Microsoft Azure

Modern Data Analytics in Azure

Aitor Murguzur
Solutions Architect @ Microsoft
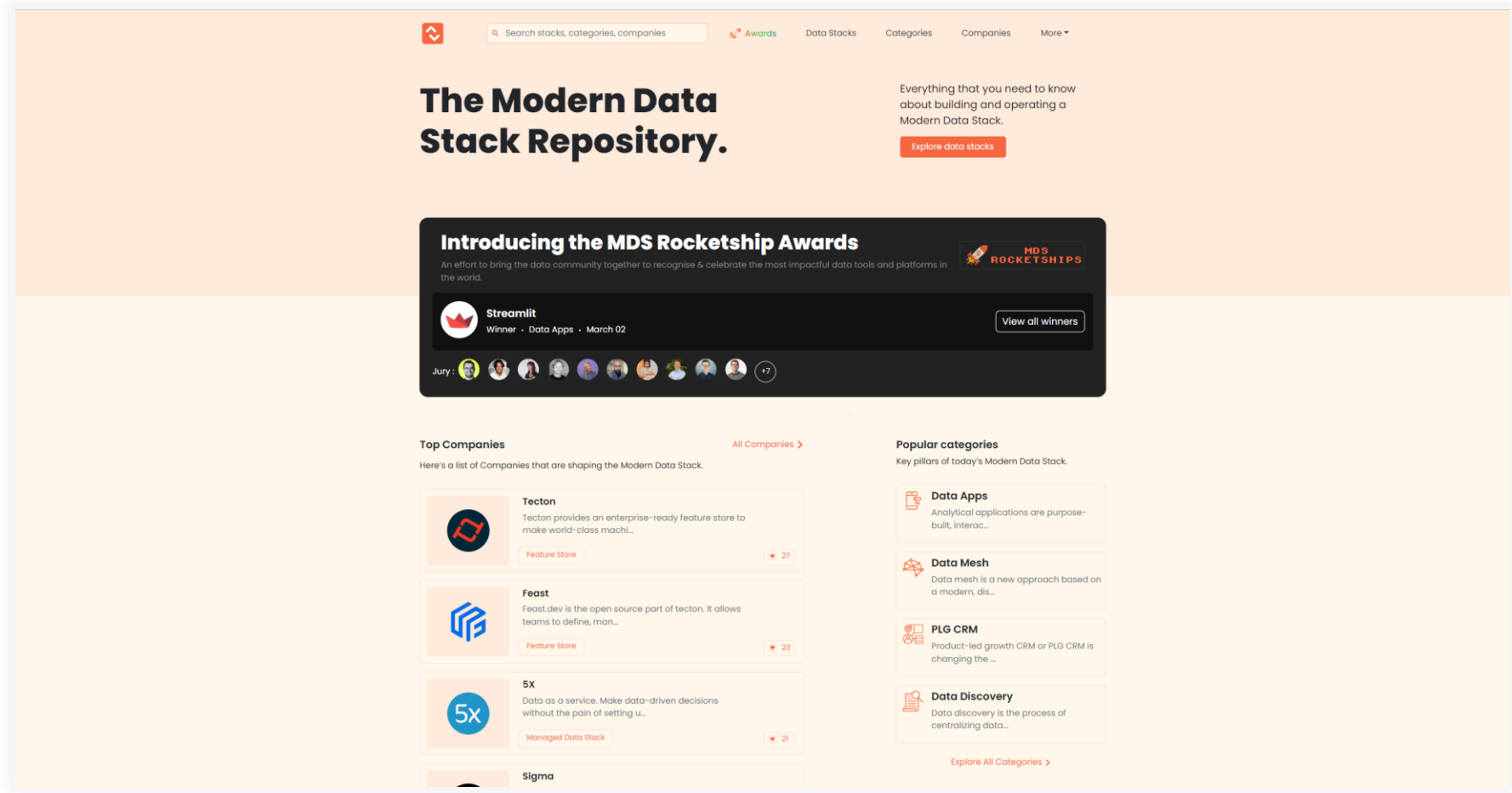
Hello! 👋

Aitor Murguzur
@murggu

Solutions Architect @ Microsoft

Ph.D. in Computer Science

Agenda

- Modern Data Stack

- Azure Data Stack

- Data Analytics Architectures

- Lessons learned & some reads

What is Modern Data Stack?

