

Study Definition Repository (SDR) Reference Implementation

High Level Design (Infrastructure)

SDR Release V2.0, March 2023

Disclaimer

These materials and information, as well as the underlying code/application they relate to, are provided by TransCelerate Biopharma Inc. AS IS. Any party using or relying on this information, these materials, and/or the underlying code/application do so entirely at their own risk. Neither TransCelerate nor its members will bear any responsibility or liability for any harm, including indirect or consequential harm, that a user may incur from use or misuse of this information, materials, or underlying code/application.

TransCelerate does not endorse any particular software, system, or service. And the use of specific brands of products or services by TransCelerate and its collaboration partners in developing the SDR Reference Implementation should not be viewed as any endorsement of such products or services. To the extent that the SDR Reference Implementation incorporates or relies on any specific branded products or services, this resulted out of the practical necessities associated with making a reference implementation available to demonstrate the SDR's capabilities.



Table of Contents

- 1. Overview**
- 2. SDR RI Platform Architecture**
- 3. Governance**
- 4. Resource Groups**
- 5. Resources**
- 6. Networking**
- 7. Connectivity**
- 8. Identity**
- 9. Security**
- 10. IaC Deployment**
- 11. Operations**

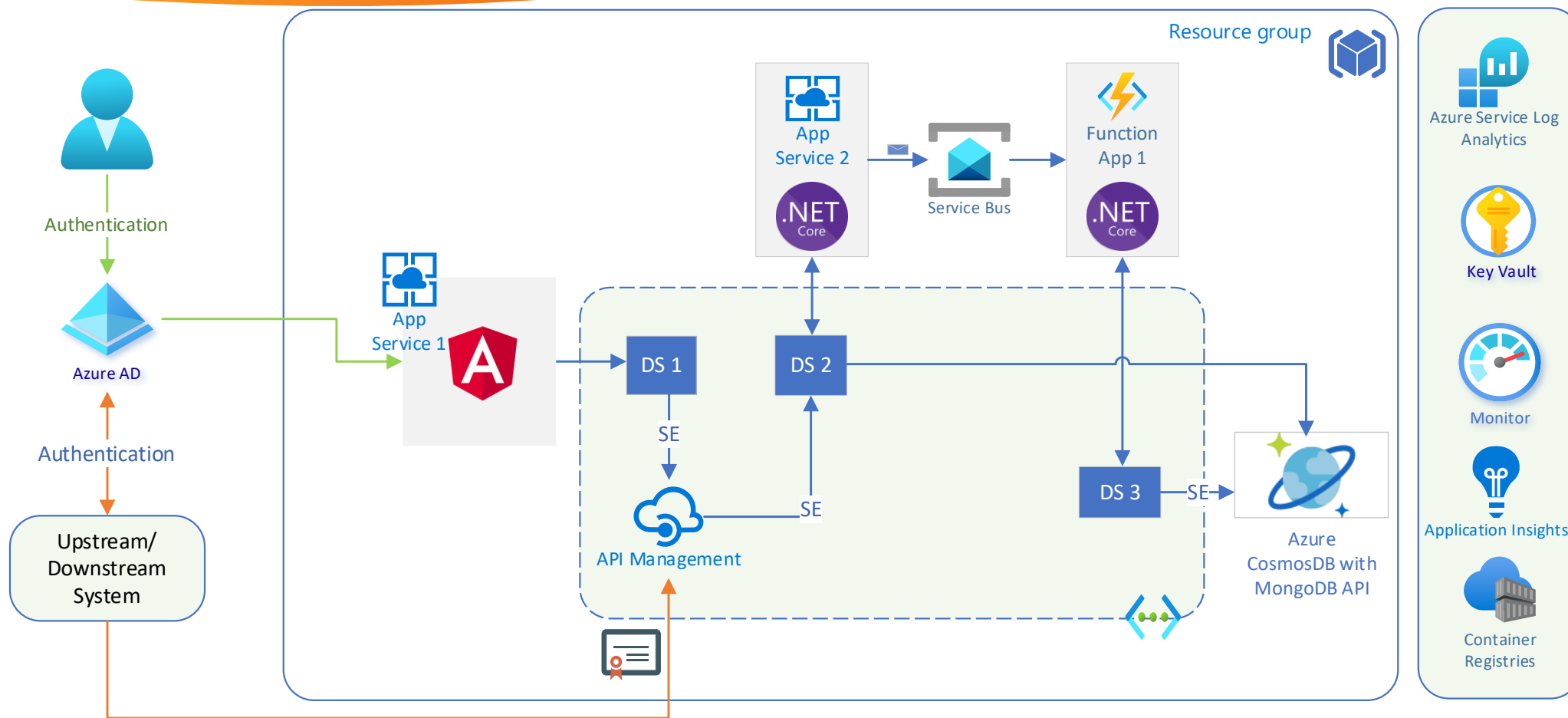


Overview

- This document captures High-Level Design of the Infrastructure Platform of Study Definition Repository Reference Implementation (SDR RI) hosted on Azure* Cloud.
 - To the extent that the SDR Reference Implementation incorporates or relies on any specific branded products or services, this resulted out of the practical necessities associated with making a reference implementation available to demonstrate the SDR's capabilities and is not an endorsement of those products and services.
 - Users remain free to download the SDR source code and design their own implementations.
 - This document outlines each of the architectural construct that make up Azure Fabric (API Management, AppService, AppService Plan, Cosmos DB, Azure Monitor, Application Insights, Key Vault).
- *
- Careful consideration has been given to ensure that the design is extensible and can adapt to evolving requirements.
 - This document is not a Build, Deployment or Operations guide. It only explains the Cloud configurations for SDR reference Implementation Infrastructure demonstrated on Azure Cloud



SDR RI Platform Architecture on Azure



- Azure Service Log Analytics
- Key Vault
- Monitor
- Application Insights
- Container Registries



Governance – Tagging

Key Decision Points for SDR Reference Implementation

For SDR Reference Implementation, following tags have been created as part of the Tagging Strategy:

Tag Name	Description	Required/ Optional	Value	Resources
Environment	Specifies the environment to which the resources belong to. Every resource belongs to a unique environment.	Required	E.g.: Dev, Test	Resource Group, VNet, API Management, App Service Plan, Application Insights, App Service, Log Analytics Workspace, Cosmos DB, Key Vault, Container Registries, Service Bus, Function App.
App_Layer	Identifies whether the resource is part of “Front End” or “Back End”	Required	Frontend/ Backend/ "N/A"	VNet, API Management, App Service Plan, Application Insights, App Service, Log Analytics Workspace, Cosmos DB, Key Vault, Function App, Service Bus.



Governance – Resource Locks

Key Decision Points for SDR Reference Implementation

SDR Reference Implementation has Delete lock on all Resource Groups. This includes all resources servicing core functionality such as network connectivity, security, and automation, as well as any resource running a workload that serves either internal or external customers (i.e., that would create impact for multiple users if it were taken offline or deleted). The locks have been applied at the Resource Group level. Locks applied to a Resource Group are to be applied to all the resources that exist within that Resource Group.

Note: SDR reference implementation has leveraged Read-only locks for archived data that need to be retained for compliance and regulatory needs.

