Azure Environment Assessment for International Analytics Firm

ORGANIZATION

Our client is an information, intelligence, and analytics service provider for large corporations and governments driving the global economy. Founded in 1959 and headquartered in London, the company has grown both organically and through acquisitions. With over 5,000 analysts, data scientists, and domain experts, our client offers a highly integrated view of their customers' world by connecting data across various variables. The insights provided help clients identify opportunities, mitigate risks, and solve complex problems.

CHALLENGE

Our client utilized Azure Table Storage and Azure Blob Storage to query and report on large volumes of infrastructure logs, including time series data from vROps, Pure performance data, WAN link performance data from SolarWinds, geographic data from Tangoe, backup data from Commvault, Azure/AWS vulnerability data from Qualys, incident and asset management data from ServiceNow, and AWS inventory data from CloudCheckr. However, the performance of Azure Table Storage proved inadequate for these use cases. Additionally, they managed Power BI user workspaces under 5 GB due to the Pro license limitation of 10 GB per user.

They transitioned some data from Azure Table Storage to Azure Data Lake Storage Gen-2 for Power BI queries due to frequent time-out issues with Table Storage. Many ad hoc queries from Azure Table Storage were running slowly. Scheduled data refreshes into Azure occurred at intervals of 30 minutes, 4 hours, or once daily. Additionally, our client began migrating from Azure Blob Storage to Azure Data Lake Analytics Gen-2 to leverage account-level storage capabilities.

Our client aimed to incorporate additional infrastructure performance and log data from sources such as on-premise SCCM, Linux server validation (Red Hat Satellite), AWS/Azure MS SQL Server licensing data (Qualys), system coverage (Splunk), costing data (Datadog and PagerDuty), server security (CrowdStrike), and Oracle database information (OEM). However, lacking in-house Azure expertise to establish a future-state data architecture for analytics and reporting of large data volumes, they required a trusted partner to validate their Azure environment and provide recommendations.

TECHNICAL SOLUTION

XTIVIA conducted an Azure environment assessment and data analysis, identifying root causes for poor Table Storage performance, issues with naming conventions, data quality, and the reliability of the existing architecture to handle large data volumes for future use cases. XTIVIA recommended migrating Table Storage to Cosmos DB, leveraging Azure Analysis Services as a semantic layer, and

utilizing Databricks for data engineering, stream analytics, and machine learning use cases. Additionally, XTIVIA provided architecture recommendations for accommodating future use cases.

XTIVIA's architectural recommendations included:

Migrating Table Storage to Cosmos: The current Table Storage faced challenges with complex ad hoc queries due to improper partition key and row key selection. We recommended migrating the data and queries to Cosmos DB, which offers multiple APIs, a robust query processing engine, and NoSQL database capabilities. Existing Table API queries could transition to Cosmos without modifications. Additionally, Cosmos supports auto-indexing, ensuring performance, and allows migration to SQL API over time for enhanced querying capabilities. We also provided performance improvement design patterns for optimizing Table Storage.
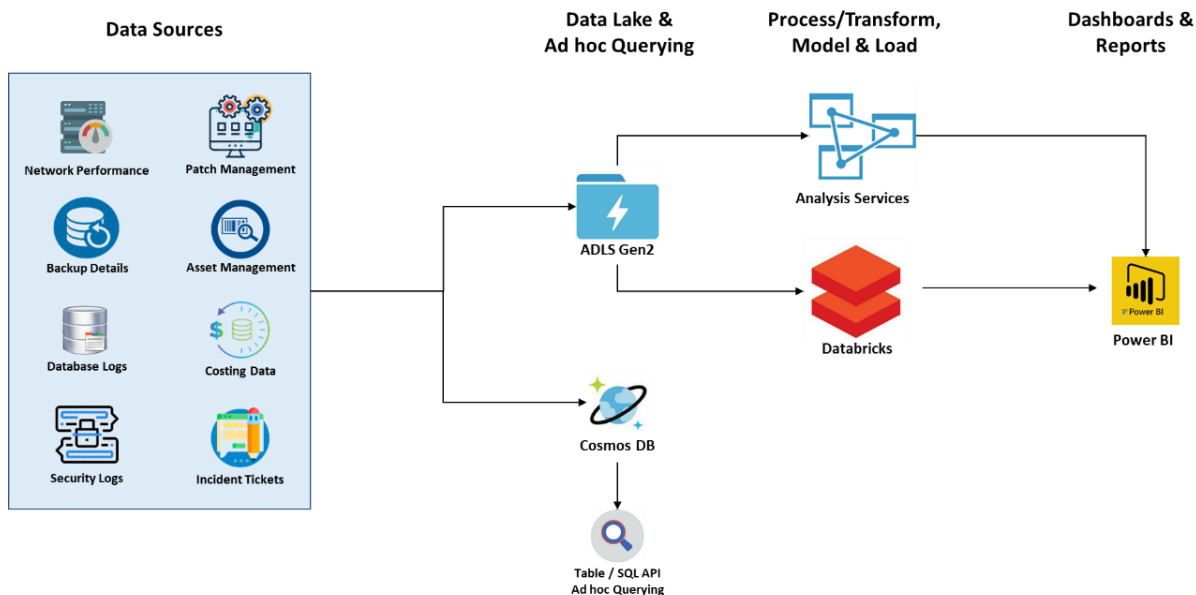
Use of Naming Conventions: XTIVIA recommended establishing naming conventions for Azure resources, data assets, and attributes, and publishing them across the organization to enforce standards and consistency. Additionally, recommendations were provided for Blob naming, emphasizing that the combination of account, container, and blob acts as a partition key, which partitions blob data into different ranges.

