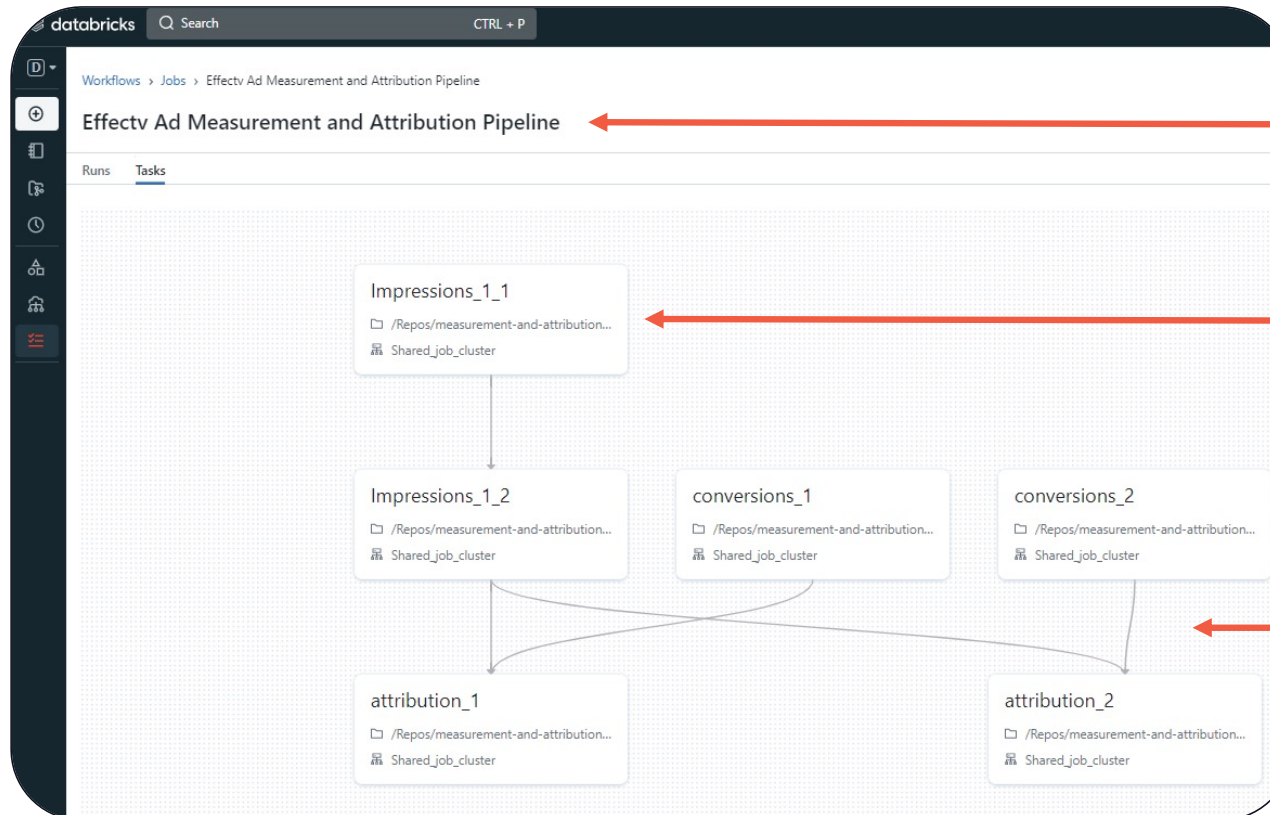
Enabling an analytics platform with Databricks

