DATA+AI SUMMIT 2022

Databricks Serverless

Nikhil Jethava

Staff Product Manager, Databricks

Matt Ryan

Director Engineering and Co-founder, Kythera Labs

Aaron Davidson

Principal Software Engineer, Databricks

ORGANIZED BY databricks

Product safe harbor statement

This information is provided to outline Databricks’ general product direction and is for informational purposes only. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all.

What is serverless and why?

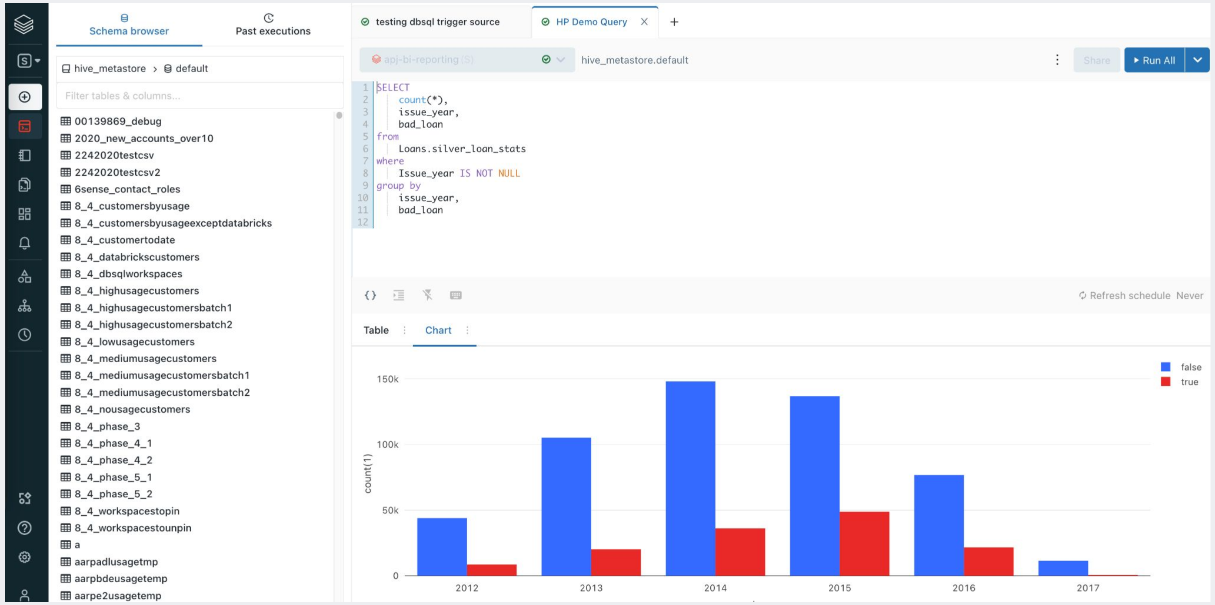
Databricks SQL

Best warehouse is Lakehouse

Databricks SQL provides an environment for:

- Running ad-hoc SQL queries
- Creating dashboards
- Connecting to external BI tools, e.g., Tableau and Power-BI

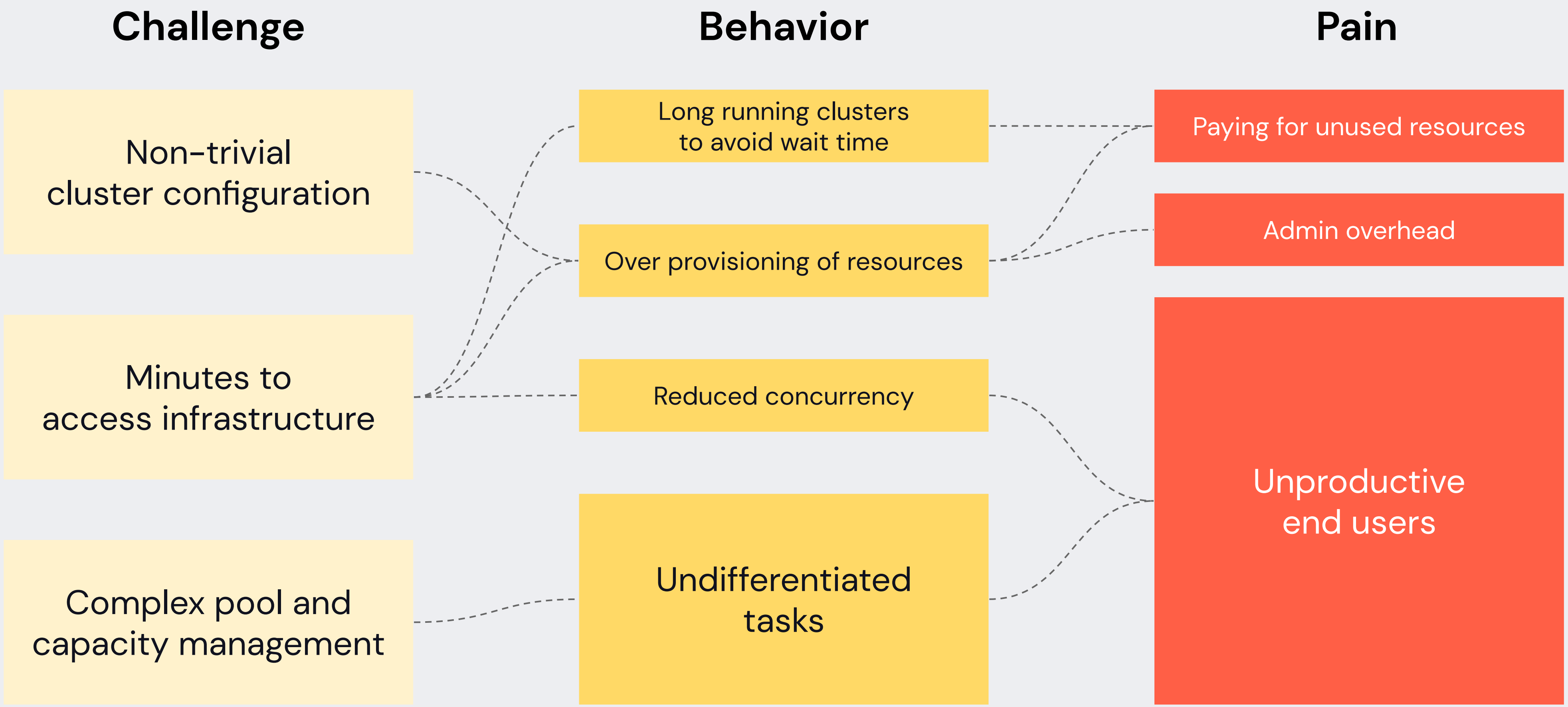
Customers create and manage the compute cluster needed to run the queries



- 00139869_debug
- 2020_new_accounts_over10
- 2242020testcsv
- 2242020testcsv2
- 6sense_contact_roles
- 8_4_customersbyusage
- 8_4_customersbyusageexceptdatabricks
- 8_4_customertodate
- 8_4_databrickscustomers
- 8_4_dbsqlworkspaces
- 8_4_highusagecustomers
- 8_4_highusagecustomersbatch1
- 8_4_highusagecustomersbatch2
- 8_4_lowusagecustomers
- 8_4_mediumusagecustomers
- 8_4_mediumusagecustomersbatch1
- 8_4_mediumusagecustomersbatch2
- 8_4_nousagecustomers
- 8_4_phase_3
- 8_4_phase_4_1
- 8_4_phase_4_2
- 8_4_phase_5_1
- 8_4_phase_5_2
- 8_4_workspacestopin
- 8_4_workspacestounpin
- a
- aarpadlusagetmp
- aarpbdeusagetemp
- aarpe2usagetemp

```sql
SELECT
    count(*),
    issue_year,
    bad_loan
from
    Loans.silver_loan_stats
where
    Issue_year IS NOT NULL
group by
    issue_year,
    bad_loan
```

Databricks Clusters (Classic)