[Modern Data Stack - Everything that you need to know !](#)

MDS Categories

📦 Data Apps	📦 Data Mesh	📦 PLG CRM	📦 Data Discovery
📦 Reverse ETL Tools	📦 Data Privacy & Governance	📦 Data Cataloguing	📦 Operational Analytics
📦 Metrics Store	📦 Feature Store	📦 ETL Tools	📦 Data Streaming
📦 Spreadsheet based BI	📦 Business Intelligence (BI)	📦 Data Warehouses	📦 Data Modelling & TBC
📦 No Code Automation	📦 Data Lakes	📦 Event Tracking	📦 Data Workspace / Collaboration
📦 Data Quality Monitoring	📦 Workflow Orchestration	📦 Synthetic Data	📦 Business Reliability / Observability
📦 Augmented Analytics	📦 Customer Data Platform	📦 Product Analytics	📦 Managed Data Stack
📦 Change Data Capture			

Azure Data Stack

Analytics

Azure Synapse Analytics	Azure Purview	Azure Data Explorer	Data Factory
HDInsight	Azure Stream Analytics	Azure Data Lake Storage	Azure Analysis Services

AI + Machine Learning

Azure Databricks	Azure Machine Learning	Project Bonsai	Azure Cognitive Services
Azure Cognitive Search	Azure Bot Service	Kinect DK	Azure OpenAI Service

Storage

Azure Blob Storage	Azure Backup	Data Box	Archive Storage
Azure Files	Azure Data Share	Azure Cosmos DB	Azure SQL Database

Databases

Azure Database for PostgreSQL	Azure SQL Managed Instance	Azure Database for MySQL	SQL Server on Azure VMs
Azure Cache for Redis	Azure Database Migration Service	Azure MI for Apache Cassandra	Azure Database for MariaDB

This list is not exhaustive. All Azure products here Directory of Azure Cloud Services | Microsoft Azure

