UNIVERSITÀ degli STUDI di CATANIA

Università degli Studi di Catania
Dipartimento di Matematica e Informatica

corso laurea informatica catania clic

Innovazione nell'analisi dei dati

8 - 15 maggio

Coordinatore: F. Stanco
DMI - Università degli Studi di Catania

M. Marroccia, G. Ursino, G. Scuderi, F. Milotta
STMicroelectronics

life.augmented

STMicroelectronics: a global presence

- One of the world's largest semiconductor companies
- 2019 revenues of **$9.56B**
- **46,000** employees of which **7,800** in R&D
- Over **80** Sales & marketing offices serving over **100,000** customers across the globe
- **11** Manufacturing sites
- Signatory of the United Nations Global Compact (UNGC), Member of the Responsible Business Alliance (RBA)

Top 10 Customers 2019	
Apple	Huawei
Bosch	Mobileye
Ciena	Samsung
Continental	Seagate
HP	Tesla

life.augmented

Where you find us

Affordable, desirable electric vehicles

Increase safety for road users
& driver comfort and convenience

Cleaner, greener Internal Combustion Engines

Making **driving** safer,
greener
and more connected

Rising demand for and usage of electrical energy

Enabling the evolution of
industry towards
smarter, safer and more
efficient factories and
workplaces

Cloud connected and data-enabled services

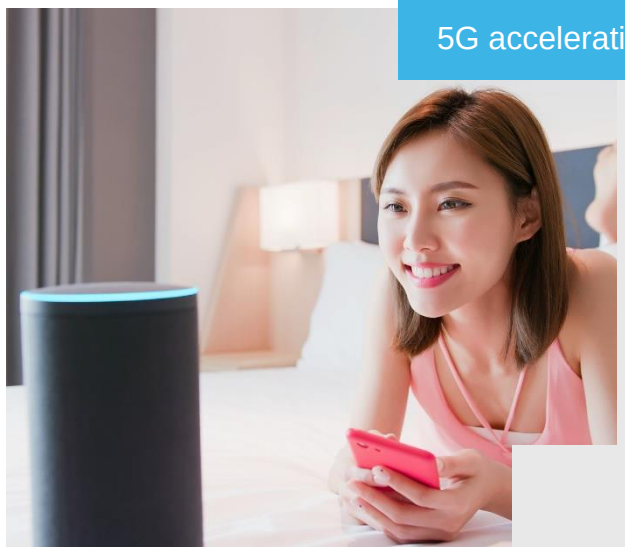
Digital security for all data

5G accelerating the connection of objects to the IoT

Decrease carbon emissions to reduce global warming impact

Increase use of renewable energy

Making **homes & cities**
smarter, for better living,
higher security,
and to get more from
available resources



Making everyday **things**
smarter, connected
and more aware
of their surroundings

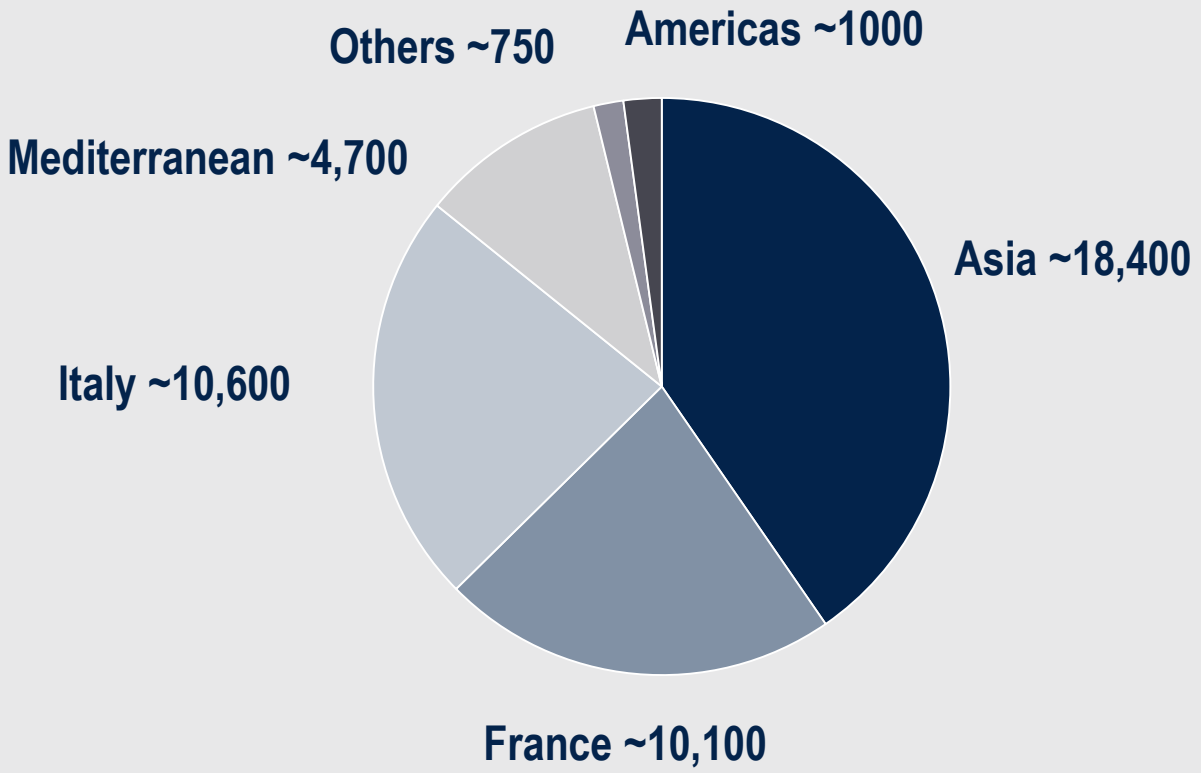
We are drivers of your innovation

Advanced R&D centers around the world for close collaboration with operations and customers

- **18,500** patents &
  590 new filings in 2019
- **16%** of revenues invested in R&D
- **7,800** people working in R&D and product design

life.augmented

Our People



Manufacturing	~ 67%
Research & Development	~ 17%
Marketing & Sales, Divisional Functions, Administration & General services	~ 16%

As of December 31, 2019

Sustainability

life.augmented

Where will we be in 5 years?

VUCA

Did you know?

- Volatile
- Uncertain
- Complex
- Ambiguous

VUCA

Did you know?

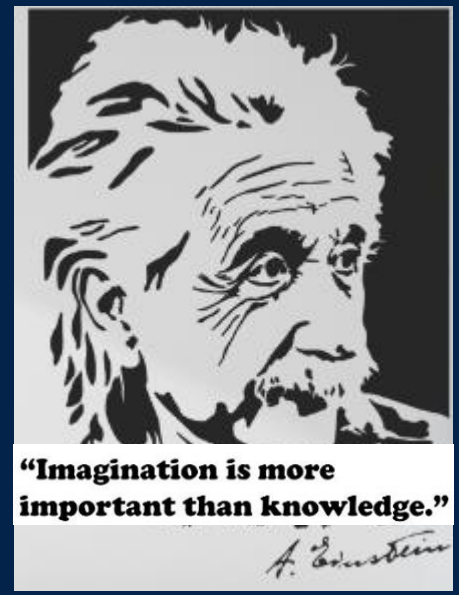
- Top 10 In-Demand Jobs in 2018
- Did not exist in 2008

- 1984: 1000 @ devices
- 2020: 75 000 000 000 @ devices

- The « PRESENT » has never been
- SO SHORT

- 5 500 000 000 searches on
- On Google per day
- 400 hours of video per minute to YouTube
- 692 000 000 Tweets per day
- 54 000 000 000 What's app per day

- We are preparing STUDENTS
- For jobs that do not yet exist…
- For technologies that have not been invented...



- 2 200 000 000 Facebook Users
- 1 100 000 000 Instagram Users
- 33 zettabytes of new data created in 2018

Did you know?

- DISRUPTION is new « Normal »
- BUT
- CONSTRUCTION is ESSENTIAL

life.augmented

UNIVERSITÀ degli STUDI di CATANIA

Università degli Studi di Catania
Dipartimento di Matematica e Informatica

corso laurea informatica catania clic

Innovazione nell'analisi dei dati

Cloud Architectures

venerdì 8 maggio - 09.00-12.00

Coordinatore: F. Stanco
DMI - Università degli Studi di Catania

M. Marroccia, G. Ursino, G. Scuderi
STMicroelectronics

life.augmented

CLOUD ARCHITECTURES - FOCUS ON SECURITY AND DATA GOVERNANCE ASPECTS

STMIcroelectronics

life.augmented

Cloud data Governance - How to Successfully Design and Implement a Data-Centric Security Architecture

Giuseppe URSINO

Agenda

- # What’s “Cloud Computing”
- # What’s “XaaS”
- # What is Data Governance
- # Data Integration
- # Data Classification

life.augmented

What is Cloud Computing?

> ***A style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service using internet technologies.***

1. Service-Based
Consumer concerns are abstracted from provider concerns through service interfaces that are well-defined. The interfaces hide the implementation details and enable a completely automated response by the provider of the service to the consumer of the service. In addition, the service could be considered **"ready-to-use"** or "off the shelf" because the service is designed to serve the specific needs of a set of consumers, and the technologies are tailored to that need rather than the service being tailored to how the technology works. The articulation of the service feature is based on service levels and IT outcomes (availability, response time, performance versus price, and clear and predefined operational processes), rather than technology and its capabilities. In other words, **what the service needs to do is more important than how the technologies are used to implement the solution.**

2. Scalable and Elastic
The service can scale capacity up or down as the consumer demands at the speed of full automation (which may be seconds for some services and hours for others). Elasticity is a trait of shared pools of resources. Scalability is a feature of the underlying infrastructure and software platforms. Elasticity is associated with not only scale but also an economic model that enables scaling in both directions in an automated fashion. This means that services scale on-demand to add or remove resources as needed.

3. Shared
Services share a pool of resources to build economies of scale. IT resources are used with maximum efficiency. The underlying infrastructure, software or platforms are shared among the consumers of the service (usually unknown to the consumers). This enables unused resources to serve multiple needs for multiple consumers, all working at the same time.

4. Uses Internet Technologies
The service is delivered using Internet identifiers, formats and protocols, such as URLs, HTTP, IP and representational state transfer web-oriented architecture.

5. Metered by Use
Services are tracked with usage metrics to enable multiple payment models. The service provider has a usage accounting model for measuring the use of the services, which could then be used to create different pricing plans and models. These may include pay-as-you-go plans, subscriptions, fixed plans and even free plans. **The implied payment plans will be based on usage, not on the cost of the equipment.** These plans are based on the amount of the service used by the consumers, which may be in terms of hours, data transfers or other use-based attributes delivered.

WHAT'S THE DIFFERENCE BETWEEN SaaS, PaaS, IaaS?

> ***'XaaS' is a style of computing where scalable and elastic IT-related capabilities are provided "as a service" to external customers using internet technologies.***

When it has made the decision to consider cloud services for an application or infrastructure deployment, it's important to grasp the fundamental differences between the core categories of cloud services available.
The cloud is a very broad concept, and it covers just about every possible sort of online service, but when businesses refer to cloud procurement, there are usually three models of cloud service under consideration, Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). Each has its own intricacies and hybrid cloud models, but today we're going to give you an understanding of the high-level differences between SaaS, PaaS, and IaaS.

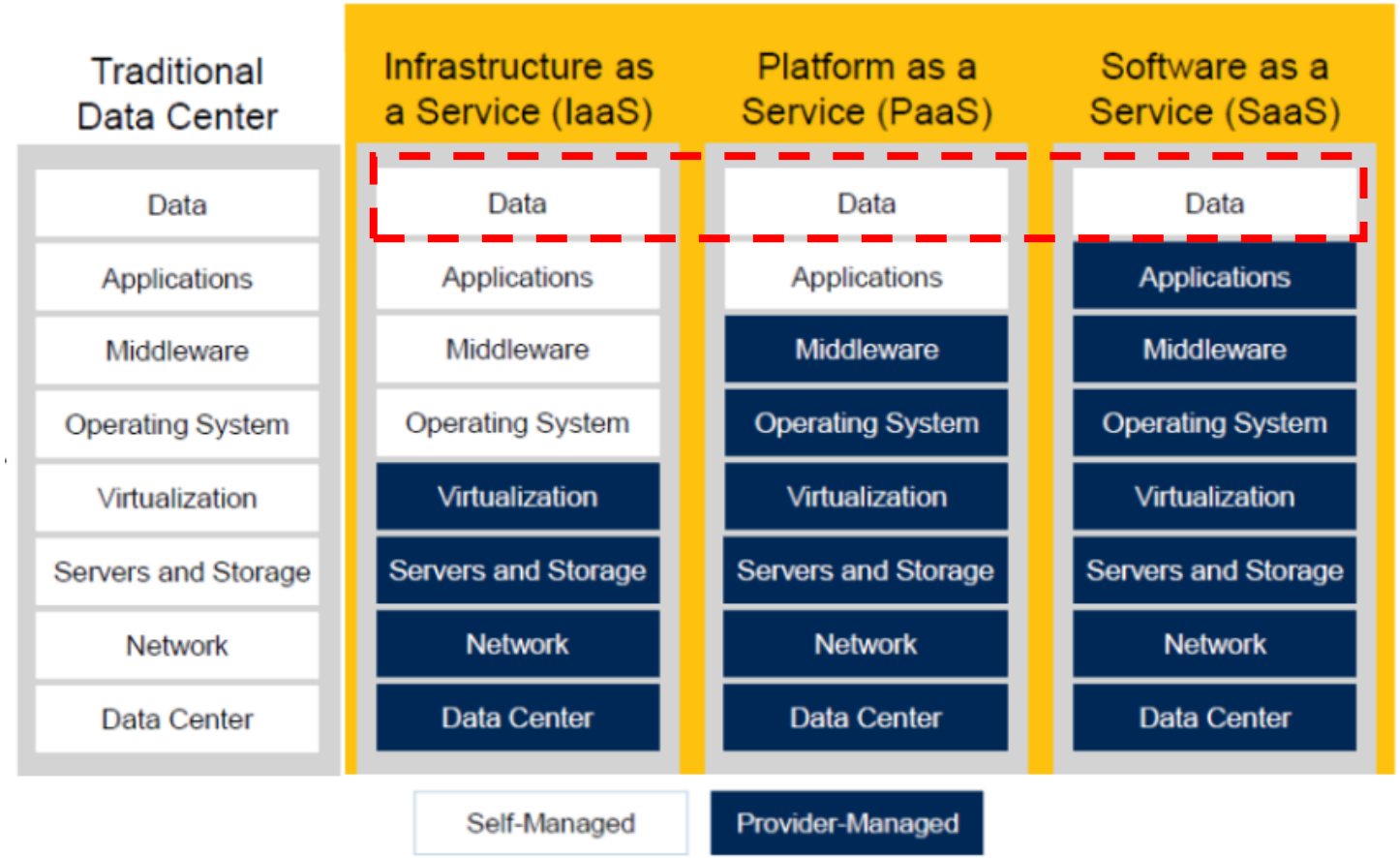
Software-as-a-Service (SaaS)
Software deployment model where applications are delivered as a service to the customer

- **CRM** as a service (e.g., Salesforce)
- **Office Productivity Software** as a service (e.g., O365)

Platform-as-a-Service (PaaS)
Software development platform that supports the full software lifecycle (design, testing, and development)

- **Development & Testing** as a service (e.g., Microsoft Azure)
- **Database** as a service (e.g., Azure SQL DB, Azure Hyperscale SQL DB)

Infrastructure-as-a-Service (IaaS)
Model for provisioning core computing power (servers, storage) for applications and data, virtualization, etc.. (provisioning of VM, containers, etc…)



life.augmented

WHAT is Data Governance?

Policies definition and orchestration

Set of roles, accountabilities and responsibilities within the organization

to define and orchestrate

the corporate guidelines, processes, policies, standards, and technologies implementations

Usage of data to foster Company's Strategy

Management, ownership, security, availability, accessibility, quality, consistency, and auditability of data

and their alignment with company strategy

life.augmented

Data is an Asset

- Protecting data has long been the primary goal for security organizations, as evidenced by the classic confidentiality, integrity and availability (CIA) triad. **Yet, security controls focus on users and systems, and are rarely architected to focus on the data itself.**
- **Data is an asset class that is difficult to secure comprehensively**, unless a broader view of its life cycle is taken. **The challenge organizations are facing today is that, unlike a server, data is pervasive and does not stay in a single silo. Data has become an infinitely mutable and movable resource that is stored and processed in a decentralized ecosystem of data silos, often for many different business purposes.**
- This semi-chaotic environment can frequently lead to the implementation of duplicate security controls in different silos or, worse, within a single environment. Besides the financial and operational implications, this situation can introduce inconsistencies in control outcomes, leading to a lower level of risk management.

HOW? Set of role in Data Governance

- Data Owner (Business representative)
  - Executive manager who has responsibility over the business domain
  - Responsible of the data owned by his/her business organization, he/she leads and promote data governance to ensure data security, availability and quality
  - Appoint the data steward(s)

- Data Steward (Business representative)
  - Expert of Business Domain data
  - Ensure the right data usage, computation, data process, monitoring, data generation through business processes by providing correct data requirements
  - Define the policies and the guidelines to preserve data value, quality, security, accessibility and ensure data governance aligned with corporate strategy
  - Can delegate one or more Data Stewardships for dedicated area/programs

- Data Stewardship
  - Same as Data Steward for particular initiatives and/or area.

- Data Gatekeeper/Custodian
  - DIT Expert of Business Domain
  - DIT counterpart of the Data Steward
  - Support Data Steward by ensuring implementation of all the requirements and policies to properly manage and integrate data in the IT Landscape

life.augmented

Details: accountabilities & responsibilities

Data Owners are business people, usually executive profiles, who have direct line responsibility for the functional area that owns the data.

- Accountabilities:
  - Implementation of data governance organization
  - Definition and implementation of processes, guidelines and standards
  - Data compliancy to external / internal regulation (GDPR, liability, SOP, contractual liabilities…)
- Responsibilities:
  - Nominate Data Steward
  - Approval of exceptions

As leaders in the user community, they are part of the team that drives the governance process since the need for governance should originate and be maintained in the business community.

Data Steward is a business senior manager, domain's data expert, that can influence business and operational decisions, and obtain stakeholder commitments.

