

 Qalytics

# PRODUCT UPDATE

End of Year: 2023 Edition



Qualytics

## PRODUCT UPDATE

Over the course of the last few months, the Qualytics team has maintained our commitment to providing enterprise businesses with best-in-class data quality innovations, while significantly enhancing our existing functionalities. This update highlights net-new capabilities, navigation enhancements, reporting improvements, and operational efficiencies that we've built into the product in Q4 2023.

70% of these updates came directly from customer suggestions, so, as always, your feedback is welcome and appreciated.

# TABLE OF CONTENTS

## 04 NEW CAPABILITIES

a. Check Templates	04
b. Metric Checks	05
c. Export Metadata Directly to Enrichment Datastore	05
d. New Check - IsAddress	06
e. New Check - Aggregation Comparison	06
f. Command Line Interface	06
g. System Health Dashboard	07
h. Tag-based Operations	07
i. Notification Testing	07
j. Three New Datastore Connectors: Athena, Presto, and Trino	07

## 08 NAVIGATION ENHANCEMENTS

a. Tree View	08
b. List View	08
c. Slide Navigation	08
d. Heatmap Filtering for Operations	09

## 09 REPORTING IMPROVEMENTS

a. Enhanced Activity Reporting	09
b. Insights Dashboard Improvements	10
c. Check and Anomaly Categories	10
d. Advanced Check Surveillance	11
e. Source Records Display Enhancement	12
f. Overhauled In-App Notifications Panel	12
g. Dynamic Quality Score Based on Anomaly Resolution	12

## 13 OPERATIONAL EFFICIENCIES

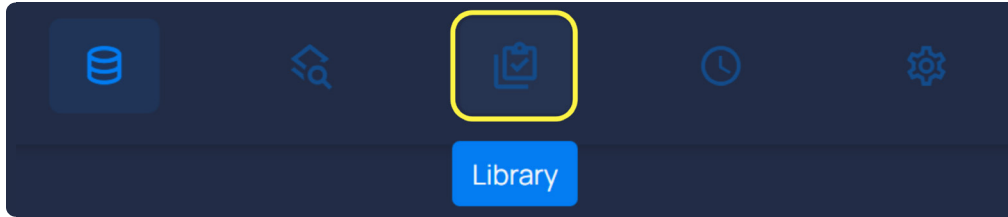
a. Specify Teams Upon Datastore Creation	13
b. New Bulk Actions	13
c. User-Defined Source Record Limits	13
d. "Select All" Fields for Checks with Multi-Select Picklist	14
e. Profile Based on Incremental Identifiers	14
f. Conditional Notifications Based on Weight	14

## 15 PRODUCT ROADMAP 2024

# NEW CAPABILITIES

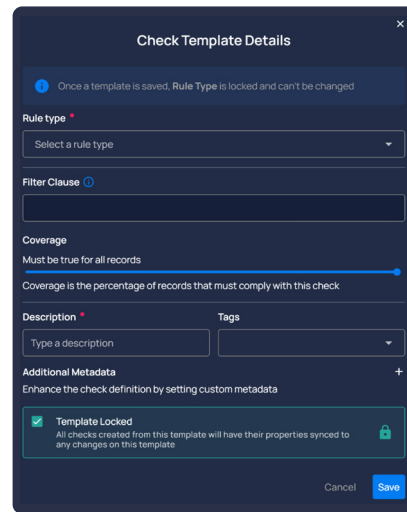
## Check Templates

We are excited to announce the release of **Check Templates**. Through this new way of check authoring, **users can centrally manage authored checks**, allowing for distributed management of individual checks while retaining centralized control over check parameters.

