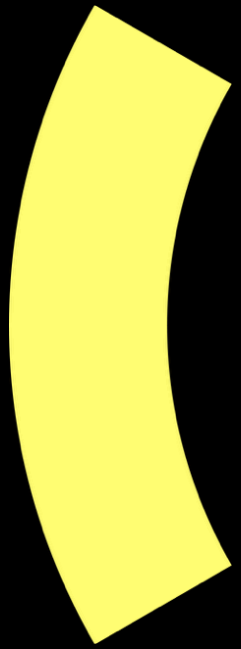
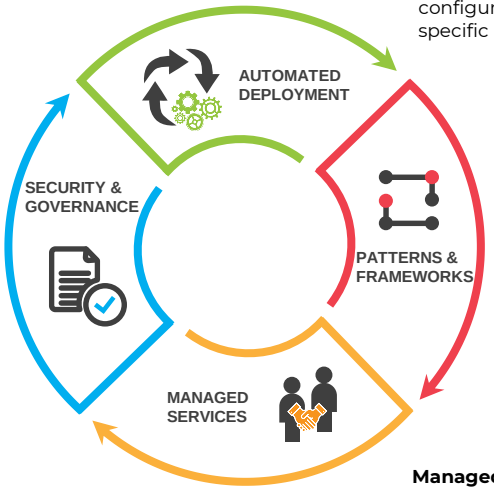


ARQ Group Experience: Solution Architecture with Microsoft Azure

Automated Deployment

The creation and configuration of the Cloud Analytics Platforms are automated using Arq Groups proprietary ADO Framework. The approach reduces the time required to provision that platform, improves governance, and reduces the risk of error due to human intervention.



Secure & Governed

The Cloud Analytics Platform leverages Azure's robust application security and governance from first point of data ingestion to output. Arq Groups Cloud Analytics Platforms support data encryption at rest and in transit and are capable of enforcing role based security privileges by plugging into your existing identity management system.

This provides you with confidence that your data is secure and only accessible by authorised users.

Defined Patterns & Frameworks

Arq Groups Cloud Analytics Platforms are configured to deliver against the specific customers user stories using defined patterns and frameworks.

Based on years of industry experience, these patterns and frameworks accelerate the configuration of the platform to enable your specific outcomes, reducing time to value.

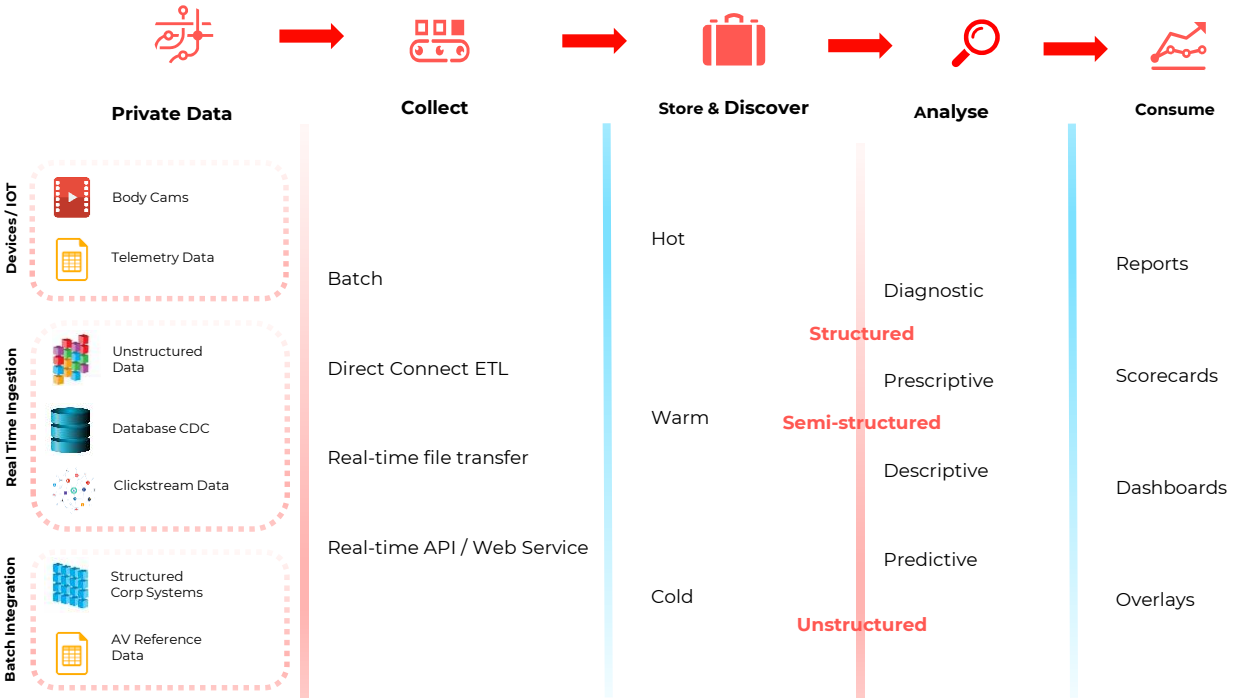
Managed Service

Arq Groups Cloud Analytics Platforms are offered as a managed service, including service desk, maintenance and support, security, and service management. This allows you to focus on the generated insights.

Arq Groups managed service offering includes:

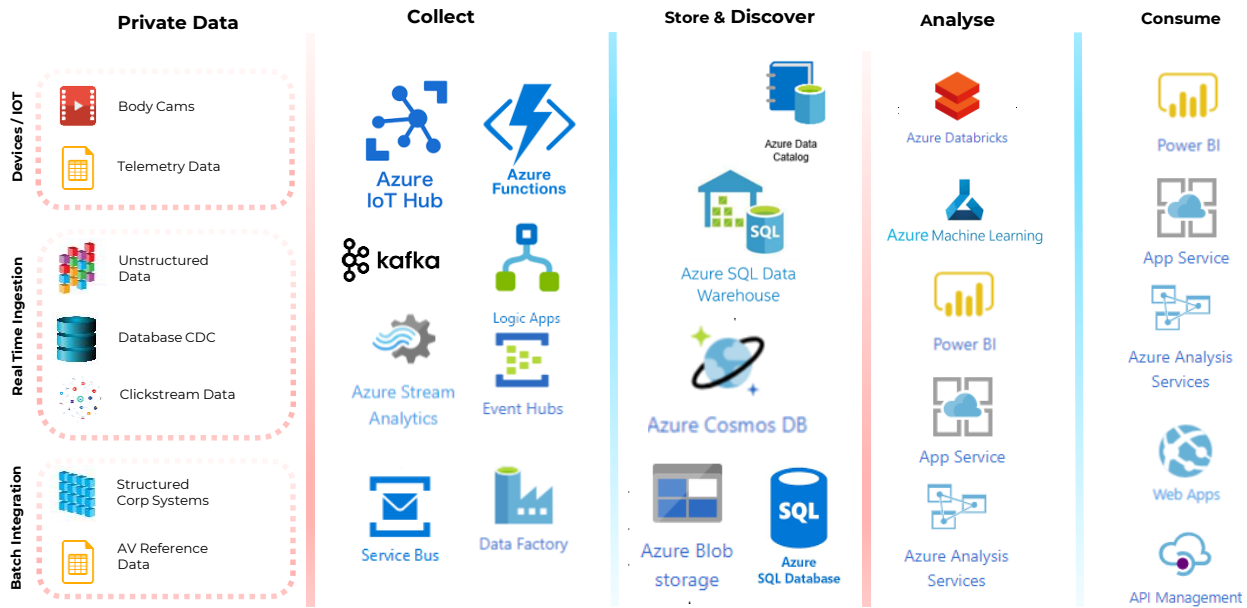
- **Service desk** - 24 x 7 local support and monitoring, escalation path to resolution with deep SME capabilities
- **Maintenance Support** - patching, backups, and updates
- **Service Management** - continuous improvement, warranty support

Arq Groups Cloud Analytics Platform Principles

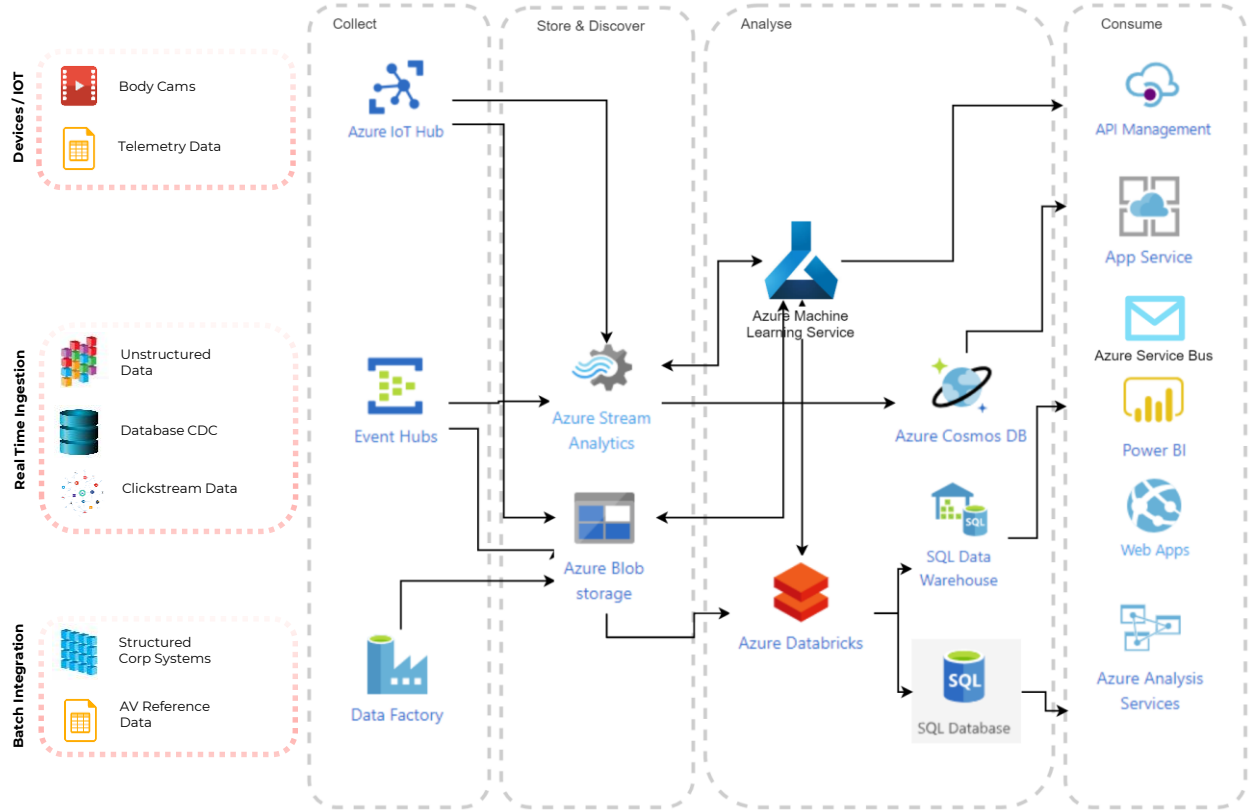


Enabling flexible pathways / Governed Data / Operational Data / Analytic Data / Structured / Semi-Structured

Arq Groups Cloud Analytics Data Process



Considered Architecture Components



ARQ group

Conceptual Architecture

Cloud Services continue to mature and innovate at a rate that is hard to match.

