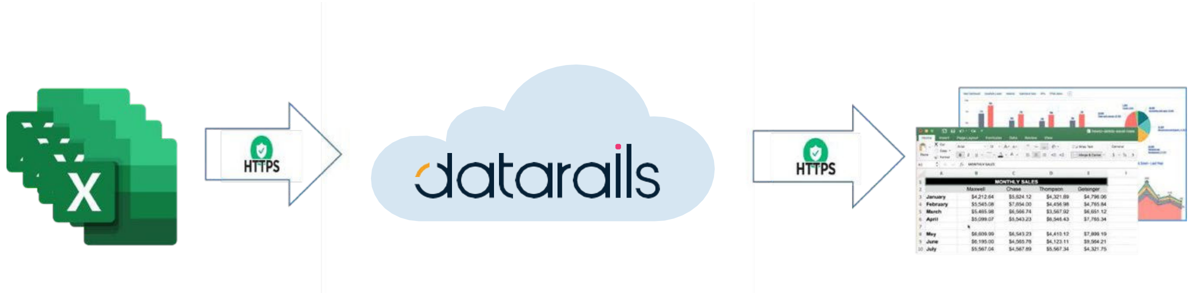
Datarails
Technical and Architectural Overview

2024

Technical and Architectural Overview

Contents

1. Solution Overview

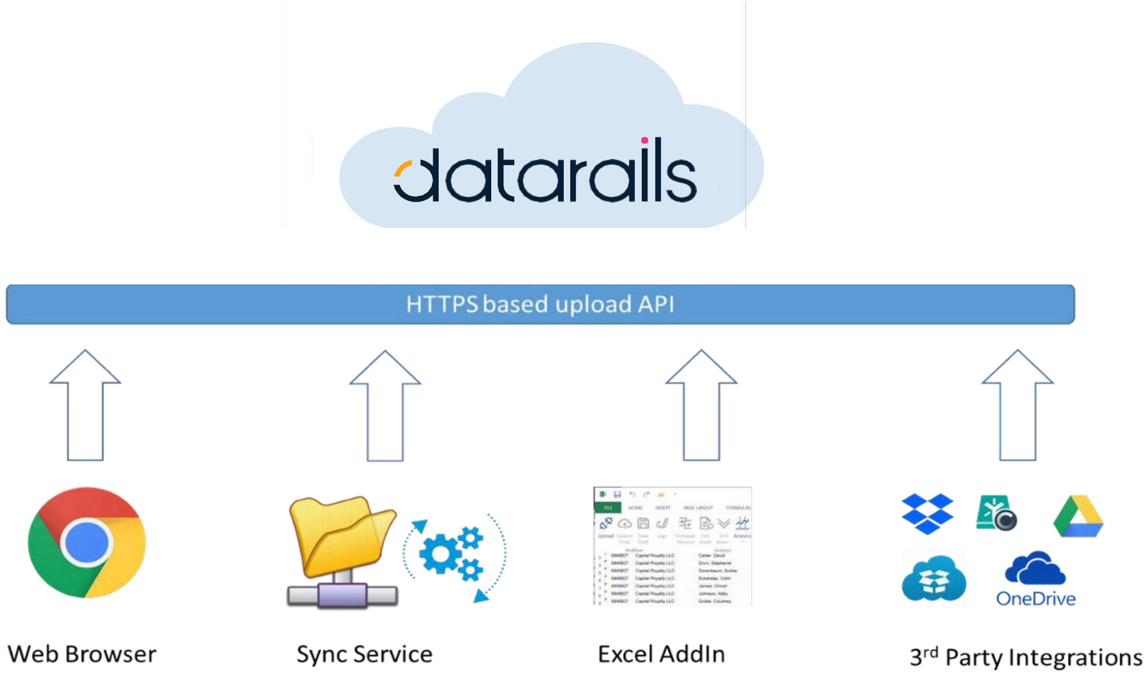


The Datarails solution is a cloud-based web platform, that is used to extract data from uploaded spreadsheets, and make it available to the customers in various ways via reporting, visualization, and audit information.