Lock Level	Description
CanNotDelete (Delete)	Authorized users can read and modify a resource, but they can't delete it
ReadOnly (Read-only)	Authorized users can read a resource, but they can't delete or update it. Applying this lock is like restricting all authorized users to the permissions granted by the Reader role



Governance – Naming Convention

<ResourceType>-<Subscription>-<Purpose/Segment/Environment>-<region>-<Instance Number or Level>

Meta Data	Field Required	Field Length
<ResourceType>	Mandatory	Up to 4 Characters
<Subscription>	Mandatory	Up to 7 Characters
<Environment>	Mandatory	Up to 20 Characters*
<Region>	Mandatory	Up to 6 Characters
<Instance Number or Level>	Mandatory	Up to 3 Characters

Usage

Resource Type	Format	Examples
VNet	vnet-<subscription>	vnet-sdr-test-eastus
Subnet	snet-<subscription>	snet-sdr-test-eastus

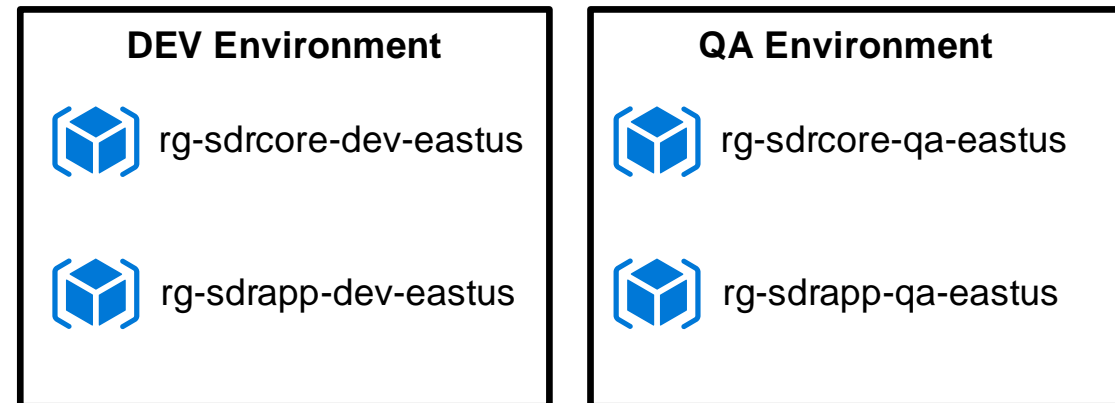
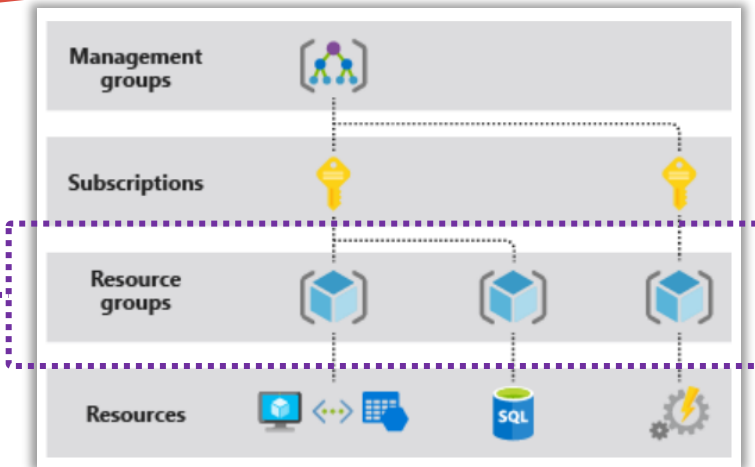
*Note: There are some exceptions to this naming convention where the resource names should be globally unique (like Azure Key Vault, Log Analytic Workspace, Storage Accounts etc.), some which do not allow any special characters (like Storage Accounts) and some have *character limitation.*



Resource Group

Key Decision Points for SDR Reference Implementation

- Resources have been grouped in a Resource Group if they have the same life-cycle such as Applications or Services.
- SDR Reference Implementation used 2 different resource groups for core and app for each environment.
- The core resource group includes the resources – Key Vault, Log Analytics Workspace, Virtual Network and all the remaining resources are part of the rg-sdrapp-<env>-eastus resource group.



Resources - Decisions

Decision Area	Decision
API Management	<ul style="list-style-type: none">• A single API Management is used per environment.• The Application Insights is integrated with API Management.
App Service Plan	<ul style="list-style-type: none">• There are 2 App Service plans used per environment.
Application Insights	<ul style="list-style-type: none">• A single Application Insights is deployed per environment.
App Service	<ul style="list-style-type: none">• There are 2 App Services used per environment – one for the frontend UI and the other for Backend API.



Resources - Decisions

Decision Area	Decision
Cosmos DB	<ul style="list-style-type: none">A single Cosmos DB API for Mongo DB per Environment is used.
Function App	<ul style="list-style-type: none">A single instance of Function App per Environment is used. (Storage account is created by default with a function app)
Service Bus	<ul style="list-style-type: none">A single Service Bus per environment