Challenge

Non-trivial cluster configuration

Minutes to access infrastructure

Complex pool and capacity management

Behavior

Long running clusters to avoid wait time

Over provisioning of resources

Reduced concurrency

Undifferentiated tasks

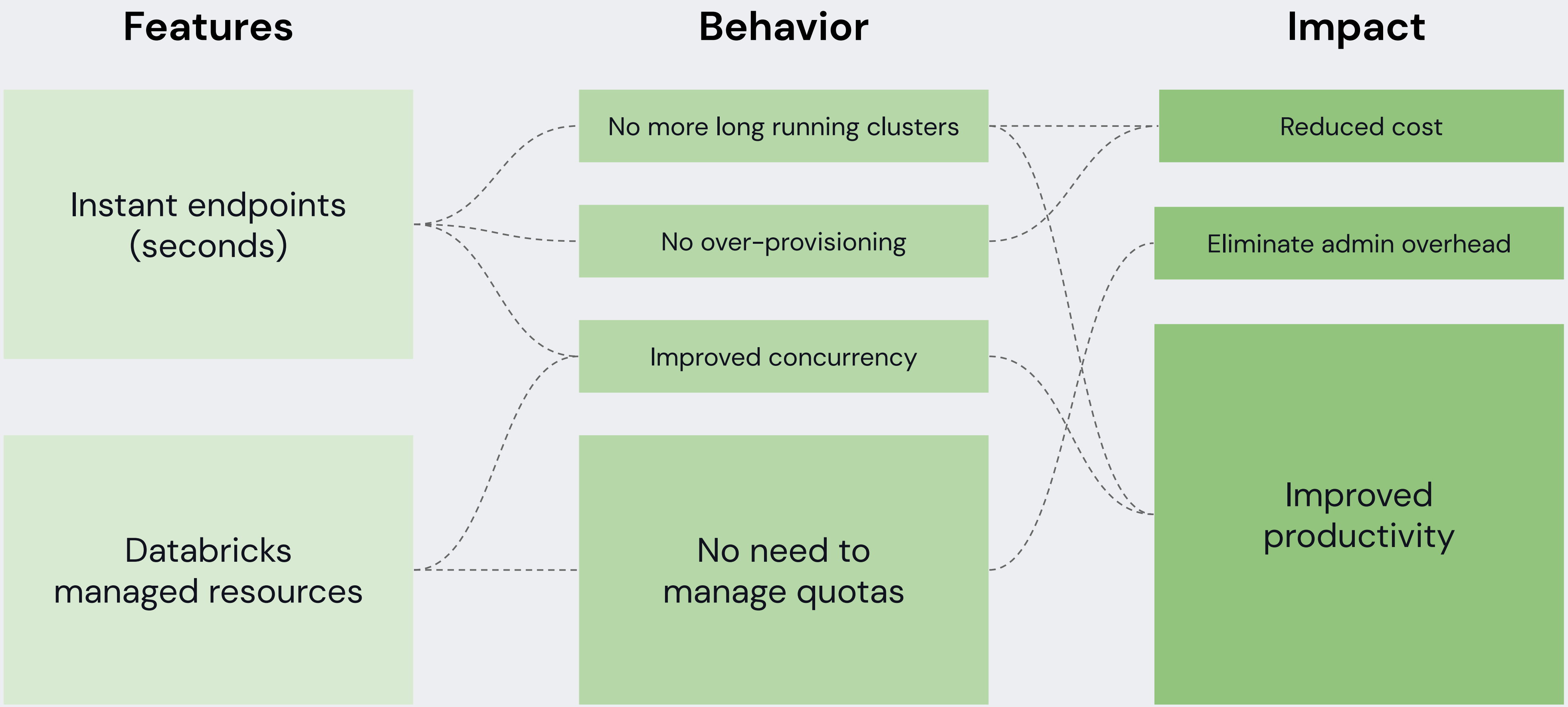
Pain

Paying for unused resources

Admin overhead

Unproductive end users

Benefits: Databricks SQL Serverless



Features

Instant endpoints (seconds)

Databricks managed resources

Behavior

No more long running clusters

No over-provisioning

Improved concurrency

No need to manage quotas

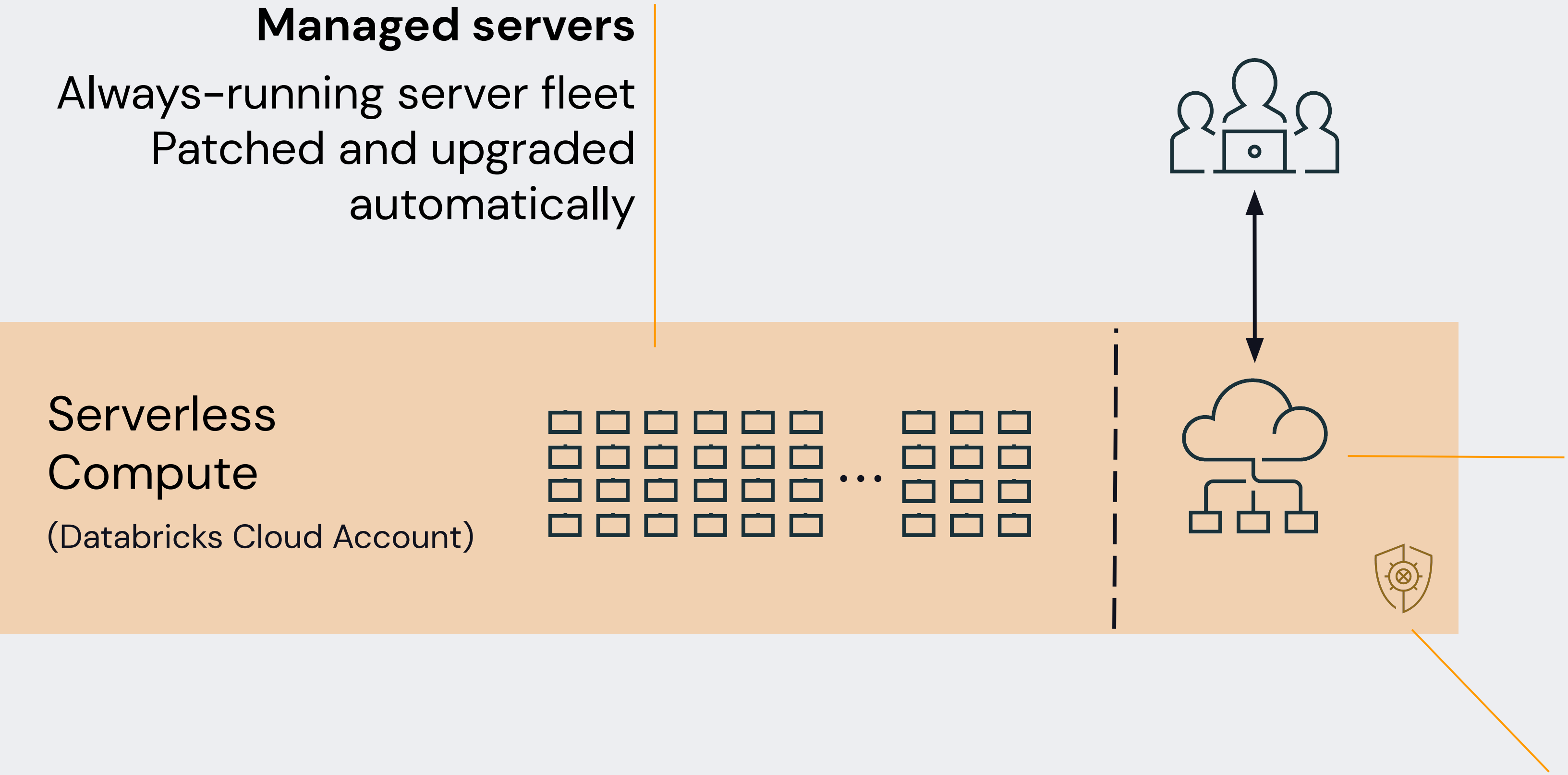
Impact

Reduced cost

Eliminate admin overhead

Improved productivity

Capabilities: Databricks SQL Serverless



Managed servers

Always-running server fleet
Patched and upgraded automatically

Serverless Compute

(Databricks Cloud Account)