Where possible we look to take advantage of the scale of investment from the public cloud providers and focus on the outcomes we can drive from services which are well defined, governed, secured, and allow us to focus on delivering against the specific use cases.

Purpose built serverless offerings with flexible data ingestion methods. Enabling ingestion from a variety of sources and types with easy integration with common tools to further refine the data as needed.

Provides relatively cheap and scalable ETL capability for data transformation of large volumes where the data is not time critical.

Provides extendable AI and ML frameworks to enable predictive modeling and intelligent validation of data. Delivered as a PAAS offering it enables us to focus on the models instead of the infrastructure.

Horizontally scalable No-SQL database which understands and respects relational data and transactions. Delivered as a PAAS Platform to enable us to focus on what we can make the data do.

ARQ group

The Right Tool
In the Right Place
At the Right time

Sometimes the maturity of the cloud services isn't where it needs to be for a critical enterprise application.

In these cases we can take advantage of the more mature IAAS services and our internal capability to deliver a service to a customer as if it was a cloud native service.

talend

Talend MDM Platform

Market leading Master Data Management platform.

Giving you confidence in the governance and security of your data.

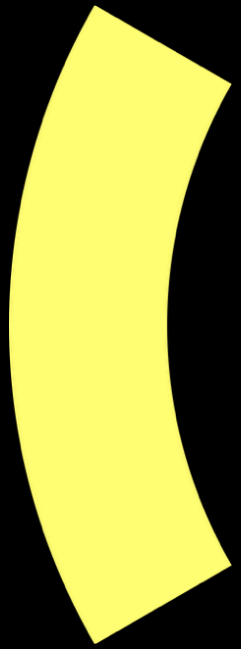
Intelligently tiering and structured data storage models sitting on top of the basic IAAS building blocks allows us to the balance the needs for long term and large file data retention with the speed and accessibility required for Realtime applications.

Custom development to enable rich, user experience lead, workflows and journeys for your users.

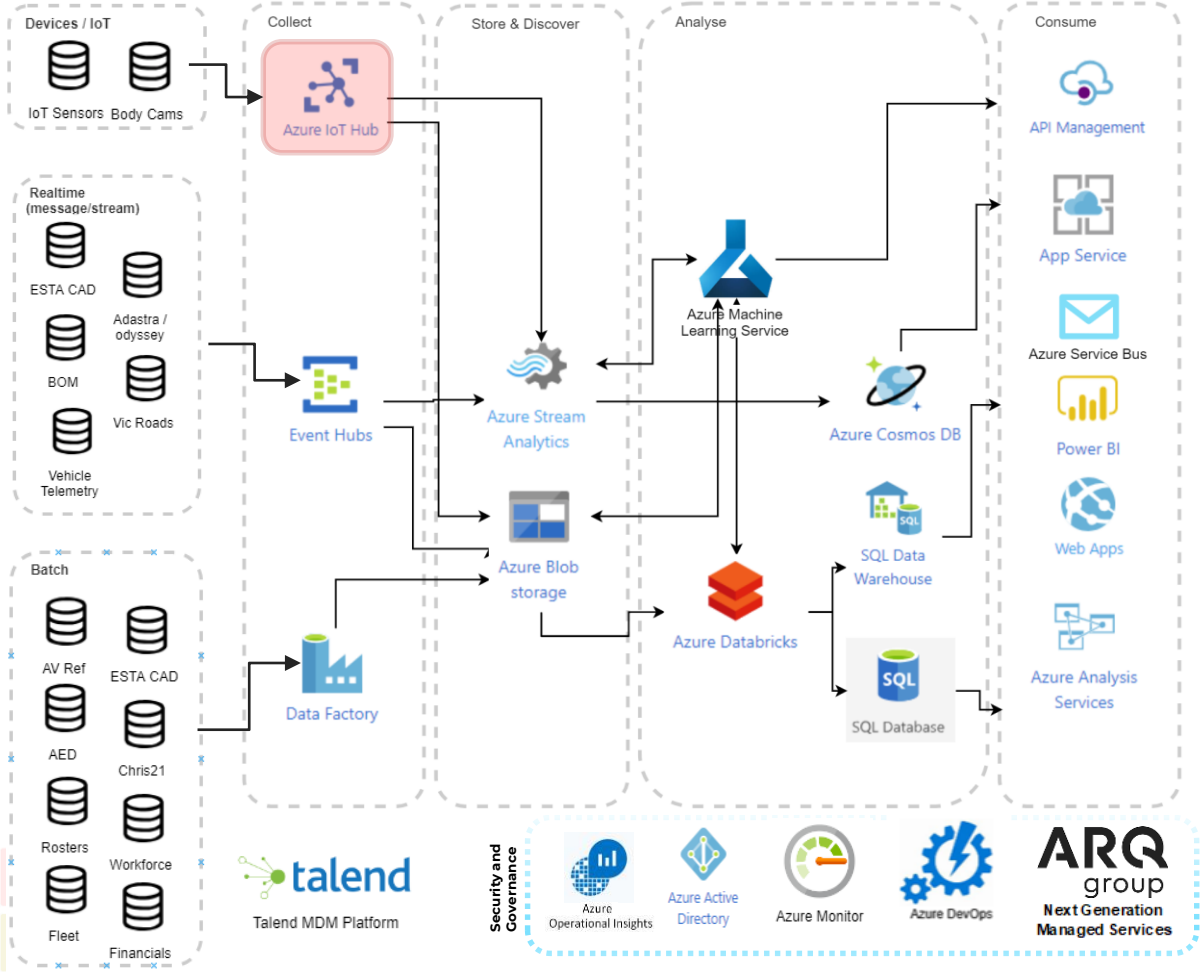
Mapping, Navigation, and Response Time requirements for such a critical system are just as important as the user experience and there isn't an off the shelf product we believe can deliver either of these key aspects with their current maturity of offering.

ARQ group

The Right Tool
In the Right Place
At the Right time



Platform Services



Service

Azure IoT Hub

Layer

Collect

Description

Ingestion of un/semi-structured data directly from large fleets of devices (e.g. body cams)

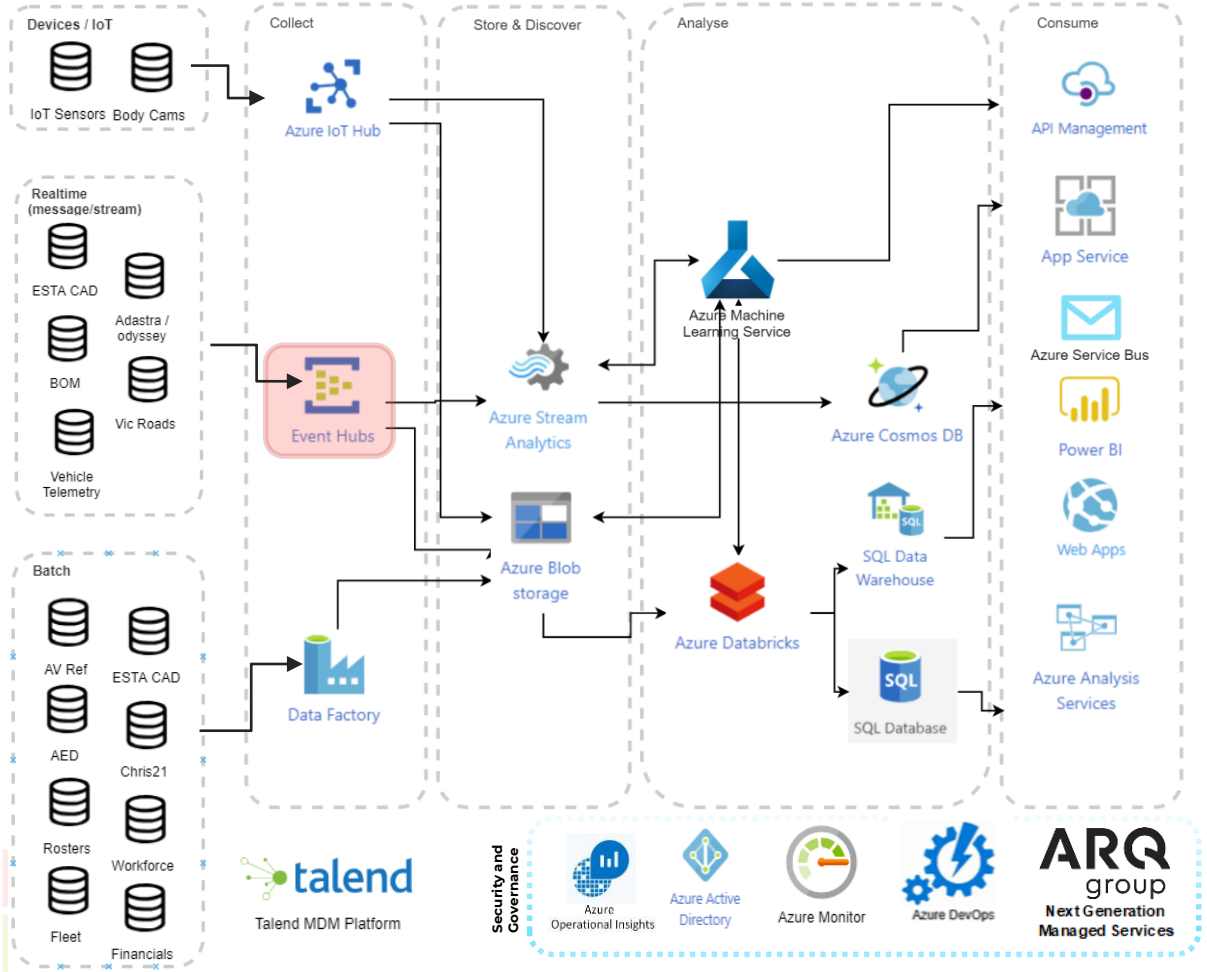
Scalable on-demand based on messages/day

Managed device connectivity and security

Support for a range of protocols

Design Considerations

Proposed assuming that Azure IOT Hubs is connected directly to devices



Service

Azure Event Hubs

Layer

Collect

Description

Ingestion of structured messages from aggregated sources (e.g. CAMS)