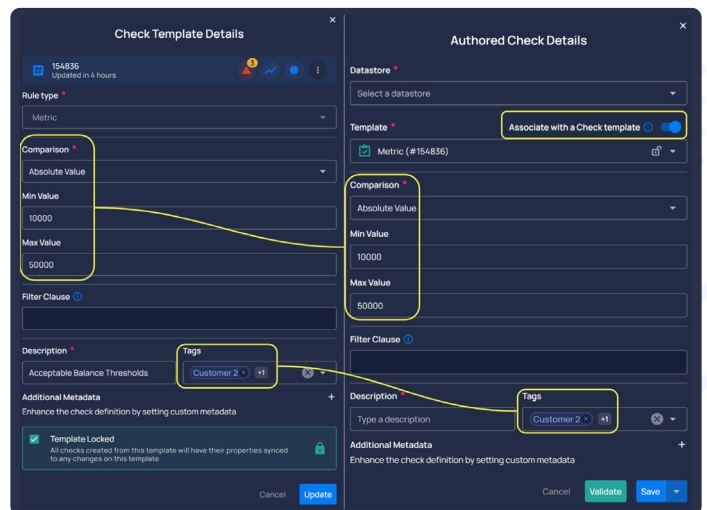


Within the library, users can view details about the template, such as associated tags, a count of checks associated with the template, number of active anomalies those checks have produced, and other relevant information such as template parameters.

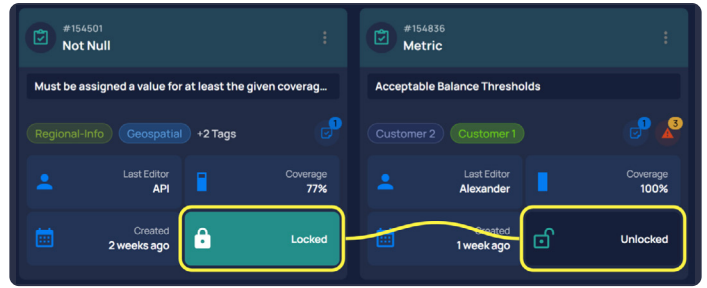
Check templates are configured similarly to individual checks; a user selects the **Rule type**, any desired **Filters**, defines **Coverage**, enters a **Description**, and applies **Tags** and any Additional **Metadata**. The main difference is that **Datastore**, **Table**, and **Field** designations are not present, as they will be specified on the individual checks created from the **Template**.



**There are two ways to leverage Templates when authoring a new check.** First, by creating a **check directly** from the Template itself, or **by toggling** on “Associate with a Check template” when creating a check anywhere else in the platform. Either way, the Rule type, any associated Tags, and the check parameters (in the below example, Min and Max values for an Absolute Value Metric Comparison) will be **automatically inherited** from the Template by the newly created check.



Finally, **Check templates** have the concept of being **Locked vs Unlocked**. This setting determines two things. First, a locked check Template **will not allow** for parameters of underlying checks associated with that Template to be changed. In the above example, the user would need to unlock the Check Template before the Min and Max values of an associated check could be altered. Second, a locked check Template **will automatically apply** any changes to all associated checks, giving users the power to dynamically adjust entire check populations from a single centralized location.

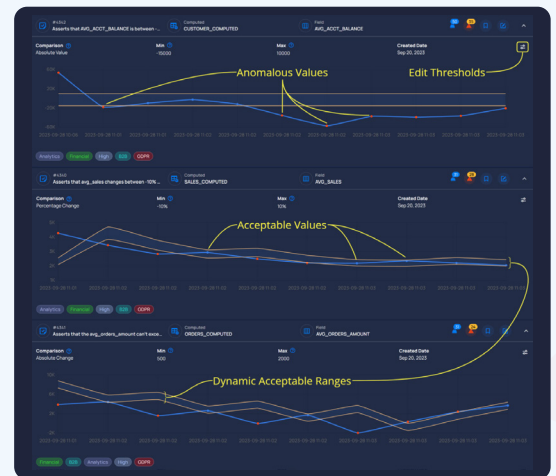
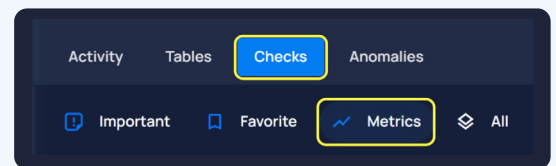


### Metric Checks

We have added time series visualizations in support of **Metric Checks**, which allow users to define, in absolute or percentage terms, **minimum** and **maximum** acceptable changes to a field's value, and identify when records **change more or less** than expected from scan to scan. They also allow for a lifetime minimum and maximum, and will identify if a value ever falls outside the defined range.

Metric Charts track these measurements over time, which are available under a newly created **"Metrics"** tab, which can be accessed via the Explore page or under "Checks" within a particular datastore.