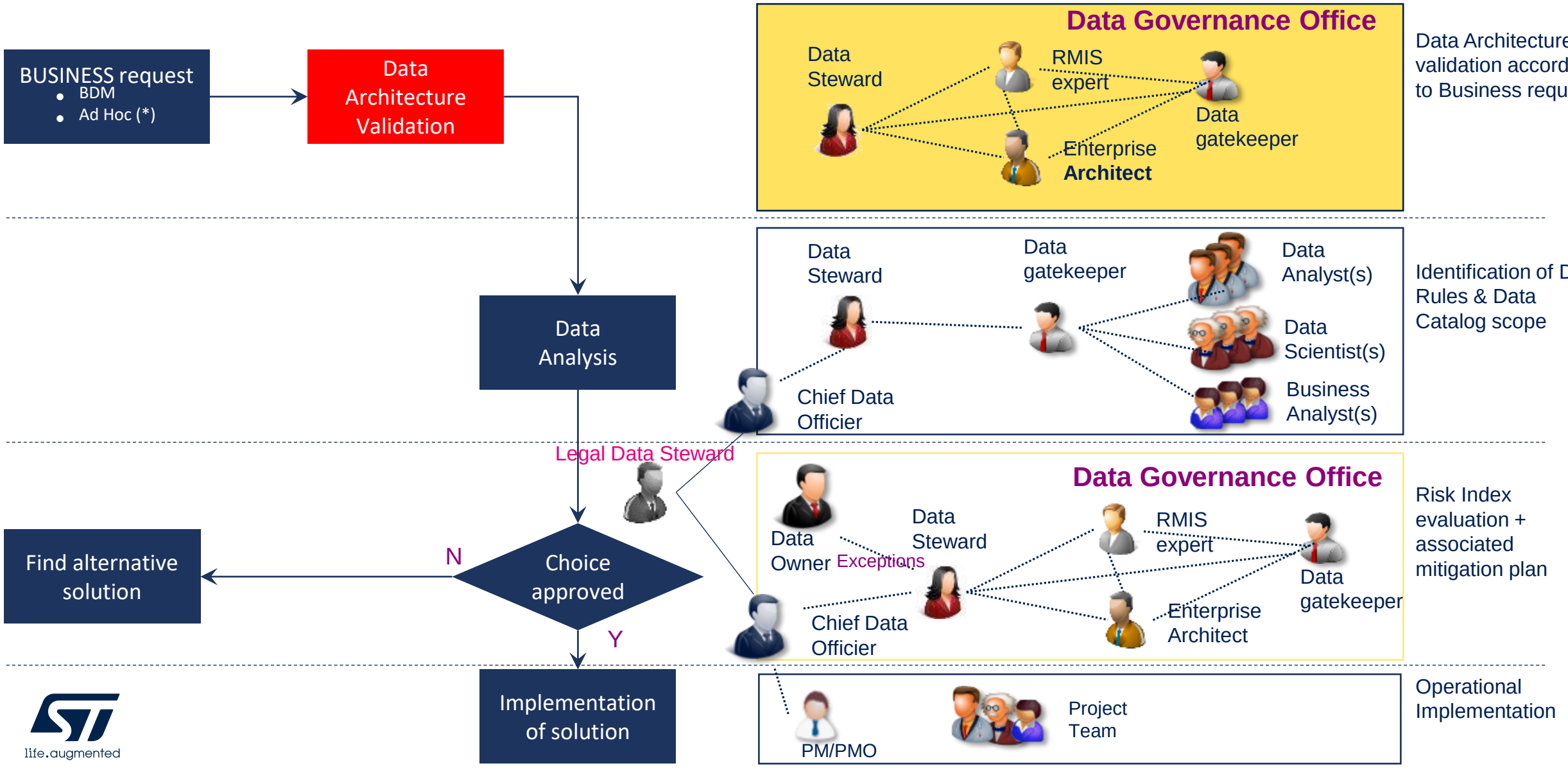
- Accountabilities:
  - Ensuring monitoring of information assets and data collection
  - Data Security, Protection, Quality, Consistency, and Availability
  - Ensuring data access controls according to corporate policies and compliancy to regulation
  - Monitoring usage, relevance and data quality (accuracy, completeness, consistency, timeliness,….) of data published
- Responsibilities:
  - Ensuring data access controls according to corporate policies and compliancy to regulation
  - Definition and implementation of processes, guidelines and standards
  - Data Compliancy to external / internal regulation (GDPR, liability, SOP, contractual liabilities…)
  - Data Security, Protection, Quality, Consistency, and Availability
  - Ensuring monitoring of information assets and data collection

Data stewards rely on a support network within the business organization to deploy processes, ensure data accuracy, timeliness, analysis and resolution of issues or to coordinate the same with external partners, customers and vendors when required.

Data Gatekeeper/Custodian ensures the technology support to data governance; collaborating with Data Steward and implement systems, integration, and monitoring for data management

- Accountabilities:
  - Data analysis and provide proper artifacts to Data Steward
  - Implementation of given policies, guidelines, and standards
- Responsibilities:
  - Support Data Steward by implementing monitoring usage, relevance and data quality (accuracy, completeness, consistency, timeliness,….) of data published
  - Granting data governance as per given policies implementing
    - Data Collection and maintenance
    - Data integration and processes
    - Data security
    - Access control
    - Data quality controls
  - Create and manage the metadata for published data sources ensuring usability and scalability
  - Trigger Data Governance Office and Data Steward when required

Data Governance Process



life.augmented

Why Data Governance?

- Data is always changing
- Data growths exponentially
- Business is always changing
- Regulatory, Compliancy, Policy etc…
- New Security Policies
- New Risks

Common questions on and with the Cloud:

Are you still doing the right things?

Which Architecture(s)? Which Governance?

life.augmented

Key challenges for data governance and security

- Continual increases in the **Volume, Velocity, Variety** and business value of data.
- The increasing tension between the various **regulatory requirements for data security and privacy** and the business requirements for agility, flexibility, sharing, deep insight and partnering with other organizations.
- The need to automate and **scale data processing using artificial intelligence (AI) and machine learning (ML)** techniques.
- The decentralization of storage into an ecosystem of distinct and unrelated data silos, where traditional system-centric approaches to data security lose their efficacy.
- **The increasing number and complexity of privacy regulations and laws requiring mandatory compliance,** based on the Organization for Economic Co-operation and Development's eight privacy principles. Examples include the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA), CLOUD ACT
- An **upward trend in the number and scale of data breaches, throughout 2018 and 2019**, that may have been subject to regulatory response, such as punitive fines.

2019 Cost of a Data Breach Report

Data breaches can cause devastating financial losses and affect an organization's reputation for years.

From lost business to regulatory fines and remediation costs, data breaches have far reaching consequences.
The annual Cost of a Data Breach Report, **conducted by the Ponemon Institute and sponsored by IBM Security, analyzes data breach costs reported by 507 organizations across 16 geographies and 17 industries**.
Read the report to discover all the factors that influence the cost of a data breach and which security measures can help organizations reduce the financial impact.

USD 3.92 million
Average total cost of a data breach

United States
Most expensive country: **USD 8.19 million**

Sensitive Data is a Risk

70%

Of organizations surveyed use live customer data in non-production environments (Development, testing, Q/A)

Database Trends and Applications. Ensuring Protection for Sensitive Test Data

25,575 records
Average size of a data breach

52%

of surveyed organizations outsource development

50%

Of organizations surveyed have no way of knowing if data used in test was compromised

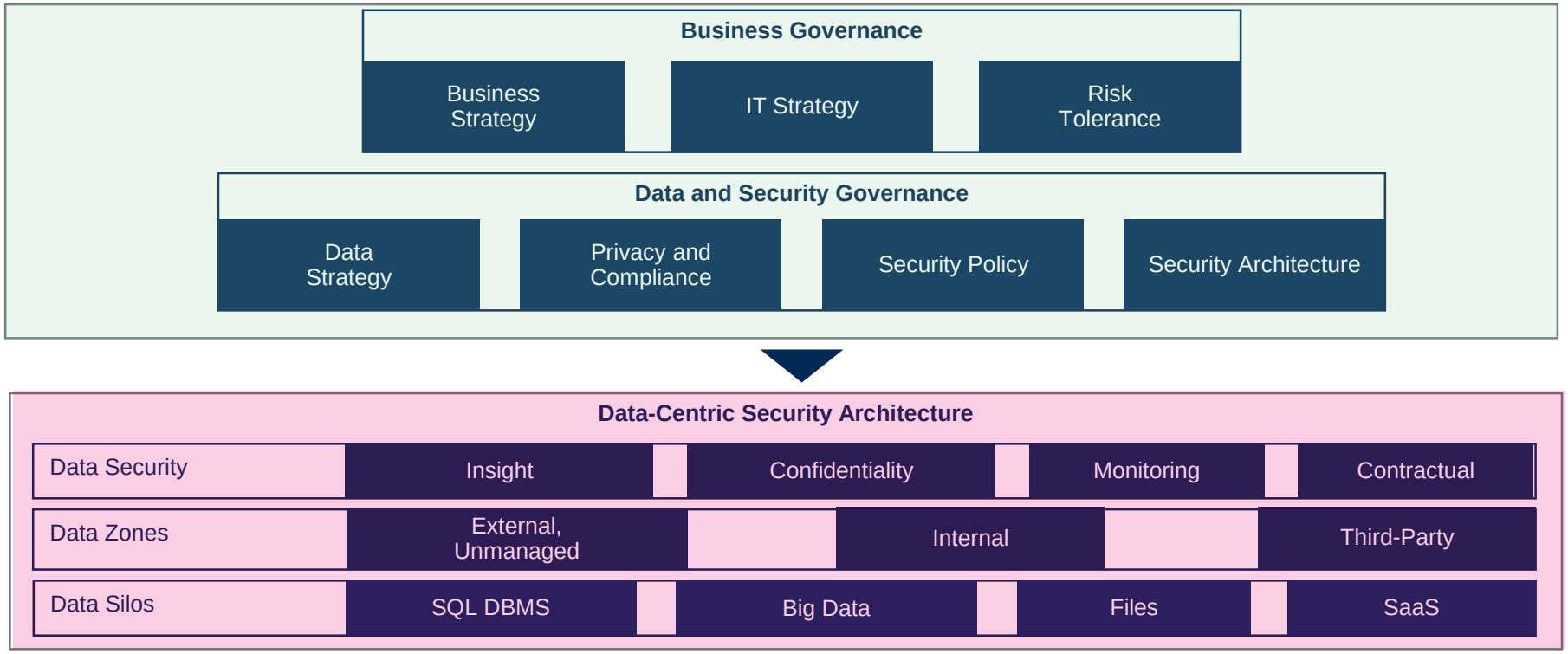
life.augmented

The Four Data-Centric Control Families, With Examples of Commonly Implemented Controls

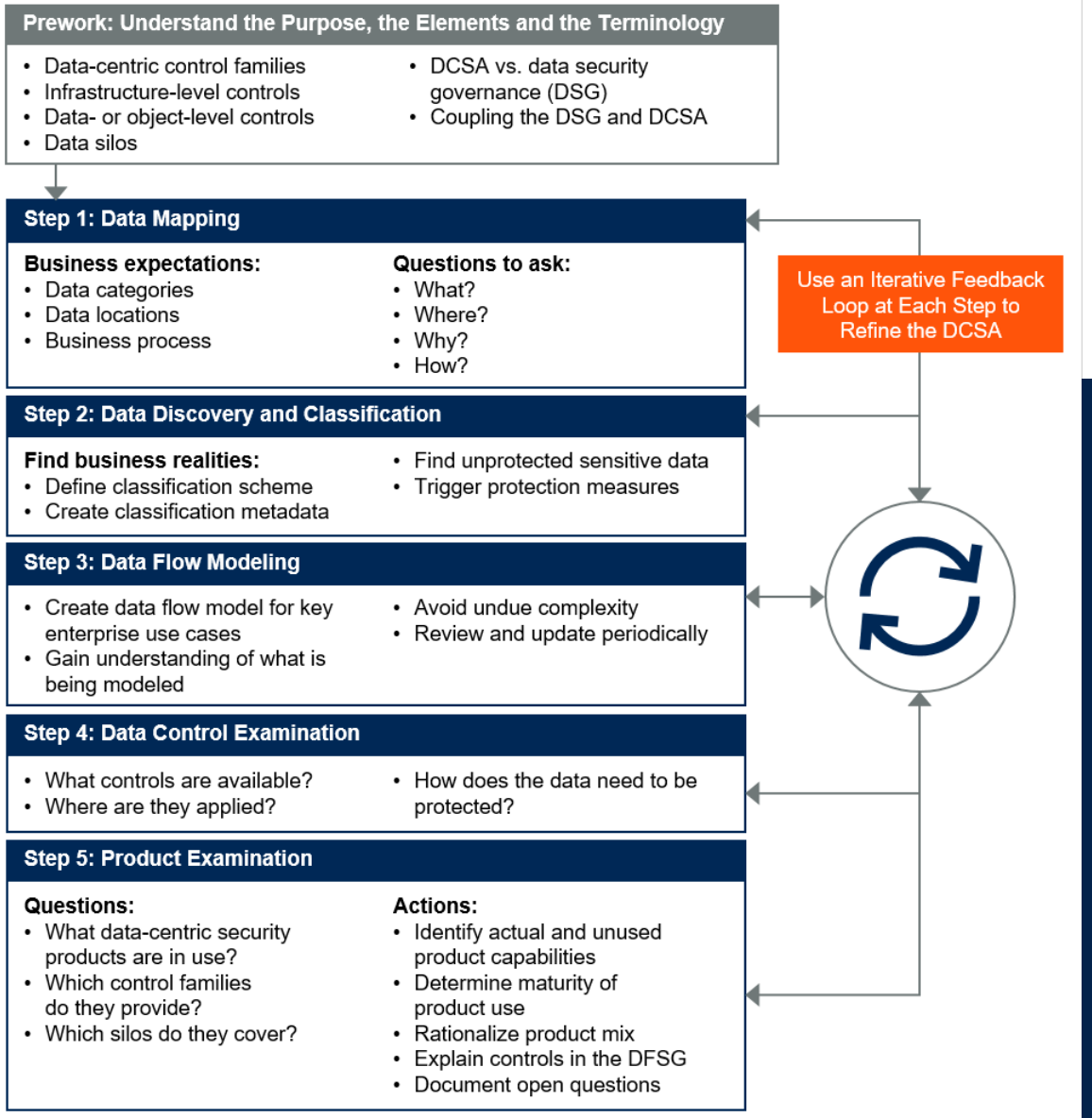
DCSA Control Families	Control Examples
Insight	Data Mapping
	Data Discovery
	Data Classification
Confidentiality	Access Control
	Data Masking and Encryption
	Enterprise Digital Rights Management (EDRM)
	Data Loss Prevention (DLP)
Monitoring and Response	Security Information and Event Management
	Database Activity Monitoring
Third-Party Governance	Contractual Controls

life.augmented

Coupling Data (Security) Governance With the Data Centric Security Architecture



How to Successfully Design and Implement a **D**ata-**C**entric **S**ecurity **A**rchitecture **(DCSA)**



Data discovery and classification. Find the data in more detail, record what is found, and codify its level of sensitivity and criticality to the organization. Assign owners to the data.

Describe how and why data moves through the system.

Examine data flows for typical data use cases. What DCSA controls are applied to data in each silo, and are they sufficient for both present operations and any known future plans?

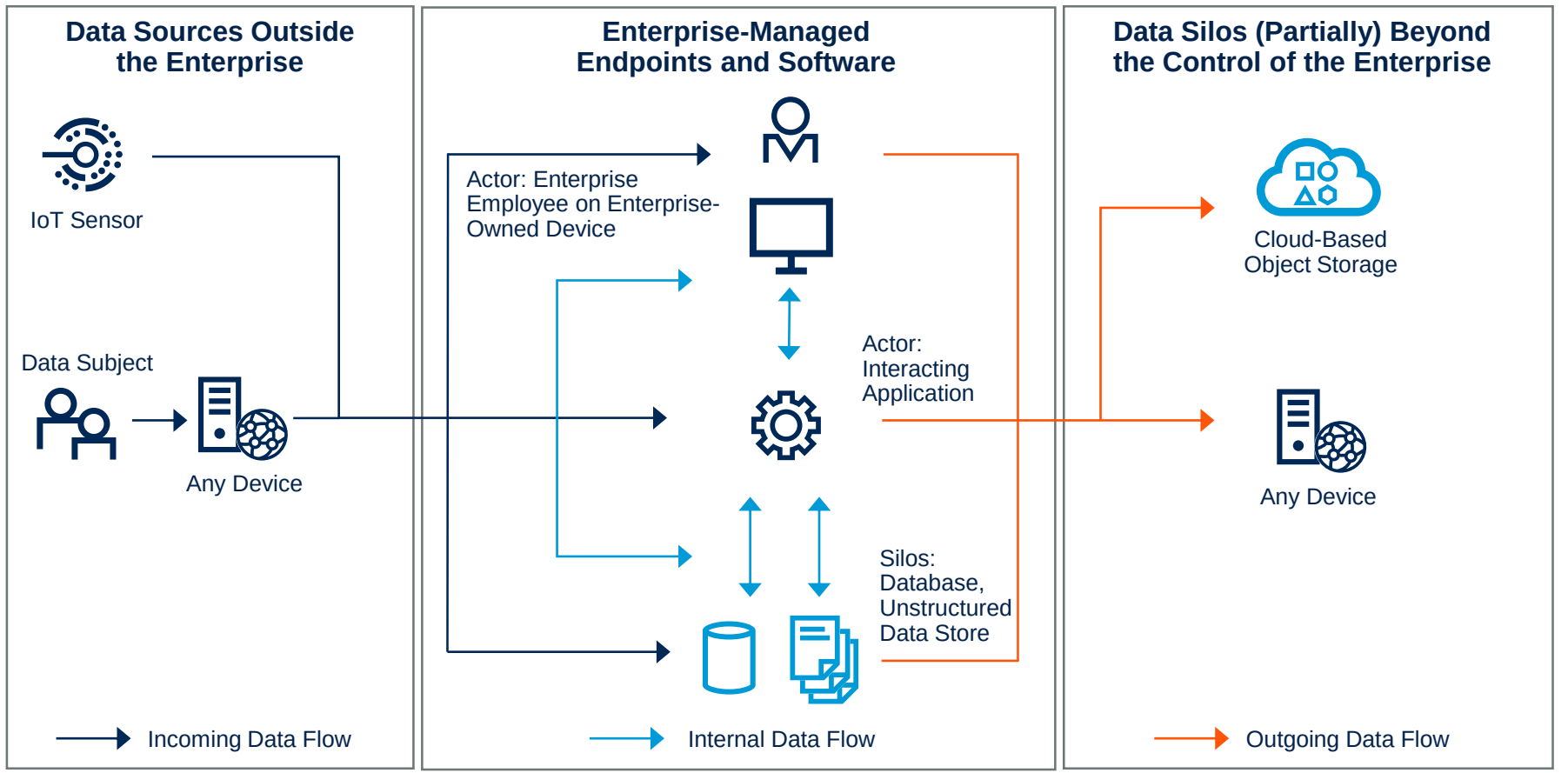
Examine the security products you have in place:
•Which of the four DCSA control families do they address, and for what silo?
•What products can easily be used across more silos?
•What critical gaps must be alleviated by acquiring new technologies and products?