Analytics Enablement: XTIVIA recommended Azure Databricks for data engineering and machine learning due to its collaborative workspace for data engineers, data scientists, and business analysts. It can also be integrated with Azure AutoML to automate and manage the end-to-end data science lifecycle, from data preparation to model deployment. Additionally, Azure Databricks, combined with Event Hub, is well-suited for handling stream analytics use cases involving large volumes of infrastructure logs.
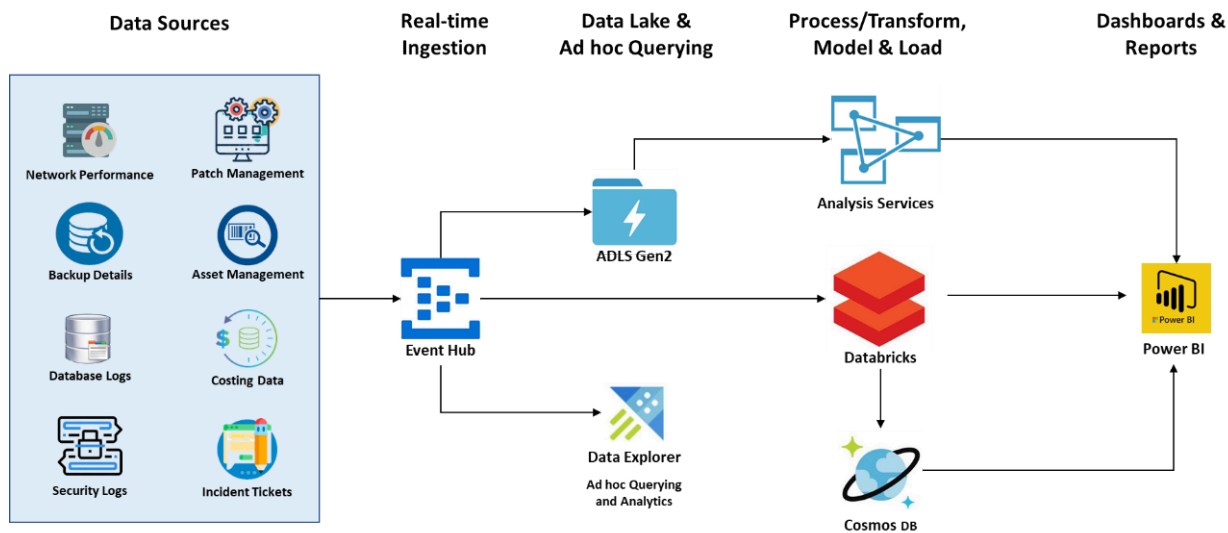
Ad hoc Capabilities on Streaming Data: XTIVIA recommended Azure Data Explorer for exploring and analyzing terabytes of data. Azure Data Explorer offers multiple querying options, including Kusto Query Language, Power BI, Tableau, ODBC, and Excel, along with robust indexing and joining capabilities.

Enterprise Semantic Layer (Optional): XTIVIA recommended building a semantic layer by integrating multiple time-series datasets using common business keys to support reporting on millions of rows. This approach was suggested as a potential solution for future use cases involving ad hoc analysis on large data volumes by a broad user base with limited infrastructure domain knowledge.

Phase 1 (Short Term)

Data Sources

- Network Performance
- Patch Management
- Backup Details
- Asset Management
- Database Logs
- Costing Data
- Security Logs
- Incident Tickets

Data Lake & Ad hoc Querying

- ADLS Gen2
- Cosmos DB
- Table / SQL API Ad hoc Querying

Process/Transform, Model & Load

- Analysis Services
- Databricks

Dashboards & Reports

- Power BI

Phase 2 (Long Term)

Data Sources

- Network Performance
- Patch Management
- Backup Details
- Asset Management
- Database Logs
- Costing Data
- Security Logs
- Incident Tickets

Real-time Ingestion

- Event Hub

Data Lake & Ad hoc Querying

- ADLS Gen2
- Data Explorer Ad hoc Querying and Analytics

Process/Transform, Model & Load

- Analysis Services
- Databricks
- Cosmos DB

Dashboards & Reports

- Power BI

BUSINESS RESULT

The detailed analysis and future-state architecture provided our client with a better understanding of Azure environment best practices, tactical, strategic objectives, and solution tradeoffs. The deliverables from the engagement provided the basis for cost-benefit analysis, budgeting, staffing, and risk management in implementing the solution. In addition to this, the client gained understanding of the process and data required for machine learning use cases relevant to infrastructure logs and performance data.

BY THE NUMBERS

- 50,000 customers in 140 countries
- 85% of customers are in Fortune's Global 500
- 94 customers are among 100 largest US corporations
- 600+ industry subject-matter experts
- 5,000+ data scientists and information analysts

KEYWORDS

Azure Environment Assessment, Azure Data Lake, Azure Storage, Azure Architecture

SOFTWARE

Azure Table Storage, Azure Data Lake Storage Gen-2, Power BI, Azure Cosmos DB, Azure Analysis Services, Azure Databricks, Azure Event Hub, Azure Storage Explorer, Azure Data Explorer

ABOUT XTIVIA

At XTIVIA, we've been providing IT solutions and consulting services for over 30 years with a wide range of services, including technology assessments, IT service and asset management, software development, data analytics, cloud migration, DevSecOps, ERP, and enterprise content management. Dedicated to each discipline, ensuring that our clients receive the best possible service. Through strategic acquisitions, we've acquired talented people who are experts in their industries, passionate about what they do, and committed to providing exceptional service to our clients. Whether you need to improve your IT infrastructure or implement new software solutions, XTIVIA is here to help you achieve your business goals. XTIVIA has offices in Colorado, New York, New Jersey, Texas, Virginia, and India. www.xtivia.com