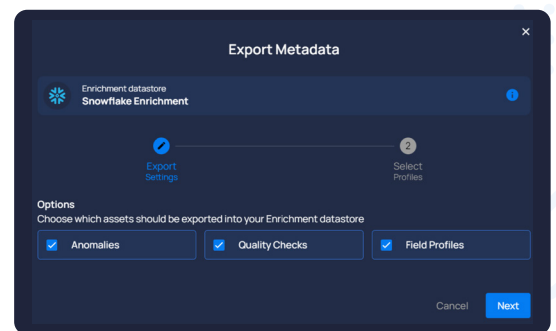
In addition to tracking the values and how they measure up against acceptable values over time, the UI has been updated to allow on-the-fly adjustments of acceptable thresholds.



### Export Metadata Directly to Enrichment Datastore

Previously, access to metadata was available exclusively via API. We now support the ability to export calculated metadata directly to your Enrichment store through our UI, eliminating the need to write code to obtain this information at scale.

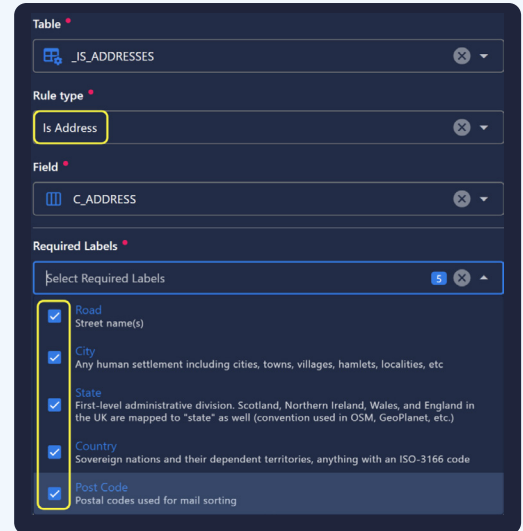
Additionally, users have complete control over what kind of metadata is exported. In the **Export interface**, users can select **metadata pertaining to Checks, Profiles, or both**, and further refine the extraction criteria by specifying individual tables, ensuring that **only relevant information is exported to the enrichment store**.



## New Check - IsAddress

This new feature ensures that **address fields** conform to required components. Users can enforce the correctly formatted presence of data representing a Road, City, State, Country, Post Code, or any combination thereof via multi-select picklist.

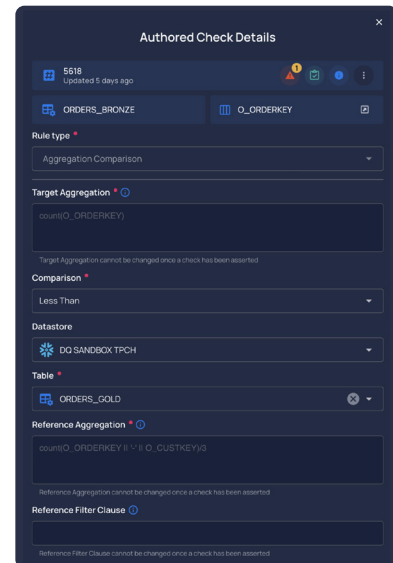
This check leverages machine learning to support multilingual, global street address parsing and normalization. It was trained on over **1.2 billion records of real address data from every inhabited country on earth, in over 100 languages**.



## New Check - Aggregation Comparison

This new rule capability allows for dynamic **analysis of aggregations** across datasets, empowering users to ensure that aggregate values meet expected comparisons, whether they are **totals, averages, counts or any other aggregated metric**.

By setting a comparison between aggregates from potentially different tables or even source datastores, this rule confirms that relationships between data points adhere to business logic or historical data patterns. This is particularly useful when trying to validate **interrelated financial reports, summary metrics, or when monitoring the consistency of data ingestion over time**.