Data Sources
- CRM
- Metadata
- Campaigns
- 3rd Party
- Finance
- Impressions

Ingestion Services

Unity Catalog

Unified assets in bronze, silver and gold layer

Feature Store

mlflow

ML Serve

GOLD — Business-level Aggregates

SILVER — Filtered, Cleaned, Augmented

BRONZE — Raw Integration

Systems — Real time & batch

Delta Sharing — Layers available to enterprise

SQL Engines — Athena / DB SQL

Internal Systems

Internal & External Consumers

Reporting & Dashboards

Dataflow

An Overview of Our Attribution Pipeline

Attribution

Impression Data

Impressions and Conversions are matched via Identity Resolution

Conversion Data

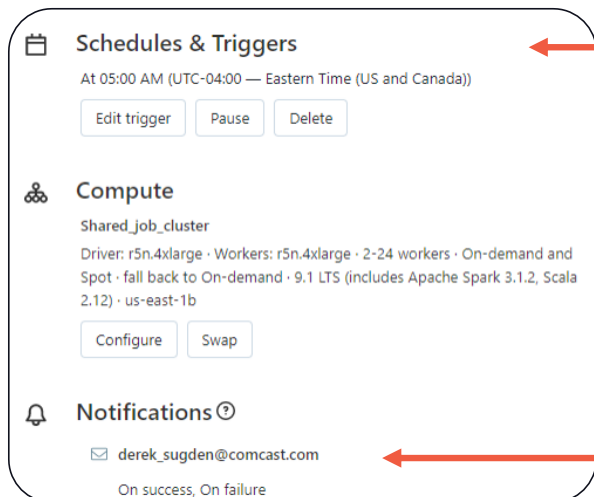# Our Attribution Pipeline in Databricks

Multiple Tasks Can Run Under a Single Job

Each Task Represents a Different Notebook Each with Several Commands

Tasks Can be Set to Run Independent or Dependent on Previous Tasks

Simplifying Automation With Scheduled Jobs and Auto-Scaling Clusters

Schedules & Triggers
At 05:00 AM (UTC-04:00 — Eastern Time (US and Canada))
Edit trigger   Pause   Delete

Compute
Shared_job_cluster
Driver: r5n.4xlarge · Workers: r5n.4xlarge · 2-24 workers · On-demand and Spot · fall back to On-demand · 9.1 LTS (includes Apache Spark 3.1.2, Scala 2.12) · us-east-1b
Configure   Swap

Notifications ⓘ
✉ derek_sugden@comcast.com
On success, On failure

Scheduling the Job Allows for Concurrent Processes to Run and by Scheduling Major Computes Outside of Peak Usage Hours We Can Ensure Enough Resources are Available for Job Needs

We Can Configure and Optimize a Cluster to Save on Costs While Allowing for Additional Resources When Needed

We Can Decide Who Receives Notification Upon Job Failure or Completion
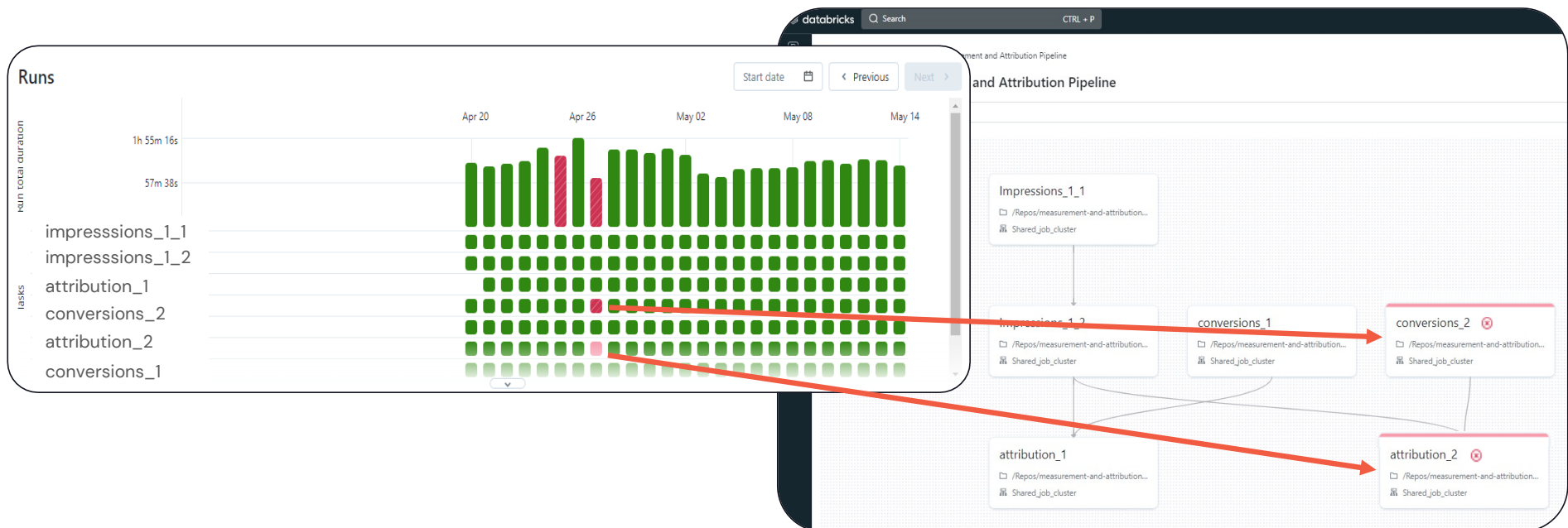
Monitor and Adjust Workflows As Needed