Data Security, Data Compliance, Data Risk

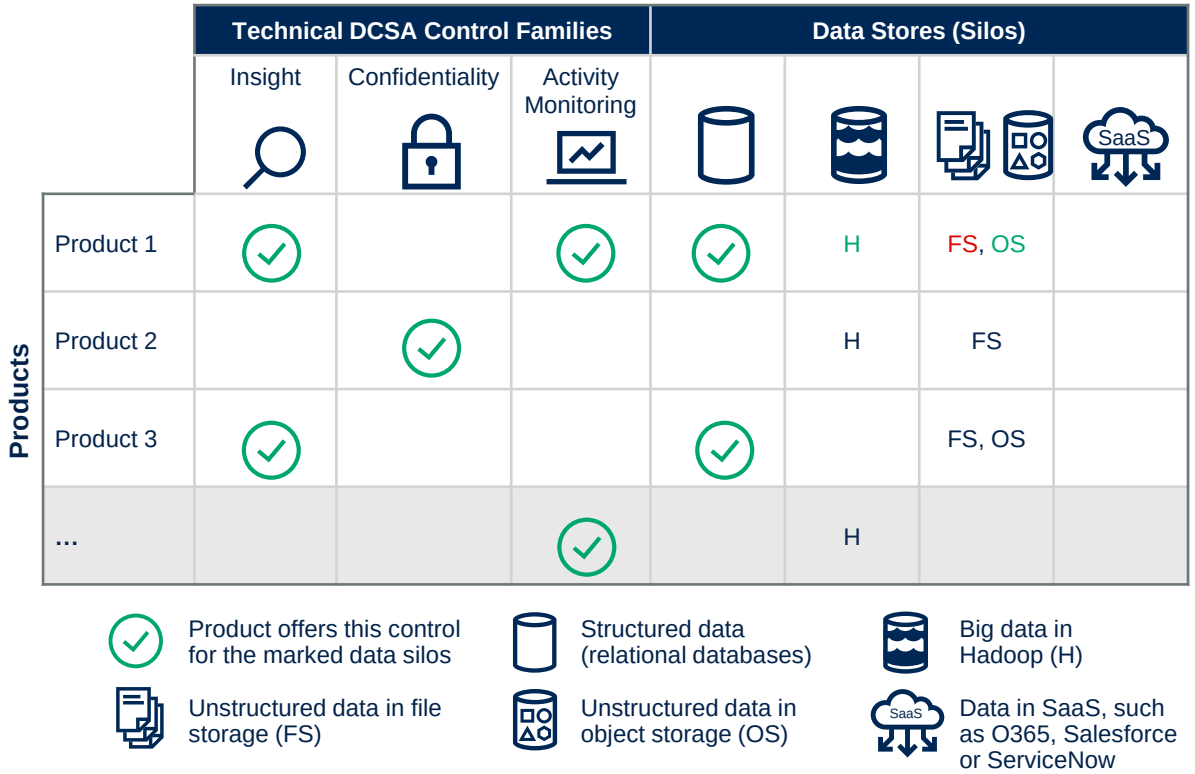
DO NOT FORGET !!!

Data Security	Data Compliance	Data Risk
• Access control • Encryption • Data Integrity • Data Leakage Prevention	• Data Access Policy • Data Retention • Forensics	• Preparing for Potential issues • Data loss • Data Inaccessibility (service outage) • Data exposure

Zones, Actors, Silos and Dataflows for Unstructured Data



Product Capability Matrix



life.augmented

Data Integration Architecture

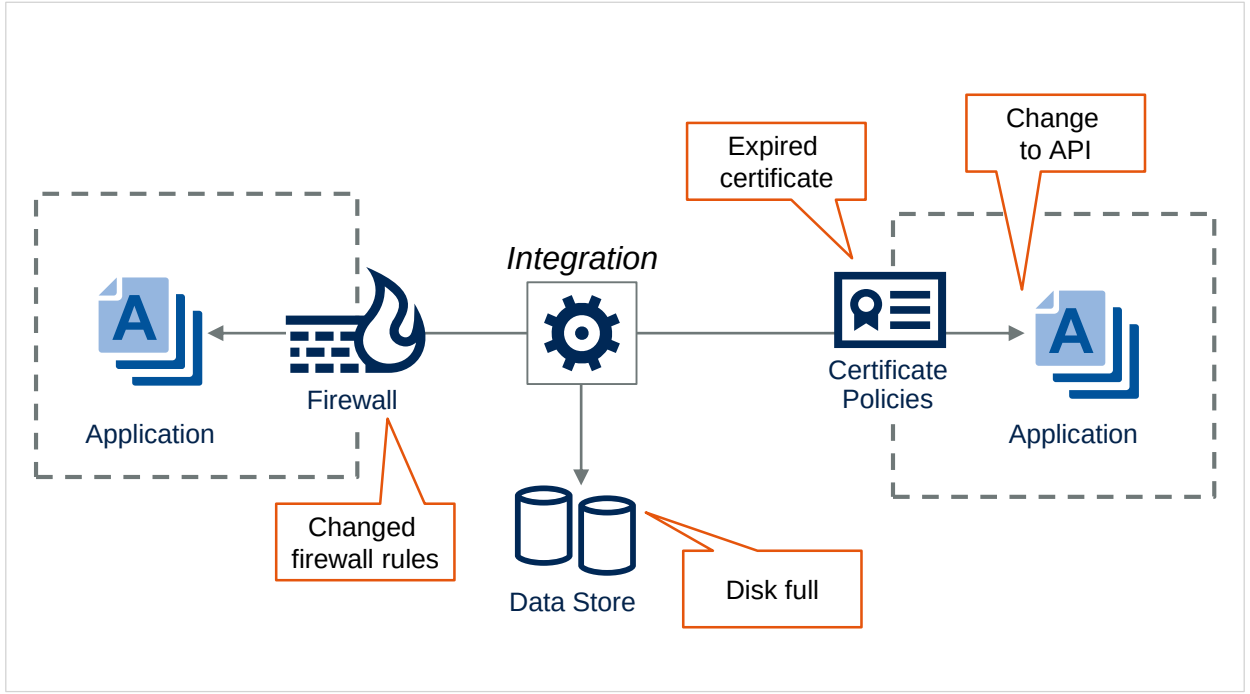
life.augmented

How to guarantee a secure Data Integration Architecture

8 assessment criteria to guarantee data integration

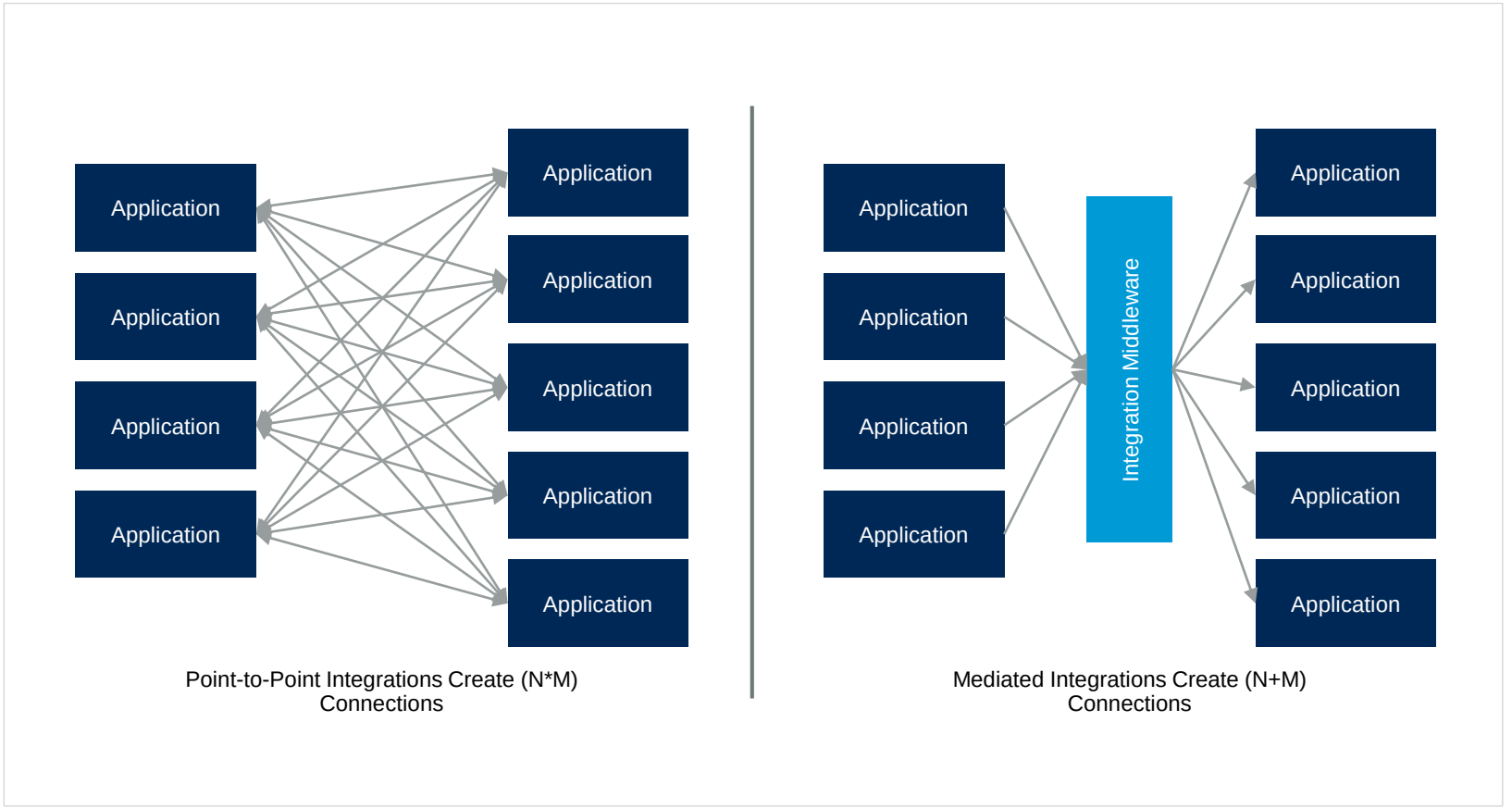
- **Ability to adapt** — How much of a solution can still be used if (or when) one or more of the applications being integrated need to be replaced, such as when you replace an on-premises application with a SaaS application?
- **Ability to extend** — How easy it is to make changes in the integration when required, such as adding fields to an integration between an ERP and an e-commerce application?
- **Ability to monitor** — How much visibility does this solution provide into integration uptime, performance and quality, such as integration with existing monitoring and alerting systems or built-in monitoring capabilities?
- **Ability to reuse** — How easy it is to reuse this solution for additional integrations, including reuse of artifacts such as business logic or data mapping? For example, can you reuse a product data integration between an ERP and an e-commerce system to share that data with search engines?
- **Security and data governance** — How much work is required to secure the solution, and how much control does it afford over sensitive or proprietary data? For example, does the integration functionality run on hardware you control?
- **Support for ad hoc integrators** — How easy it is for staff members who are not integration specialists to implement or update integrations? For example, some developers may need to create integrations in order to deliver a project.
- **Support for mediation** — How well does the solution support connecting endpoints with disparate styles, such as connecting an API to a messaging system, or converting XML to JSON?
- **Support for reliable communications** — How well does the solution deal with intermittent or unreliable connections between endpoints, such as with equipment in the field or mobile devices?

Some Possible Reasons for Integrations to Break



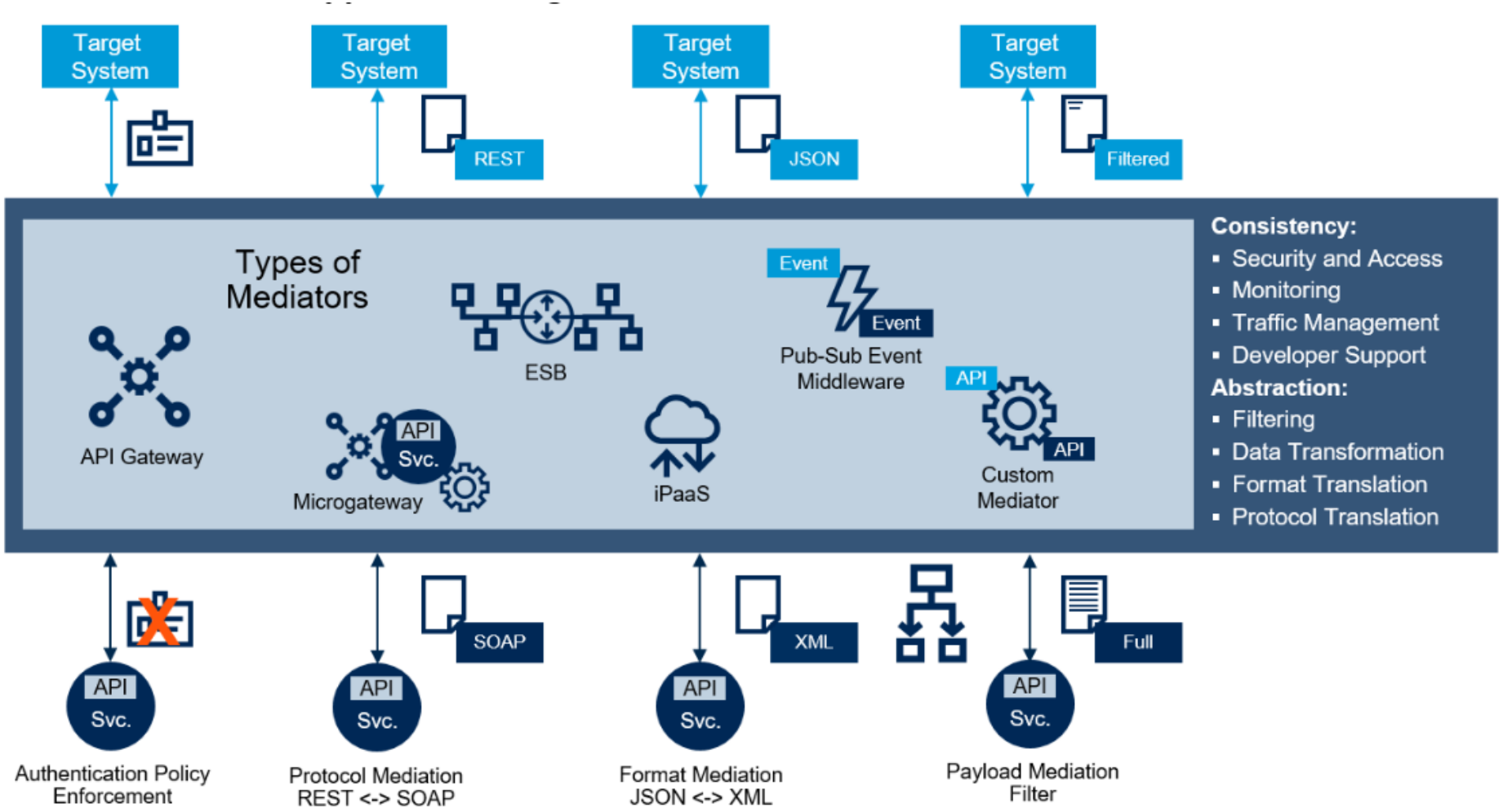
life.augmented

Point-to-Point Integrations Scale Poorly



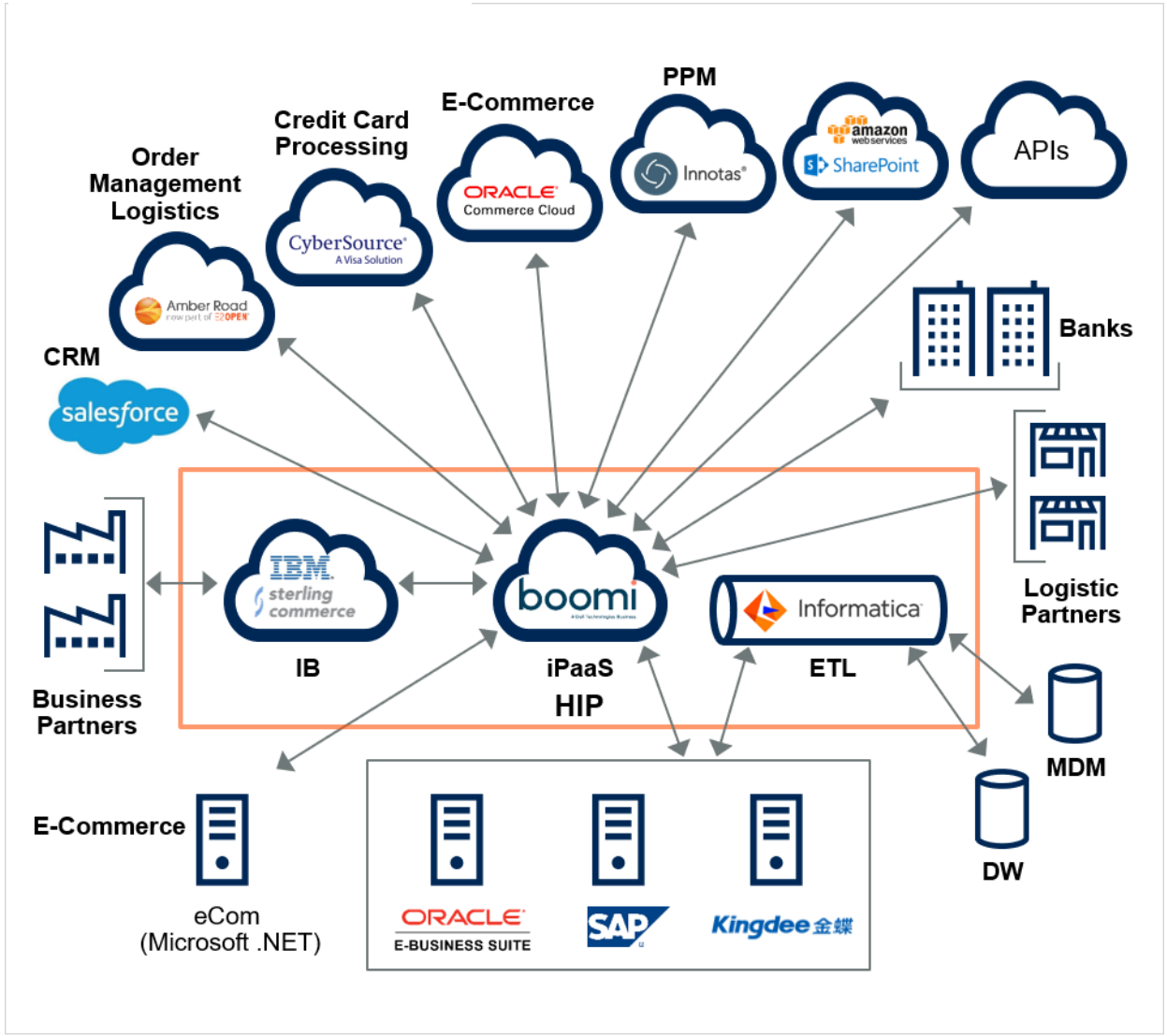
life.augmented

Mediation Enables Integration and Guarantee proper Governance and Security



life.augmented

Example of Hybrid Integration Platform



Data Classification

Data Classification Definition

Data classification is "the process of organizing information assets with an agreed-on categorization glossary, enabling effective and efficient prioritization for information governance policy spanning **quality, security, access, privacy, storage and retention.“**