Scalable based on provisioned throughput

Kafka-compatible with broad skills pool

Configurable with highly durable configurations

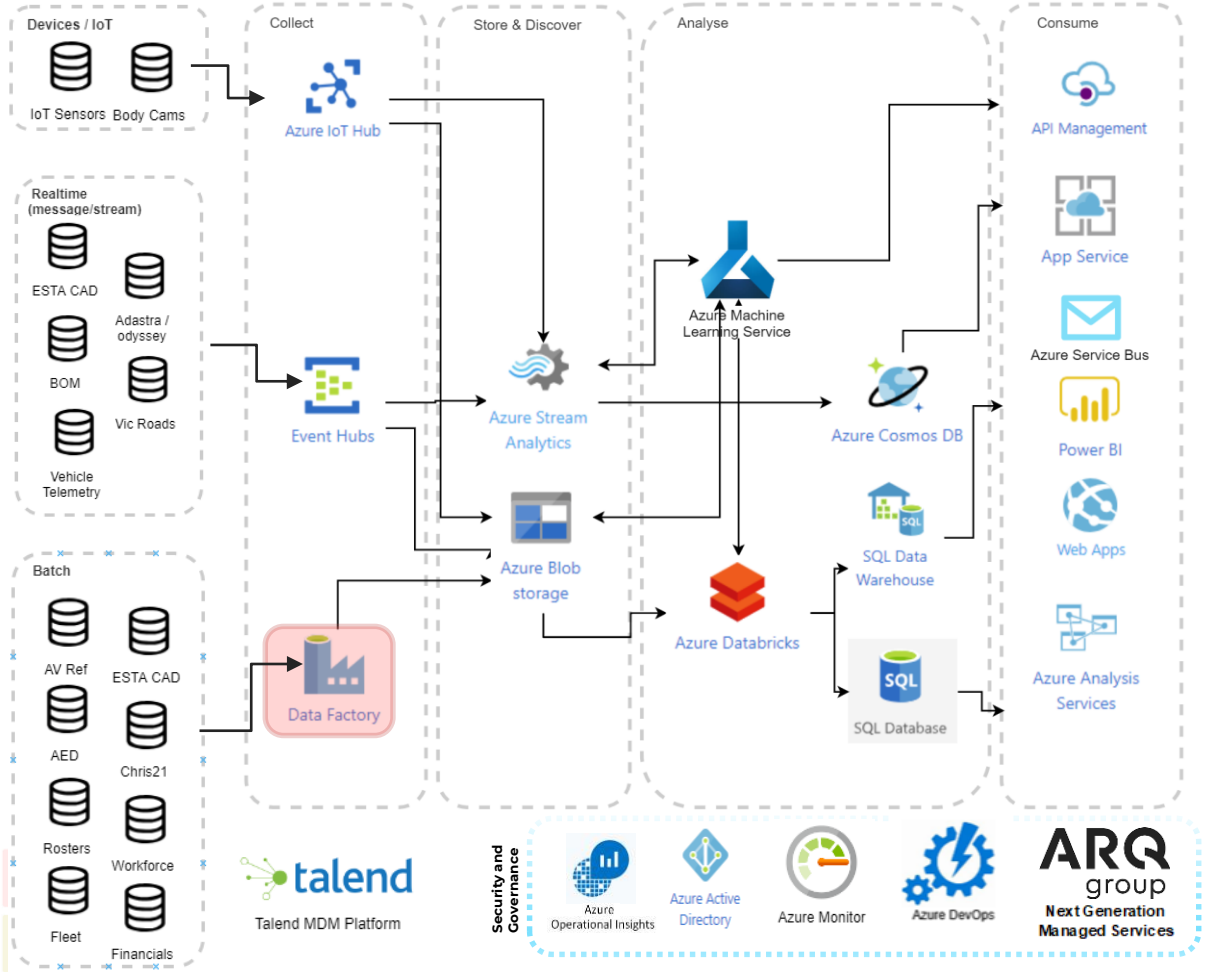
One-way at-least-once data streaming

Stream directly to Azure blob storage for a persistent record

Design Considerations

All real time source systems can push data.

For pulling data a service like Azure Functions could be used



Data Factory

Service

Azure Data Factory

Layer

Collect

Description

Ingestion of structured data from Batch sources

Scalable Integration Runtime (Compute) based on provisioned throughput

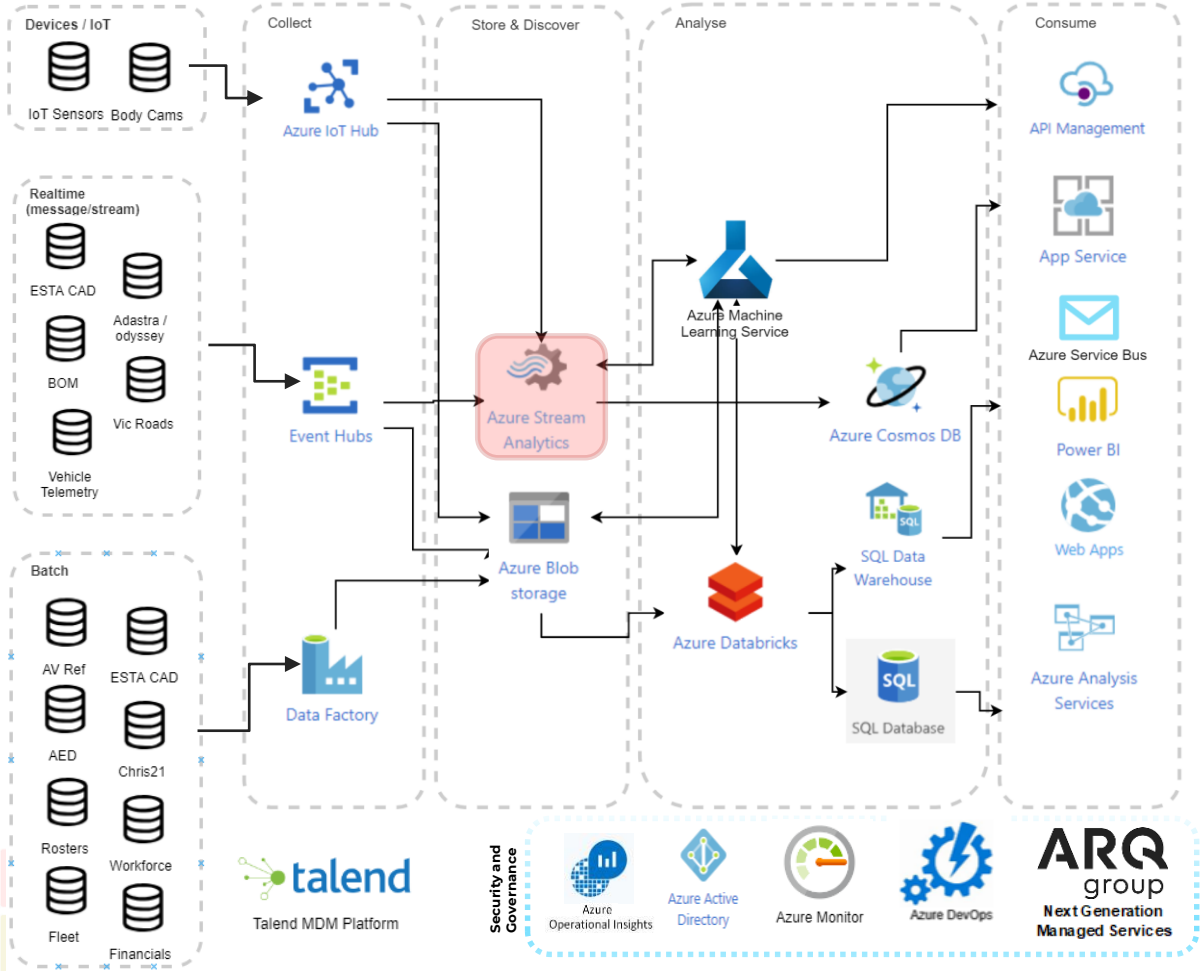
Can dispatch and monitor transformations running on other data services

Offers more than 80 natively supported connectors

Design Considerations

All data sources extracted are supported by standard Data Factory connectors. Any custom sources may require using services like Azure Functions to integate data

Specifically incorporated for batch oriented processing (periodic)



Azure Stream Analytics

Service

Azure Stream Analytics

Layer

Store & Discover

Description

Integrates with Event Hubs, IOT Hubs, Blob Storage as common streaming sources

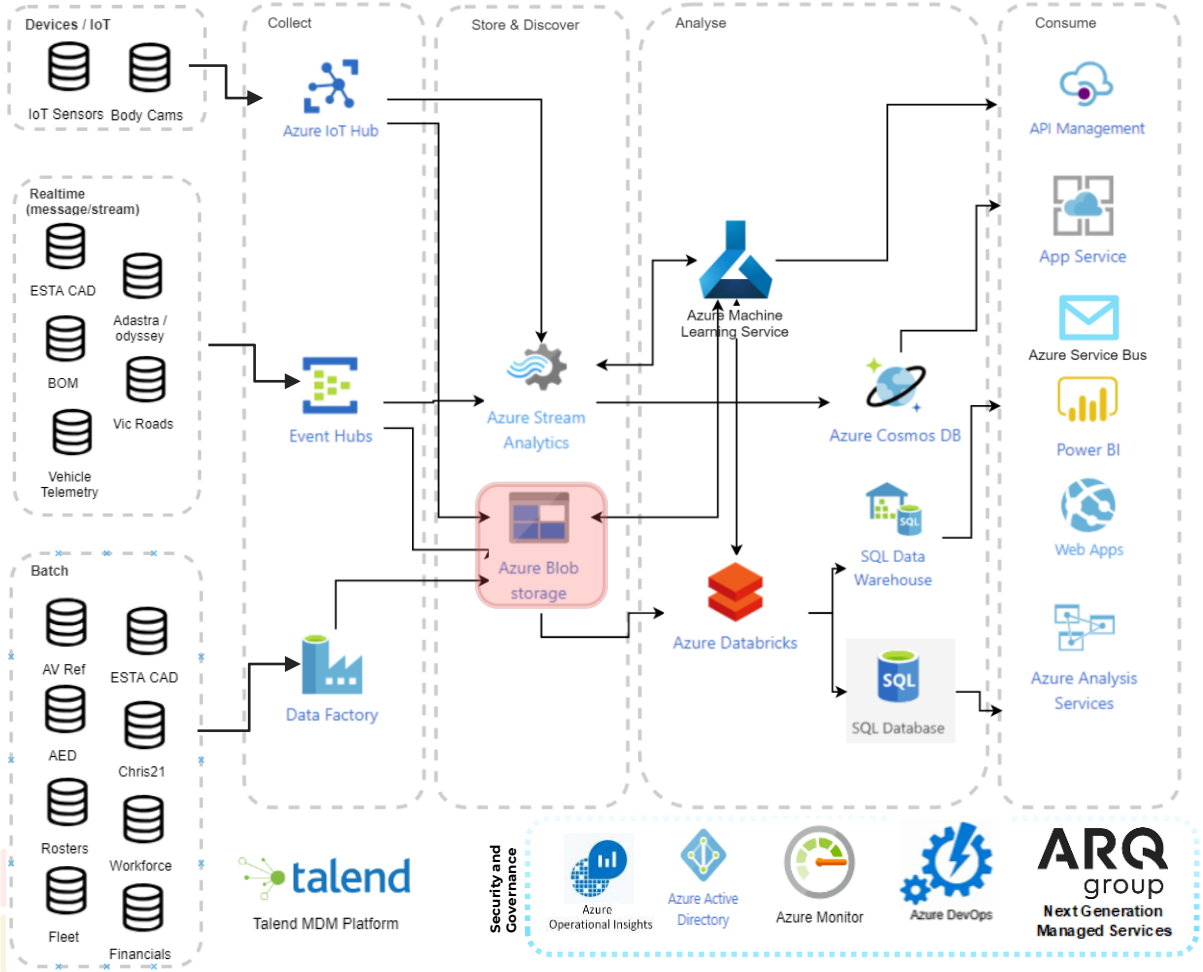
Scalable based on provisioned throughput

