Talend Cloud Data Catalog

Security architecture overview

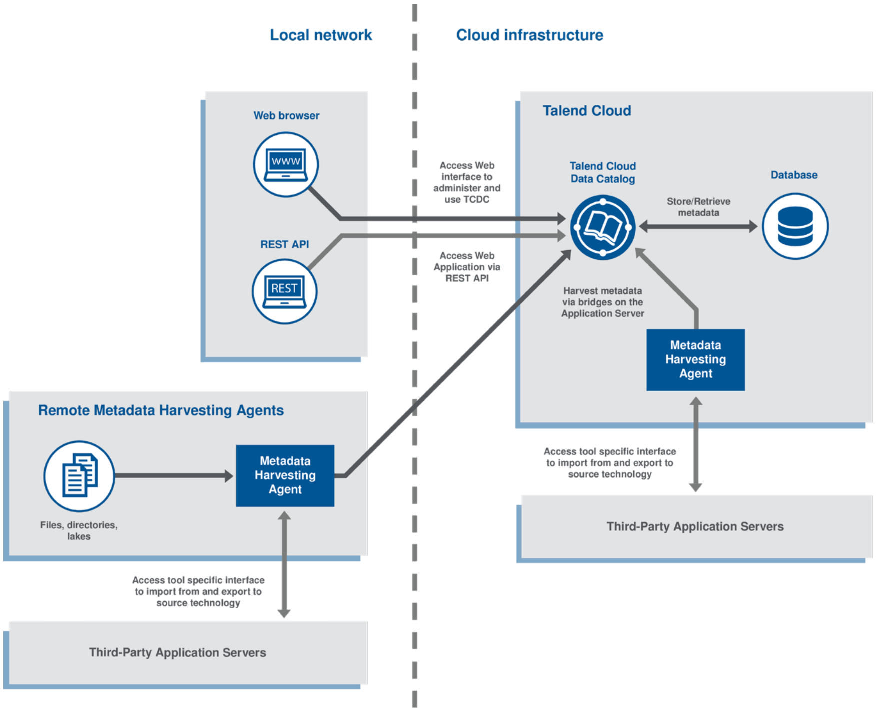
Contents

Summary

Talend Cloud Data Catalog is a managed platform that helps you to create a central, governed catalog of enriched data. It can automatically discover, profile, organize, and document your metadata and make it searchable. Talend leverages security and privacy best practices to protect both the Talend platform and Talend, the company. Talend implements a combination of policies, procedures, and technologies to ensure your data is protected and secured. Talend's chief information security officer (CISO) defines the Talend security strategy, architecture, and program. This document provides an overview of the Talend internal architecture and our policies and procedures as they pertain to employee, physical, network, infrastructure, platform, architecture, and data security.

Talend Cloud
Data Catalog architecture

Talend Cloud Data Catalog is a multitenant managed platform that helps you to create a central, governed catalog of enriched data. All managed



Local network
Cloud infrastructure
Web browser
WWW
REST API
REST
Access Web interface to administer and use TCDC
Access Web Application via REST API
Talend Cloud
Talend Cloud Data Catalog
Database
Store/Retrieve metadata
Harvest metadata via bridges on the Application Server
Metadata Harvesting Agent
Access tool specific interface to import from and export to source technology
Third-Party Application Servers
Remote Metadata Harvesting Agents
Files, directories, lakes
Metadata Harvesting Agent
Access tool specific interface to import from and export to source technology
Third-Party Application Servers

Figure 1: Talend Cloud Data Catalog functional architecture

Talend Cloud Data Catalog infrastructure

Each Talend Cloud Data Catalog customer has its own account to access the environment. The account contains the number of users defined by

Computation resources

Talend Cloud Data Catalog gives separate computation resources to each tenant. Each tenant is hosted in an AWS Elastic Kubernetes Service (EKS) pod.

