7.2.1.2 Data Fabric Strategy

Overview

L1 - Data and Analytics Components

Nike's Data Fabric is our critical unlock to the business looking to accessing Nike's data. One of the key functionalities of Data Fabric is Unified Access. **Unified Access** a virtual layer for seamless access to data for consuming systems. This virtual layer helps abstract the need for specific access technologies by providing more standardized protocols. For the user, this allows them to work with different data sources as if they were the same.

Problem Statement

The Nike Data Foundation (NDF) has a core concept of the Data Lake. The data lake over time has the Global Data & Analytics Platform evolved to be more than data and has evolved into Data Fabric. Data Fabric represents the following logical functions: **Unified Data Access, Registry & Discovery, Persistence, Security & Access Contro**l.

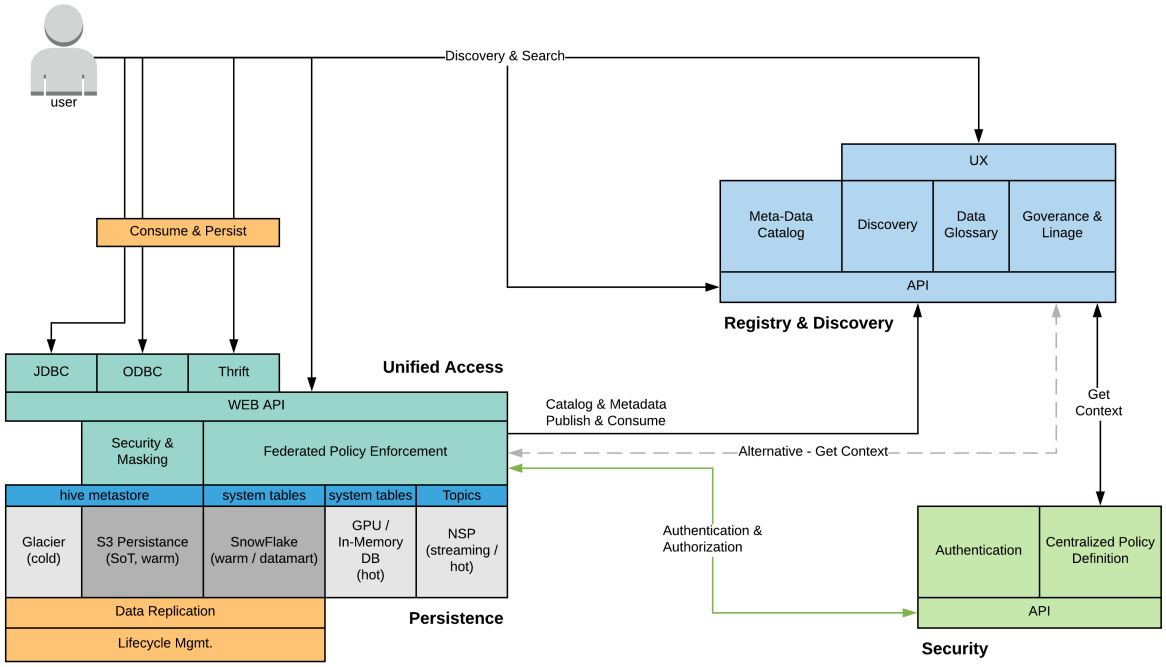
Notable Advantages

- Single point for data consumption, registry and discovery.
- Operational tasks such as data replication and lifecycle management are handled in the fabric layer and no longer a concern for the analytics community.
- Consistent, fine-grained and integrated access control and security.
- Scalability and Risk Reduction.

Business Objectives, Goals, and Strategies

DATA FABRIC COMPONENT VIEW

Updated | October 13, 2020



Unified Access

The unified access layer is designed to be the single point of access and enforcement to data stored in the data fabric. The technology posistioned to fullfill this layer is RANGER.

Persistence

The persistance layer is the varioius forms of data storage for the data fabirc. There are differing persistance layers based on usage requirements.

NOTE: The source of truth for all data is S3 and therefore all data must be persisted at this layer. This does not imply all data must be consumed from this layer.

Registry & Discovery

The layer which users publish and discover data. Apart from knowing where data is, this layer also services as the outward communication of all data attribution and classification.

