IBM **Guardium**
Discover and Classify

SECURITY OVERVIEW

IBM GUARDIUM DISCOVER AND CLASSIFY

VERSION 4.3.2

TABLE OF CONTENTS

SECURITY OVERVIEW

This document provides a step-by-step breakdown of how each component contributes to defense-in-depth, ensuring comprehensive protection against threats while maintaining operational resilience, and an overall description of IBM Guardium practices and approaches to encryption, catalog protection, authentication, and authorization, etc.

Executive Summary:

Our commitment to security is underscored by industry-leading certifications and practices. We are ISO 27001 and SOC 2 Type 2 certified, adhering to the highest standards of information security. As active participants in the CISA Secure-by-Design program and the Security by Design group, we integrate secure development practices from the outset, ensuring robust protection for your data.

Key highlights include:

- **End-to-End Encryption**: AES-256 and TLS 1.2/1.3 secure data both in transit and at rest.
- **Data-At-Rest protection**: Raw/Column/Value level encryption, masking, tokenizing, anonymizing stored catalog data to comply with various financial, health and privacy regulations
- **Authentication & Authorization**: OAUTH2/OpenID standards with multi-factor authentication (MFA) provide comprehensive access control.
- **Advanced API Security**: Leveraging API gateways and rate limiting to ensure resilience against cyberattacks and misuse.
- **Proactive Monitoring**: Logs and user activity audits detect and respond to threats in real-time.
- **Compliance Integration**: Aligns with global data protection standards through continuous monitoring and incident response frameworks.

Our solutions deliver peace of mind with security baked into every layer, going beyond standard measures and leveraging programs like Google security enhancements for our SaaS product as secondary add-ons.

PRODUCT ARCHITECTURE

The diagram below illustrates the data flow in the IGDC application from the security perspective.

IGDC includes 3 deployment units: network analytic engine (NAE), analytic engine (AE) and console manager (CM).

Table 1: Security data stored in the deployment units

DEPLOYMENT UNIT	STORED DATA
NAE	-
AE	Root data asset credentials if RDA loader was used
CM	Data subject catalog, metadata, credentials for data sources.

The table below specifies the ports used for communication between the product components. All ports are encrypted as they use HTTPS transport for communication.

Ports marked with * can be different depending on the Kubernetes cluster configuration. For details, refer to Kubernetes documentation.

Table 2: Communication ports

GENERIC KUBERNETES	AWS EKS, IBM ROKS
Node Port in range **30000-32767*** for communication between CM and AE/NAE.	Load Balancer port **9096** for communication between CM and AE/NAE.

GENERIC KUBERNETES	AWS EKS, IBM ROKS
Port **443** for access to CM/AE UI. Port **6443*** Kubernetes API for maintenance (install, upgrade, etc.).	Port **443** for access to CM/AE UI. Port **6443*** Kubernetes API for maintenance (install, upgrade, etc.).

