Retail Data Estate Assessment – ML, Databricks & Security

- **Problem**
  - The client had enabled a wide range of data services on Azure, including Snowflake, Azure Data Lake, SQL, Databricks, Cosmos and plan to leverage Azure ML (Machine Learning) and expand Databricks adoption soon. Before moving forward, the client needed to perform a review of the existing data estate within the Azure cloud to ensure they were building a good ecosystem and foundation for data ingestion, storage, and analytics and that the environment and services are integrated and connected securely to support operational and ad-hoc analytics.
- **Solution**
  - Spyglass worked with the client and delivered their Data Estate Assessment to evaluate their current maturity level towards accomplishing cloud-scale analytics. Through a series of discovery workshops, Spyglass' Microsoft Data specialists evaluated their Microsoft cloud data estate against critical guiding principles and best practices. Spyglass scored against several dimensions of Microsoft cloud-scale analytics and formalized a roadmap which could be leveraged to accelerate their adoption of Azure data services.
- **Benefits**
  - Databricks and other Azure data services were managed to a very high maturity of cloud standards. However, even with best practices being applied, their security team lacked the trust in cloud services. The assessment formulated a plan forward to educate the security team on Azure security concepts to enable moving forward with new Azure data platforms.
  - The client often struggled with having many platforms and overlapping capabilities. The assessment mapped out and categorized which use-cases were fit for which data platform. This provided guidance that established core data warehousing on Snowflake and Data Science on Databricks & Azure ML as "best of breed" architectures.
  - With driving adoption of Databricks, there was concern on the scale as more and more use-cases were implemented. The assessment guided the better use of a standard data format (Delta), adoption of Unity catalog for collaboration and governance, and a distribution hub data mesh architecture. This allowed core data provisioning and standardization to be automated and enabled agility within the product groups through decentralization. Likewise, data products are easily discoverable and governed through a single tool, Unity Catalog.