The main data flow of the Datarails Platform is as following:

1. Spreadsheets are uploaded into the platform using standard HTTPS protocol.
2. Data is processed on the Datarails servers, where data is extracted and stored.
3. The stored data can be consumed in various ways and tools, over standard HTTPS protocol.

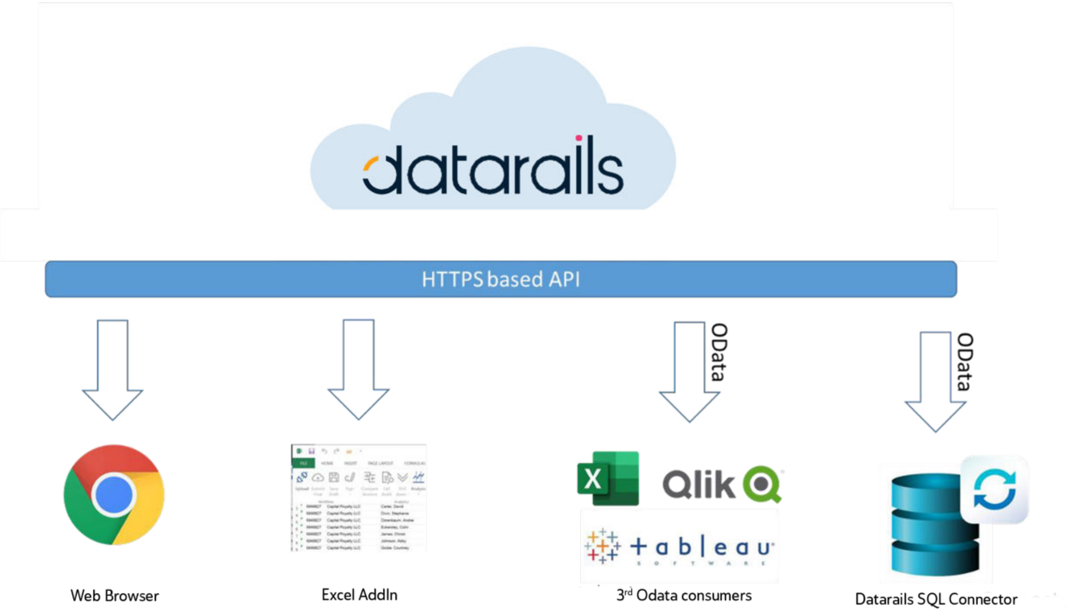
2. Uploading files into Datarails platform



Spreadsheet files are uploaded to the Datarails platform using an HTTPS based authenticated API, in one for the following ways:

1. Manually: An authenticated and logged in user can upload files into Datarails platform through the Datarails web interface.
2. Sync Service: A windows service that can be installed on any Windows machine, and that can sync files automatically to the Datarails platform. The sync service can be configured to monitor changes to specific files and folders.
3. Excel Add-in: The Datarails Excel Add-in is a simple Windows platform that is installed on the user's endpoint. This platform allows authenticated and logged in users both to consume data from Datarails platform as well as to sync files to the Datarails platform on a click-of-a- button.
4. Datarails offers several 3rd party integrations with common file-sharing vendors. When using such integration, a "server to server" communication is initiated to sync files from the 3rd party service into Datarails platform.

