



Architecture Summary

Data Sources:
 primary data sources
HubSpot: Storing its data in PostgreSQL.
Five9: Housing its data in Amazon RDS.
 Additionally, we also collect **semi-structured flat files** and **unstructured data**.

Ingestion:
 Data ingestion process encompasses data from the above sources and for orchestration and processing, we deploy **Azure Data Factory (ADF)** and **Azure Databricks**.

Storage:
 The ingested data finds its home in Azure Data Lake Storage (**ADLS Gen2**, a premier high-performance polyglot storage platform that ensures expansive scalability and robust data lake capabilities).

Process/Transformation
 We harness the dedicated SQL pool in Azure Synapse Analytics using its MPP power for our data processing and transformation endeavors. Techniques such as **PolyBase, CTAS (Create Table As Select)** for **merge operations** to work seamlessly and incremental loads to distributed tables.

Raw Zone: The initial landing zone for ingested raw data.

Cleansed Zone: Contains data post basic cleaning and preprocessing

Archive Zone: A repository for our historical data, meeting compliance and other long-term archival requirements.

Aggregation Zone: Hosts data that's been aggregated or summarized, primed for in-depth analysis.

- Platform**
- Azure Active Directory
- Microsoft Cost Management
- Azure Key Vault
- Azure Monitor
- Microsoft Defender for Cloud
- Azure DevOps & GitHub
- Azure Policy