Data storage

Talend works with two general types of data: data that we collect and data that customers use with the software.

Data that we collect

Talend, across its cloud applications, collects only customer information that it needs to provide its services or to manage customer accounts.

All personally identifiable information collected by Cloud Data Catalog (such as name, country, and email address) is protected with best security practices: It is encrypted at rest via AES-256 and in transit via TLS 1.2.

Figure 1: Talend Cloud Data Catalog functional architecture

Secrets such as passwords, keys, and certificates are managed via third-party technologies and products. Read the Key Management section for more details.

No payment information is stored in Talend Cloud Data Catalog. We rely on third-party vendors to collect and manage payment information.

Data that customers use with Talend Cloud Data Catalog

Metadata and any other objects that Talend stores to provide services or for security reasons are isolated via tenant-specific schemas.

Network

Talend networks and systems are protected via network and application firewalling, visibility mechanisms, and micro segmentation strategies.

Data flows

This section gives an overview of the data flows between Talend Cloud Data

Data flows from remote metadata harvesting agents and Talend Cloud Data Catalog clients

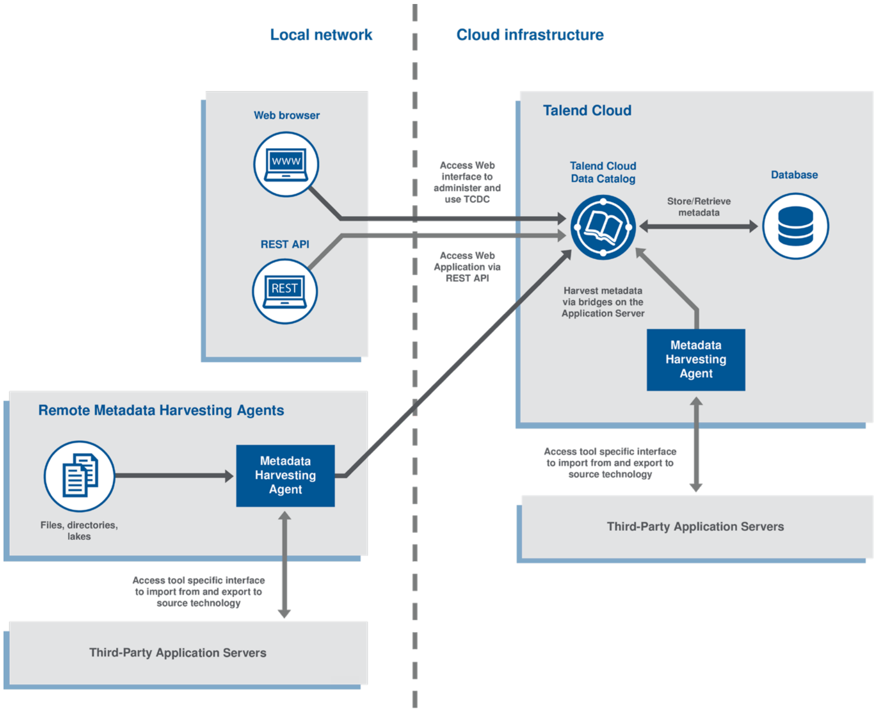

Local network
Cloud infrastructure
Web browser
WWW
REST API
REST
Access Web interface to administer and use TCDC
Access Web Application via REST API
Talend Cloud
Talend Cloud Data Catalog
Database
Store/Retrieve metadata
Harvest metadata via bridges on the Application Server
Metadata Harvesting Agent
Access tool specific interface to import from and export to source technology
Third-Party Application Servers
Remote Metadata Harvesting Agents
Files, directories, lakes
Metadata Harvesting Agent
Access tool specific interface to import from and export to source technology
Third-Party Application Servers

Figure 2: Talend data flows when using Cloud Engines

From the remote agent to the Catalog server: data sampling, data profiling, technical metadata such as data model schema, agent version and status, agent logs, and harvesting results.