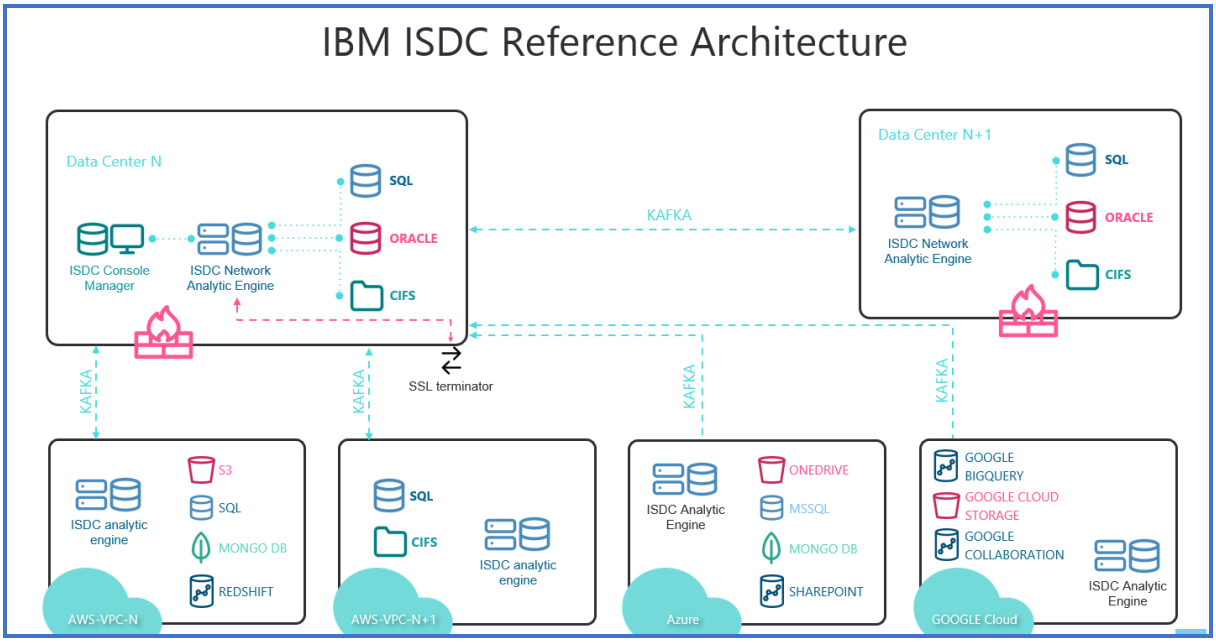

Figure 3: IGDC architecture

CATALOG PROTECTION

In Discover & Classify mode, the master catalog stores the sensitive and sensitive personal data: discovered data element types and metadata of the data source (data source/file), meaning that the system will not store any values of the sensitive data, only the type and metadata.

In Catalog mode, the master catalog stores the sensitive personal data: discovered data element types and values, as well as metadata on all the data sources (data source/file).

See how the data is encrypted in the Encryption section.

LOGS

IGDC has a **Security Audit Log** for audit and reporting purposes. The collected logs do not contain any sensitive or sensitive personal information.

See how the data is encrypted in the Encryption section.

USER ACTIVITY AUDIT

IGDC uses the **Security Audit Log** to monitor the user activity and record the user actions, which may introduce risks to the system. The log is a KAFKA topic, so each IGDC service has *write* access to it. See the logged information in the table below.

Table 4: Security audit log

PARAMETER	DESCRIPTION
ID	Event identifier
TS	Event timestamp - date & time
Name	Event name

PARAMETER	DESCRIPTION
	Options: login, logout, logout_due_inactivity, network_capture_start, network_capture_stop, repository_create, repository_delete, manual_job_back_to_queue, manual_job_stop
Error	Optional field containing the error message in case the intended action caused error
User	Actor that caused the event
Source	Component, which the event is originated in
Data	Optional field containing context attributes and additional information

Security audit log message example:

```
{

  "schema_version": "0.1",

  "id": "uuid",

  "ts": "20/01/2020 16:34:12'23",

  "name": "login|logout|logout_due_inactivity|network_capture_start|network_capture_
stop|repository_create|repository_delete|manual_job_start|manual_job_stop",

  "error": "",

  "user": "user@app.com",

  "source": {

     "applianceId":"xxxxx-xxx",

     "component":"RM|CM|DSAR|APGW"

   },

  "data": {

    "key":"value"

   }

}
```

You can export the **Security Audit Log** to the SysLog server of Splunk over Syslog protocol.

CREDENTIALS TO DATA SOURCES

To enable connection to the data sources for Discovery, Classification and Cataloging, IGDC stores access credentials of a user with read-only permissions. See how the data is encrypted in the Encryption section.

AUTHENTICATION AND AUTHORIZATION

Authentication & authorization are based on open standards such as OAUTH2/OpenId Connect. IGDC uses Authorisation Code + PKCE for user authentication, which is the most secure flow created for SPAs.

1) WebApplication (Client in OAuth terminology) authenticates itself against external/internal OIDC provider it might be Okta/Auth0/KeyClock and others.

2) WebApplication will receive Auth Token. The token will be stored in sessionStore on the Client's side.

3) With each request to the API service WebApp will add an Authentication Request Header with the value of the token.

4) API GW validation incoming token against OIDC provider or with Public Certificate issued for OIDC provider.

5) If a token is valid, API GW will exchange the external token for an internal JWT and add it as an Authorization header to subsequent requests.

