



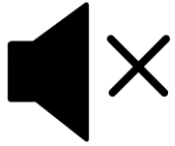
Feb 06, 2024

APIs in Cloud Data Quality Governance and Profiling with Case Studies

- Khadarvali Shaikh, Senior Solution Architect.
- Rakesh Varma Kucharlapati, Principal Solution Architect.
- Avanish Srivastava, Manager, Solution Architect.

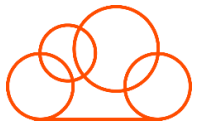
Where data & AI come to **LIFE**

Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our [Success Portal](#) - where you can download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

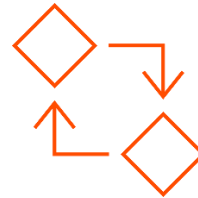
Feature Rich Success Portal



**Bootstrap trial and
POC Customers**



**Enriched Customer
Onboarding
experience**



**Product
Learning Paths
and Weekly
Expert Sessions**



**Informatica
Concierge**



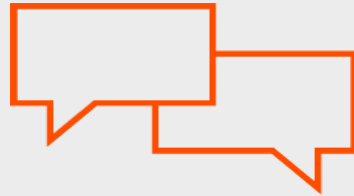
**Tailored training
and content
recommendations**

More Information



Success Portal

<https://success.informatica.com>



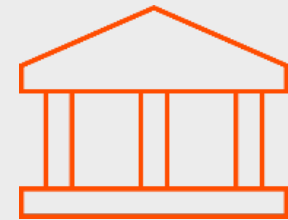
Communities & Support

<https://network.informatica.com>



Documentation

<https://docs.informatica.com>



University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

Agenda

1 Introduction to Cloud Data Governance Rest API's

2 Introduction to Cloud Data Profiling Rest API's

3 Case Study for Governance APIs

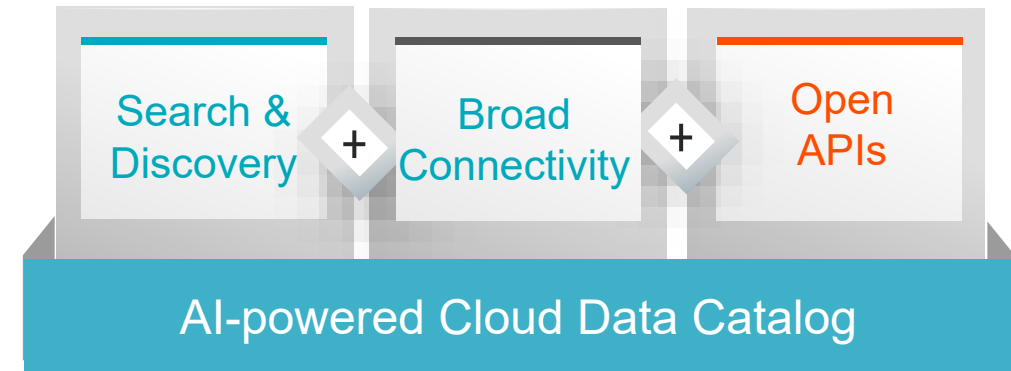
4 Case Study for Profiling APIs

5 Demos

6 Q&A

Cloud Data Governance APIs

- **Why APIs:** APIs are the connectivity tissues to streamline data governance, automating processes, enforcing standards, and ensuring seamless external integrations for effective management and quality assurance
- **What APIs:**
 - Create and Manage assets
 - Search Assets
 - View assets details
- **How:** Through IDMC secured API auth layer with following common approaches
 - Informatica CAI
 - Any Programming interface
 - Curl tools like Postman



CDGC Public APIs with Endpoints

| TYPE | API | CALL | ENDPOINT | DESCRIPTION |
|--------------------------|-----------------------------|------|--|--|
| Search API | Asset List API | POST | /data360/search/v1/assets | Provides list of Assets based on search query or body parameters |
| | Asset Details API | GET | /data360/search/v1/assets/{id} | Provides details of an Asset. |
| | Multiple Assets Details API | POST | /data360/search/v1/assets/details | Provides details of multiple Asset passed in body parameter. |
| Bulk Import / Export API | Bulk Export API | POST | /data360/search/export/v1/assets /data360/search/export/v1/assets/details | Search and export assets in a file based on query or body parameters (specific assets using Asset Ids in /assets/details API). |
| | Bulk Import API | POST | /data360/content/import/v1/assets | Import assets from a file. |
| JOB API | Job Monitoring API | GET | /data360/observable/v1/jobs/{id} | Monitor a CDGC Job. (Import / export Jobs, Catalog Source Jobs) |
| | Download Exported file | GET | /data360/observable/v1/jobs/{id}/outputProperties/files/<outputfile properties Key : Name> | Download exported file |

