Roadmap

Achieved to Date (V1 in Development)

- **Architecture & Ingestion**
  - Hub-and-spoke model in Azure, aligned with NHSPS enterprise standards.
  - IoT Hub, API, and MQTT ingestion operational.
- **Medallion Data Lakehouse**
  - **Blob (Cold)**, **Cosmos (Hot)**, **SQL (Warm)**, **ADX (Analytical)** working as designed.
- **Data Aggregation & Cleansing**
  - Azure Data Factory pipelines for aggregation.
  - “Dough Cutter” app operational for canonical data standardisation.
- **Integration & Event Push**
  - Event hub, IoT Hub & Event Grid supported throughout event journey to allow for fan-in fan-out messaging.
- **CI/CD Pipelines**
  - Centralised in Azure DevOps, with automated builds and deployments into dev environment.
- **Platform Monitoring**
  - Application insights and Log Analytics monitoring setup across applications.

Platform Design

Security

- **Implement a Unified Entry Point Strategy:** The architectural design already leverages central points for data flow, mitigating risks associated with scattered access
- The **IoT Hub** acts as the central ingestion point for all device telemetry
- Includes integration with the **shared central APIM (hub)**, ensuring API ingress is controlled through a centralised gateway
- The entire infrastructure is built on an Azure **Hub-and-spoke model**, which aligns with NHSPS enterprise standards and provides a foundational networking structure for security management
- **Secure Configuration Management:** Configuration for all environments (Dev, Test, Staging, Prod) will use unique keys/secrets configured through **App Config per environment**
- **Governance via RBAC:** Ensure strict access control through implementing **RBAC and least privilege across subscriptions**
- Managed Outbound Traffic (**Zero Egress Principle**)
- **Data Redundancy as Resilience: Geo-redundancy for data** (e.g., Cosmos multi-region writes, SQL geo-replication, blob GRS)

Scalability

- **High-Volume Ingestion:** The architecture uses **IoT Hub** as the central ingestion point for device telemetry, which is designed to handle massive scale.
- **Flexible Event Routing:** Scalable message processing is ensured by using **Event Hub**, **IoT Hub**, and **Event Grid** throughout the event journey, allowing for flexible **fan-in fan-out messaging** to decouple services and manage varying loads.
- **Data Streaming:** Raw telemetry data and alert-driven events are streamed through dedicated **Event Hubs**, enabling high throughput and the ability for multiple consumer groups (like the Telemetry consumer group or DoughCutter consumer group) to process data in parallel

- **Massive Analytics Engine: Azure Data Explorer (ADX)** is employed specifically for **large-scale, near real-time analytics**. ADX is responsible for performing the heavy-lift aggregations, ensuring that analytical demands do not strain the primary data stores
- **Cost Efficiency Strategy:** The overall data flow balances real-time needs (Cosmos + alerting) with **long-term storage optimisation** (Blob + ADF), ensuring scalability is achieved cost-efficiently
- **Azure Function Scaling**: The nature of azure functions in azure lends to scaling by SKU. Most functions live on a consumption SKU outside of production keeping the costs low whilst allowing for ramp-up if needed. Production SKUs will be premium to remove any problems around cold-starts.

Storage & Analytics

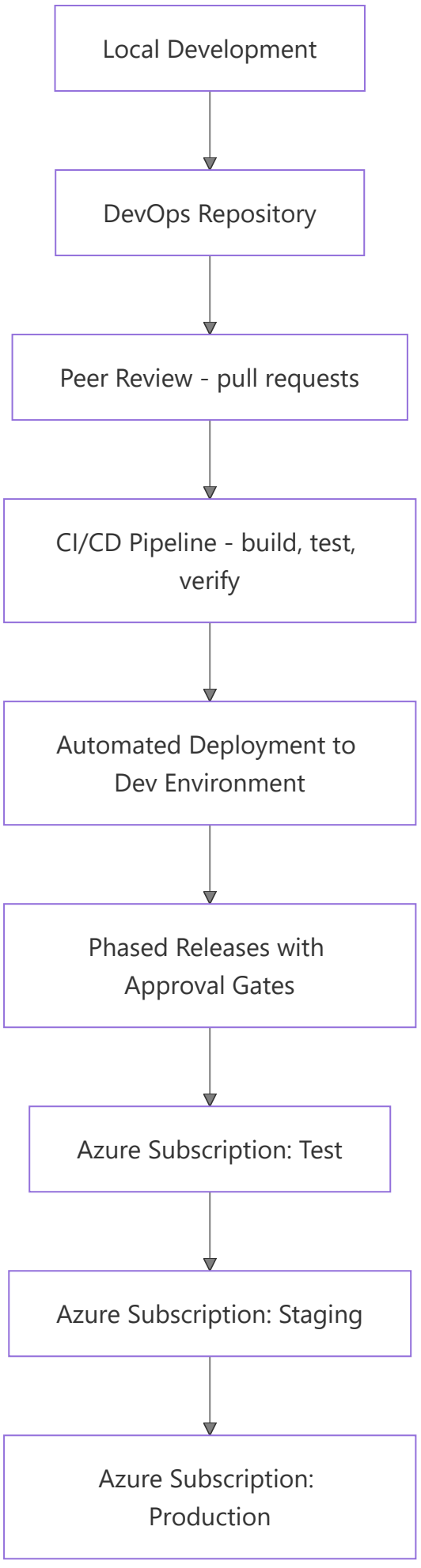
The architecture successfully implements a **Medallion Data Lakehouse** approach, dividing data across three optimised tiers:

- **Hot Storage (Cosmos DB):** This tier is used for **fast-access, real-time telemetry data**. It supports instant querying and device interactions.
- **Warm Storage (SQL Database):** This layer stores **periodically aggregated data** and is explicitly defined to *not* store raw, high-frequency data. It is optimised because business queries executed here are cheaper and faster than querying raw data.
- **Cold Storage (Blob Storage):** This tier is used for **cost-efficient long-term archival**. It stores raw historical data in cheaper storage.
- **Cost Optimization: Azure Data Factory (ADF)** periodically batches daily telemetry into single blob files to **minimise Azure storage transaction costs**
- **Core Analytical Engine (ADX): Azure Data Explorer (ADX)** is utilised for **large-scale, near real-time analytics**. ADX performs the essential data aggregation and querying capabilities and handles the "heavy-lift aggregations"
- **Reporting Workflow (ADF and SQL): Data Factory (V2)** plays a critical role by moving aggregated telemetry results *from* ADX **into Warm SQL storage periodically**. This strategic movement ensures that subsequent business reporting queries are **cheaper and faster**
- **Visualization: Power BI** connects across the storage tiers (ADX, SQL, and Blob) to **visualise telemetry insights** and provide business dashboards for decision-making
- **Future Scope:** connecting to **Microsoft Fabric** will hopefully allow for benefits such as **machine learning** as well as easier to share reporting and analytics across functions
- **Future Scope:** setting up a process with **Integration Layer** to ensure **system-to-system** communication is achieved effectively, efficiently and is contractually managed to ensure a canonical data stream.