The Importance of data visibility and Classification

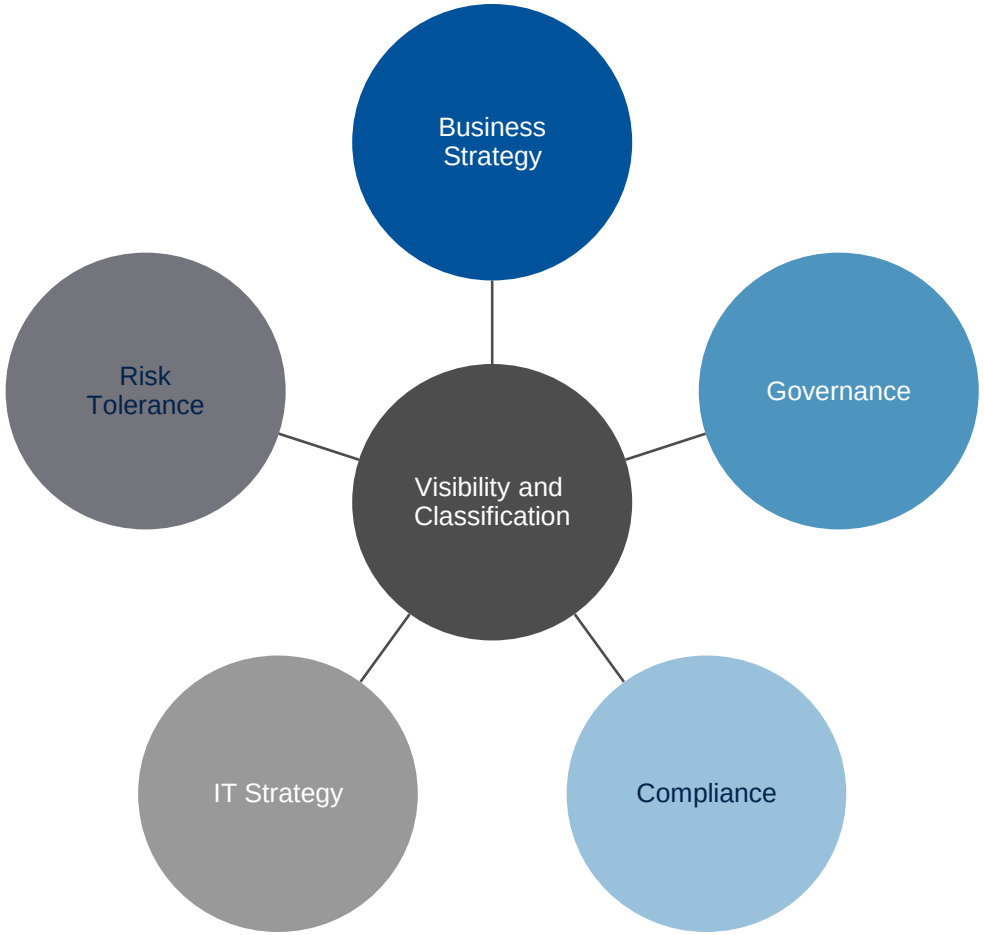
Organizations increasingly understand the need to better manage their data in order to extract value and justify the costs of security, collection and storage.

Regulations such as the **European General Data Protection Regulation (GDPR)** are requiring:

- **Robust and risk-based data governance and security controls**.
- **Driving a need for better visibility of data**
  Some data may be so sensitive so to be considered - **CLASSIFIED** as secret,

Being able to find, analyze and classify data is increasingly important to drive business value, support governance and achieve compliance.

The figure illustrates the business functions that benefit **from good data insight, including visibility and classification.**



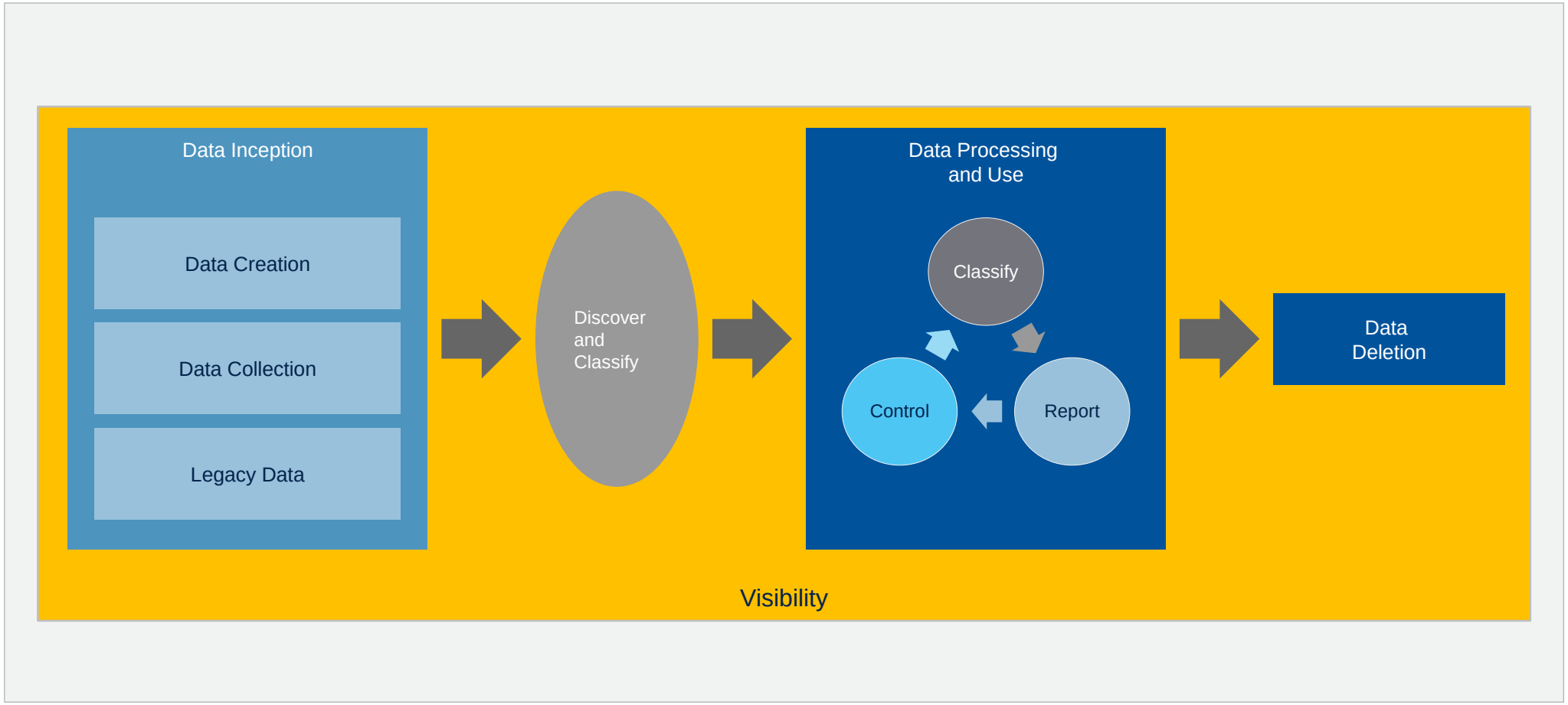
life.augmented

Common Drivers for Data Governance and Data Classification

Source	Example	Applicable Data
Privacy regulation	European GDPR	▪ Processing of Personal data and rules relating to the free movement of personal data for data subjects in the EU
Cybersecurity regulation	New York State Department of Financial Services 23 NYCRR 500: Cybersecurity Requirements for Financial Services Companies	▪ Consumers' private data
Health and insurance	Health Insurance Portability and Accountability Act (HIPAA)	▪ Protected health information (PHI)
Contract	IT or business process outsourcing. Cloud Service Provision (SaaS or infrastructure as a service [IaaS])	▪ Any data shared or accessible as a result of the contract agreement
Financially sensitive information	Merger and acquisition data Corporate performance	▪ Target acquisition information, including the fact of the existence of the intent ▪ Financial performance report prior to release to stock market
Intellectual property	Engineering or research records Process documentation Exploration data Customer data and sales opportunities	▪ Any data deemed by the organization to contribute to the competitive advantage, revenue and profitability of the organization
Security data	Security logs Enterprise architecture (EA)	▪ Information material to the protection of the organization, including the organization's technical infrastructure

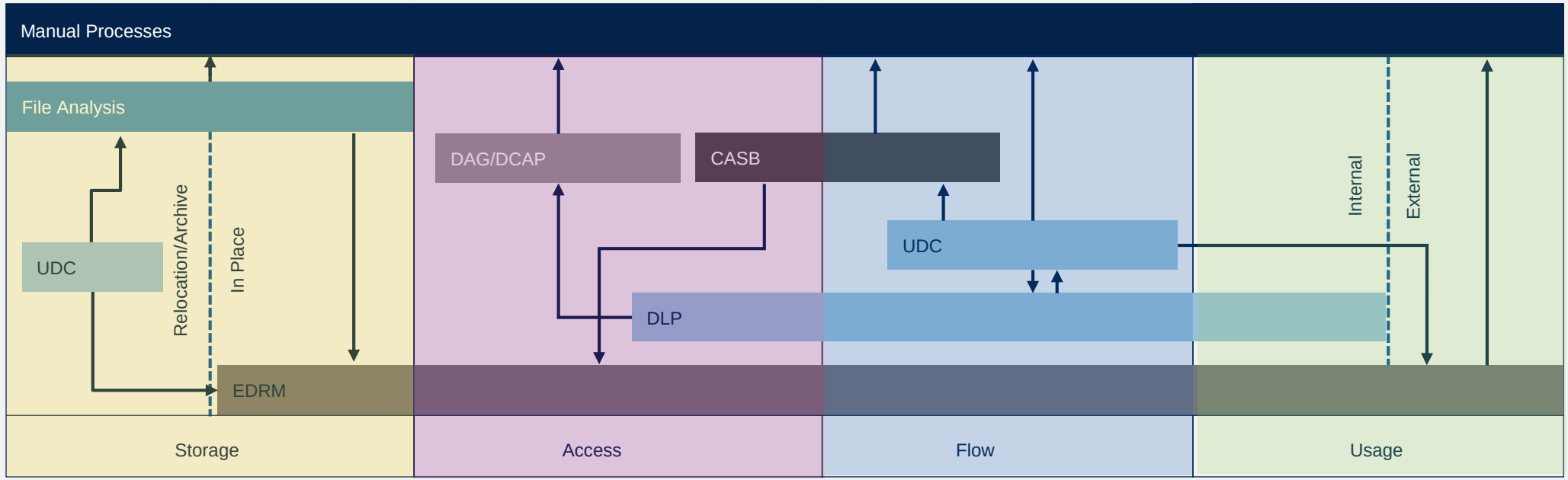
life.augmented

Data classification in the data life cycle



life.augmented

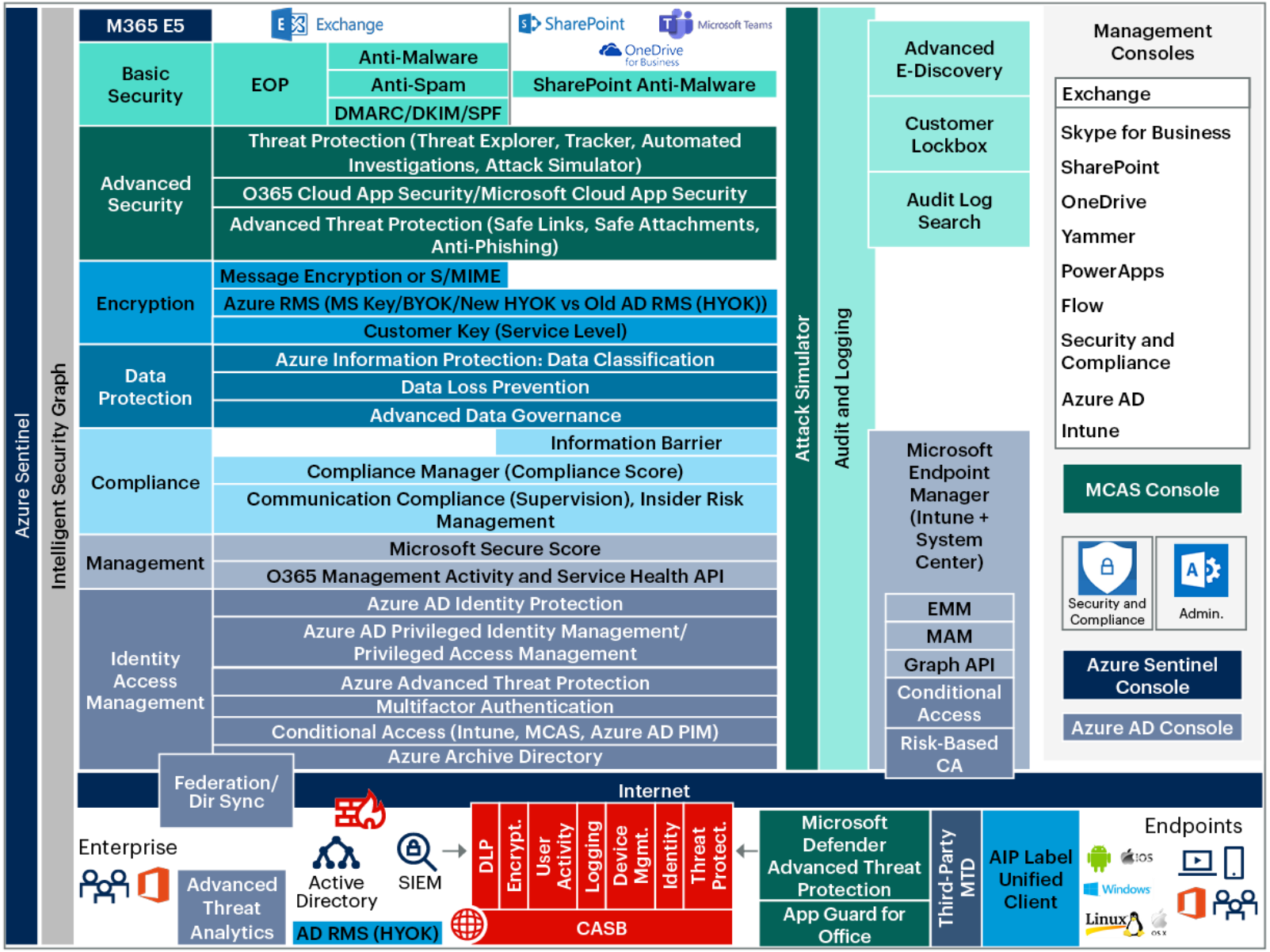
Data classification tools and their relation to the data life cycle



CASB = Cloud Access Security Broker; DAG = Data Access Governance; DCAP = Data Centric Audit and Protection; EDRM = Enterprise Digital Rights Management; FA = File Analysis; DLP = Integrated Data Loss Prevention; UDC = User Driven Classification

life.augmented

O365 Data Security Native Tools



life.augmented

Five categories of requirements for classification tools

Repositories	▪ Where can the tool detect sensitive data? ▪ Does the repository type change the capabilities of the tool? ▪ Can the system inspect file objects embedded in a database?
Content Detection	▪ What technologies are used to identify sensitive data? ▪ What "out of the box" policies does the tool provide? ▪ Does the tool support context-based classification, such as user or file location? ▪ What language limitations does the tool have?
Control	▪ Is the tool able to impose controls on the data depending on classification without other toolsets being required?
Recording	▪ Does the tool provide any permanent record of the classification outcome, or is the metadata only used to implement a control at that time? ▪ If metadata is persistent, what method is used?
Interoperation	▪ What capability does the tool have to share or consume metadata from other toolsets? ▪ What capability does the tool have to "trigger" other controls or actions?

Example data repositories

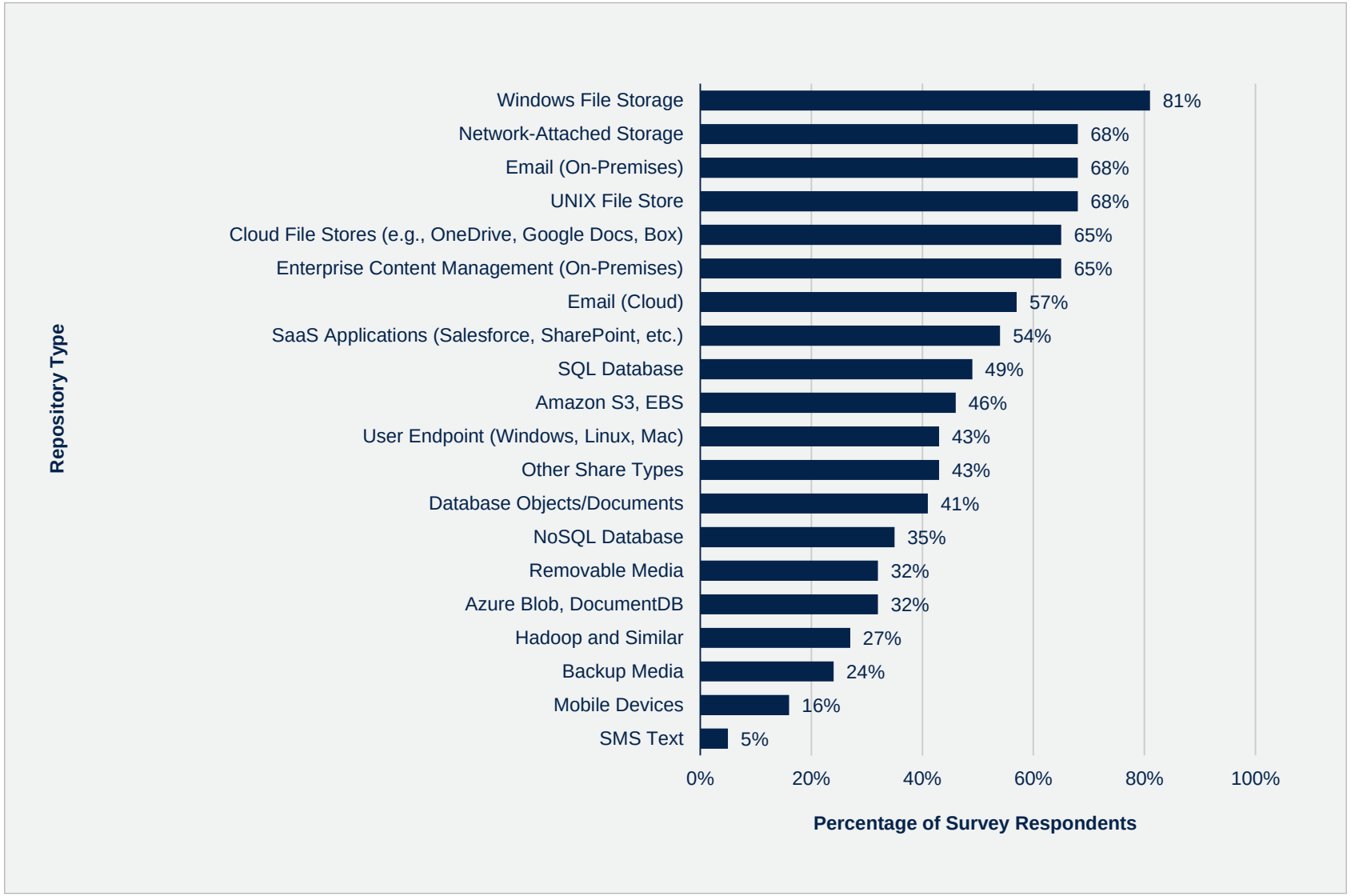
Data Scenarios Defined

Repository Types Include:

