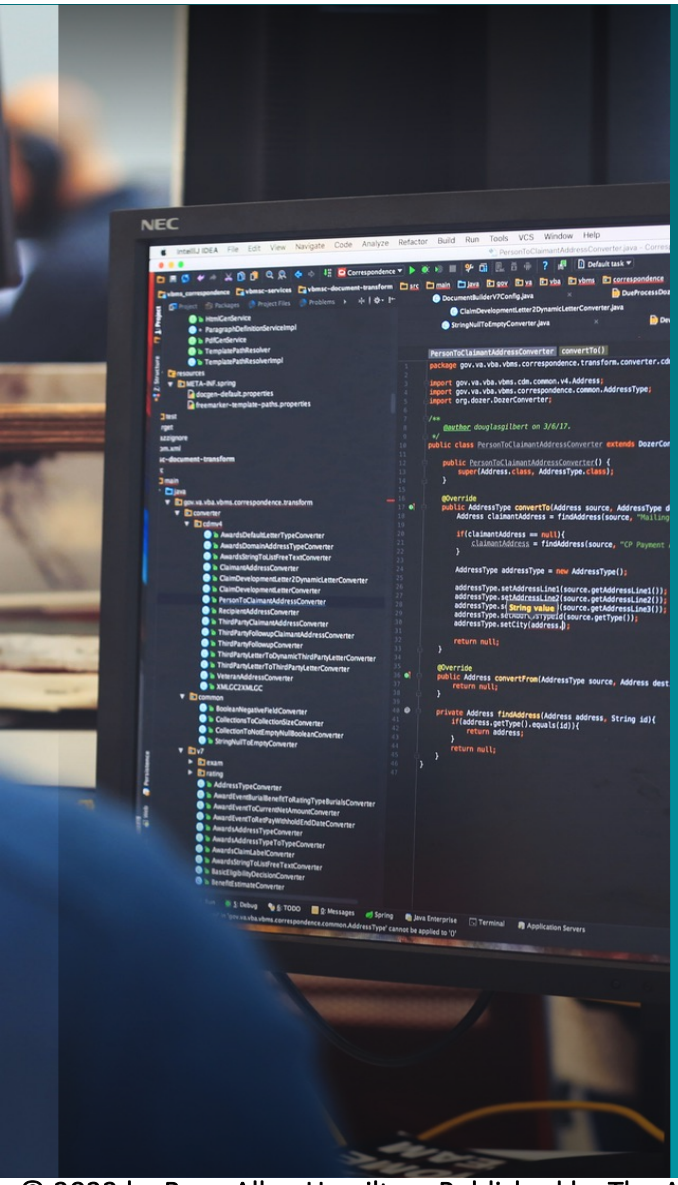
Booz | Allen | Hamilton®



FREEING AND SECURING DATA THROUGH DATA CENTRICITY

Reducing barriers to data while ensuring data integrity and protection

MARCH 2022

CONSULTING | ANALYTICS | DIGITAL SOLUTIONS | ENGINEERING | CYBER

WHAT IS DATA CENTRICITY

"An architecture where data is the primary and permanent asset, and applications come and go" – Dave McComb - The Data Administration Newsletter – The Data-Centric Revolution: Data-Centric vs Data Driven

- **Focused on coming with a data-first approach. How and where the data comes from might change but the data should persistent and be available**
- Benefits:
  - Removing the dependency on developing individual, one-off point-to-point connections
  - Ingest new data events without needing to redeploy for speed-to-mission delivery
  - Reduce barriers to the data – i.e. easy data access and data discovery
  - Data Stewards & Data Owners have more granular control over access to their data

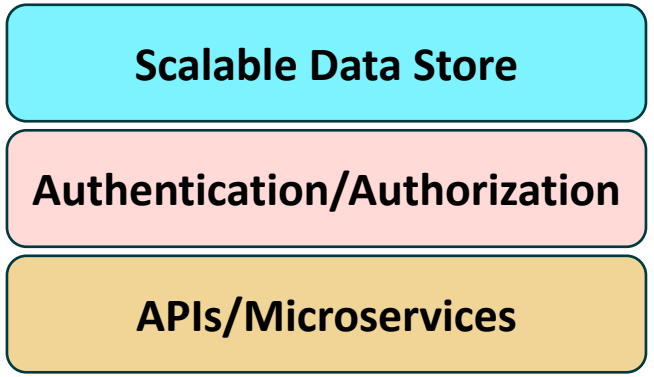
CHALLENGES

- There are several challenges that "Freeing the Data" faces. Most challenges fall under the category of security.

