

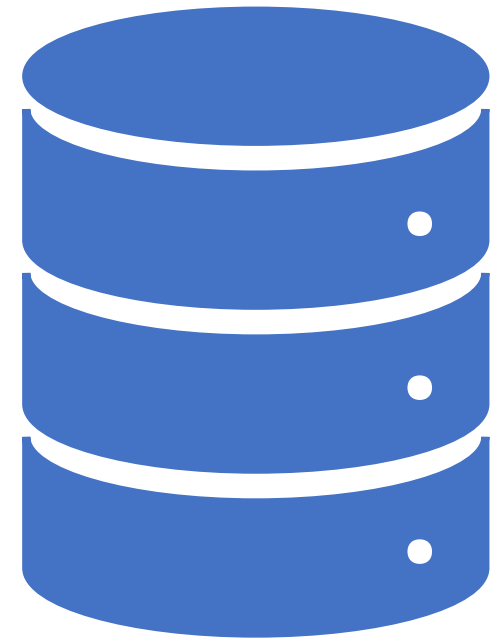
Reference Solution Architecture

Data Architecture – Enterprise Data Warehouse (EDW) and Data Lakehouse

Acquire and Organize - Azure Data Factory, Databricks and Azure Synapse

Eric Frayer
March 2024

www.EricFrayer.com – www.ThinkDataAnalysis.com



Building a Data Driven Organization


Enabling Self Service - building Data Models and organizing visualizations to support end-users creating their own dashboards and finding their own insights.

Storytelling with Data – Insights on how to be most effective with visualizations of data.

Citizen Data Scientists – BI professionals who have mastered one of the Modern Analytics and BI toolkits could be looking to adopt data science skills.

Advanced Analytics – solid Descriptive Statistics are required so the underlying data supports the ability to effectively develop and deploy AI and ML models.

Data Literacy Programs – building business user's abilities to analyze data and practices.



Why “Data Warehousing” in the Cloud?

Scalability

MPP – Massive Parallel Processing

Cost

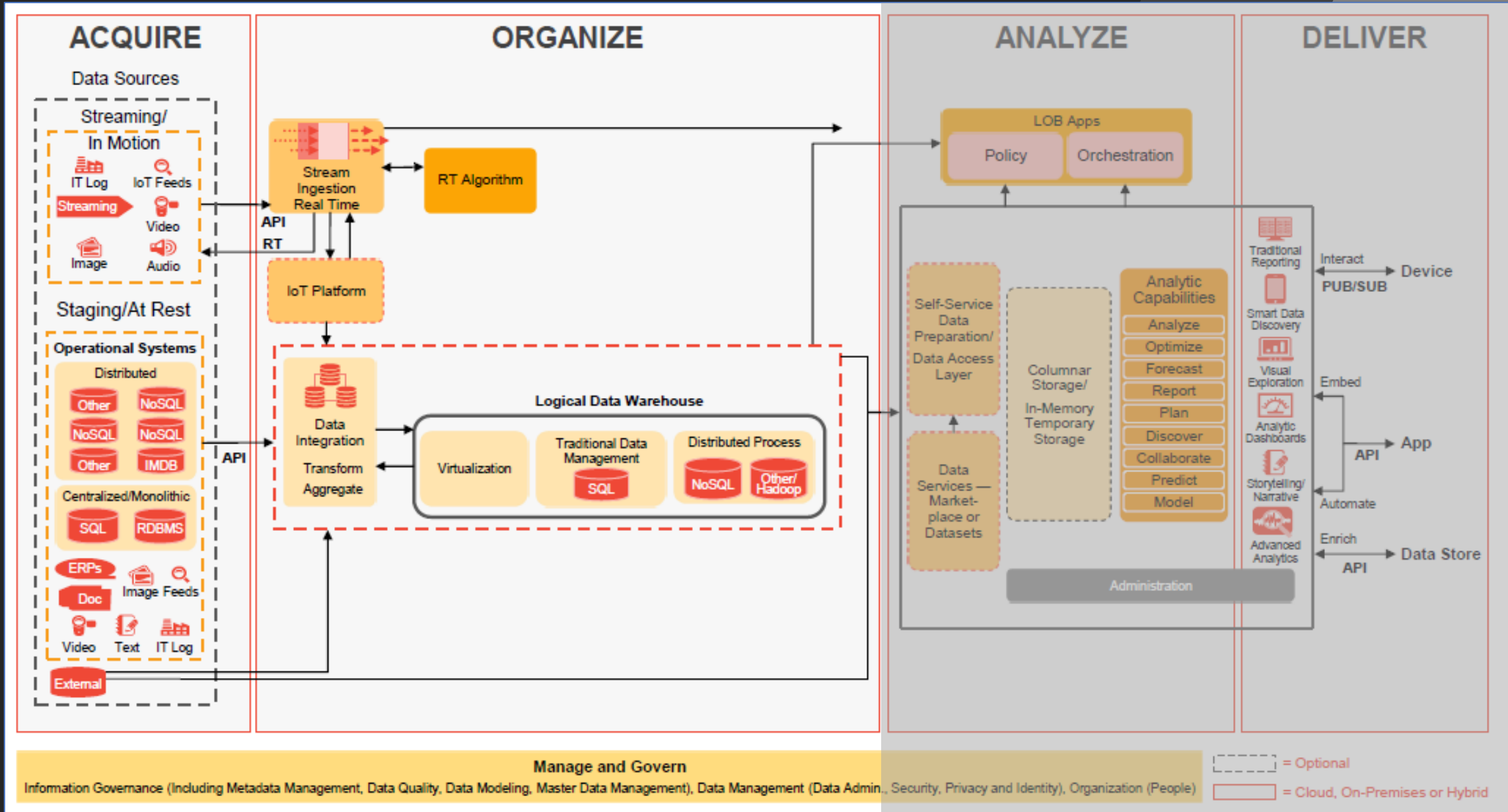
No Capital Expense

No Hardware Maintenance

Integration with other services (SaaS)