On-Premises	Cloud
▪ Windows (CIFS, SMB) File Share ▪ Server Local Storage ▪ Network-Attached Storage ▪ UNIX File Stores ▪ Sharepoint ▪ Email Server ▪ Databases (SQL) ▪ Database (NoSQL) ▪ Objects and Documents Stored Within Databases ▪ Hadoop and Other HDFS Storage ▪ Enterprise Content Management (On-Premise) ▪ Backups ▪ User Endpoint (Laptops, Desktops; Including BYOD) ▪ User Mobile Devices (Including BYOD) ▪ Removable Media	▪ IaaS and PaaS Such As: – Amazon S3, EBS – Azure Blob, DocumentDB ▪ DBaaS, PaaS and Other Middleware Solutions ▪ Cloud Sharing Such As: – O365, Teams, OneDrive, GoogleDocs, Box ▪ Cloud SaaS (SalesForce, Concur, Docusign, etc.) ▪ Cloud Email Services Such As Exchange Online

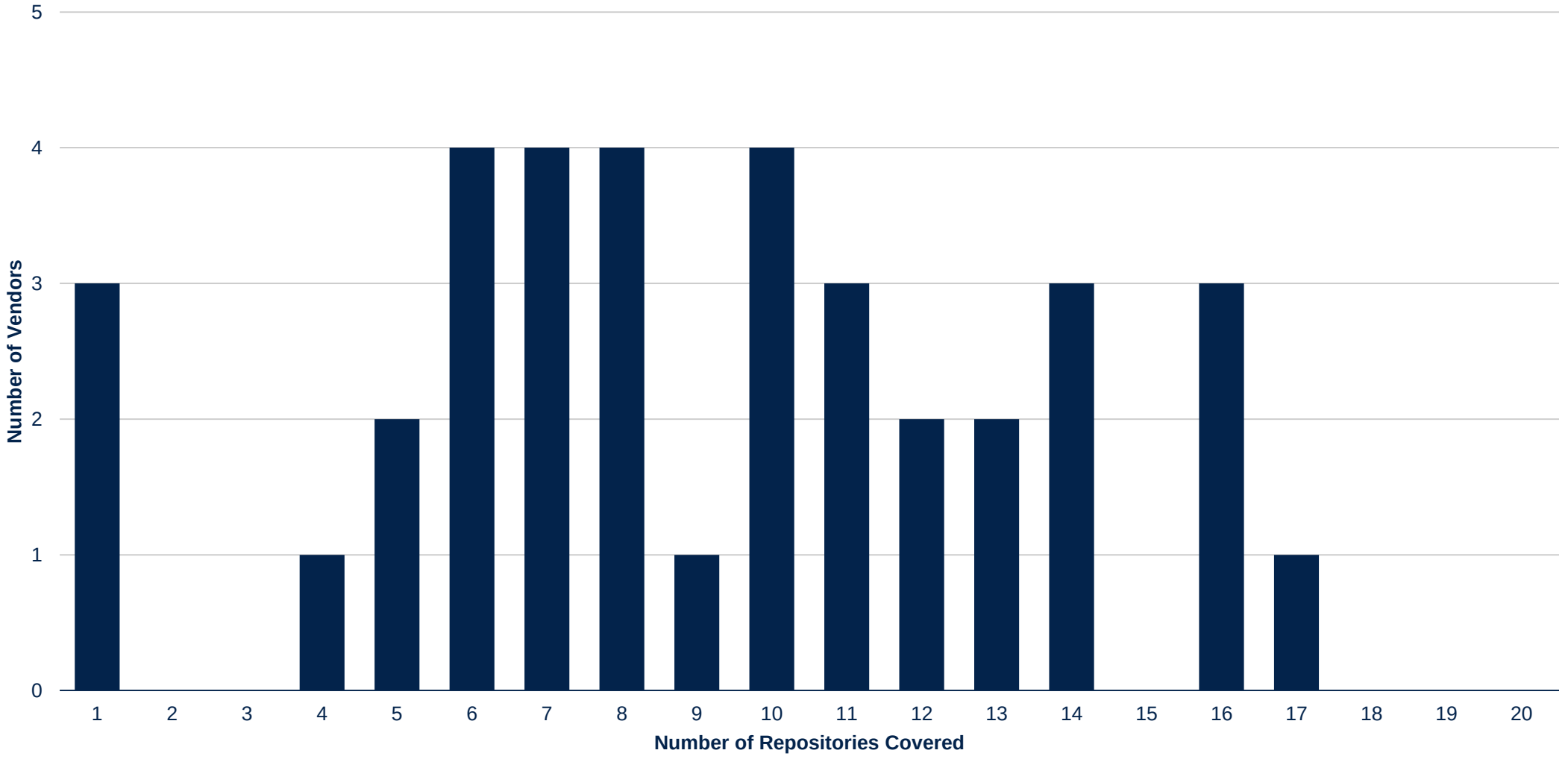
File Types Include:

Distribution of repositories covered by classification tools



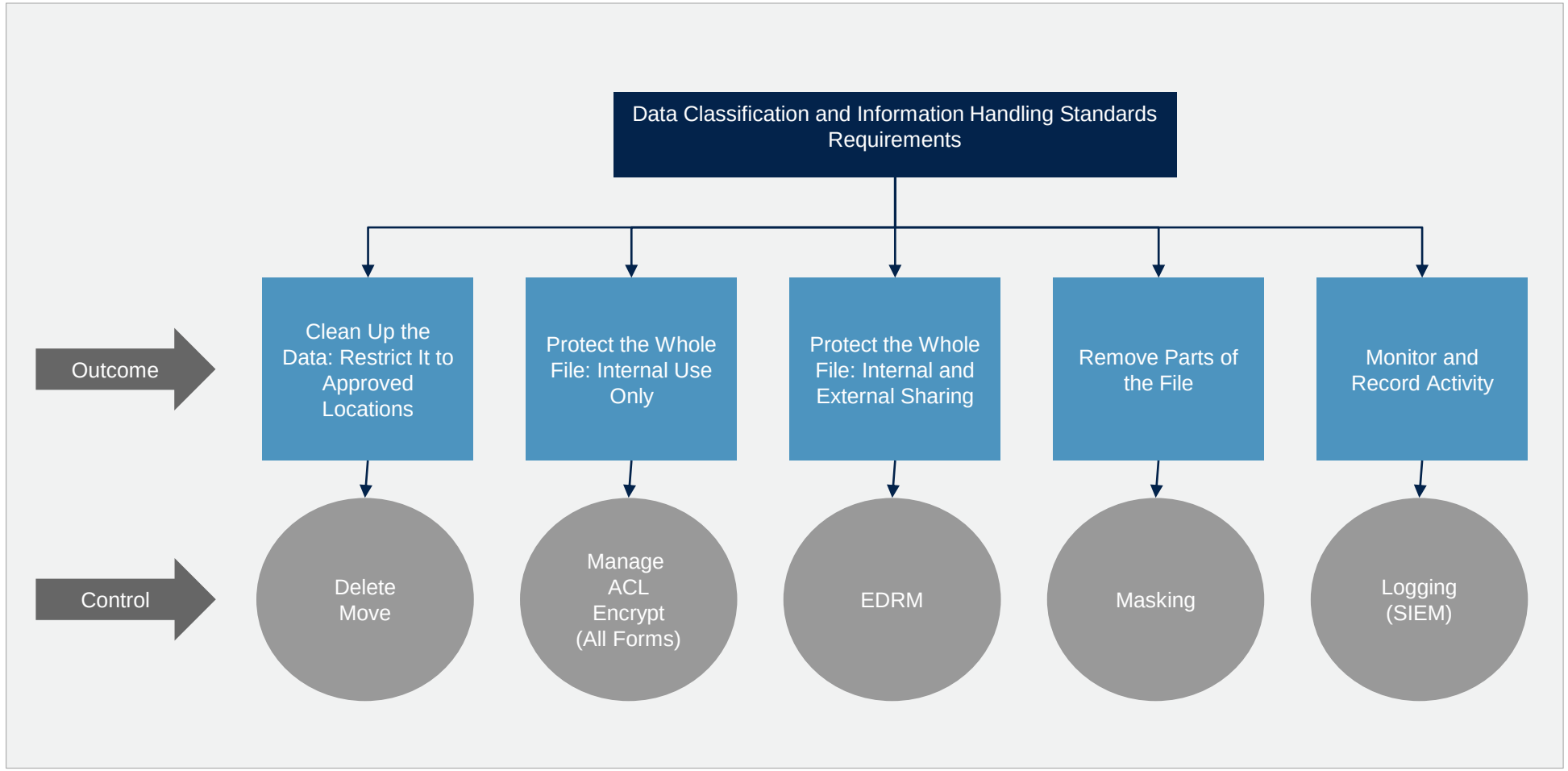
life.augmented

Vendors do not have universal coverage of repositories



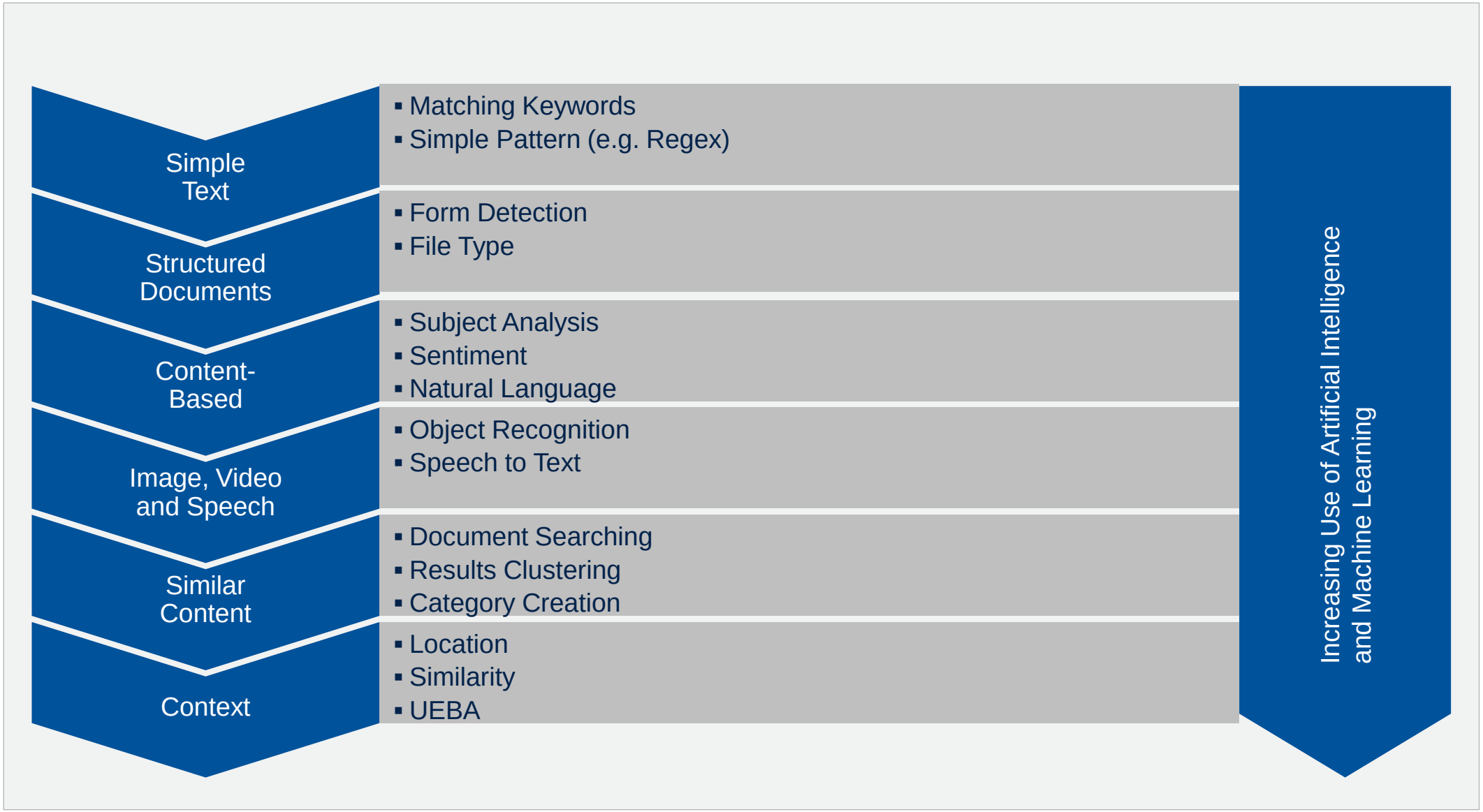
life.augmented

Security outcomes and the control choices

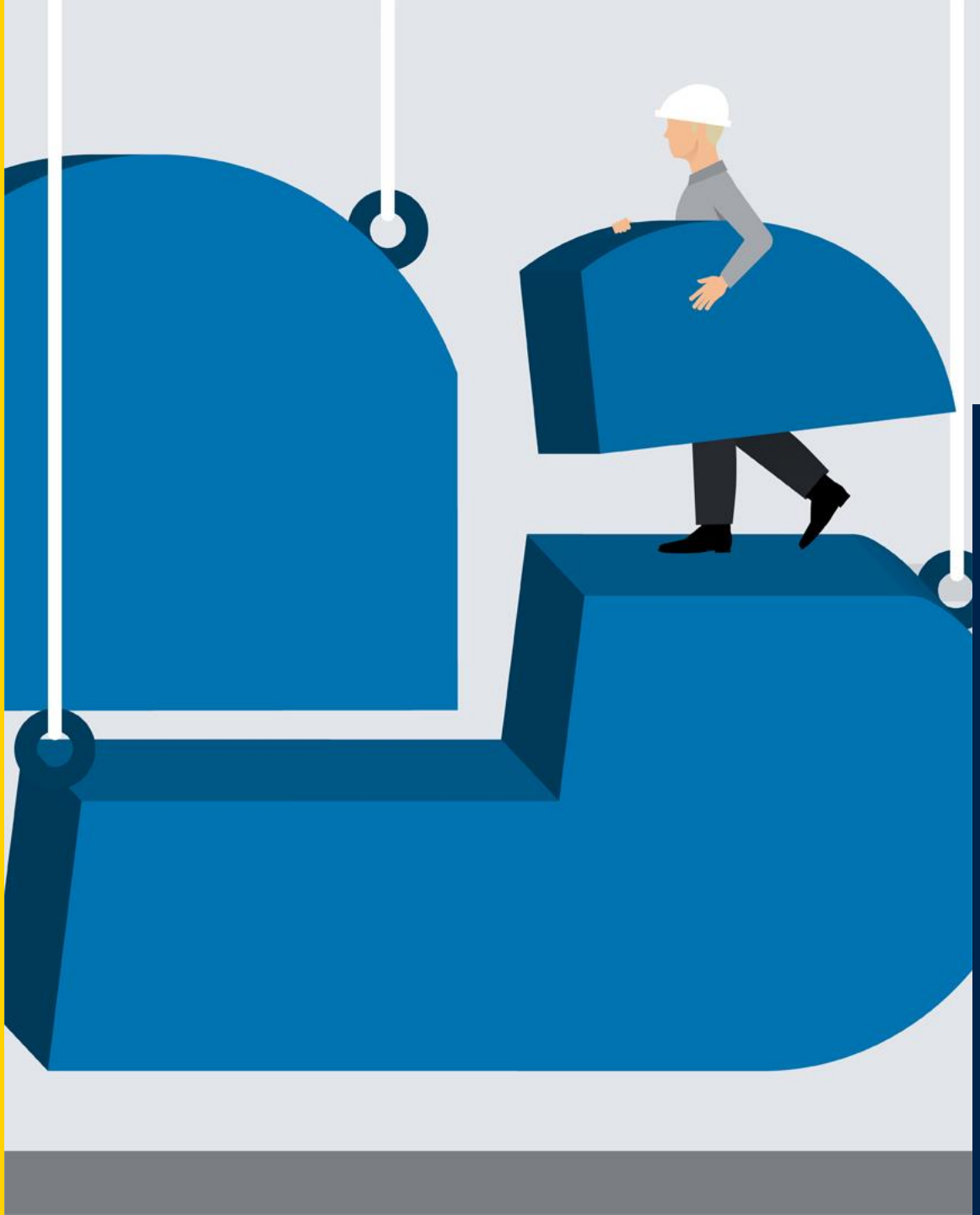


life.augmented

How machine learning is increasingly involved in data classification



life.augmented



life.augmented

Design Patterns for Cloud Architecture

Giordano Scuderi

ST Microelectronics – Digital Fab

Agenda

1. Introduction
2. Cloud Design Patterns
3. First architecture example
4. Second architecture example
5. Second architecture improvement using iPaaS
6. Q&A

life.augmented

Introduction

life.augmented

Architecture styles

Architecture style	Dependency management	Domain type
N-tier	Divide application into logical layers	Traditional business domain. Frequency of updates is low.
Web-Queue-Worker	Front and backend, decoupled by async messaging.	Relatively simple domain with some resource intensive tasks.
Microservices	Vertically (functionally) decomposed services that call each other through APIs.	Complicated domain. Frequent updates.
Event-driven	Producer/consumer. Independent view per sub-system.	IoT and real-time systems
Big data	Divide a huge dataset into small chunks. Parallel processing on local datasets.	Batch and real-time data analysis. Predictive analysis using ML.
Big compute	Data allocation to thousands of cores.	Compute intensive domains such as simulation.

Focus area in cloud architecture


Availability


Data Management


Design and Implementation


Messaging


Management and Monitoring


Performance and Scalability


Resiliency


Security

life.augmented

Focus area in cloud architecture

AVAILABILITY

- Time where system is functional and working, measured as a percentage of the uptime.
- **Applications must be designed to maximize availability.**

DESIGN AND IMPLEMENTATION

- Decisions made during the design and implementation phase have a huge impact on the quality and the total cost of ownership of cloud hosted applications and services

DATA MANAGEMENT

- Data is typically hosted in different locations and across multiple servers
  - data consistency must be maintained, data will typically need to be synchronized across different locations.
- key element of cloud applications, influences most of the quality attributes.

MESSAGING

- Applications requires a messaging infrastructure that connects the components and services, ideally in a loosely coupled manner in order to maximize scalability.
- Challenges: ordering of messages, idempotency, …

Focus area in cloud architecture

MANAGEMENT & MONITORING

- Applications should help exposing information operations team can use to manage and monitor the system
- Applications should support changing business requirements and customization without requiring the application to be completely stopped

RESILIENCY

- Resiliency is the ability of a system to gracefully handle and recover from failures.
- Due to the nature of cloud hosting there is an increased likelihood of both transient and more permanent failures.

PERFORMANCE AND SCALABILITY

- Performance is an indication of the responsiveness of a system to execute any action within a given time interval
- Scalability is the ability of a system to handle increases in load without impact on performance

SECURITY

- Capability of a system to prevent malicious or accidental actions outside of the designed usage, and prevent disclosure or loss of information.
- Applications must be designed and deployed in a way that protects them from malicious attacks, restricts access to only approved users, and protects sensitive data.

