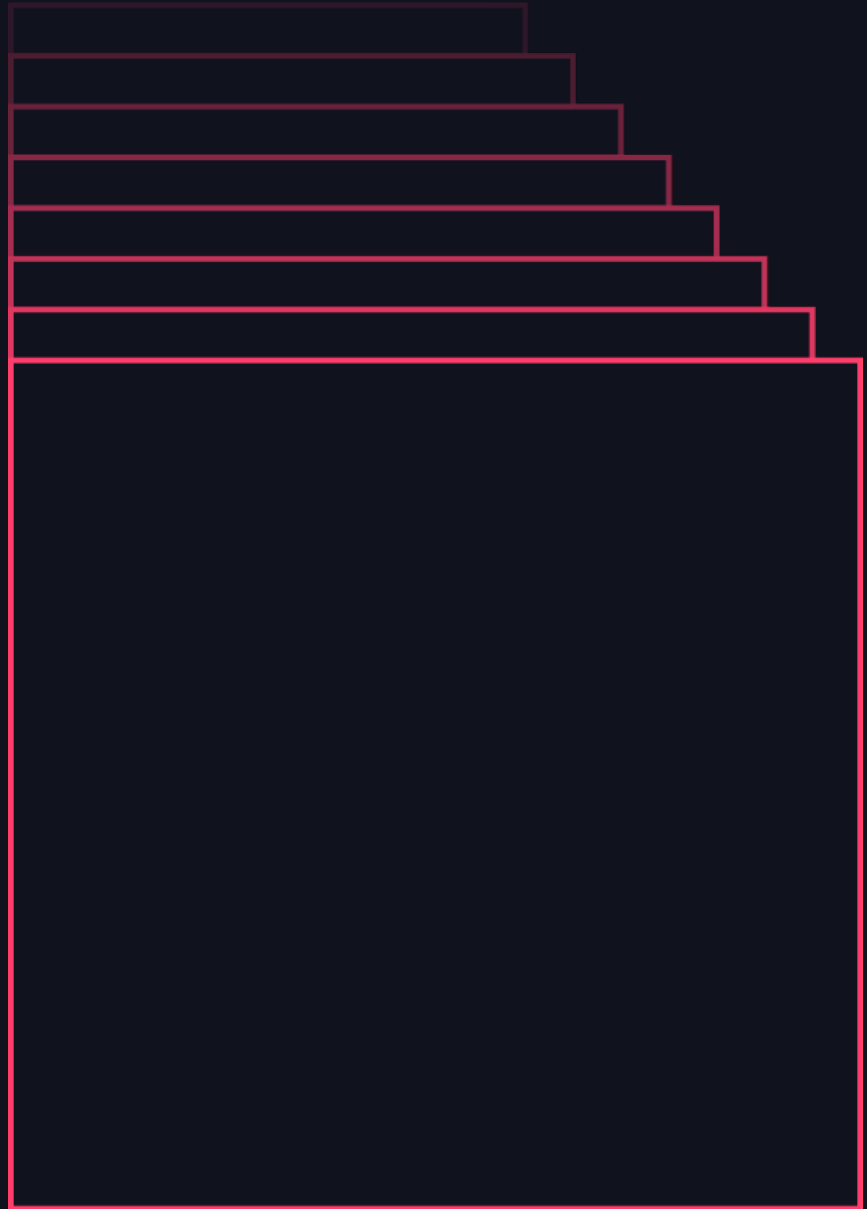


# How migrating to Unity Catalog unlocked the Diageo data roadmap

---

Redkite by Accenture  
Abigail Bennett

Diageo  
James Turner



# Introduction



Abigail Bennett  
Industry Director - CPG  
Redkite



James Turner  
Head of Data & Analytics  
Diageo

# Who are Diageo?

- Global leader of alco-bev, with over 200 premium spirits and beer brands
- Headquartered in London, UK
- Sell our brands in over 180 countries
- 30k employees, and 132 production sites worldwide