## Command Line Interface

To better serve our power users who wish to leverage the Qalytics API, we now have our own **Command Line Interface (CLI)**. As you know, the Qalytics API is fully fledged and enables advanced capabilities - for example, it can be used to seamlessly copy existing checks from one datastore to another, run iterative operations at scale, and much, much more. Now, you can leverage the **CLI directly without having to write additional code** to hit the API. Check it out [here](#).

```
shindiogawa@shindiogawa > qalytics checks --help
Usage: qalytics checks [OPTIONS] COMMAND [ARGS]...

Commands for handling checks

Options
--help          Show this message and exit.

Commands
export          Export checks to a file.
import          Import checks from a file.

shindiogawa@shindiogawa >
```

```
shindiogawa@shindiogawa > qalytics checks --help
Usage: qalytics checks [OPTIONS] COMMAND [ARGS]...

Commands for handling checks

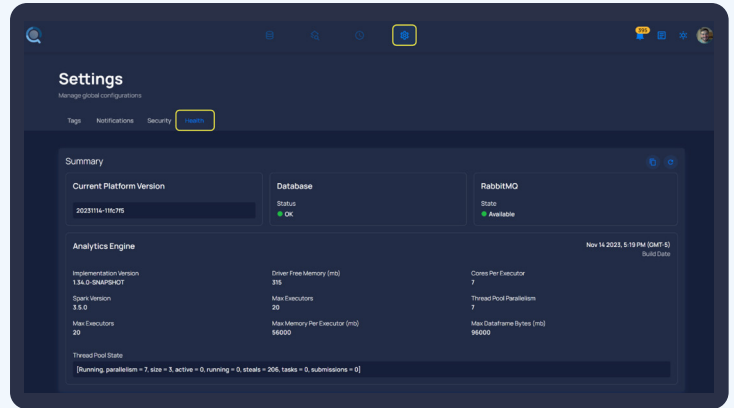
Options
--help          Show this message and exit.

Commands
export          Export checks to a file.
import          Import checks from a file.

shindiogawa@shindiogawa >
```

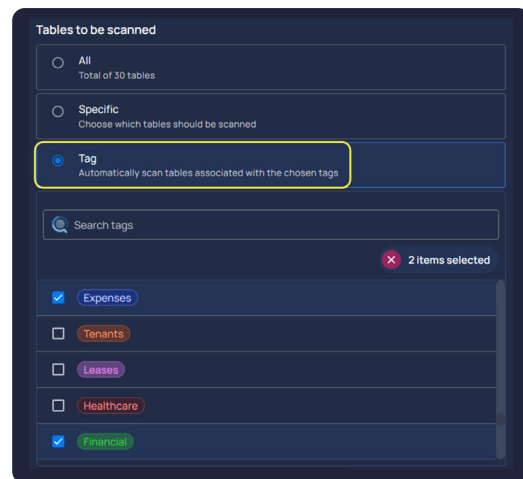
## System Health Dashboard

A new **Health Dashboard** has been added to the **Admin menu**, offering a comprehensive view of each deployment's operational status. This feature encompasses critical details such as the **status of app services, current app version, and analytics engine information.**



## Tag-based Operations

**Scan and profile operations** can now be **triggered and scheduled** to target the tables associated with tags of the user's choosing.

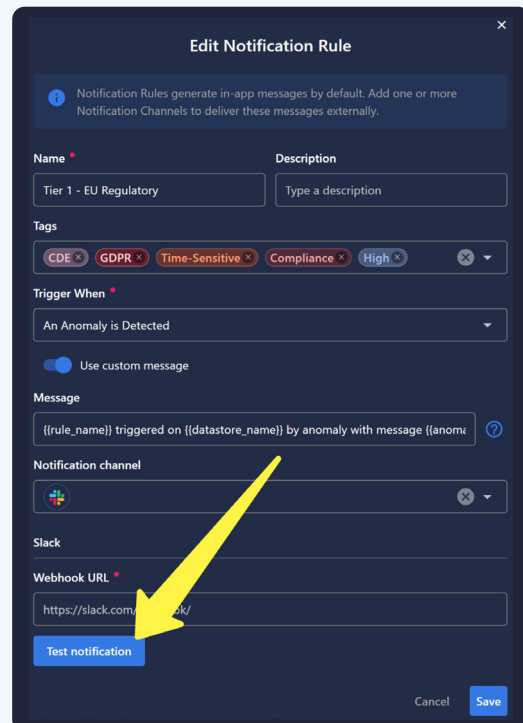


## Notification Testing

We have added a **"Test notification"** button to the notification interface, which will **trigger an ad-hoc test notification**, ensuring that the right people get the right information at the right time from the moment of configuration. We will also display any error messages that the system receives in the event of a failure, significantly streamlining the testing and validation process of downstream notifications.