CDGC Public API – Limit & Pagination

Rate Limits regulate the number of API calls each tenant can make within a specific time frame, preventing any single tenant from overwhelming the system.

- Rate Limit :
 - 120 API calls/minute
 - 10000 API calls/day

When retrieving many assets, pagination is used to retrieve the results in smaller chunks. API pagination ensures efficient data retrieval, managing large datasets effectively. Thus, single API call may not give all the results.

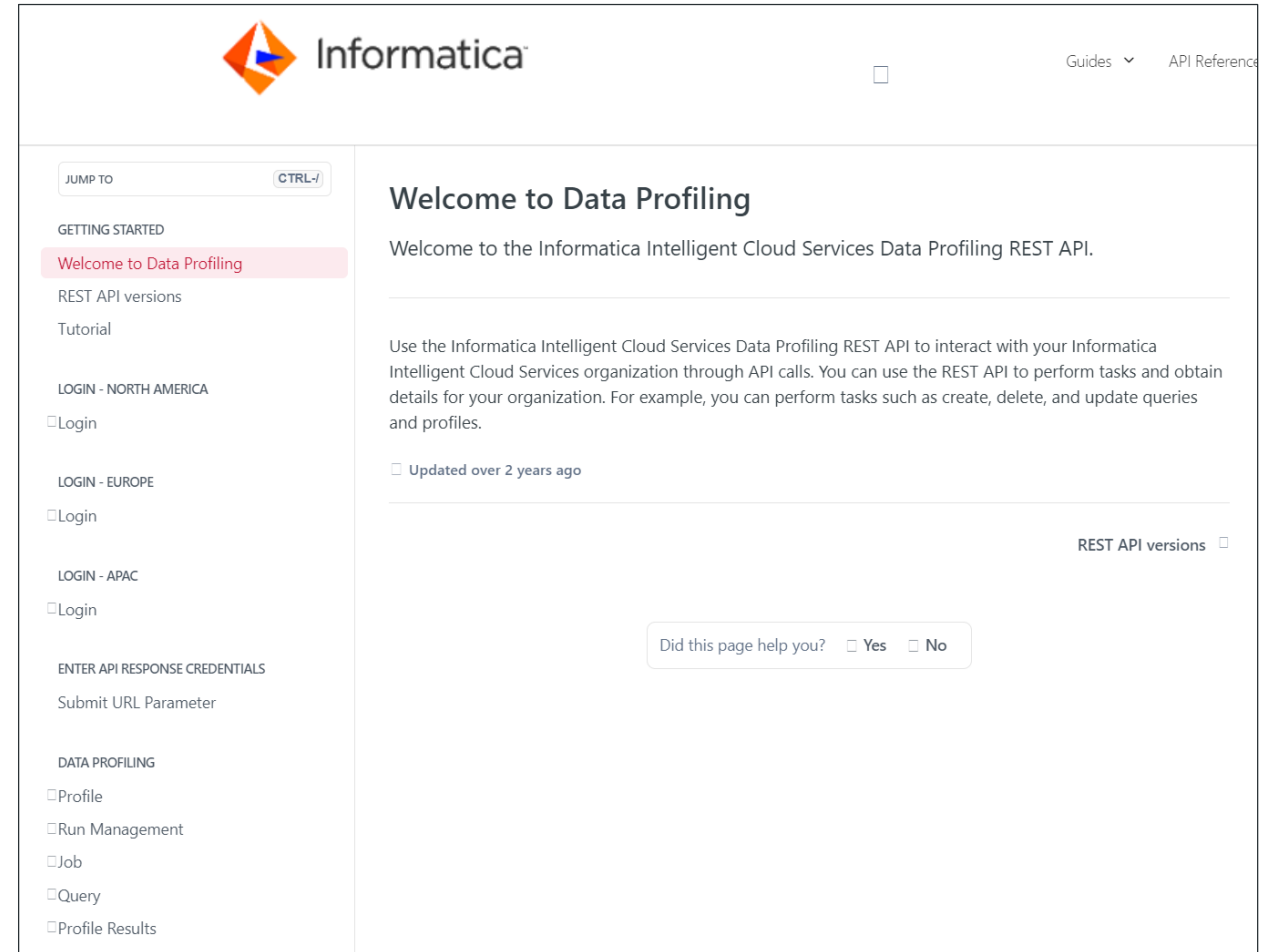
- Search API will return max 100 assets per page.
 - 1 API call per 100 assets
- Bulk Export API
 - 1 API call per 100 assets exported
- Bulk Import API
 - 1 API call per 25 assets imported