Time to Market

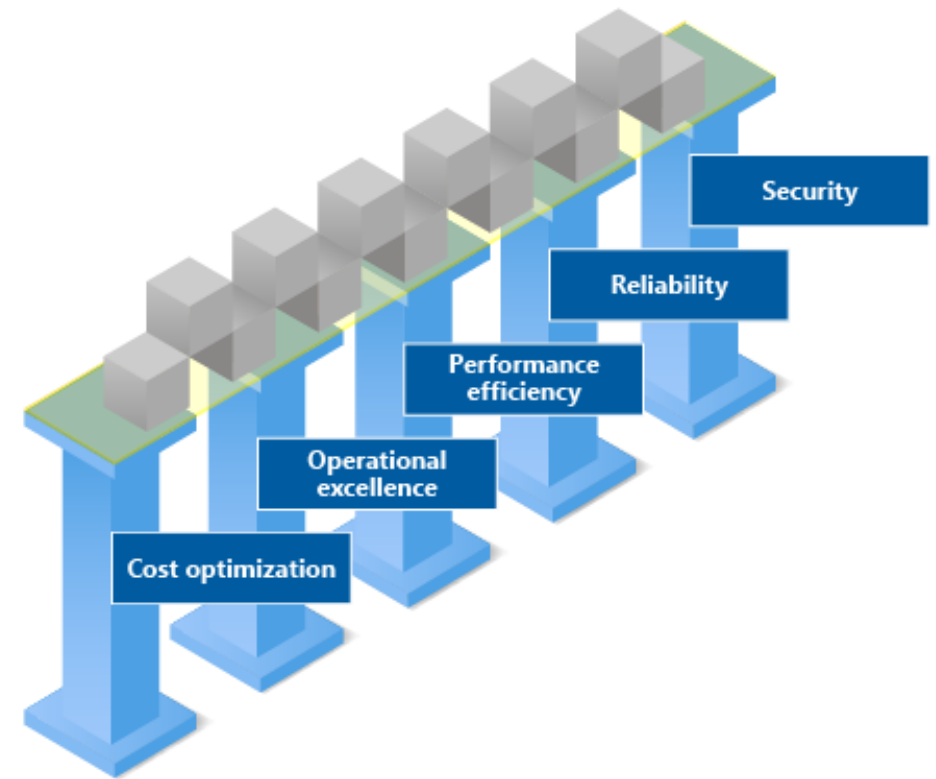
Acquire and Organize



Azure Well Architected Framework

In addition to each of these pillars, there are some consistent design principles that you should consider throughout your architecture.

- **Enable architectural evolution:** No architecture is static. Allow for the evolution of your architecture by taking advantage of new services, tools, and technologies when they're available.
- **Use data to make decisions:** Collect data, analyze it, and use it to make decisions surrounding your architecture. From cost data, to performance, to user load, using data will guide you to make the right choices in your environment.
- **Educate and enable:** Cloud technology evolves quickly. Educate your development, operations, and business teams to help them make the right decisions and build solutions to solve business problems. Document and share configurations, decisions, and best practices within your organization.
- **Automate:** Automation of manual activities reduces operational costs, minimizes error introduced by manual steps, and provides consistency between environments



Design Choices

The selected architecture should take into consideration many variables. The following is guidance from Microsoft:

“In an ideal architecture, you would build the most secure, high-performance, highly available, and efficient environment possible. However, as with everything, there are trade-offs.

To build an environment with the highest level of all these pillars, there's a cost. **That cost might be in money, time to deliver, or operational agility.** Every organization will have different priorities that will affect the design choices that are made in each pillar. As you design your architecture, you'll need to determine which trade-offs are acceptable, and which are not.

When you're building an Azure architecture, there are many considerations to keep in mind. **You want your architecture to be secure, scalable, available, and recoverable.** To make that possible, you'll have to make decisions based on cost, organizational priorities, and risk.”

Key Variables: Cost Optimization, Operational Excellence, Performance Efficiency, Reliability and Security.

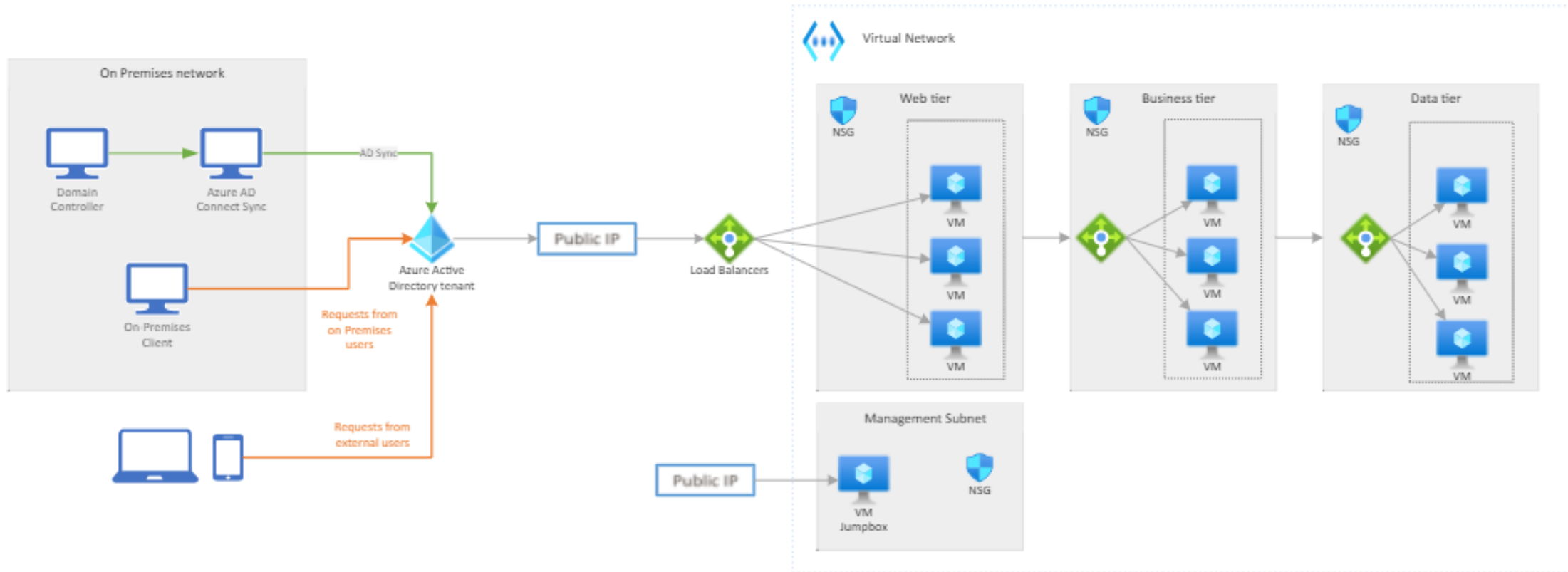
Enterprise Architecture's Role in Building a Data-Driven Organization

