Contents

Summary ................................................................. 4

Talend architecture .................................................. 5
  Talend Management Console .................................... 7
  Talend Data Inventory .............................................. 7
  Talend Data Preparation ........................................... 8
  Talend Data Stewardship .......................................... 9
  Talend API Designer ................................................ 10
  Talend API Tester .................................................... 10
  Talend Pipeline Designer .......................................... 11

Hybrid infrastructure .................................................. 12

Talend Cloud Data Fabric infrastructure ............................ 13
  Computation resources .............................................. 13
  Data processing and privacy ....................................... 13
  Data that we collect ................................................ 13
  Data that customers process with Talend Cloud Data Fabric .... 14
  Network perimeter .................................................... 14

Data flows ..................................................................... 15
  Data flows between Talend Studio and
  Talend Cloud Data Fabric ........................................... 15
  Data flows between Talend Remote Engine and
  Talend Cloud Data Fabric ........................................... 16
  Data flows between Talend Remote Engine Gen2 and
  Talend Cloud Data Fabric ........................................... 17
  Data flows in hybrid deployment between Talend Data Preparation,
  Talend Data Stewardship, and Talend Cloud Data Fabric .......... 18
Security at Talend

Physical security ............................................. 19
Security awareness training .............................. 19
Secure software development ............................ 19
Cloud workload protection and monitoring ............ 20
Authentication, access control, and audit ............... 20
  User access ................................................ 20
  Public API access ........................................... 20
  Administrative access .................................... 20
  Audit trails ............................................... 21
Encryption and Key management .............................. 21
  Unique Key pairing with Talend Management Console .............................. 21
  Job Artifact Signature .................................... 21
Vulnerability management .................................... 21
Disaster recovery and business continuity ................. 22
  Recovery strategy ......................................... 22
  Data backup & replication ............................... 22
  Monitoring ............................................... 22
  Testing .................................................. 22
Talend compliance and security certifications .......... 23
Talend Cloud Data Fabric (https://www.talend.com/products/data-fabric/) is a managed cloud integration platform that makes it easy for developers and data constituents to collect, transform, and clean data. 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 architecture

Talend Cloud Data Fabric is a multitenant integration environment that allows businesses to collect, transform, govern, and share data.

**Talend Cloud Data Fabric architecture distinguishes 3 layers that ensure by design best practices around security and privacy:**

- The Control Plane, known as Talend Management Console (TMC), manages all the administrative and operational aspects of the platform.
- The Data Plane performs all the data processing and is composed of execution engines managed either by Talend or by the customer.
- The Applications layer implements business domain logic (Job design, Data Quality ruleset, etc.).
All managed components are hosted on either Amazon Web Services or Microsoft Azure according to customer preference.

**Talend Cloud Data Fabric comprises the following applications:**

- Talend Management Console
- Talend Data Inventory
- Talend Data Preparation
- Talend Data Stewardship
- Talend API Designer
- Talend API Tester
- Talend Pipeline Designer

Additionally, Talend Studio, which runs on a local workstation, allows users to design data integration flows (or Talend Jobs) and publish them to Talend Cloud Data Fabric.

*Figure 1: Talend functional architecture*
Talend Management Console
Talend Management Console (TMC) is a browser-based application that provides access to all Talend Cloud Data Fabric applications and components as well as the administrative features and configurations that surround them.

TMC lets users schedule the execution of Talend Jobs via discrete components called execution engines. There are two types of engines:

- Cloud Engines are fully managed components that are provisioned, deployed, and controlled by Talend within our platform. Cloud Engines do not share jobs from multiple tenants; they are provisioned at execution time (per job schedule), per tenant.
- Remote Engines are execution agents deployed and managed by customers on their own systems, within their own physical or virtual (cloud) networks.

Talend Data Inventory
Talend Data Inventory provides automated tools for dataset documentation, quality proofing, and promotion. It identifies data silos across data sources and targets to provide visualization of reusable and shareable data assets.

Figure 2: Talend Data Inventory functional architecture
Talend Data Preparation

Talend Data Preparation (TDP) allows customers to simplify and speed up the process of preparing data for analysis and other tasks. TDP allows customers to create, update, remove, and share datasets, then create preparations on top of the datasets that can be incorporated into Talend Jobs with Talend Studio.

Figure 3: Talend Data Preparation functional architecture

Figure 4: Talend Data Preparation functional architecture in hybrid deployment
Talend Data Stewardship

Talend Data Stewardship (TDS) allows customers to collaboratively curate, validate, and resolve conflicts in data, as well as address potential data integrity issues.

![Figure 5: Talend Data Stewardship functional architecture](image)

![Figure 6: Talend Data Stewardship functional architecture in hybrid deployment](image)
Talend API Designer

Talend API Designer lets users design APIs collaboratively and visually, then run simulations to test APIs and generate reference documentation.