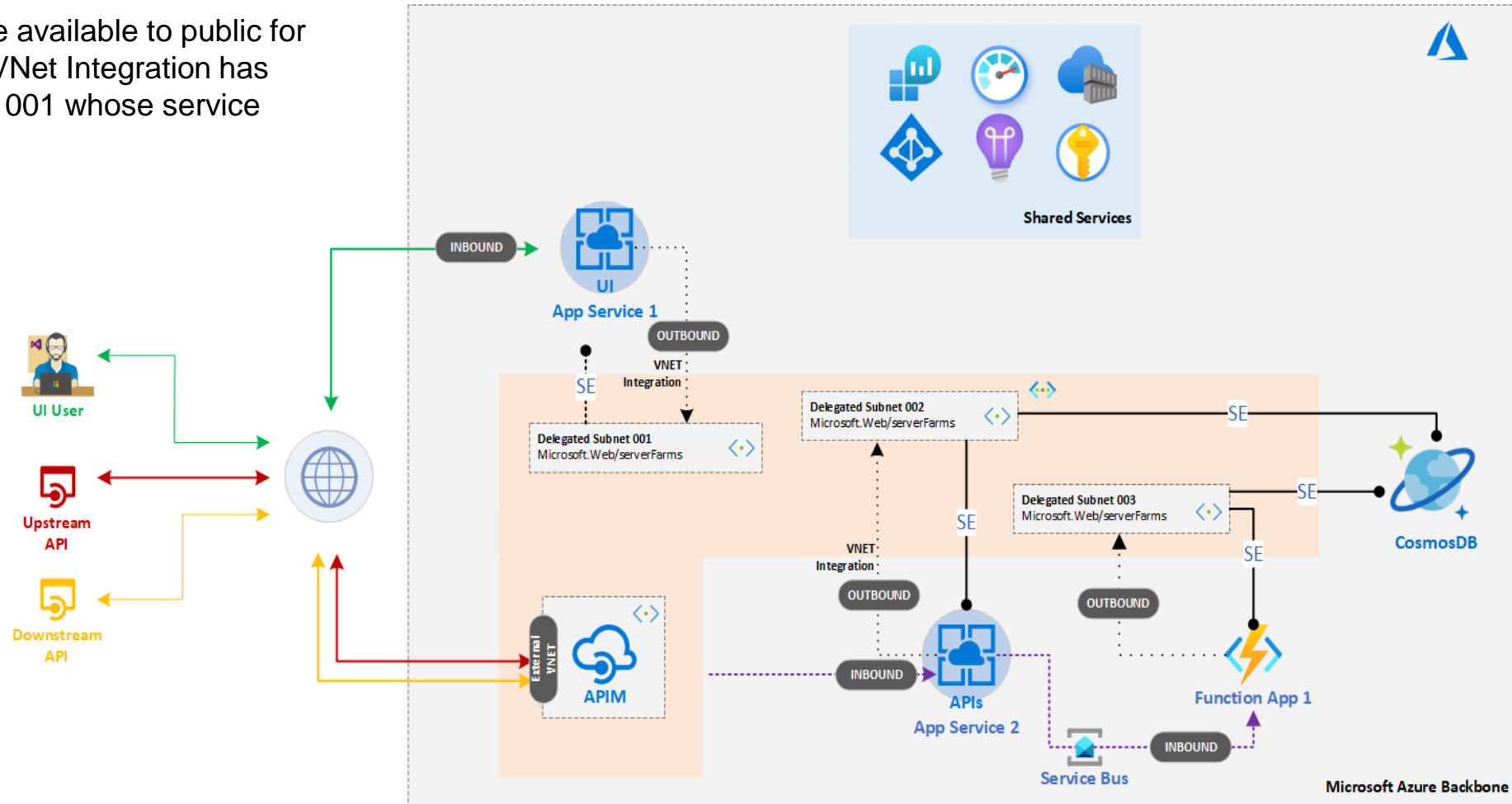
Networking - Key Decisions

Decision Area	Decision
VNet Architecture	Each environment has its own VNET
Subnet	Each environment has a single subnet assigned.
Delegated Subnet	Each environment will have 3 delegated subnets with Microsoft.Web, Microsoft.AzureCosmosDB, Microsoft.Storage as the service endpoints.



Networking cont.

- The Inbound access has been made available to public for App Service1 and Outbound has a VNet Integration has been done to the Delegated Subnet 001 whose service endpoint is Microsoft.Web.
- The Inbound access for App Service2 and Function App has been restricted to VNet through Access Restriction and Outbound has a VNet Integration done to the Delegated Subnet 002 and Delegated Subnet 003 respectively, whose service endpoints are Microsoft.AzureCosmosDB (for both DS) and Microsoft.Web (for DS 002 only)
- API Management is the API Gateway for the Upstream and Downstream APIs. It has been deployed in the VNet and made External.



Key Decisions for SDR reference implementation

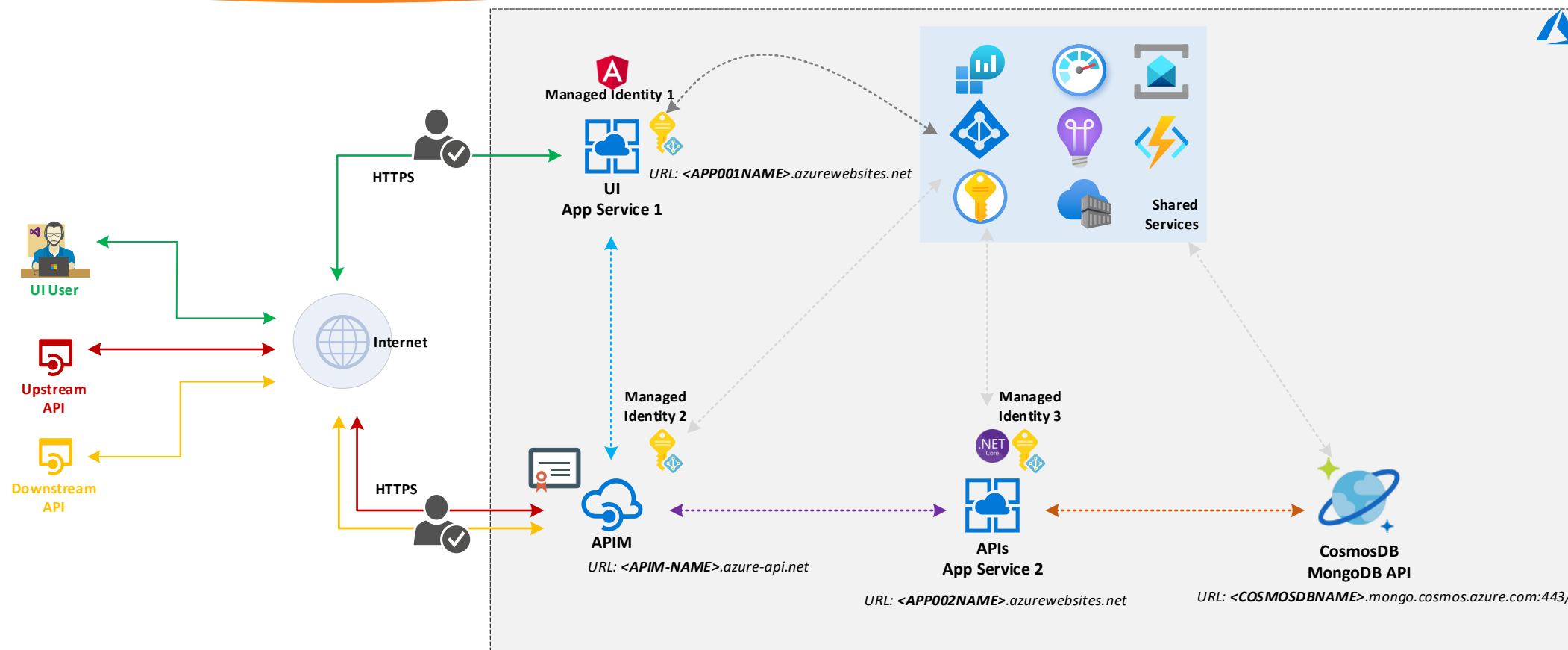
- Principle of least privilege access has been applied. Only minimal permissions needed to accomplish the task have been granted.
- Process has been established around this, as well as a process around periodic access reviews to ensure users who don't need elevated roles are removed.

Communication Flow in SDR

- Communication from Internet to UI App Service is enabled through HTTPS Protocol
- Upstream communication through the internet is handled by APIM which in turn communicates with API App Service. The API App Service forwards the call further to Cosmos DB.
- Downstream communication is also handled by APIM via Internet. APIM handles the calls to UI App Service and API App Service.
- Function App trigger is a queue-based trigger and can be invoked only from messages pushed to the Service Bus.
- API App Service, and CosmosDB Mongo API App cannot be reached directly from the internet. All the calls to these resources are handled by APIM.
- The shared services in Azure are enabled to handle the requests from UI App Service, APIM, API App Service for Logging, Monitoring etc.



