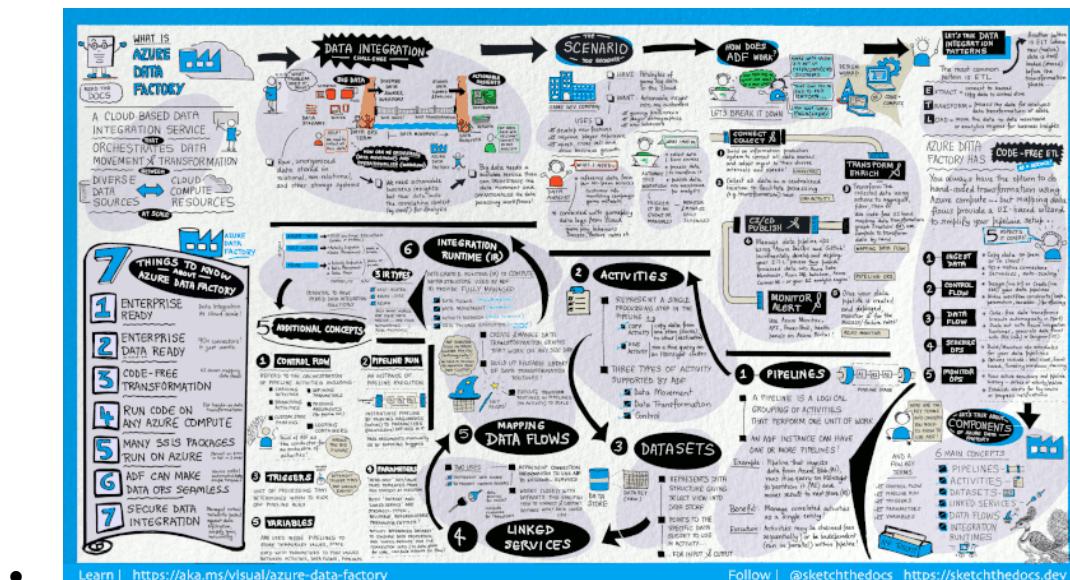


Azure Synapse Analytics

- Enterprise'owe narzędzie służące do przyspieszanie procesu „zaglądania do danych” w data warhouse'ach i innych systemach Big Data
- Integruje ze sobą SQL, apache Spark i inne usługi takie jak np. PowerBI
- Synapse SQL - distributed query system dla T-SQL, jest **serverless**
- Support dla C#
- Narzędzie do interaktywnego querowania danych, przy pomocy Data Explorera
- Jedno „kompletne” rozwiązanie

Azure Data Factory

- serverless data integration service
- UI z 90 wbudowanymi connectorami do data source'ów
- Wymaga konta azure(jak zapewne każdy tool azure 😊)
- Features(the rest in English, since it's much easier this way + interview is going to be in English anyway 😊):
 - Data Compression
 - Extensive Connectivity Support for Different Data Sources
 - Custom Event Triggers
 - Data Preview and Validation
 - Customizable Data Flows
 - Integrated Security
- It contains a series of interconnected systems that provide a complete end-to-end platform for data engineers