Stream directly to any of the Azure Data Services (Blob, Cosmos, SQL DW, Database etc)

Design Considerations

Azure Stream Analytics can be replaced with Azure Data Bricks. With Azure Databricks provide more powerful custom development and machine learning integration

Stream Analytics can be easier to implement but more expensive for large scale data volumes



Azure Blob storage

Service

Azure Blob Storage

Layer

Store & Discover

Description

Optimised for storing large volumes of unstructured data

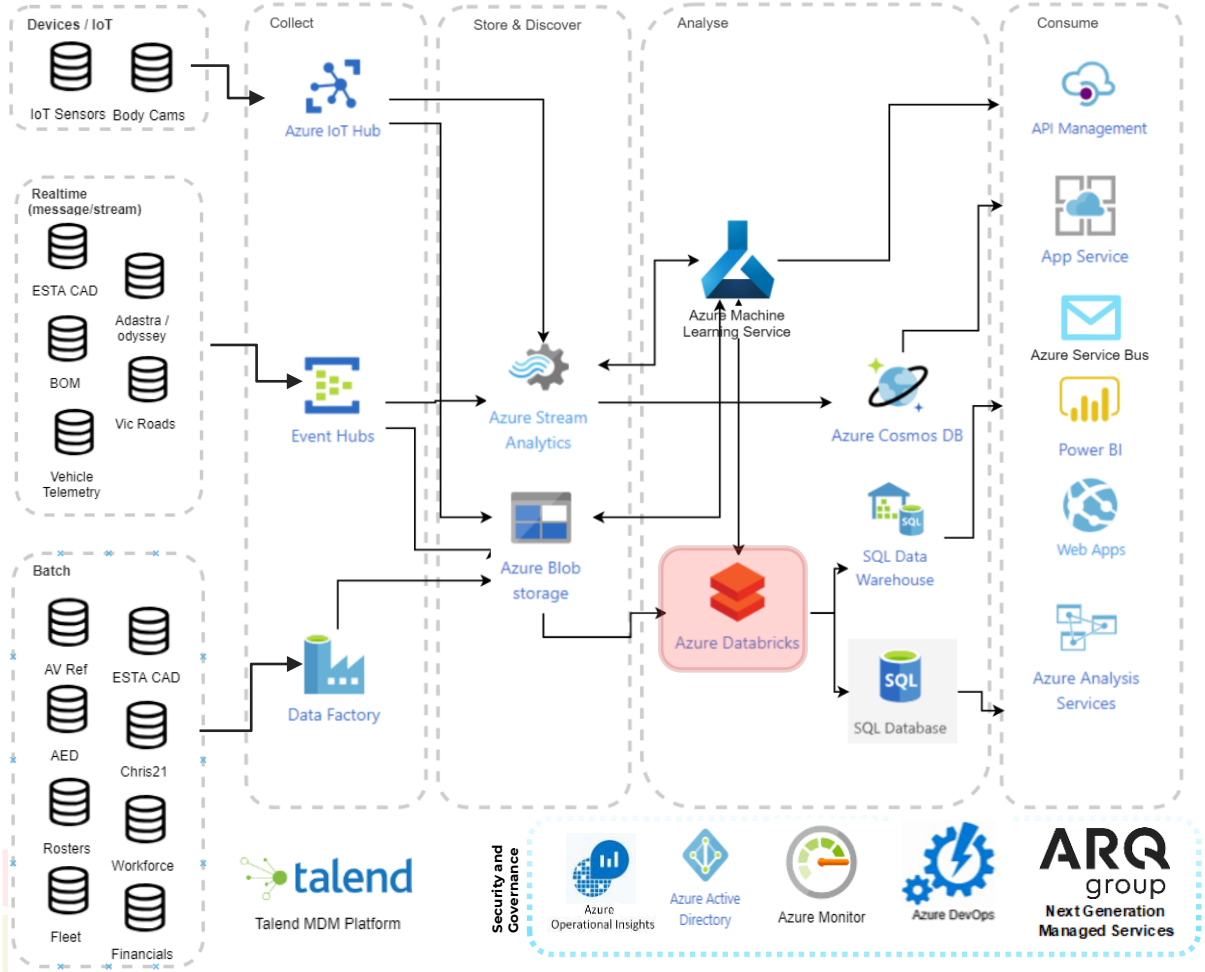
Highly durable and low cost data storage (Hot, Cool, Archive)

Good storage mechanism for documents, audio & video and log files

Easily scalable

Design Considerations

Blob storage is used to primarily store unstructured data types and non-relational data. SQL Database and Cosmos are better suited for fast retrieval of data and transactional querying



Service

Azure Databricks

Layer

Analyse

Description

Apache Spark based analytics service

Highly-flexible, scalable data integration

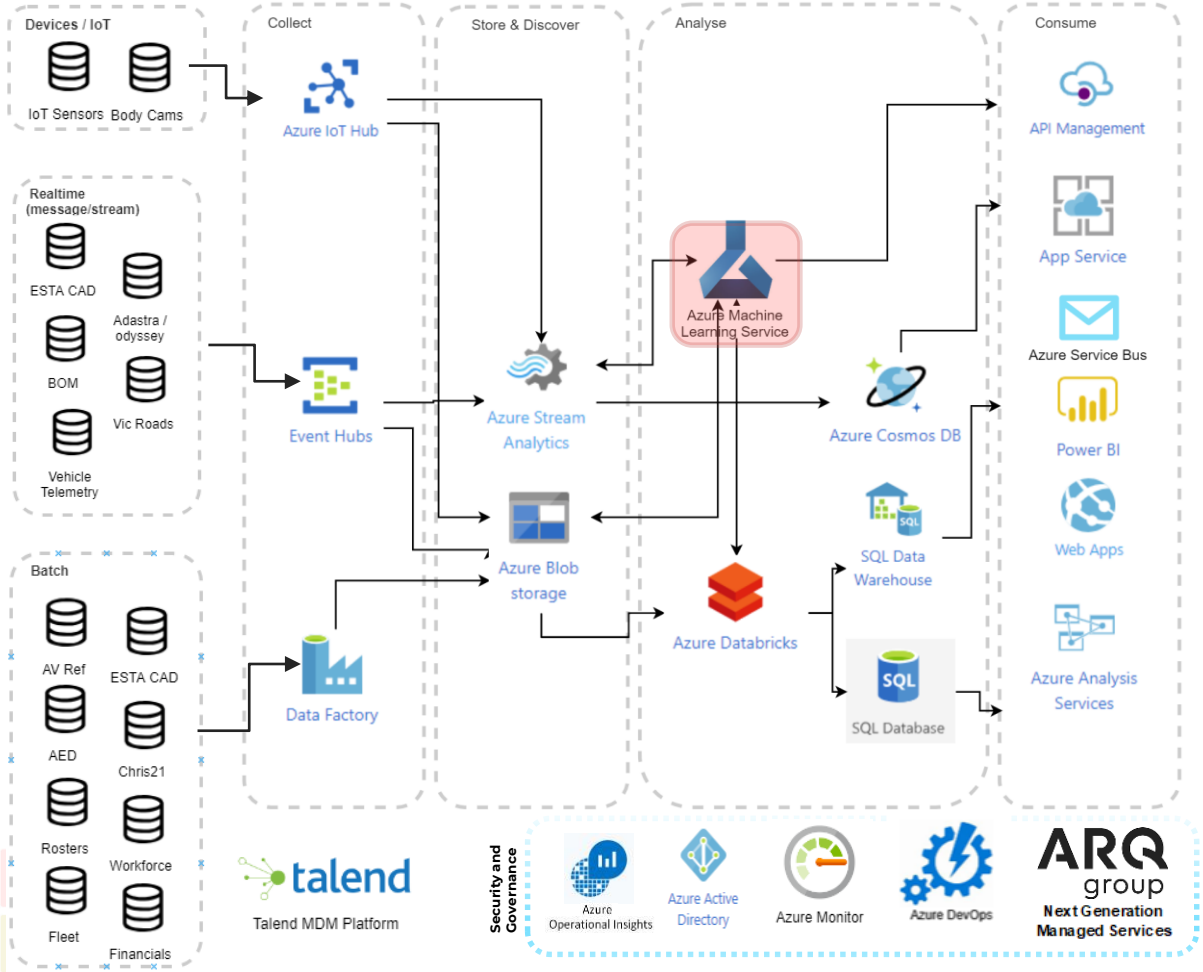
Can integrate data from multiple hot & cold sources and orchestrate other Azure services

Supports multiple scalable compute clusters with auto-shutdown features

Design Considerations

Azure Databricks provides batch processing and real-time capabilities with strong integrations with Azure data services and machine learning services.

Simpler uses cases may be supported using Azure Streaming Analytics



Service

Azure Machine Learning Services

Layer

Analyse

Description

This service provides SDKs and services to quickly prep data, train, and deploy machine learning models

Use automated machine learning to identify suitable algorithms

Deploy trained models as a web API

Integrated with Databricks and Stream Analytics for real-time streaming. Also Databricks for batch oriented processing.