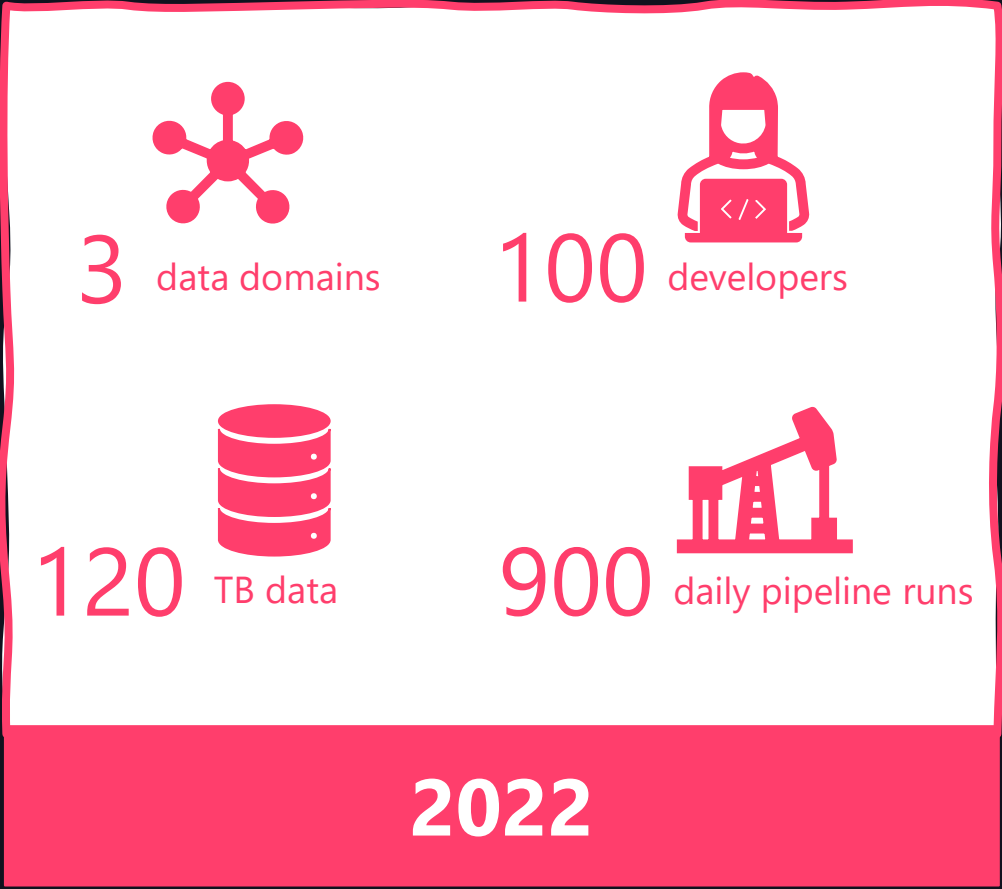
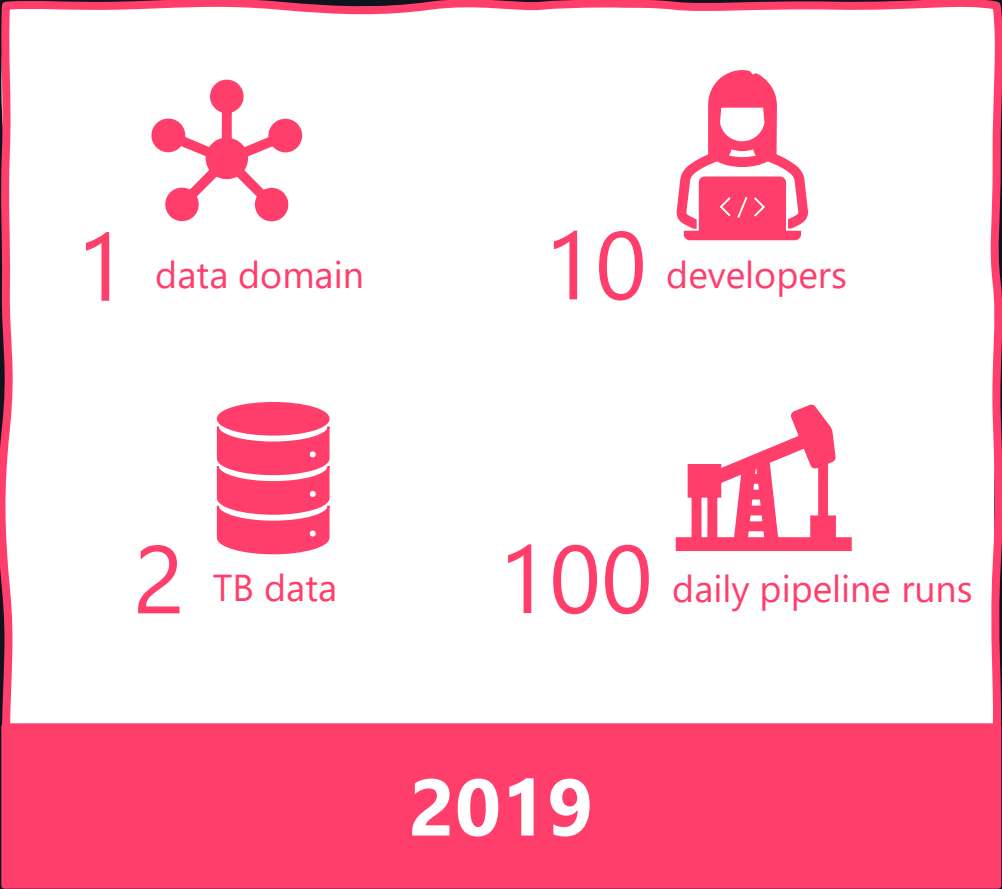