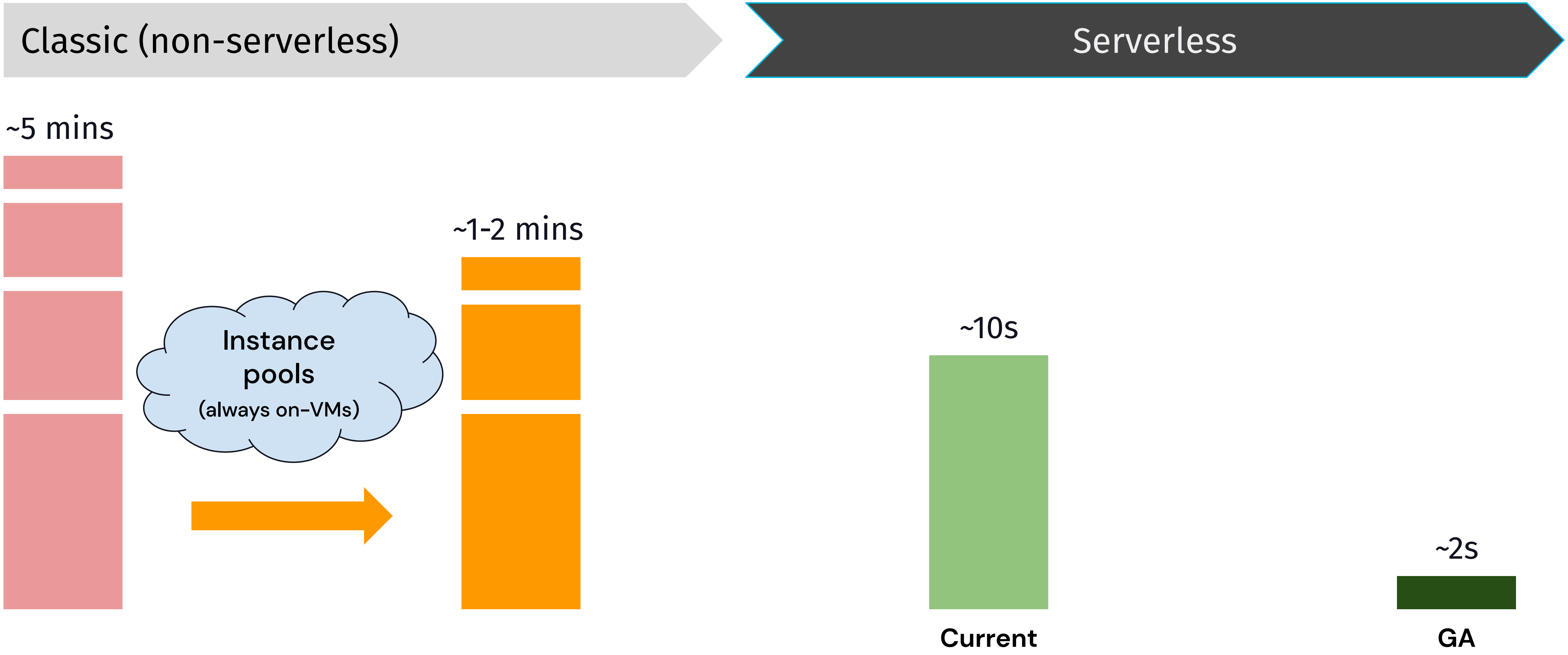
Elastic

Scale up and down automatically

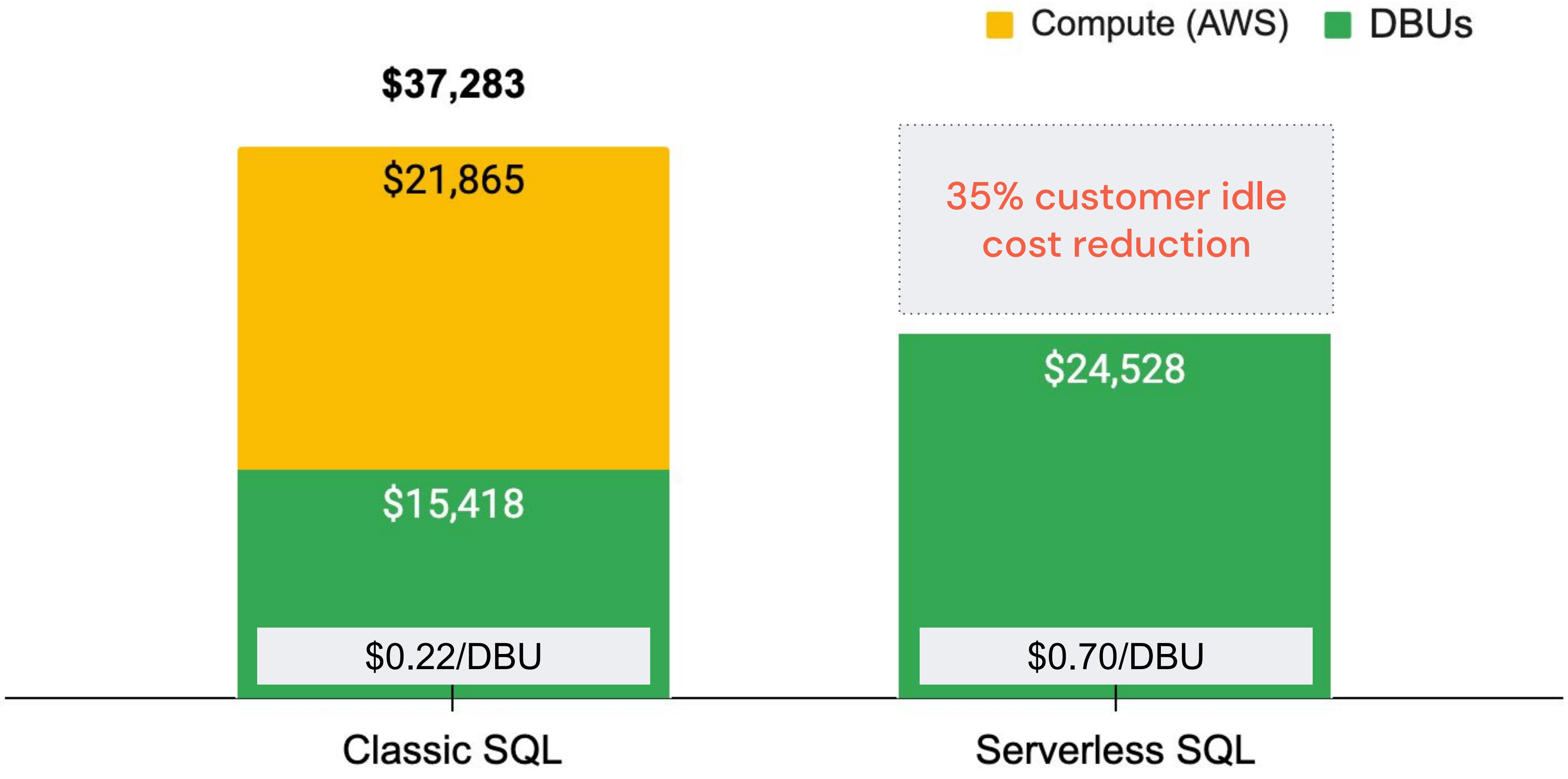
Secure

3-layer isolation with data encryption

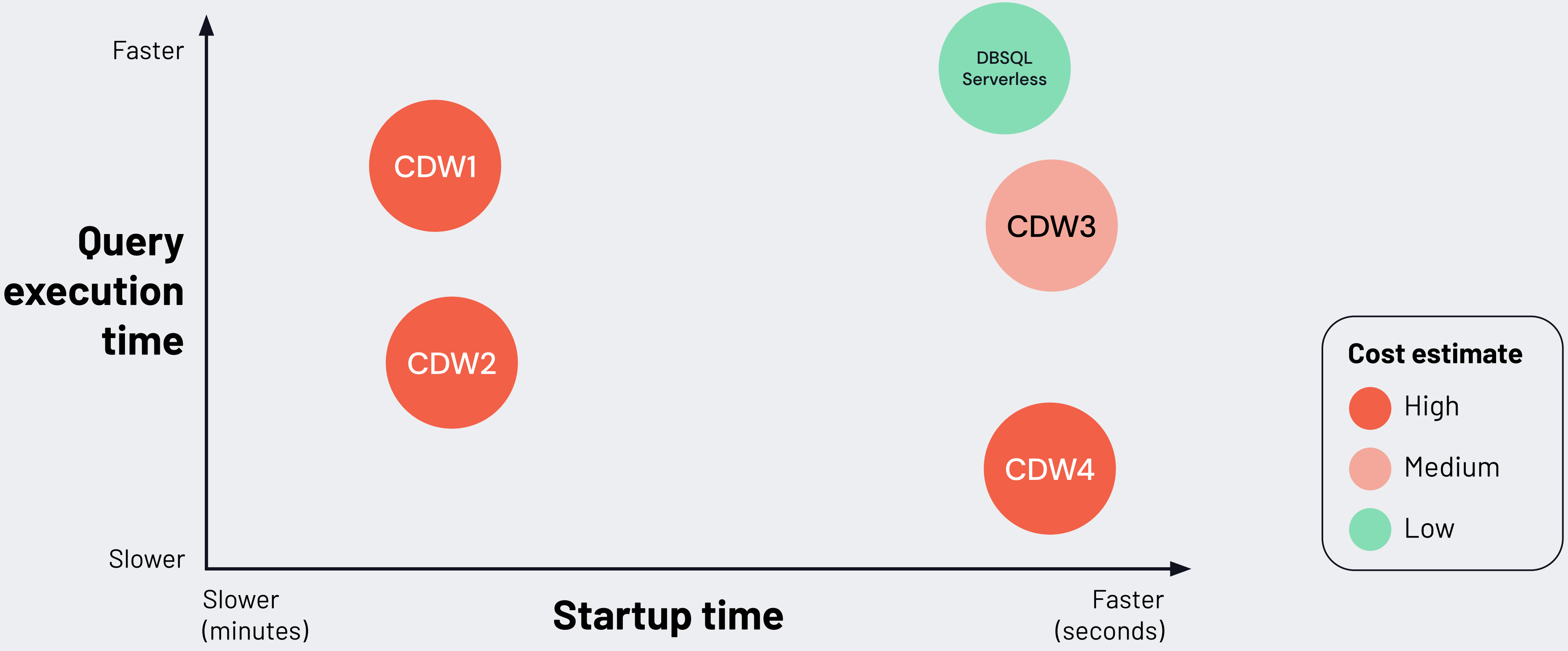
Fast 1st query performance… only getting better



Cost savings: Serverless platform helps reduce TCO by ~20-40%



Compete: Databricks Serverless SQL Serverless



Source: 2022 Cloud Data Warehouse Benchmark Report

How does it work?