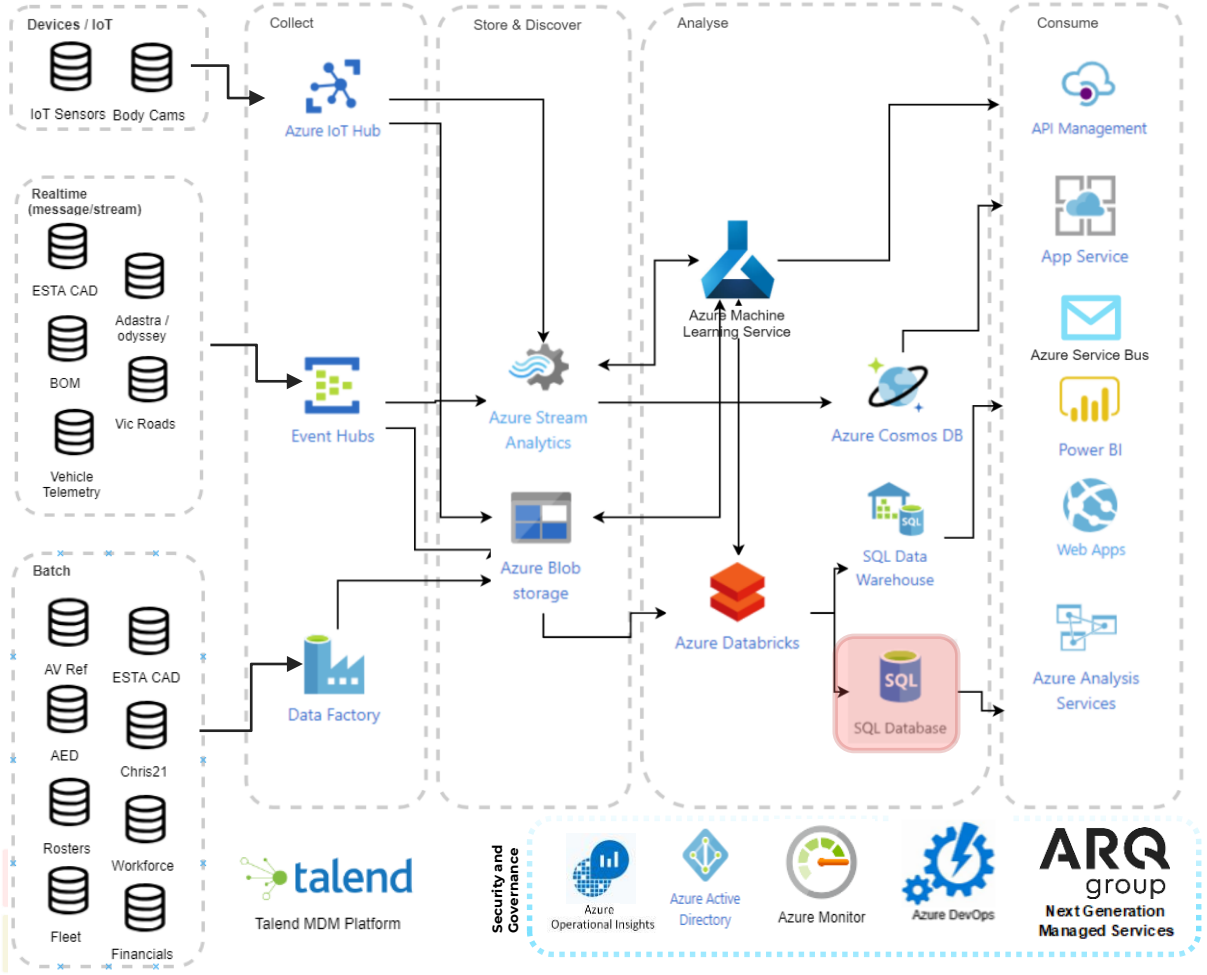
On-demand querying using services such as Stream Analytics and Databricks

Simplifies model management and deployment to Azure

Autoscaling cloud compute and built-in DevOps

Design Considerations

Additional applications can call the service using web APIs



Service

SQL Database

Layer

Analyse

Description

Fully managed cloud relational database service

PaaS service supporting SQL Server, Postgres and Geo-spatial libraries (PostGIS support)

Supports MDM data storage

Used for transactional processing of structured data

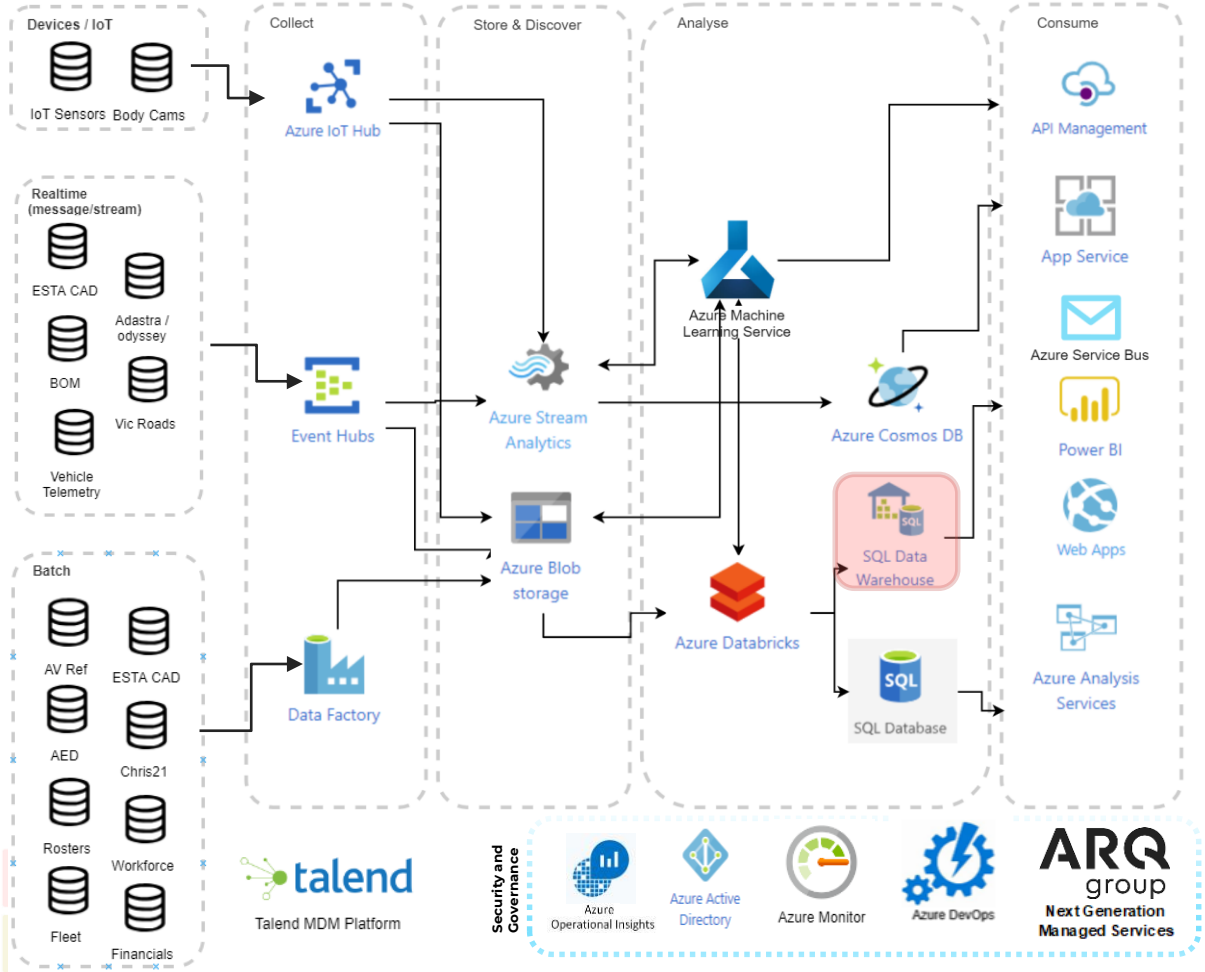
Scalable based on provisioned throughput

High availability options

Design Considerations

SQL Database to be used with SPOG over SQL DW to better support SPOG query patterns of a huge number of queries retrieving a small data sets (not analytical queries)

To supply functions not natively supported by Cosmos DB.



Service

SQL Data Warehouse

Layer

Analyse

Description

Easily scalable cloud data warehouse

Leverages Massively Parallel Processing (MPP) to quickly run complex queries across petabytes of data

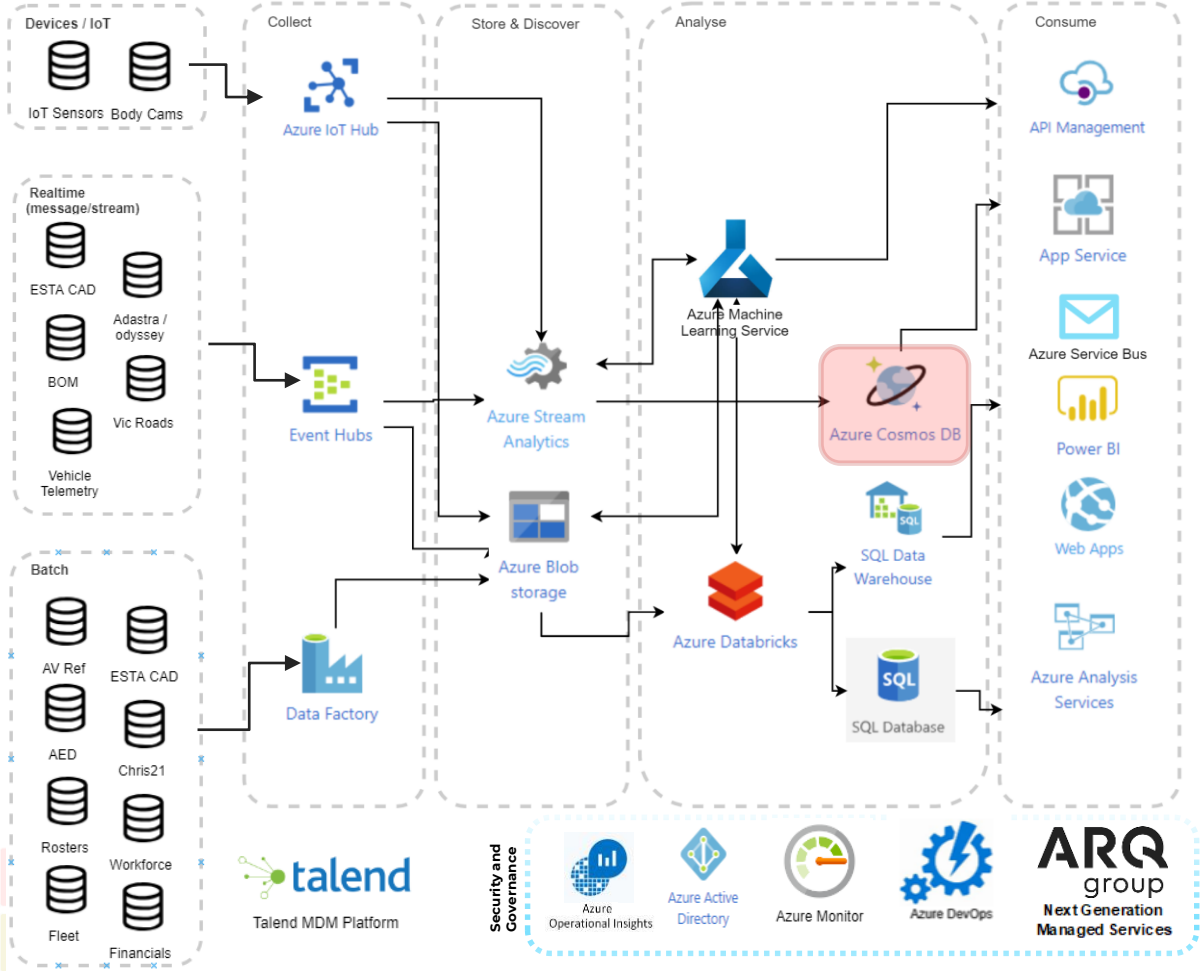
Query and ingest data directly from Azure blob storage