Connectivity- Diagram



INDEX

- ↔ End User call to UI App Service
- ↔ Upstream call API to APIM
- ↔ Downstream API call to APIM

- ↔ UI App Service call to APIM
- ↔ APIM call to API App Service
- ↔ API App Service call to Cosmos DB

- ↔ Virtual network
- ↔ OAuth 2.0



Identity - Decisions

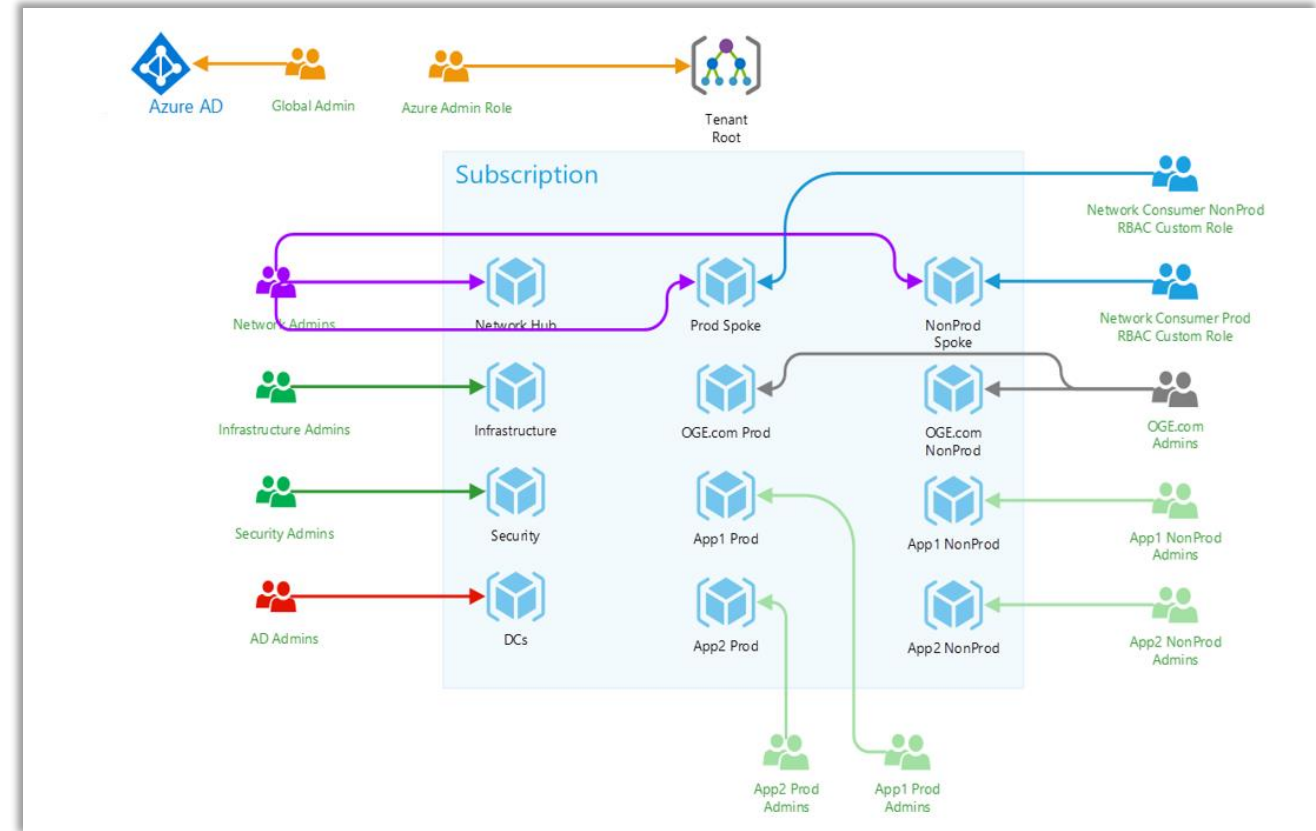
Decision Area	Options	Decision
Azure AD	<ul style="list-style-type: none"> • Use Existing Azure AD Tenant • Use New Azure AD Tenant • AD Integration/ Synchronization 	<ul style="list-style-type: none"> • The SDR Reference Implementation application has leveraged Azure AD for Authentication.
Admin Scope	<ul style="list-style-type: none"> • Limited owner permissions 	<ul style="list-style-type: none"> • Subscription owner permissions assigned to limited users.
Service Principal	<ul style="list-style-type: none"> • Integration with GitHub Actions • Authentication/Authorization of App 	<ul style="list-style-type: none"> • For IaC code deployment, a service principal is configured with appropriate privileges to handle the resource deployment in azure.
Managed Identity	<ul style="list-style-type: none"> • System assigned • User assigned 	<ul style="list-style-type: none"> • System Managed Identities are used, and it's configured for 3 resources – APIM, UI App Service, API App Service.
RBAC	<ul style="list-style-type: none"> • Principle of least privilege • Establish process around to strive this goal 	<ul style="list-style-type: none"> • Security Groups are leveraged • Multiple groups have been created to handle the access requirements and segregate the user permissions accordingly



Role based access control

Key Decision for SDR Reference Implementation

- Principle of least privilege access has been applied while giving access. Minimal permissions have been granted that are needed to accomplish the task.
- Process has been established around this, as well as a process around periodic access reviews to ensure users who don't need elevated roles are removed.
- All Users have not been given access to the resources. Privileged access for select users has been given with assigned roles via groups and are revoked once the necessary work is completed.















Security - Decisions

Decision Area	Options	Decision
Azure Key Vault	<ul style="list-style-type: none">• Key Vault Breakdown based on Environment	<ul style="list-style-type: none">• Cloud native SDR deployment has single Key Vault per Environment
Security Monitoring	<ul style="list-style-type: none">• User Management & Activity Monitoring• Restriction of Security log Access• Regulatory Compliance Monitoring	<ul style="list-style-type: none">• Application and User Activity Monitoring - Cloud Native Tool – Azure Monitor has been configured• Compliance Monitoring is being performed by Accenture Cloud Platform
Certificates	<ul style="list-style-type: none">• Certificate based authentication• Password based authentication	<ul style="list-style-type: none">• Certificates through APIM have been leveraged.• Root CA and Client certificates are configured in APIM



Perimeter Security – Decisions

Area	Features	Decision
DDoS Protection	<ul style="list-style-type: none"> DDoS protection @ edge 	<ul style="list-style-type: none"> Azure Basic DDoS Protection has been used