# Diageo's Data & Analytics Transformation

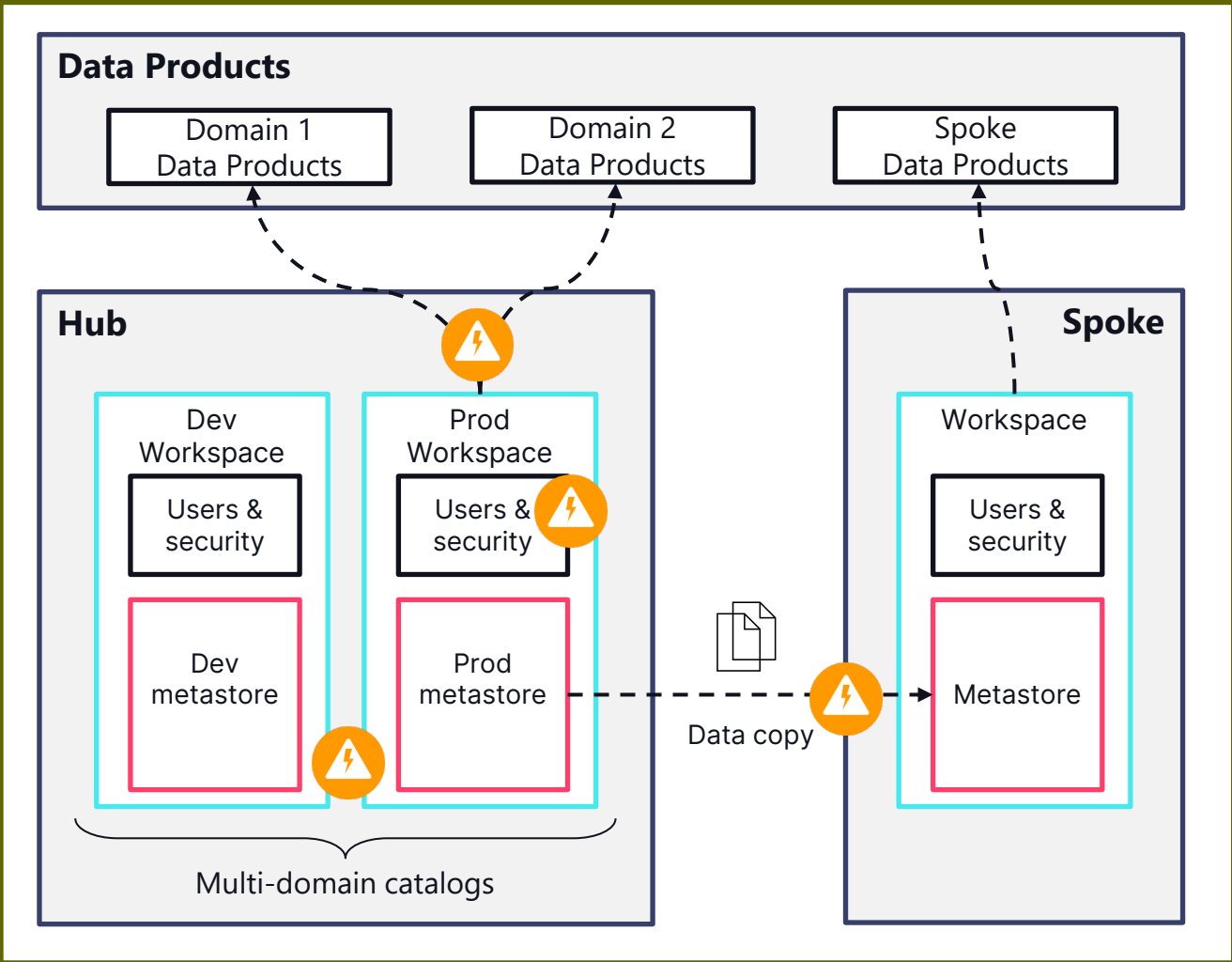
- Achieving many of our strategic business goals rely on decisions and measurement enabled by trusted data
- On a data and analytics journey since 2018 to become more data-driven & insight-led
- At the heart of the transformation world class approach to data delivery & management through our Diageo Data Hub (DDH), one of our crown jewels
- DDH is our enterprise data platform architecture delivering centralised, harmonised & curated data of over 200 data sources