6) Service will validate the local JWT against the local Certificate and make sure that the AUD file is correct.

7) Service will use Roles filed of jwt to determine User Permissions.

8) Service will use a subfield to identify the user.

9) If the service has to invoke another service in the same domain, the token should be propagated.

ccess to the IGDC solution is implemented via the Kong API gateway that supports two authorization and authentication services.

Table 5: IBM Guardium authorization and authentication services

SERVICE	DESCRIPTION
KEYCLOAK	Default auth service allowing configuring local and LDAP users for the IGDC services. For details on how to configure the authorization via KeyCloak including the user management, see the How to Configure Authorization with KeyCloak guide.
OKTA	Optional integration of the OKTA Single Sign On solution allowing granting access to the IBM Guardium application to your OKTA users.

The authorization procedure provides full or partial access to the Console Manager UI (Inventory, Candidates search, CM configuration) and to the IGDC platform (Personal information search).

IGDC supports 15 preconfigured user roles as detailed in the table below.

Table 6: IGDC user roles

#	ROLE NAME	DEPARTMENT	DESCRIPTION
1	DC_SUPER_ ADMIN	IT Operations	A user with full access to all modules of the console manager (Inventory, Settings) and to the IGDC platform. Restricted: file preview.
2	DC_ADMIN	IT Operations	A user with access to most pages and configurations with some limitations. Restricted: master catalog re-calculation, export of search results, export of candidates, data subject's profile, file preview.
3	DC_PRIVACY_ OFFICER	Privacy Department	A user responsible for the privacy of sensitive data in your network. Privacy officers can: • Review and export data sources, files. • Review **Privacy** tab on data protection dashboards, billing info and re-calculation history.

#	ROLE NAME	DEPARTMENT	DESCRIPTION
			• Manage data assets, sensitivity classes, data categories. • Search for data subjects, review their profiles, decrypt sensitive data, export reports with data subject's info. • Search for data sources in **Advanced Search** and export reports. • Train candidate virtual views and create root data assets in **Supervised AI**, manage accuracy reports. Full access: data assets, master catalog re-calculation history, billing info. Limited access: data source page, data protection dashboards, data subject search, advanced search, supervised AI, ecosystem, file page.
4	DC_DSR_ OPERATOR	Privacy Department	A user responsible for preparing and sending reports based on DSR requests. DSR operator can: • Generate, edit and send DSR reports by email. • Configure PDF and email templates for DSR reports. • Search for data subjects, review their profile, decrypt sensitive data. • Review the reasons for processing, business units, roles. • Export a file list. Full access: sending DSR report. Limited access: data assets, data subject search, ecosystem, DSR settings, file page.
5	DC_DATA_ GOVERNOR	Data Governor	A user responsible for managing the governance of specific datasets or domains within an organization. Data governor can: • Review and export data sources, files, network map. • Review data assets, **Privacy** and **Risk** tabs on data protection dashboards, attack surface reduction dashboard. • Work with advanced search (with reports), discovery & classification search (without reports), data subject search (encrypted data will be masked). • Train candidate virtual views and create root data assets in **Supervised AI**, generate an accuracy report. • Manage the data source catalog and export reports, excluded data sources. • Re-calculate the master catalog of sensitive data. • Configure and manage sensitivity classes, data categories, data elements, data sources TLS, document classifiers, tags. • Manage cloud discovery. Full access: attack surface reduction dashboard, network map, master catalog re-calculation, data source catalog, data recognition settings, document classification settings, general settings. Limited access: data source page, data assets, data protection dashboards, discovery& classification search, data subject search,