All Azure services

https://azurecharts.com/overview
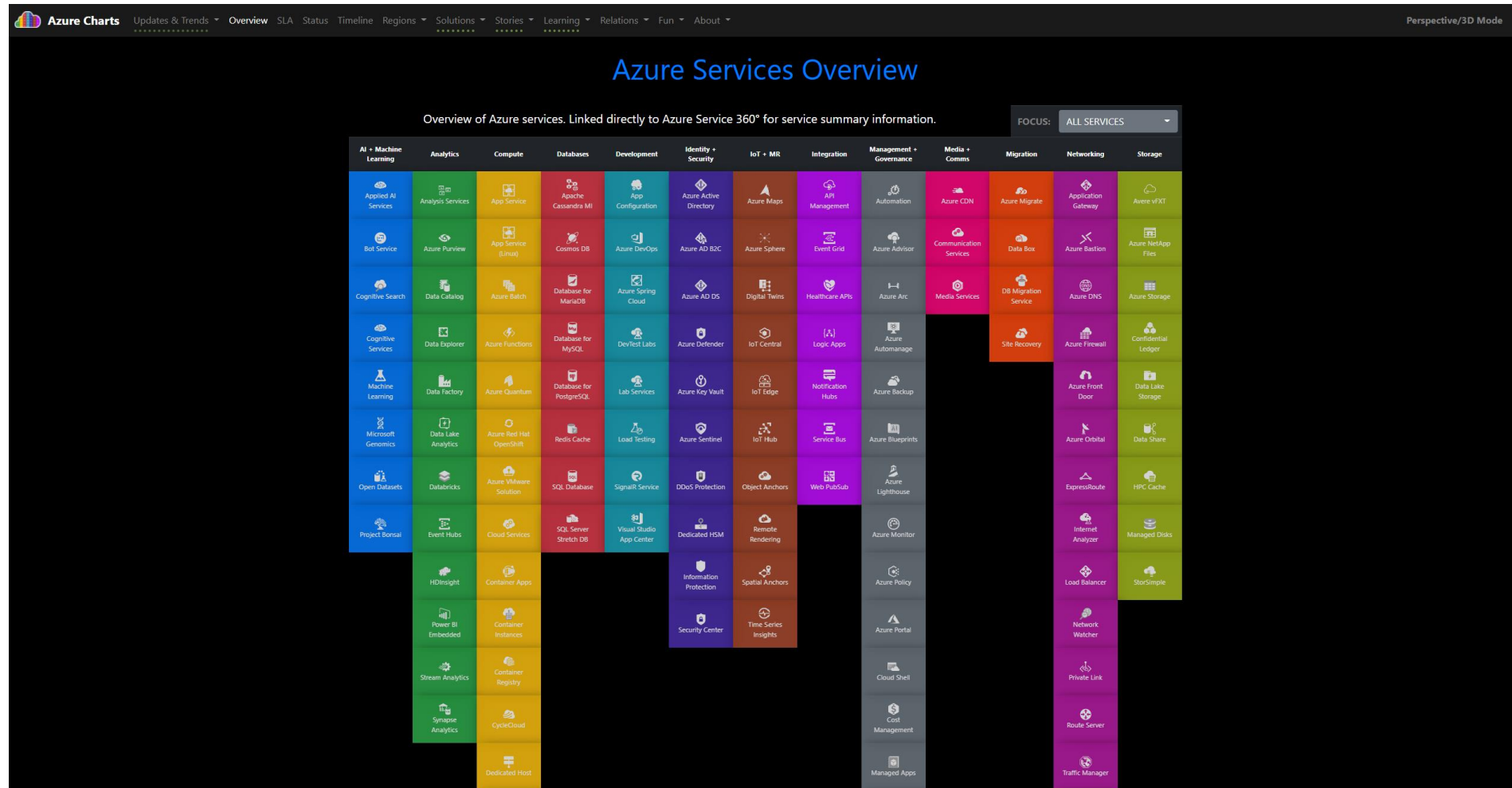