3. Consuming Data from Datarails Platform



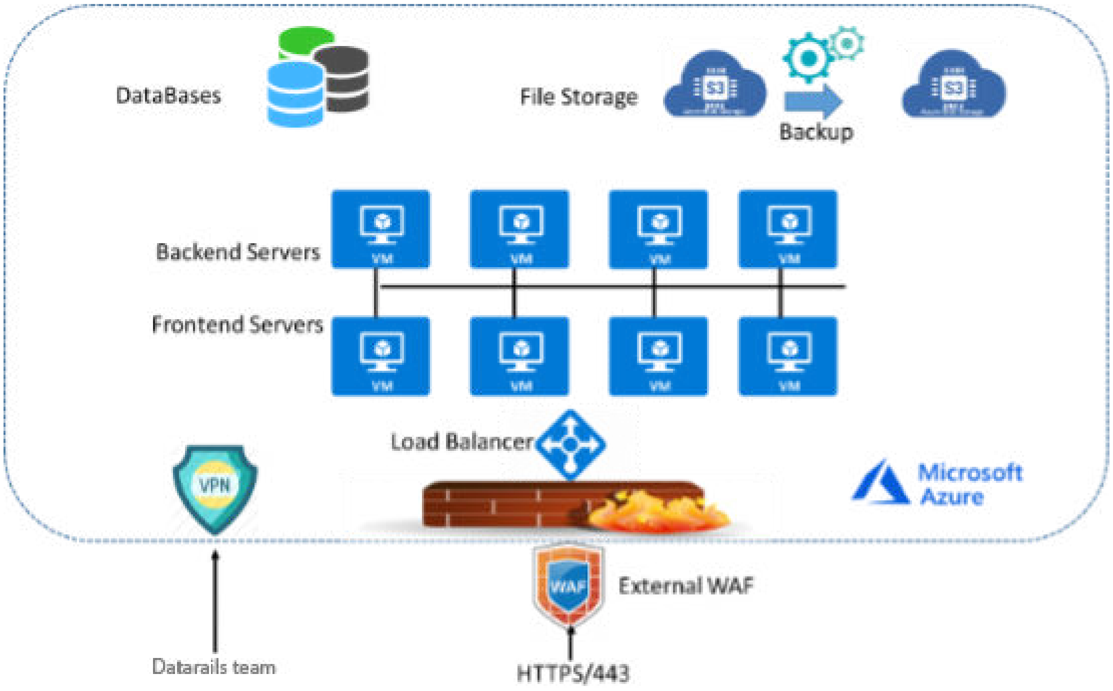
Interfacing and data consumption from Datarails platform is done via HTTPS based REST APIs in one or more of the following ways:

1. Using a standard web browser to interact with the Datarails web interface at http://app.datarails.com
2. Using the Datarails excel Add-in to import data using standard MS-Excels tools.
3. Consuming data via authenticated OData protocol (https://www.odata.org/) – this is a standard XML based protocol for querying Web-hosted databases, developed by Microsoft. This protocol is natively

supported within Microsoft Excel (no installation required) and in most common BI tools.

4. The Datarails SQL Connector is a simple windows service, which can be installed on any Windows machine within the organization's network and can be used to pull data from the Datarails platform (via OData protocol) and push it into an internal SQL database.

4. Cloud Infrastructure



The Datarails platform is hosted on Microsoft Azure cloud and includes the following components:

1. Frontend servers – in charge of serving HTTPS request from various clients. The traffic to the frontend servers is managed via an HTTP load balancer.
2. Backend servers – in charge of extracting, processing, and storing the data from the uploaded spreadsheets.
3. The uploaded files are stored in Azure blobs storage (each file is replicated 6 times in two geo-separated datacenter). A backup process also copies the uploaded files to a secondary blob once a day.
4. Several databases are used by the system:
    a. Operational database for the web platform stores information about accounts, users, permission, and metadata for the uploaded files. The operational DB is a MS managed database service, with a continuous backup providing a point-in-time restore option for 14 days.
    b. Analytic database is used for data aggregation and visualization. This database is replicated for redundancy and data resiliency. The data resides on this database can be re-generated from the

files + data residing in the operational database.

c. Audit database is used for storing all changes made to each synced file .This database is replicated for redundancy and data resiliency. The data resides on this database can be re-generated from the files + data residing in the operational database.

5. A private network is used for communicating between the various components of the network. The only component with a public IP address is the load balancer, which is configured via firewall IP-whitelisting to allow inbound traffic only from the external WAF.
6. Access to all other infrastructure components is allowed only via VPN tunnel using a dedicated client certificate.
7. CloudFlare WAF is used in front of the load-balancer for providing additional layer against DDOS and other common web attacks.

