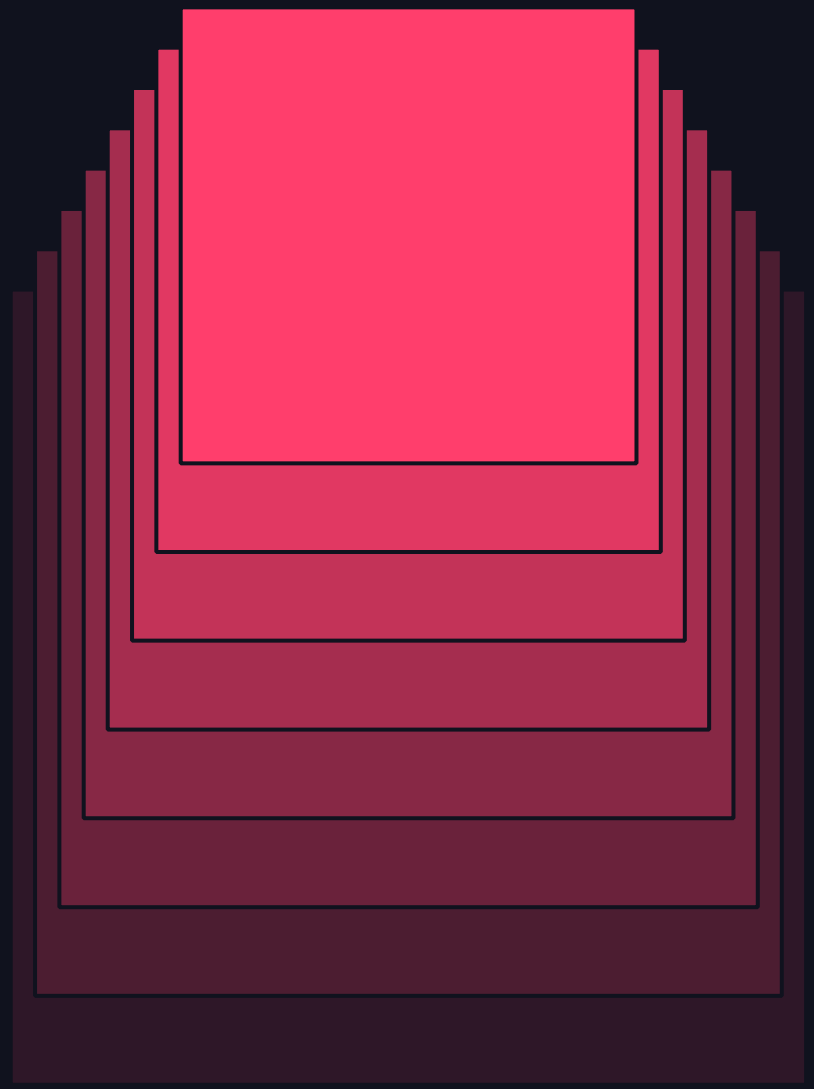
DATA+AI SUMMIT
BY databricks

Default Storage: Fast Track to Databricks Data Intelligence Platform



Emma Liu, Staff Product Manager @ Databricks

Samrat Ray, Sr. Staff Product Manager @ Databricks

Data and AI Summit 2024

©2024 Databricks Inc. — All rights reserved

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

Benyue (Emma) Liu

- **Staff Product Manager** – Default Storage and Delta Lake @ Databricks
- **Previous Experiences:** Product Manager at TigerGraph, MarkLogic; Software Engineer at Oracle
- **Domain Focus:** Delta Lake, Ingestion, DBaaS, Multi-Model & Graph Databases, Developer/DBA Tools, Complex System Design

databricks

©2024 Databricks Inc. — All rights reserved

databricks

Samrat Ray

- **Sr. Staff Product Manager** – Enterprise Security @ Databricks
- **Previous Experiences:** Group PM at Google Cloud – launched VPC Service Controls, Cloud SQL. Founding PM of Istio, a successful open source project.
- **Domain Focus:** Security for the Databricks platform including security of the serverless platform, encryption, data protection and privacy

©2024 Databricks Inc. — All rights reserved

Have you ever set up
a new Databricks account,
a new workspace,
or enabled Unity Catalog?

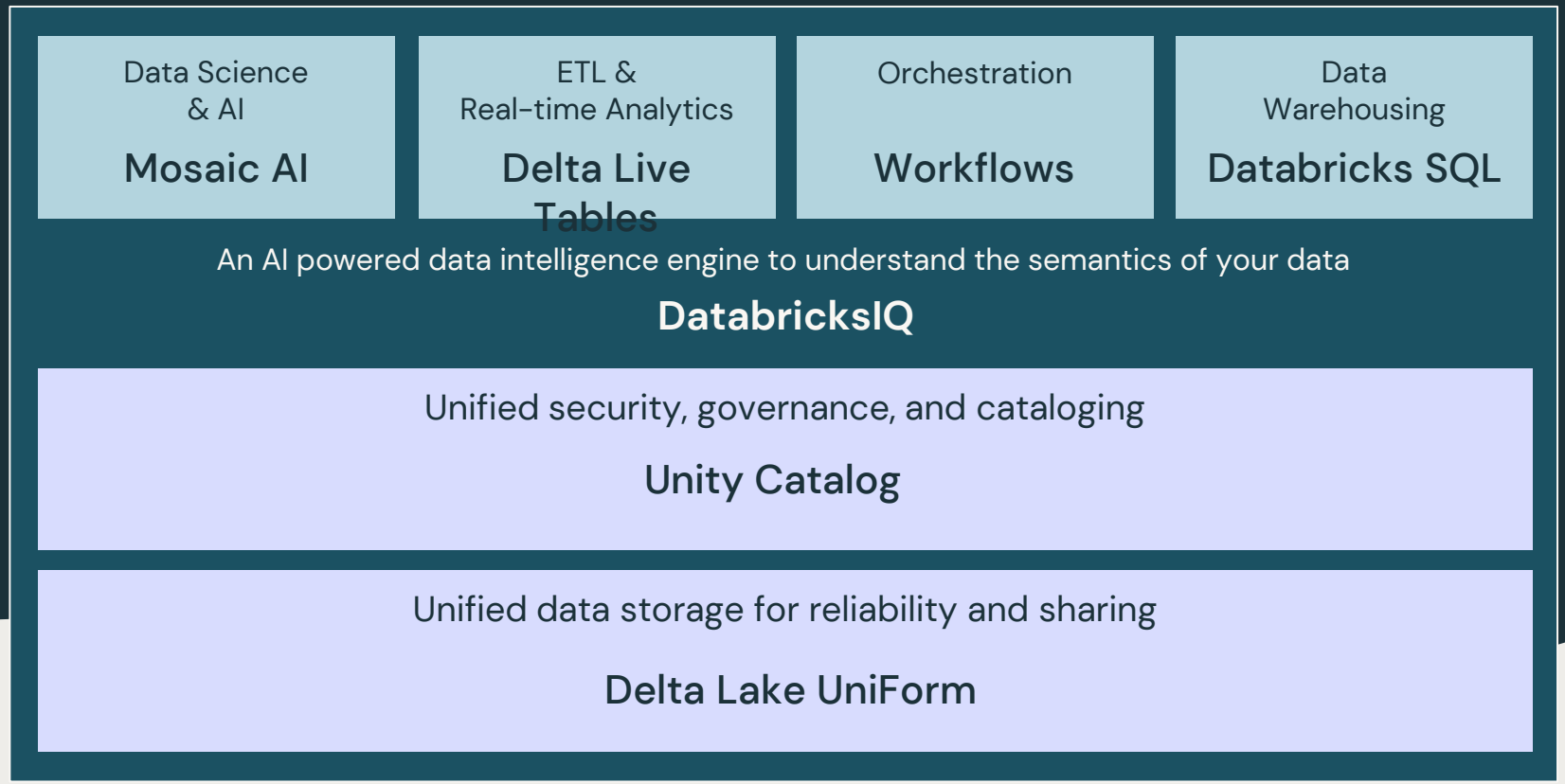
©2024 Databricks Inc. — All rights reserved