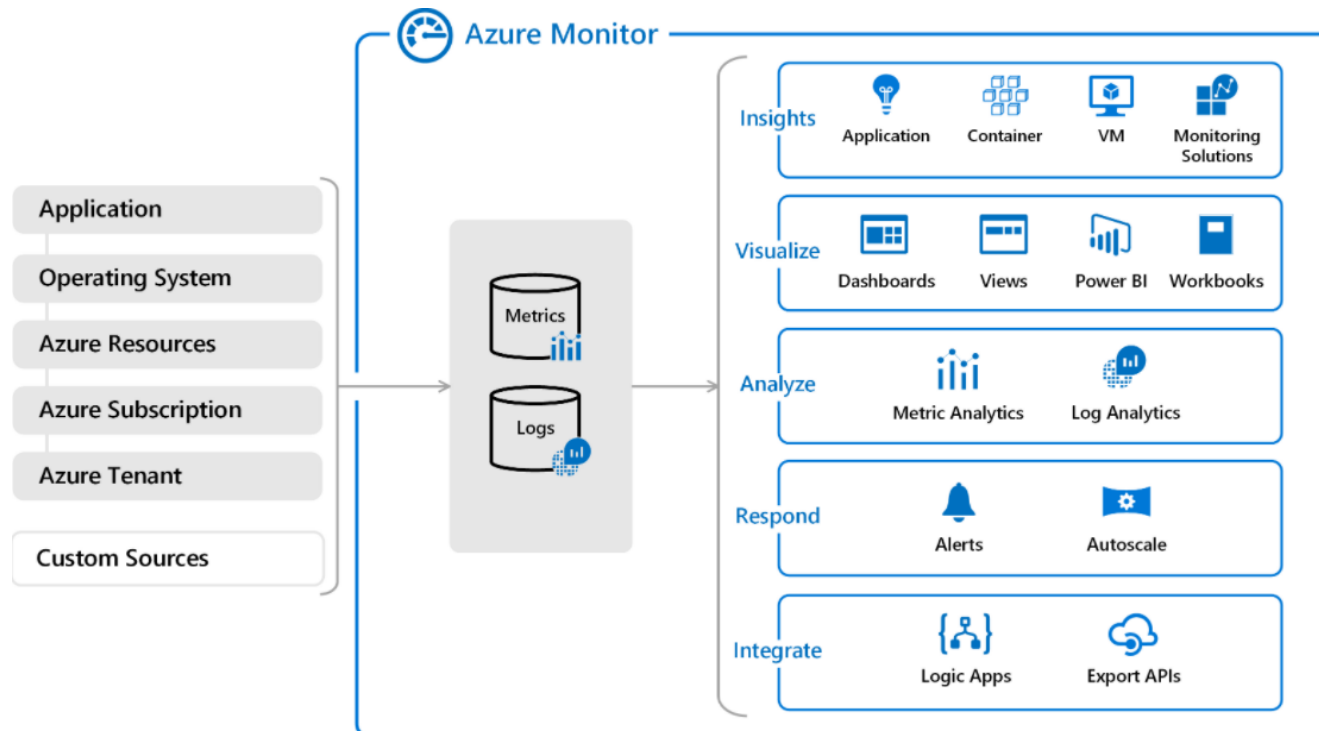
Feature	 DDoS Protection Basic	 DDoS Protection Standard
 Active traffic monitoring & always on detection	Yes	Yes
 Automatic attack mitigations	Yes	Yes
 Availability guarantee	Azure region	Application
 Mitigation policies	Tuned for Azure region traffic volume	Tuned for application traffic volume
 Metrics & alerts	No	Real time attack metrics & diagnostic logs via Azure monitor
 Mitigation reports	No	Post attack mitigation reports
 Mitigation flow logs	No	NRT log stream for SIEM integration
 Mitigation policy customizations	No	Engage DDoS experts
 Support	Best effort	Access to DDoS Experts during an active attack
 SLA	Azure region	Application SLA guarantee & cost protection



Security – Logging & Monitoring – Azure Monitor

Key Decision for SDR Reference Implementation

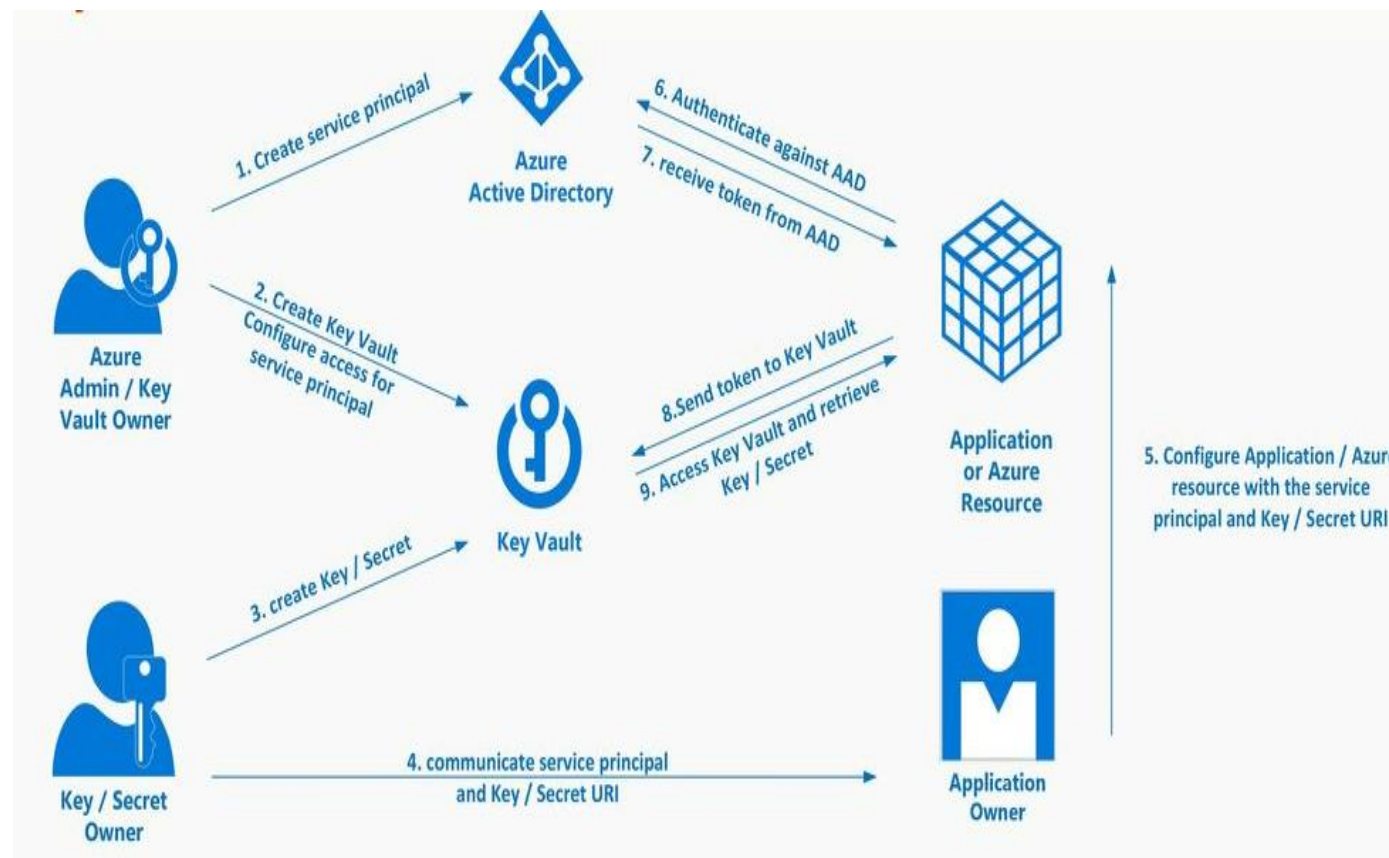
- Centralized Log Analytics workspace per environment to log all the events and metrics.
- SDR Reference Implementation has leveraged built-in cloud native tools Azure Monitor, Application Insights for Monitoring with its default configuration.



Security - Azure Key Vault

Key Information

- Key Vaults have been separated by environment since environments should not share any credentials.
- Access policies have been configured for each vault to determine what operations a user, group, app, or Service Principal can perform.
- Key Vaults have been locked to CanNotDelete. Purge protection and soft delete have been enabled for recovery in case of a malicious or accidental deletion attempt.
- SDR RI UI and API have been integrated with KeyVault to store or retrieve secrets like connection strings, credentials etc.



