ENTERPRISE DATA MESH INGESTION

# DECISIONS AT MISSION SPEED

Automating Real-Time Data Pipelines with Qlik Replicate

DEPARTMENT OF THE AIR FORCE
72 ABW • Tinker AFB

PRESENTED BY

**Mr. Neil Goldstein**

Program Manager, Tinker Center of Automation

72 ABW/CCPD

# THE CHALLENGE

**BLUF: DAF's ambition to weaponize Data as a strategic asset is at risk.**



## THE GROUND TRUTH

- **Excel/CSV**
- **Hand-tooled scripts**
- **Robotic Process Automation**

## THE BOTTLENECK

- Every source is custom work
- "Pull data out, then **piece it back together**"
- **Demand** outpaces Capacity

## CORE ORGANIZATIONAL R

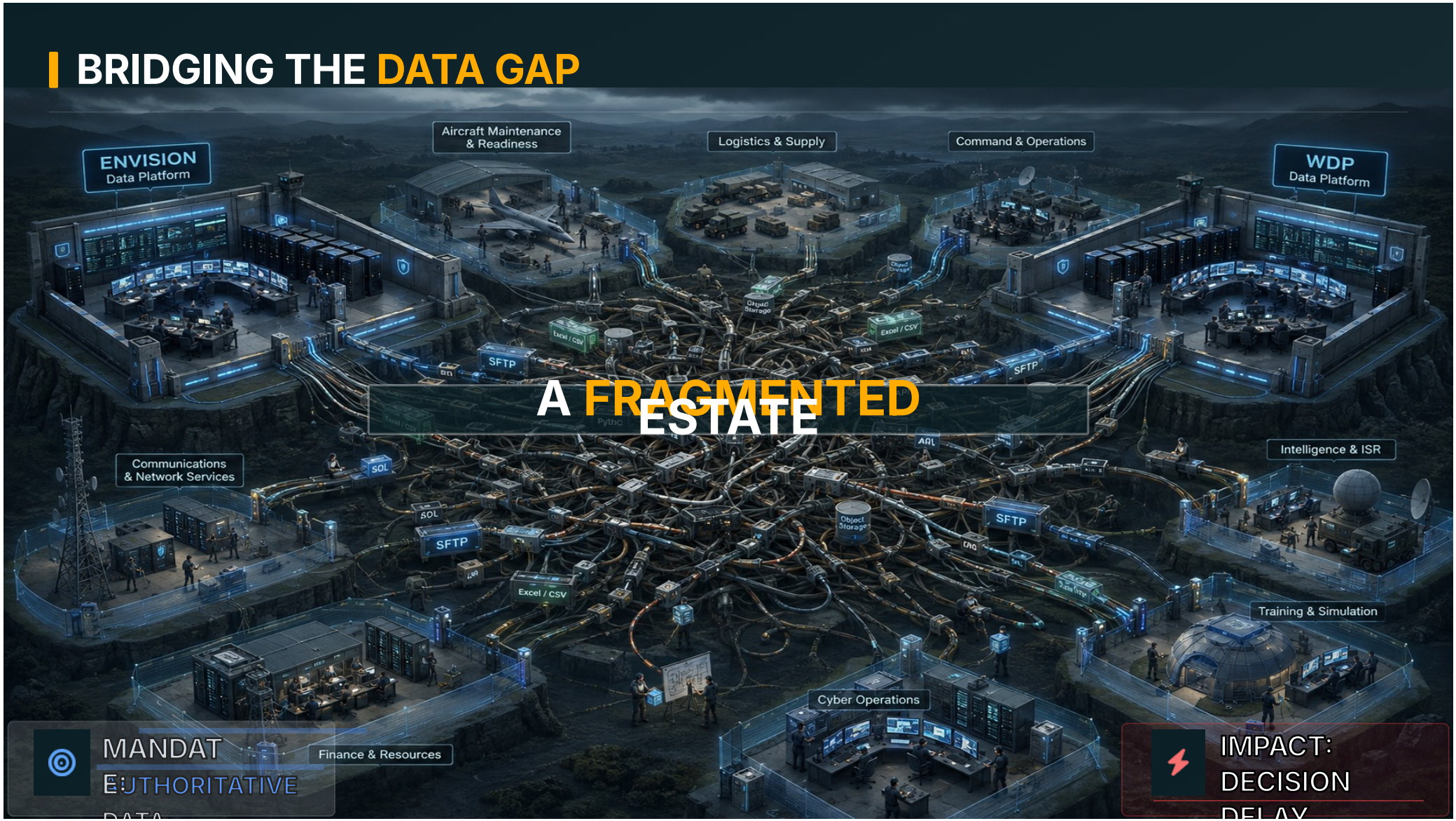
### Data Source Availability

- Delayed onboarding
- Unclear refresh tempo

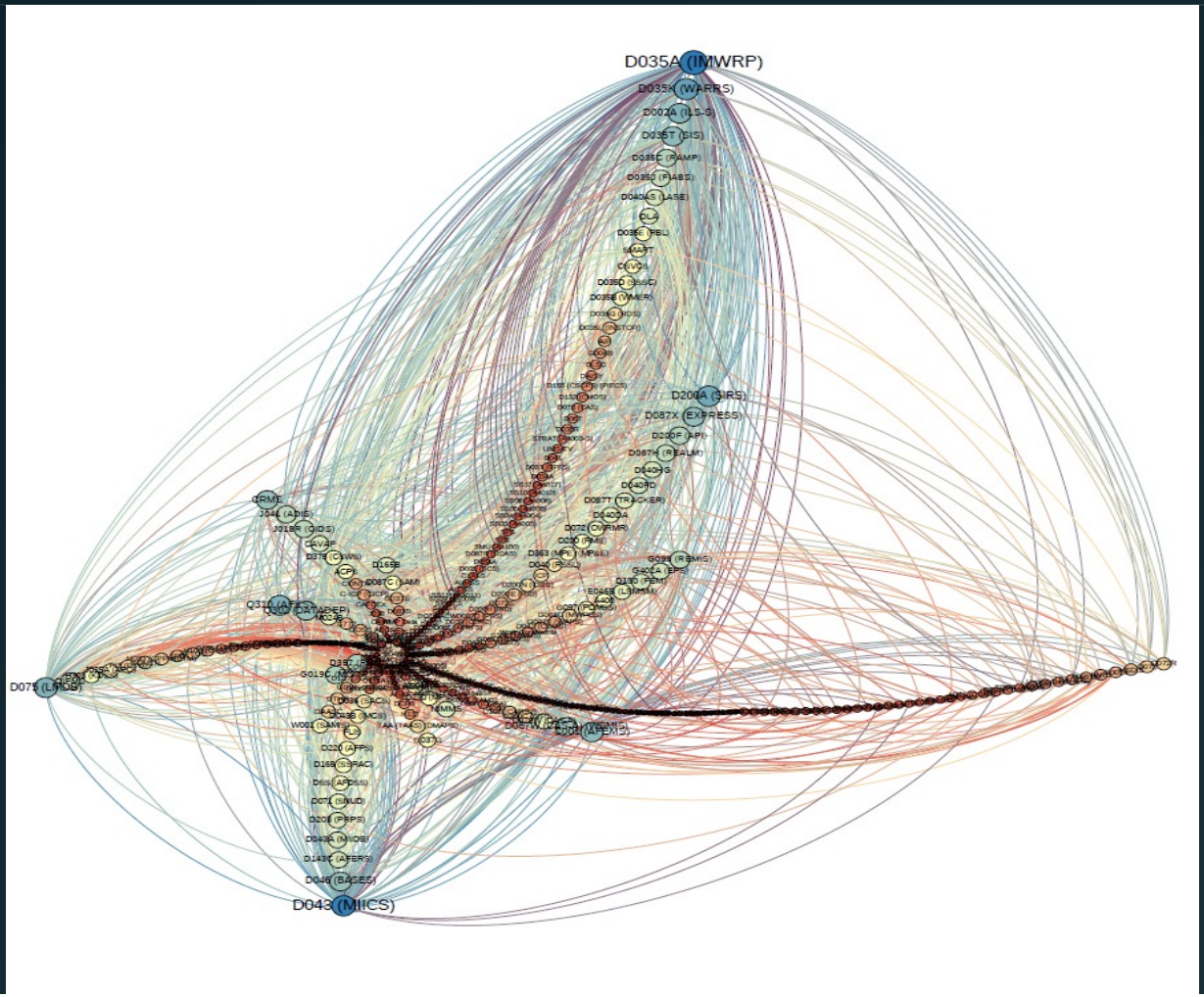
### Hostage to "Batch Scripts"

- Constant technical debt
- Maintained via daily heroics

# BRIDGING THE DATA GAP



# A GLIMPSE AT THE PIPES



# THE HUMAN COST & WORKFORCE ROI

## TODAY
## EXPERTS MOVE DATA

Mission experts

Lose DAYS

Being Data Plumbers.

THE SUNK COST

**Thousands of hours**
**Lost to manual scripting.**

## THE GOAL
## EXPERTS USE DATA

Pipeline automation

Gives WARRIORS

Hours back on

Mission.