Agenda

- Background
- What is Default Storage?
- Use Cases & Demos
- Architecture
- Roadmap & Vision
- Takeaways

©2024 Databricks Inc. — All rights reserved

Databricks Data Intelligence Platform



Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)

2020
Databricks pioneered the lakehouse architecture

Today
74% of global enterprises have adopted lakehouse

MIT Technology Review Insights, 2023

Delta Lake Uniform and Delta Sharing

Scalable

4+

Exabytes

processed / day

Reliable

>10K

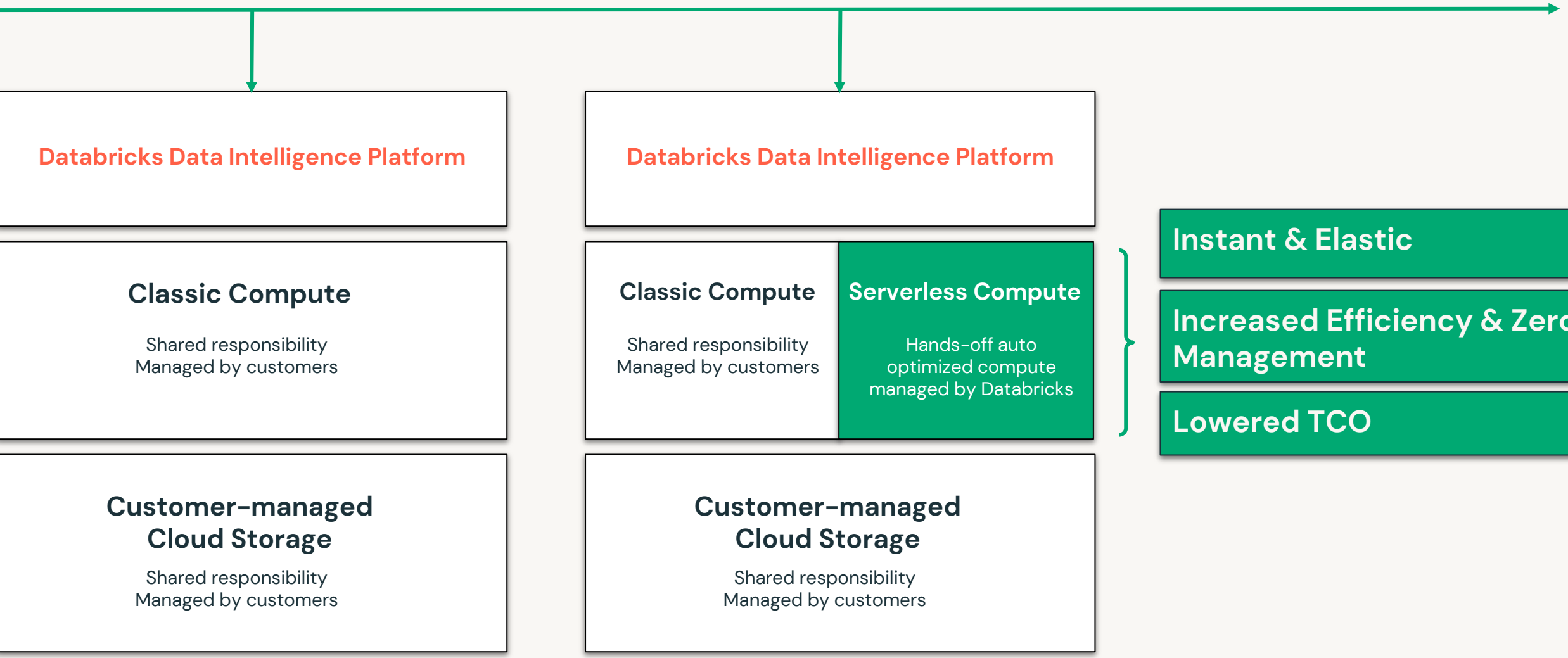
companies in production

Open

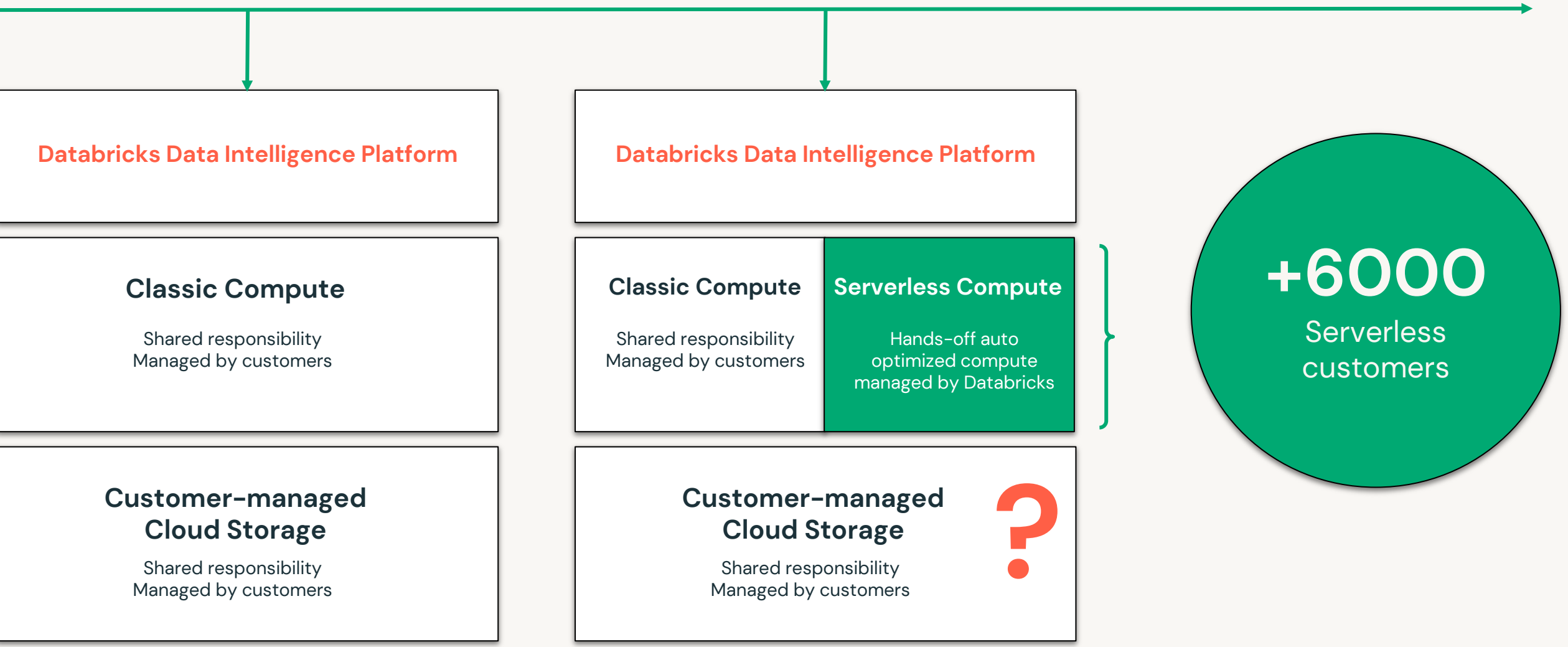
+32K

data recipients

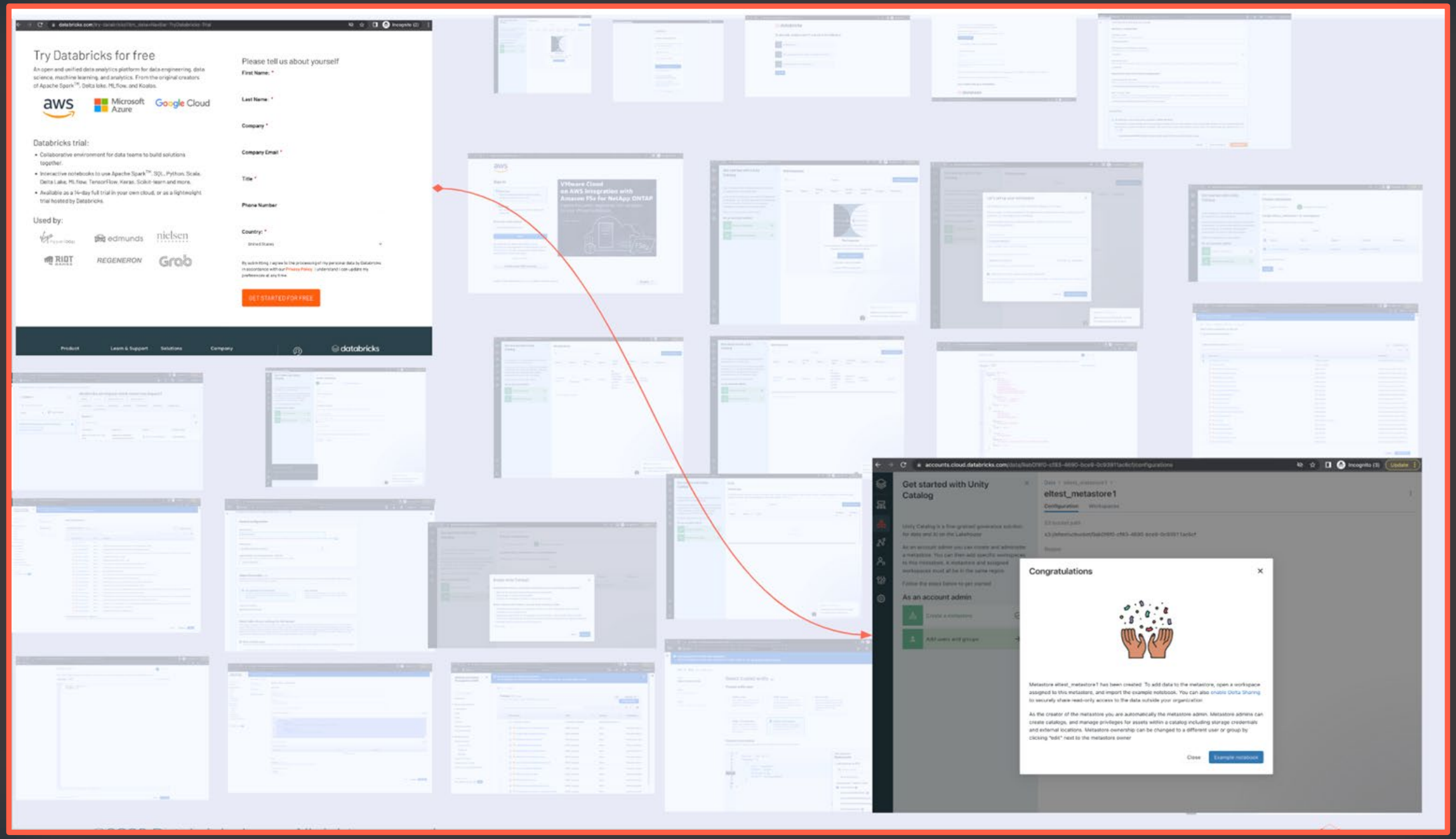
©2024 Databricks Inc. — All rights reserved

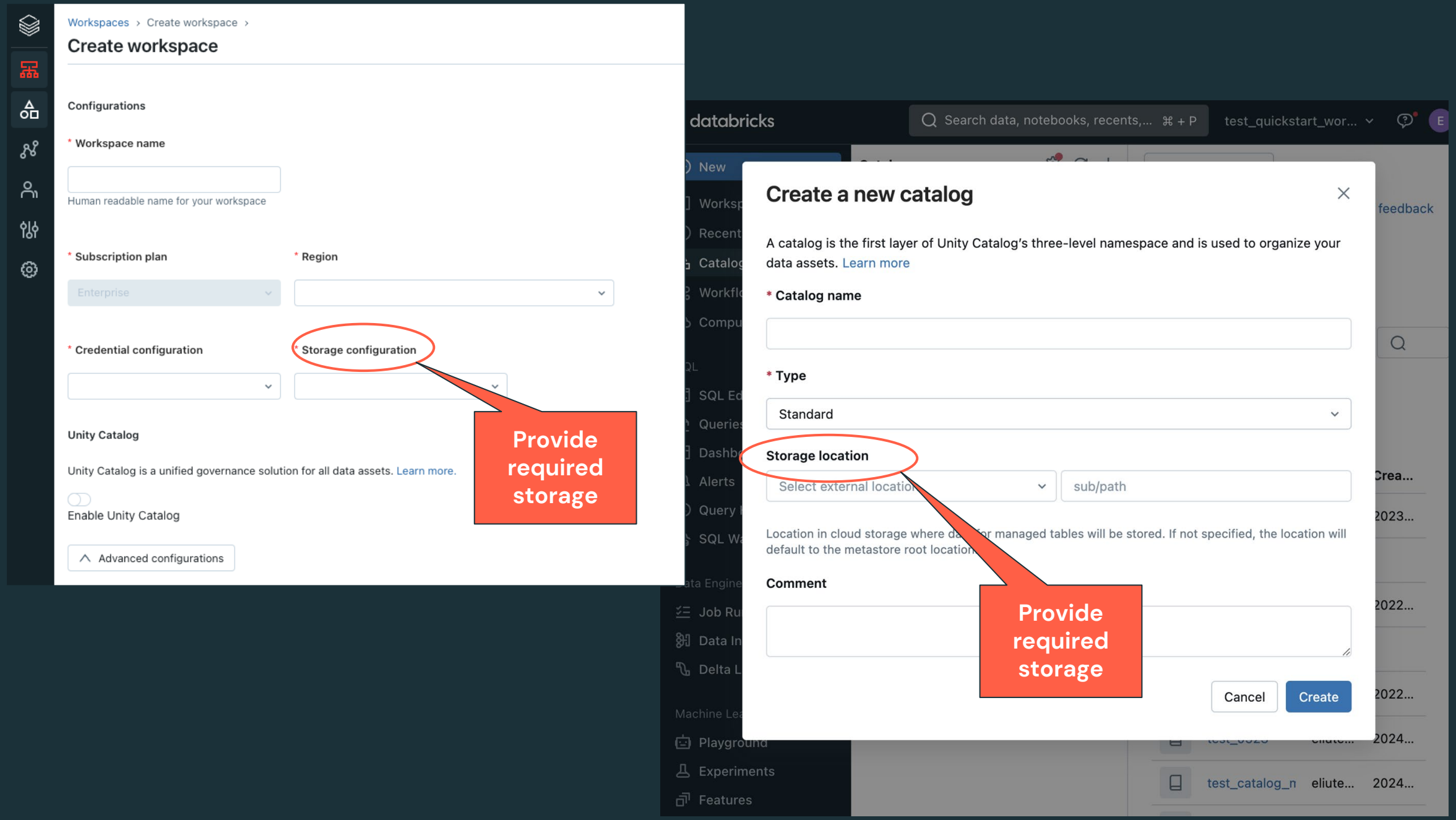


©2024 Databricks Inc. — All rights reserved



©2024 Databricks Inc. — All rights reserved


Try Databricks for free
An open and unified data analytics platform for data engineering, data science, machine learning, and analytics. From the original creators of Apache Spark™, Delta lake, MLflow, and Koalas.
aws
Microsoft Azure
Google Cloud
Databricks trial:
Collaborative environment for data teams to build solutions together.
Interactive notebooks to use Apache Spark™, SQL, Python, Scala, Delta Lake, MLflow, TensorFlow, Keras, Scikit-learn and more.
Available as a 14-day full trial in your own cloud, or as a lightweight trial hosted by Databricks.
Used by:
edmunds
nielsen
RIOT GAMES
REGENERON
Grab
Please tell us about yourself
First Name: *
Last Name: *
Company *
Company Email *
Title *
Phone Number
Country: *
GET STARTED FOR FREE
Product
Learn & Support
Solutions
Company
databricks
Get started with Unity Catalog
eltest_metastore1
Configuration
As an account admin
Congratulations
Close
Example notebook