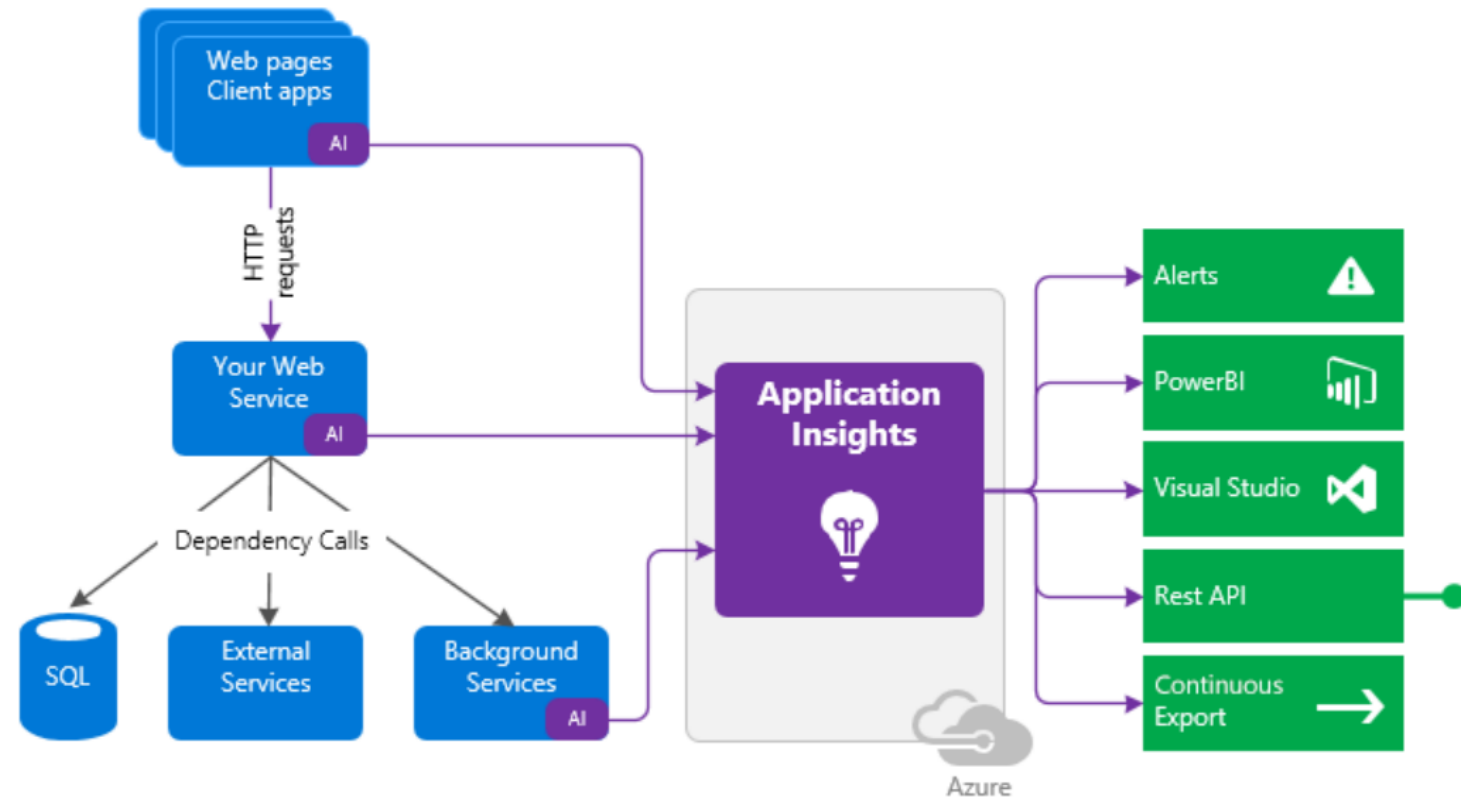
API Management

- The SDR RI APIs are exposed via API Management
- SDR RI UI and external API - Users can interact with APIs only through APIM.
- The SDR Reference Implementation architecture prevents calling APIs directly. All web service requests are channeled through the API Management layer.



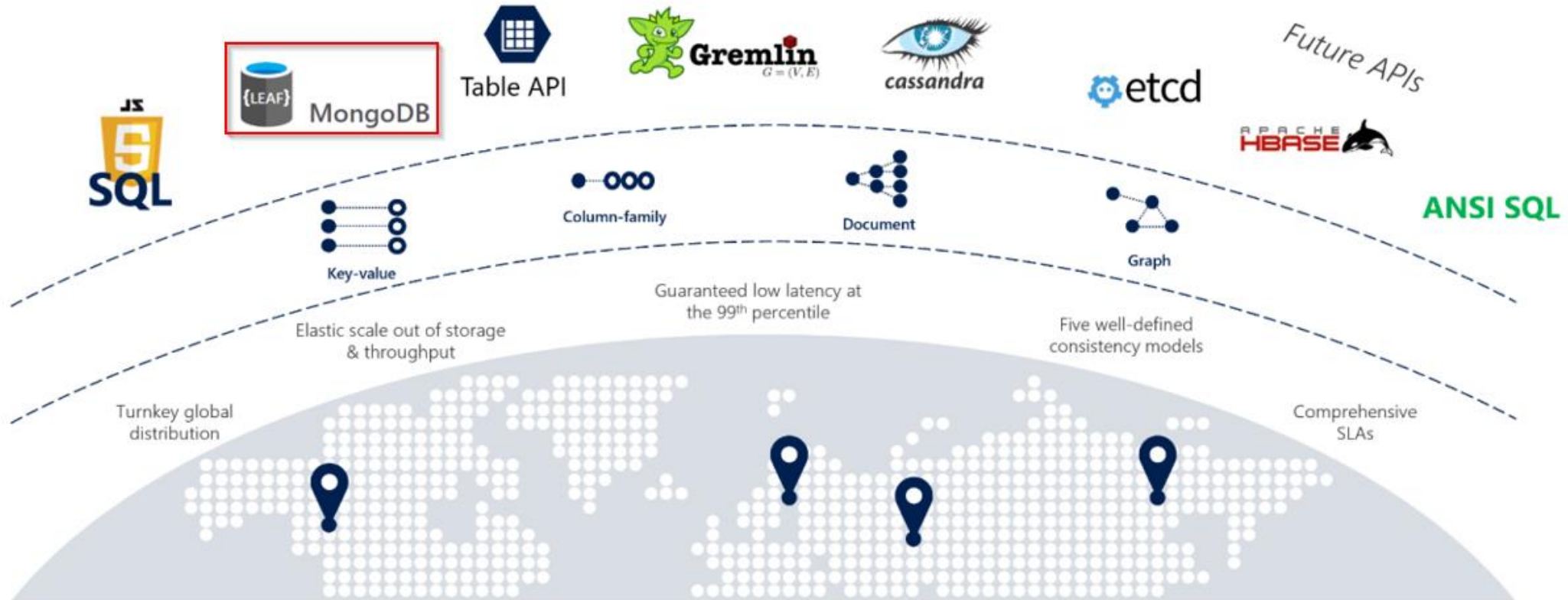
Application Insights

- SDR Reference Implementation utilizes 1 Application Insights per Environment. Application Insights, a feature of Azure Monitor, is an extensible Application Performance Management (APM) service for developers and DevOps professionals. It is used to monitor live applications.
- Automatically detects performance anomalies, and includes powerful analytics tools to help diagnose issues and to understand what users do with your app. It's designed to help continuously improve performance and usability.