Data architecture - basic components
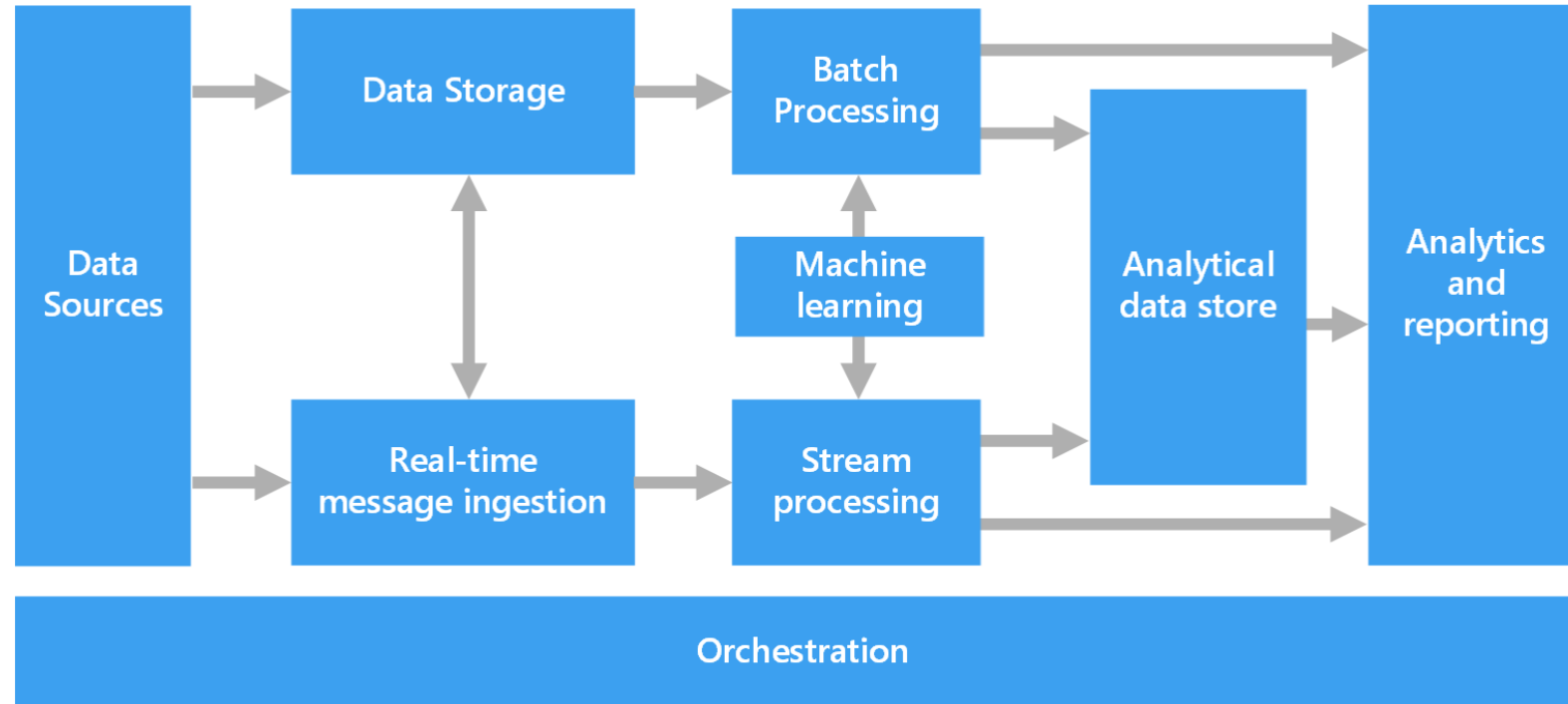
This could be complex