#	ROLE NAME	DEPARTMENT	DESCRIPTION
			advanced search, supervised AI, ecosystem, file page.
6	DC_ FINGERPRINTING_ DATA_OWNER	Fingerprinting Data Owner	A user responsible for document classification: "tagging" files with certain text fragments. Fingerprinting data owner can: • Configure and manage document classifiers, tags, sensitivity classes and data categories. • Generate accuracy report. Full access: sensitivity, data categories, document classification settings, general settings. Limited access: ecosystem.
7	DC_SECURITY_ OFFICER	Security department (SecOps)	A user responsible for sensitive data compliance with the applicable data protection rules. Security officers can: • Review and export data sources, files, network map. • Work with data protection dashboards, discovery & classification search (with reports). • Manage the data source catalog and export reports, excluded data sources. • Configure data sources TLS. • Review data assets. Full access: data source page, data protection dashboards, discovery & classification search, network map, data source catalog. Limited access: data assets, advanced search, ecosystem, file page.
8	DC_RISK_OFFICER	Security department (SecOps)	A user responsible for managing sensitive data risks within an organization. Risk officer can: • Review and export data sources, files, network map. • Work wish discovery & classification search (with reports). • Review **Risk** tab on data protection dashboards, attack surface reduction dashboard, reasons of processing, business units, roles. • Manage the data source catalog and export reports, excluded data sources. • Configure and manage sensitivity classes, data categories, data elements, data sources TLS, analysis, discovery, tags. • Manage cloud discovery. Full access: data source page, discovery & classification search, network map, data source catalog, data recognition settings, discovery settings, general settings. Limited access: data assets, data protection dashboards, advanced search, ecosystem, data source catalog settings, file page.
9	AUDITOR	External	A user responsible for sensitive data audits. Auditor can: • Review and export data sources, files, network map, data source catalog, generate accuracy report. • Work with data protection dashboards, discovery & classification

#	ROLE NAME	DEPARTMENT	DESCRIPTION
			search, data subject search, advanced search with reports. • Review attack surface reduction dashboard, reasons of processing, business units, roles, DSR templates and history. • Configure and manage sensitivity classes, data categories, data elements, tags, manage accuracy reports. Full access: data source page, data protection dashboards, discovery & classification search, network map, advanced search, billing info, data recognition settings, general settings. Limited access: data assets, data subject search, ecosystem, DSR settings, data source catalog, file page.
10	DC_ INTEGRATION_ MANAGER	External	A user responsible for IGDC integrations with 3rd party solutions. Integration manager can: • Configure integrations. • Create policies for the integrations. Full access: ecosystem, policy, billing info, product version info.
11	DC_ CREDENTIALS_ ADMIN	IT Operations	A user responsible for credentials for IGDC access to the data sources for analysis. • Configure and manage credentials. Full access: product version info, credentials in data source catalog settings. Limited access: ecosystem.
12	DC_FILE_ PREVIWER	File Previewer	A user responsible for validating the data elements discovered in files: • Preview files on the file page. Full access: file preview popup and preview generation buttons. Restricted: file page. **Note:** This role cannot be used independently, only in combination with at least one other role having access to the file page. To be able to preview files, a user must have one of the following role combinations: • DC super admin + DC file previewer; • DC privacy officer + DC file previewer; • DC data governor + DC file previewer; • DC security officer + DC file previewer; • DC risk officer + DC file previewer; • DC auditor + DC file previewer. Before assigning the DC file previewer role to a user, please consider the following risks: • A user can make a screenshot(s) of the file, and it cannot be restricted by the system. • A user can make photos of the documents/files, and it cannot be restricted by the system.

#	ROLE NAME	DEPARTMENT	DESCRIPTION
			• APIs with the files previewed might be hacked by the users in order to download all previewed reports. • The sensitive data values are not masked in the preview window. When previewing one data element, the user might see some other (possibly more critical) values of other data elements in the context frames.
13	DC_ INTEGRATIONS_ BASE	Integrations	Technical user role for integration implementation. The user can: • Configure and manage sensitivity classes, data categories, and regulations; • Manage data element categories, data element configuration, and tags; • View Billing Info, product version info, and IAM page; • Read and manage the Data Source Catalog, read/filter the "Data element" column. Full access: sensitivity, data category, regulation, tag settings, billing info, data recognition. Limited access: data source catalog.
14	DC_ INTEGRATIONS_ PRIVACY	Integrations	Technical user role for integration implementation in the privacy field. The user can: • Review Privacy and Classification tabs in data sources, and export the data distribution table (classification tab); • View the data asset list including widgets, view data assets, review data asset details and sources; • Execute and review the search results, export the metadata and personal data for the list of data subjects, review the data subject profile; • View discovered, trusted, and conflict sources for data subject profile, export the data subject data with encrypted value. Limited access: data subject search, data asset, data source page.
15	DC_ INTEGRATIONS_ SECURITY	Integrations	Technical user role for integration implementation in the security field. The user can: • Review the Classification tab in a data source content page, view and export the list of content drill down in file system data sources; • Execute, view, and export classification search results (both files and data sources); • Execute advanced search, view, and export the search results. Limited access: data source content page, classification search, advanced search.