Enterprise Architects are uniquely positioned to:

- Identify enterprise data opportunities that people working within silos cannot.
- Understand end-to-end data flows to ensure optimization.
- Ensure that digital and business strategies are aligned to meet the data needs of the present and future.

The Enterprise Architects works with all stakeholders. To build an Enterprise Data Warehouse, they will need to work closely with the Power BI Analytics, Data and Reporting Teams.

Source: "Enterprise Architecture's Role in Building a Data-Driven Organization"
Smarter With Gartner - May 28, 2021 -
Contributor: Ashutosh Gupta



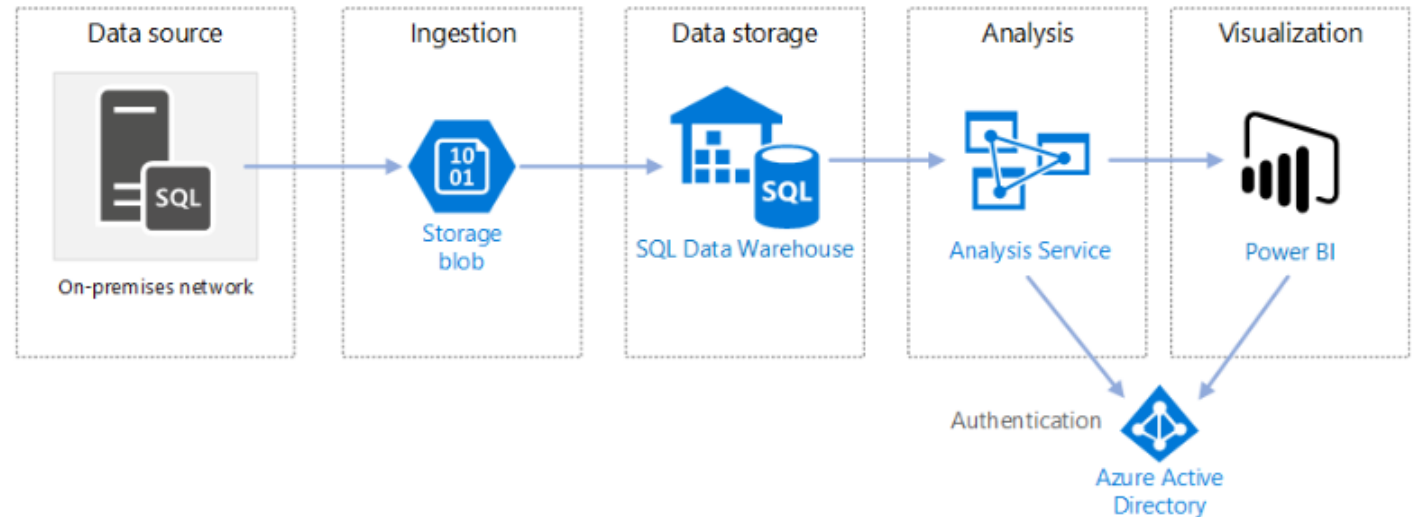
Source: placeholder – to be updated!

Legacy Architecture

Reference Architecture

Enterprise BI with SQL Data Warehouse

This reference architecture implements an [ELT](#) (extract-load-transform) pipeline that moves data from an on-premises SQL Server database into SQL Data Warehouse and transforms the data for analysis.



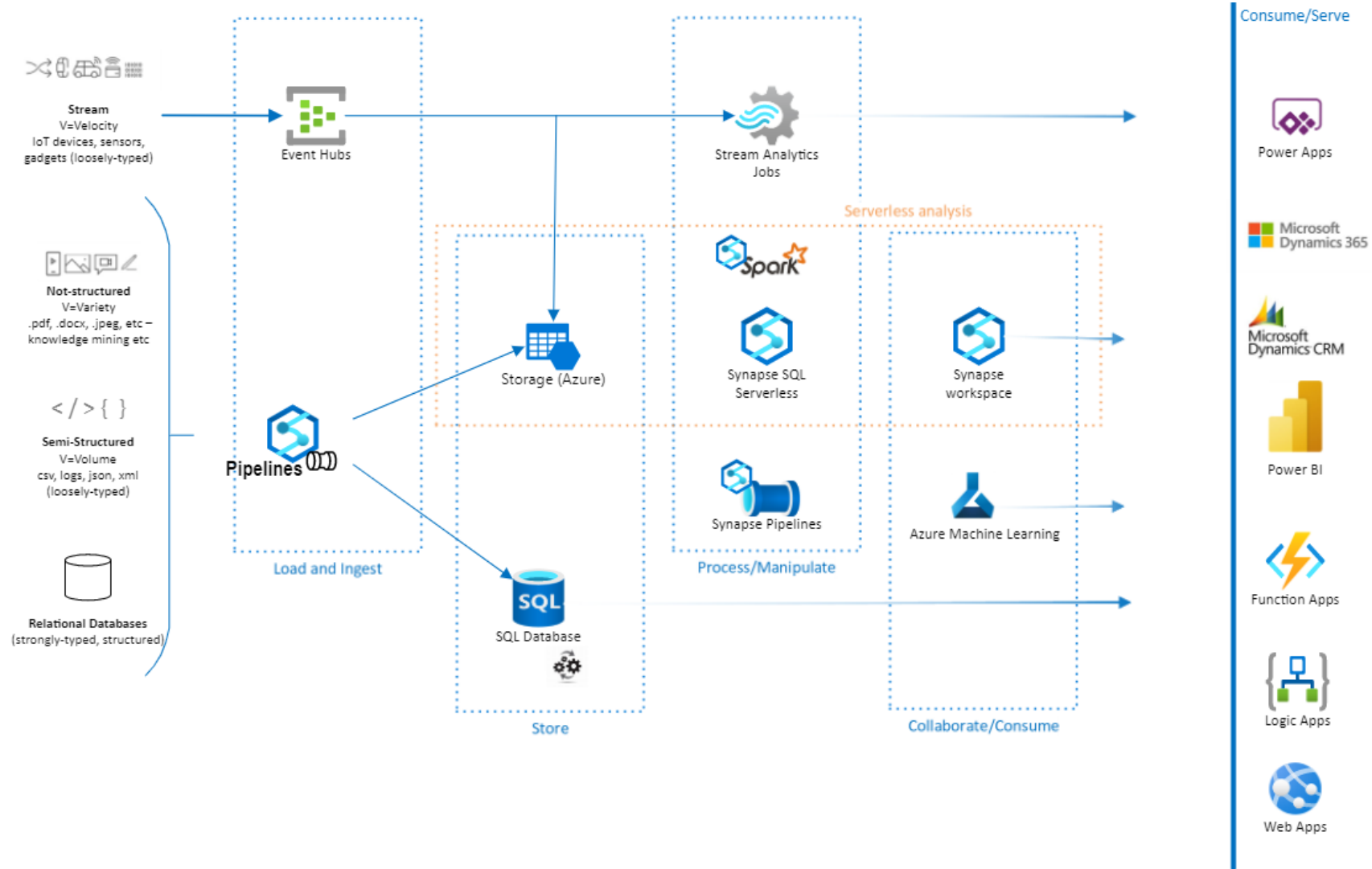
For more information about this reference architectures and guidance about best practices, see the article [Enterprise BI with SQL Data Warehouse](#) on the Azure Architecture Center.