Recap: goals of Serverless

✓

Improve reliability

- While reducing TCO
- While reducing management burden
- While maintaining (or improving) security posture

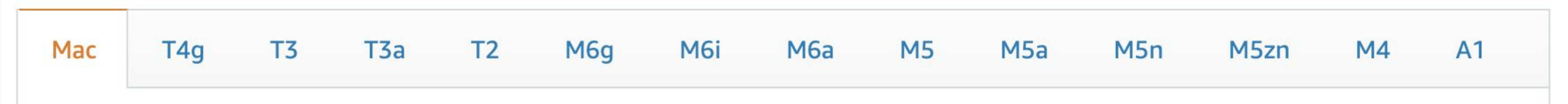
✓

Key mechanisms:

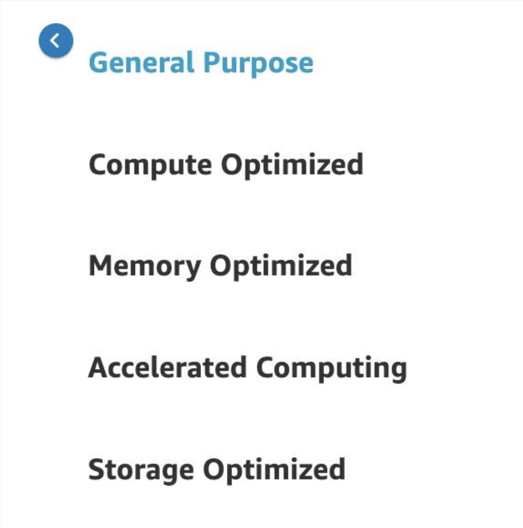
- Instance types
- Warm pooling
- Administrative simplicity

Instance types: classic

- Instance types critical for reliability, cost, and performance



Mac | T4g | T3 | T3a | T2 | M6g | M6i | M6a | M5 | M5a | M5n | M5zn | M4 | A1



General Purpose

Compute Optimized

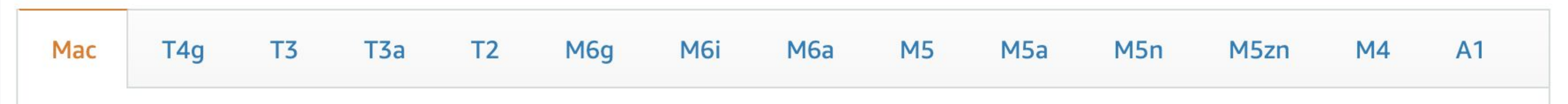
Memory Optimized

Accelerated Computing

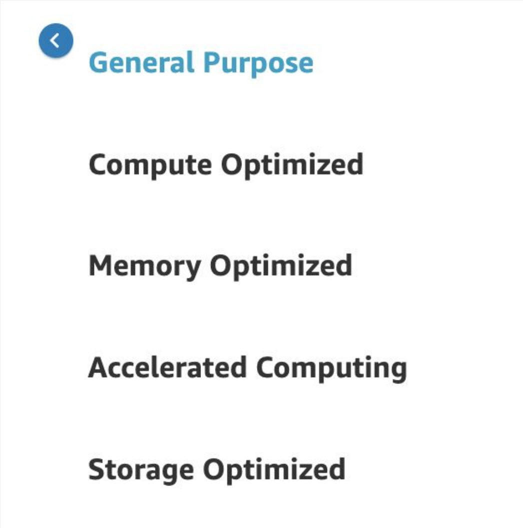
Storage Optimized

Instance types: classic

- Instance types critical for reliability, cost, and performance



- Picking the right one(s) for a workload is hard
- Keeping up to date with new instances is hard
- Databricks can help, but…
  - Capacity
  - Reservations

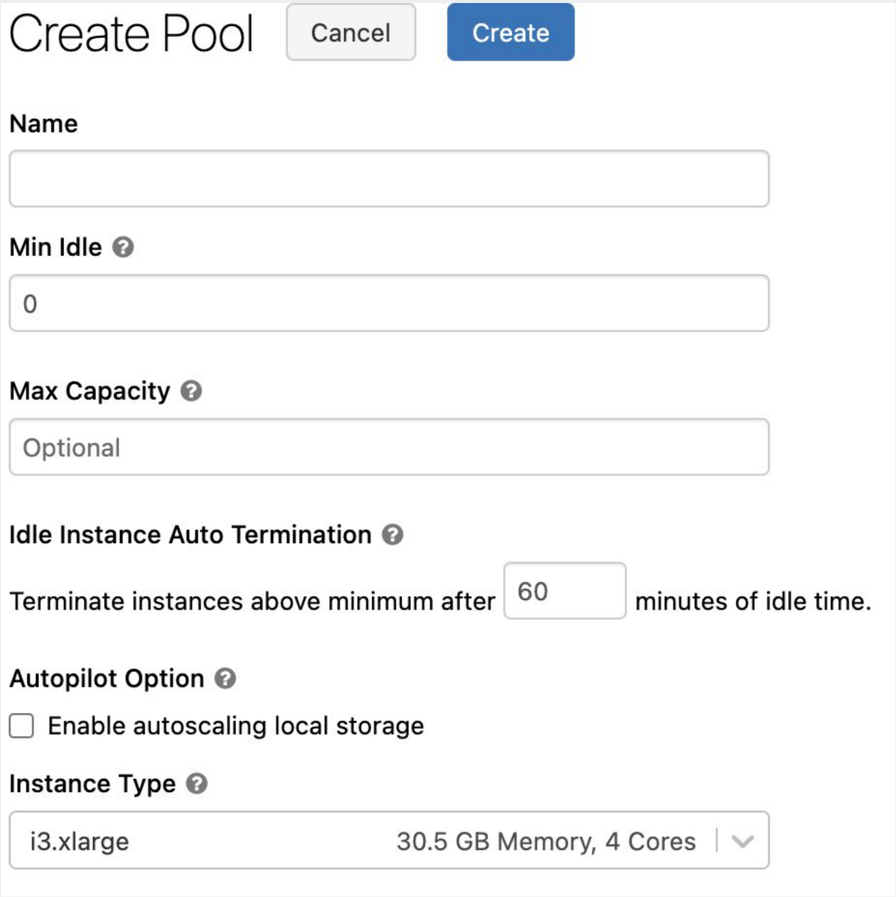


Instance types: serverless

- For SQL, we offer sizes (Small, Medium, Large, etc.)
- Leverage benchmarks to identify best cost/perf instances
- Manage capacity & reservations, enabling us to
  - Adopt new instances quickly as they become available
  - Leverage heterogeneous pool for availability
- Work with cloud vendors to understand regional/AZ capacity

Warm pooling: classic

- Databricks offers a pooling abstraction
  - Faster startup
  - More predictable
- But: you pay for this
- Management is complicated



Create Pool

Name

Min Idle: 0

Max Capacity: Optional

Idle Instance Auto Termination

Terminate instances above minimum after 60 minutes of idle time.