Configurable resource allocations (compute)

Ability to snapshot and restore data warehouses

Design Considerations

SQL Data Warehouse is primarily used in batch oriented processing. Data can be inserted in real-time or periodically but typically processing happens on a periodic basis



Azure Cosmos DB

Service

Azure Cosmos DB

Layer

Analyse

Description

A database for supporting fast, large scale apps with native support for SQL

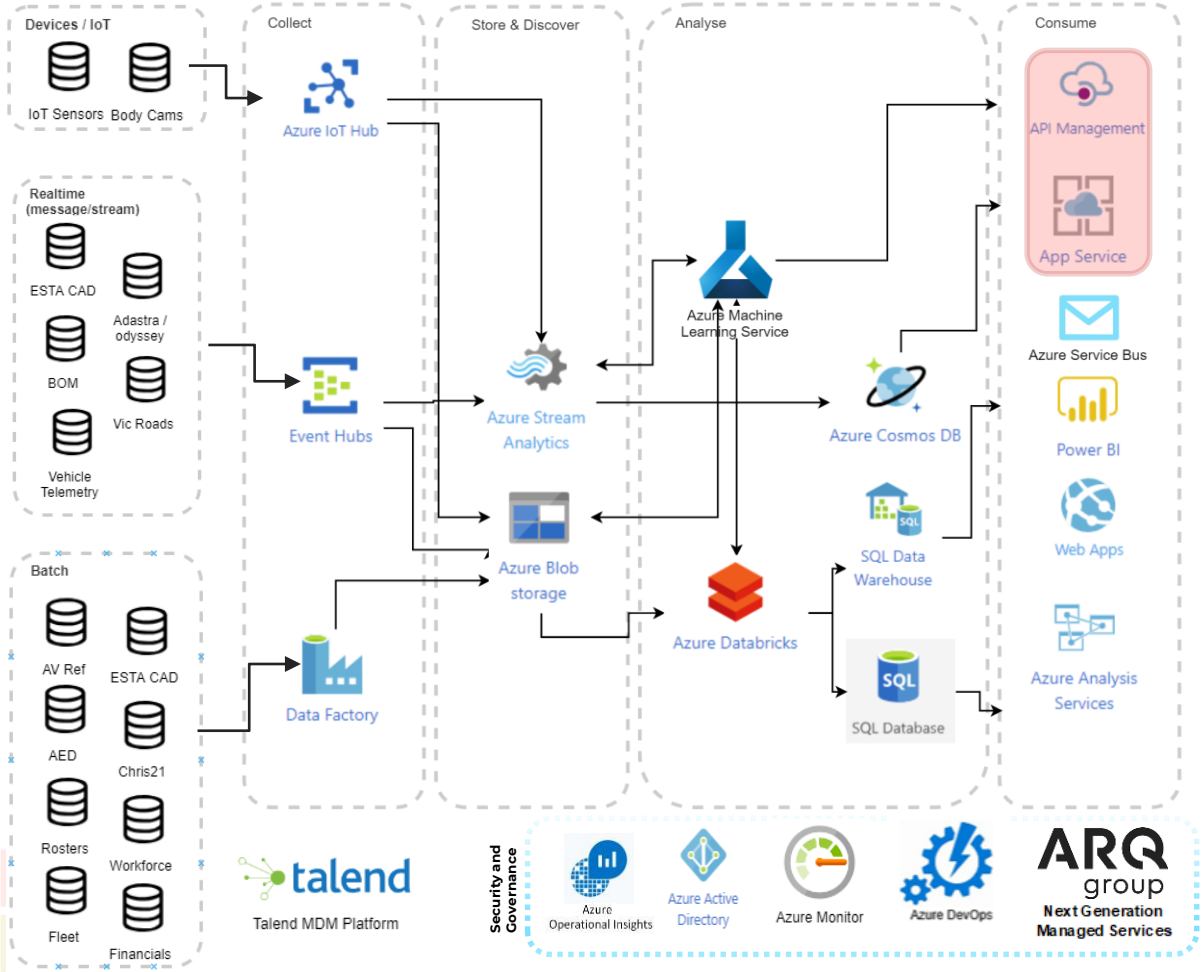
Provides a scalable database for small real-time querying features with high read/write performance

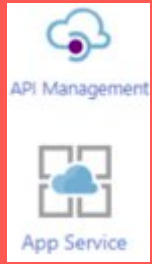
Optimised for specific real-time reporting scenarios

Industry-leading comprehensive SLAs for 99.999% high availability, latency in the 99th percentile, guaranteed throughput and consistency

Design Considerations

Cosmos DB is used by the Single Pane of Glass to support real-time data retrieval





Service

API Management/API Service

Layer

Consume

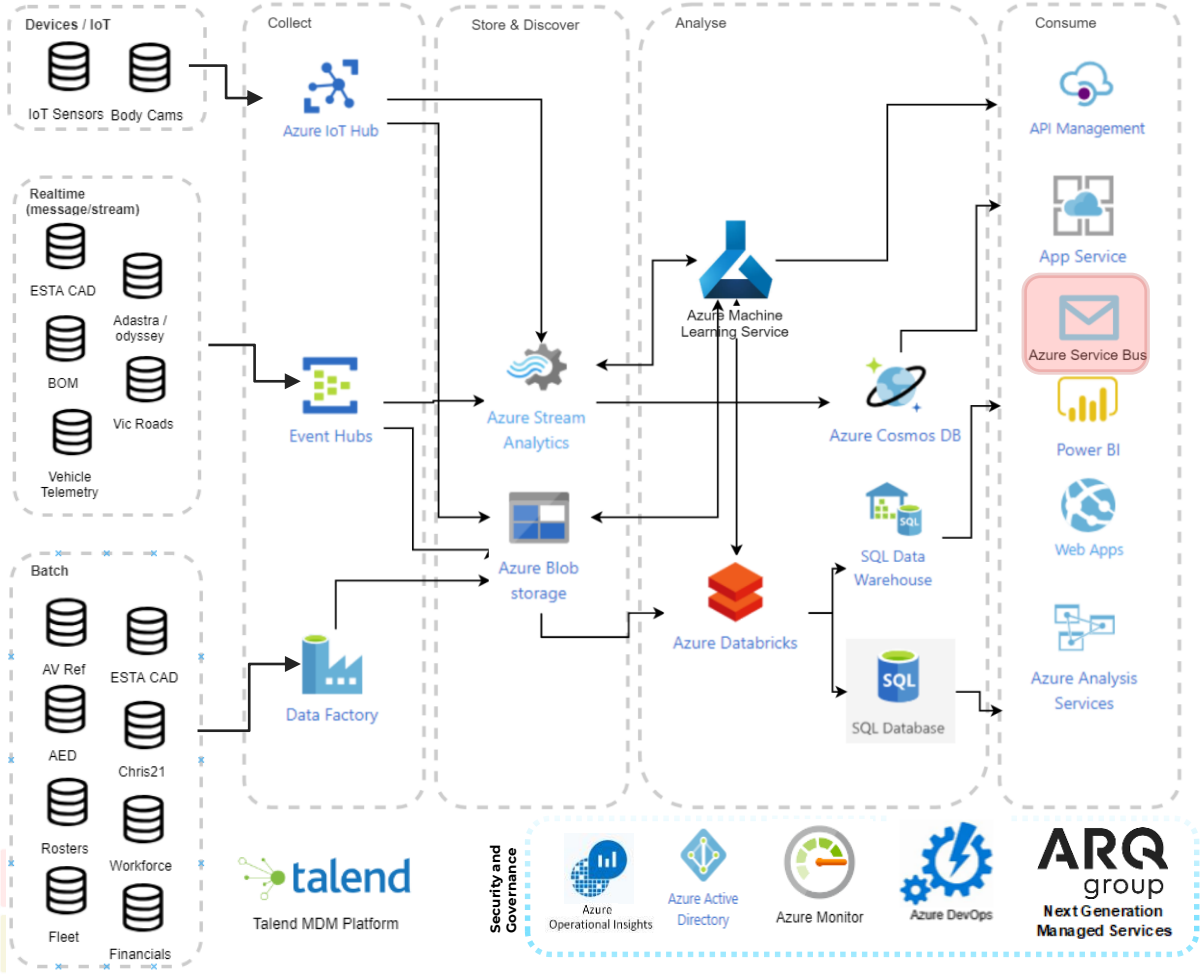
Description

A service for hosting web applications, REST APIs and mobile back ends.

Design Considerations

This represents the API layer where all the APIs published from the Predicative Analytical platform will be exposed.

Provides security, load balancing, autoscaling, and automated management.