Workspaces › Create workspace ›
Create workspace
Configurations
* Workspace name
Human readable name for your workspace
* Subscription plan
Enterprise
* Region
* Credential configuration
* Storage configuration
Unity Catalog
Unity Catalog is a unified governance solution for all data assets. Learn more.
Enable Unity Catalog
Advanced configurations
Provide required storage
databricks
Search data, notebooks, recents,... ⌘ + P
test_quickstart_wor...
feedback
Create a new catalog
A catalog is the first layer of Unity Catalog's three-level namespace and is used to organize your data assets. Learn more
* Catalog name
* Type
Standard
Storage location
Select external location
sub/path
Location in cloud storage where data for managed tables will be stored. If not specified, the location will default to the metastore root location
Comment
Cancel
Create
Provide required storage
Playground
Experiments
Features
test_catalog_n
eliute...
2024...

These are not problems only in Day 1 journey.
What about ongoing management?

©2024 Databricks Inc. — All rights reserved

Meet Sally: Your Favorite Gen AI ML Scientist

Scenario: Sally has been assigned to implement an end-to-end Machine Learning pipeline incorporating LLMs to accelerate the time-to-market of their new AI-powered products. She needs:

Instant access to DBRX in Databricks Data Intelligence Platform without having a cloud account

Auto provisioned storage, compute and Unity Catalog in an isolated workspace

Secured environment for data storage, access and sharing

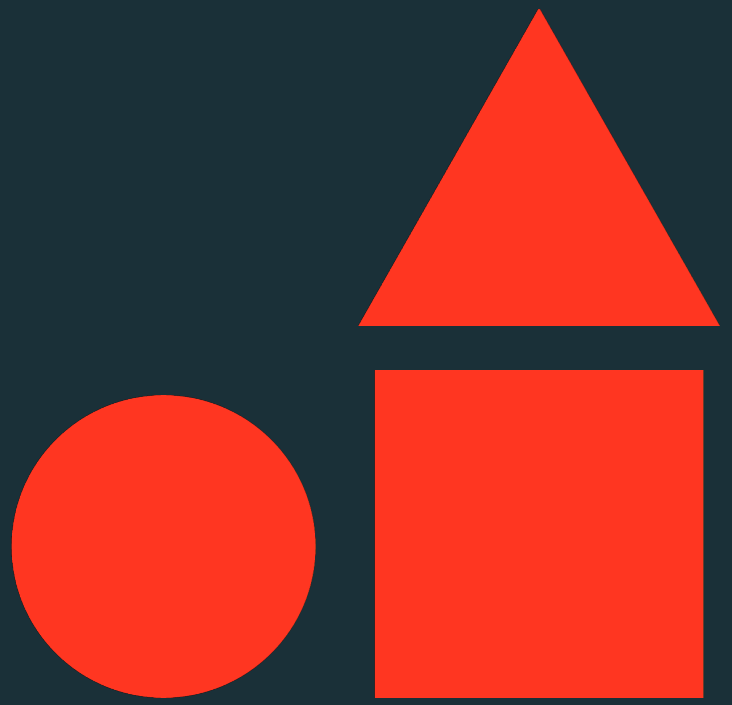
Wants automatic maintenance

Simple dashboard to report the cost on the data assets and workloads

©2024 Databricks Inc. — All rights reserved

Democratizing Data and AI demands a new option

Introducing Default Storage

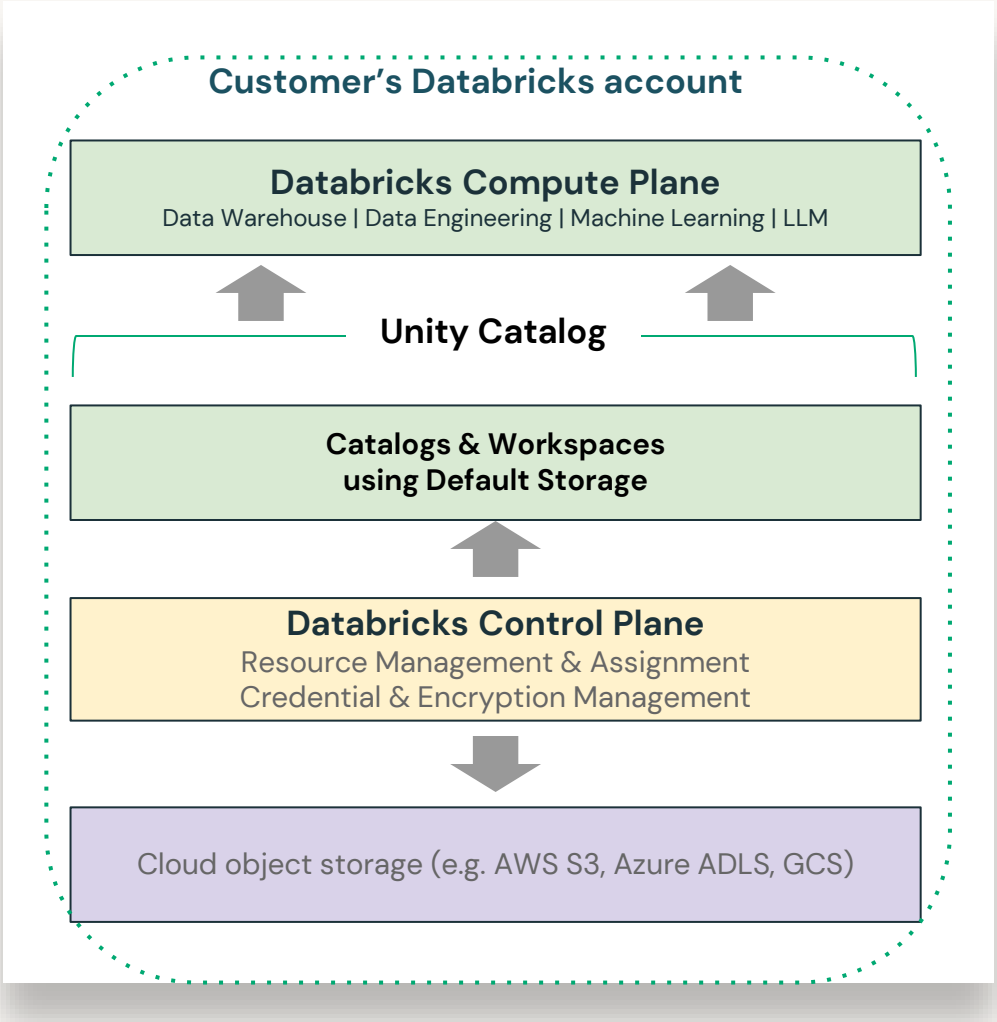


Default Storage

One-click set up for new workspaces and catalogs

Pre-configured storage for workspaces and Unity Catalog

Set & forget: zero ongoing maintenance



©2024 Databricks Inc. — All rights reserved

Default Storage

Fast Track to
Data Intelligence
Platform;
Zero Cloud Infrastructure

Simple and Efficient

Fast onboarding and optimized with Serverless Compute for a hassle-free Lakehouse experience

