

# Redesigning Kaizen's Cloud Data Lake for the Future



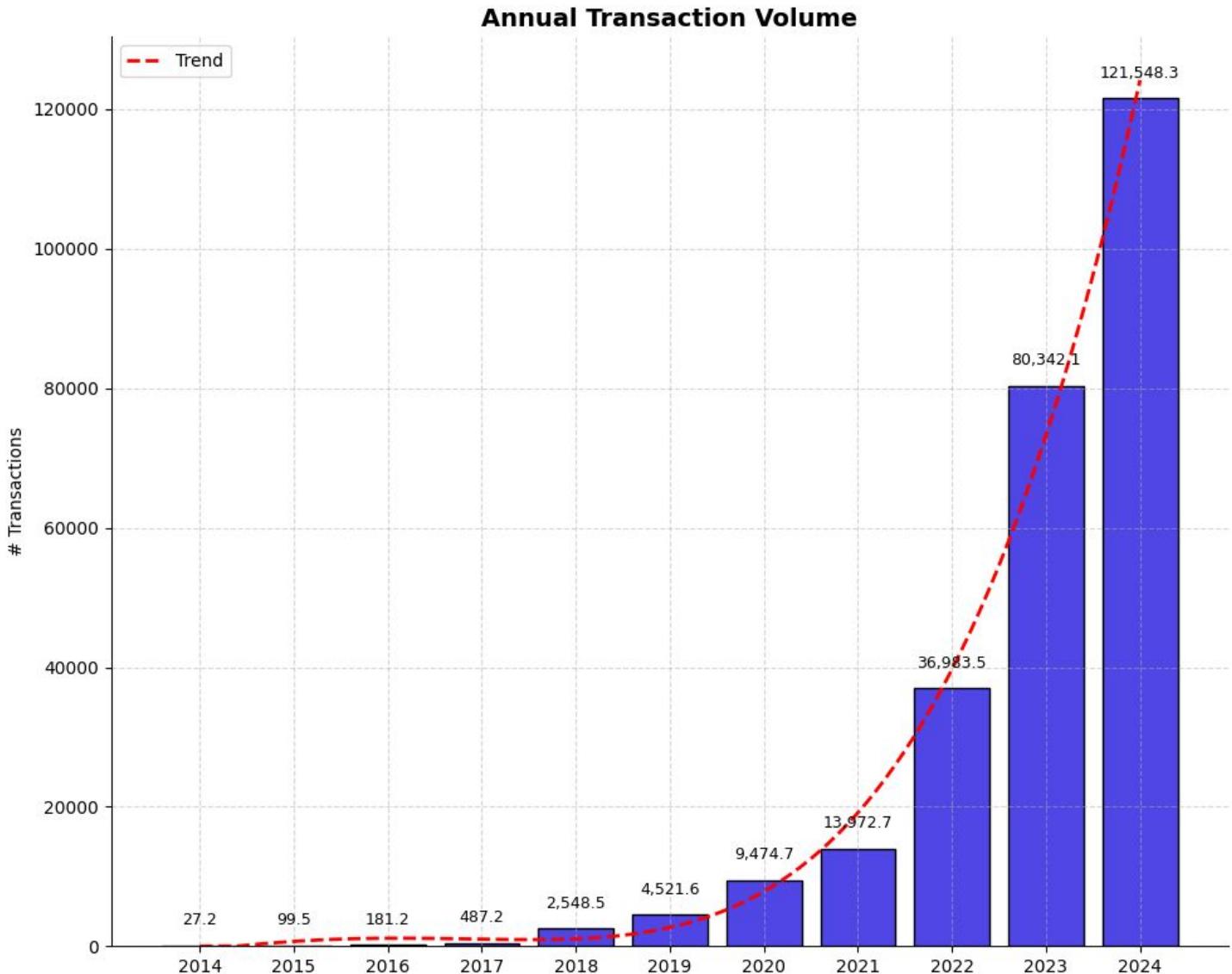


- GameTech company
- Betano Brand in 18 markets
- 13M+ active users
- 1.5M+ Gaming events per year



# Data in Kaizen

- **2 PetaBytes** in Data Lake
- Lakehouse Azure DBx
- Daily: ~400M Tx
- Minute: ~700K Tx (peak)
- From 27M to **121B in 10 years**  
(+4,500%)