## Three New Datastore Connectors: Athena, Presto, and Trino

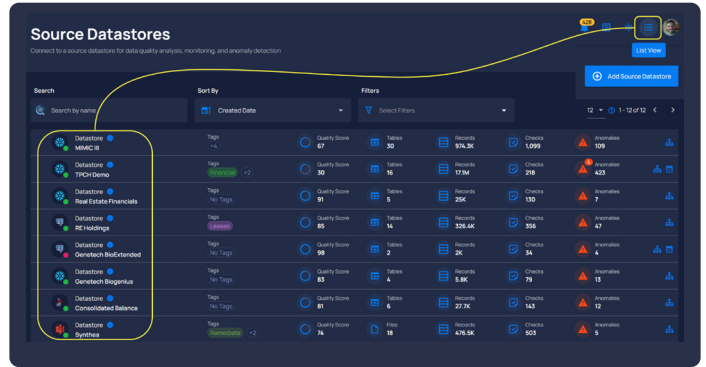
We continue to add to our population of **datastore connectors**, and are pleased to introduce **Athena, Presto, and Trino** - bringing our total number of supported platforms to **21**.



# NAVIGATION ENHANCEMENTS

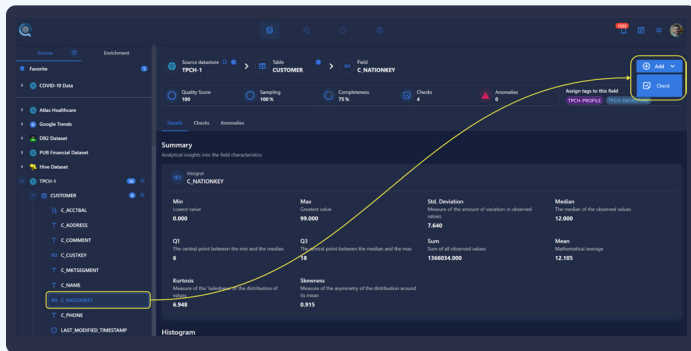
## List View

Introduced **list view** layouts for **Datstores, Profiles, Checks, and Anomalies**, providing users with a more compact display for navigating their data quality elements.



## Tree View

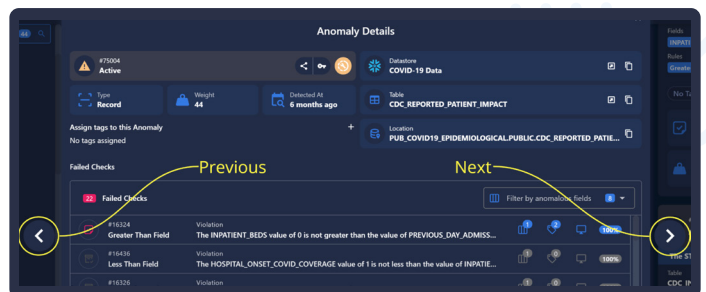
New **simplified, streamlined navigation** organizes your data assets into intuitive, nested, searchable dropdowns, allowing users to easily drill down to tables, files, and even fields that are of interest. Within the **Tree View**, users are dynamically presented with stats and options that are relevant to their current view (eg: Datstore, table, and field-level tag management).



Additionally, users can author checks directly from the Field Details Page.

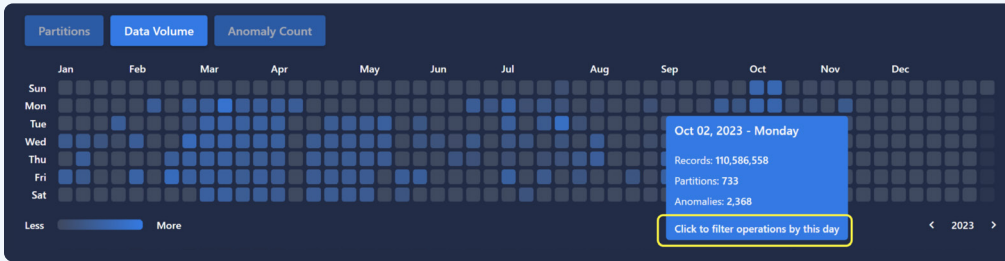
## Slide Navigation

Added **"Previous"** and **"Next"** buttons for **Checks and Anomalies** details, allowing users to efficiently navigate between items without the need to close and open individual cards.