Secure by Default

for all your governed data assets

Unified data access, billing and observability

for all your workloads

©2024 Databricks Inc. — All rights reserved

Databricks Data Intelligence Platform

Fast Track = Default Storage + Serverless Compute

Default Storage Use Cases

©2024 Databricks Inc. — All rights reserved

1. Summit Test Drive with Default Storage

Simpler onboarding experience optimized for data and AI practitioners

- Get started in seconds: no cloud account, admin access, or credit card required
- Preconfigured with default storage, serverless compute and Unity Catalog
- Playground for Summit Contents

Find Test Drive
invite codes at Databricks Booth

©2024 Databricks Inc. — All rights reserved

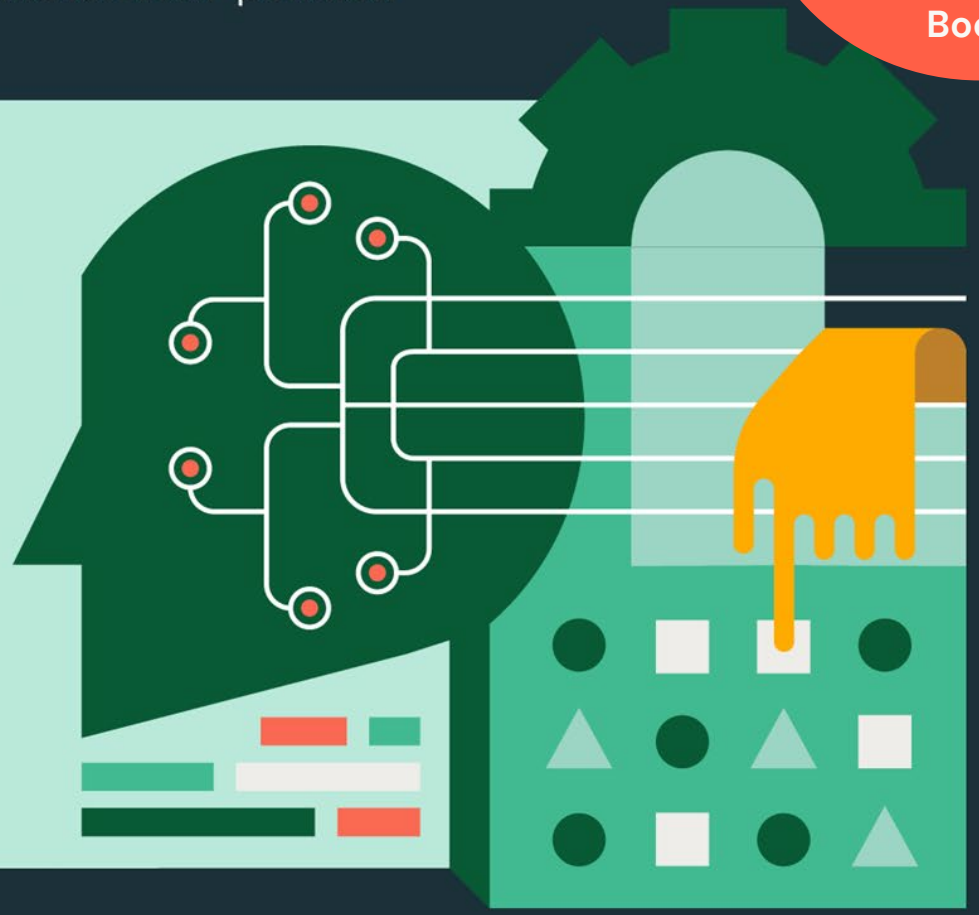
DATABRICKS TEST DRIVES

$100
In free compute and storage credits June 10–24 Databricks Booth

Created exclusively for Data + AI Summit attendees, Databricks Test Drives is your chance to get hands-on with the latest AI-powered features of the Data Intelligence Platform.

WITH TEST DRIVES, YOU CAN:

- Compare DBRX, Llama 3, Mixtral and other language models head-to-head
- Create drag-and-drop charts — or ask Databricks Assistant to visualize data for you
- Intuitively analyze data by asking natural language questions



©2024 Databricks Inc. — All rights reserved

DEMO

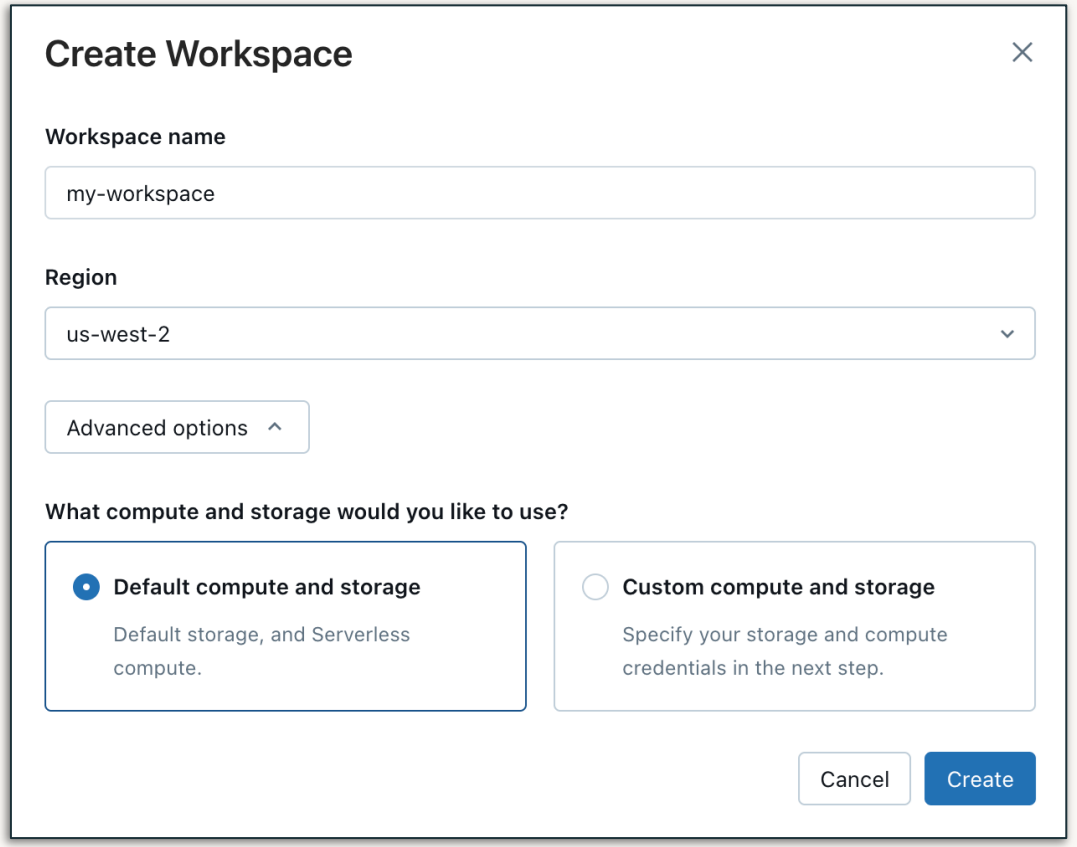
Summit Test Drive
Powered by Default Storage

©2024 Databricks Inc. — All rights reserved

2. New Workspaces with Default Storage

One click set up for new workspaces

- Start new workspaces in seconds
- Preconfigured with default storage, serverless compute and Unity Catalog by default
- Enable self-service teams and isolated workspaces