Having Robust, Secure & Scalable Authentication
Scalable Authorization For Varying Types of Data
Reliable Architecture That Can Support Mission Critical Operations
Integrating Legacy Systems

BUILDING A SECURE DATA ECOSYSTEM

- **Three Core Components**
  - ***Scalable Data Store***
    - Cloud Native/Capable
    - Secure Data Storage
  - ***Authentication/Authorization***
    - Centralized Authentication
    - Put power in the hands of data stewards
  - ***Open APIs & Microservices***
    - Discovery of data; data availability
    - Allows growth with minimal impact to development on the underlying layers
    - Integrate legacy systems

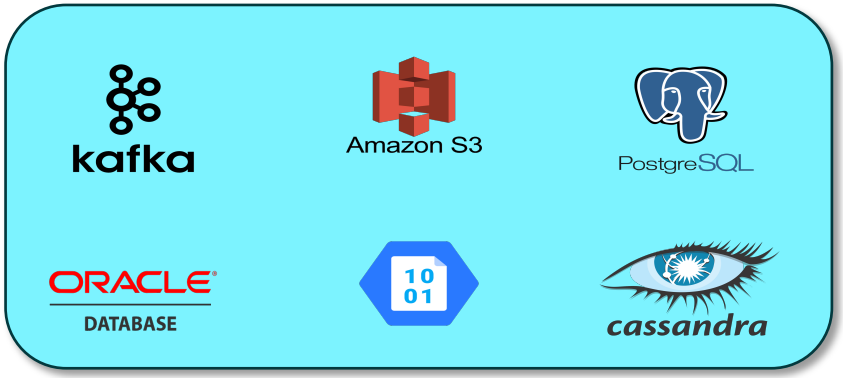


DATA CENTRICITY EXAMPLE

- ***Challenge:***
  - Build a Data Event Data Broker (DEDB) to provide near real time streaming of Key Data Event Messages from various Data Producers to Data Consumers
  - ***Requirements:***
    - Small Event Messages (1 MB or less)
    - Data Producers & Consumers from various Organizations, Policies & Data Security Levels
    - Automatically Onboard New Data Events & Data Producers
    - Provide ability for Data Producers or Data Stewards to Modify ACL access to Data Event Events
    - Data Producers aren't always the Data Stewards

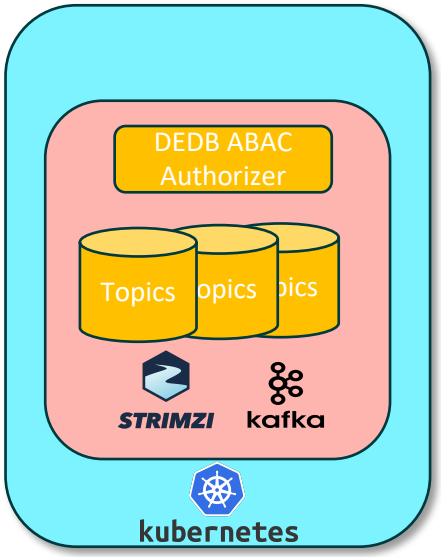
SCALABLE DATA STORE

- Start with 1 or more Data Store.
  - Message Queue
  - Relational or non-Relational Database
  - Key-Value Store
- The key is to pick the right data store(s) based on the data needs.
  - Scalable
  - Highly available and reliable
  - Able to meet performance needs
  - Support for different authorization needs (ABAC, RBAC, etc)
  - Encryption at Rest