# Why Unity Catalog?

# From launch in 2019 to 2022, Diageo's Data Hub saw high adoption and rapid growth



# By 2022 the Hub was facing issues in keeping up with demand



Difficulty sharing and understanding data



Fragmented and imprecise user permissions



Proliferation & duplication of views for downstream consumption



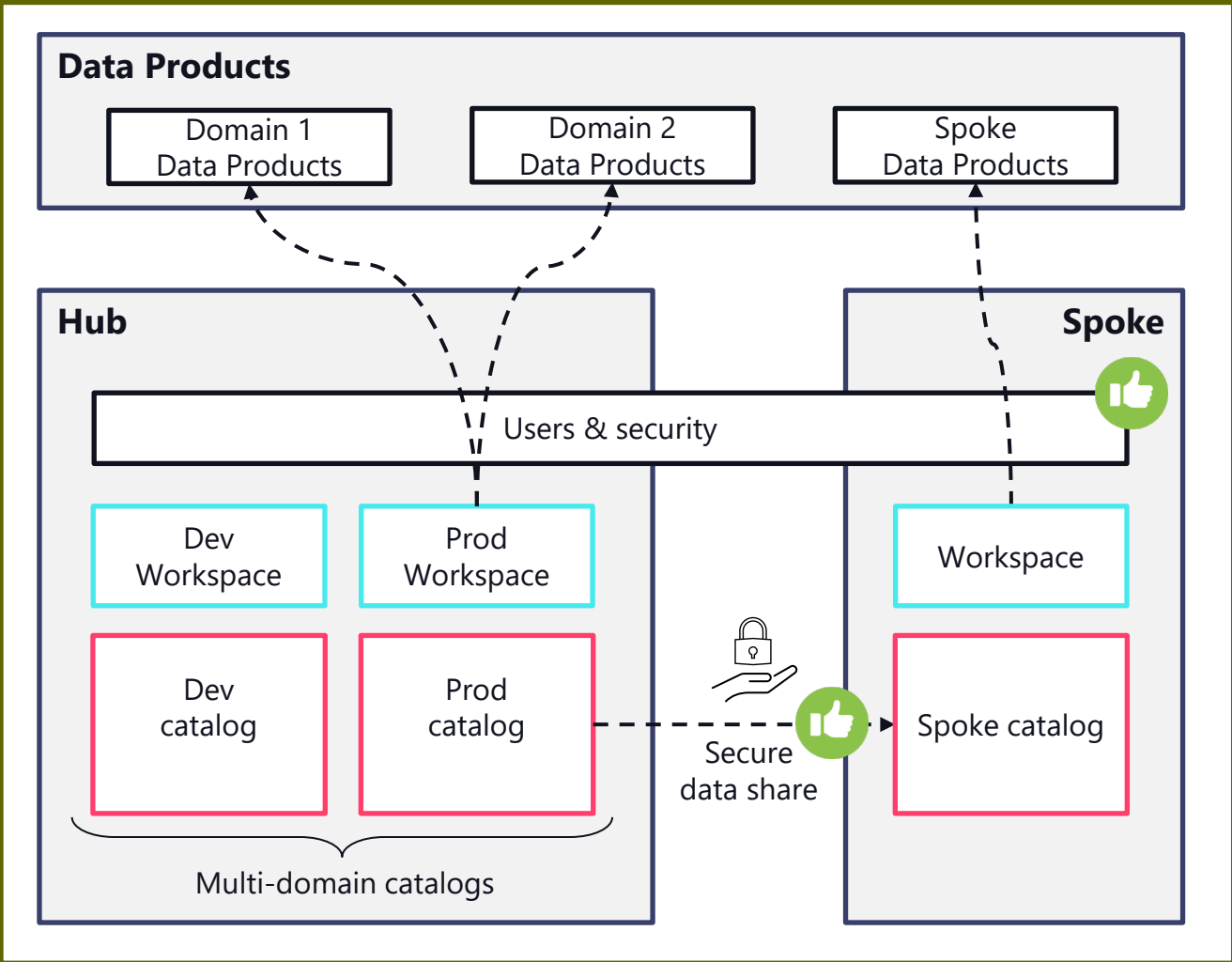
Congested workspace for data consumers



Bottlenecks in releasing change



# Migrating our workspaces to Unity Catalog offered immediate value to data consumers



Tighter controls and more precision with data sharing, critical to manage PII in our platform



Centralised security management



Vastly improve ability to find and understand data within platform



Accelerated workstream to rationalise our data landscape



Bottlenecks in releasing change



Congested workspace for data consumers



Lineage reporting enabled





# How did we implement Unity Catalog?

# Diageo set a clear goal to deliver the migration without disrupting Data Hub activities



Target zero downtime for end users



Minimise reliance on developer teams



Ensure smooth transition for live assets



Build and maintain stakeholder confidence

# Diageo designed and adopted a staged approach to Unity Catalog migration



**Prepare**



**Roll out**



**Switch**



**Accelerate**

**Unity Catalog**

Prototype deployment

Copy of Hive metastore, read-only, synced weekly

Data Asset dev work switches to UC

All dev work and product access now via UC