Azure

Modern Data Warehouse – Small Medium Business (SMB) pattern

Azure SQL DB and Synapse workspace



Azure Data Factory (ADF)

- Fully Managed
 - Serverless
 - Data Integration Service
 - Ingesting, Preparing and Transforming Data
 - Scalable and Schedulable
-
- ADF Can read and write to:
 - Azure SQL Database
 - Azure Blob Storage
 - Azure Data Lake Storage Gen2

Source: Microsoft

Azure Databricks

- Industry Leading Spark (Databricks Runtime) built on a highly optimized version of Apache Spark offering 50x performance
- Already has support for Spark 3.1.2 with DBR 8.4
- Allows users to opt for GPU enabled clusters and choose between standard and high-concurrency cluster mode
- Supports for streaming data
- Databricks Notebooks has as real-time co-authoring (both authors see the changes in real-time)
- Automated versioning with CI/CD

Source: <https://docs.microsoft.com/en-us/answers/questions/461614/what-are-the-major-pros-and-cons-of-using-synapse.html>

Azure Synapse

- Open-source Apache Spark (thus not including all features of Databricks Runtime)
- Supports Apache Spark 2.4 (GA) and 3.0 (Preview)
- Has built-in support for .NET for Spark application.
- Unified security and monitoring features including Managed VNets.
- Nteract Notebooks has co-authoring of Notebooks, but one person needs to save the Notebook before another person sees the change.
- Automated versioning with CI/CD
- Allows for interactions with Data Warehouse and Data Lake with different tools and technologies in one Workspace

Source: <https://docs.microsoft.com/en-us/answers/questions/461614/what-are-the-major-pros-and-cons-of-using-synapse.html>

Azure Synapse advantages over Azure Databricks

- Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing and big data analytics.
- It gives you the freedom to query data on your terms, using either serverless or dedicated resources—at scale.
- Azure Synapse brings these worlds together with a unified experience to ingest, explore, prepare, manage and serve data for immediate BI and machine learning needs.

Source: <https://docs.microsoft.com/en-us/answers/questions/461614/what-are-the-major-pros-and-cons-of-using-synapse.html>

Acquiring Data – Loading the Data Warehouse/Data Lake

Single client loading methods

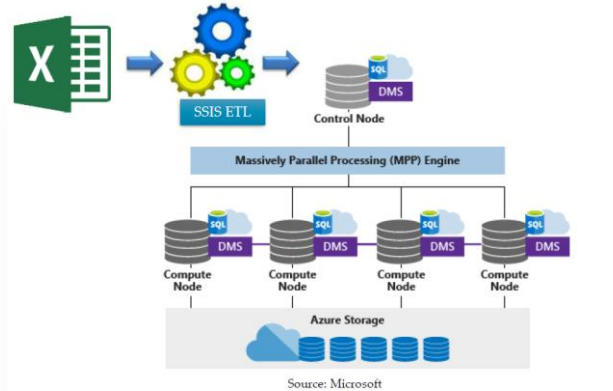
- SSIS
- Azure Data Factory
- BCP
- Can add some parallel capabilities but are bottlenecked at the control node

Parallel readers loading methods

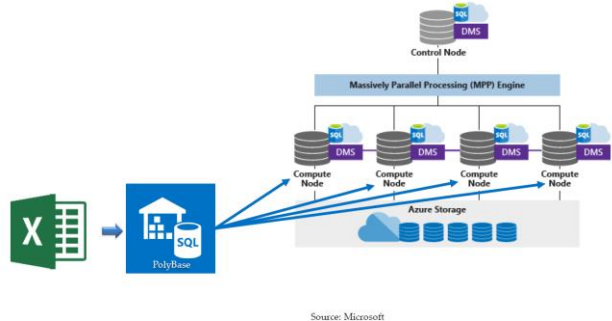
- PolyBase
- Reads from Azure blob Storage and loads the contents into Azure SQL DW
- Bypasses the Control Node and loads directly into the Compute Nodes

Single client loading methods

- SSIS
- Azure Data Factory
- BCP
- Can add some parallel capabilities but are bottlenecked at the control node