When a user opens IGDC, the landing page shows a list of all modules. If the user role has access to some functionality within the module, the module name will be a clickable link, otherwise the name is unclickable text.

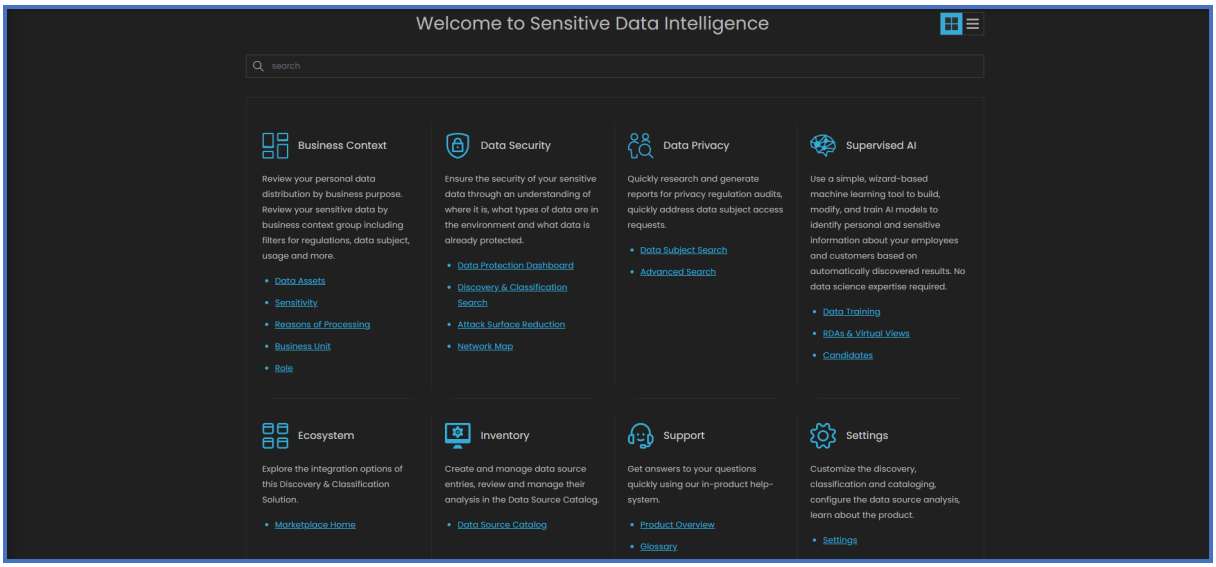

Figure 7: IGDC user roles

For the detailed permissions on each page per user role, see User Roles' Permissions.

API

IGDC exposes API to 3rd party solutions, providing access to the master catalog data. See the data stored in the Catalog protection section. See how the data is encrypted in the Encryption section.

ENCRYPTION

IGDC uses encryption to protect all types of sensitive data stored and exchanged within the application. The protected data includes the stored / in transit data, user data, authentication & password, etc. See the encryption usage details in the table below.

Table 8: Encryption usage

ENCRYPTION CATEGORY	MECHANISM / ALGORITHM / STANDARD	ENCRYPTION KEY LENGTH
Authentication & password protection	PBKDF2 (KeyCloak) / RSA	2048
Signature	RSA 2048	2048
Integrity check	RSA	2048
Confidentiality (user data)	AES	256
Stored / in transit data	AES 256 / RSA 2048 (option)	256 / 2048
Databases	AES 256 (from 2.3)	256

TLS USAGE

IBM Guardium services serve their traffic via HTTPS for internal communication. Versions such as TLSv1.2 and TLSv1.3 are supported by default. Supported TLS ciphers are:

- TLS_AES_256_GCM_SHA384,
- ECDHE-ECDSA-AES256-GCM-SHA384,
- ECDHE-RSA-AES256-GCM-SHA384.

TLS versions and ciphers can be configured using the Helm chart values per chart.