Autopilot Option

□ Enable autoscaling local storage

Instance Type: i3.xlarge 30.5 GB Memory, 4 Cores

Warm pooling: serverless

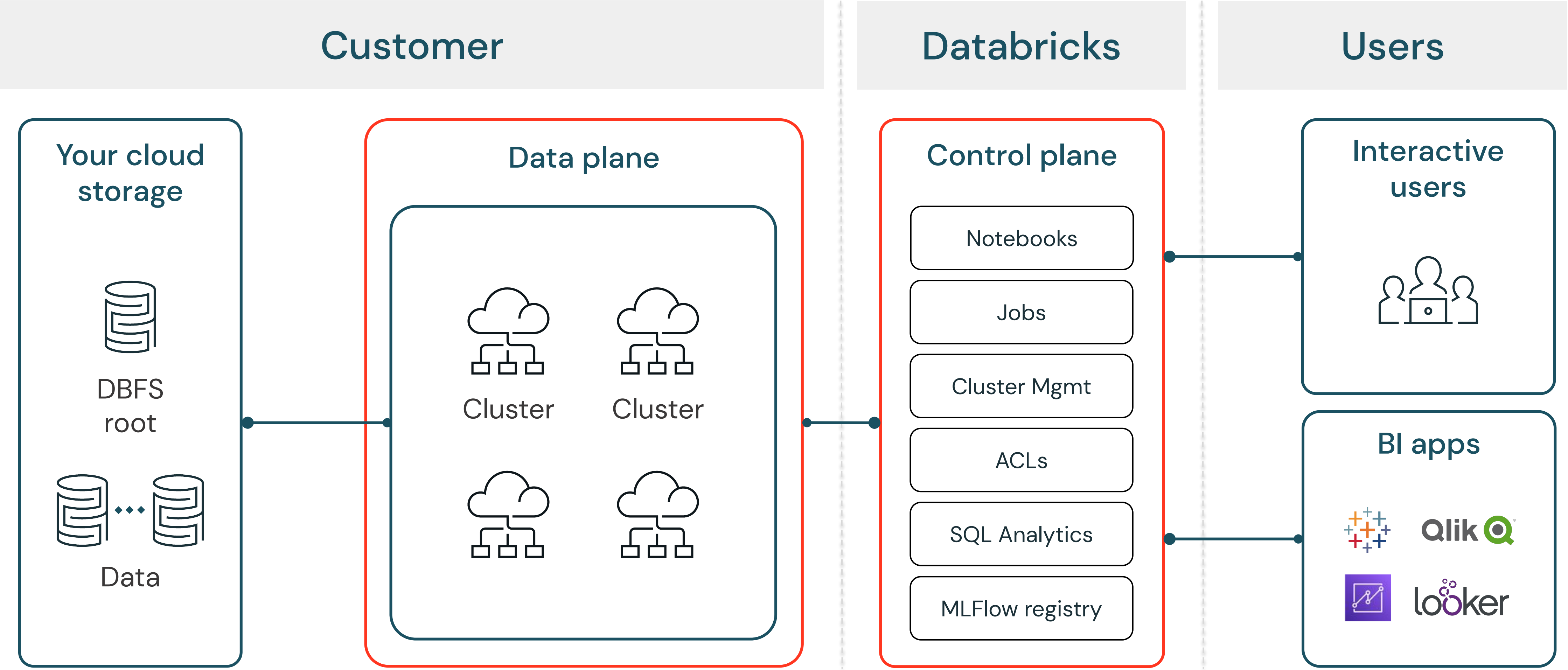
- Pool by default
- Amortized cost across customers
- Our responsibility to optimize pool

Administrative: Classic vs. Serverless

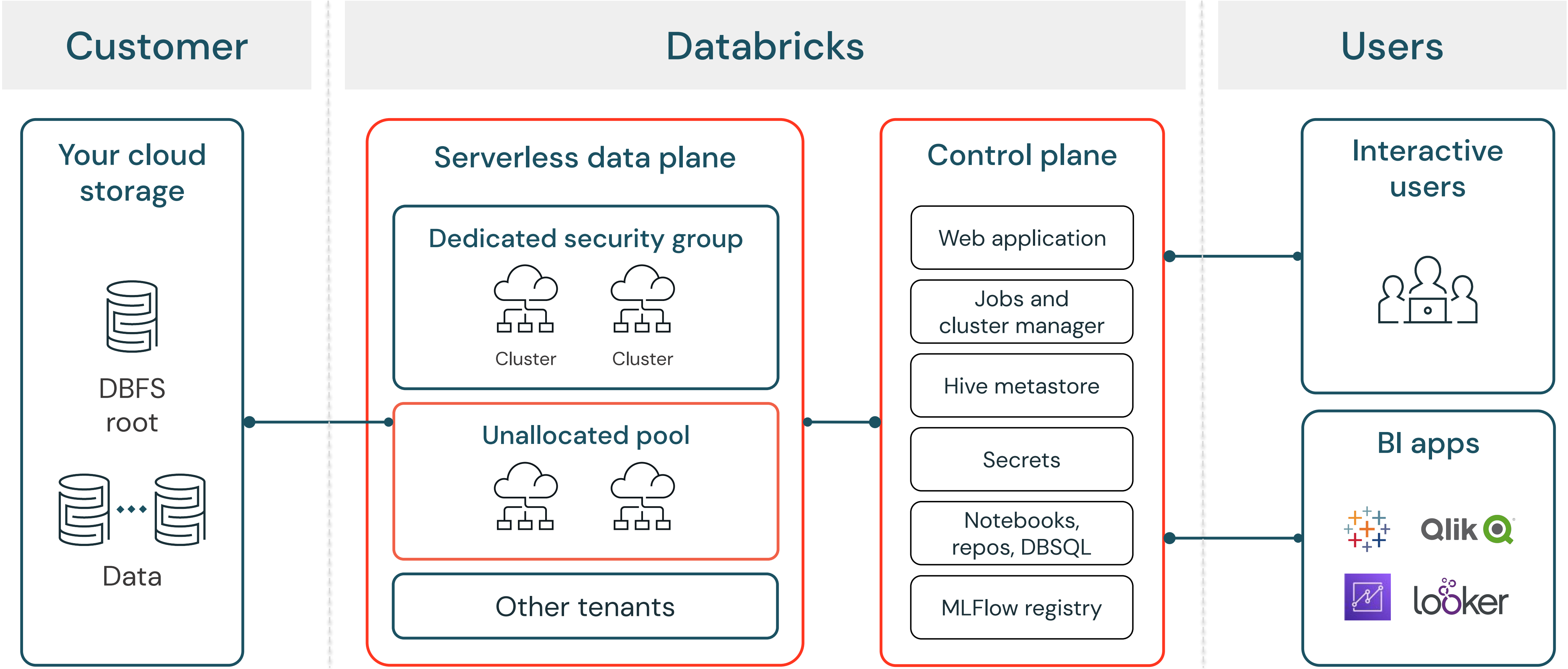
- VPC/VNet CIDRs
- Account/subscription sharding