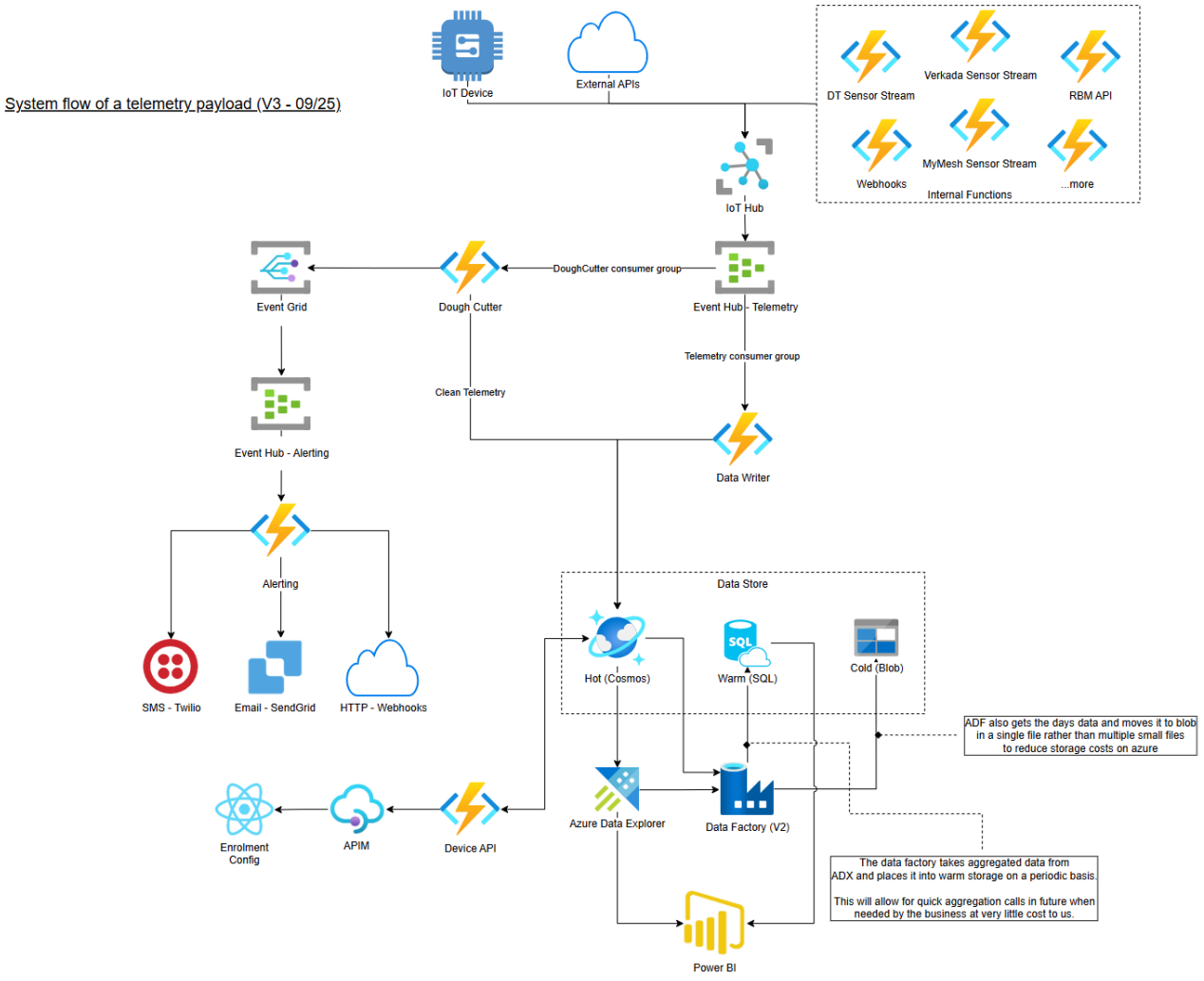
Development Cycle

- **Agile Methodology:** The development process follows **agile development** principles.
- **Local Development:** Functions and web applications are initially **developed locally**.
- **Version Control:** Developed code is then **pushed to the repository (repo)**.
- **Peer Review:** Code changes are mandated to undergo **peer review using pull requests** to maintain quality and collaboration
- **Automated Verification:** The centralised CI/CD pipeline handles the essential steps of **build, test, and verify** for all functions
- **Initial Automated Deployment (Dev):** The system currently has operational CI/CD pipelines with **automated builds and deployments into the dev environment**
- **Phased Releases:** Once the development team is satisfied with the deployment in the dev environment, **release pipelines** are utilised to move the releases forward
- **Approval Gates:** The releases are strictly controlled through **approval gates** before they can be pushed into the subsequent environments

- **Target Subscriptions:** The gated release process pushes the verified applications into the subsequent **subscriptions in Azure**, specifically targeting the **test, staging, and production** environments



Platform Design Diagram



Next Steps – Roadmap to BAU

Phase 1 – Infrastructure as Code (Q3 FY)

- **Terraform Modules**
  - Develop reusable modules for IoT Hub, Cosmos, SQL, Blob, ADX, ADF, networking.
  - Ensure modules support **parameterised deployments per environment** (subscription IDs, resource groups, SKUs).
- **Disaster Recovery (DR) Strategy**
  - Multi-region design where required.
  - Geo-redundancy for data (Cosmos multi-region writes, SQL geo-replication, blob GRS).
  - Document DR invocation procedures.
- **CI/CD Extension**
  - Update pipelines to trigger Terraform against the correct subscription/spoke.
  - Implement approval gates between dev → test → staging → prod.

Phase 2 – Environment Build-Out (Q3–Q4 FY)

- Re-provision **Test, Staging, and Production subscriptions** using Terraform.
- Deploy:
  - Ingestion functions and pathways (IoT Hub, API, MQTT, Function Apps).
  - Medallion architecture components (Cosmos, SQL, Blob).
  - Analytics (ADX, ADF).
  - Dough Cutter app.
- Integrate with **shared central APIM (hub)**.
- Configure **App Config per environment** with unique keys/secrets.

Phase 3 – Validation & Governance (Q4 FY)

- **Environment Validation**:
  - End-to-end data flows across all ingress methods.
  - Failover/DR testing.
  - Correct SKUs for production level data flow are met.
- **Operational Readiness**:
  - Runbooks (deployments, DR, incident response).
  - Monitoring dashboards (App Insights, Log Analytics, ADX queries).
- **Governance**:
  - RBAC and least privilege across subscriptions.
  - Naming conventions, tagging, and cost governance.

Phase 4 – BAU Transition (End of FY)

- Final deployment into **Production subscription**.
- Handover to BAU team with:
  - Operational playbooks.
  - Agreed SLAs.
  - Cost management ownership.
- Programme closure with formal sign-off.