From the Catalog server to the remote agent: schedule of harvesting, command to start harvesting, and semantic types definitions.

Incoming requests from remote metadata harvesting agents to Talend Cloud Data Catalog Server require a secret credential.

Talend uses custom connection procedures between Talend Cloud Data Catalog Server and third-party applications servers. Procedures differ depending on the third-party implementation.

Metadata is transferred to Talend Cloud via the following URLs:

Cloud	Region	Talend Cloud Data Catalog service URL
AWS	US	https://{tenant}.datacatalog.us.cloud.talend.com/MM/rest/v1/
	Europe	https://{tenant}.datacatalog.eu.cloud.talend.com/MM/rest/v1/
	Asia-Pacific	https://{tenant}.datacatalog.ap.cloud.talend.com/MM/rest/v1/

Public APIs

In addition to the data flows between Talend applications, Talend exposes public APIs that let developers automate workflows. These APIs are secured with personal access tokens generated at login time.

Cloud	Region	Talend Cloud Data Catalog service URL
AWS	US	https://{tenant}.datacatalog.us.cloud.talend.com/MMDoc/
	Europe	https://{tenant}.datacatalog.eu.cloud.talend.com/MMDoc/
	Asia-Pacific	https://{tenant}.datacatalog.ap.cloud.talend.com/MMDoc/

Security at Talend

Talend's security organization consists of a dedicated team of security experts distributed across the company who work closely with the Talend CISO. Their mission is to protect Talend and its clients with security best practices. This team supports all aspects of Talend business, including Talend development and operations. The responsibility of Talend security rolls up to the CISO, who also defines Talend security strategy, architecture, and program.

Physical security

Talend maintains security controls to prevent unauthorized physical access to buildings and data centers and to protect its systems and software, and by extension the Talend environment, from damage, interruption, misuse, or theft.

Authorizations are granted only to those people who need them for work; they are reviewed regularly, and access is monitored continuously.

Security training

All Talend employees are trained on security best practices. All Talend employees involved in the Talend development lifecycle, from creation to deployment and operation, are guided through training, reviews, and drills.

Secure software development

Talend's security organization is involved throughout the creation of any new application, capability, or feature.

Our security experts conduct architecture, design, and code reviews.

Software composition analysis (SCA) and static security vulnerability (SAST) scans are integrated in the software development lifecycle.

Talend implements a Top 10 Open Web Application Security Project (OWASP) awareness program during application development, and schedules regular internal and external audits to assess compliance with OWASP best practices.

Cloud workload protection and monitoring

We use a combination of security services from third-party vendors to protect Talend Data Fabric.

Our security experts use external scanning tools to ensure that systems and containers are hardened, configured, and patched according to Talend guidelines and best practices.

Talend uses NIST Cybersecurity Framework as part of its global security strategy.

Our deployments leverage the built-in segmentation capabilities of AWS EC2 Security groups and Microsoft Azure Network Security groups to restrict inter-resource communication.

Talend Cloud's perimeter security is composed of (but not limited to):

- Web Application Firewall (WAF) — validates, monitors and filters all web application and API traffic
- Network-based intrusion detection system (IDS) and intrusion prevention system (IPS) — alert on rogue activity and protect against threats such as zero-day attacks
- Security information and event management system (SIEM) — monitoring and observability of system status, performance and detection of rogue processes

Authentication, authorization, and access control

Standard access

Tenant users are authenticated with their own unique credentials: username plus password.

Talend uses TLS certificates issued by Talend's approved Certificate Authority (CA), GoDaddy to secure and encrypt all communications between user systems and Talend Cloud Data Catalog. Talend Cloud Data Catalog supports HTTPS over TLS.

The authentication process follows the OpenID Connect standard and uses either the authorization code or the implicit flow. Once connected, a session is managed using cookies.

Administrative access

Talend Cloud Data Catalog administrative access requires management review and approval. Elevated privilege access requires the same level of approval by management.