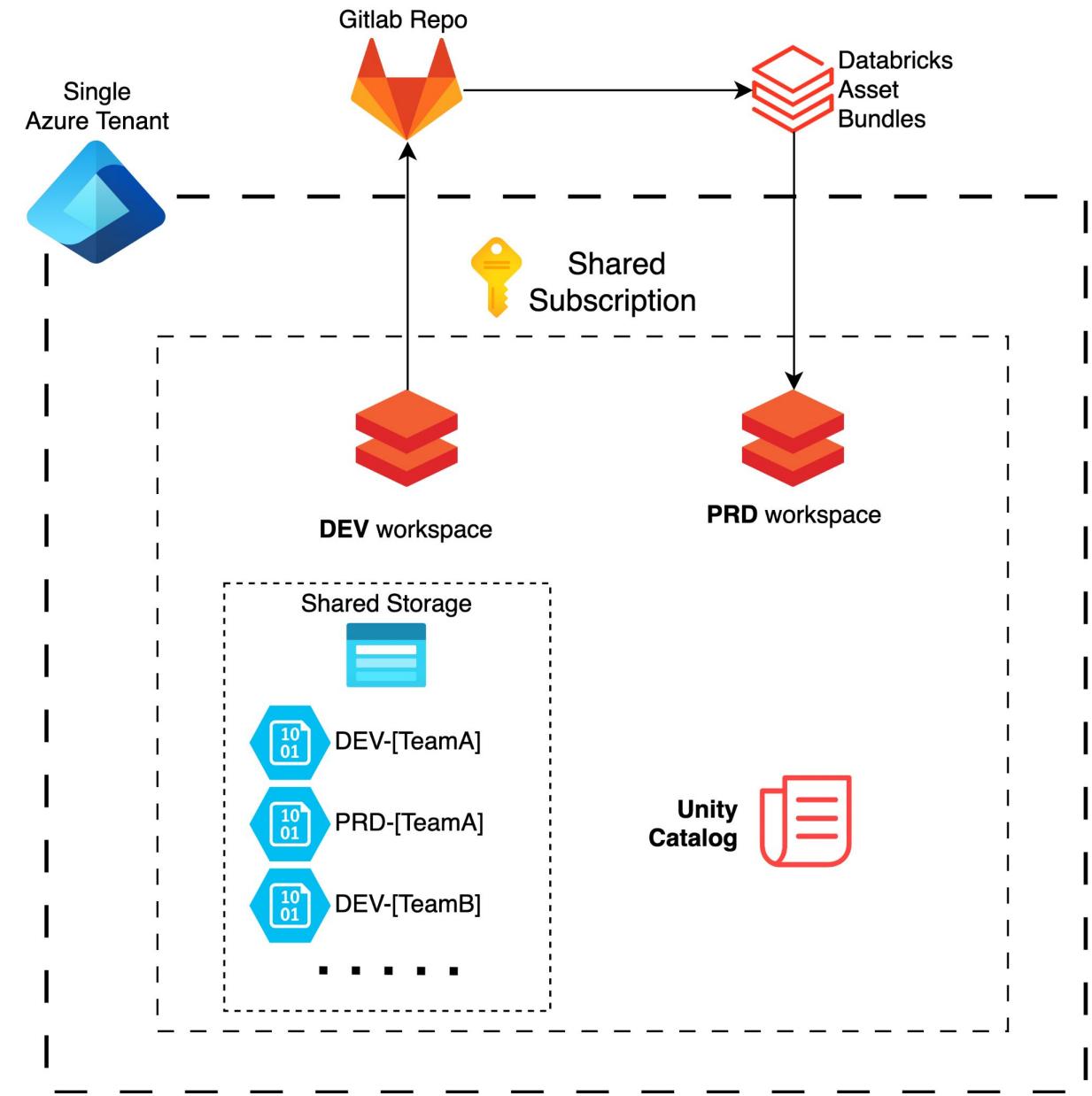
# Cloud Architecture



# Old Architecture

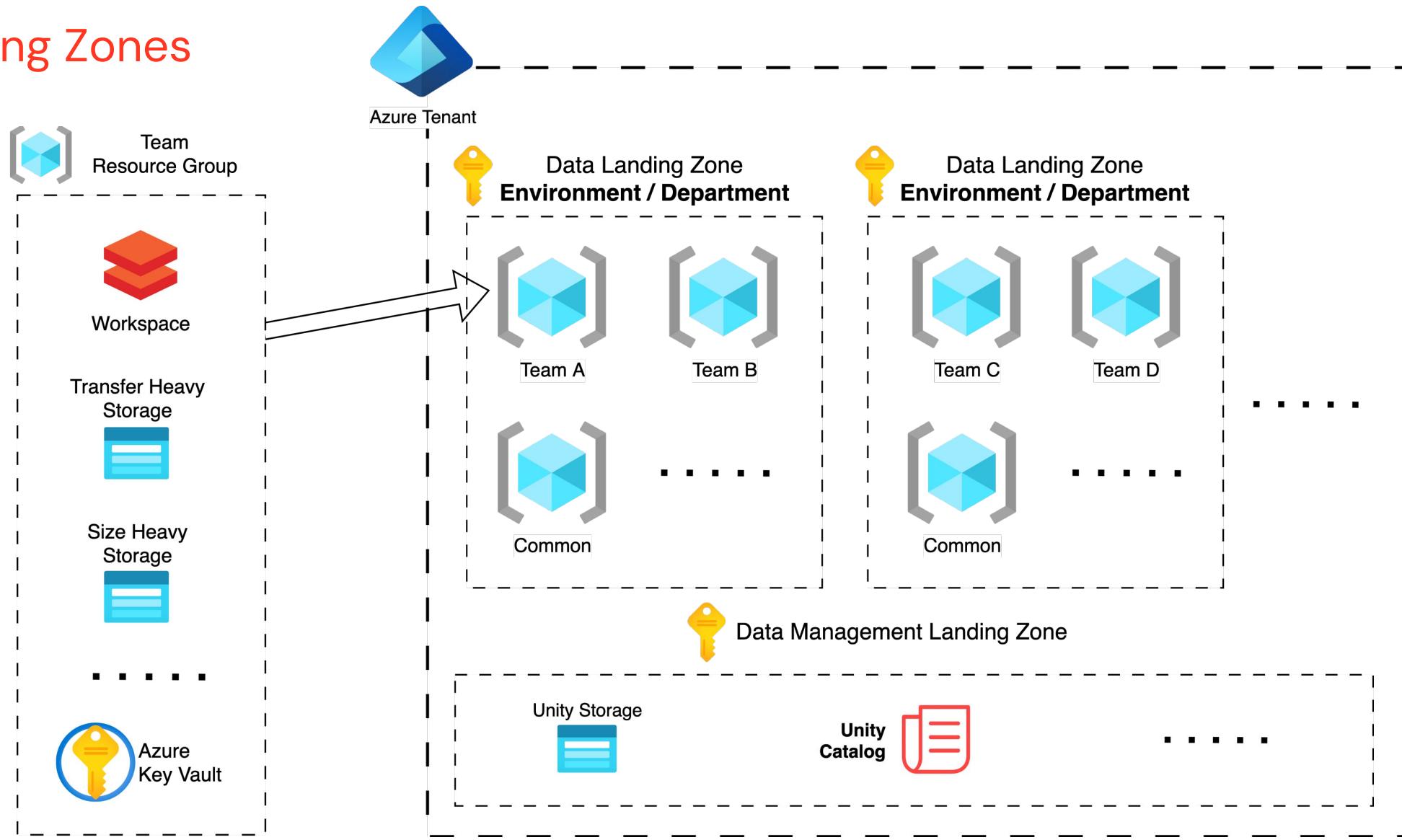
## Challenges

- Common Quotas
- Single Storage Type
- Manually Created Resources
- FinOps
- **Difficult to Scale**