Cloud design patterns

- Patterns are a widely used concept in computer science
  - Describe **good solutions to recurring problems in an abstract form**.
  - Can be applied regardless of used technologies

- Let’s see some patterns as example…

- References:
  - https://www.cloudcomputingpatterns.org
  - https://docs.microsoft.com/en-us/azure/architecture/

Cloud Design Patterns

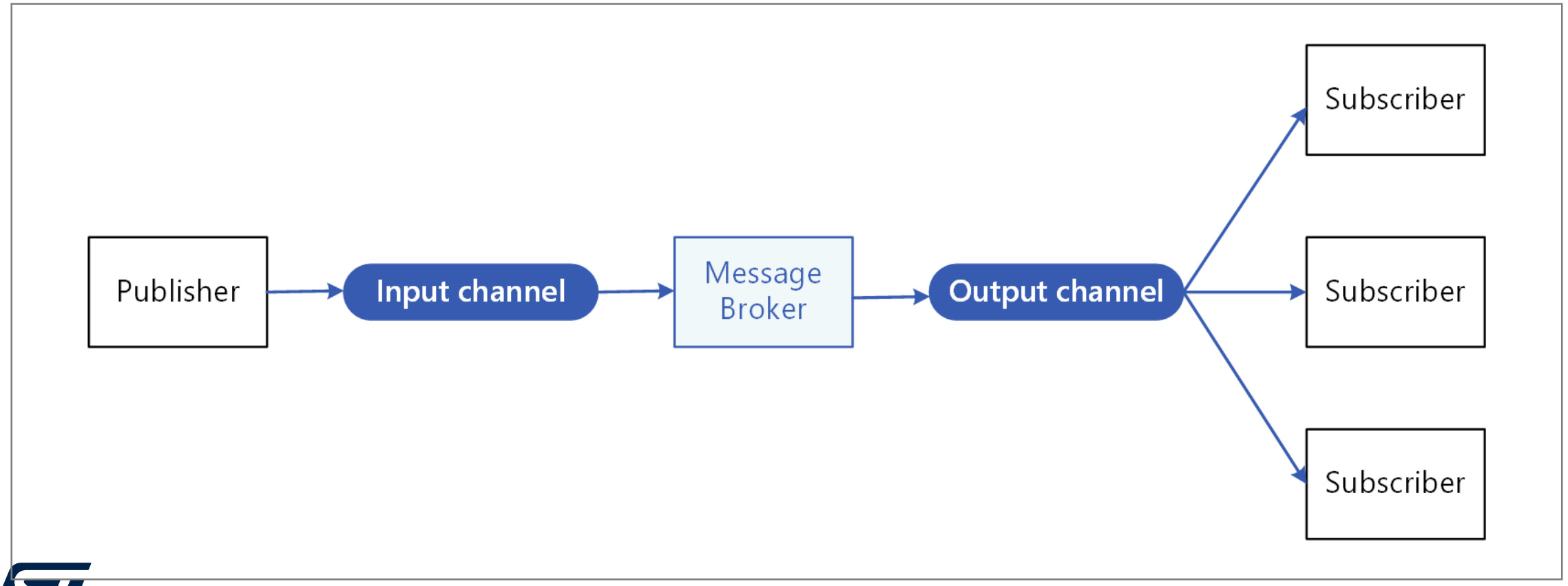
life.augmented

Messaging patterns

The distributed nature of cloud applications requires a messaging infrastructure that connects the components and services, ideally in a loosely coupled manner in order to maximize scalability.

Publisher-Subscriber pattern

- Enable an application to announce events to multiple interested consumers asynchronously, without coupling the senders to the receivers



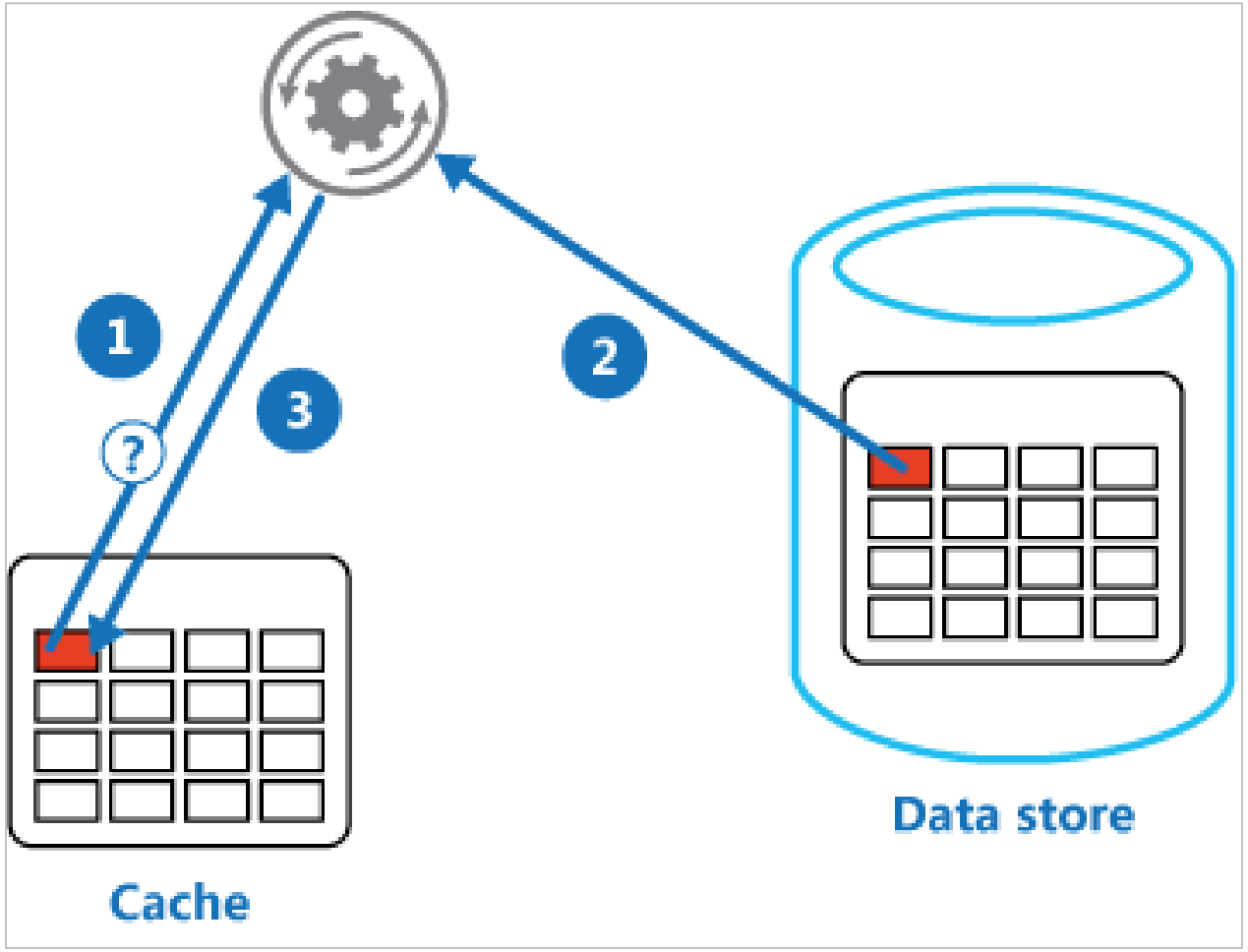
Scalability Patterns

Scalability is the ability of a system to handle increases in load without impact on performance.

life.augmented

Cache-Aside Pattern

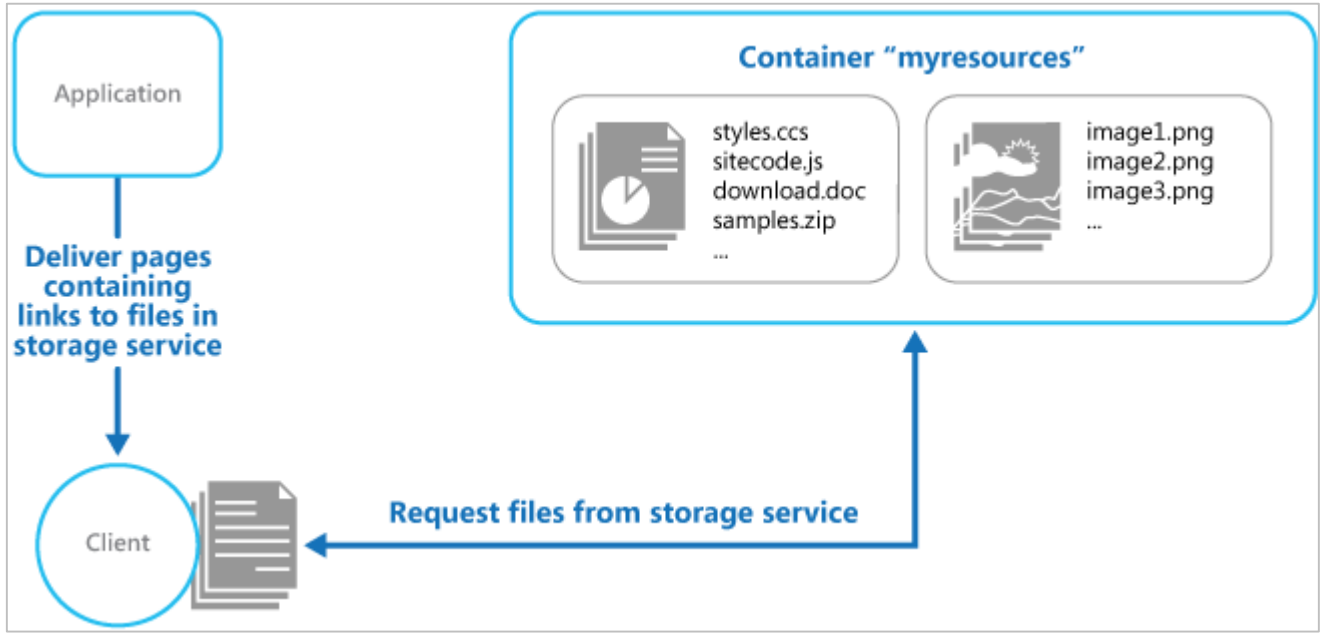
- Load data on demand into a cache from a data store
- Improve performance
- helps to maintain consistency between data held in the cache and data in the underlying data store



life.augmented

Static Content Hosting Pattern

- Minimizes web hosting compute costs
  - Especially so for web sites that consist only of static content
- Improves end-user content performance
  - Serving static content from a CDN (Content Delivery Network) can save on compute and memory utilization of web servers



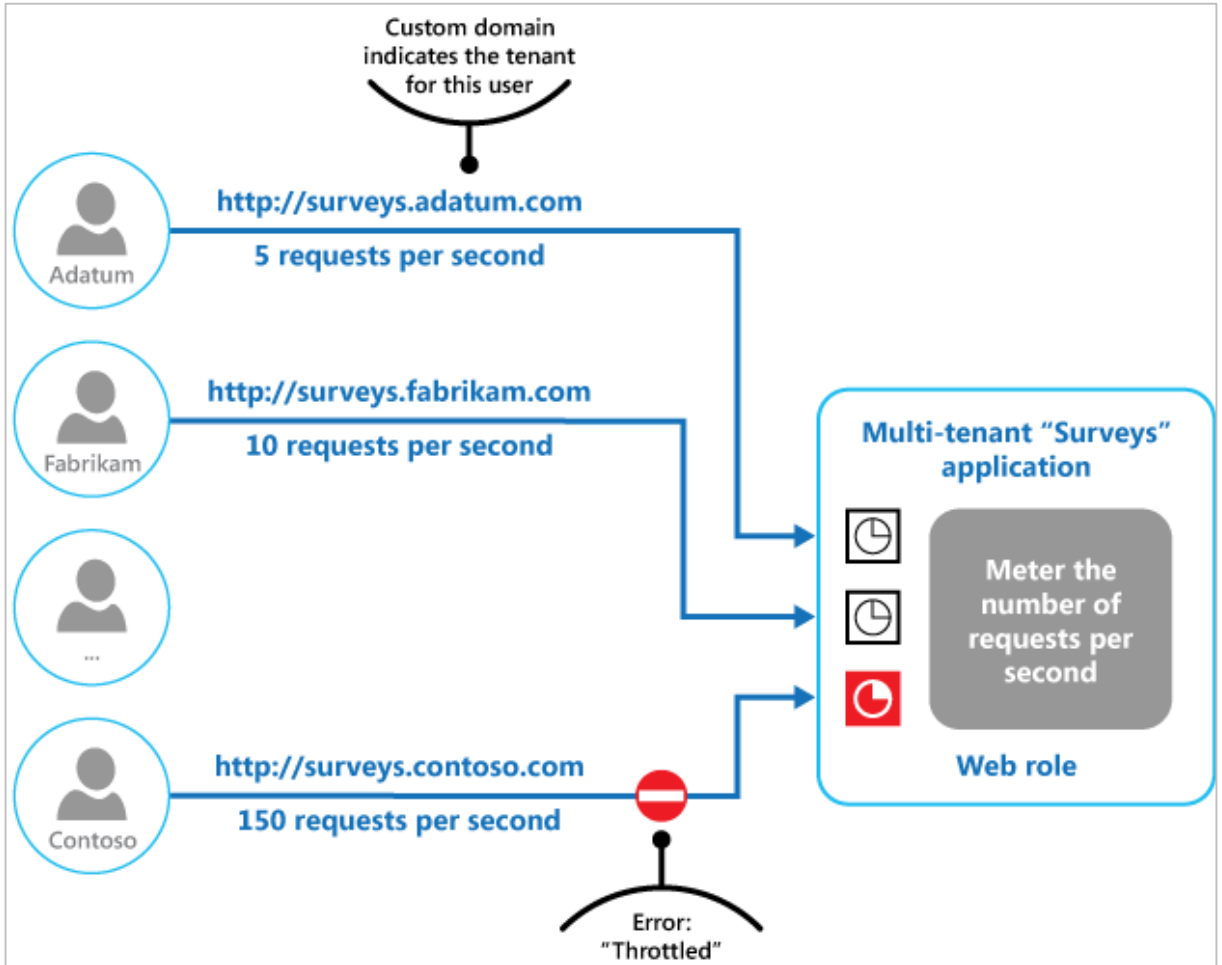
Performance

Performance is an indication of the responsiveness of a system to execute any action within a given time interval.

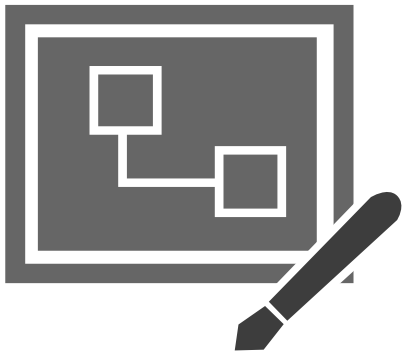
life.augmented

Throttling Pattern

- Control the consumptions of resources used by a single account, or even an entire application which is using the service