Talend API Tester

Talend API Tester lets users automatically generate test cases from API contracts, then field test APIs by grouping tests together that simulate real-world examples. Users can integrate unit tests into a managed CI/CD process to ensure quality.

Figure 7: Talend API Services functional architecture
**Talend Pipeline Designer**

Talend Pipeline Designer (TPD) allows customers to design and run data pipelines in the cloud.

- A data pipeline is a data integration process: a series of transformation steps applied to data. It extracts data from customer-specified sources, transforms it step by step using prebuilt processors, and loads it into other datasets (destinations).
- Data pipelines can be started directly from TPD or scheduled in Talend Management Console.
- Data pipelines can be executed on Cloud Engines or Remote Engines.

![Diagram of Talend Pipeline Designer functional architecture](image-url)
Hybrid infrastructure

Organizations can deploy Talend in a hybrid configuration, with some components running on-premises and others running on cloud platforms. The only required component for running Talend in a hybrid environment is the Talend Studio development environment, which is installed on local workstations. Users may install additional applications or components in a hybrid configuration:

- Talend Data Preparation
- Talend Data Stewardship
- Talend API Tester — web browser extension
- Remote Engine — Java-based runtime to execute Talend Jobs on premises or on a cloud platform that the customer controls
- Remote Engine Gen2 — a Docker-based runtime to execute Talend Pipeline Designer data pipelines or dataset sharing APIs on premises or on a cloud platform that the customer controls
Talend Cloud Data Fabric infrastructure

Each Talend Cloud Data Fabric customer has its own account to access the Talend environment. The account contains the number of users defined by the customer’s license. In the following section, “tenant” is equivalent to account; we use the terms interchangeably.

Computation resources

Talend Cloud Data Fabric is a multitenant platform and customers can set up isolated execution environments for computation resources.

- Remote Engines are deployed by customers on their own systems and therefore serve as computation resources that they manage and control.
- Cloud Engines are deployed within a Talend Cloud Data Fabric tenant-specific AWS or Azure Kubernetes cluster. Each tenant gets its own Cloud Engine instance.

The live preview feature of Talend Pipeline Designer, which allows users to preview the output of processors while designing a pipeline, is executed in a dedicated Remote Engine or Cloud Engine.

Talend Management Console, Talend Data Inventory, and Talend Pipeline Designer give separate computation resources to each tenant.

Data processing and privacy

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

**Data that we collect**

Talend, across its cloud applications, collects only the customer information needed to provide its services or to manage customer accounts.

All personally identifiable information that we collect (such as name, country, and email address) is protected with encryption at rest via AES-256 and in transit via HTTPS TLS 1.2.

Secrets such as passwords, keys, and certificates are managed via third-party technologies and products. For more additional information, see the Key Management section below.

No payment information is stored in Talend Cloud Data Fabric. We rely on third-party vendors to collect and manage payment information.
Data that customers process with Talend Cloud Data Fabric

Transfer of customer data to Talend systems depends on the product configuration selected by the customer. For customers with hybrid configurations or remote engine configurations, the Talend software resides on customer’s infrastructure, and customer data remains within the customer’s environment and systems at all times.

For cloud customers using a fully managed configuration (non-hybrid), customer data may be transferred to Talend systems depending on the Talend Cloud Services components used. Schedule A of Talend’s Assessment under CJEU Schrems II: Compliance with EU International Data Transfer Requirements document identifies the Talend Cloud Services components where customer data may be transferred to Talend Cloud.

For these components, please note that:

• There is no physical transfer of EU customer data outside of the EU: Talend Cloud services for EU customers are hosted on Amazon Web Services’ SSAE 16 certified data centers. Talend Cloud AWS’s primary data center is located in Germany, while its backup is in Ireland. Thus, EU customer data is stored in the EU at all times.

• Customers retain full control of the data transferred to Talend Cloud: Customers may delete their data from Talend Cloud at any time.

• Access to the Talend Cloud production environment is limited to our Site Reliability Engineering and Information Security teams, which abide by strict data access policies: Talend secure infrastructure is a closed network protected by multi-factor authentication and is accessible only to qualified members of our Site Reliability Engineering (SRE) and Information Security teams. All members of our SRE and Information Security teams have signed non-disclosure agreements and receive regular data privacy and security training.

Additional information with regard to cross-border data transfer and data processing can be found on https://www.talend.com/privacy/ and https://www.talend.com/legal-agreements/

Network perimeter

To function properly and deliver its services, Talend Cloud Data Fabric may need to communicate with external third-party solutions. All communications between Talend Cloud Data Fabric and such external solutions need to be authorized and initiated by Talend Cloud Data Fabric. No external solution can communicate with Talend Cloud Data Fabric unless the communication was initiated by Talend Cloud Data Fabric.