Access to Amazon management console requires multifactor authentication (credentials plus secret keys).

Access to the AWS console is restricted to select members of the Talend Site Reliability Engineering (SRE) and Information Security teams. New account creation follows a strict approval process. Accounts are reviewed quarterly.

System access is provided via Kubernetes administration management, relying on AWS authentication.

Password management

Talend maintains a password management policy that all employees must comply with. It ensures the creation of strong passwords, the protection of those passwords, and that passwords are never reused.

Key management

Talend relies on AWS-managed Customer Master Keys (CMK) for encryption. Talend uses its own AWS CMK to generate unique data encryption keys (DEK).

Most DEKs are tenant-specific and are managed (including rotation) by Talend. DEKs that do not need to be tenant-specific are managed via the AWS Encryption SDK.

Front-end TLS endpoints are managed through the AWS Certificate Manager (ACM). The private key is generated by Talend and the associated certificate signed by Talend's approved Certificate Authority (CA), GoDaddy. The certificates are then published as part of the Certificate Transparency program and uploaded to the ACM.

Vulnerability management

All applications are tested by Talend's security experts (dynamic application security testing (DAST) and penetration tests) at least twice a year.

In addition, Talend leverages internal and third-party security services to perform external penetration tests.

Third-party penetration tests are scheduled twice a year and prior to any new Talend Data Fabric release and deployment. The penetration tests cover a wide range of security aspects of the application and address modern web best practices.

All detected vulnerabilities are logged by the Talend Quality Assurance team and analyzed by the Talend Information Security team, which then supports, tracks, and tests their remediation.

Talend follows the Security Content Automation Protocol (SCAP) framework. Vulnerabilities are rated according to the Common Vulnerability Scoring System (CVSS) v3.0 equation. Vulnerabilities are resolved depending on their severity rating and their potential impact on the infrastructure.

Third-party penetration test reports are available upon request at Talend's discretion.

Backups

Talend uses AWS services for both mirroring and long-term storage. All storage processes are automated, monitored, and tested. Mirrors and snapshots are performed twice daily.

Disaster recovery and business continuity

Talend maintains disaster recovery/business continuity (DR/BC) plans that are reviewed, updated, and tested at least annually.

Talend operates in multiple AWS and Azure regions globally. The redundant infrastructure has primary and disaster recovery data centers in each of the Talend Cloud regions, and multiple Availability Zone (AZ) architecture per region.

We are in close contact with both vendors and carefully monitor their service levels to make sure that they meet our required service levels. Latest uptime per region is available on https://trust.talend.com.

Talend R&D and Operational teams span multiple geographical locations: US, Europe, and Asia. Every function and duty can be fulfilled by at least two people.

Security certifications

Talend is SOC 2 Type 2 compliant and eligible to sign HIPAA (Health Insurance Portability and Accountability) Business Associate Agreement (BAA).

We use the Cloud Security Alliance (CSA) Security Trust Assurance and Risk (STAR) program to assess our security practices and validate the security posture of our cloud offerings.

A comprehensive list of security certifications and privacy compliances are available on https://www.talend.com/security/.

Refer to AWS and Azure websites for more details about their security certifications and compliance information.

About Talend

Talend, a leader in data integration and data integrity, is changing the way the world makes decisions.

Talend Data Fabric is the only platform that seamlessly combines an extensive range of data integration and governance capabilities to actively manage the health of corporate information. This unified approach is unique and essential to delivering complete, clean, and uncompromised data in real-time to all employees. It has made it possible to create innovations like the Talend Trust Score™, an industry-first assessment that instantly quantifies the reliability of any data set.

Over 6,500 customers across the globe have chosen Talend to run their businesses on healthy data. Talend is recognized as a leader in its field by leading analyst firms and industry media.

For more information, please visit www.talend.com and follow us on Twitter: @Talend.

talend