Enroll in Private Preview

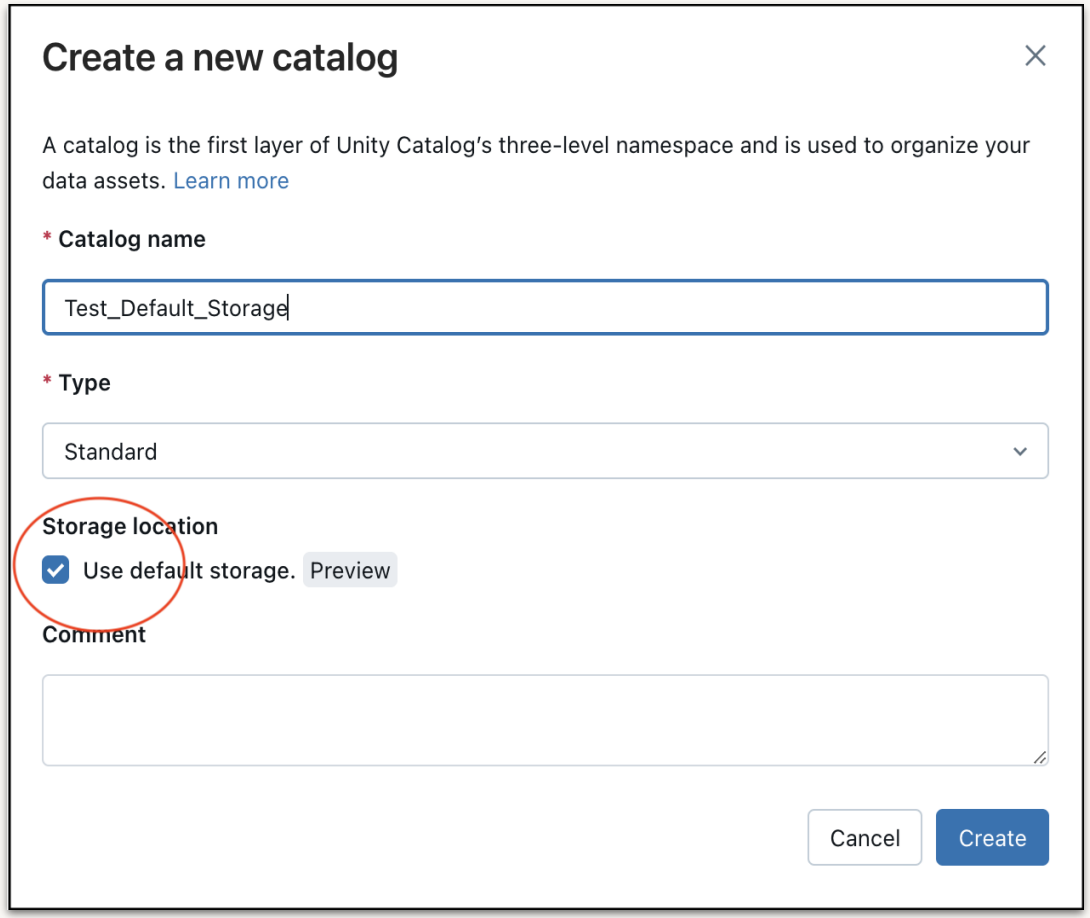
©2024 Databricks Inc. — All rights reserved

3. Default Storage for Unity Catalog

One click set up for new catalogs

- Start new catalogs in seconds
- Unified data access, billing, and observability
- Enable new data projects

Enroll in Private Preview



©2024 Databricks Inc. — All rights reserved

DEMO

- A New Catalog
- A New Workspace

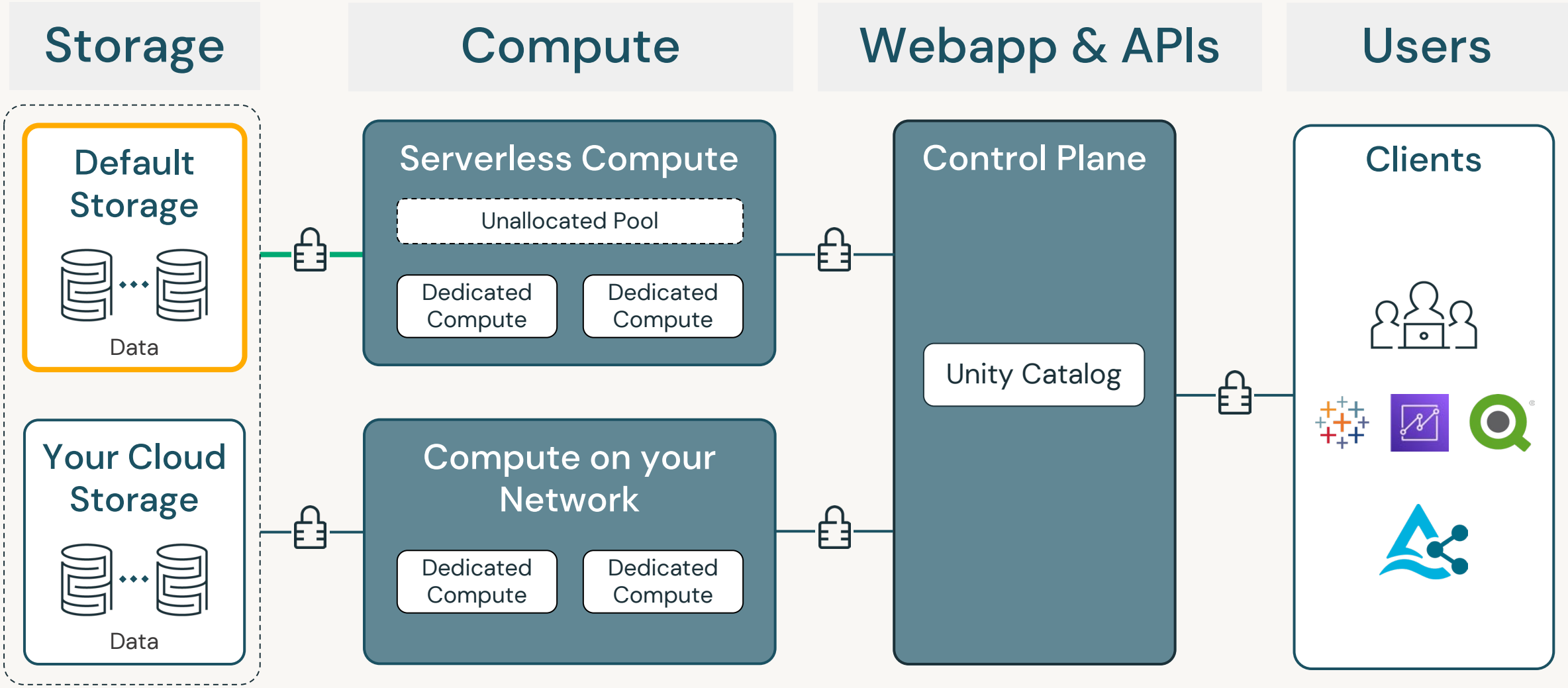
Powered by Default Storage

©2024 Databricks Inc. — All rights reserved

How does Databricks Default Storage protect your data?