Acquiring Data –
Loading the Data
Warehouse/Data Lake



Parallel readers loading methods

- PolyBase
- Reads from Azure blob Storage and loads the contents into Azure SQL DW
- Bypasses the Control Node and loads directly into the Compute Nodes

Azure Synapse – Acquire and Organize – POC Exercises

The following actions are needed to evaluate the proposed Architecture:

1. PolyBase Data Loading Example – File Flat loading to Data Lake with Access by External Tables in Azure Synapse - SQL Server SSMS
<https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-guide>
2. Copy and transform data in Azure Synapse Analytics by using Azure Data Factory or Synapse pipelines
<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse?tabs=data-factory#use-polybase-to-load-data-into-azure-synapse-analytics>
3. Copy and transform data in Dynamics 365 (Microsoft Dataverse) or Dynamics CRM using Azure Data Factory or Azure Synapse Analytics
<https://docs.microsoft.com/en-us/azure/data-factory/connector-dynamics-crm-office-365?tabs=data-factory>

The method(s) for Data Transfer and ETL will likely change with the move to a “Cloud only” Architecture. Additional discovery and research is needed.

Resources and Additional Materials:

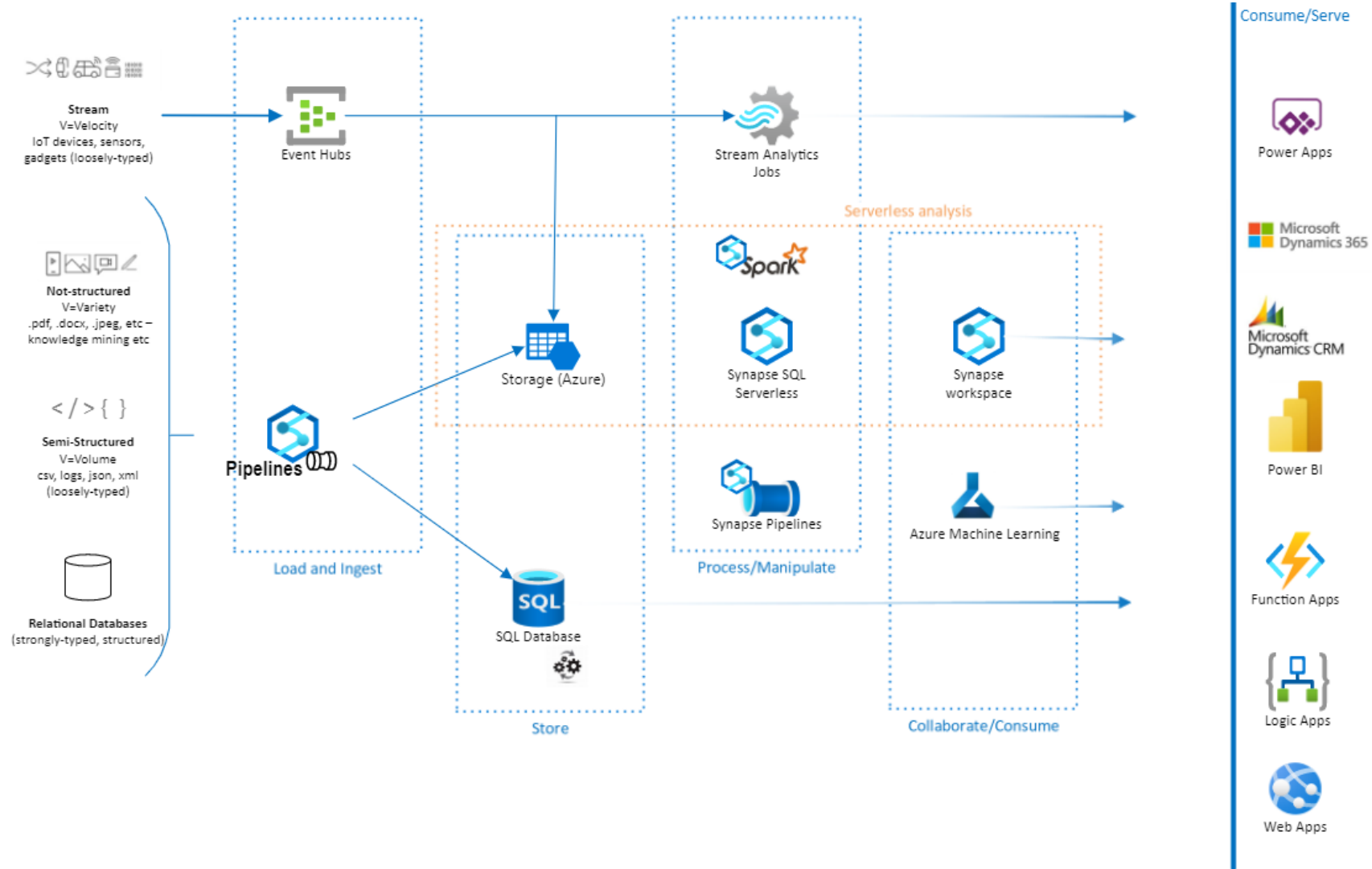
- BI, Data and Analytics SharePoint Site:
<https://<client>.sharepoint.com/sites/BIDataAnalytics>
- Microsoft Architecture:
<https://docs.microsoft.com/en-us/azure/architecture/browse/>
- Gartner Analytics Architecture:



Azure

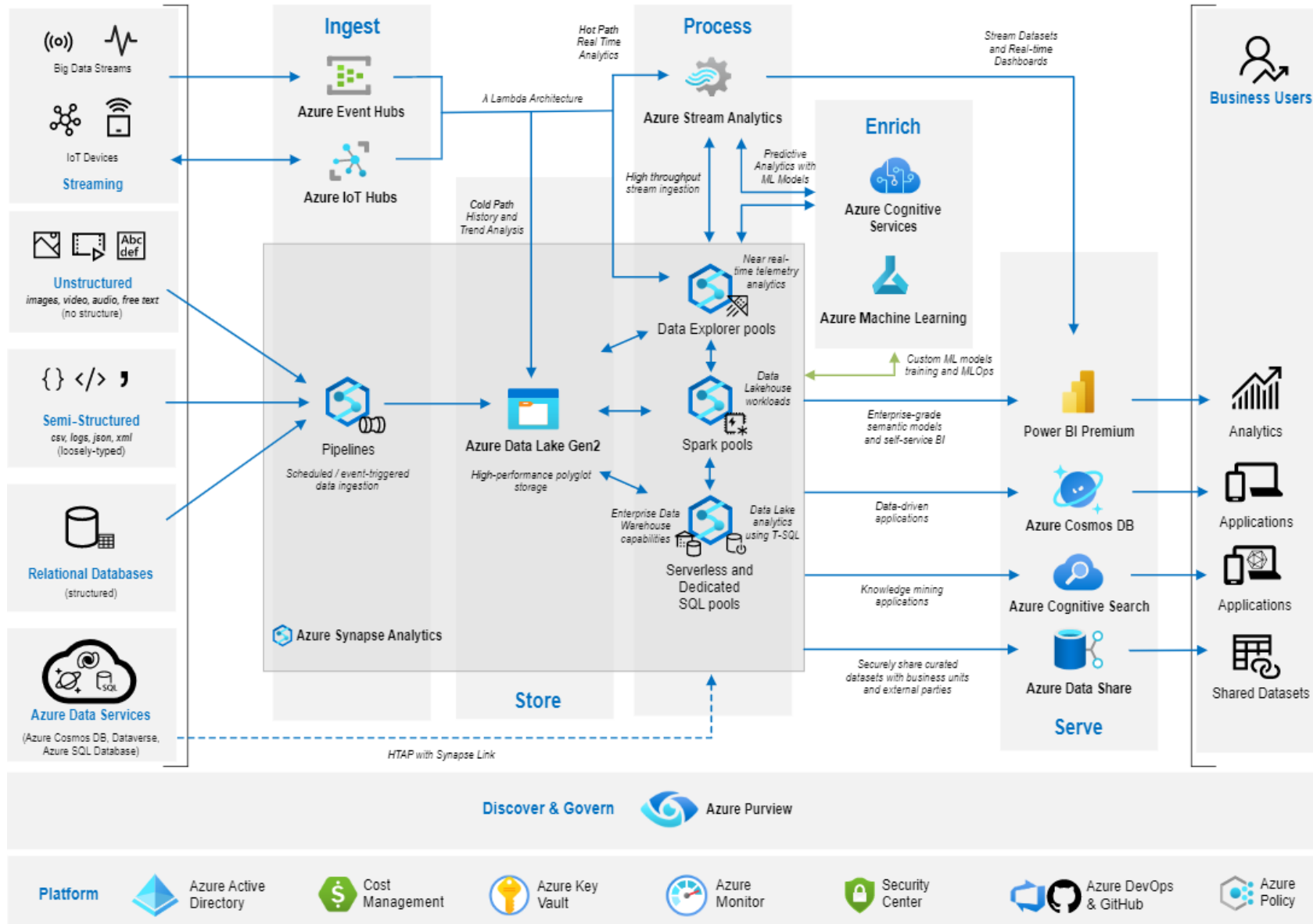
Modern Data Warehouse – Small Medium Business (SMB) pattern

Azure SQL DB and Synapse workspace



<https://docs.microsoft.com/en-us/azure/architecture/example-scenario/data/small-medium-data-warehouse>

Complete Azure Synapse Architecture



Appendix:

- Power Platform – Low Code/No Code
 - Power Automate
 - Power Apps
 - Power Virtual Agent
 - Power BI
- <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/load-data-overview>

- Dashboard
- All services
- FAVORITES
- Azure Active Directory
- Users
- Enterprise applications

Dashboard > Enterprise applications

Enterprise applications | All applications

ERICFRAYER - Azure Active Directory

Overview

- Overview
- Diagnose and solve problems

Manage

- All applications
- Application proxy
- User settings
- End user app launchers

Security

- Conditional Access
- Consent and permissions

Activity

- Sign-in logs
- Usage & insights
- Audit logs
- Provisioning logs
- Access reviews







[+ New application](#)
[Refresh](#)
[Download \(Preview\)](#)
[Preview info](#)
[Columns](#)
[Preview features](#)
[Got feedback?](#)

View, filter, and search applications in your organization that are set up to use your Azure AD tenant as their Identity Provider.

Search by application name or object ID

Application type == **Enterprise Applications** X
 Applications status == **Any** X
 Application visibility == **Any** X
 [Add filters](#)

6 applications found

Name	Object ID	Application ID	Homepage URL	Created on
 Microsoft Clarity	0086daa3-1c2c-40ee-80fb-75...	2f5fdc21-34c4-4039-881e-b8...	https://clarity.microsoft.com	2/25/2022
 Microsoft Business Sol	03585afc-af7e-4f30-b938-08...	dd4c082e-4e78-4792-a772-4...		4/22/2022
 databricks-service-app	2423e4ca-d9a5-44a2-b24f-0...	40c8e351-a771-44df-bffd-cd...		1/31/2022
 power-bi-scanner-app	5ff97108-0a7b-498a-a1c2-99...	087e52a0-1922-4051-be83-0...		4/26/2022
 iOS Accounts	828883a0-8fd5-4fe8-b2b9-0e...	f8d98a96-0999-43f5-8af3-69...		9/4/2019
 Power BI Community	d23500db-e2fb-4dd6-a81c-b...	04147409-c066-4a0c-964e-8...	https://community.powerbi.c...	11/17/2019

Dashboard

All services

FAVORITES

Azure Active Directory

Users

Enterprise applications

My Dashboard

Private dashboard

+ New dashboard | Refresh | Full screen | Edit | Export | Clone | Delete

ERICFRAYER

NETORGFT5008576.onmicrosoft.com



Azure AD for Office 365 | [Try Azure AD Premium](#)

Welcome to the Azure AD admin center



You already have Azure Active Directory. Use it to make Office 365 even better.

[See how Azure AD can help your organization](#)

Azure AD quick tasks

- [Add a user](#)
- [Add a guest user](#)
- [Add a group](#)
- [Find a user](#)
- [Find a group](#)
- [Find an enterprise app](#)

Users and groups



Users Sign-ins

Sign-ins for 'All Users' between 4/4/2022 and 5/4/2022.