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

This section gives an overview of the data flows between Talend Cloud Data Fabric applications and components.


Data flows between Talend Studio and Talend Cloud Data Fabric

The types of data that can be exchanged between Talend Studio and Talend Cloud Data Fabric include:

- Task artifact binaries
- Task artifact metadata (such as context variables and parameters)
- Talend API Designer definitions

Figure 9: Talend Studio data flows
Talend Studio Jobs have the following components that can communicate with Talend Cloud Data Fabric:

- Data Preparation Job components — more details here
- Data Stewardship Job components — more details here

**Users’ credentials (login name and password or API token generated in TMC) are required to authorize the exchange.**

**Data flows between Talend Remote Engine and Talend Cloud Data Fabric**

![Talend Architecture – Remote Engines](image)

*Figure 10: Talend data flows when using Remote Engine*

Talend Cloud Data Fabric never initiates connections to Remote Engines. Remote Engines always initiate outbound connections to Talend. Once a connection is established, all data is sent encrypted over HTTPS.

Here are the types of data that can be exchanged between Remote Engines and Talend:

- Status information and metrics
- Lifecycle commands
- Task artifact metadata
- Job logs (optional)
- Task artifact binaries
Data flows between Talend Remote Engine Gen2 and Talend Cloud Data Fabric

Talend Remote Engine Gen2 is a new generation, secure execution engine based on container technologies that serve Talend Dataset Service, Talend Cloud Data Inventory, Talend Cloud Data Preparation, and Talend Cloud Pipeline Designer.

Talend Cloud Data Fabric never initiates connections to Remote Engines Gen2. Remote Engines Gen2 always initiate outbound connections to Talend. Once a connection is established, all data is sent encrypted over HTTPS and Secure WebSocket.

The types of data that can be exchanged between Remote Engines Gen2 and Talend are:

- Status information and metrics
- Lifecycle commands
- Task Artifact binaries
- Preparations tasks
- Pipeline designs
- Connection metadata
- Datasets (sampling or Data API)

Figure 11: Talend data flows when using Remote Engine Gen2
Data flows in hybrid deployment between Talend Data Preparation, Talend Data Stewardship, and Talend Cloud Data Fabric

**Guiding principle** — Talend applications and components always initiate outbound HTTPS connections. Talend Cloud Data Fabric never initiates any inbound connection to these applications. Here are the types of data that can be exchanged between hybrid applications and Talend Cloud Data Fabric:

- **a)** During user login: Client ID and client secret (as defined by the OIDC specification) of the installed application is used to authorize its communication with Talend Cloud Data Fabric.

- **b)** After user login: A JSON Web Token (JWT) that represents the user’s identity, metadata, and claims is transferred back to the application.
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 reviewed regularly and access is monitored continuously.

Security awareness training
Talend conducts security training programs annually and upon hire for all employees. System administrators receive training on their legal responsibilities with regard to security and data integrity. Talend conducts a complementary secure coding training program dedicated to Product and Engineering teams.

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 reviews, threat modeling, and code reviews. Automated security scans and testing such as Software Composition Analysis (SCA), SAST, and DAST 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 biennial external audits and pentests. Talend also runs a continuous Bug Bounty program.
Cloud workload protection and monitoring

Talend uses a combination of security services from third-party vendors to protect Talend Cloud 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 Security groups and Microsoft Azure Network Security groups to restrict inter-resource communication.

All code is tested in Talend’s lower environments before promotion to production (DEV, QA, Staging). Production is physically and logically separate from lower environments according to the principle of segregation of duties.

Talend Cloud Data Fabric’s perimeter security is composed of (but not limited to):

• Anti-DDoS protection — prevents distributed denial-of-service type of attacks
• 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) — alerts on rogue activity and protects against threats such as zero-day attacks
• Security information and event management system (SIEM) — monitoring and observability of system status and performance and detection of rogue processes
• Cloud Security Posture Management (CSPM) — continuously monitors cloud infrastructure to identify misconfiguration issues and compliance risks

Authentication, access control, and audit

User access

Tenant users are authenticated with their own unique set of credentials: username plus password by default. Talend also supports and recommends integration with external SAML-based single-sign-on (SSO) and multifactor authentication (MFA) providers. In addition, source IP-based access control can be applied to restrict access to Talend Cloud Data Fabric from unauthorized locations.

Public API access

Talend exposes public APIs that let developers automate workflows or auditors populate regulatory compliance reporting. Access to these APIs is secured with access tokens either bound to an individual (Personal Access Token) or bound to a programmatic client (Service Account).