This example scenario demonstrates how to use Azure Synapse Analytics with the extensive family of Azure Data Services to build a modern data platform that's capable of handling the most common data challenges in an organization.
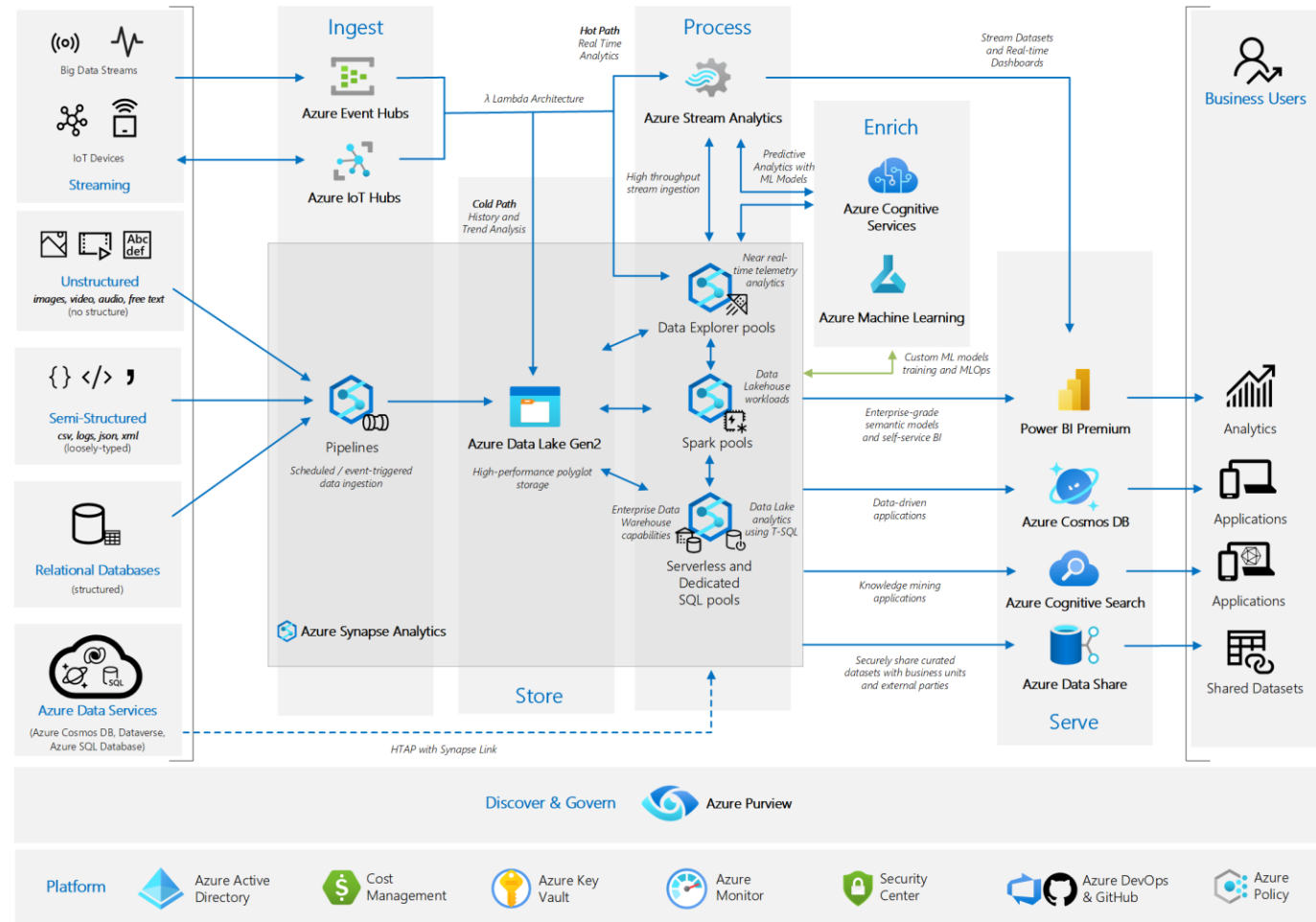
I´m small, do I need such complexity? Basic stack
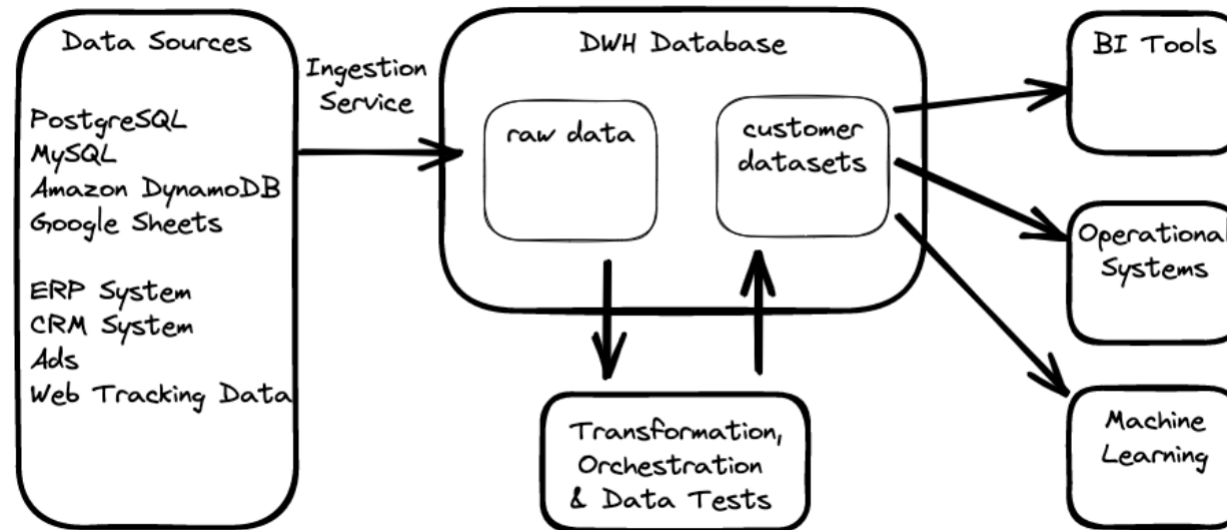