Security

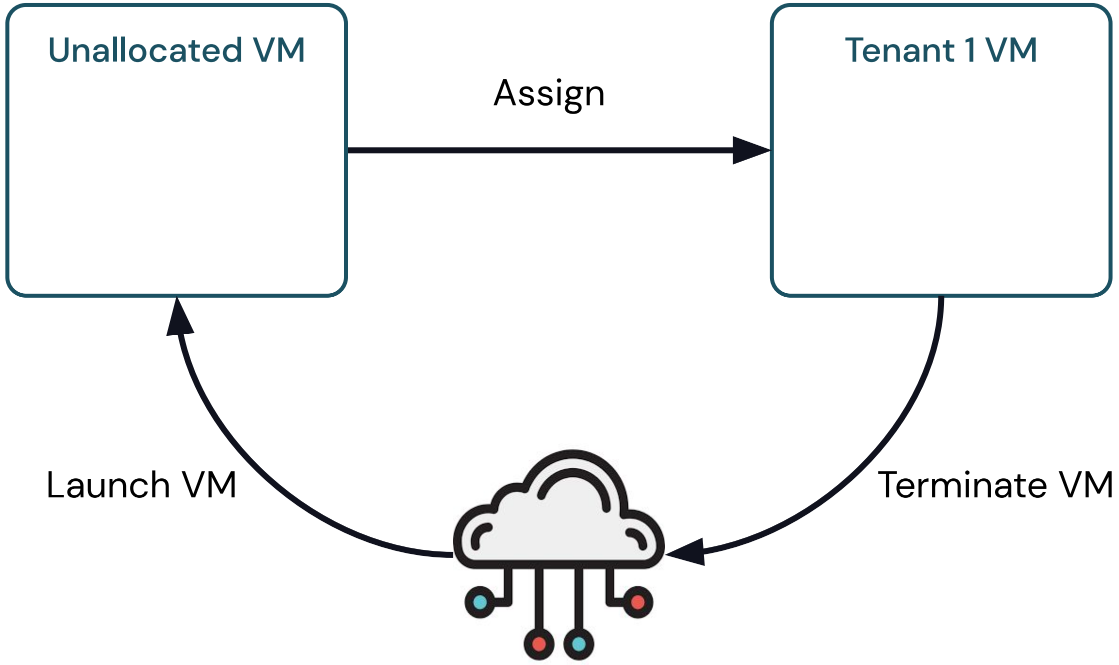
Classic architecture

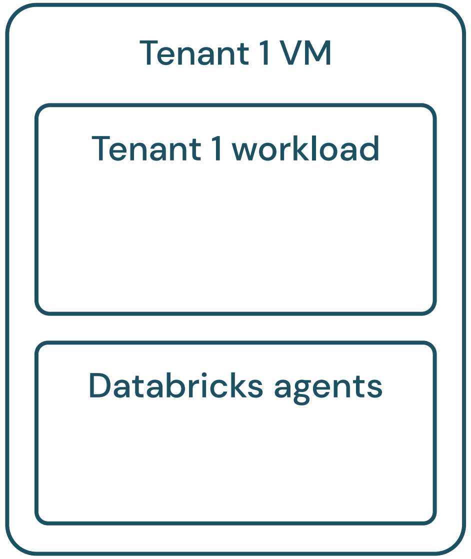


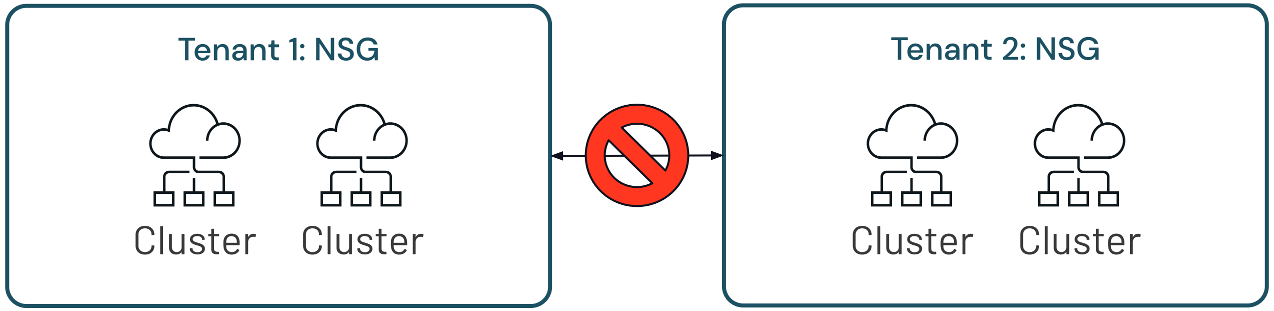
Serverless architecture



Serverless VM-level security

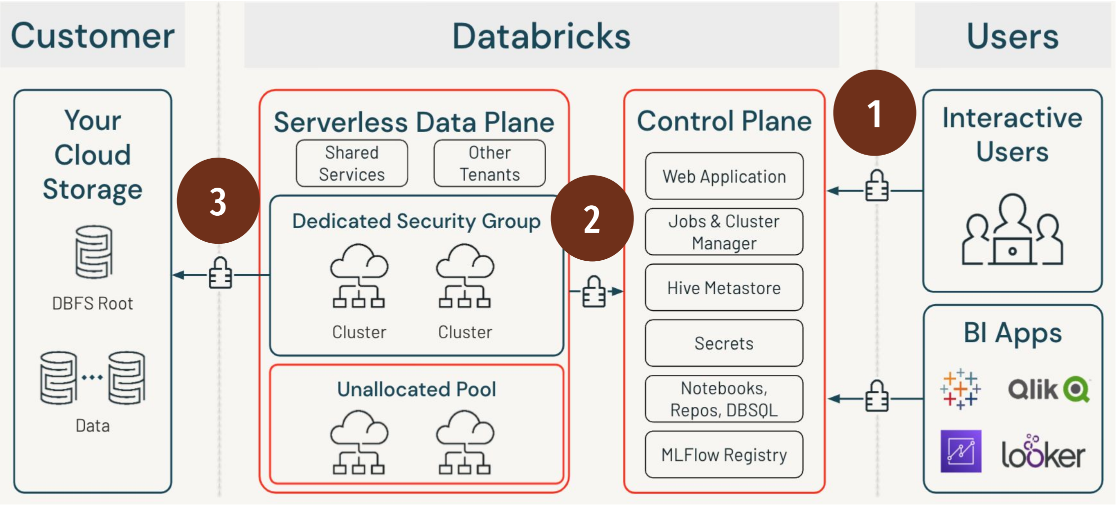






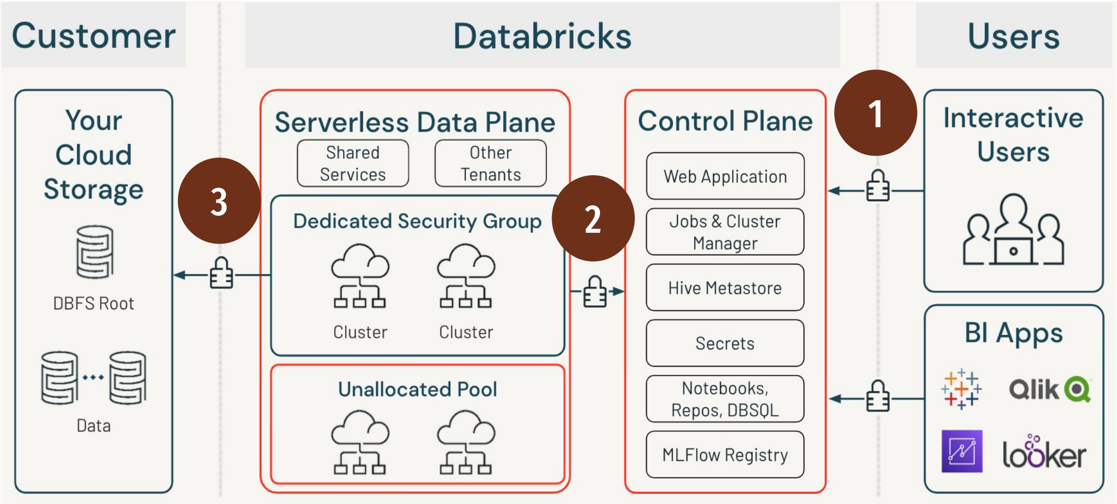
Network access

1. User to Control Plane Options
   - Open to all (Typical)
   - **IP Access Lists**
   - **Private Link** (Preview)



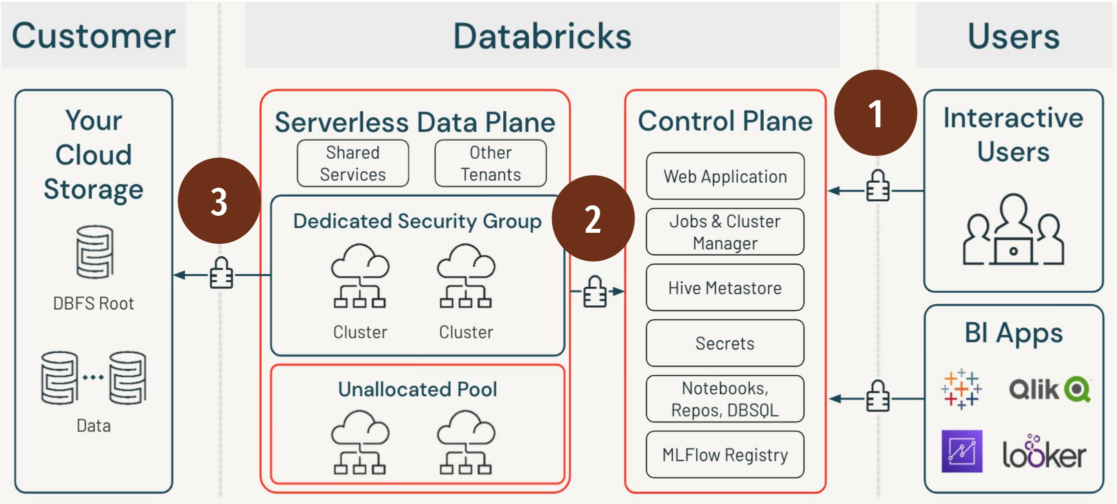
Network access

1. User to Control Plane Options
   - Open to all (Typical)
   - **IP Access Lists**
   - **Private Link** (Preview)
2. Data Plane to Control Plane managed by Databricks
   - mTLS 1.2+
   - **All workers are private**



Network access

1. User to Control Plane Options
   - Open to all (Typical)
   - **IP Access Lists**
   - **Private Link** (Preview)
2. Data Plane to Control Plane managed by Databricks
   - mTLS 1.2+
   - **All workers are private**
3. Data Plane to Customer Storage
   - Assume Role via Public IP
   - **Private network connectivity to blob storage**



Three layers of isolation controls

1. Container Isolation

- Hardened container images per industry best practice
- Disable privileged access in the container