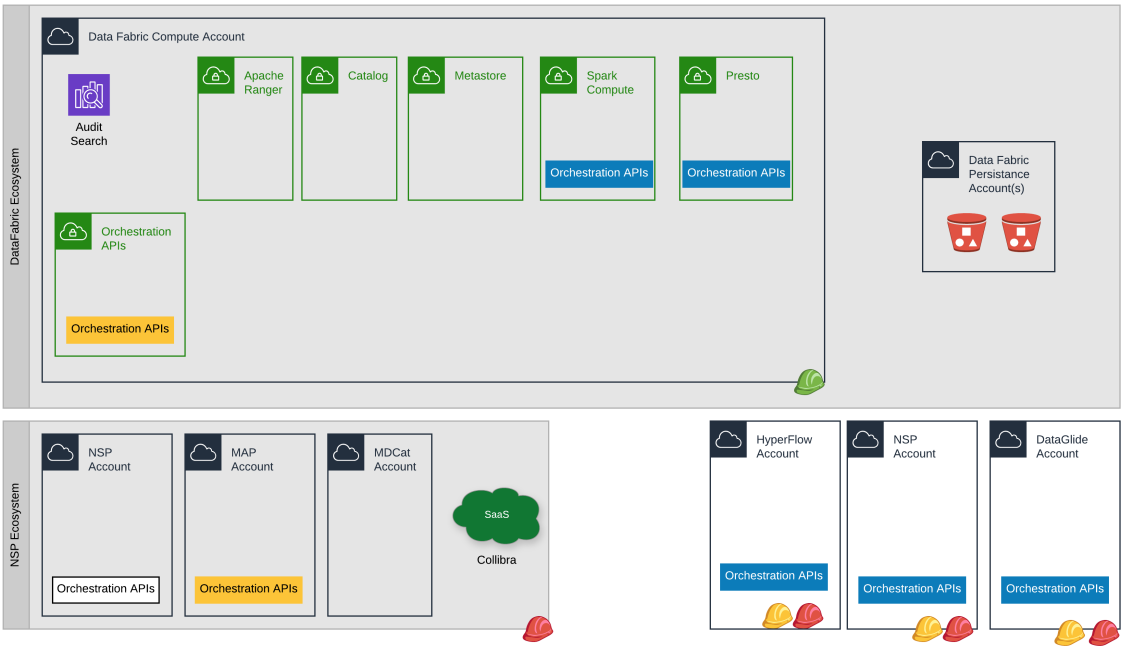
Security

Consistent layer for we manage all entitlments across assets at Nike. The need for consistency here is that wether we're daling with a file or data the both carry similar data sensitivty requirements. Defining these in a singular location provides a consistent user experience across all Nike environments.

NOTE: the point for context injection is still to be defined based on entitlments design discussions.

ACCOUNT TOPOLOGY

Updated | October 13, 2020



Only DataFabric Compute will have access DataFabric persistance.

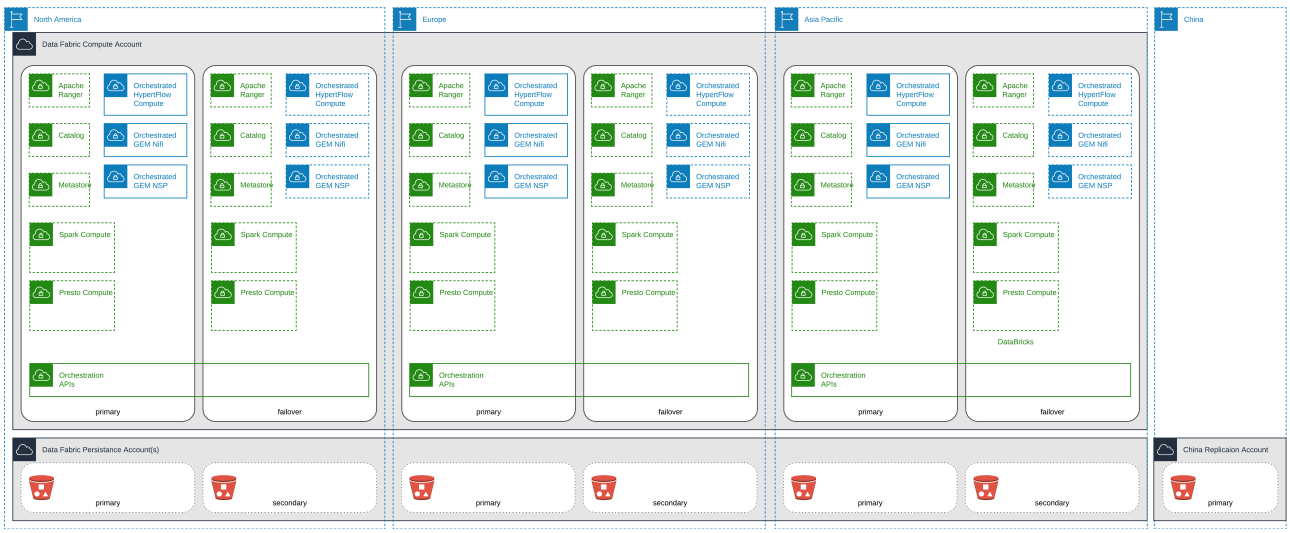
Compute layers for GEM, that require access to S3 and/or require direct kafka broker access should live in the DataFabric Compute Account.