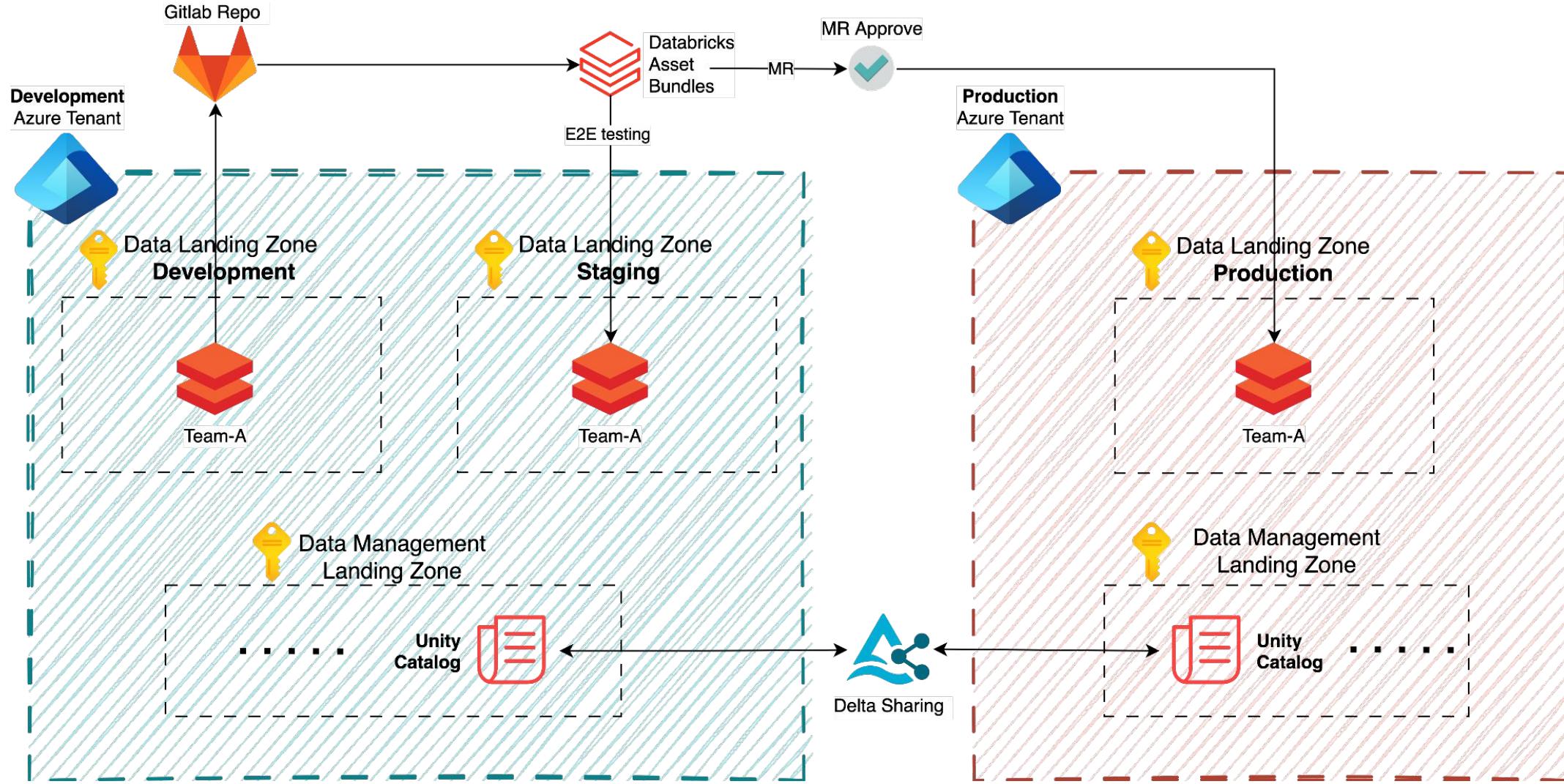
# New Architecture

## Landing Zones



# New Architecture

## Environments



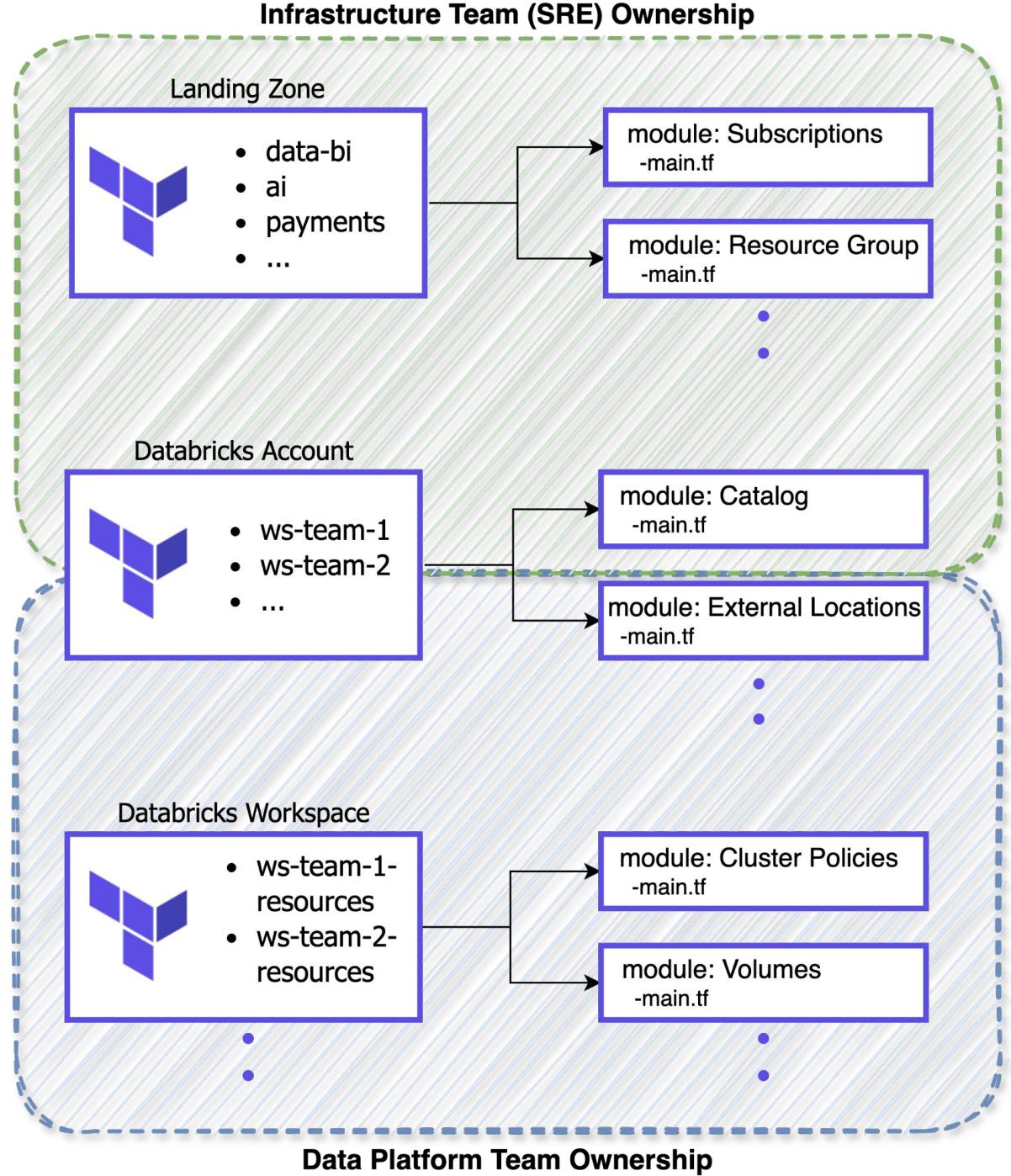
# Migration



# Migration

## Infrastructure

- Terraform Modules
- Abstraction
- Reduce Change Amplification
- Team boundaries
- Self serve infrastructure