Administrative access

The Cloud environment is totally separated from corporate IT resources and assets – only designated members of the SRE team can access the Cloud environment governed by the Principle of Least Privilege. Privileged access for the Cloud environment must be requested, is time-constrained, and only performed via a bastion host.
Administrators receive training on their legal responsibilities with regard to security and data integrity. New account creation follows a strict approval process. Accounts privileges are reviewed quarterly.

**Audit trails**

Talend Cloud Data Fabric provides always-on audit trail capabilities to help monitoring user activities. The audit logs are made available via a REST API. The logging service tracks all users and their actions in the system with the timestamps and outcome of those actions.

With this API, you can manage regulatory compliance risks by collecting and storing those logs on your own system. More details at https://help.talend.com/r/en-US/Cloud/api-user-guide/audit-logging

**Encryption and Key management**

Talend follows state of the art requirements on Encryption and Key management practices. As audited by third-party auditors, Talend has defined policies, procedures, and controls covering the entire Encryption Key lifecycle such as Key provisioning, rotation, destruction, and auditing.

Customer data and metadata stored on Talend infrastructure is always encrypted in transit with TLS 1.2 and at rest with minimum AES-256. A unique data encryption key (DEK) is used to encrypt tenant-specific information.

Talend relies on third-party key management services such as AWS KMS and HashiCorp Vault to manage the encryption key lifecycle, including key rotation.

Talend relies on trusted certificate services such as AWS Certificate Manager or Let’s Encrypt to provision, manage, and deploy SSL/TLS certificates.

**Unique Key pairing with Talend Management Console**

During the pairing process of a Remote Engine with Talend Management Console (TMC), a unique Key-Pair (RSA/4096 bits) is generated on the engine side. The public part of the key is shared to TMC for further information exchange, where it is used to encrypt any sensitive information that is sent to the Remote Engine.

**Job Artifact Signature**

To ensure workload integrity, Talend Studio signs task artifacts before deployment to Talend Management Console and signature verification is performed by Remote Engines prior to task execution.

**Vulnerability management**

Talend partners with an external vendor for Static Application Security Testing (SAST) and Software Composition Analysis (SCA). Their product is used to scan our software for security vulnerabilities in third party or community software and in our own code. Scans are automated and integrated in the development process of every Talend product.
Talend partners with an external vendor to conduct Dynamic Application Security Testing (DAST). Scans are automatically run weekly and monthly.

Talend partners with an external vendor for penetration testing against the OWASP standard. Talend also conducts a private bug bounty program to extensively test the security of our applications.

All detected vulnerabilities are analyzed by the Talend Information Security team, which then supports 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.

**Disaster recovery and business continuity**

**Recovery strategy**

Talend’s recovery strategy is a set of predefined actions implemented in response to a business interruption from a disaster. These can be natural (earthquake, flood, hurricane, fire, etc.) or cloud provider region-wide outages. Talend Cloud Data Fabric Disaster Recovery strategy is using the “pilot light” type approach. For faster recovery of the cloud network infrastructure, data backends and application infrastructure are pre-provisioned at the failover region and necessary capacity reservations are set. The pilot light site infrastructure remains mostly turned off.

**Data backup & replication**

All Talend Cloud data backends are backed up regularly and replicated continuously to the failover region.

**Monitoring**

Talend monitors all data backend backups and data replication to the failover region and follows up the backup status using an internal dashboard to ensure the RPO target is respected.

Latest uptime per region is available on [https://trust.talend.com](https://trust.talend.com).

**Testing**

Talend performs regular (at least, annual) tests of the below plans:

- Paper test: involved stakeholders review and update recovery plans
- Structured walkthrough: step by step review of disaster recovery plans and configurations
- “War Game Day” simulation: conduct scenario-based practice execution of plans
- Automatic biweekly backup data restoration and integrity tests at failover region
Talend compliance and security certifications
Talend is SOC 2, ISO 27001:2013 and ISO 27701:2019 certified and eligible to sign HIPAA (Health Insurance Portability and Accountability) Business Associate Agreements (BAA).

Talend uses the following programs and vendor platforms to assess the security maturity level and transparently share the ratings:

- Cloud Security Alliance (CSA) Security Trust Assurance and Risk (STAR) program
- BitSight Security Rating
- SecurityScorecard
- CyberVadis
- RiskLedger

A comprehensive list of security certifications and privacy compliance status is available on [https://www.talend.com/security/](https://www.talend.com/security/).
About Talend

Talend, a leader in data integration and data governance, 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 enabled the creation of innovations like the Talend Trust Score™, an industry-first assessment that instantly quantifies the reliability of any dataset.

Over 7,250 customers across the globe have chosen Talend to run their businesses on healthy data. Talend is recognized as a leader in its field by both leading analyst firms and industry media.

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