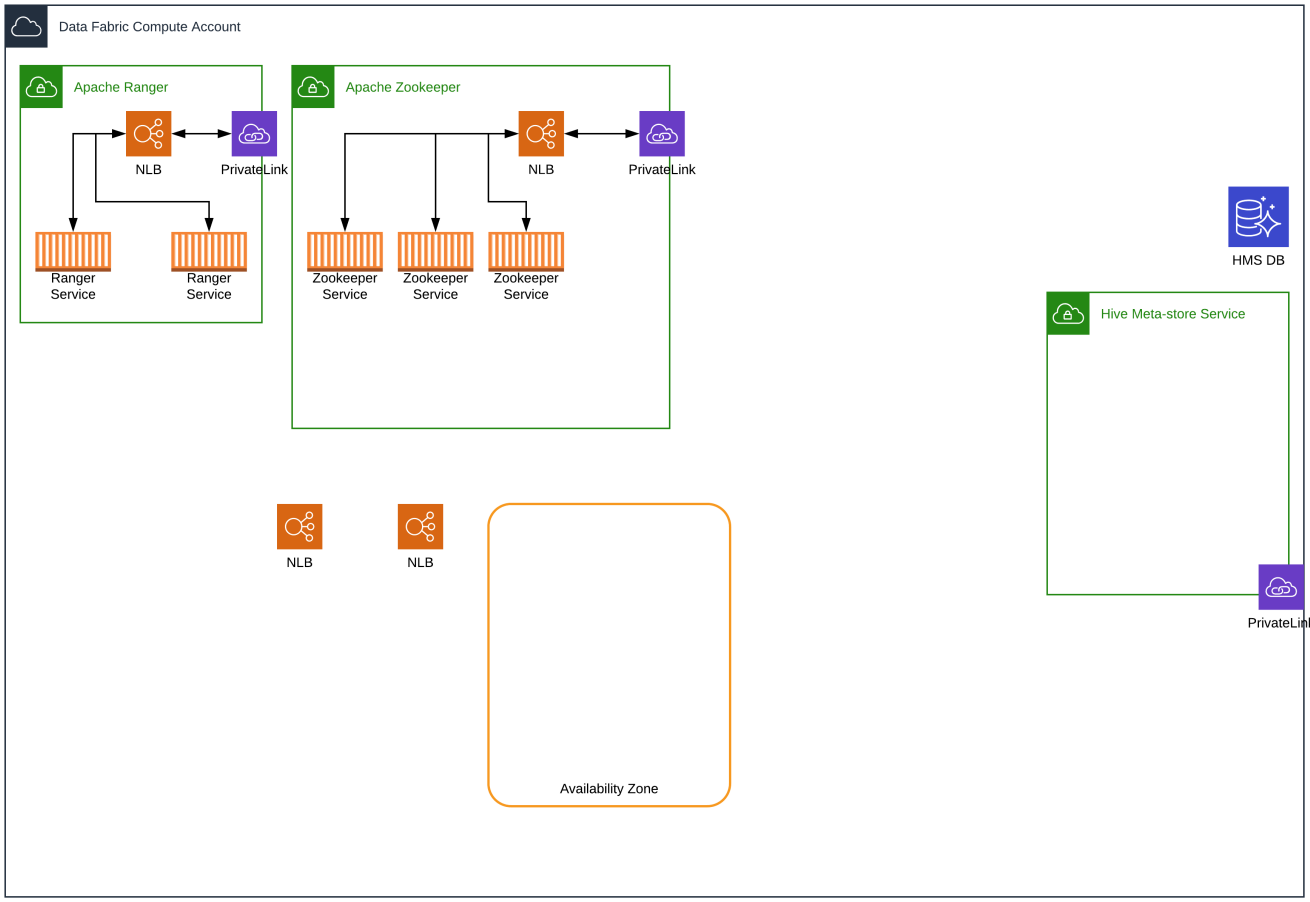
Data Fabric Permissions Key

- Orchestration Permissions
- No Data Fabric S3 Access
- Data Fabric S3 Access

GLOBAL FOOTPRINT

Updated | October 13, 2020





Business Capabilities Impact

Today, there is a significant majority that has an immediate need for this collection of capabilities.

As Nike becomes more mature in becoming a data-driven company nearly every business capability will require a place to ingest, modify and publish their data and associated data products.

People and Culture

TBD

Ecosystem

TBD

Technology Component Impact

- Persistence
- Ingestion
- Consumption
    - Ad-Hoc
    - System Level
- Access Control
- Security
- Registry
- Discovery
- Catalog
    - Meta-Data
    - Logical
    - Registry & CRUD
    - Publication
- Data Governance & Stewardship

- Glossary
- Data Quality & Heuristics
    - Accuracy and Precision
    - Legitimacy and Validity
    - Reliability and Consistency
    - Timeliness and Relevance
    - Completeness and Comprehensiveness
    - Availability and Accessibility
    - Granularity and Uniqueness
- Movement
- Lineage
- Replication
- Lifecycle Management

Strategic Principles

TBD

Strategic Choices

Catalog

Data Masking

Data Modeling

Discovery

Ingestion

Lifecycle Management

Persistence

Registry & Publication

Replication

Data Row Filtering

Unified Access Integration and use cases

Data Fabric Lake Persistence

Supporting Reference Architectures