



A Women Owned/Women Led Company

Spyglass Presents: The Data Estate Assessment



Infrastructure Azure
Data & AI Azure
Digital & App Innovation Azure
Modern Work
Security

Spyglass's Data Estate Assessment

Spyglass's Data Estate Assessment is a 4-week \$25,000 program helps your organization evaluate your current maturity level for accomplishing cloud-scale analytics. Through a series of discovery workshops, Spyglass will evaluate your estate against critical guiding principles and best practices.

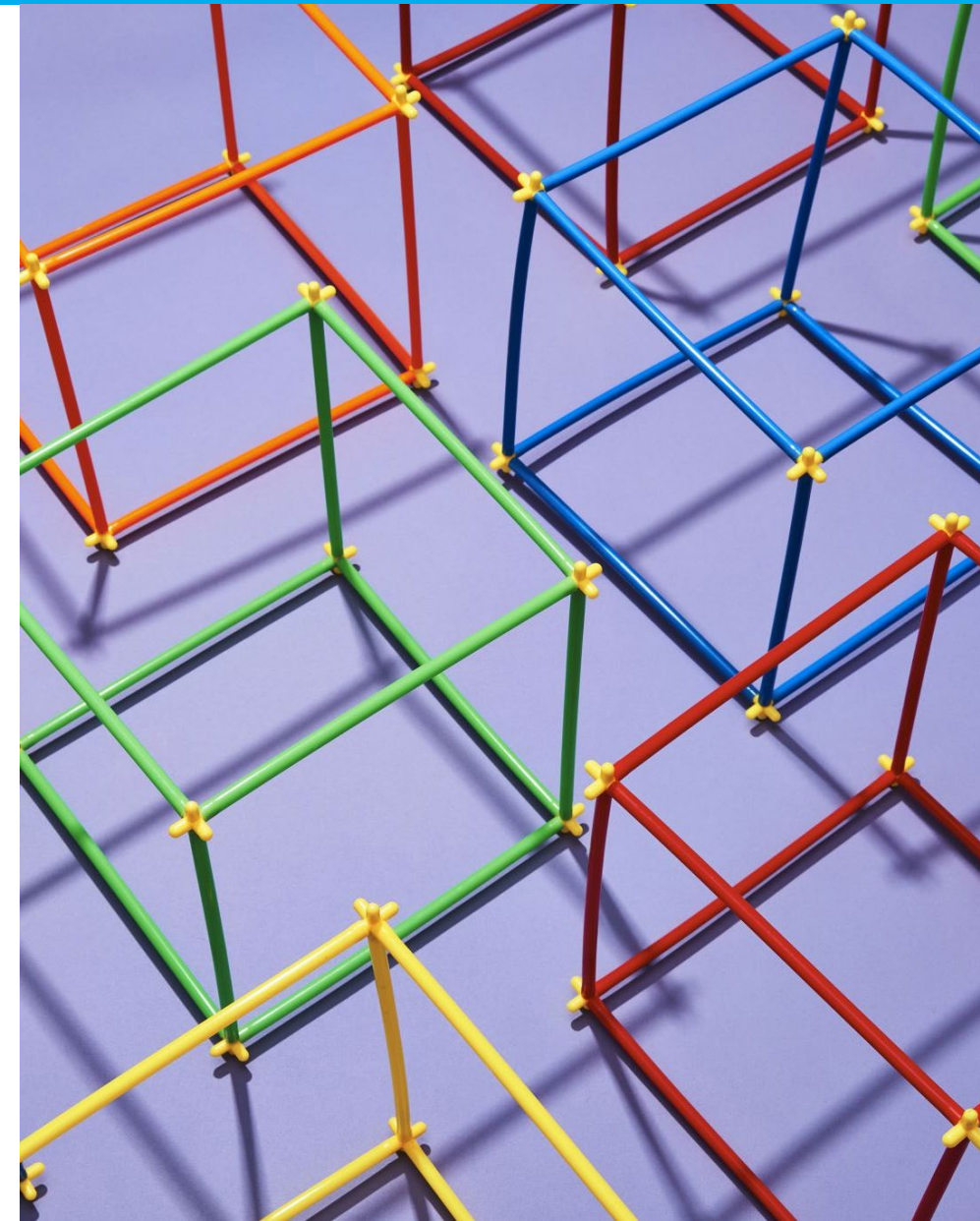
The approach to the assessment follows Microsoft's Cloud-scale analytics and Cloud Adoption Framework. The Microsoft Cloud Adoption Framework provides prescriptive guidance and best practices on cloud operating models, reference architecture, and platform templates while Cloud-scale analytics paves the way for customers to build and operationalize landing zones to host and run analytics workloads.