CDGC Public API - Metering

- CDGC Public API is GA in 2023.08 release
 - Metadata consumption API (Search API, Bulk Export API, Bulk Import API), Job Monitoring API
 - Public API Rate limit are applied to ensure level of services for all tenant on the pod.
- Metering
 - Public API usage is metered, pricing & details are published in PDS & measurement guide (Metadata record consumption)
 - **Metadata API are Metered**
 - First 100000 API calls/month are included.
 - 0.32 IPU / 1000 API calls beyond 100K API calls
 - **Admin API are not metered.** API such as :
 - Job Monitoring API
 - Catalog Source Execution API
 - Configuration API (future release)
 - Connection Assignment API (future release)

Cloud Data Profiling APIs

- **Why APIs:** Data profiling APIs streamline access to stored results, fostering seamless integration and actionable insights. They enhance interoperability, enabling efficient retrieval of comprehensive data profiles for informed decision-making and governance.
- **What APIs:**
 - Create and Manage Profiles
 - Execute and Monitor Profiles
 - List Profiles and Retrieve Profile results
- **How:** Through IDMC secured API auth layer with following common approaches
 - Informatica CAI
 - Any Programming interface
 - Curl tools like Postman



The screenshot shows the Informatica Data Profiling REST API documentation page. The page features the Informatica logo at the top left and navigation links for 'Guides' and 'API Reference' at the top right. A 'JUMP TO' search bar is located at the top of the left sidebar, with a 'CTRL-F' icon. The sidebar is organized into sections: 'GETTING STARTED' (with 'Welcome to Data Profiling' highlighted), 'REST API versions', 'Tutorial', 'LOGIN - NORTH AMERICA' (with a 'Login' link), 'LOGIN - EUROPE' (with a 'Login' link), 'LOGIN - APAC' (with a 'Login' link), 'ENTER API RESPONSE CREDENTIALS' (with a 'Submit URL Parameter' link), and 'DATA PROFILING' (with links for 'Profile', 'Run Management', 'Job', 'Query', and 'Profile Results'). The main content area is titled 'Welcome to Data Profiling' and contains a welcome message: 'Welcome to the Informatica Intelligent Cloud Services Data Profiling REST API.' Below this, it explains that the REST API is used to interact with the Informatica Intelligent Cloud Services organization through API calls, and lists tasks such as creating, deleting, and updating queries and profiles. A 'REST API versions' link is visible on the right side of the main content area. At the bottom right of the main content area, there is a feedback form: 'Did this page help you?' with 'Yes' and 'No' options.

CDP REST API – Endpoints

| TYPE | API | CALL | ENDPOINT | DESCRIPTION |
|------------------|--------------------------------|------|--|--|
| List All Profile | Profile List API | GET | profiling-service/api/v1/profile/{IDS-SESSION-ID} | Provides list of profiles. |
| Create Profile | Create Data Profile | POST | /profiling-service/api/v1/profile/{IDS-SESSION-ID} | Create Data Profile Task |
| Run Profile | Run Profile task | POST | profiling-service/api/v1/profile/{id}/execute | Run the profile to extract column statistics, patterns, data types, value frequencies, and outliers for columns and rules. |
| Monitor Profile | Monitor Job for profiling Task | GET | /profiling-service/api/v1/job/{id} | Monitor to learn how to view and monitor jobs that are running or have run in Data Profiling. |
| Profile Results | View Profile Results | GET | /metric-store/api/v1/odata/{Profiles('profileId')}/Columns | Displays all the columns and rules and their statistics in the profile run. |
| CLAIRE Insights | Enable CLAIRE Insights | POST | /profiling-service/api/v1/profile | Use the POST method to create a profile with CLAIRE configuration. You can enable CLAIRE to generate insights recommendations for your data that you can approve or reject |