**Hive metastore**

Data Asset dev work and product access continue

Data Asset dev work and product access continue

Table structures frozen; access continues for unchanged products

Access revoked and resources re-allocated



# Prepare: "Shift Left" approach ensured maximum readiness

## Prepare

- Enabled Unity Catalog at Diageo Level
  - ✓ Enabled Unity Catalog in selected Databricks workspaces
  - ✓ Reviewed & refined permissions & access levels across all Diageo workspaces
- Created Data Hub specific Catalogs
  - ✓ Utilized 3-level namespace for improved naming conventions in Data Hub
  - ✓ Confirmed approach to splitting UC
- Developed & Deployed "Alpha" version
  - ✓ Assessed capability by creating, deploying & consuming an object in a Catalog
  - ✓ Granted access for select users for exploration & early feedback

# Roll Out: "Beta" Version availability provided early feedback and adoption

## Roll out

- Developed & Deployed "Beta" version of Data Hub Catalogs
  - ✓ Published all Hive metastore objects with improved naming conventions
  - ✓ Synced code weekly to keep the Catalogs updated
  - ✓ Completed automated testing
- Rolled out to all Stakeholders
  - ✓ Enabled Read-Only access to Catalogs for exploration
  - ✓ Encouraged stakeholders to adapt to naming conventions and features
- Encouraged early adoption
  - ✓ Identified downstream candidates to switch over to Unity Catalog Beta version
  - ✓ Established lines of communication to offer assistance, assurance & gather feedback
  - ✓ Incorporated lessons learned