During this engagement, Spyglass will look across and score several dimensions of cloud-scale analytics including:











- Strategy and Vision
- Cloud Governance & Compliance
- Cloud Identity & Access
- Cloud Architecture
- Cloud Security
- Data Architecture
- Data Management Practices
- Data Governance & Security
- Data Application Security

Deliverables:

- Maturity Scorecard & Executive Summary
- Data Estate Evolution Roadmap
- Detailed Assessment & Recommendations



Data Assessment Maturity Model Scoring

Maturity Score Level	Description
  Ad-hoc	<p>The processes for creation, gathering, sharing of data, or information is not defined. There is a lack of definition of common established standards for data gathering or storage for metadata management. Data exchange, storage, and archiving take place mainly over email. Strategic decisions are often made without enough information. The existence of silos and ad-hoc data managing approaches hinder the performance of the teams</p>
  Defined	<p>There is a well-recognized need for a standard set of tools, processes, and models in place to establish uniformity across the organization. The business finally understands the importance and value of Data Ecosystem and Assets. Sharing of information takes place between the internal teams in the organization.</p>
  Managed	<p>At this stage, the information management system is accepted and adopted. Now, this becomes imperative to support crucial business decisions. Information owners are assigned to govern the data. Information sharing between teams is finally considered as a pivot for enterprise-wide projects. The policies and standards defined earlier are now employed organization-wide. Data governance becomes a part of every project in the organization</p>
  Optimized	<p>Data Ecosystem, at this stage, is viewed as an asset to the company. Data Governance standards and policies are well understood and implemented throughout the organization. Assets are categorized, optimized and metrics are defined. The cost of data management is reduced, and data becomes easier to manage. Operations are more comfortable to navigate through and are streamlined.</p>
  Innovating	<p>Data Assets are now considered to provide the company with an added edge over its competitors. Data strategies are linked with improved productivity and efficiency.</p>

Data Assessment Maturity Model Scorecard (Example)

Governance & Compliance	Identity & Access Management	Cloud Architecture	Data Architecture	Data Governance & Security
<p style="text-align: center;">Optimized</p>	<p style="text-align: center;">Managed</p>	<p style="text-align: center;">Managed</p>	<p style="text-align: center;">Defined</p>	<p style="text-align: center;">Ad-hoc</p>
<ul style="list-style-type: none"> • Customer has established a cloud governance baseline and is actively enhancing it. • The complexity of security policies, data access management, and networking pathways within Customer often poses challenges to innovation. • Striking a balance between stringent security measures and the need for flexibility in innovation is a significant challenge. 	<ul style="list-style-type: none"> • Azure AD is employed to manage identity and access in both cloud and hybrid environments. • Identity centralization is a key component of the well-established cloud adoption plan. • Nevertheless, the use of granular permissions presents access challenges. 	<ul style="list-style-type: none"> • Customer has deployed hardware, virtualization, applications, and services in the cloud, all of which are actively utilized. • Databricks has undergone a security comparison against the Azure security baseline and aligns seamlessly with our technical deployment standards. 	<ul style="list-style-type: none"> • Multiple teams are independently developing and maintaining their data models and reports. • The emergence of other platforms and Snowflake's limitations in the data science realm are driving the adoption of alternative solutions like Databricks, leading to the creation of additional subsystems and data duplication. • There is a lack of consistency in the framework and format used for data storage within Customer. 	<ul style="list-style-type: none"> • Currently, there are no established standards for data management or methods for discovering data silos within Customer. • There is a lack of a centralized data governance policy and framework enforced throughout the organization. • The tracking of data duplication and its limits is not systematically managed, and concerns regarding data classification and sensitive data discovery persist, particularly in relation to security.