Apr 10 Apr 17 Apr 24 May

Recommended



Sync with Windows Server AD
Sync users and groups from your on-premises directory to your Azure AD

Azure portal

portal.azure.com

- Home
- Environments**
- Analytics
- Dataverse
- Power Automate
- Power Apps
- Resources
- Capacity
- Dynamics 365 apps
- Portals
- Help + support
- Data integration
- Data (preview)
- Policies
- Admin centers

+ New Refresh Recover deleted environments

Search

Environments

Environment	Type	State	Region	Created on ↓	Created by
Think Data Analysis	Microsoft Teams	Ready	United States	02/13/2022 7:43 AM	Eric (info) Frayer
ERICFRAYER	Microsoft Teams	Ready	United States	02/13/2022 7:42 AM	Eric (info) Frayer
ERICFRAYER (default)	Default	Ready	United States	12/01/2019 11:17 AM	SYSTEM

Dynamics 365 apps

Feedback

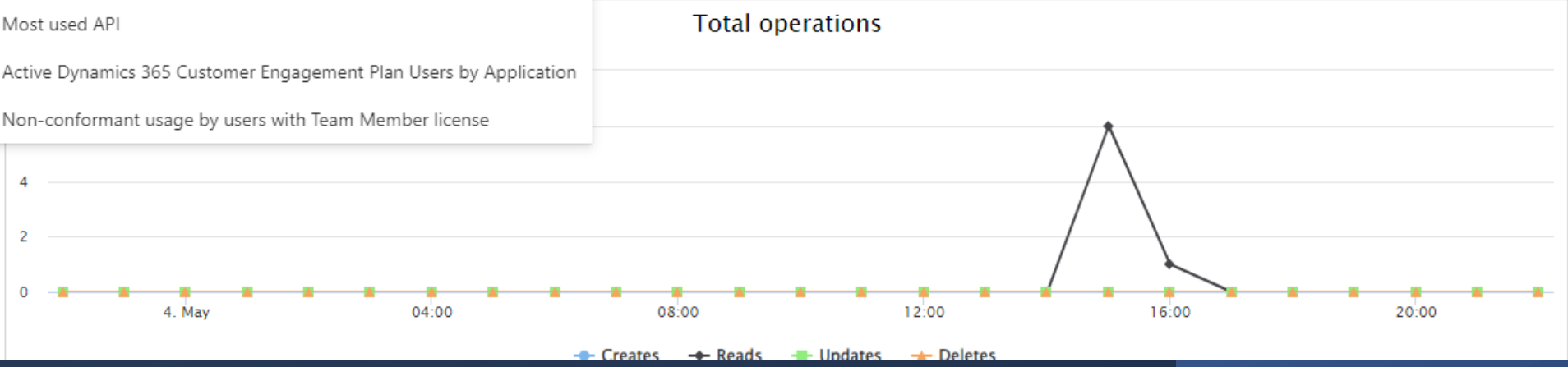
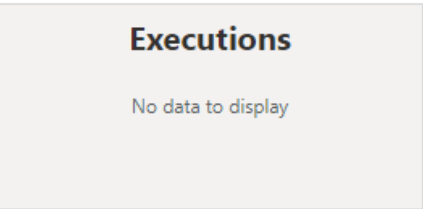
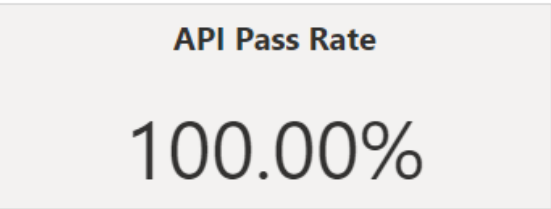
- Home
- Environments
- Analytics
- Dataverse**
- Power Automate
- Power Apps
- Resources
- Capacity
- Dynamics 365 apps
- Portals
- Help + support
- Data integration
- Data (preview)
- Policies
- Admin centers

Download Learn more - about Download reports

- Active users by device type
- Active users by business unit
- Active users by security role
- Active users by client
- Active users by entities
- Most active users performing operations
- Most used custom entities
- Most used OOB entities
- Most active workflows
- Most active plug-ins
- Most used API
- Active Dynamics 365 Customer Engagement Plan Users by Application
- Non-conformant usage by users with Team Member license

Plug-ins API calls statistics Mailbox usage

2022 6:00 PM to 5/4/2022 6:00 PM Change filters



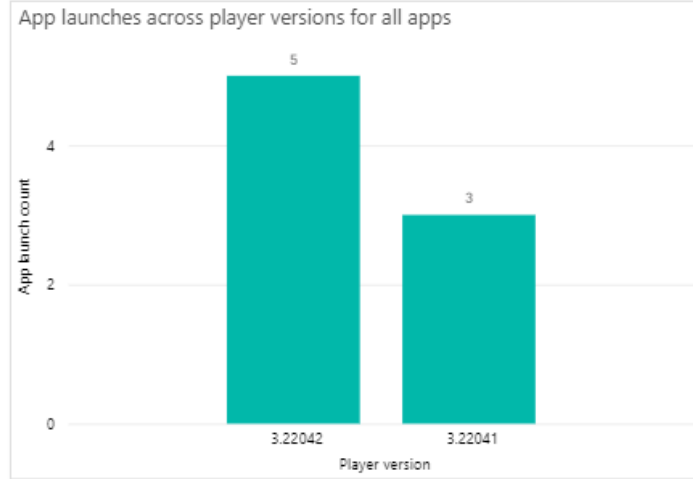
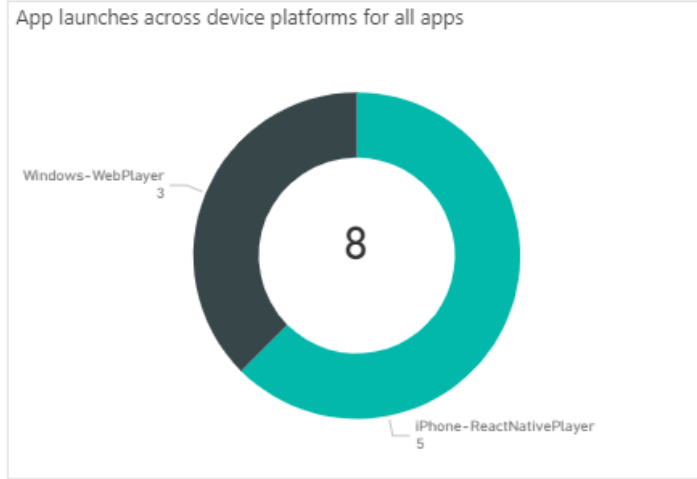
- Home
- Environments
- Analytics
- Dataverse
- Power Automate
- Power Apps**
- Resources
- Help + support
- Data integration
- Data (preview)
- Policies
- Data policies
- Tenant isolation (preview)
- Lockbox policy (preview)

Power Apps analytics

Learn more · Who can view these reports?

