

# Azure Data Discovery and Data Governance

How do I add Data Discovery and Data Governance capabilities to the proposed solution.

Response ;

Adding Data Discovery and Data Governance capabilities to your proposed solution is crucial for ensuring data quality, security, compliance, and accessibility. Both can be seamlessly integrated into the custom data platform using Azure-native tools or Microsoft Fabric capabilities. Here's how you can enhance the solution:

---

## 1. Data Discovery

Objective: Enable users to understand and discover data assets, their lineage, and relationships across the platform.

Azure Services for Data Discovery

### 1. Azure Purview (Now Part of Microsoft Purview):

**Metadata Management:** Automatically scan and register data assets from Azure Data Lake, Azure Synapse, Azure SQL, Cosmos DB, and external sources (e.g., Salesforce, MongoDB).

**Search and Catalog:** Enable a searchable, centralized catalog of all data assets with metadata.

**Data Lineage:** Track data flows across your pipelines (e.g., from ingestion to transformation to reporting).

### 1. Microsoft Fabric Integration:

**Built-in Data Discovery:** Fabric natively integrates with Microsoft Purview, enabling seamless metadata management, cataloging, and lineage tracking for all assets

stored in OneLake, Data Warehouse, and other Fabric components.

## Key Features to Implement

### Central Data Catalog:

Use Microsoft Purview to create a centralized, searchable catalog of all data assets, including Salesforce, ReactJS/NodeJS, and Azure components.

Register datasets, tables, pipelines, and Power BI reports for unified access.

### Automated Scanning and Classification:

Configure Purview to scan Azure and external sources regularly.

Use built-in AI models to classify data automatically (e.g., PII, financial data).

### Data Lineage Tracking:

Ensure end-to-end lineage visibility for all assets, including the flow of data from Salesforce and D2C systems to the data platform and downstream analytics.

## Implementation Steps

### 1. Configure Microsoft Purview:

Connect Purview to Azure services and external data sources.

Enable scanning for metadata extraction and lineage tracking.

### 1. Establish Search and Discovery:

Build custom data classifications and tags for better discoverability.

Enable user-friendly search for business and technical users.

### 1. Integrate with Power BI and Fabric:

Publish Power BI datasets and reports to Purview for unified insights.

Automatically track lineage for Fabric workflows and pipelines.

---

### 1. Data Governance

Objective: Ensure data security, compliance, and proper usage across the organization.

## Azure Services for Data Governance

### 1. Microsoft Purview:

**Policy Enforcement:** Define and enforce access control policies, ensuring only authorized users can access specific datasets.

**Compliance Tracking:** Monitor data compliance with regulations (e.g., GDPR, CCPA) using built-in compliance tools.

#### 1. Azure Active Directory (AAD):

Manage access control for users and groups with Role-Based Access Control (RBAC).

Integrate AAD with Purview, Fabric, and Azure services for unified authentication and authorization.

#### 1. Microsoft Fabric Integration:

**Data Security:** Fabric supports AAD-based access control for all components, including OneLake, Data Factory, and Power BI.

**Unified Policy Management:** Govern data across the platform using Purview integration.

### Key Features to Implement

#### Access Control and Data Security:

Use AAD for RBAC, restricting access to sensitive data (e.g., PII, financial records).

Encrypt data at rest and in transit using Azure-native encryption tools.

#### Data Policies and Compliance:

Define data retention, masking, and usage policies in Microsoft Purview.

Enforce masking for sensitive fields using Dynamic Data Masking in Azure SQL or Cosmos DB.

#### Auditing and Monitoring:

Enable activity logging in Azure Monitor and Microsoft Purview to track access and changes.

Use Azure Sentinel for advanced threat detection and security monitoring.

### Implementation Steps

#### 1. Define Governance Policies:

Identify sensitive datasets and define policies for access, masking, and retention.

#### 1. Enforce Policies with Purview:

Create and assign data policies in Purview for access control and compliance.

##### 1. Enable RBAC with AAD:

Assign roles and permissions to users and groups for datasets, pipelines, and reports.

##### 1. Monitor and Audit:

Set up regular audits of data usage and compliance in Purview.

Use Purview's compliance dashboard to track adherence to regulations.

---

#### 1. Enhanced Data Platform with Governance and Discovery

Proposed Workflow

##### 1. Data Ingestion:

Automatically classify and tag data assets from Salesforce, ReactJS/NodeJS, and other sources upon ingestion using Purview.

##### 1. Data Transformation:

Track lineage through Fabric Data Factory or Azure Data Factory pipelines and register transformed assets in Purview.

##### 1. Data Storage:

Secure data in OneLake or Azure Data Lake with encryption and RBAC.

##### 1. Data Analytics:

Publish datasets and Power BI reports with lineage and metadata registered in Purview.

##### 1. Compliance and Auditing:

Continuously monitor data access and usage against defined governance policies.

---

Benefits of Adding Data Discovery and Governance

---

Tools Comparison: Purview vs. Other Azure Services

---

By incorporating Microsoft Purview and integrating its capabilities with Azure services or Microsoft Fabric, you can achieve robust data discovery and governance for your custom data platform. This approach ensures better operational efficiency, regulatory compliance, and confidence in your data's quality and security.