Basic stack / an example
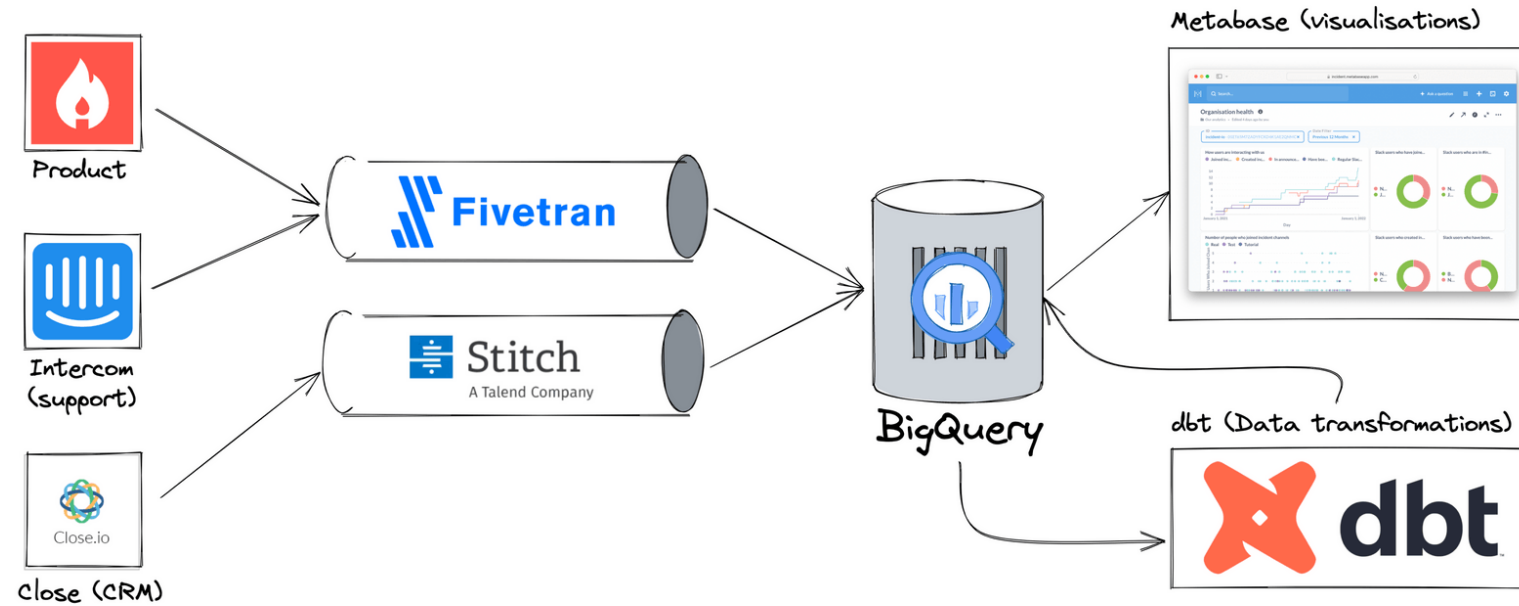
Basic stack / an Azure example
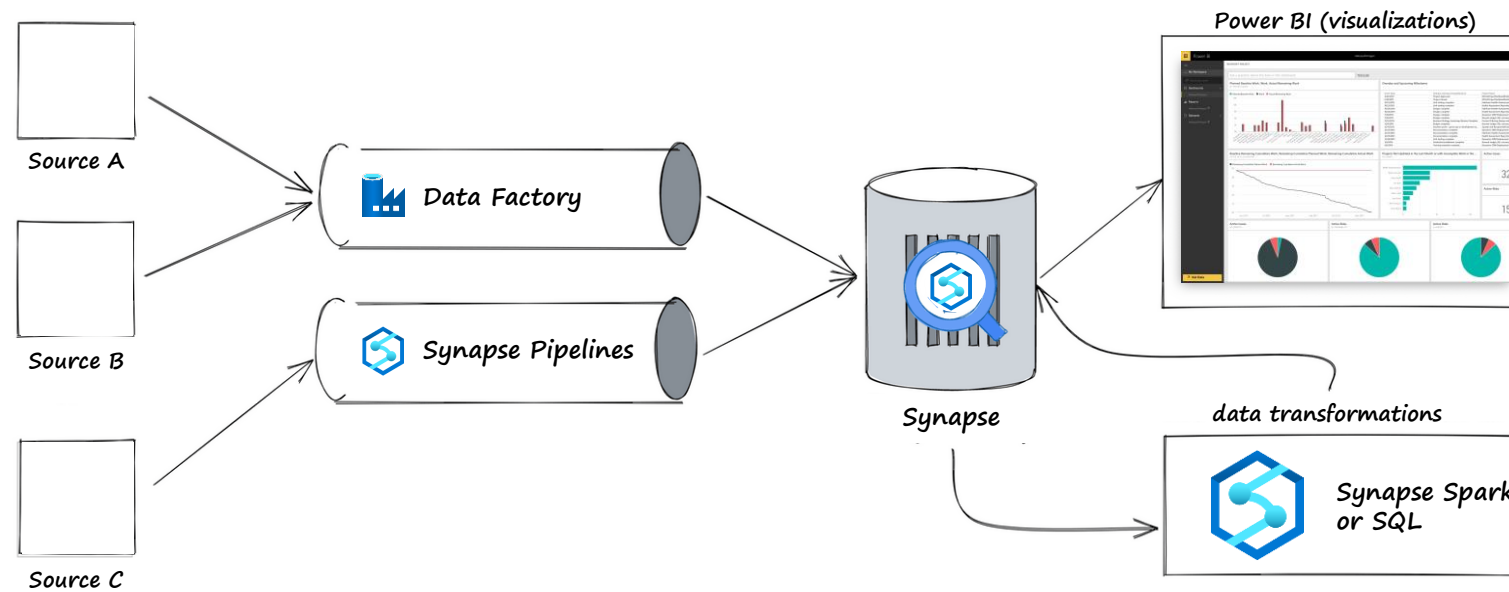
This is just an example, there are multiple options

