The low-down on Azure Purview

Microsoft's Enterprise Data Catalogue tool, Azure Purview, was released as "Generally Available" last week so I thought it was timely to do a review of its focus, features, functions, fees, and future.

Focus:

Purview is Microsoft's entry into the Enterprise Data Catalogue market. It's a growing market, with a recent BusinessWire article estimating that the market will reach USD$8.2bn by 2027. The key focus is in cloud-based catalogues and metadata management. Purview is positioned to be Microsoft's entry into this valuable market.

The Enterprise Data Catalogue market is still finding its direction, though, as there is not yet a clear and agreed scope of what these tools should cover. Gartner quotes the following capabilities as available in some or all the competitors in the market:

- Metadata Repositories: Contains information about the columns, tables, views, fields, files, and processes that are used in your IT systems
- Business Glossaries: Contains business terms and their definitions, policies, procedures, regulations, and other business items of note
- Data Lineage: Shows the movement of data from source to consumption, through your systems
- Impact Analysis: Uses Data Lineage information to show the impact of upstream change on downstream data consumption
- Rules Management: Automates the enforcement of business rules and processes in relation to data
- Semantic Frameworks: Structures business and technical glossaries into structured hierarchies

On top of these business features, the participants of this market need some way to automate as much of the above functionality as possible, through automated tools that scan your IT systems and data estates to collect and retrieve the information needed to deliver the above functionality.

In my view, Gartner should also add Data Profiling and/or Data Quality capabilities as these very neatly complement the metadata information in the Enterprise Data Catalogue tools.

Microsoft Azure is a Cloud-based Platform-as-a-Service product that provides Metadata Repositories, Business Glossaries, Data Lineage, and Impact Analysis.

Features:

As you would expect in a product that is cloud-native to the Microsoft Azure cloud environment, Purview is super easy to install and connect to scan your Azure data estate. Installation is literally a matter of a few minutes and, assuming security configurations are in place, getting your first scan up and going is also easily completed in 5-10 minutes. This is a big advantage over the non-cloud-hosted data catalogues that can take days or weeks to install.

Purview has already released connectors for the following technologies:

- **File Store**: Azure Blob Storage, AWS S3, Azure Data Lake Storage Gen 1+2,
- **Database**: Azure Cosmos DB, Azure Dedicated SQL Pool, Azure SQL Database, Azure Synapse Analytics, Cassandra, Google BigQuery, Hive Metastore DB, Oracle DB, SQL Server, Teradata
- **Services and Apps**: Erwin, Looker, PowerBI, SAP ECC and SAP S4 Hana
- **Transformation Engines**: Azure Data Factory, SqlServer Integration Services, Spark

This set of connectors for automated scanning of catalogue and lineage will cover customers with a modern Microsoft architecture. Customers who have technologies outside this list can still benefit from Purview by using the REST API to load in additional catalogue and lineage information.

Purview also comes with an extensive range of pre-programmed patterns in a function called Classification. Purview uses these to search for specific types of data (e.g., driver's license numbers, credit cards, email addresses, etc) normally associated with sensitive or private data, and then can flag where they have been found. This is a real useful function in finding where sensitive data is stored in your organisation so that it can be correctly controlled. Microsoft has provided a great set of approx. 200 different global patterns for sensitive data, and this can be customised for any additional local patterns in your data.

Functions:

Once technical and business data has been loaded into Azure Purview, it can support the following basic use cases:

- **Understand business terms**. Purview provides a hierarchical business glossary that can be searched and viewed to understand the business meaning of terms. Additionally, these terms can be linked to the technical asset that houses the relevant data.
- **Find data**. If a data analyst or data scientist was searching for data, Purview can assist. The analyst could search for the relevant term and then find where that data is stored by linking to the technical asset.
- **Understand sourcing**: Someone who has a query about the meaning of data on a report can use Purview to understand the definition of the data, and also the sourcing method for the data. Purview can provide the business definition, as well as the lineage report to show how the data was transformed and the steps, stages, and systems it has passed through to get to the report.

Fees:

Purview is charged on a usage basis, so actual pricing will depend on amount of data, number of systems scanned and how many people are using the instance. A likely low end price point will be for a single platform with approx. 4 systems to be scanned weekly, generating less than 2GB of metadata. This would cost between AUD$5,000 and AUD $10,000 per year. This increases if you are looking at Dev/Test/Prod instances, more systems to scan, more metadata to store, or more users to use the results.

These price points compare very favourably against the Enterprise Data Catalogue market, which can range from $50k to $1m depending on volume of metadata, number of users and number of platforms.

Future:

Microsoft is continuing to build out the features and capabilities of Purview. Publicly available information includes proposed connectors for Snowflake, Postgres, MongoDB, SAP HANA and MySQL. Additional functionality can also be expected. Keep tuned as this newly production-ready platform is building towards something very exciting.

Finally:

Microsoft Azure Purview is an excellent Enterprise Data Catalogue, with great applicability for those organisations that are Microsoft-centered but have a hybrid cloud and on-premise enterprise application base. It's got features for business in the business glossary, for IT security with its classifications capability and for IT architecture and analysts with the ability to track data flows and lineage from source to target.

If your organisation is interested in Azure Purview, or you would like to find about more, please contact a Fujitsu Data & AI specialist now.

Contact

Fujitsu Data & AI

+61 3 9924 3000