- Usage**
- Location
- Toast Errors
- Service Performance
- Connectors

Showing data for **ERICFRAYER (default) (orgef7ee3a1)** for the **Last 28 Days**. [Change filters](#)

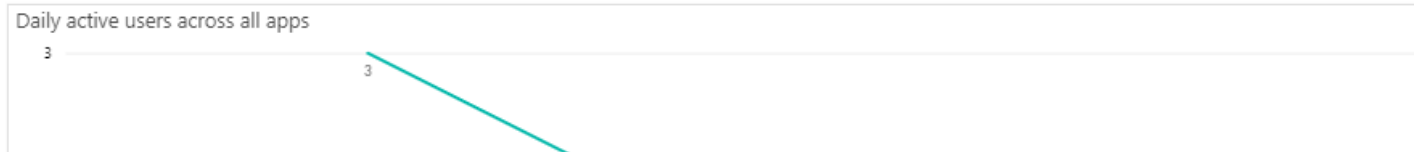


Device platform

- Select all
- iPhone-ReactNativePlayer
- Windows-WebPlayer

Player version

- Select all
- 3.22041
- 3.22042



Embedded BI

Eric (Info) Frayer

This is a page to share information and materials on how to "embed" data solutions into webpages. Being able to combine "knowledge management" tools like Confluence and SharePoint with data visualizations using Power BI, Qlik or Tableau has shown to improve adoption and decision making.

The CPG App below highlights a data set hosted in Azure and Snowflake and presented for analysis using Power BI.

- Pages
- Intro
- Executive Dashboard
- Sales
- Analysis
- Reporting
- StoryTelling
- Case Study
- About the Data
- Template



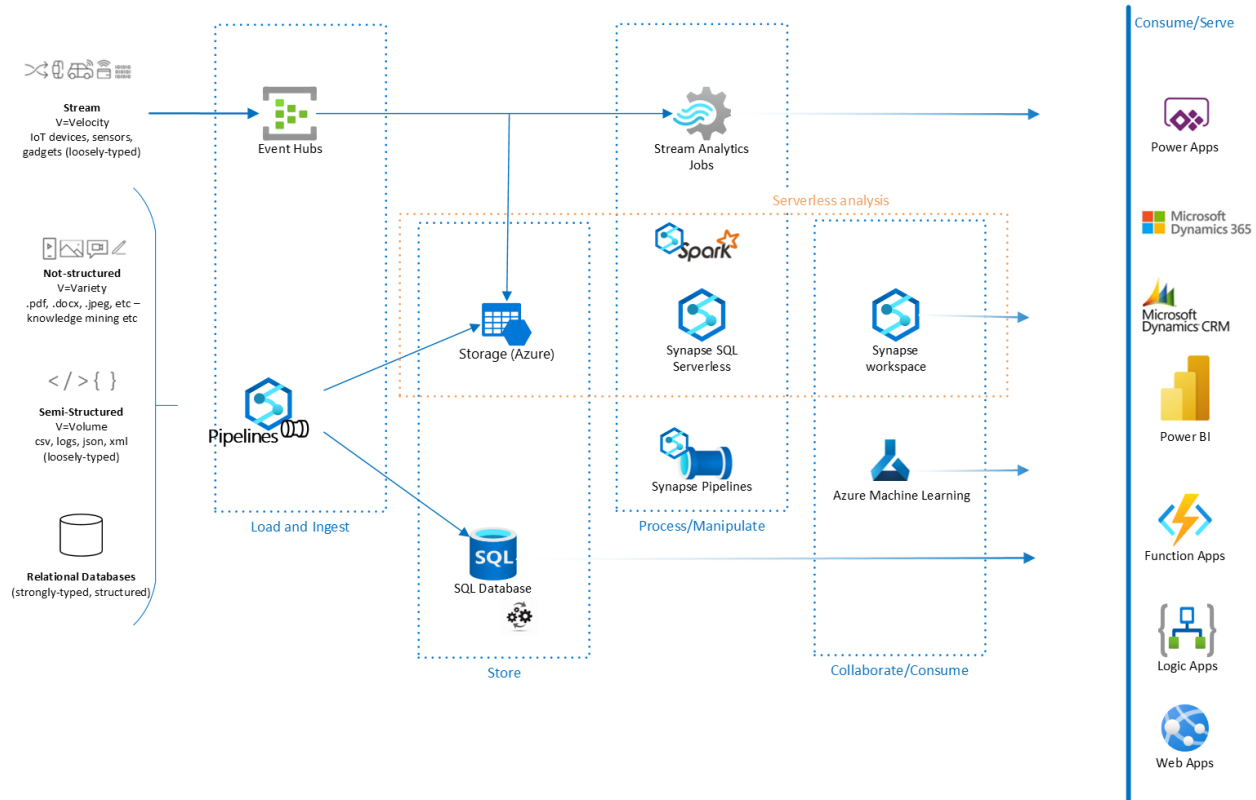
CPG Case Study

The data presented here is from an actual CPG company. It is "representative" of actual store sales, products and campaigns. All customer, product, vendor, store and company information has been anonymized.



Modern Data Warehouse – Small Medium Business (SMB) pattern

Azure SQL DB and Synapse workspace



Discover & Govern Azure Purview

Platform

