**STRADA Downstream Environment Provisioning**

**Objective**: Provision isolated and secure dev, tst, and prd environments for a downstream data team at Van Lanschot Kempen, enabling controlled access to STRADA platform data via Unity Catalog and Azure Databricks, while integrating essential Azure services for data processing, storage, and governance.

**Key Design Principles**

**1. Environment Isolation & Lifecycle Management:** Each environment is deployed in its own Resource Group, ensuring clear separation of resources and lifecycle control. Supports independent development, testing, and production workflows.Naming conventions (e.g., rg-data-dev, rg-data-tst, rg-data-prd) aid traceability and automation.

**2. Core Azure Services per Environment:**

Each environment includes:

1. Azure Databricks Workspace
2. Azure Storage Account
3. Azure Data Factory
4. Azure SQL Server + Database
5. Azure Key Vault

Provisioned using Terraform modules with environment-specific .tfvars files.

**Networking Architecture:**

1. Each environment has its own VNET and subnets.
2. Private Endpoints for secure access to services.
3. VNET Peering with STRADA platform VNET.
4. NSGs and Route Tables for traffic control.

**Identity & Access Management (IAM):**

* 1. Azure RBAC for least privilege access.
  2. Entra ID (Azure AD) for identity federation.
  3. Managed Identities for service authentication.
  4. Unity Catalog for centralized data governance.

**Data Access via STRADA & Unity Catalog:**

* 1. STRADA’s Storage Account registered in Unity Catalog.
  2. Downstream Databricks workspaces access data via catalog policies.
  3. Ensures governance, lineage tracking, and auditability.

**Automation & CI/CD:**

1. -Provisioned using Terraform via Azure DevOps Pipelines.
   1. Pipelines include init, plan and apply stages.
   2. OIDC-based authentication for secure access.
   3. .tfvars files drive the deployments.

**Assumptions**

* 1. STRADA platform is operational and exposes data via Unity Catalog.
  2. Downstream teams have Databricks workspaces.
  3. Network policies allow VNET peering and private endpoints.
  4. Service Principal or OIDC authentication is available.
  5. Governance policies support Unity Catalog integration.

**Prompts Used and Their Purpose:**

1. **STRADA and Unity Catalog Integration:** Can you also show how the STRADA storage account and Unity Catalog is connected here?. Purpose: To illustrate data flow and governance model.
2. **IAM and Entra Groups:** Can you indicate how Entra ID groups map to resource roles such as Storage Blob Contributor, SQL Reader, and Databricks Admin?. **Purpose:** To emphasize least-privilege access and identity-driven governance.
3. **Unity Catalog Administration:** Can you represent how catalogs, schemas, and external locations in Unity Catalog map to the downstream team’s storage containers?. **Purpose:** To make the governance and data access model clear across environments.