# Migration

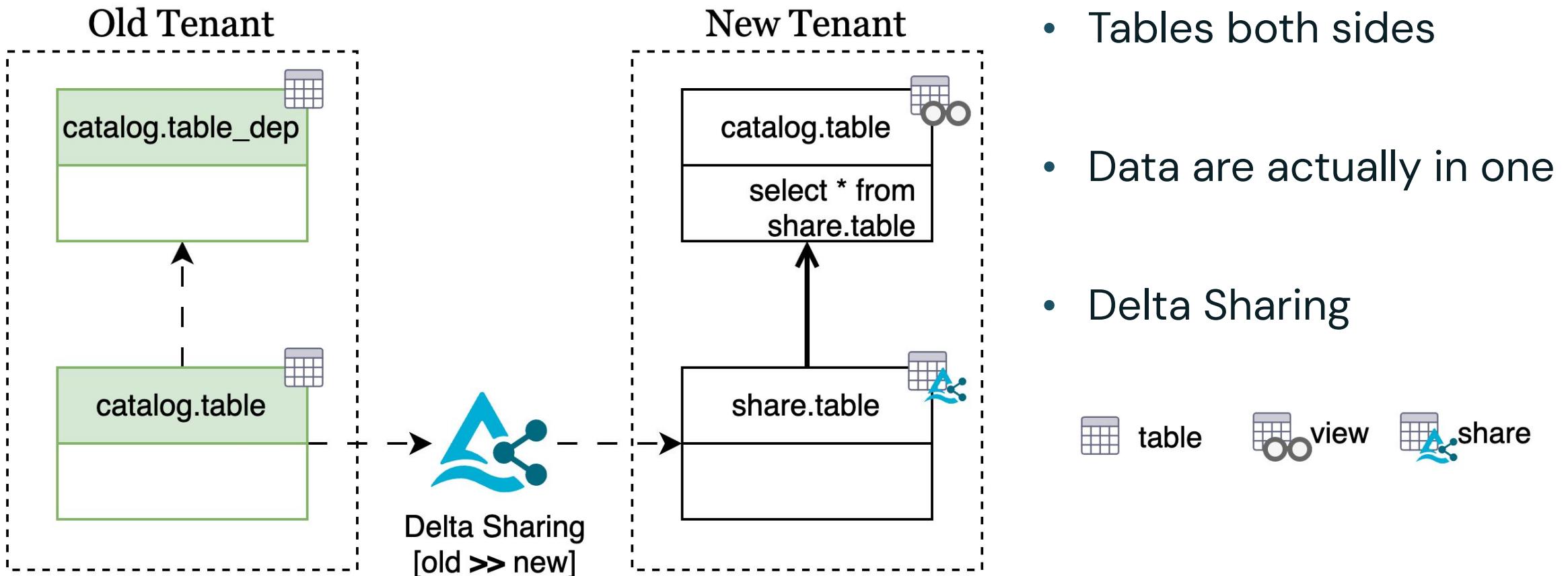
## Code

- Databricks Asset Bundles
- Multi environment setup
- Central configurations
- Service principals, Runtime, Catalogs, Defaults, ACL
- Migrate to **Volumes** from DBFS

```
databi-pltf-stg:  
  mode: development  
  workspace:  
    host: https://adb-...azuredatabricks.net  
    root_path: ...  
  variables:  
    spark_version: 15.4.x-scala2.12  
    policy_id: ...  
    single_user_name: ...  
    cluster_permission_level_user_group: CAN_RESTART  
    container_path_bronze: ...  
    ...  
    spark_conf_catalog_name: "catalog_stg"  
  permissions:  
    - level: CAN_RUN  
      group_name: ${var.user_group}  
  
databi-pltf-prd:  
  mode: production  
  workspace:  
    host: https://adb-...azuredatabricks.net  
    root_path: ...  
  variables:  
    spark_version: 15.4.x-scala2.12  
    policy_id: ...  
    single_user_name: ...  
    cluster_permission_level_user_group: CAN_ATTACH_TO  
    container_path_bronze: ...  
    ...  
    spark_conf_catalog_name: "catalog_prd"  
  spark_env_vars:  
    ...  
    PYSPARK_PYTHON: "/databricks/python3/bin/python3"
```