5. Security Highlights

Infrastructure Security

1. Platform is hosted on MS Azure, which complies will all known security standards for IAAS and PASS platforms.
2. All the cloud infrastructure resides within a private network, accessible only via VPN tunnel.
3. Infrastructure access is enabled via Azure IAM to minimal number of Datarails technical stuff. Access to infrastructure is authenticated and authorized via Azure AD and requires multi-factor authentication.
4. Azure security center is user for monitoring infrastructure threats and applying relevant security patches.
5. Firewall is used to allow only a communication between an external WAF (CloudFlare) to the Azure load balancer (via HTTPS).

Network Security

1. Web platform firewall (CloudFlare) is used against DDOS and other common web attacks.
2. IP whitelisting can be enabled to allow customer access their data only from a predefined IP range.
3. Admin access (by Datarails administrator) is blocked in the network layer by CloudFlare access and requires multi-factor authentication for network access.

Data Encryption

1. Encryption in Transit: all communication with the Datarails endpoints is done via HTTPS using strong cyphers.
2. Encryption in rest:
   a. The operational database uses storage AES-256 encryption for data in rest and is FIPS- 140-2 compliant.
   b. Audit and analytic database use disks encrypted by AES-256 encryption.
   c. Uploaded and intermediate files are encrypted by the platform using Fernet algorithm (AES128 in CBS mode and PKCS7 padding, using HMAC using SHA256). And stored on

Azure blobs storage which is encrypted by AES-256 encryption.

3. Passwords and secrets management:
    a. User passwords are not stored in Datarails servers, a hashed key using pbkdf2_sha256 algorithm is stored instead
    b. Platform secrets (encryption keys, db credentials, etc.) are stored on Azure Valut service.

Authorization and Authentication

1. All calls to the Datarails platforms must be authenticated, both basic and session-based authentication can be used.
2. Authorization can be done either via the platforms built-in authorization module or via integration to Azure-AD (integration with other SSO providers is possible).
3. Build-in authorization module:
    a. Strong passwords policy is enforced (8 characters minimum, must include upper and lower case, as well as special characters. Common passwords are also not allowed.
4. The platform includes a flexible permission mechanism, with different access levels, enabling the access control for each platform entity.
5. A proactive permission policy is in place, initializing each entity (such as uploaded file) with the minimal possible permission, and only the owner of the object can grant permission to other users.
6. Session time out is used for logging out users that are not active more than 3 days.
7. IP whitelisting can be enabled to allow customer access their data only from a predefined IP range.

Platform Security

1. The following HTTP headers are used: x-xss-protection, x-content-type-options, strict-transport-security, X-Frame-Options.
2. DB access by the platform is done using an ORM, for preventing SQL injection
3. Cross Site Request Forgery (CSRF) mechanism is used for all PUT and POST calls.
4. All inputs are validated by the platform against XSS attacks, all outputs are automatically escaped by the platform.

Data Backup and Recovery

1. All files are stored in a geo-redundant blob storage, replicated in 6 locations over two geo-separated data centers.
2. A backup process is used daily for copying all new files into a separate (redundant) blob storage.
3. Operations database has a continuous backup mechanism and enables a point-in-time restore backup for 14 days.
4. Analytic and audit databases are back upped using replication and can be restored from the files and operation database as well.
5. There is a well-defined documented procedure for data restore in cases of data lost.

Datarails Personnel Access

1. Datarails support team can access customer's accounts for the purpose of support and maintenance, however, has no access to the files uploaded by the customers.
2. Datarails support team members authorization is managed via Azure active directory and requires multi-factor authentication.
3. Admin access privileges are limited to qualified personnel and require access via multi-factor authentication and VPN. No admin endpoint is accessible via the public web.

Certification

1. Datarails is certified for SOC2 type II by EY. The SOC2 auditing process is performed on an annual basis.
2. 3rd Party Penetrations tests are performed on an annual basis, both in platform and infrastructure levels.