2. VM Isolation

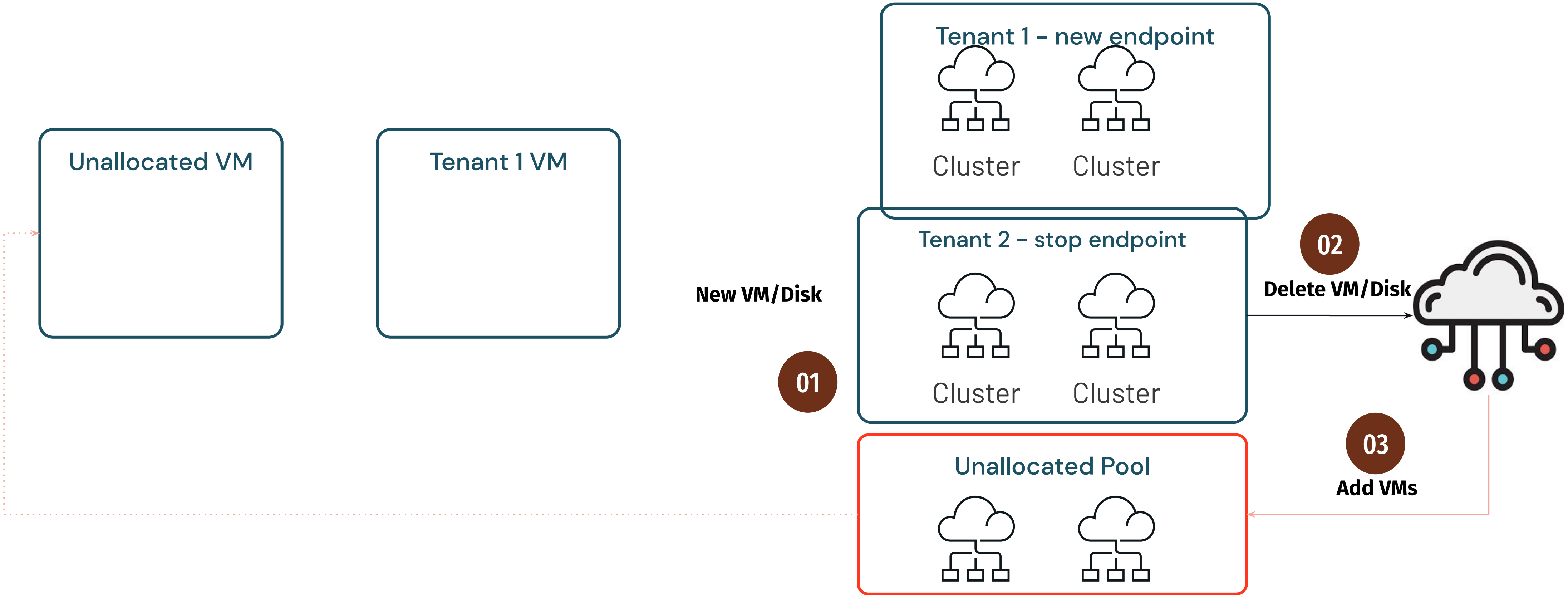
- No VMs reuse
- No privileges within broader environment

3. Network Isolation

- Federated access via temporary security tokens
- Only intra-cluster traffic allowed

Tested by internal teams and external vendors

Serverless Resource Life-cycle



Features and limitations

- Internal Hive Metastore is **supported**
- External Hive Metastore access via public IP is **supported**
- Unity Catalog is **supported**
- Private Link (DP -> S3/ADLS)
- Python/Scala are UDF not supported
- 5min auto-stop and 1min via API **in Q2**

Data location and encryption

- Data stays in customers cloud account (S3 buckets)
- Databricks managed Data Plane is in the same regions customers data plane—**no egress cost**

KYTHERA™

DECIPHERING HEALTHCARE

Serverless Access to Big Healthcare Data

DATA+AI SUMMIT 2022

KYTHERA LABS & DATABRICKS

Provided via a Big Data Platform

- Databricks' First healthcare OEM
- Ability to spin up and manage branded instances for prospects and clients (Wayfinder)
- Wayfinder gets clients access to refined data assets within hours - not months
- Increases return on data investments, 8x faster results, and richer, more granular insights

We Provide Remastered Healthcare Data and Analytics

- 330+ million patients
- 12.5 billion healthcare claims
- 12.9 billion Rx claims
- 27 billion claim lines
- BYO Data

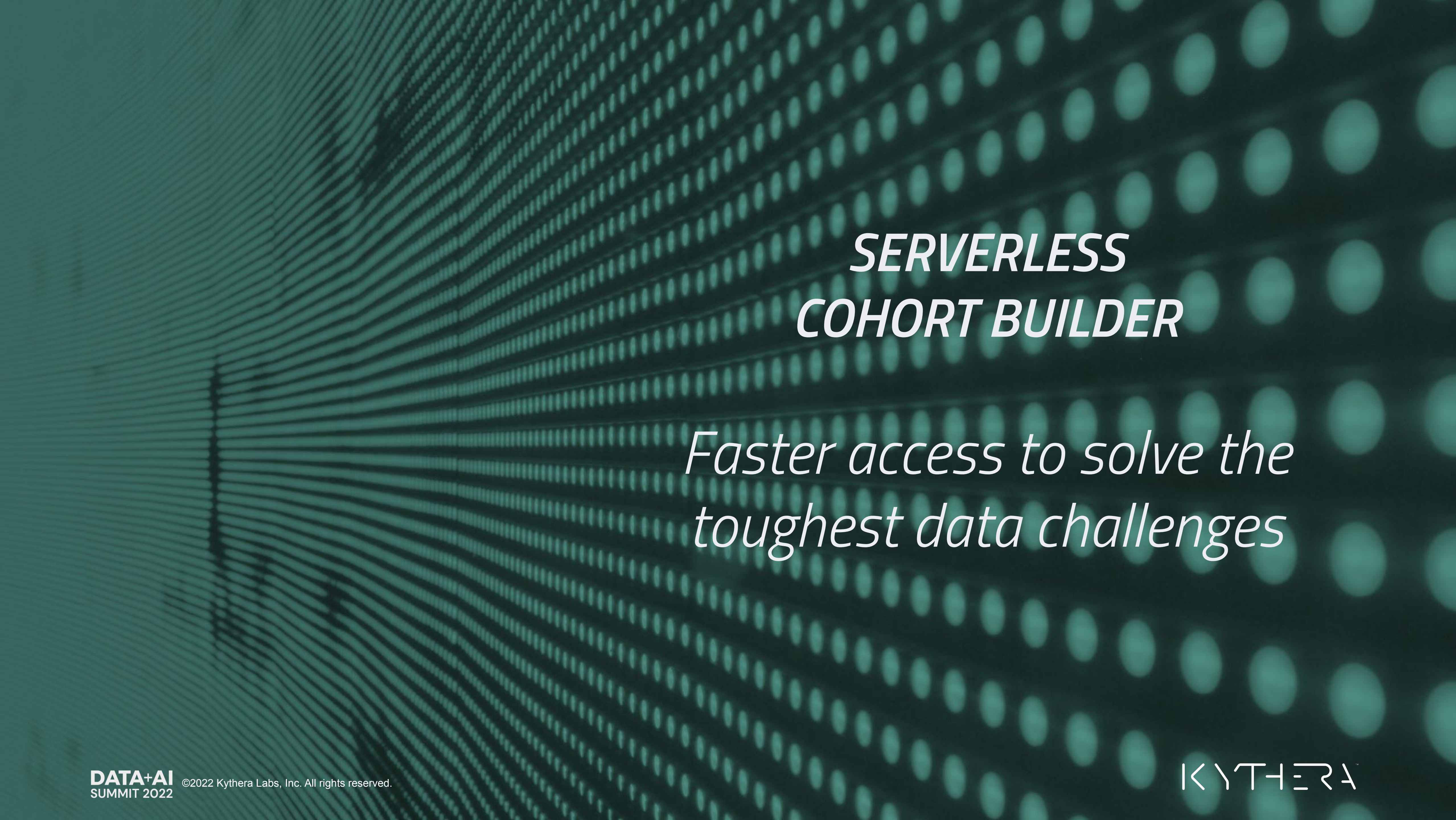
With Rich Analytics and Expertise

- Unified and simplified data model
- Corrected, standardized and flattened data models (Remastered)
- Foundational patient pathway analytics
- Machine learning models
- BI Visualizations
- Useful interfaces (Datavant, Salesforce, Mirador, etc)

DISCOVER TARGET POPULATIONS

Healthcare and Life Sciences companies often need to identify and validate patient populations and markets to understand how patients behave over time - where they receive care, who provides care, and how care is reimbursed - to make informed business decisions.

Patient Cohort dashboards empower business users to access and analyze data without wading through billions of rows of data, storing idle data, or waiting for a cluster to start.



SERVERLESS COHORT BUILDER

Faster access to solve the toughest data challenges

BETTER INSIGHTS.
FASTER.