THE STRATEGIC ROI

**AI • Modeling • Analysis**

# THE SOLUTION: CDC VS. BATCH




## CAPABILITY COMPARISON

**LATENCY**

| | |
|---|---|
| CDC (Qlik) | Milliseconds |
| Legacy Batch | Hours / Days |

**SYSTEM IMPACT**

| | |
|---|---|
| CDC (Qlik) | Near-Zero |
| Legacy Batch | Severe |

**MECHANISM**

**CDC:** Real-time log reading.

**Batch:** Heavy full queries.

**IDEAL USE CASE**

**CDC:** Mission-Ready Data.

**Batch:** Historical Archiving.

# WE WEREN'T SHOPPING FOR SOFTWARE

**MISSION:** MOVE AUTHORITATIVE DATA INTO BLADE



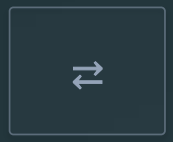
## S3 → S3

OBJECT STORAGE



## API / ODBC SFTP

CUSTOM INTERFACE



## SFG

SECURE FILE GATEWAY FILE TRANSFER

SELECTED

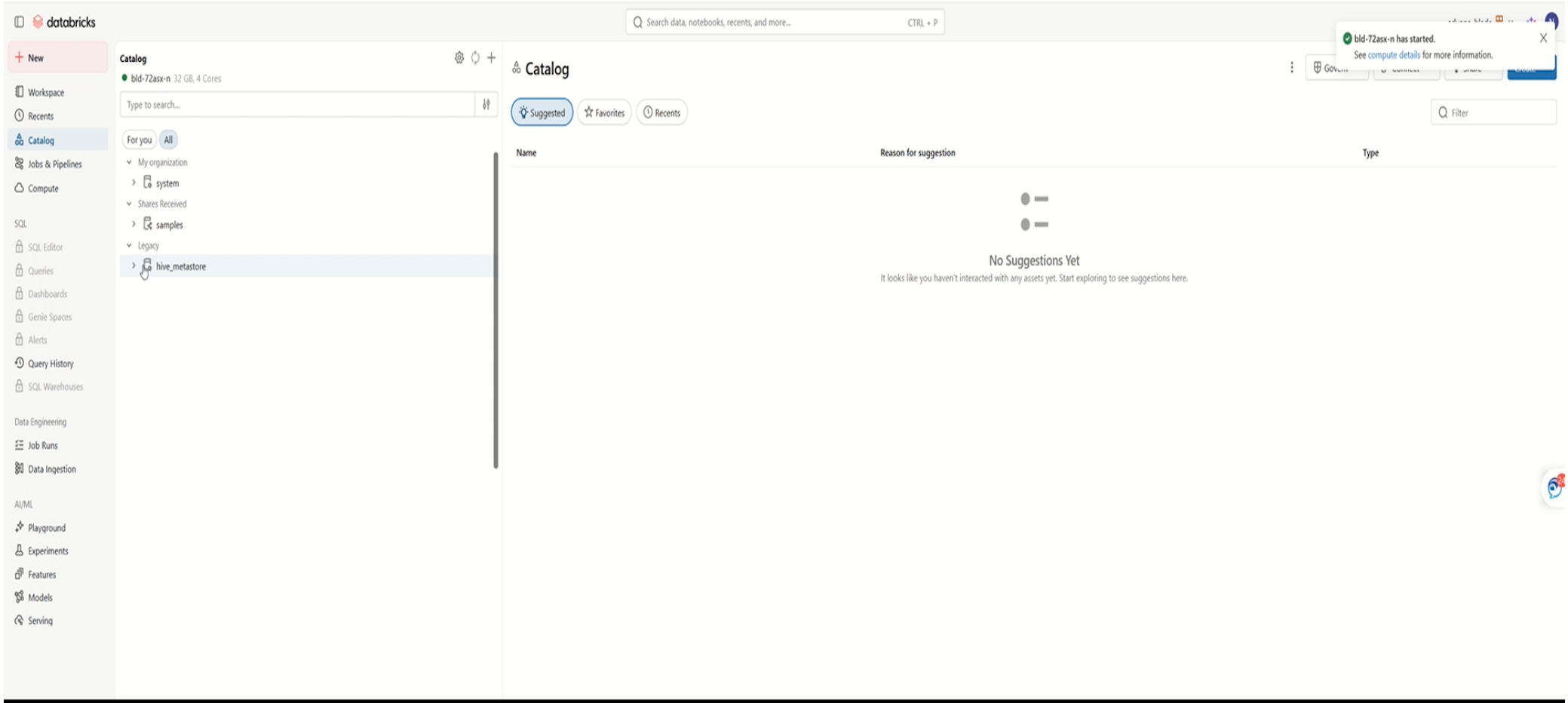


## QLIK REPLICATE

AUTOMATED CDC

**WE CHOSE THE PATH THAT DIDN'T REQUIRE ANOTHER BESPOKE DATA PIPELINE.**

# THE PROOF


databricks
Search data, notebooks, recents, and more... CTRL + P
bld-72asx-n has started.
See compute details for more information.
New
Workspace
Recents
Catalog
Jobs & Pipelines
Compute
SQL
SQL Editor
Queries
Dashboards
Genie Spaces
Alerts
Query History
SQL Warehouses
Data Engineering
Job Runs
Data Ingestion
AI/ML
Playground
Experiments
Features
Models
Serving
Catalog
bld-72asx-n 32 GB, 4 Cores
Type to search...
For you
All
My organization
system
Shares Received
samples
Legacy
hive_metastore
Catalog
Suggested
Favorites
Recents
Filter
Name
Reason for suggestion
Type
No Suggestions Yet
It looks like you haven't interacted with any assets yet. Start exploring to see suggestions here.