©2024 Databricks Inc. — All rights reserved

Architecture



©2024 Databricks Inc. — All rights reserved

Default Storage – security is baked in

Always encrypted and isolated at rest

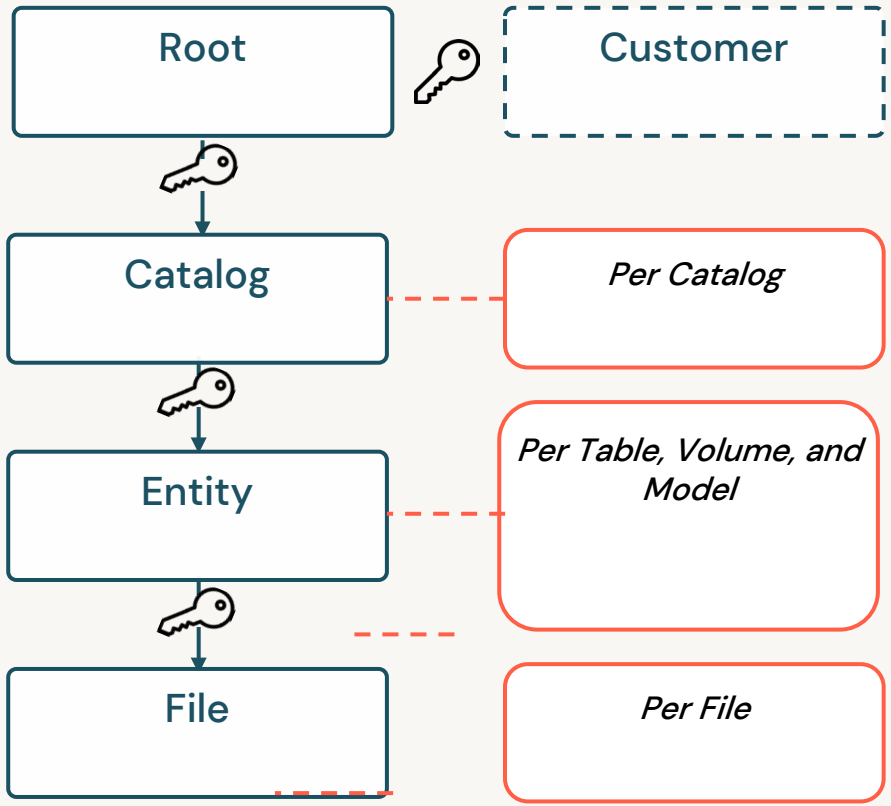
Can only be accessed via Unity Catalog on authorized Workspaces

Secured by default – no configuration needed!

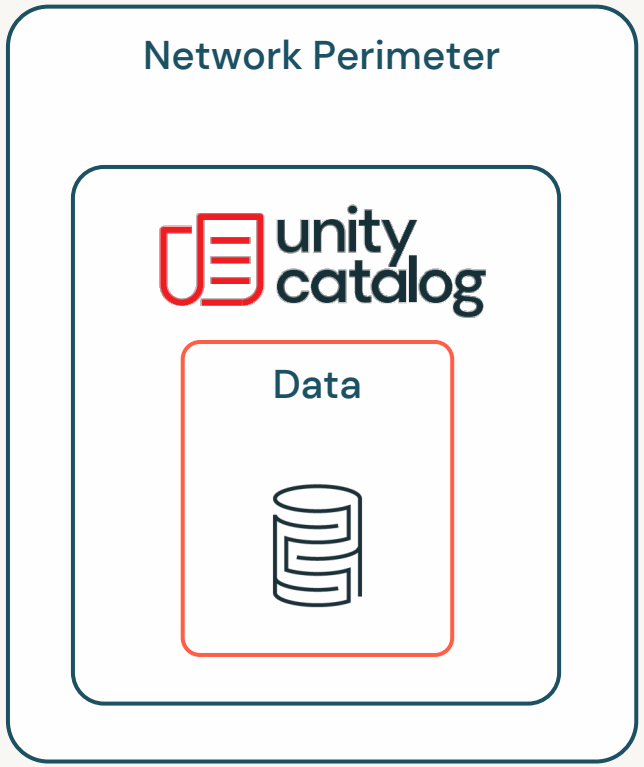
©2024 Databricks Inc. — All rights reserved

Protecting your data with defense in depth

Isolation at Rest
Per-tenant encryption keys



Isolation at Access
UC & Networking Controls



©2024 Databricks Inc. — All rights reserved

Encryption Process Flow

Isolation at rest in Default Storage

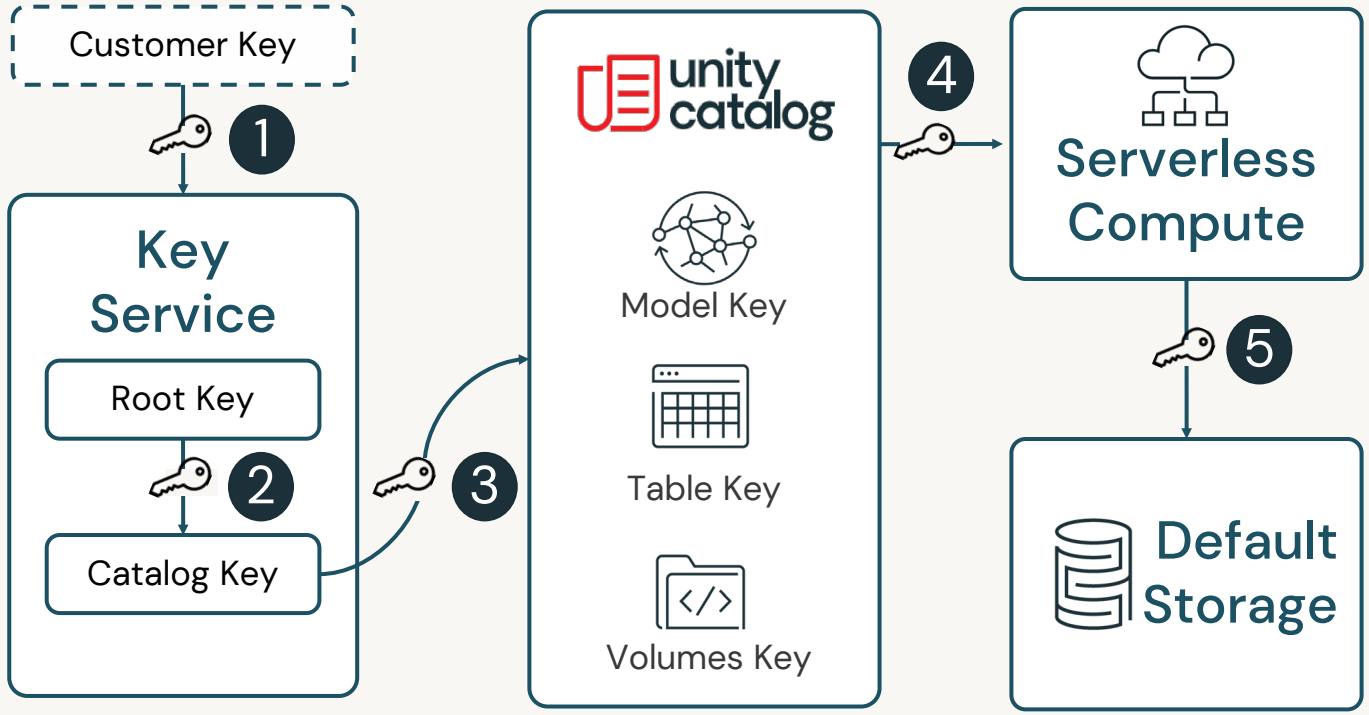
1: [roadmap] Each Catalog can be protected by a customer-managed key.

2: Separate key hierarchy per Catalog..

3: Catalog key encrypts the entity keys, such as table, model, and volumes. These keys never leaves the key service.

4: File level keys are derived in memory from the key hierarchy.

5: The file key is used to encrypt each file at rest in the Default Storage.