IGDC uses the TLS (Transport Layer Security) protocol in several places: Kafka, Ingress.

KAFKA

Kafka uses TLS for client-broker communication, inter-broker communication, and Zookeeper communication. For more details, refer to the vendor documentation.

The enabled protocol versions are TLSv1.2, TLSv1.3.

Table 9: TLS usage in Kafka

COMMUNICATION	DEFAULT PROTOCOL
Client to broker	TLSv1.3
Inter broker	TLSv1.3
Zookeeper	TLSv1.2

INGRESS

Ingress TLS settings may depend on your configuration of the Ingress Nginx controller and Load Balancer (LB) configuration in case you configured an LB to terminate TLS.

For the ISO installation, the default enabled protocol versions are TLSv1.2, TLSv1.3.

API SECURE OPERATIONS

1. Audit and update regularly. Conducting regular security audits and keeping APIs up to date using Security audits, Vulnerability assessments, prompt Patch management, API Version control, and Regular audits.

2. Implement robust authentication mechanisms to ensure the right individuals or systems access our API. Keycloak allows us to cover widely adopted authorization frameworks that provide robust authentication for both applications and users like Oauth2, SAML, OpenID etc, usage of API keys for simpler client authentication, if needed, Multi-Factor Authentication (MFA) and usage of JWTs to securely transmit information between parties

3. Code to protect against common cyber attacks. Our APIs are resilient against common attacks like SQL injection, cross-site scripting (XSS), and cross-site request forgery (CSRF) using: Input validation, Content Security Policy CSRF tokens

4. Encryption of sensitive data. Data at rest: we encrypt sensitive data when it's stored, whether in databases or on disk. We implement strong encryption algorithms and manage encryption keys securely. Data in transit: data transmission happens over secure channels only using protocols like HTTPS. We also provide tools to ensure proper certificate management.

It guarantees that even if data is intercepted, it remains unreadable and confidential.

5. Usage of API gateways. API gateways act as intermediaries between clients and our API services. They offer centralized management and security enforcement:

Authentication and authorization: Implement authentication and authorization logic within the API gateway. This centralizes security controls and simplifies management.

Logging and monitoring: API gateways can capture detailed logs and metrics, aiding security incident detection and response.

Traffic control: Control and manage network traffic destined for our APIs, enabling features like rate limiting and content caching.

Security plugins: Many API gateways offer security plugins that can be customized to enforce security policies designed to optimize API security.

Leveraging API gateways can simplify and streamline security management, ensuring consistent security enforcement across all our APIs.

6. Implement rate limiting. Kong Gateway imposes rate limits on clients through the use of the Rate Limiting plugin. When rate limiting is enabled, clients are restricted in the number of requests that can be made in a configurable period of time. The plugin supports identifying clients as consumers or by the client IP address of the requests. Rate limiting is an essential mechanism to prevent abuse and misuse of our API. It controls the number of requests a client can make within a specified timeframe:

Thresholds: Defining reasonable request thresholds for different types of clients (e.g., free users, premium users, applications).

Granularity: Implementing per-endpoint rate limiting to ensure that a spike in requests to one endpoint doesn't affect the entire API.

Error handling: Returning appropriate error responses when rate limits are exceeded, allowing clients to adjust their behavior.

SAAS SECURITY INSIGHTS

The customer environment is architected with security as a foundational pillar, leveraging Google Cloud Platform's (GCP) advanced services such as **Security Command Center (SCC) Premium** and **Coralogix** for centralized logging and analytics.

NETWORK SECURITY

Table 10: Network security

VPC Configuration	The Virtual Private Cloud (VPC) architecture enforces strict segmentation and isolation of resources.	
	Private Subnets	• Nodes in the GKE cluster reside in private subnets without external IPs to eliminate direct exposure to the internet. • Subnets are tagged with specific purposes to maintain tiered isolation.
	Private Google Access	Enabled to allow resources in private subnets to securely access Google APIs over internal IP addresses, bypassing the internet.
	Firewall Rules	Enabled to allow resources in private subnets to securely access Google APIs over internal IP addresses, bypassing the internet.
Ingress Traffic Control	**Load Balancer**	• External HTTPS Load Balancer terminates SSL/TLS, providing encrypted entry points for users. • Uses Cloud Armor to enforce **OWASP Top 10** rules, blocking threats such as SQL Injection, XSS, and CSRF.
	Firewall Policies	• All ingress traffic is permitted only from trusted IP ranges or through VPN tunnels. • Logging is enabled for rule hits to provide visibility into access patterns.
Egress Traffic Control	**Proxy VM**	Outbound internet traffic is routed through a scalable proxy VM that applies strict policies, allowing communication only to trusted destinations.
	Firewall Rules	Egress traffic is denied by default unless explicitly