## THE FIRST SYSTEM TOOK TWO YEARS.

## THE SECOND TOOK AN AFTERNOON.

### PATHFINDER
**ANDON**

200K Records
2 YEARS
**Proved:**
- Authorization
- Architecture
- Integration
- Troubleshooting

### DEMO SYSTEM
**DCAST**

80 Million Records
30 Perfect Delta Tables Configured & Replicated
**ONE AFTERNOON.**

## THE DIFFERENCE WASN'T JUST THE TOOL. IT WAS THE PLAYBOOK.

# WHAT WE HAD TO UNLEARN

LESSONS FROM THE **TWO-YEAR PILOT**

## AN OPTION IS NOT A CAPABILITY

### AVAILABLE ON PAPER
### OPTION

- LISTED
- AVAILABLE
- TECHNICALLY POSSIBLE



NO ESTABLISHED PLAYBOOK

### OPERATIONALIZED
### CAPABILITY

- AUTHORIZED
- SECURE
- SUPPORTED
- REPEATABLE

**THE TECHNOLOGY EXISTED. THE OPERATING PATTERN DID NOT.**

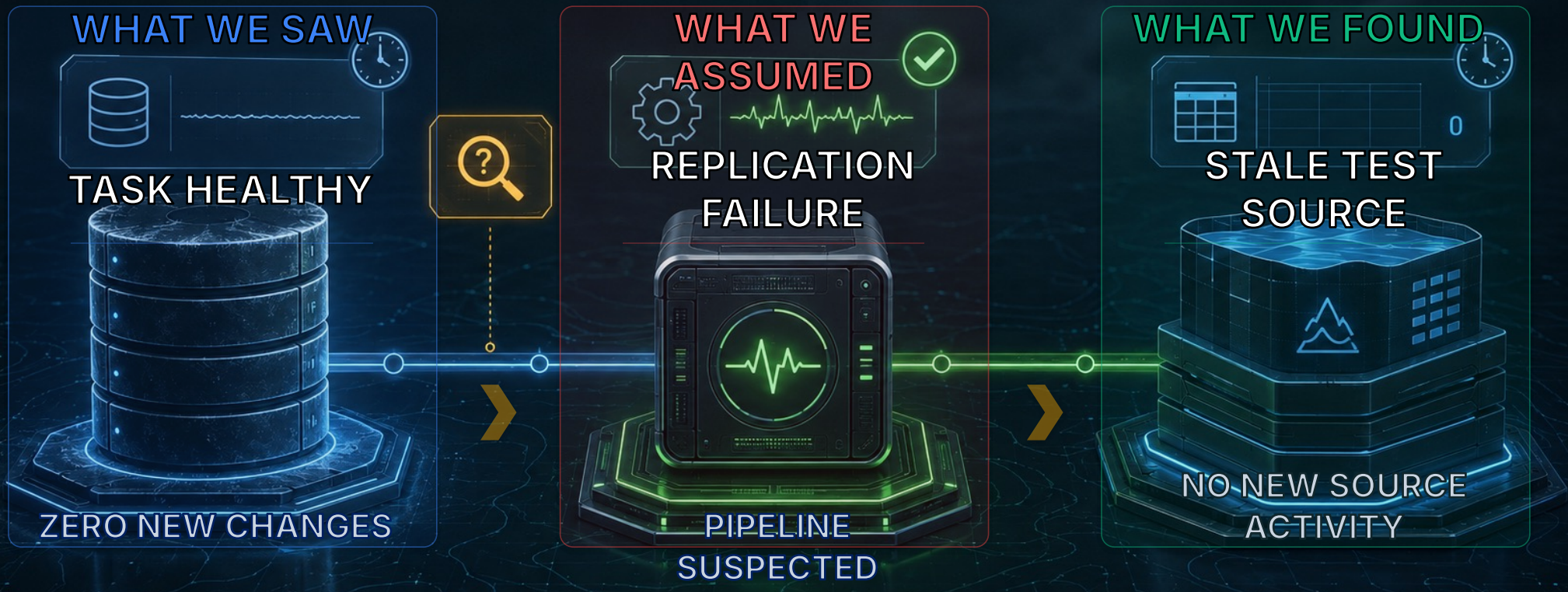
# BUILDING THE PLAYBOOK




**EVERY DEAD END BECAME A DOCUMENTED PATH FORWARD.**

TWO YEARS TO GO FROM "POSSIBLE" TO "PROVEN"