Azure Service Bus

Service

Azure Service Bus

Layer

Consume

Description

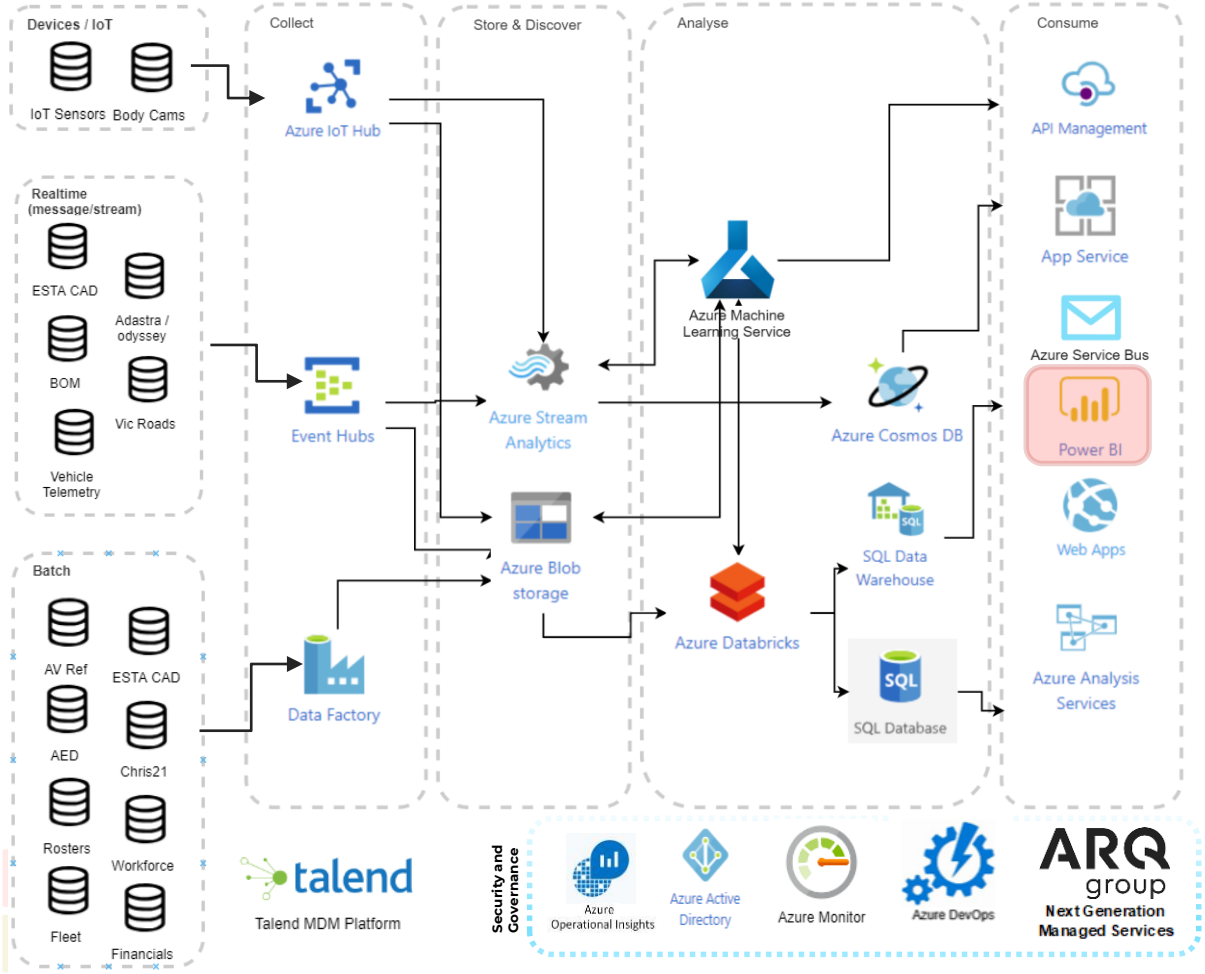
Enterprise Messaging service

A reliable and secure platform for asynchronous data and state transfer

Design Considerations

Used for Command and Control scenarios.

To be used to send notification messages raised from the predictive analytical platform / SPOG to users and other applications



Service

Power BI

Layer

Consume

Description

Data and Analytics reporting tool. Ideal for ad-hoc analytical reporting and interactive dashboarding.

Connectors for all common Azure Data Services

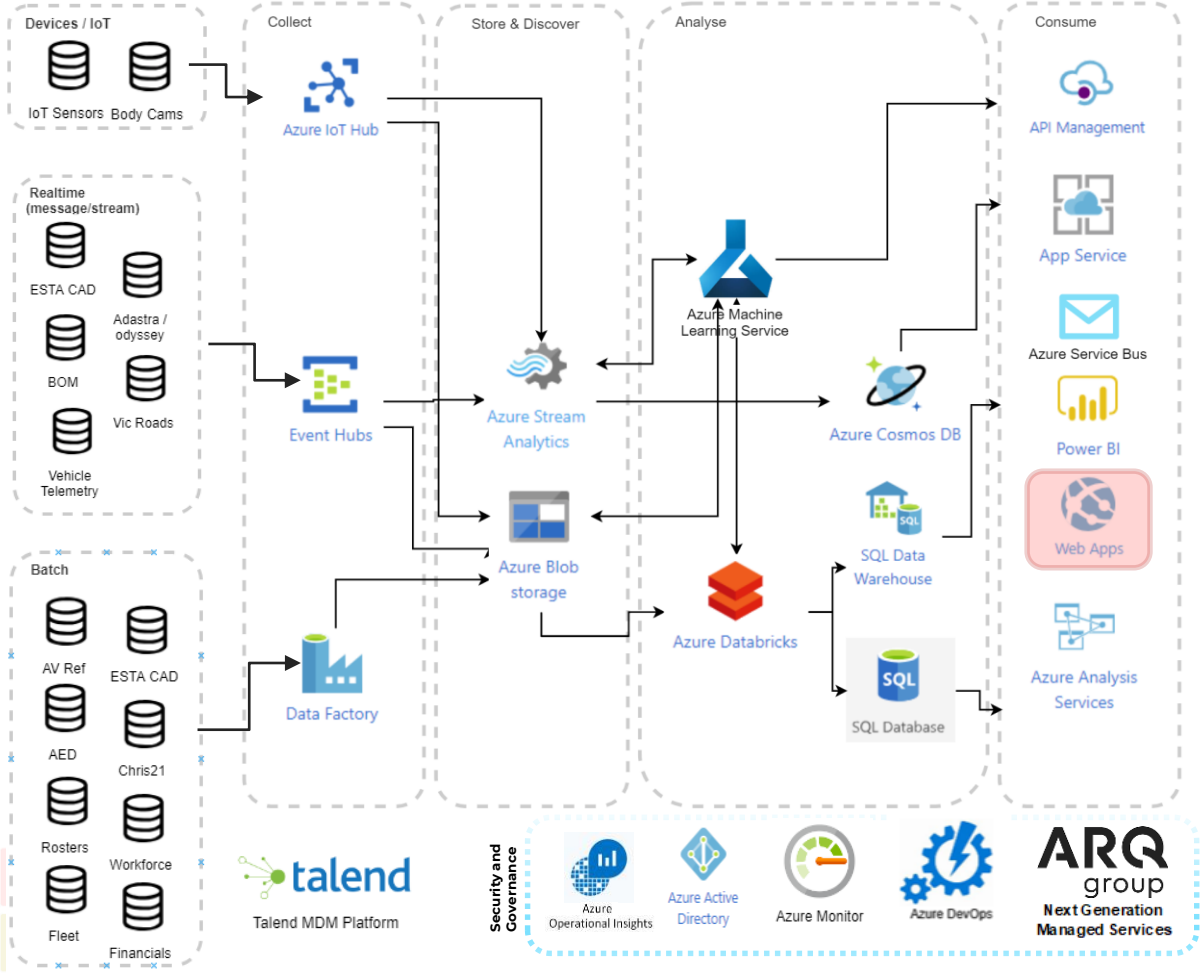
Design Considerations

Power Bi reports will be accessed via the power bi web service. Reports and dashboards can be embedded in to web applications for interactive reports – SPOG where appropriate.

Realtime power bi connection to be used to represent only simple KPI visuals based on a small data sets.

Power BI Premium is expected to meet performance requirements.

Data governance to restrict the use of Power BI Import models in the production environment to ensure control. Preferred approach to refer to an SSAS model or direct query against SQL DWH



Service

Web Apps

Layer

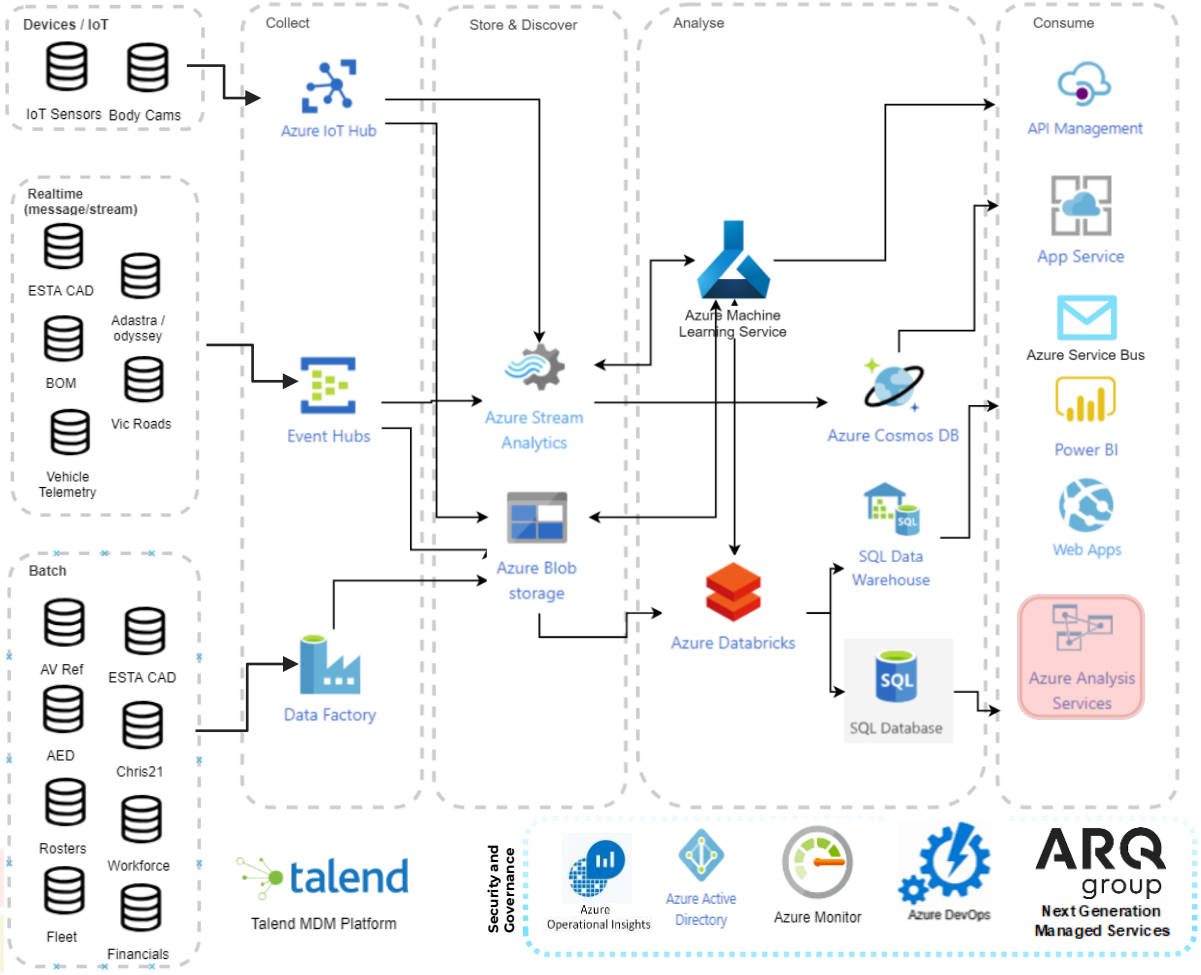
Consume

Description

A service for hosting web apps

Design Considerations

Web Apps is used to host the SPOG application



Service

Azure Analysis Services

Layer

Consume

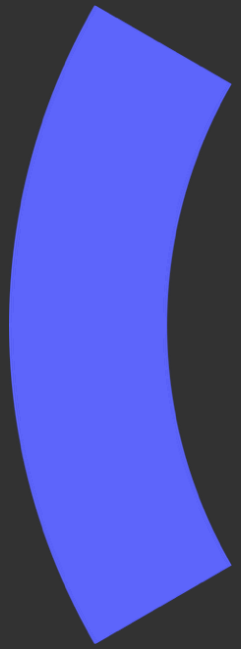
Description

In memory analytical data store that provides a business semantic model over data. Business Measures and KPIs created in models which are used for analytical performance reporting from a Data Warehouse.