By tracking the metrics for a job run, we've been able to fine-tune our notebooks and cluster policy to allow auto-scaling to work at its best.
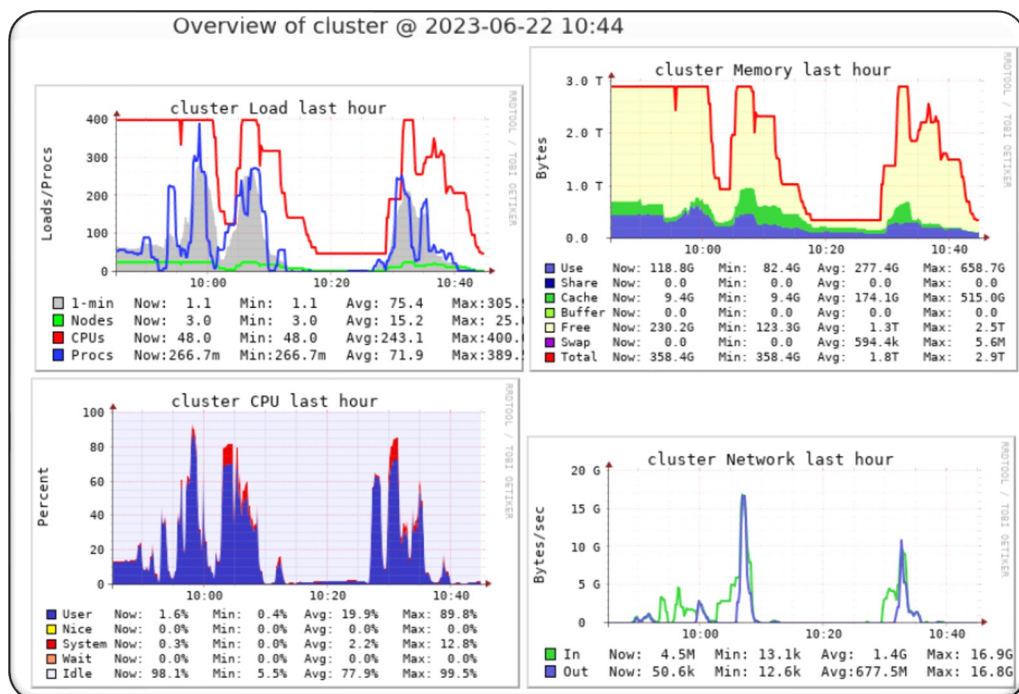
Realtime Feedback and Performance Over-Time

By monitoring when and where jobs fail, we can quickly diagnose and fix pipeline issues to prevent future job failures.

The Use of Confidential Cluster Policies

aws

STEP 1: On AWS IAM Role

- We created a new IAM role with limited or otherwise confidential access to sensitive datasets. Privileges to use this role is limited to select users.
- An instance profile is created off of this role.

databricks

STEP 2: On Databricks

- Add the instance profile created above in Databricks > Admin > Instance Profile.
- Create a new cluster via a previously established cluster policy linked to the instance profile.
- Limit access to the cluster to specific users via defining confidential user group.
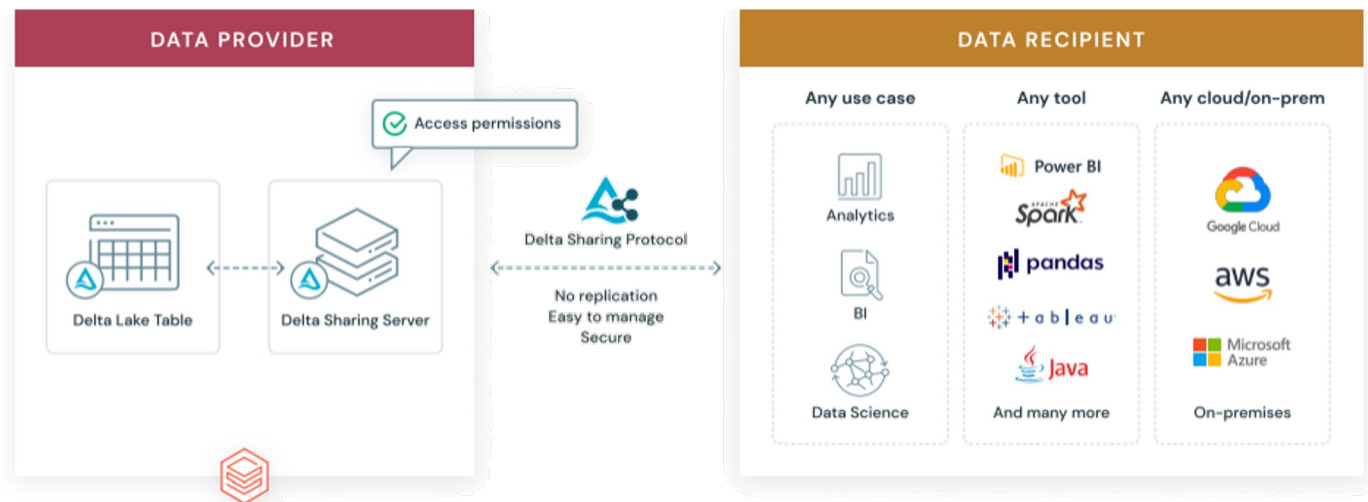
These steps ensure