# THE PIPELINE WAS WORKING. OUR TEST WASN'T.




**VALIDATE THE SYSTEM — NOT JUST THE TOOL.**

# IDENTITY DEPENDS ON WHERE YOU STAND

## WORKLOAD INSIDE AWS




WORKLOAD

IAM ROLE

TEMPORARY CREDENTIALS

## WORKLOAD OUTSIDE AWS



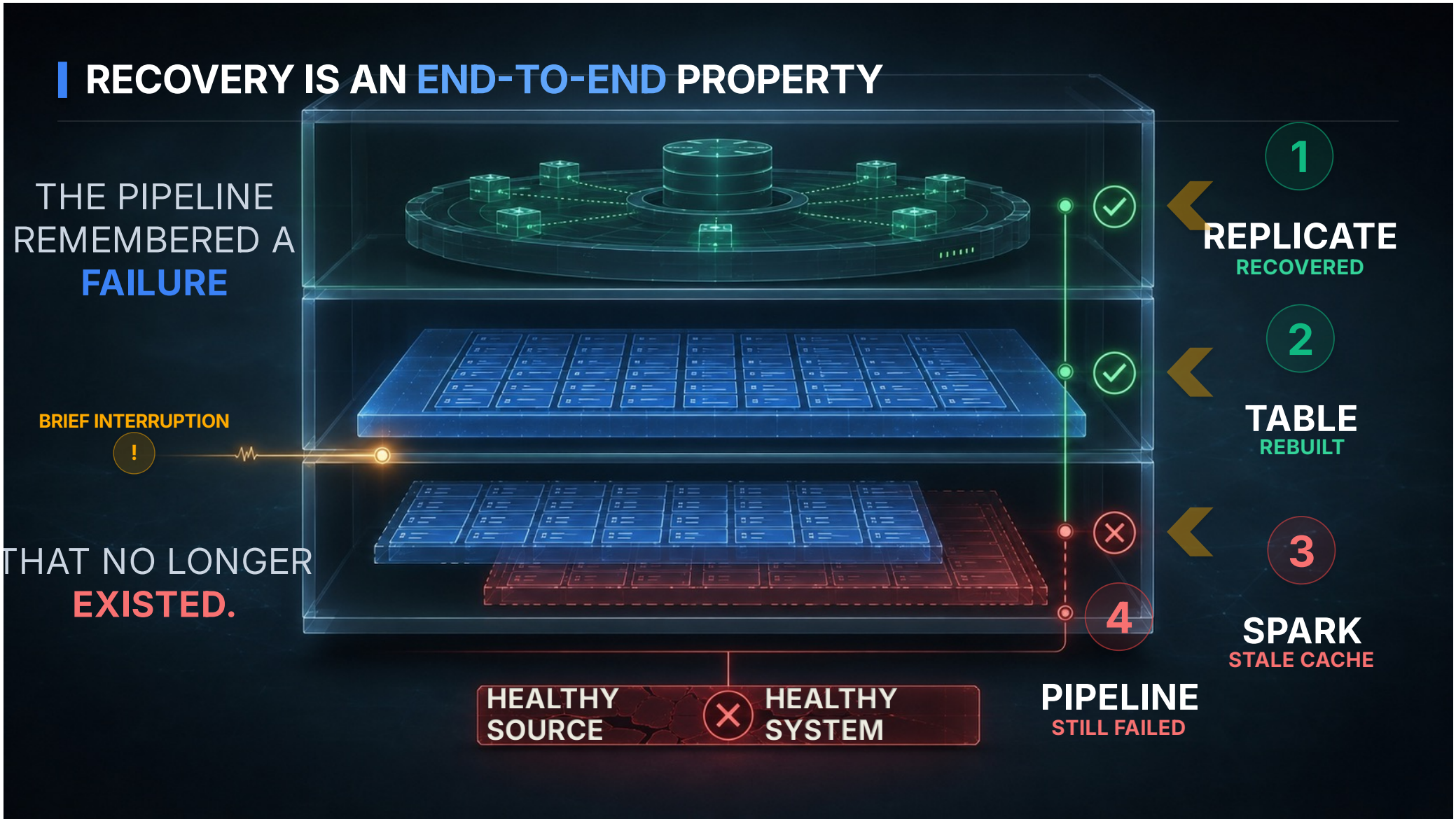
REPLICATE SERVER

TRUST BOUNDARY

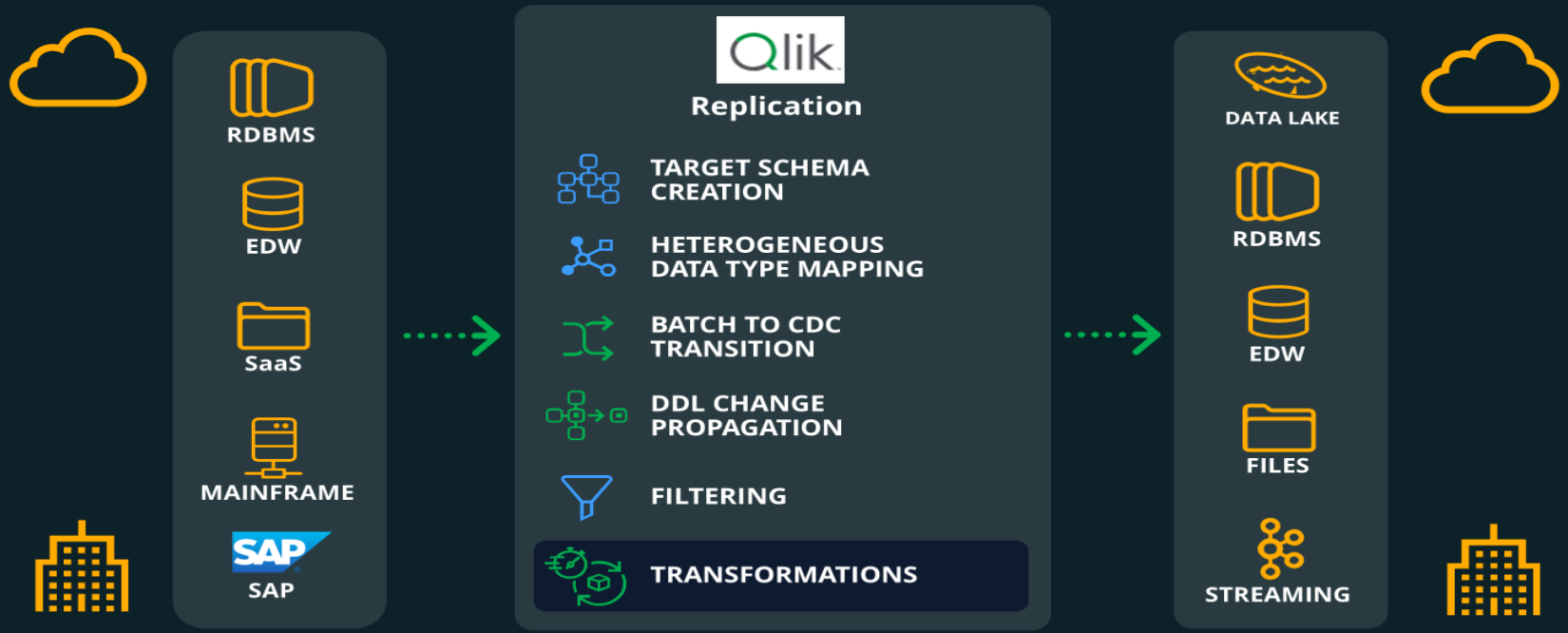
AWS GOVCLOUD

DIFFERENT TRUST PATH

IDENTITY DESIGN CHANGES WITH WORKLOAD LOCATION.


RECOVERY IS AN END-TO-END PROPERTY
THE PIPELINE REMEMBERED A FAILURE
BRIEF INTERRUPTION