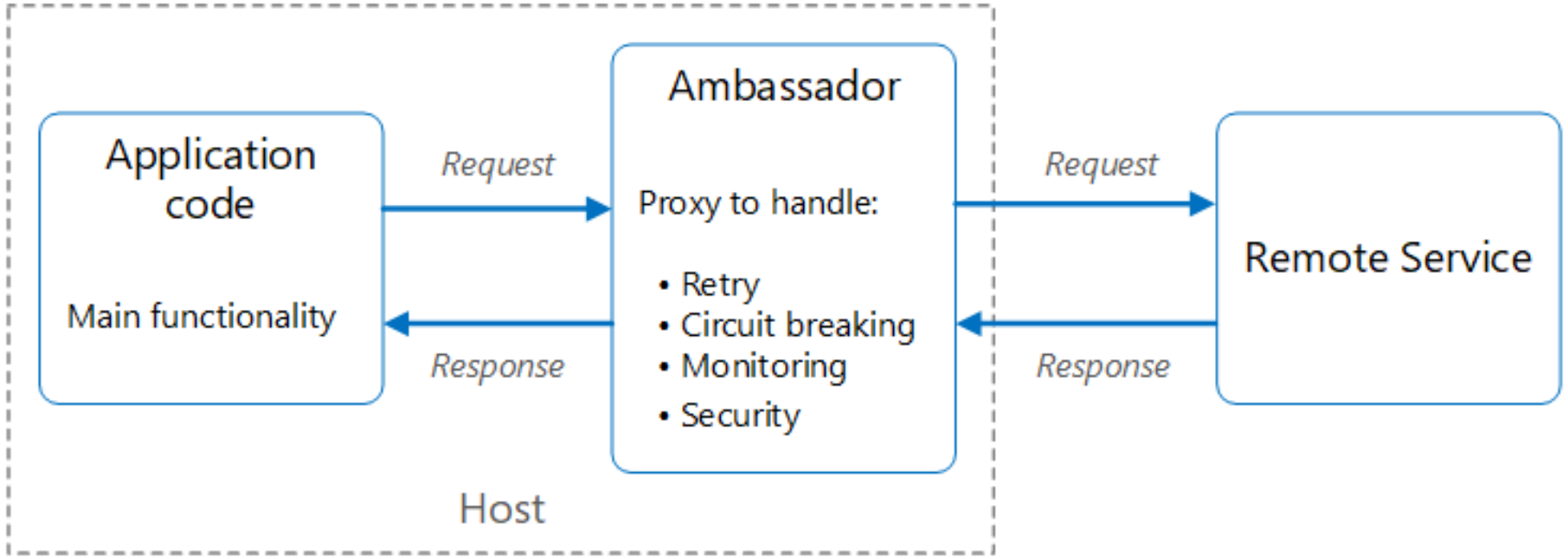
life.augmented



Design and Implementation patterns

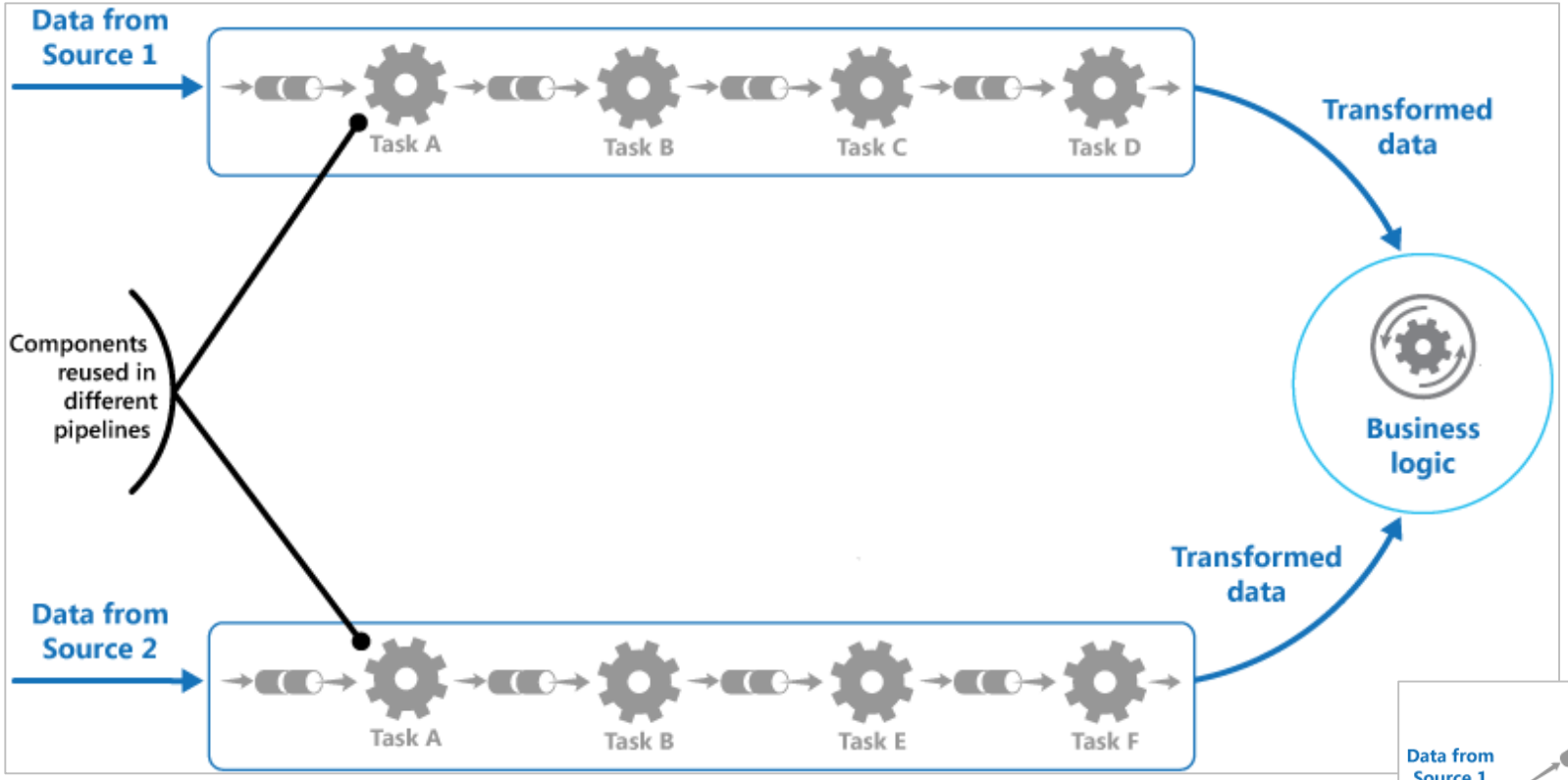
Ambassador pattern

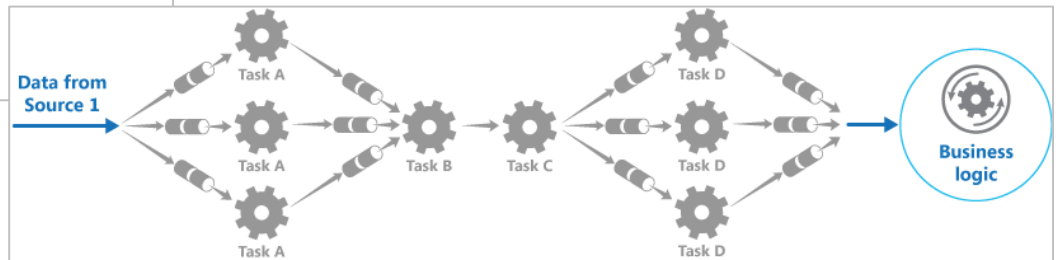
- Useful for off-loading common tasks such as monitoring, logging, routing and resiliency patterns in a language agnostic way.



Pipes and Filters pattern

- Decompoperformancese complex processing tasks into a series of separate elements that can be reused





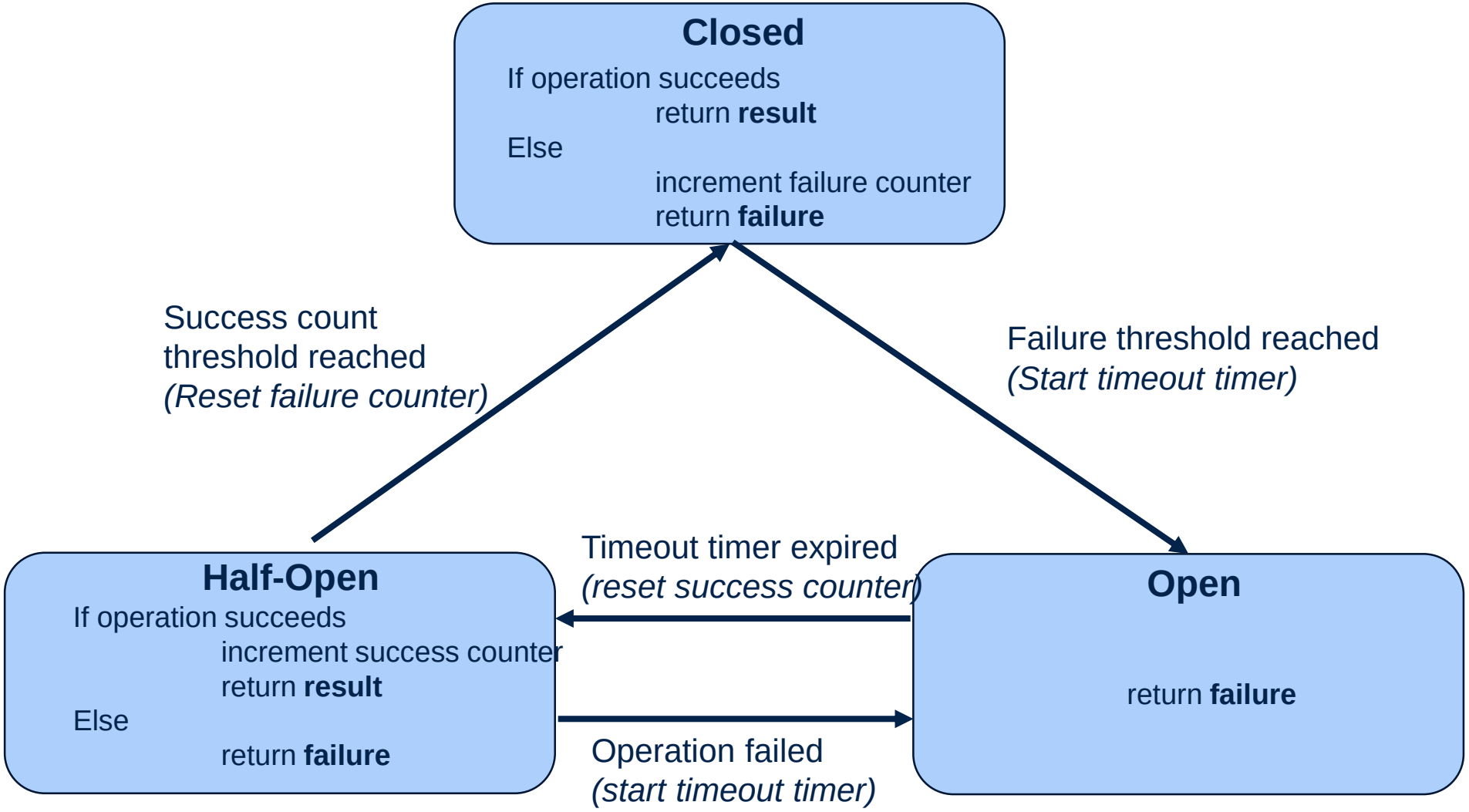
life.augmented

Resiliency Patterns

Resiliency is the ability of a system to gracefully handle and recover from failures.

life.augmented

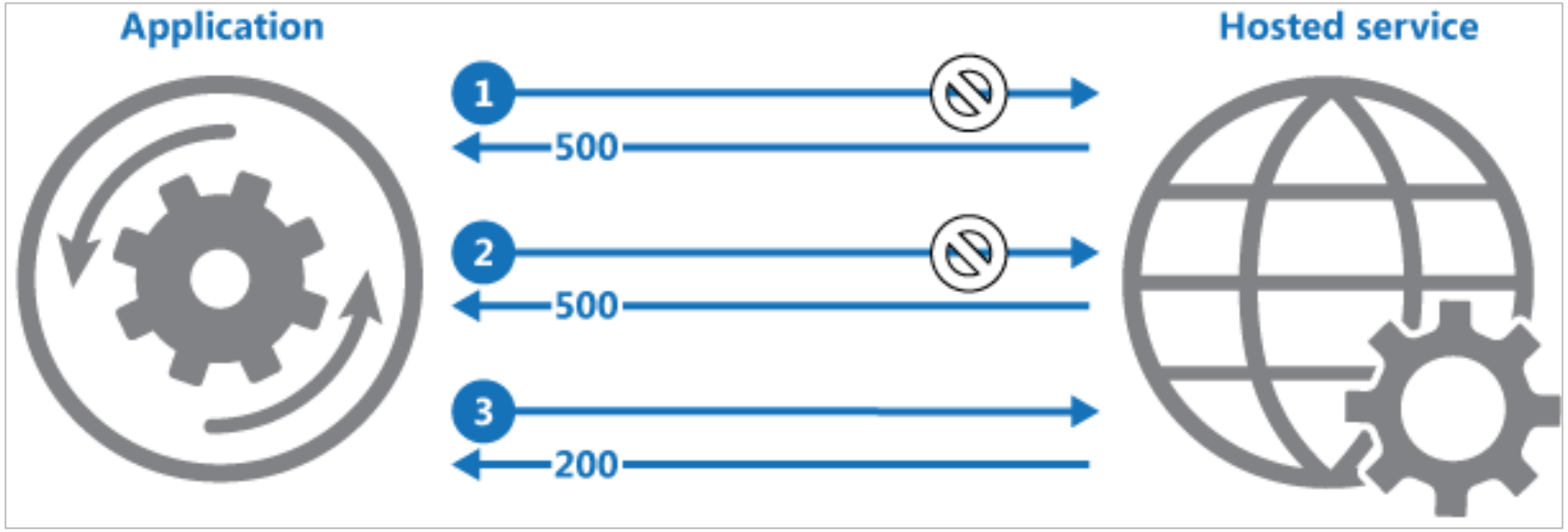
Circuit Breaker Pattern



life.augmented

The Retry Pattern

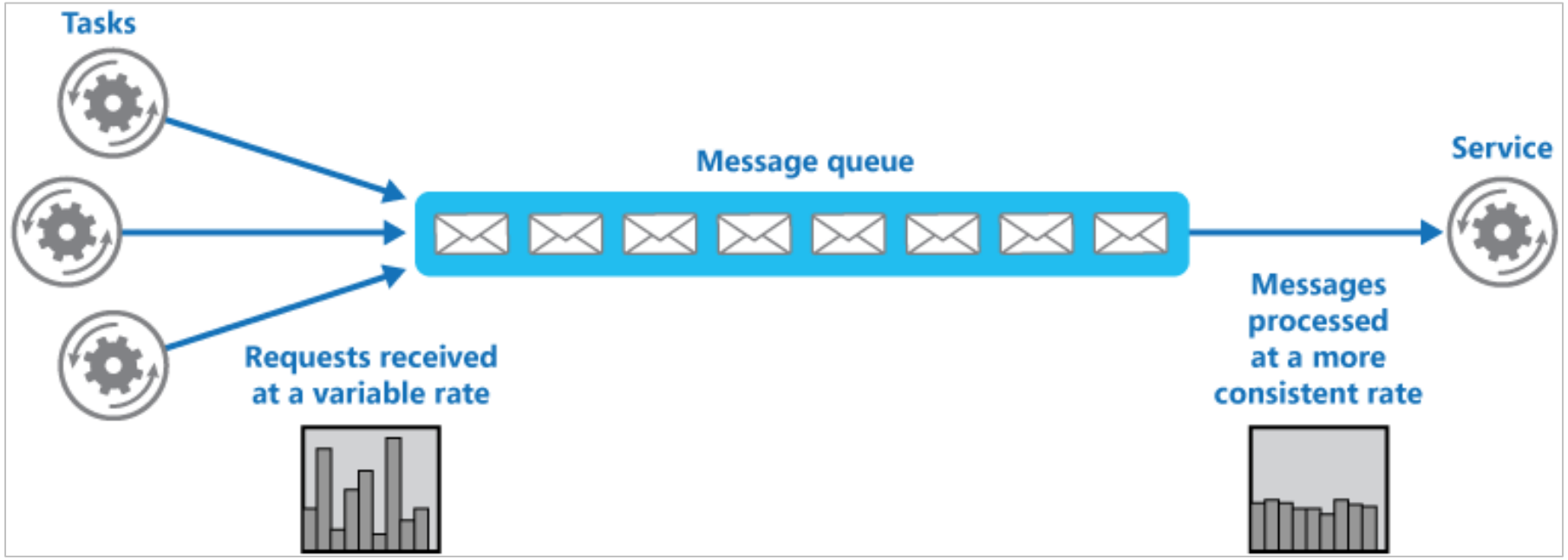
- The Retry pattern is designed to handle temporary failures (Transient Errors)
- Failures are assumed to be transient until they exceed the retry policy



life.augmented

Queue-Based Load Leveling Pattern

- Use a queue to act as a “buffer” between requestor generators and request services
- Queue decouples the tasks from the service
  - Services can work at their own pace



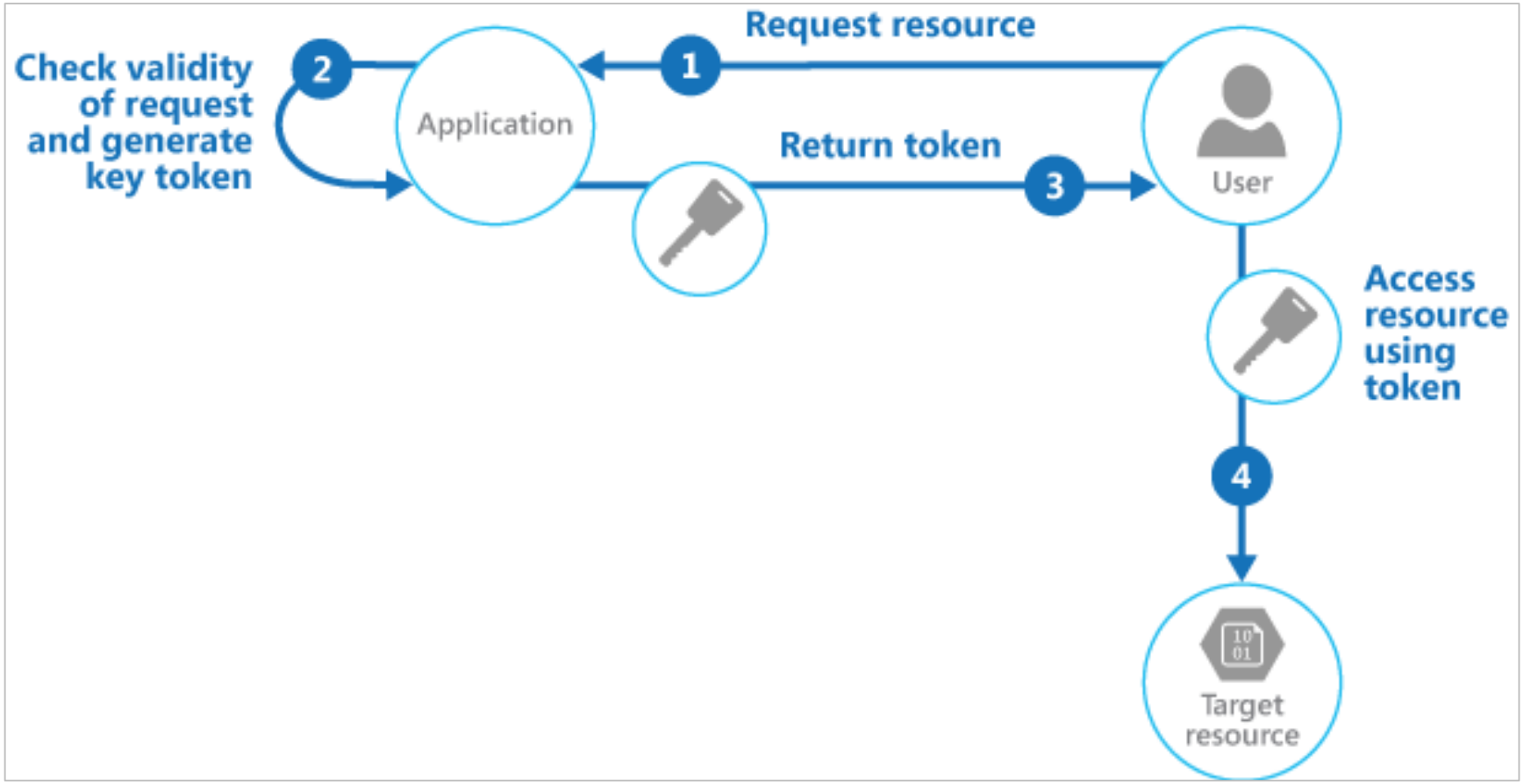
Security Patterns

Security is the capability of a system to prevent malicious or accidental actions outside of the designed usage, and to prevent disclosure or loss of information.

life.augmented

Valet Key pattern

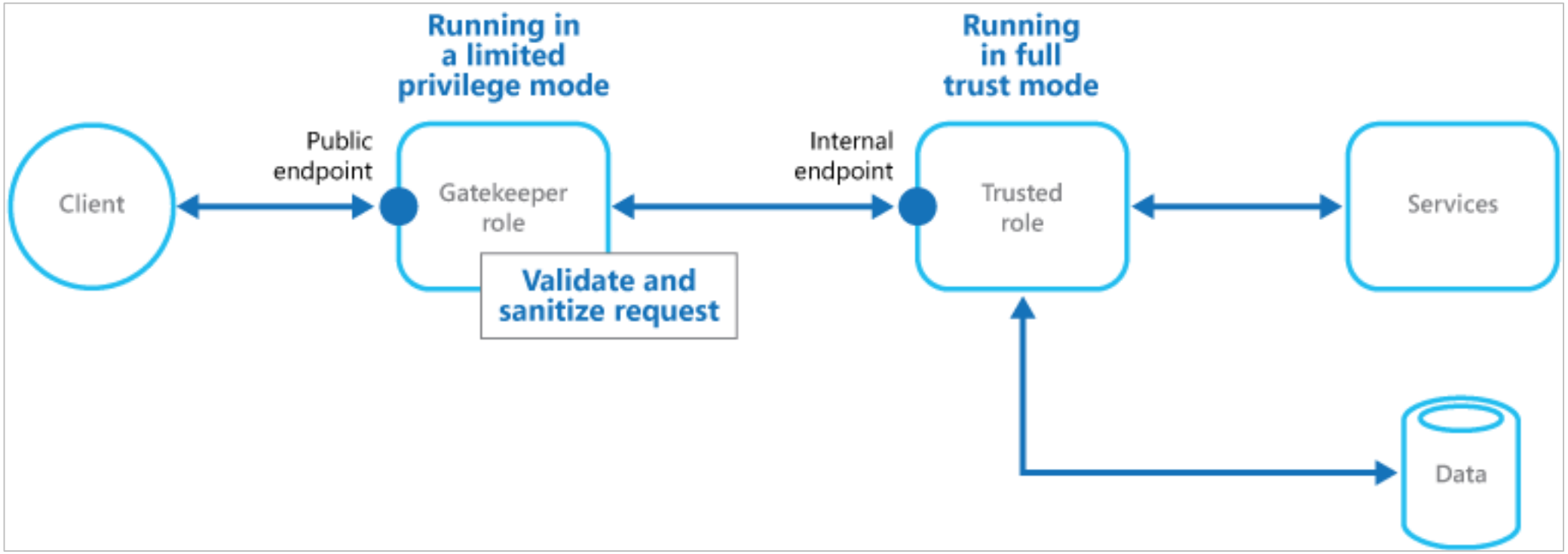
- solve the problem of controlling access to a data store where the store can't manage authentication and authorization of clients.



