Day 28: Azure Purview for Data Governance

Welcome to Day 28 of our Azure Data Engineer interview series! Today, we'll focus on Azure Purview, a unified data governance solution that helps organizations manage and govern their on-premises, multi-cloud, and SaaS data. Azure Purview is essential for ensuring data compliance, security, and discoverability across the enterprise.

Azure Purview for Data Governance

1. What is Azure Purview, and how does it support data governance?

Answer: Azure Purview is a unified data governance platform that provides visibility into your data landscape. It supports data governance by enabling organizations to catalog, classify, and manage data across various sources. Purview ensures that data is discoverable, secure, and compliant with regulatory requirements.

2. How does Azure Purview help in data discovery and classification?

Answer: Azure Purview helps in data discovery by scanning and cataloging data from various sources, such as Azure services, on-premises databases, and SaaS applications. It automatically classifies data based on predefined rules and machine learning models, enabling users to search and discover data assets quickly.

3. What are the key components of Azure Purview, and how do they function?

Answer: The key components of Azure Purview include:

- **Data Catalog:** Central repository where data assets are registered and organized.
- **Data Map:** Visual representation of the data estate, showing how data flows through the organization.
- **Data Lineage:** Tracks the flow of data from source to destination, providing insights into data transformations and dependencies.
- **Classification and Labeling:** Automatically classifies data based on sensitivity, ensuring proper handling and compliance.

4. Explain how data lineage works in Azure Purview and its significance in data governance.

Answer: Data lineage in Azure Purview shows the flow of data from its origin to its final destination. It captures the transformations, data sources, and dependencies involved in the data lifecycle. Lineage is significant in data governance as it helps organizations understand the impact of changes, track data movement, and ensure data integrity and compliance.

5. How can Azure Purview be integrated with other Azure services?

Answer: Azure Purview integrates seamlessly with other Azure services such as Azure Data Factory, Azure Synapse Analytics, and Azure SQL Database. This integration allows for automated data scanning, lineage tracking, and governance across the Azure ecosystem. Additionally, Purview can connect with third-party tools and on-premises systems, providing a comprehensive view of the data landscape.

LinkedIn - Anubhav

6. Describe the role of sensitivity labels in Azure Purview and how they contribute to data security.

Answer: Sensitivity labels in Azure Purview classify data based on its sensitivity level (e.g., confidential, public, or private). These labels enforce data security by restricting access, controlling data sharing, and ensuring that sensitive data is handled appropriately. Sensitivity labels are crucial for maintaining compliance with data protection regulations.

7. What are the benefits of using Azure Purview in a multi-cloud environment?

Answer: In a multi-cloud environment, Azure Purview offers the following benefits:

- **Unified data governance:** Centralized management of data across different cloud platforms and on-premises systems.
- **Cross-cloud data visibility:** Provides a holistic view of data assets, regardless of their location.
- **Consistency in governance:** Ensures consistent data classification, labeling, and security policies across all environments.

8. How does Azure Purview support regulatory compliance, and what are some examples of regulations it helps to address?

Answer: Azure Purview supports regulatory compliance by providing tools to classify, label, and track sensitive data. It helps organizations comply with regulations such as GDPR, HIPAA, and CCPA by ensuring that data is stored, processed, and accessed in accordance with legal requirements. Purview's auditing and reporting features also aid in demonstrating compliance during regulatory reviews.

9. Explain how Azure Purview can be used to establish data ownership and stewardship within an organization.

Answer: Azure Purview enables organizations to assign data ownership and stewardship roles to individuals or teams. Data owners are responsible for managing data assets, ensuring data quality, and overseeing access controls. Data stewards ensure that data governance policies are followed and that data is used ethically and compliantly. This role-based approach enhances accountability and clarity in data management.

10. What are the key challenges in implementing Azure Purview, and how can they be mitigated?

Answer: Key challenges in implementing Azure Purview include:

- **Data integration:** Integrating data from diverse sources can be complex. This can be mitigated by planning a phased implementation and leveraging Purview's built-in connectors.
- **Change management:** Ensuring that the organization adopts data governance practices requires strong leadership and clear communication.
- **Cost management:** Managing the cost of scanning and cataloging large data estates can be challenging. Optimizing scan frequency and focusing on critical data assets can help manage costs effectively.

LinkedIn - Anubhav