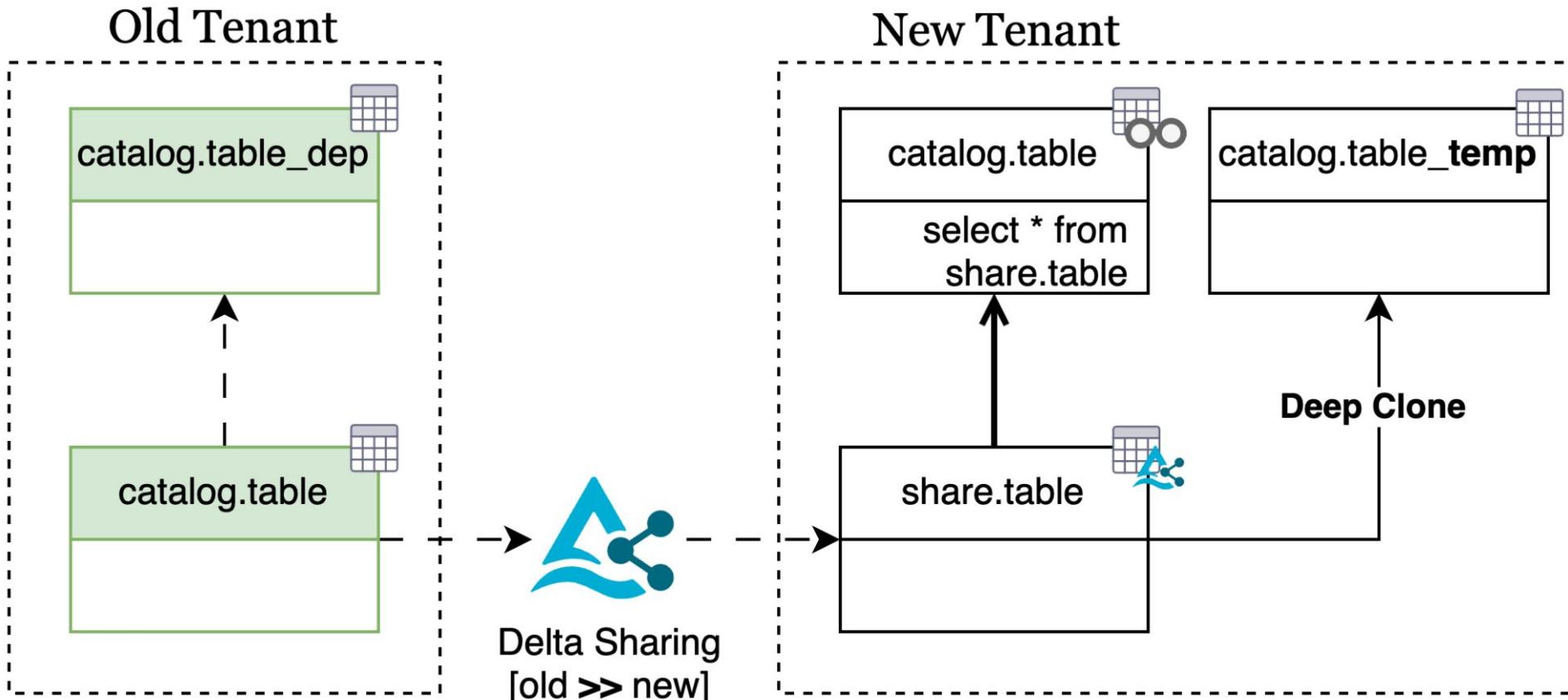
# Migration Tooling

## Phase 1



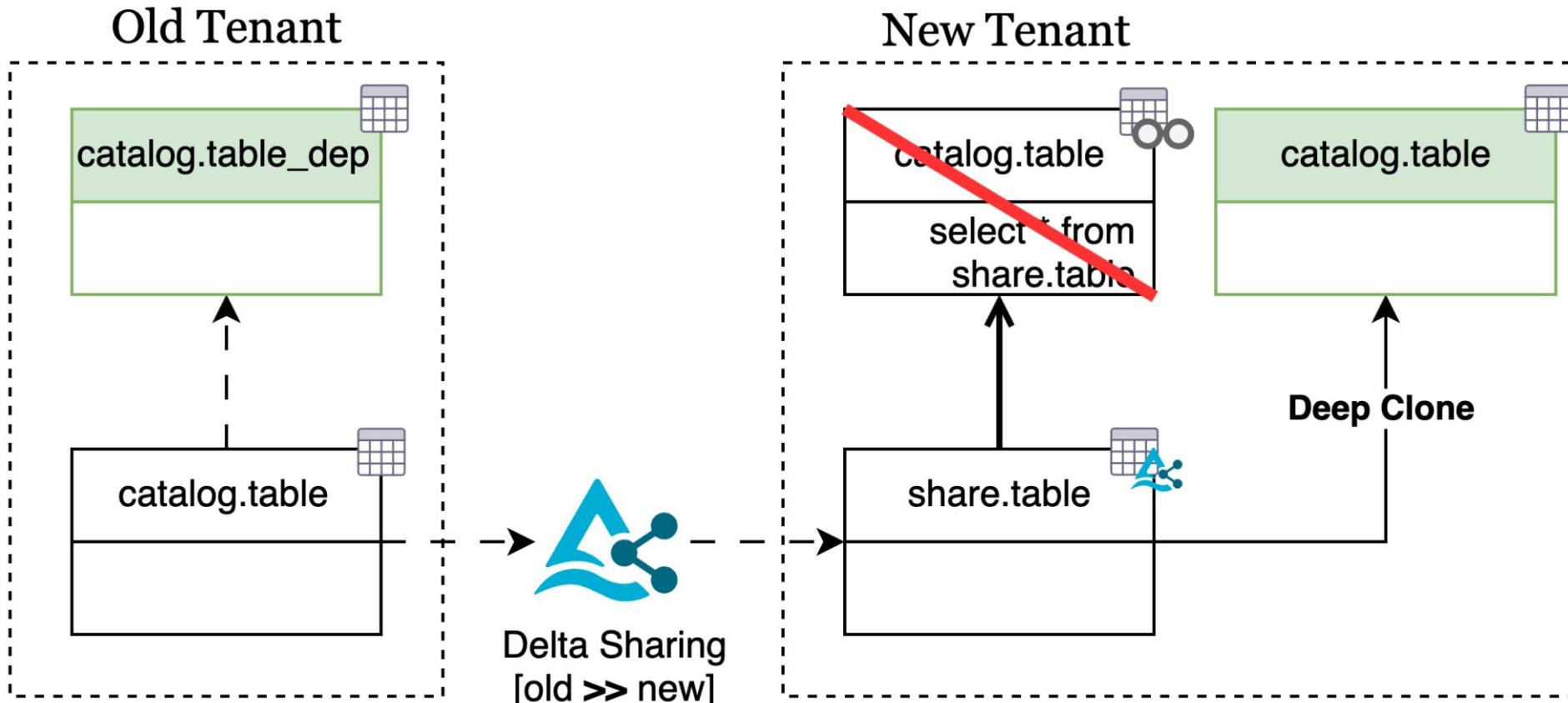
# Migration Tooling

## Phase 2



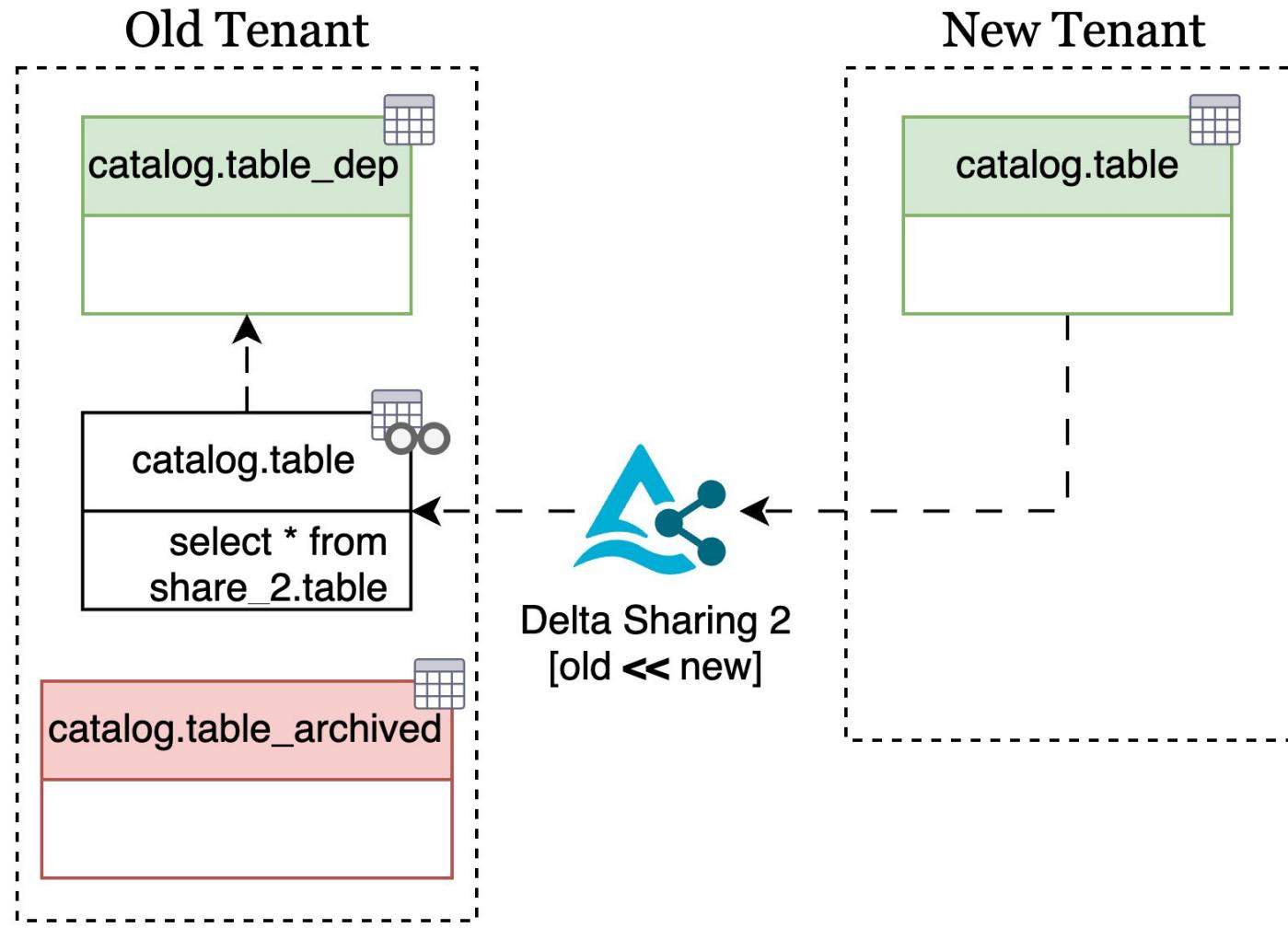
# Migration Tooling

## Phase 3.1



# Migration Tooling

## Phase 3.1



- **Dependencies did not change** throughout
- Second Delta Share
- Minimum data duplication