!
THAT NO LONGER EXISTED.
1
REPLICATE
RECOVERED
2
TABLE
REBUILT
3
SPARK
STALE CACHE
4
PIPELINE
STILL FAILED
HEALTHY SOURCE
HEALTHY SYSTEM

# WHAT REPLICATE ACTUALLY DOES



# ONE CHANGE, END TO END




**SOURCE RDBMS**
ONE ROW CHANGES

**TRANSACTION LOG**
DATABASE ALREADY RECORDED IT

**QLIK CAPTURE**
READ THE CHANGE

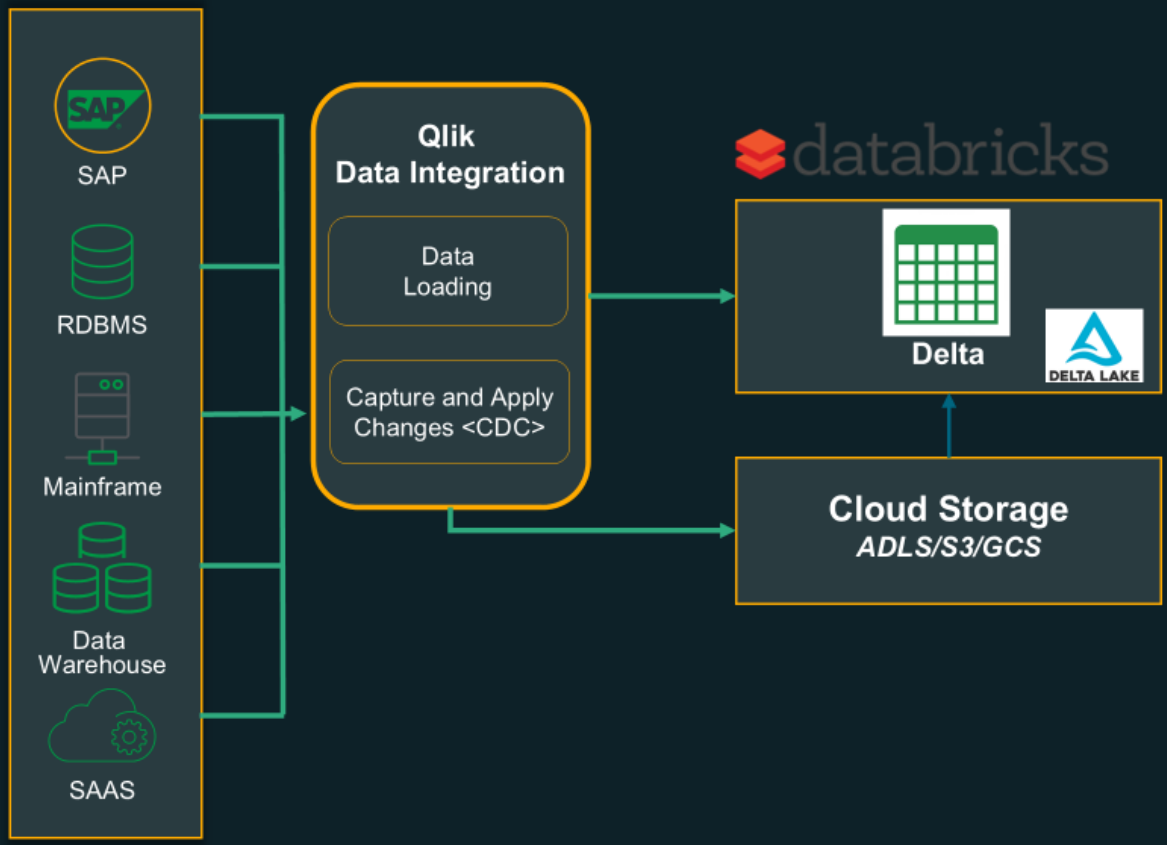
**DATABRICKS WRITE**
APPLY USING DATABRICKS WORKER

**DELTA TABLE**
WRITE THE DELTA

DON'T RE-READ THE DATABASE. READ THE RECORD OF WHAT CHANGED.