Design Considerations

Tabular Model version to be used. Ease of development and performance. Consider OLAP only if data volumes and performance become prohibitive

Azure Analysis Services is primarily used to host the in-memory analytical models for Power BI reporting and ad-hoc data analysis



Non - Functional Requirement Themes

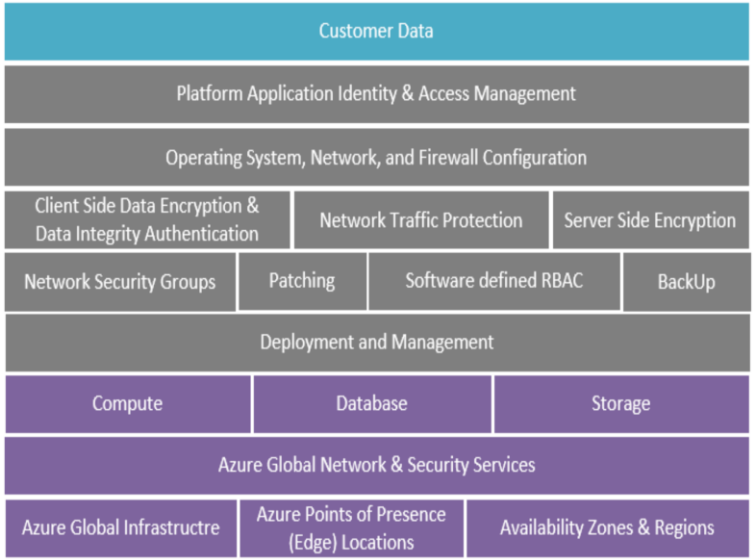
Public Cloud enables security it doesn't guarantee it.

Azure provides a number of tools to extend your security models and protect your data but it requires you to take responsibility for the use and configuration of those tools.

The Victoria Ambulance Service, as the owners of the data, provide classification information and the data itself.

Arq secure the cloud data platform, associated pipelines, and data itself once in the platform.

Microsoft Azure protects the cloud itself and supporting virtual and physical services.



Victorian Ambulance Service Responsible for the Data **in** and **before** it gets to the Cloud Platform.

Ambulance Victoria

Arq Responsible for the security **of** Transit **within** the Cloud Data Platform.

Arq Responsible for the security **of** the Cloud Data Platform.

ARQ group

Gold Microsoft Partner

Azure Expert MSP

ISO/IEC 27001

Microsoft Responsible for the security **of** the Cloud.

Microsoft Azure

ISO 27001

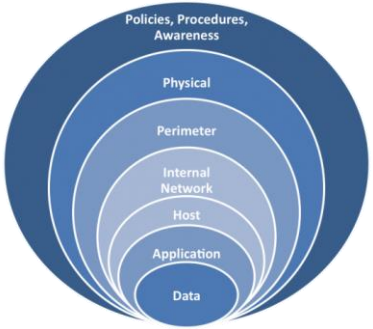
PCI

AICPA SOC

Shared Responsibility

Security is Job 0 for Arq Group.

Our design and automation frameworks are guided by two core principles



Defense in Depth

Least Privilege

We layer security into every aspect of the solution.

From users, through to infrastructure, to systems and applications, to data.

To be able to do this we not only need to consider the classification of the data, but also the required roles and responsibilities when interacting with that data to ensure appropriate controls with fine grained permissions for the defined interactions at every layer of the solution.

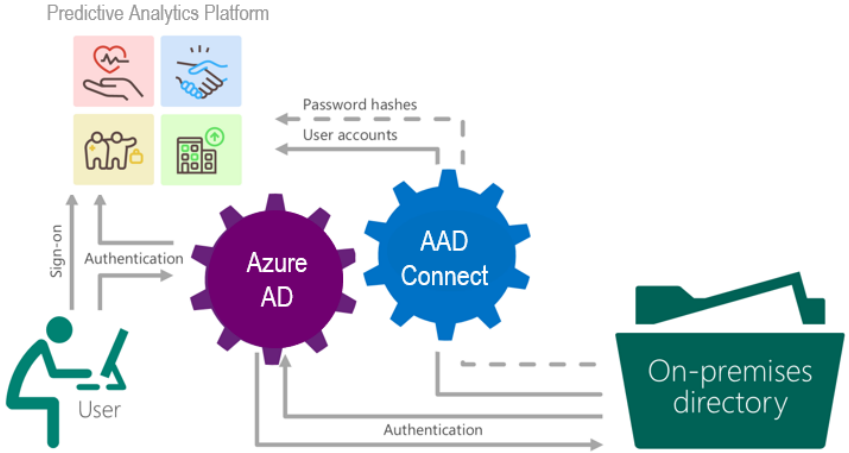
Secure Design Engineering. Controls Based. Continual Compliance.

Data Protection and Security

Identity Management

Azure provides a framework for enforcing role based security privileges by plugging into your existing identity management system.

Existing AD service with existing configured roles are synchronised with Azure AD.



Azure AD provides federated access and single sign on integrated back with your on premise directory.

AAD Connect

- ADFS (for Proxy Password Auth)
- AAD Connect Health
- Support for High Availability
- Multi-Forrest Support
- Supports Writeback
  - Passwords
  - Users
  - Devices
  - Groups
- Support for Custom Attributes

Data Classification

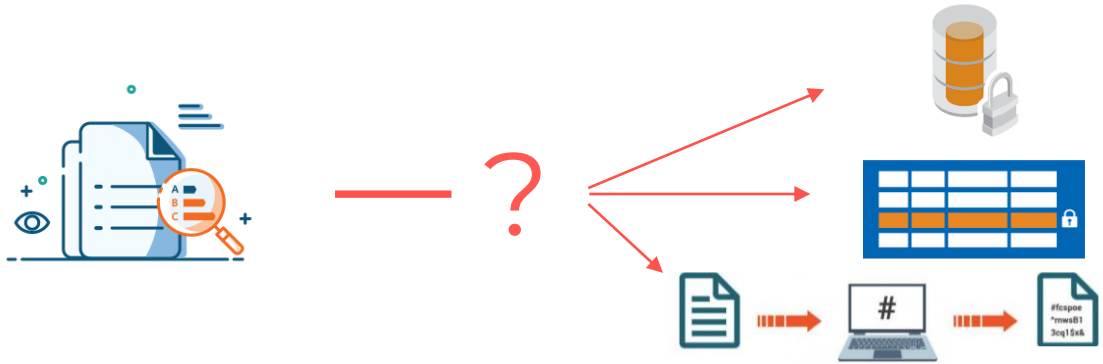
Azure gives you the tools to enforce security over your data, but these controls must reflect the sensitivity of the information and agreed acceptable use.

For each classification of data, controls need to be present for

- Data in use: The data in use to preform a function
(typically stored in non persistent digital states like temp tables or RAM)

- Data at Rest: Data held or stored in a location
(typically blob storage, database tables, or physical media)

- Data in Transit: Data which is transitioning between locations
(typically streaming data, APIs and web interfaces, or backups)

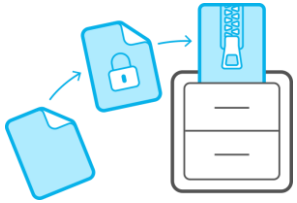


Protecting your data doesn't just mean making sure it's secure from attack.

Data Availability

Data Resiliency

Data Recovery



Microsoft Azure enforces a distance of 480km between regions with high speed private transit. Azure currently offers 3 regions in Australia; No other Major Public Cloud provider currently has this.

This allows the design of highly resilient, highly available services and means your data and your backups can remain secure within an Australian datacenter without the need for a second provider.

Data Protection

Data Protection

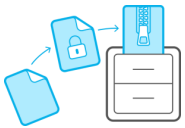
A single region is not a single datacenter.

Within a given region there are multiple
- Availability Zones
- Fault Domains
- Update Domains

As a general rule, all information copied to storage is written at least 3 times within a single region and once to a “backup” region.

A single region, is sufficient to deliver a 99.9% availability SLA and has sufficient durability that you can store your data with confidence in most cases.

When you require high Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO) you need to consider if a single region is viable.



For critical systems which require higher availability or extremely high RTO, warm standby options or options which allow services to be spread over multiple regions are required.

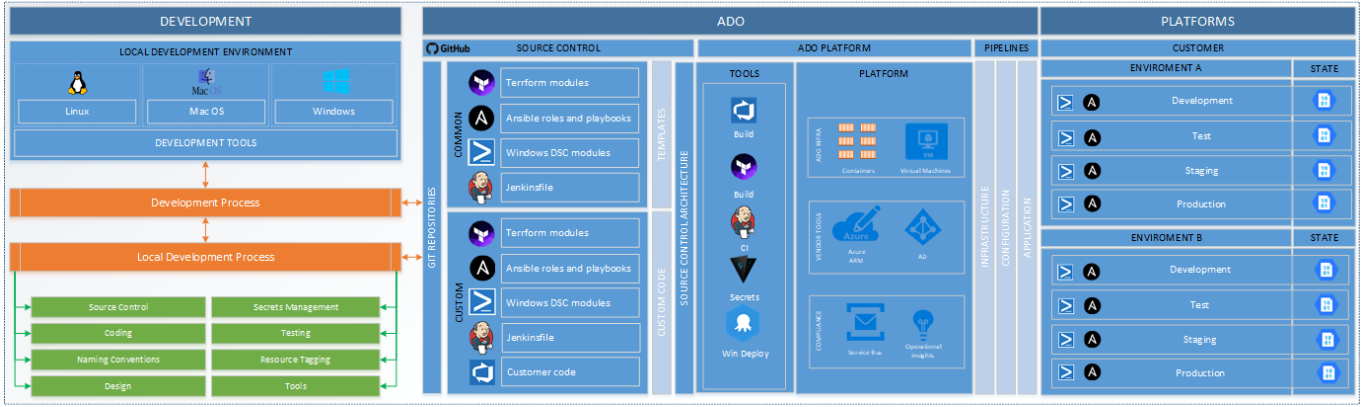
Azure provides PAAS services which offer geo-redundant, cross-region services to be deployed and automated failover mechanism to enable rapid recovery from events.

For less critical data/services, the geographic isolation of the Australian regions allows you to store “cold” backups in more cost effective methods within Australia.

To enable agility without sacrificing operational excellence and required governance you need a framework which allows CI/CD to work at the pace of innovation.

Interchangeable toolsets, governed by the Arq Group Automated Deployment & Operation (ADO) framework.

Allowing ideation to move into and constantly iterate on operational environments securely and with confidence.



Smarter Dev[Sec]Ops.

Intelligent, Self-Healing Systems. Insight, Analysis, and Optimisation.

Data Protection. Role Based Access Controls.

Integrated Identity Management. Visibility and Transparency.

Continuous Integration & Delivery

Thankyou