References and useful documentation links :

- CDGC API Article with Postman Collections

https://knowledge.informatica.com/s/article/CDGC-Public-API-Postman-Collection?language=en_US

- Product Documentation for CDGC APIs

<https://docs.informatica.com/data-governance-and-quality-cloud/data-governance-and-catalog/current-version/api-reference/introduction.html>

- Cloud Data Profiling Documentation

<https://docs.informatica.com/data-governance-and-quality-cloud/data-profiling/h2/1547-getting-started-with-cloud-data-profiling-rest-api/abstract.html>

- Developer's portal to try Cloud Data Profiling APIs

<https://developer.informatica.com/tpn/v4/reference/welcome>

- API based solution for Data Profiling and Quality Reporting

https://marketplace.informatica.com/listings/cloud/solutions/CDQ_Report_Dashboard_n_Profile_Warehouse_Template.html

- Informatica Cloud Platform and other API documentation

<https://docs.informatica.com/cloud-common-services/administrator/current-version/rest-api-reference/preface.html>

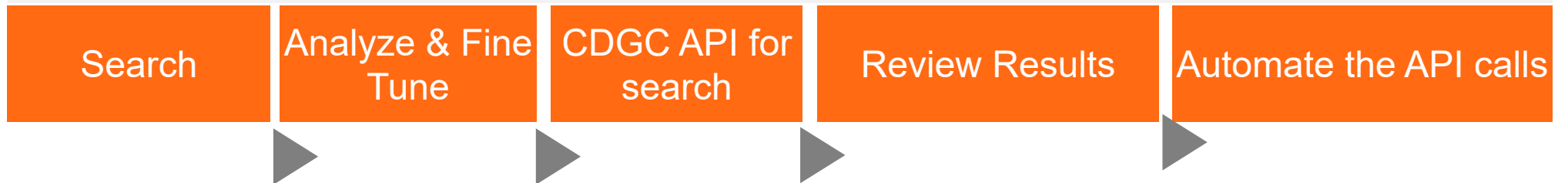
CDGC REST API Case Study



Data Modeler

Mary is a data modeler working for an MNC. In their organization, they have a centralized custom application to maintain all aspects of Enterprise architecture and related assets, such as schemas, tables, and columns. As part of a Data Governance program, the organization adheres to industry standards, policies, processes, legal requirements, and data protection laws to mitigate risks and ensure data security. This custom application is seamlessly integrated with their data governance and management system, getting updated whenever new data sources (schemas, tables, columns) are added to their enterprise systems. The governance team scans, catalogs, and approves the new data sources for use.

In this case study, we will explore how CDGC Rest API's and queries play a major role in integrating their custom application with their governance system, helping data modeler Mary access the latest and governed data assets to design and build new data models.



CDP & CDQ Rest API Case Study



Data Analyst

Bob is a Data Analyst working for one of the largest health insurance providers in the country, and he is part of the Claims department. Dealing with critical health information, they strictly follow Data Governance principles.

As part of the Data Governance program, their organization adheres to industry standards, policies, processes, legal requirements, and data protection laws to mitigate risks and ensure data security. Bob's primary job duties are to ensure that health insurance claims data, which comes from heterogeneous data systems, is accurate and up-to-date, as this information is critical to processing customer claims\bills.

In this case study, we will see how Bob utilizes Cloud Data Quality and Profiling services to evaluate the quality of the claims\bills data and then uses Rest APIs to access these results on their healthcare system, which is seamlessly integrated with their governance system(CDP/CDQ/CDGC) to process the claims\bills.

Prepare data from
different systems

Create Profile

CDP/CDQ API
for search

Review Results

Automate the
API calls



We would love to hear from you

- supportvideos@informatica.com
- <https://twitter.com/INFAsupport>

Thank You

Where data & AI come to **LIFE**