Biggest Pain Points Addressed by Serverless

- Perfect for casual access by citizen analyst
  - Infrequent access to data, but need for quick response time
  - Serverless cuts out the cluster startup time
- Manage costs
  - Sure, with lots of money we could just have a big cluster waiting for that infrequent access
  - Serverless eliminates the cost of standby compute by leveraging shared resources across the customer base
- Accelerate the time to value and illuminate critical insights

Contact:
Matt@kytheralabs.com
Kytheralabs.com

THANK YOU

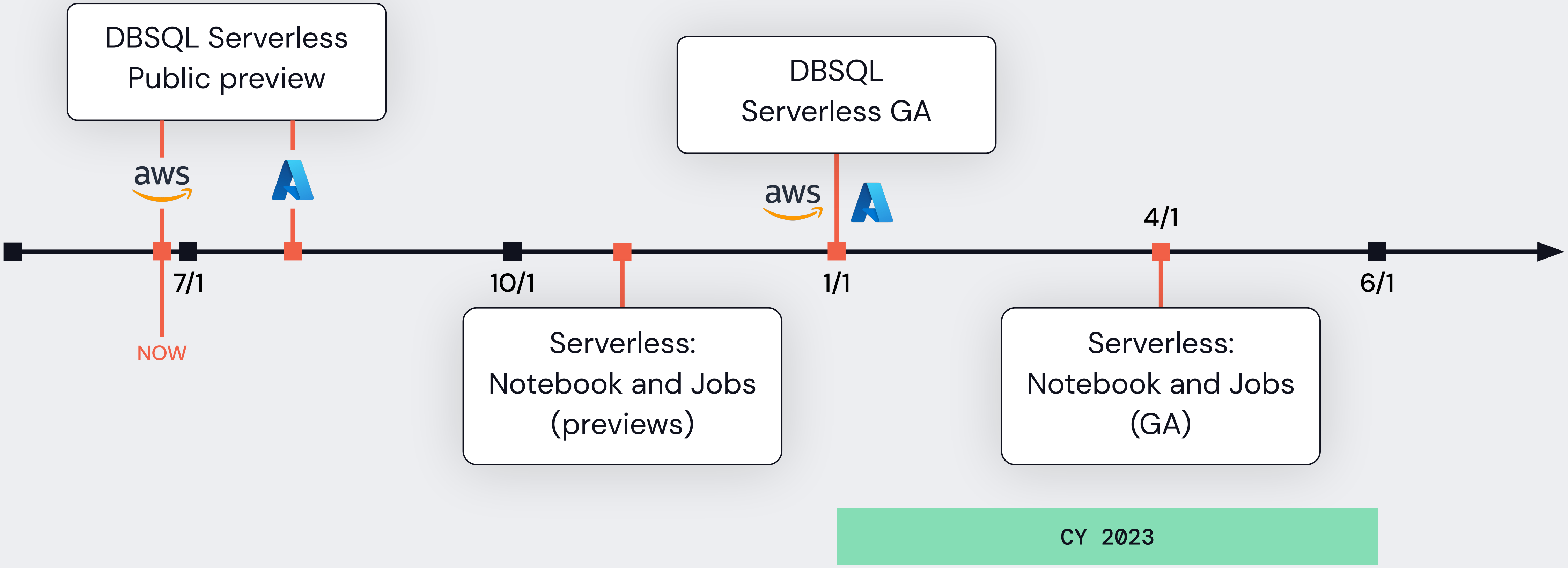
matt@kytheralabs.com

@kytheralabs

linkedin.com/company/kytheralabs

Roadmap

Roadmap: serverless platform



DBSQL Serverless Public preview

DBSQL Serverless GA

Serverless: Notebook and Jobs (previews)

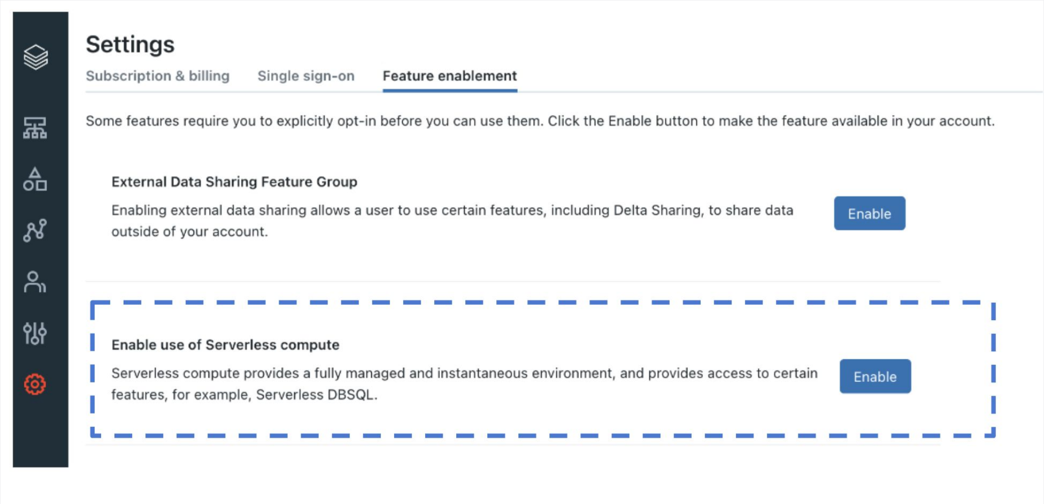
Serverless: Notebook and Jobs (GA)

Try it now!

1. Databricks SQL customers on AWS:
   a) Enable Serverless from account console
   b) Upgrade all endpoints from classic to Serverless

2. Databricks SQL customers on Azure:
   a) Submit your interest: bit.ly/DBSQLServerless-Azure

3. Learn more, talk to your Databricks representative



Settings

Subscription & billing | Single sign-on | **Feature enablement**

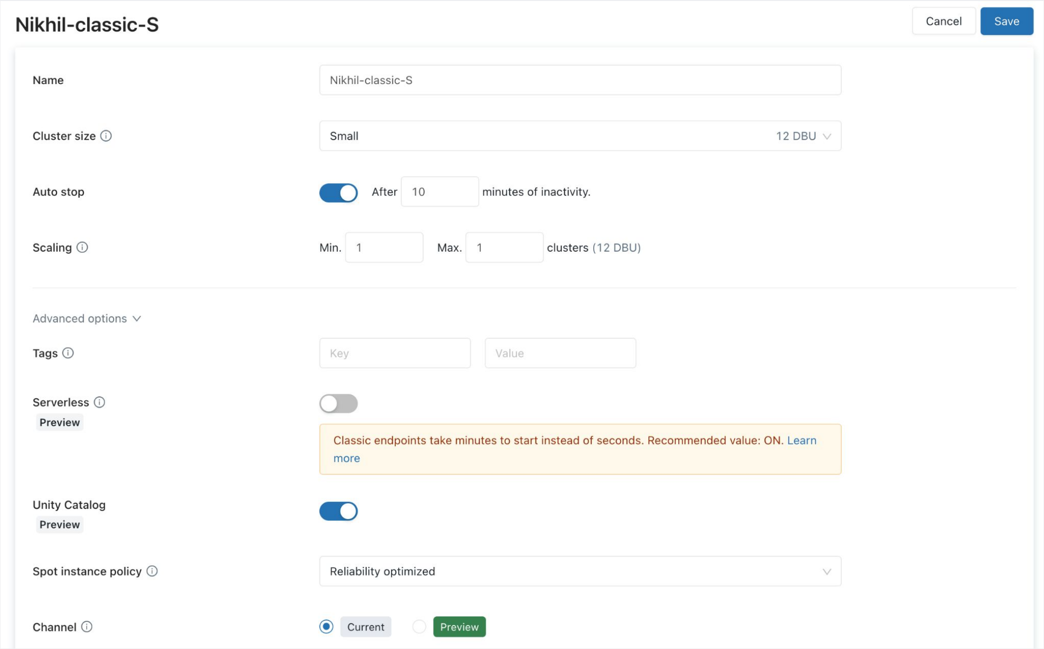
Some features require you to explicitly opt-in before you can use them. Click the Enable button to make the feature available in your account.

External Data Sharing Feature Group

Enabling external data sharing allows a user to use certain features, including Delta Sharing, to share data outside of your account. Enable

Enable use of Serverless compute

Serverless compute provides a fully managed and instantaneous environment, and provides access to certain features, for example, Serverless DBSQL. Enable



Nikhil-classic-S Cancel Save

Name	Nikhil-classic-S
Cluster size ⓘ	Small 12 DBU
Auto stop	After 10 minutes of inactivity.
Scaling ⓘ	Min. 1 Max. 1 clusters (12 DBU)

Advanced options

Tags ⓘ	Key Value
Serverless ⓘ Preview	Classic endpoints take minutes to start instead of seconds. Recommended value: ON. Learn more
Unity Catalog Preview	
Spot instance policy ⓘ	Reliability optimized
Channel ⓘ	● Current ○ Preview

DATA+AI
SUMMIT 2022

Thank you