		allowed by a rule, reducing the risk of data exfiltration.

WORKLOAD SECURITY

Table 11: Workload security

GKE Cluster Configuration	**Private Cluster**	• Nodes communicate with the control plane using private DNS. • Admin access is restricted to a bastion VM via internal IPs only.
	Node Security	**Secure Boot** and **Integrity Monitoring** are enabled to ensure nodes are free from unauthorized modifications.
	Encryption	Persistent disks and secrets are encrypted with Customer-Managed Encryption Keys (CMEK).
Workload Identity and Access	Kubernetes Service Accounts (KSA) are mapped to Google Service Accounts (GSA) with minimal permissions, ensuring workloads adhere to the principle of least privilege.	
Image and Vulnerability Security	• Images are sourced exclusively from GCP's Artifact Registry. • Vulnerability scanning is configured to identify and remediate known Common Vulnerabilities and Exposures (CVEs).	
Node Security	• **Secure Boot** ensures only verified operating system images are loaded. • **Integrity Monitoring** alerts for unauthorized changes during runtime.	
WAF	WAF (Web Application Firewall) protects from cyber attacks by analyzing incoming traffic.	

DATA PROTECTION

Table 12: Data protection

Encryption	**At Rest**	Data is encrypted using CMEK for enhanced control and auditability.
	In Transit	TLS 1.2 is enforced for all inter-component communication, ensuring end-to-end encryption.
Centralized Logging	All logs are collected at the organization level and exported to Coralogix via Pub/Sub for real-time analysis.	

SECURITY COMMAND CENTER (SCC) PREMIUM

Table 13: Security Command Center (SCC) Premium

Built-In Remediation	• Automated playbooks are configured for high-severity findings, such as: • Automatically displaying overly permissive IAM roles. • Enforcing firewall rule changes to block malicious IPs.
Threat Detection	• All curated rules for anomaly detection are enabled, monitoring for: • API abuse. • Misconfigurations in IAM and VPCs. • Compromised credentials.
Continuous Risk Engine	A dashboard provides actionable insights into misconfigurations, highlighting trends and compliance gaps.
Cloud Posture Management	Integrated with SCC's native CSPM for continuous compliance checks against standards such as CIS, PCI, and GDPR.

INTEGRATION WITH CORALOGIX

Table 14: Integration with Coralogix

Log Streams Sent to Coralogix	The following logs are exported in real time to Coralogix: • Audit Logs (Admin Activity, Data Access, System Events). • Access Logs (Authentication and resource access events). • Network Logs (VPC Flow Logs, Firewall Hits). • Execution and Request Logs.	
Data Flow and Processing	• Logs are streamed from Pub/Sub to Coralogix, where: • Alerts are triggered for suspicious activities. • Data is visualized in real-time dashboards.	
Use Cases and Dashboards	**Incident Dashboards**	Real-time views of critical security events.
	Compliance Reports	Automated reports on adherence to regulatory standards.

COMPLIANCE AND MONITORING

Table 15: Compliance and monitoring

Compliance Standards	• Regular assessments ensure alignment with CIS, PCI-DSS, GDPR, and HIPAA. • SCC and Coralogix provide compliance evidence and reports.
Continuous Improvement	Policies and configurations are reviewed monthly to adapt to emerging threats.

INCIDENT RESPONSE FRAMEWORK

- SCC and Coralogix collaborate to detect, log, and escalate incidents.
- Automated workflows ensure incidents are triaged and resolved promptly.

APPENDIX A: HTTP CERTIFICATE

The default version of IGDC is deployed with self-issued certificates. To improve system performance and usability, you can generate and upload your certificates.