# WHAT RIGHT LOOKS LIKE

- Direct delivery to **Databricks Delta**
- Replication process –
  1. Create Table USING DELTA
  2. Full load file(s) to cloud storage
  3. Loaded to Delta via COPY INTO select from "storage_file"
  4. CDC External table created (attrep_changes)
  5. CDC file(s) to cloud storage
  6. Change applied with SET based
     1. DELETE
     2. MERGE



# AUTOMATION IS ABOUT MORE THAN MOVING ROWS.

1 **STARTS COMPUTE**

Invokes target resources



2 **BUILDS TABLES**

Creates target definitions



3 **AUTOMATES LOAD ORDER**

Manages dependencies

1→2

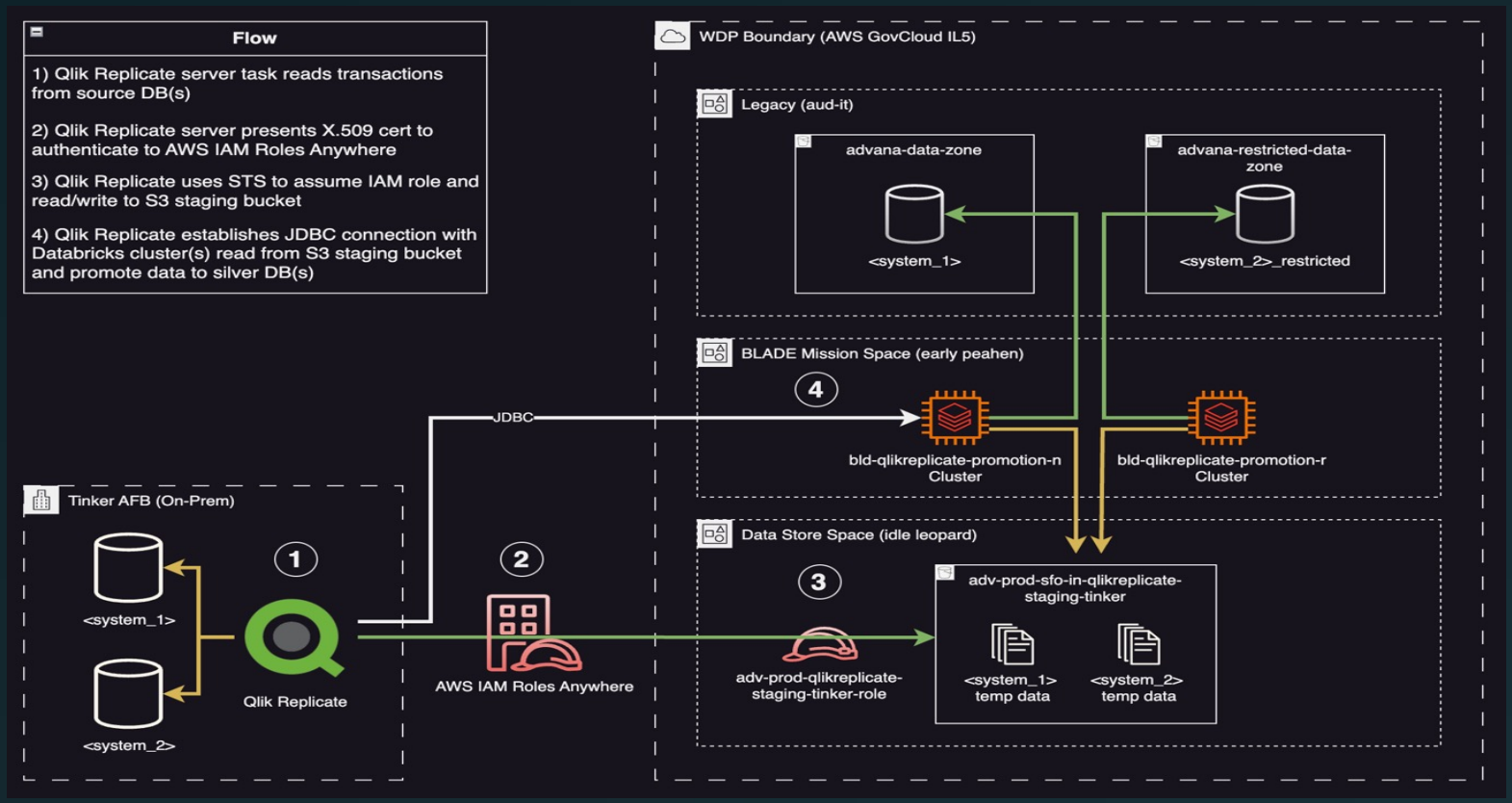
4 **HANDLES EDGE CASES**

Source readiness & reloads



LESS HAND-BUILT DATA ENGINEERING.

# AUTOMATION IS ABOUT RAISING THE BAR.




MORE ON-DEMAND MISSION CAPABILITY.