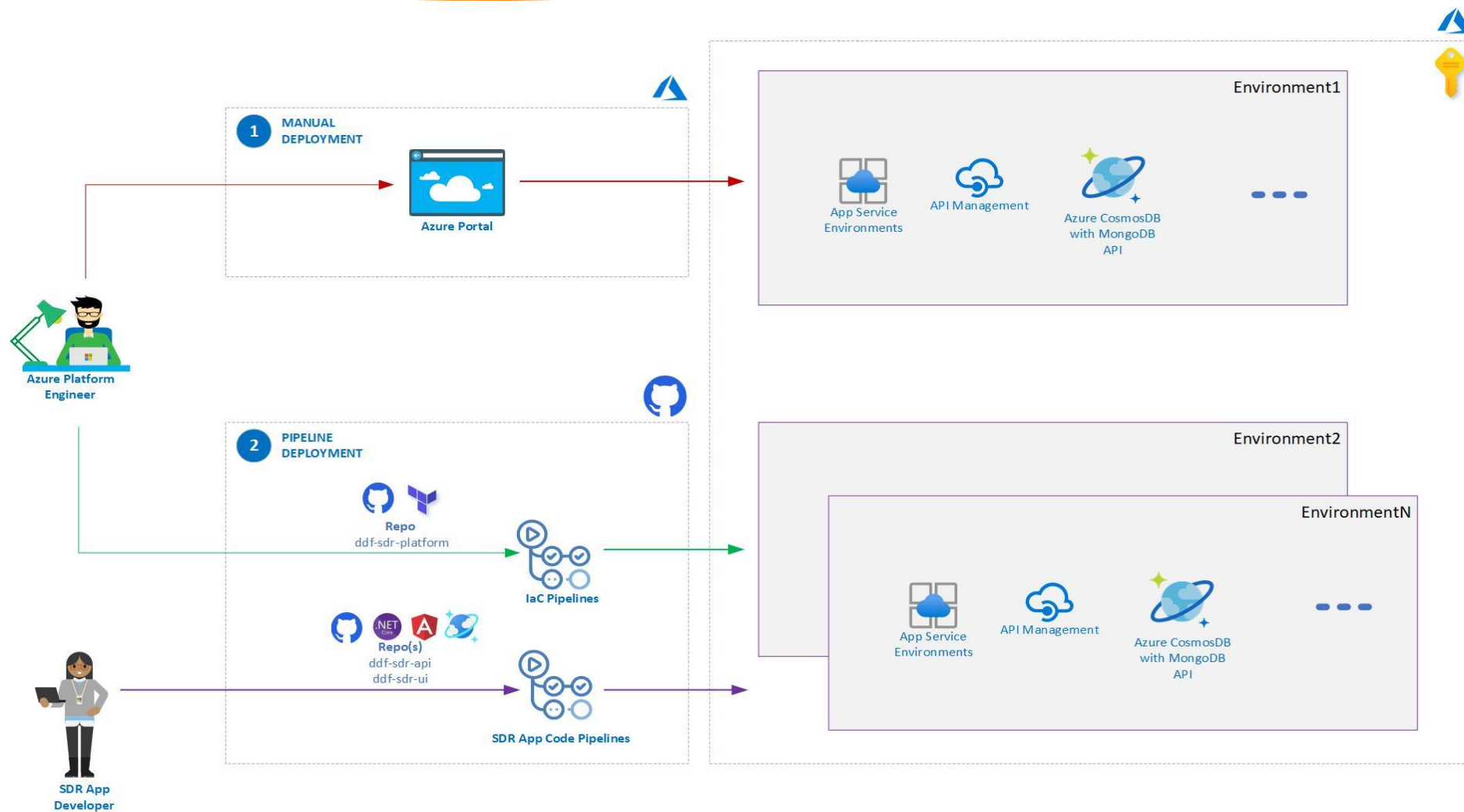


COSMOS DB

- SDR Reference Implementation is built on one Azure Cosmos DB API for MongoDB per Environment.



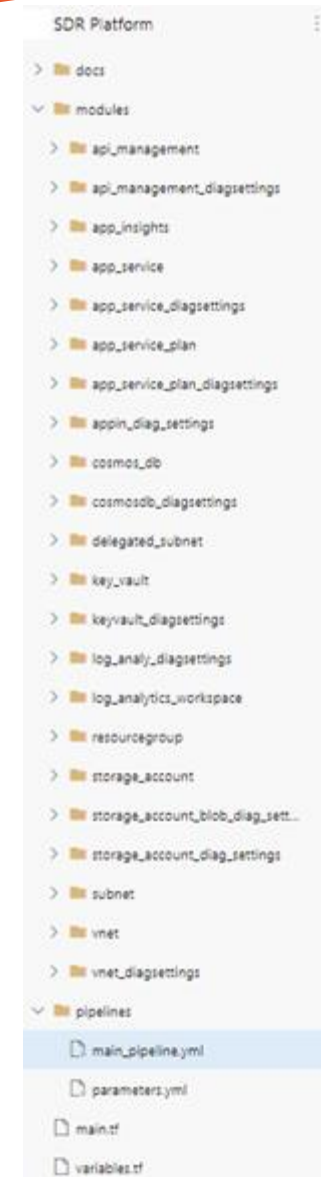
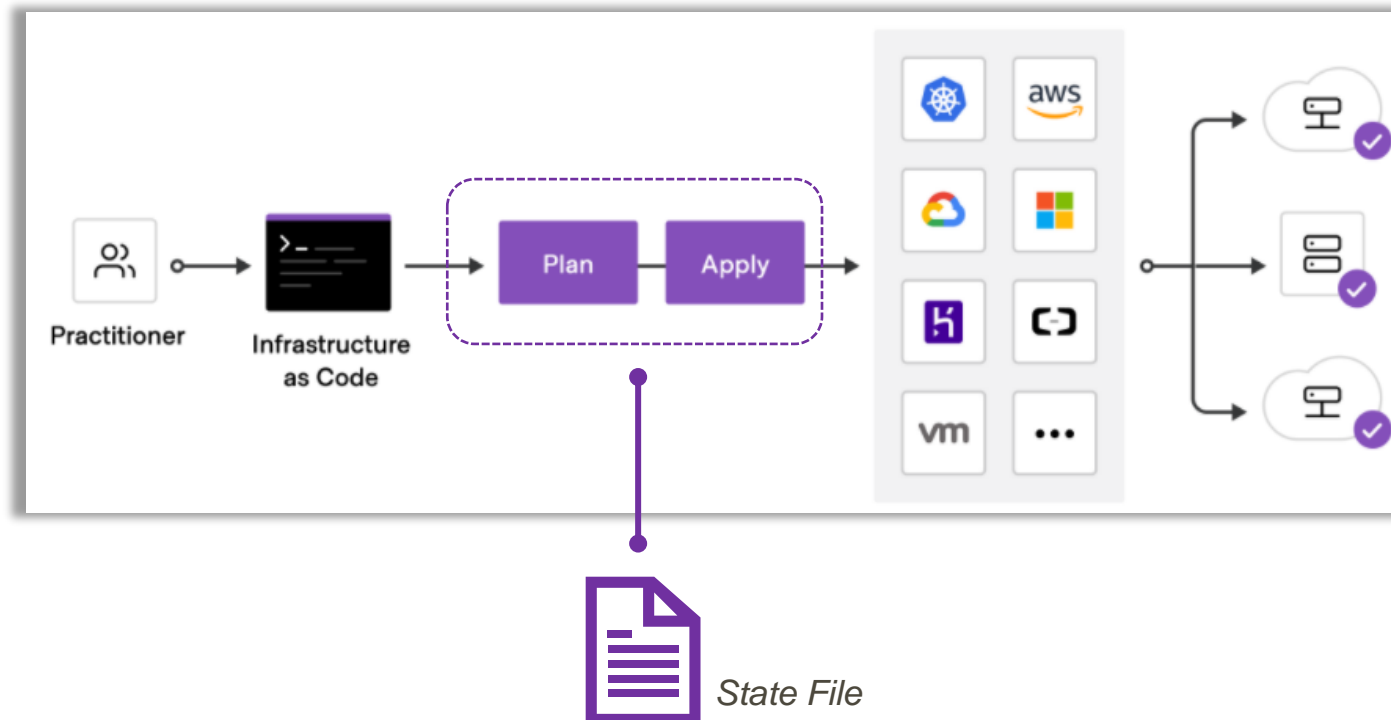
Infrastructure as Code (IaC) – Page 1/2