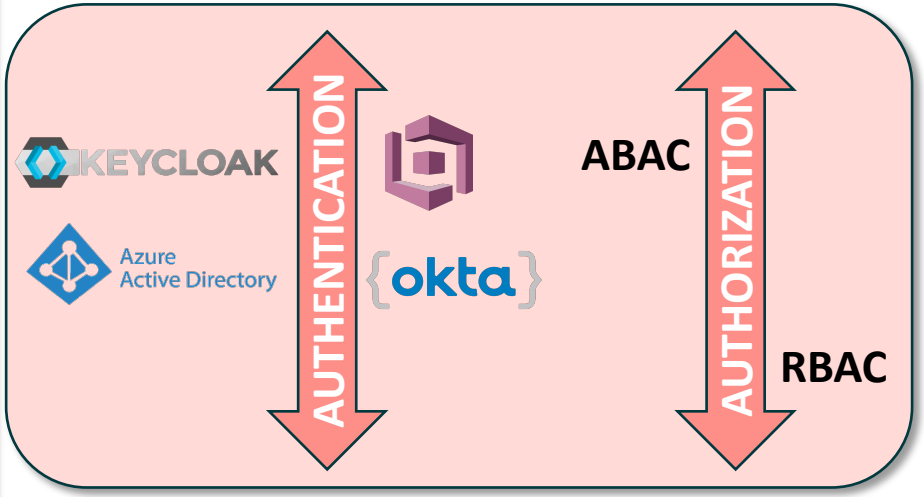
DEDB SCALABLE DATA STORE

- DEDB uses Kafka for its Data Store
  - Data Event Messages are Stored in Topics & Partitions
  - Can Configure Each Topic Specifically for the Data Needs
  - Support for OAuth2
  - Basic ACLs to Control Access to Topics
  - Extensible: Created an ABAC ACL Plugin
- Strimzi Kafka is Kafka on Kubernetes and Provides some Benefits
  - On-the-Fly Storage Increases
  - On-Demand Adding Brokers
  - Resiliency
- Deployed in the Cloud
  - Autoscaling Groups for HA
  - Encrypted EBS volumes



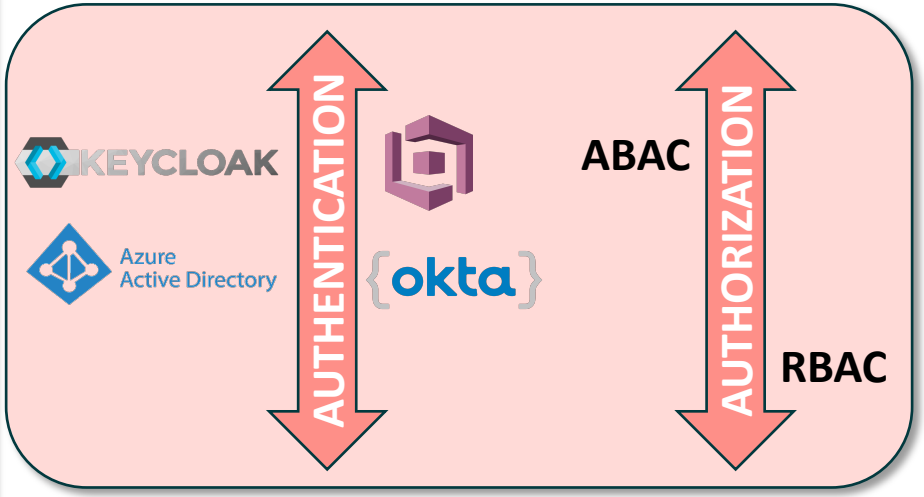
AUTHENTICATION & AUTHORIZATION

- Authentication and Authorization are separate but connected pieces of the Security
- OAuth2 Authorization Framework with OpenID Connect
  - Centralized Auth Server
  - Authenticate users and systems alike with their PKI certificates
  - Issue & use secure tokens
  - Resource server can validate access/ID tokens as well as request user info (attributes)
- Getting away from point-to-point authorization or whitelists.