# Switch: One-time migration ensured clean cutover to Unity Catalog

## Switch

- Published cutover plan along with impact & timelines
  - ✓ Create detailed cutover & rollback playbook as best practice
  - ✓ Ensured all dependencies/blockers are resolved
- Performed cutover to Unity Catalog
  - ✓ Repointed all CI/CD pipelines to deploy to Unity Catalog in all environments
  - ✓ One-time migration of all pipelines using Hive Metastore to create Derived objects
- Provided necessary support for smooth downstream transition
  - ✓ Hive Metastore available as “Read-Only”, to be progressively outdated
  - ✓ Downstream product teams to completely switch over to Unity Catalog

# Accelerate: Unity Catalog live, benefits unlocked

## Accelerate

- Transitioned to BAU
  - ✓ Published “How to” guides to follow industry best practices & Data Hub standards
  - ✓ Provided Handover to Operations team following existing process
- Deprecated Hive Metastore
  - ✓ Revoked access & reallocated resources
  - ✓ Stopped support & sunset Hive Metastore
- Utilized Unity Catalog features
  - ✓ Secure data sharing enabled faster spoke onboarding, accelerating market delivery

# The migration was delivered on time and with zero downstream impact



Target zero downtime for end users



Minimise reliance on developer teams



Ensure smooth transition for live assets



Build and maintain stakeholder confidence



# The migration was delivered on time and with zero downstream impact



Target zero downtime for end users

600 users  
transferred



Minimise reliance on developer teams



Ensure smooth transition for live assets



Build and maintain stakeholder confidence

# The migration was delivered on time and with zero downstream impact



Target zero downtime for end users

600 users transferred



Minimise reliance on developer teams

500 tables migrated



Ensure smooth transition for live assets



Build and maintain stakeholder confidence

# The migration was delivered on time and with zero downstream impact



Target zero downtime for end users

600 users transferred



Minimise reliance on developer teams

500 tables migrated



Ensure smooth transition for live assets

750 regression tests run



Build and maintain stakeholder confidence

# The migration was delivered on time and with zero downstream impact



Target zero downtime for end users

600 users transferred



Minimise reliance on developer teams

500 tables migrated



Ensure smooth transition for live assets

750 regression tests run



Build and maintain stakeholder confidence

Fortnightly meetings



# What did we learn from the migration?

## Delivery approach

- Create detailed map of downstream dependencies
- Adopt automated testing
- Use agile delivery principles
- Remember ongoing platform enhancements
- Leverage downstream teams where sensible
- Use UCX

## Change management

- Lead with benefits and business value
- No such thing as too early or too much communication
- Avoid crunch periods
- Consider change appetites

# Benefit Realization of Unity Catalog Implementation?

# Diageo has realized multiple benefits since implementation

Feature	How We Are Using It	Value Realization
<b>Access Control</b>	Using UC's Centralised Access Control Framework, Implementation step-change <b>data access control governance</b>	<ul style="list-style-type: none"> <li>Managing Data <b>access control is 90% faster</b></li> <li>Speed to <b>data democratisation has improved by 60%</b></li> <li>Data governance <b>controls, audit management of data access is 75% faster</b></li> </ul>
<b>Lineage</b>	<b>Leveraged natural generation of data lineage in UC</b> , used to democratize data discovery, sharing and understanding. Also includes integration with Enterprise Data catalogue	<ul style="list-style-type: none"> <li>Scale out of data discovery and analysis self-service resulting in <b>~50% time saving on central team</b></li> <li>Expected to <b>elevate the level of data literacy in the organisation</b></li> </ul>
<b>Metadata Management &amp; Lakehouse Monitoring</b>	Built capabilities around metadata management of UC for <b>data observability, searchability, monitoring and data quality</b>	<ul style="list-style-type: none"> <li><b>Increased data quality and trust in data</b> across our major programs</li> <li>The capability aims to <b>improve data operations, support and response times up to 4x</b></li> <li><b>From Reactive to Proactive data remediation</b> through alert automation</li> </ul>
<b>Data Sharing</b>	<b>Delta sharing enablement</b> with UC, complemented with the robustness of Access Control mentioned above. Store once, share multiple times	<ul style="list-style-type: none"> <li>For projects which deal with 5-10 data assets, <b>through Delta sharing we have been able to improve time to market by 15x</b></li> <li>Estimated <b>reduction of support cost up to 3 FTE per month for similar data assets</b></li> </ul>



# DATA+AI SUMMIT