# THE ZERO TRUST PIPELINE

**BLUF: Maximum security accelerates speed to data.**

1

## EXTRACT
### Agentless CDC

- TLS-ENCRYPTED stream reads transaction logs
- NO intrusive agent on source system
- STREAMLINES Risk Management

2

## AUTHENTICAT
### Ephemeral Identity

- ZERO standing credentials
- NO custom auth code needed
- MODERN short-lived scoped STS tokens

3

## GOVERN
### Unity Catalog

- Streams REAL TIME changes
- Directly into DELTA TABLES
- Via integration with UNITY CATALOG

# RESILIENCY & SCHEMA DRIFT



# FROM PATHFINDER TO ENTERPRISE

**SCALING** THE PLAYBOOK

# THE VALLEY OF DEATH



# PILOT TO PRODUCTION PLAYBOOK

1

## STANDARDIZE

**Reusable intake patterns**

- TEMPLATES
- SOURCE READINESS
- IaC

ONBOARDING KIT

2

## ALIGN

**Clear ownership**

- APPROVALS
- IDENTITY
- ROLES

AUTHORITY MAP

3

## OPERATE

**Run it like a service**

- VALIDATE
- OBSERVE
- RECOVER

OPS RUNBOOK

THE NEXT SOURCE STARTS WITH THE PLAYBOOK – NOT FROM ZERO.

## THE DESTINATION: DATA AT MISSION SPEED

# 2026 DAF DATA STRATEGY

## AUTHORITATIVE DATA

## AT THE SPEED OF MISSION

DISCOVERABLE CONSUMABLE AUTHORITATIVE SECURE

"A department-wide north star for data readiness, sharing, and decision advantage."

SOURCE: DEPARTMENT OF THE AIR FORCE DATA STRATEGY • 17 APR 2026

# DAF DATA ON DEMAND

"A FOUNDATION OF POLICY, GOVERNANCE & STANDARDS — PLUS A THIN ENTERPRISE TECHNOLOGY LAYER."




**DOMAIN-OWNED DATA. ENTERPRISE-SCALE DISCOVERY & SHARING.**

SOURCE: DAF DATA ON DEMAND BBP • 6 JAN 2026

# WHAT A DATA MESH ACTUALLY MEANS



## DOMAIN OWNERSHIP

Experts closest to the data own readiness.

## FEDERATED DISCOVERY

One enterprise front door.
Distributed data.

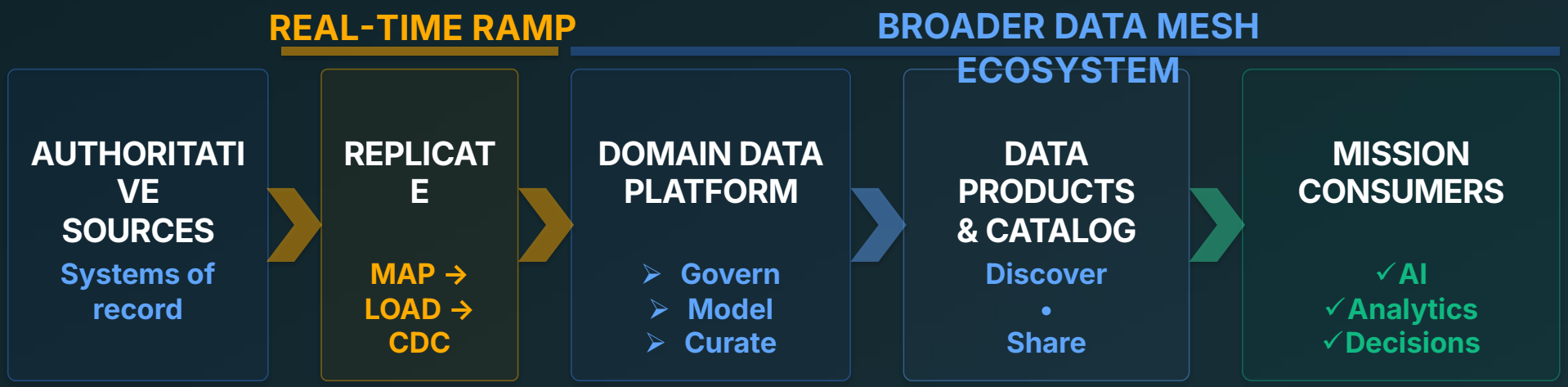
## ENTERPRISE GOVERNANCE

Common standards.
Security. Interoperability.

**KEEP OWNING YOUR DATA.**
**STANDARDIZE SECURE SHARING.**

# WHERE REPLICATE FITS

*A PRECISE ROLE, INSIDE A MUCH LARGER ENTERPRISE DATA FABRIC.*




**A HIGH-SPEED ON-RAMP INTO THE DATA MESH.**

# WHAT THIS PILOT CONTRIBUTED

## WE PROVED IT

**80M RECORDS**

**< 1 HOUR LOAD**

**< 1 SECOND UPDATE**

Mission-speed CDC.
All day, every day.

## WE SECURED IT

**IDENTITY**

**SECURITY**

**RECOVERY**

**SOURCE READINESS**

Cutting-edge capabilities.
Zero Trust by design.

## WE SCALED IT

**200K → 2 YEARS**

**80M → 1 AFTERNOON**

Pathfinding cost once.
Reuse thereafter.

**DATA SHOULD MOVE AT THE SPEED OF APPROVAL — NOT CUSTOM ENGINEERING.**