life.augmented

Gatekeeper pattern

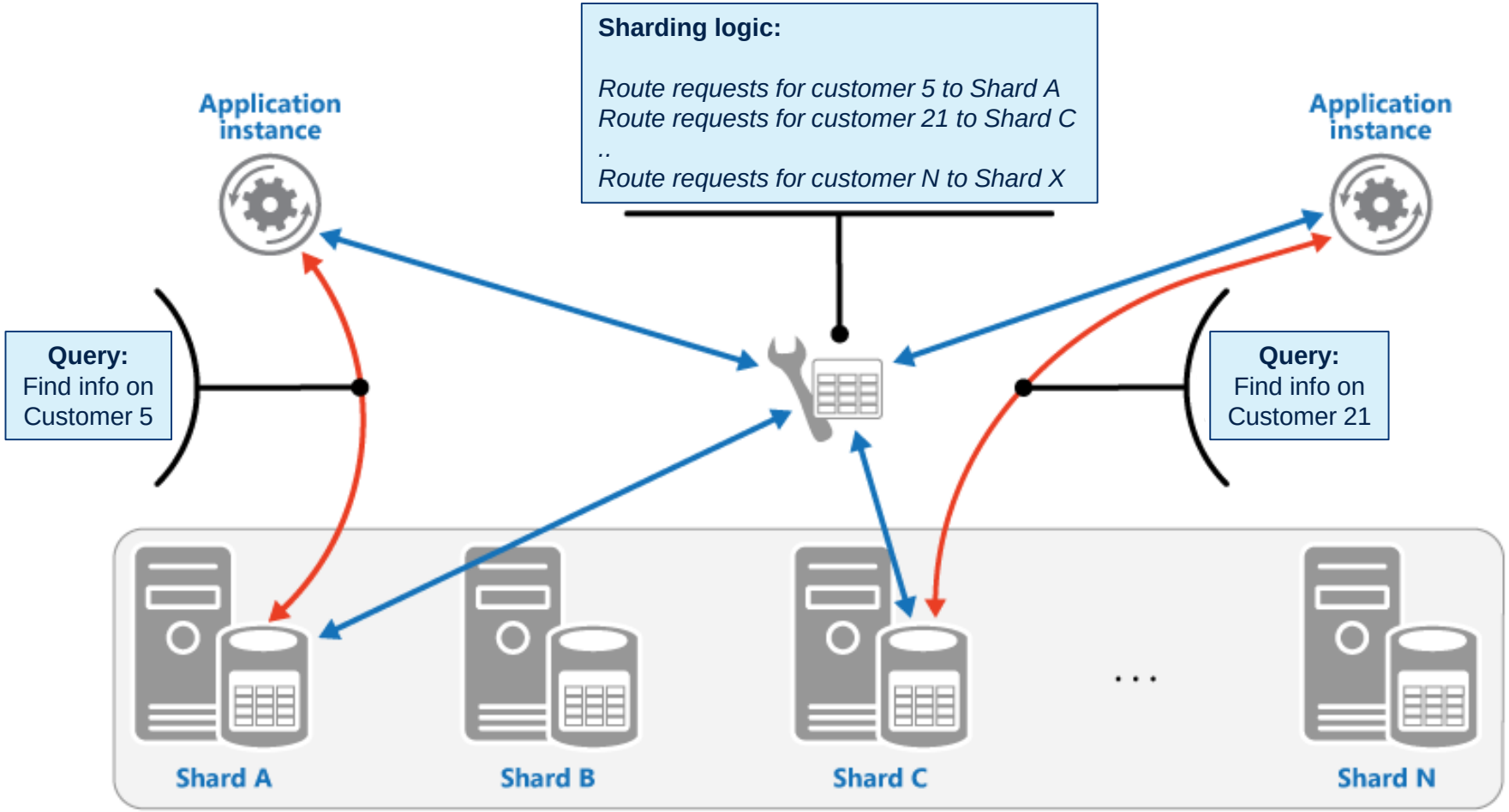
- Provide an additional layer of security, and limit the attack surface of the system



Data Patterns

Sharding Pattern - Lookup

- Divide a data store into a set of horizontal partitions or shards. This can improve scalability when storing and accessing large volumes of data



life.augmented

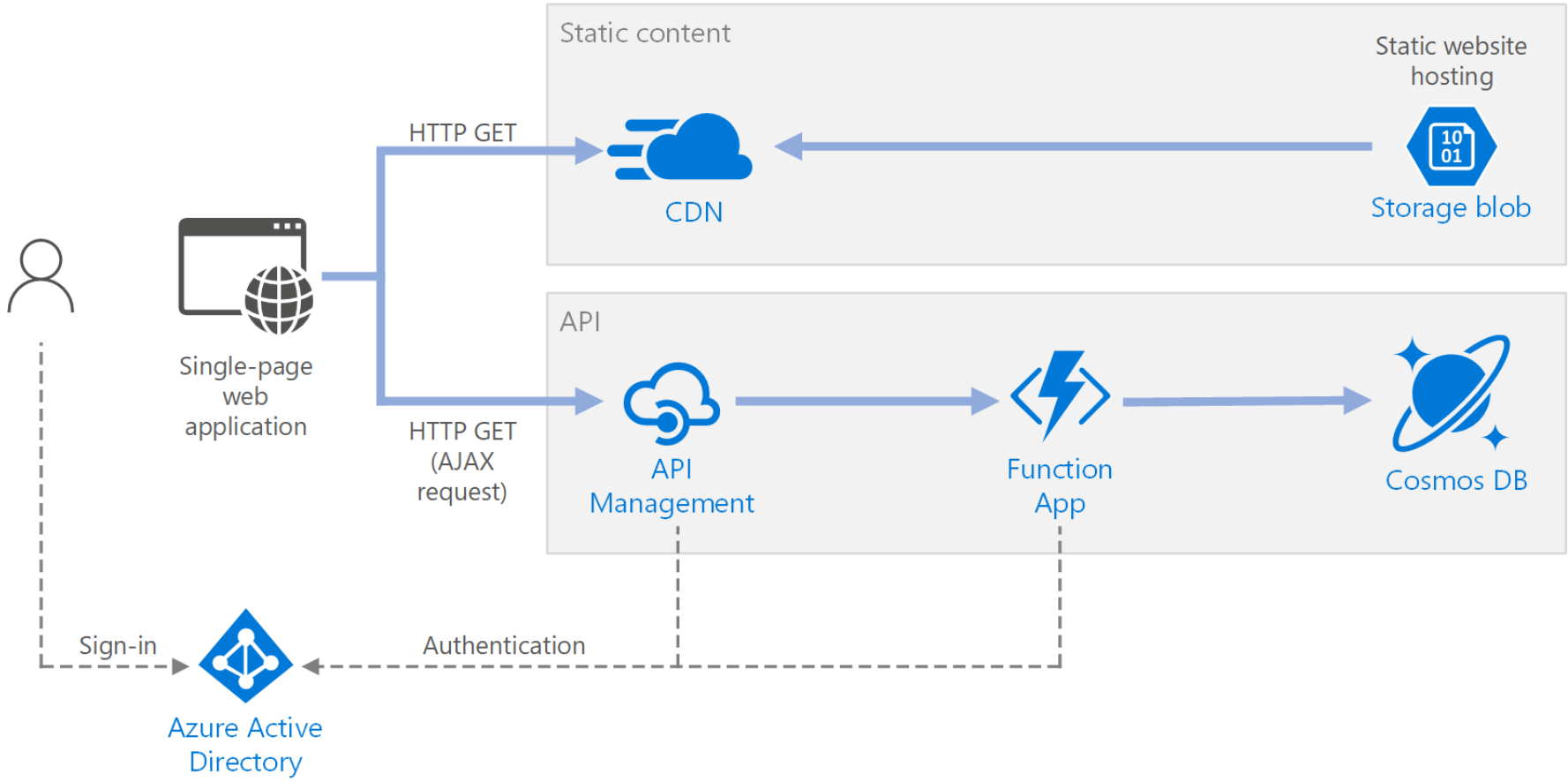
Architecture 1: serverless web app

life.augmented

Example 1: Serverless web application

The example architecture describes a generic single-page application.

YOU CAN DEPLOY THIS ARCHITECTURE:
https://github.com/mspnp/serverless-reference-implementation/blob/v0.1.0-update/README.md



life.augmented

Example 1: architecture considerations

SCALABILITY

- **Functions.** Functions scales based on the traffic (unless fixed plans are used). There is a limit to the number of concurrent function instances, but each instance can process more than one request at a time.
- **API Management**. API Management can scale out and supports rule-based autoscaling. The scaling process takes at least 20 minutes which must be considered depending on which kind of traffic is expected in your application.
- **Cosmos DB**. Use partitioning to scale individual containers in a database to meet the performance needs of the application. Select the best partition key is an important decision that will affect the application performance.

SECURITY

- **HTTPS** should be enforced throughout all the components to improve security.
- Re-use standard services or libraries to implement the **authentication**.
- To secure the back-end components, backend services should **restrict access** to the API service only, this reduces the attack surface of the application.

OPERATIONAL

- **Separate resources** for production, development, and test environments in dedicated and segregated area to allow a correct assignment or access rights.
- Implement **monitoring and logging capabilities** in your app and configure the logging, monitoring and alerting of the services you are using.
- **Deploy using automation** (IaC) to simplify the operational aspects and avoid human errors when deploying in production

BUSINESS CONTINUITY

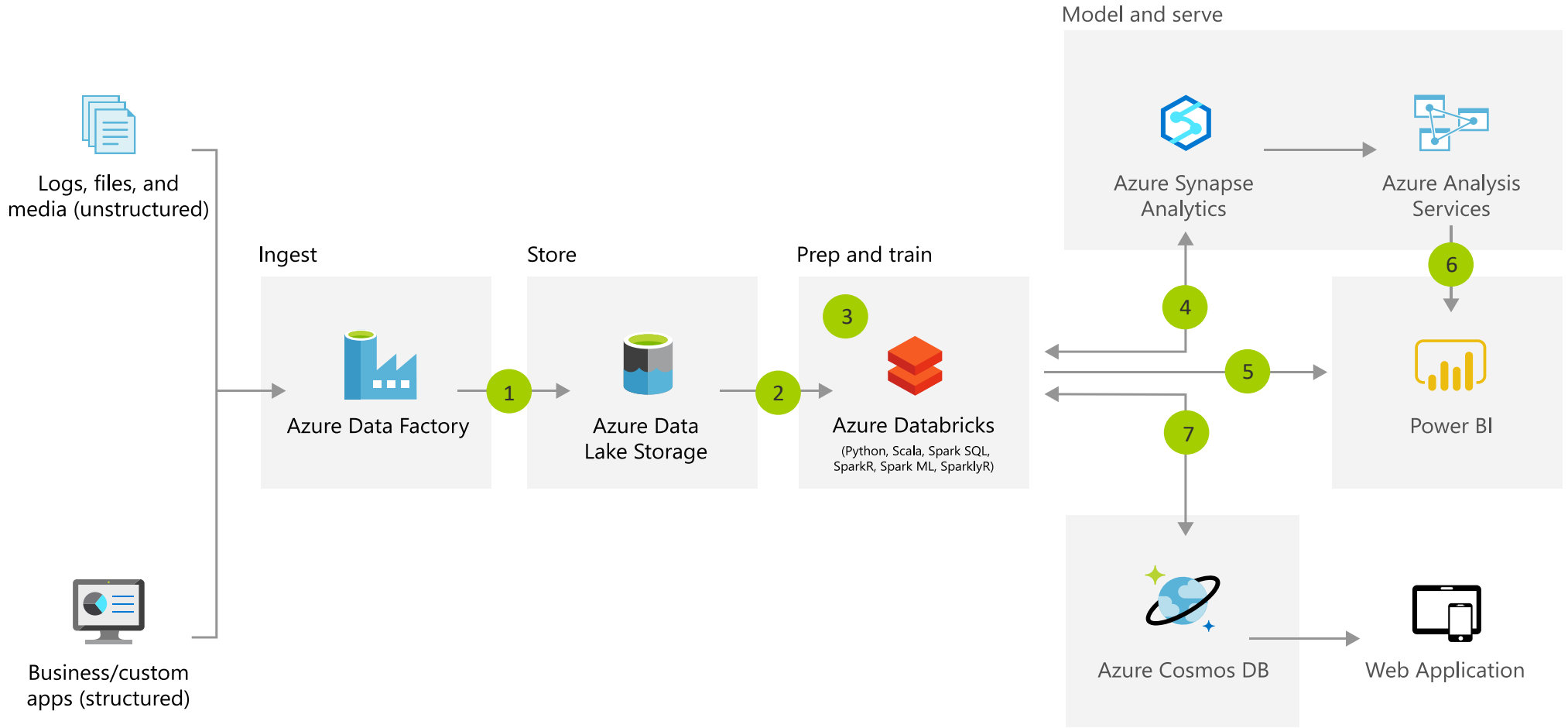
- **Functions.** Functions scales based on the traffic (unless fixed plans are used). But there is a limit to the number of concurrent function instances
- **API Management**. API Management can scale out and supports rule-based autoscaling. The scaling process takes at least 20 minutes which must be considered depending on which kind of traffic is expected in your application.
- **Cosmos DB**. Use partitioning to scale individual containers in a database to meet the performance needs of the application. Select the best partition key is an important decision that will affect the application performance.

Architecture 2: Advanced analytics

life.augmented

Architecture 2: Advanced Analytics

- This architecture allows you to combine data at different scale, and to build custom machine learning models.



Example 2: architecture considerations

SCALABILITY	SECURITY
• **?**	• ?
OPERATIONAL	**BUSINESS CONTINUITY**
• ?	• **?**

life.augmented

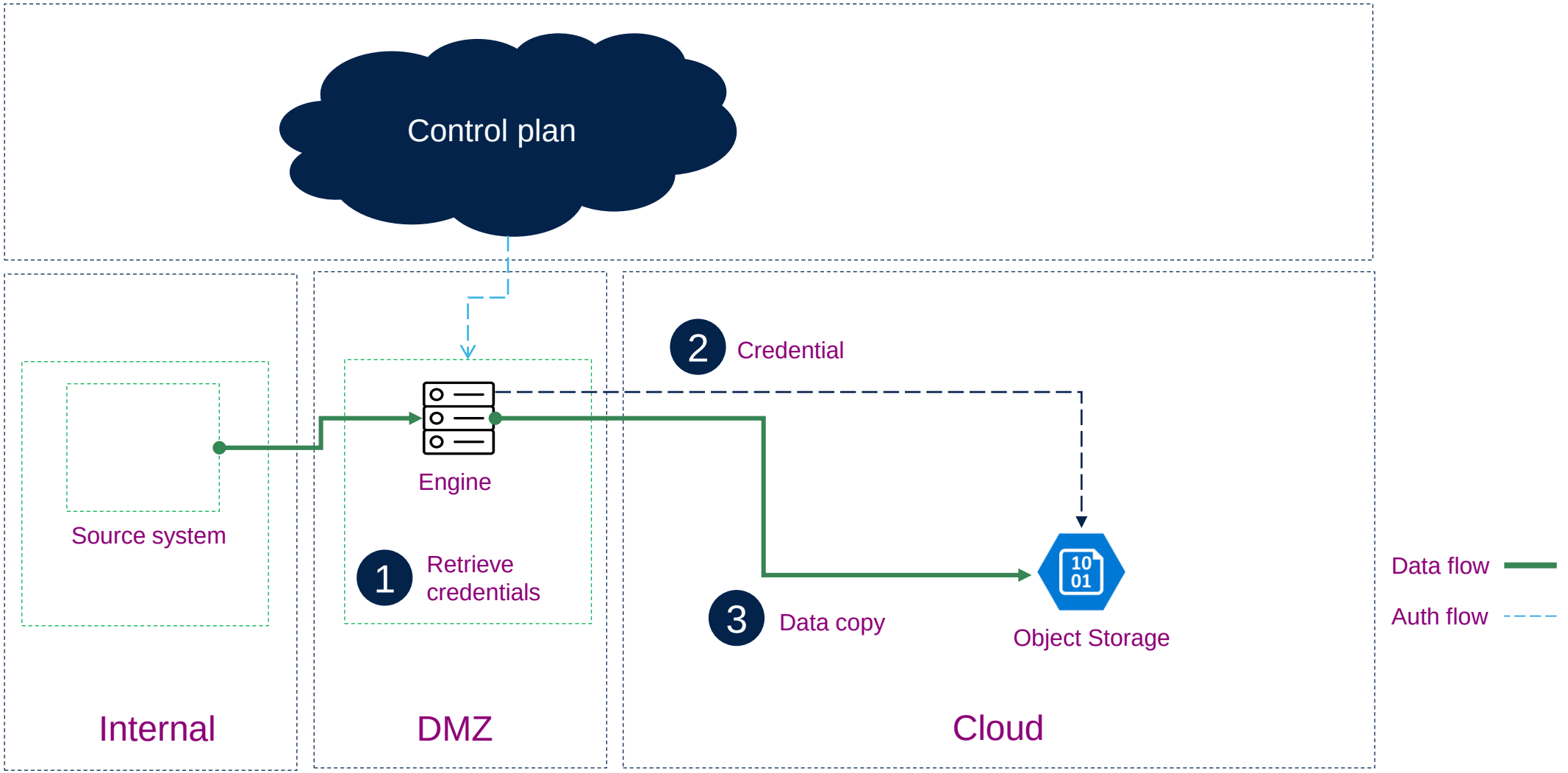
Architecture 2: improve ingestion layer

Data Pattern using an iPaaS for ingestion

Use Case:
Decouple ingestion layer by leveraging an iPaaS

Flow is:

1. iPaaS Engine retrieve the credentials to connect to the source and destination
2. Credentials are used to authenticate to Object Storage and source system
3. iPaaS Engine manage the data transfer from source system to destination (data masking might be done in this stage before writing data into destination)



life.augmented

Q&A

life.augmented

Giuseppe Ursino
Digital Transformation
Enterprise Architect Cloud, Data

Filippo Milotta
Digital Transformation
Data Scientist

Giordano Scuderi
Digital Transformation
Solution Architect

Mario Marroccia
IT Director
Head of ERP Factory

Thank you

life.augmented