Stack technologiczny pod Interview odnośnie Big Data

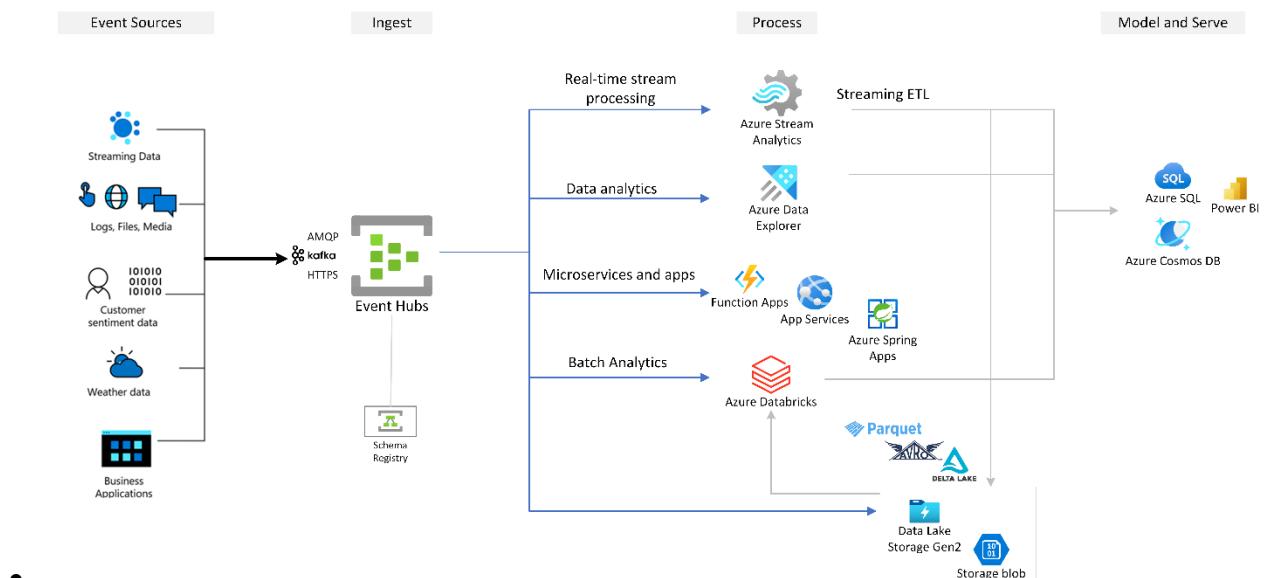
- **Copy Activity** cab be used in a data pipeline to move data from both on-premises and cloud source data stores to a centralization data store in the cloud
- Full support for CI/CD pipelines
- Built-in support for pipeline monitoring via Azure Monitor, API, PowerShell, Azure Monitor logs, and health panels on the Azure portal.

Azure Databricks

- It's a unified, open analytics platform for building, deploying, sharing, and maintaining enterprise-grade data, analytics, and AI solutions at scale.
- Use cases include:
 - Building an enterprise data lakehouse
 - ETL and data engineering
 - ML, AI and data science
 - LLMs and GenAI
 - Data warehousing, analytics, and BI
 - Data governance and secure data sharing
 - And so on...
- Has both enterprise and community editions
- It is based on Data Lakehouses
- Medallion architecture is promoted(Bronze -> Silver -> Gold) for data processing
- Heavily uses Apache Spark

Azure Event Hubs

- It's a native data-streaming service in the cloud that can stream millions of events per second, with low latency, from any source to any destination
- Businesses can use Event Hubs to ingest and store streaming data



Stack technologiczny pod Interview odnośnie Big Data

- It natively supports Advanced Message Queuing Protocol (AMQP), Apache Kafka, and HTTPS protocols.

Azure Stream Analytics

- It's a fully managed stream processing engine that is designed to analyze and process large volumes of streaming data with extremely low latencies
- A streaming data pipeline using Stream Analytics to identify patterns and relationships in data that originates from various input sources including applications, devices, sensors, clickstreams, and social media feeds, can be built using the tool
- Scenarios of usage:
 - Anomaly detection in sensor data to detect spikes, dips, and slow positive and negative changes
 - Geo-spatial analytics for fleet management and driverless vehicles
 - Remote monitoring and predictive maintenance of high value assets
 - Click stream analytics to determine customer behavior
 - Analyze real-time telemetry streams and logs from applications and IoT devices
- It's a fully managed (PaaS) offering on Azure
- Azure Stream Analytics uses a SQL query language that's augmented with powerful temporal constraints to analyze data in motion

Power BI

- Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights.
- Basic parts:
 - A Windows desktop application called *Power BI Desktop*.
 - An online software as a service (SaaS) service called the *Power BI service*.
 - Power BI Mobile apps for Windows, iOS, and Android devices.
- One common workflow in Power BI begins by connecting to data sources in Power BI Desktop and building a report. You then publish that report from Power BI Desktop to the Power BI service, and share it so business users in the Power BI service and on mobile devices can view and interact with the report.
- Desktop app is available
- In the Power BI service, you can use the deployment pipeline tool to test your content before you release it to your users.

POC(znowu po polsku) – fabryka samochodów i wykrywanie anomalii

1. **Azure Event Hubs** – zbieranie danych np. z IOT
2. **Azure Stream Analytics** – analiza danych niemalże w czasie rzeczywistym
3. **Azure Data Factory** – przetwarzanie danych np. co jakiś ustalony czas
4. **Azure DataBricks** – końcowa obróbka danych np. przy pomocy struktury medalionu i tworzenie pipeline'ów pod ML
5. **PowerBI** – tworzenie interaktywnych dashboardów, które mogłyby być przedstawiane klientom