©2024 Databricks Inc. — All rights reserved

Catalog Access

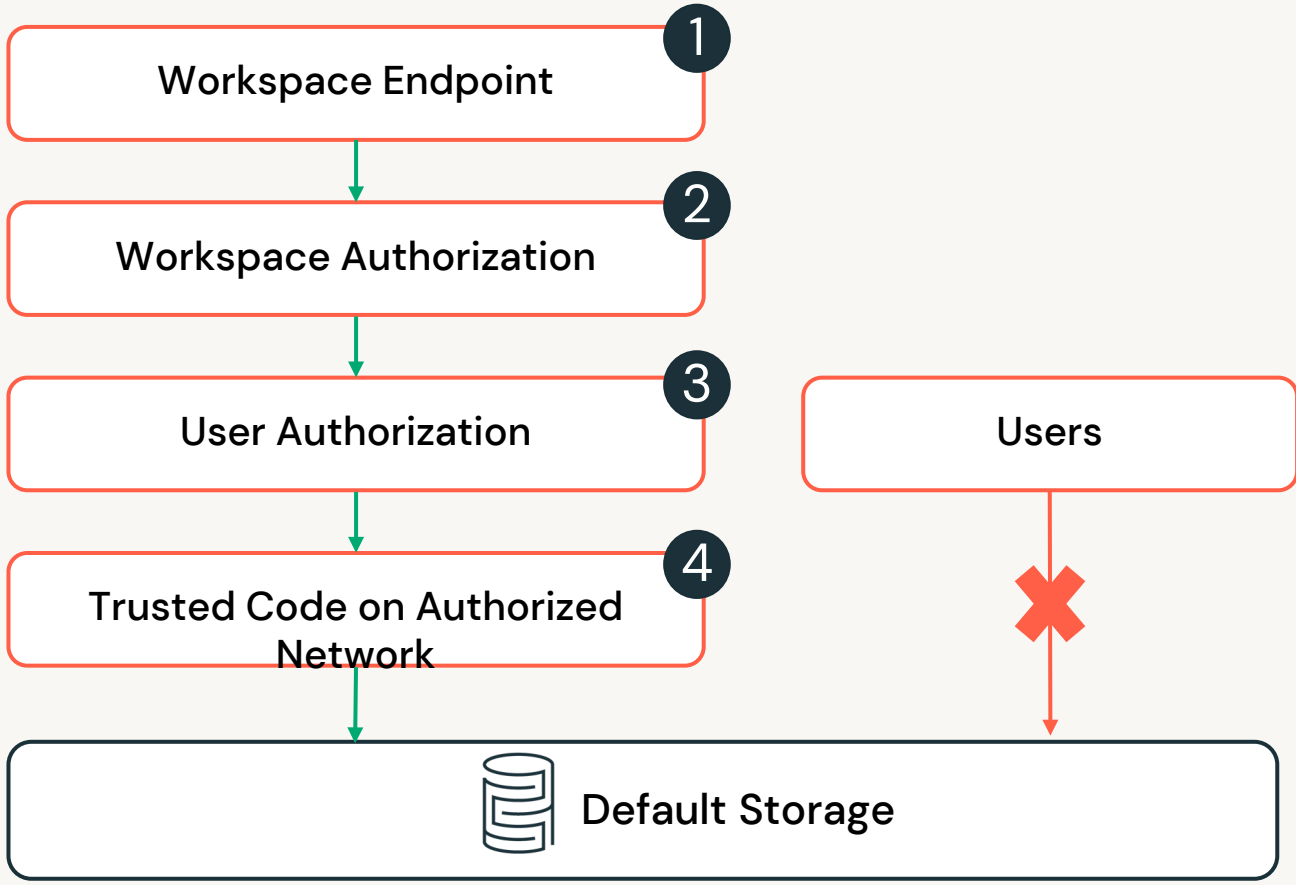
Isolation at access in Default Storage

1: A user must have access to a workspace; which can be restricted to authorized networks or IP addresses.

2: The workspace must be authorized to access the UC catalog.

3: The user must have access to the entity granted to them through Unity Catalog permissions.

4: Data encryption keys are not exposed to user code. Only trusted code – SQL engine, UC clients – running on authorized networks have access to the data encryption keys.



©2024 Databricks Inc. — All rights reserved

Secured for you by Databricks

- Databricks secures Storage and data based on industry best practices.
- You don’t need to worry about Storage policies, firewalls, connectivity, isolation, ...
- You can establish custodianship over your data with customer managed key support (roadmap).

©2024 Databricks Inc. — All rights reserved

Roadmap & Vision

©2024 Databricks Inc. — All rights reserved

Roadmap & Vision

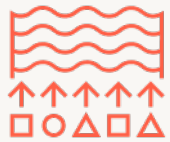
Customer-managed Keys for data assets in Default Storage

Classic clusters connecting to Default Storage

Streamlined Backup/DR with Default Storage

Push-based Ingestion to Default Storage

Intelligent Data Lifecycle & Storage Tiering

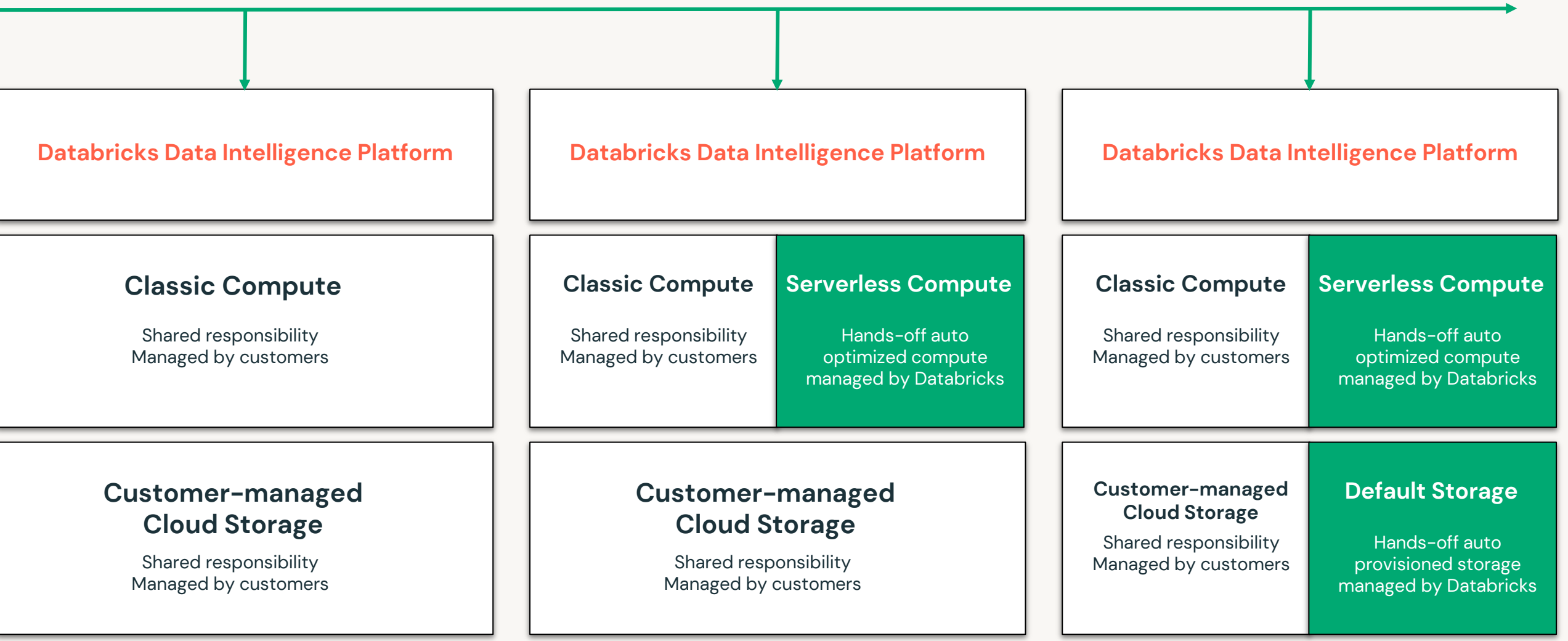


Performance Optimization

©2024 Databricks Inc. — All rights reserved

Takeaways

©2024 Databricks Inc. — All rights reserved



©2024 Databricks Inc. — All rights reserved

Call to Action

Try Databricks Test Drive

Find invite codes

at Databricks Booth

Enroll in Default Storage Private Preview

©2024 Databricks Inc. — All rights reserved

Thank you!

©2024 Databricks Inc. — All rights reserved