Okay, but in reality – requirements come first

Data Sources & Datasets	Streaming Ingestion	Batch Ingestion	Other Jobs / Ingestion
Data Storage & Modeling	Reporting	Other Data Consumers	Non-functional Requirements
Backup & Disaster Recovery	Non-Production Environments	DevOps / DataOps / MLOps	Monitoring & Auditing
Data Access	Connectivity / Cloud	Data Encryption & Masking	Authentication & Authorization
Data Lineage	Data Dictionary	Data Quality	Data Privacy
Cloud Competence Center	Responsible AI Office	Business divisions (incl. IT)	Other Supporting Domains

MVP mindset, show value quick

Depending on the use case different patterns / architectures
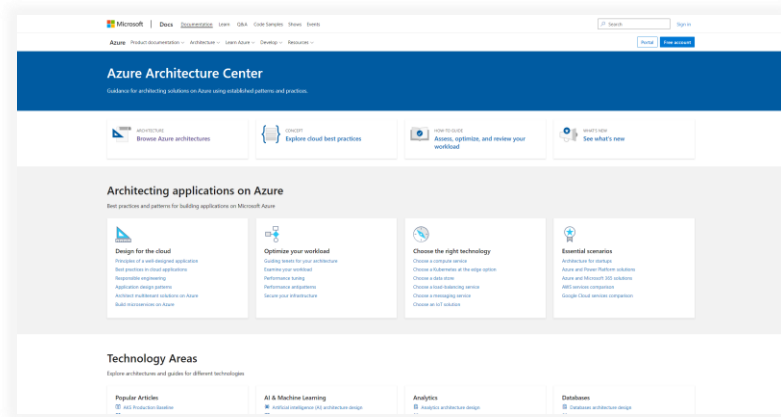
[Azure Architecture Center - Azure Architecture Center | Microsoft Docs](#)
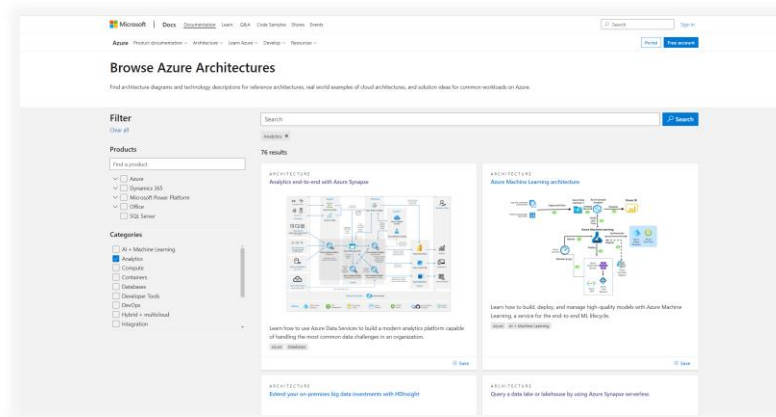
[Browse Azure Architectures - Azure Architecture Center | Microsoft Docs](#)
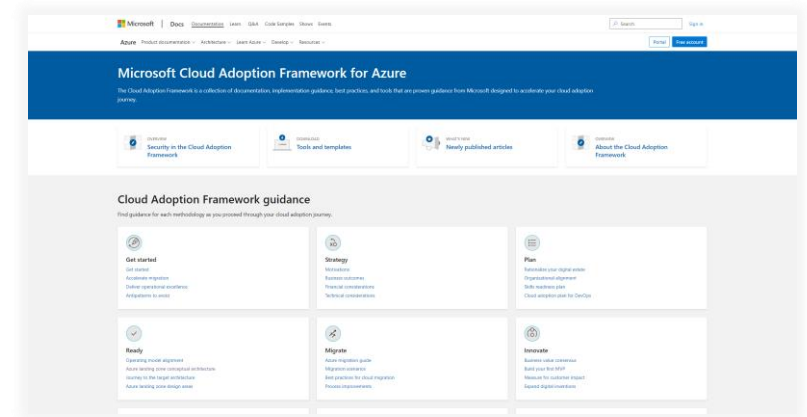
[Microsoft Cloud Adoption Framework for Azure - Cloud Adoption Framework | Microsoft Docs](#)

Lessons learned

- **Start small, iterate quickly**
  - **Provide value quick**, so people can see value from data/data products
  - From **pilots/POC** to gain momentum to **production**
  - **Learn/hands-on with the services/products**, pick up the most suitable one
  - Don´t **set up for failure**. You can´t go from 0 to "macro data mesh architecture" all at once
  - **Buzzwords** and **hype**, "you mean lakehouse"?
- **It´s about people and culture**
  - Data at **all levels** of the organization (sponsors, advocates)
  - Build an **in-house team**, focus on skills/ability gap
  - Broad data/AI **training**, no excuses
- **No one understand data, and what about quality?**
 - **Schema problems** don´t manifest immediately
  - Data pipelines needs to deal with **schema changes and data formats**
  - **Data quality** is an ongoing issue. Garbage in, garbage out
  - Domain knowledge is key, **understand the data**
- **Track all related data metadata**
  - Who owns what data, **data lineage**, data content, data access,…

Some trends & reads

- [Data & Analytics Trends to Watch in 2022 | by Taylor Brownlow | Geek Culture | Feb, 2022 | Medium](#)
- [The Modern Metadata Platform: What, Why, and How? (metaphor.io)](#)
- [Launching and Scaling Data Science Teams: Three Years Later | by Ian Macomber | Feb, 2022 | Medium](#)
- [Is BI dead? - by Benn Stancil - benn.substack](#)
- [Data Lake or Data Warehouse or a Combination of Both — Choices in Azure and Azure Synapse | by Inderjit Rana | Microsoft Azure | Medium](#)
- [Azure Synapse and Delta Lake | James Serra's Blog](#)
- [Technically: the analytics stack](#)
- [19 Data Newsletters to Read in 2022 | Graphext](#)
- [State of the Cloud 2021 · Bessemer Venture Partners (bvp.com)](#)
- [Designing Data-Intensive Applications (DDIA) — an O'Reilly book by Martin Kleppmann (The Wild Boar Book) (dataintensive.net)](#)

And many more.

Thank you! Q&A

- Azure free pass: aka.ms/try-aml

Contact

- aimurg@microsoft.com | @murggu