DEDB AUTHENTICATION & AUTHORIZATION

- Kafka Supports several Authentication Methods
  - Oauth2 libraries
- Authentication Servers
  - Keycloak
  - Broadcom API Gateway
- External Identity and Authorization Providers associate Users & NPE certificates with entitlements
- Once Users are provided an Access Token, DEDB:
  - Validates the Token
  - Retrieves User's Entitlements from a User Info Endpoint
  - Passes the Entitlements to the DEDB ABAC Authorizer to check authorization



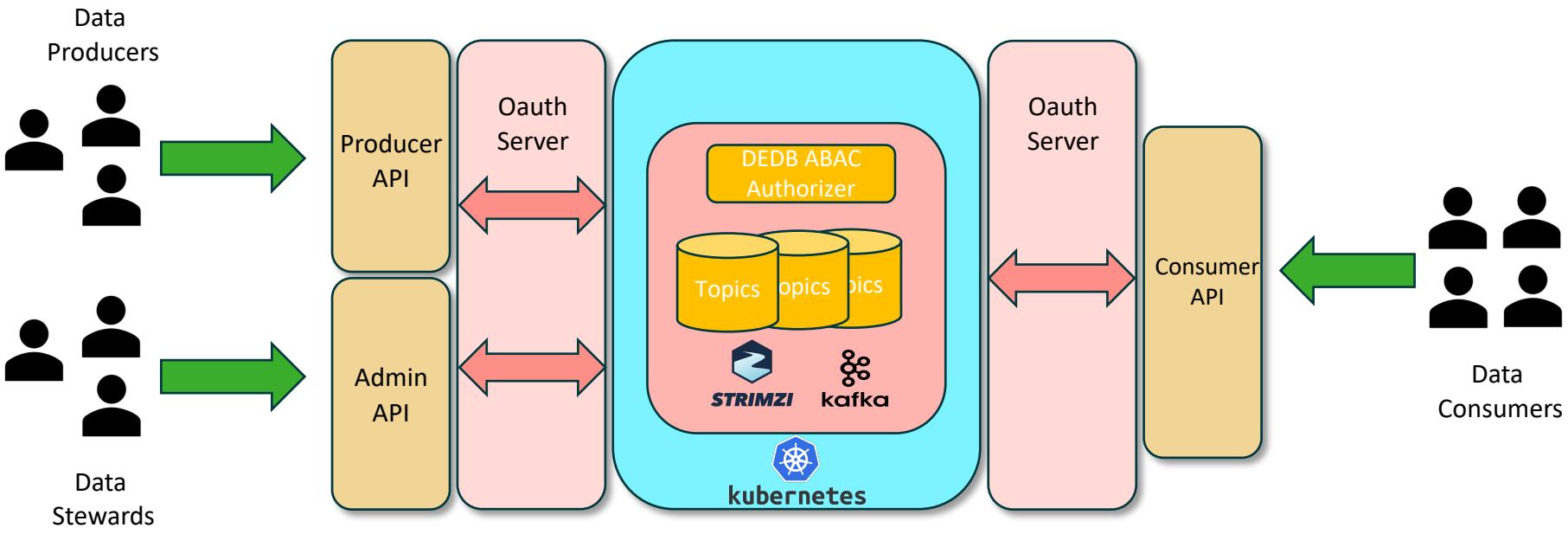
OPEN APIS & MICROSERVICES

- ❑ Easy to document and write to
- ❑ Allows for future growth
- ❑ Can add additional services without requiring end-user recode
  - ➢ Dataset Catalog
  - ➢ Provide schemas and ontology
  - ➢ Add in log aggregation or metrics
  - ➢ Ingest/egress
  - ➢ Administrative functions

DEDB APIS

- DEDB Producer and Consumer APIs are the basic Kafka APIs used for publishing and consuming from Kafka
- DEDB Admin API is used to provide limited access for Data Stewards to the Broker
  - Create Topics
  - Modify Topics
  - Modify ACLs
- Use OAuth to authentication & authorize users as valid Data Stewards
- Consists of a Frontend & Backend API
  - Frontend is an OpenAPI spec to allow create/modify
  - Backend is a wrapper around the native Kafka Admin API

DATA EVENT DATA BROKER



FINAL THOUGHTS

- Use the Cloud and Containers
  - Services from Cloud Providers can help ease scalability and reliability issues
  - Container Orchestration systems can make deployments and handling failover more robust
- Gather Metrics and Logging
  - Metrics give insight into the Frontend & Backend API performance as well as the Data Store
  - Logging is essential to provide activity and access logs for monitoring
- Integrate with other data stores to expand capabilities
  - Nifi can be used for translation & wrapping & data manipulation
  - S3/Azure Data Lake/Blob can be used for long term storage of larger data messages
  - Tools like Elasticsearch for Logging & Metrics