Data Assessment Maturity Model Scorecard (Example)

Threat Protection	Business Continuity	Product Security	Development Practices
Managed	Ad-hoc	Managed	Defined
<ul style="list-style-type: none">• Customer is employing Microsoft Defender on a selective basis.• Databricks has implemented minimum-level logging to align with Customer's baseline requirements.• DataOps teams face limitations in logging and monitoring capabilities, lacking access to existing logs for analysis purposes.	<ul style="list-style-type: none">• Customer currently lacks an organization-level business continuity strategy, which extends to its analytical platforms.• An evaluation and documentation of business continuity standards for Databricks, with a focus on its analytical purpose, are needed.• Questions arise, such as whether it is necessary to mandate data asset backups when Snowflake serves as the 'source of truth' for all analytics processes.	<ul style="list-style-type: none">• Penetration Testing is routinely conducted for the external-facing website.• Proactive monitoring is in place to ensure security measures are effective.• Data access security is robustly established, although complexity arises in managing ACL-based access to the data lake.	<ul style="list-style-type: none">• A centralized design framework for reliable, repeatable, and dependable data management is lacking.• In the case of Databricks, the recommended best practice is to utilize Parquet and Delta Lake for enhanced performance, security, and reliability within the analytics platform.• The shift in Data Science use cases toward Databricks highlights the necessity to establish clear guidance regarding when to develop on Snowflake and when to opt for Databricks.

Data Assessment Security Controls Posture (Example)

Network Security	Identity & Access Management	Cloud Architecture	Logging and Threat Detection	Data Protection
Compliant	Compliant	Compliant	Compliant	Compliant
<ul style="list-style-type: none"> ✓ NS-1: Establish network segmentation boundaries <ul style="list-style-type: none"> ✓ Virtual Network Integration ✓ Network Security Group Support ✓ NS-2: Secure cloud services with network controls <ul style="list-style-type: none"> ✓ Azure Private Link ○ Disable Public Network Access (Optional) 	<ul style="list-style-type: none"> ✓ IM-1: Use centralized identity and authentication system ✓ IM-3: Manage application identities securely and automatically ✓ IM-7: Restrict resource access based on conditions ✓ IM-8: Restrict the exposure of credential and secrets ✓ PA-7: Follow just enough administration (least privilege) principle ✓ PA-8: Determine access process for cloud provider support 	<ul style="list-style-type: none"> ✓ CAF: Resource Group Usage ✓ CAF: Infrastructure as Code ✓ AM-2: Use only approved services ✓ Azure Policy ○ PV-3: Define and establish secure configurations for compute resources** **Recommended to define standard cluster configurations for Databricks adoption within Customer machine requirements 	<ul style="list-style-type: none"> ✓ LT-1: Enable threat detection capabilities ✓ LT-4: Enable logging for security investigation **Recommended to consider logging for application and support teams beyond security 	<ul style="list-style-type: none"> ✓ DP-3: Encrypt sensitive data in transit ✓ DP-4: Enable data at rest encryption by default ○ DP-5: Use customer-managed key option in data at rest encryption when required **This is not required nor recommended ✓ DP-6: Use a secure key management process **Key vault is enabled for key management

Next Steps



Scoping Session to understand client's data landscape



Process paperwork



Start Assessing!



Thank You