- **Environment Build Reliability**
 - *Idempotency*
 - *Immutability*
- **Speed**
- **Modularity**



Infrastructure as Code (IaC) – Page 2/2

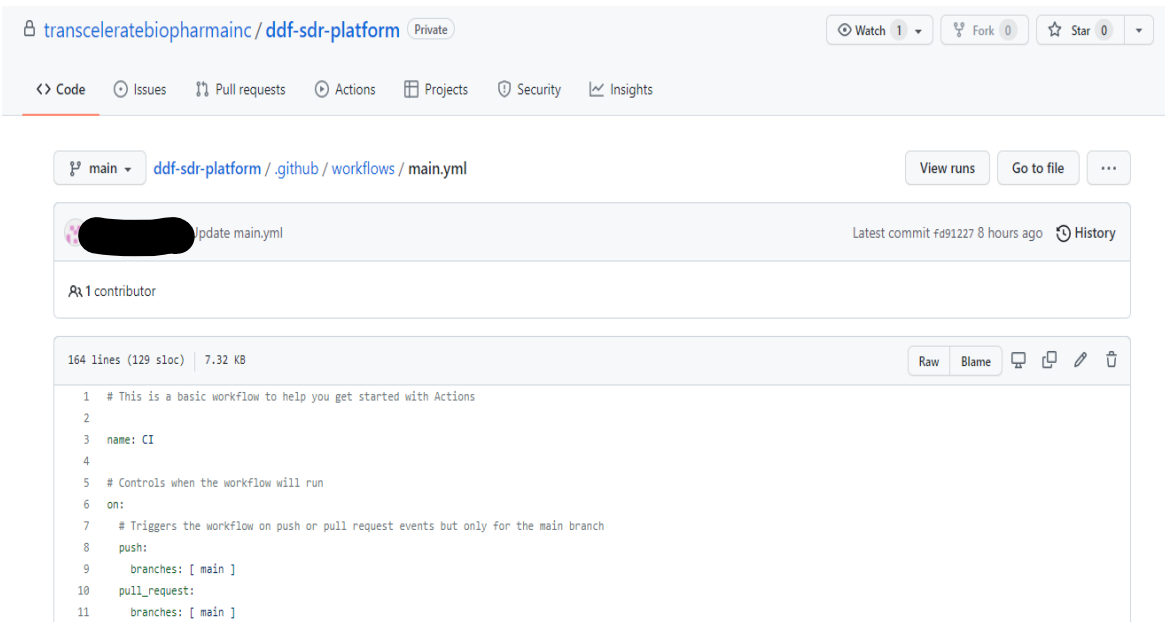
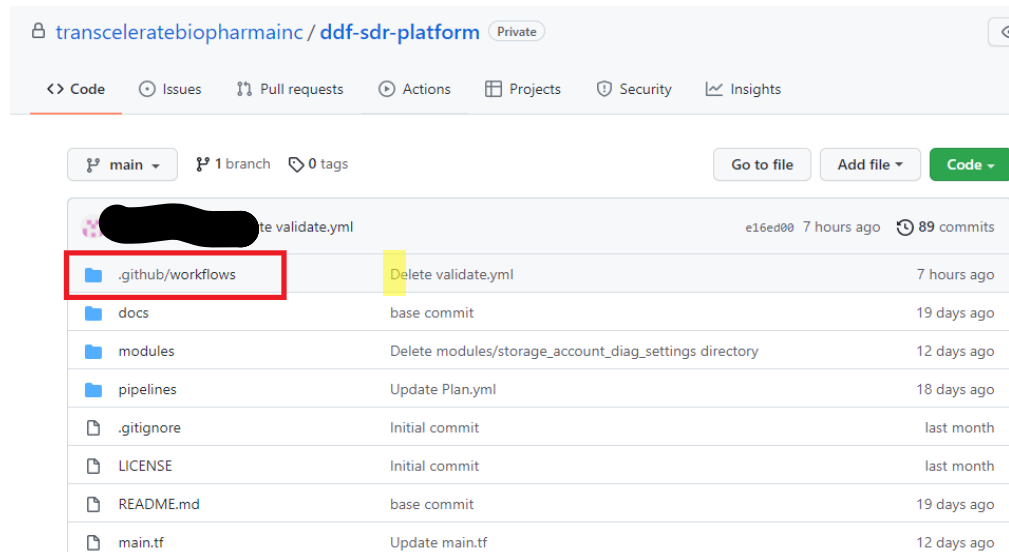


- **IaC Components**
 - *State File*
 - *Plan & Apply Code*
 - *Environment Variables*
- **Capabilities**
 - *Environment Build*
 - *Environment Destroy*



IAC Code Deployment

- SDR RI IAC code is deployed using GitHub Workflows & Actions:
- The GitHub actions are run directly in the GitHub environment which would automatically deploy the resources in Azure based on the configuration in the “main.yml” and “main.tf” file.



Operations - Decisions

Decision Area	Options	Decision	Rationale
Infrastructure Logging & Monitoring	<ul style="list-style-type: none">Azure Resources (IaaS, PaaS)<ul style="list-style-type: none">Application Logging & MonitoringEndpoint Logging & Monitoring	<ul style="list-style-type: none">Cloud Native Infrastructure Monitoring (Log Analytics & Azure Monitor)	<ul style="list-style-type: none">Diagnostic MonitoringApp Insights MonitoringLog Analysis
Cosmos DB Backup	<ul style="list-style-type: none">Backup IntervalBackup RetentionBackup Storage Redundancy	<ul style="list-style-type: none">Continuous Restoration policy enabled.Allows to do restore to any point of time within the last 30 days.Geo-redundant backup storage enabled.	