1. Locked Down Access in AWS

2. Locked Down Access to Processing of Data in Databricks

effectv
A COMCAST COMPANY

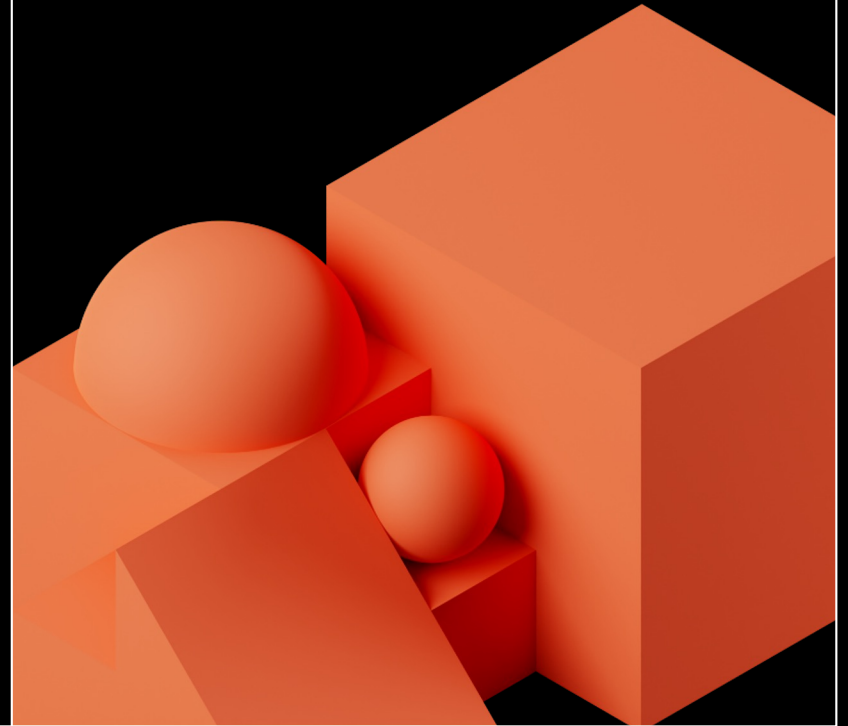# Delta Sharing Empowers Other Technical Teams

Legacy, Batch Copy
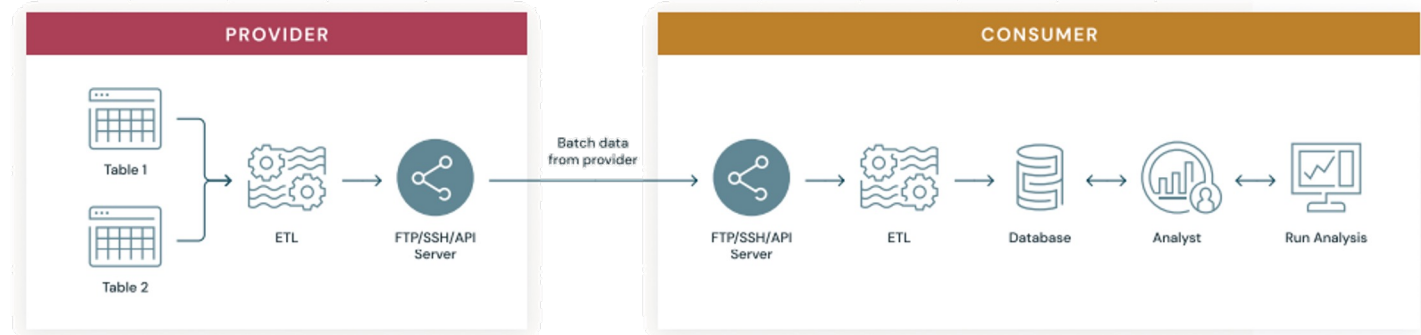
Improved, Managed Data Access with Delta Sharing

Results: Tying It All Together

Self–Serve Analytics
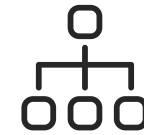
Integration With Reporting Tools (e.g. Looker)

- *Filter Functionality*

- *Campaign Level Reporting with Dimensional Breakouts*

- *Campaign Performance Over Time*

- *Incremental Lift Insights*

- *Multi–Tactic insights*

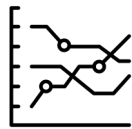# Why This Is So Important To Our Business...

Understand Behavior without Sacrificing Privacy

Simplify Cross Platform Reporting

Deeper Insights from Attribution

Speed to Market with Industry Leading Knowledge