## Heatmap Filtering for Operations

Selecting a date in the [heatmap now filters](#) the list of operations to include only those that were **conducted on that date**, allowing for faster, easier investigation of historical data quality events.



## REPORTING IMPROVEMENTS

### Enhanced Activity Reporting

[Operation details](#) have been **significantly expanded** to include comprehensive information about the profiles scanned, with the ability to easily **drill down** to individual tables/files and associated Anomalies, enhancing the user experience of interrogating each scan's results. We also now display and directly link to the Schedules that triggered each operation, improving traceability, explainability and visibility.

Additionally, we have added a **multi-select picklist** based on status, giving users the ability to instantly find operations that may require attention.

The screenshot shows a scan operation summary with the following metrics:

- Tables Requested: 8
- Tables Scanned: 8
- Partitions Scanned: 8
- Records Scanned: 59K
- Anomalies Identified: 3

Callouts highlight the following features:

- "Click to view scanned table details" points to a table icon.
- "Click to view all Anomalies produced by the scan operation" points to an anomaly icon.
- "Anomaly counts, as well as the Anomalies themselves, are displayed per table" points to an anomaly entry for the table 'FL\_INSURANCE\_REPLICA'.

The screenshot shows a scan operation summary with the following metrics:

- Tables Requested: 10
- Tables Scanned: 10
- Partitions Scanned: 9
- Records Scanned: 1.6M
- Anomalies Identified: 26

A callout highlights the "View Schedule" button.

The screenshot shows a profile operation summary with the following metrics:

- File Patterns Requested: 18
- File Patterns Profiled: 18
- Files Profiled: 18
- Records Profiled: 476.5K
- Field Profiles Updated: 177
- Inferred Checks Synchronized: 503

A callout highlights the "Files Profiled" metric.

**DFS Only:** We now report on the **total number of File Patterns and Files Profiled**. This capability enables users to better track our globbing capabilities and track profiling results in more detail.

The screenshot shows the "Operations" page with a "Status" picklist. The picklist options are:

- Success
- Failure
- Running
- Aborted

A yellow arrow points to the "Status" picklist, and a yellow box highlights the options.

## Insights Dashboard Improvements

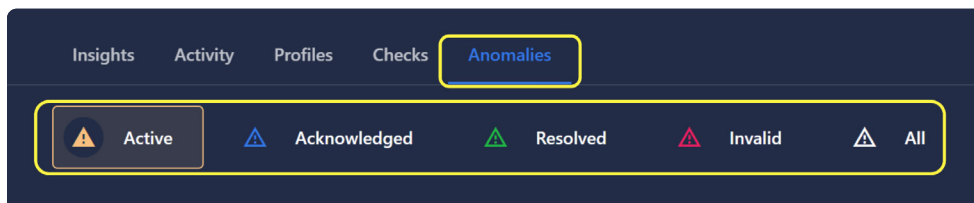
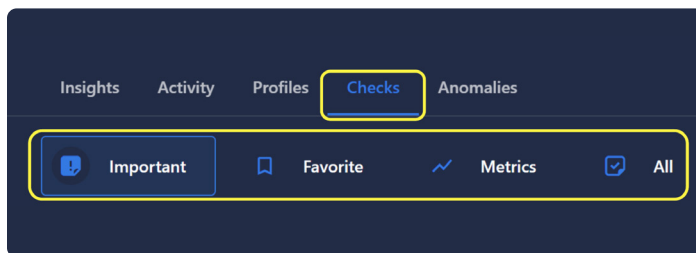
Previously, the **Insights Dashboard** was only filterable using Datastore-level tags. We now support this same functionality for **tagged tables and files** in order to provide more granular, targeted data quality reporting.

Insights now features **a new KPI** indicating the total number of checks **passed** and **failed**. This enhancement offers a visual representation through a chart, detailing the passed and failed checks over a specified reporting period.



## Check and Anomaly Categories

We have added additional control over how **Checks are displayed** in the UI, giving users the ability to easily find the ones most relevant to their current task. While we continue to support sorting and filtering by several criteria including tags, for streamlined navigation of high-quantity checks, we now group checks into 4 categories: **Important**, **Favorite**, **Metrics**, and **All**.

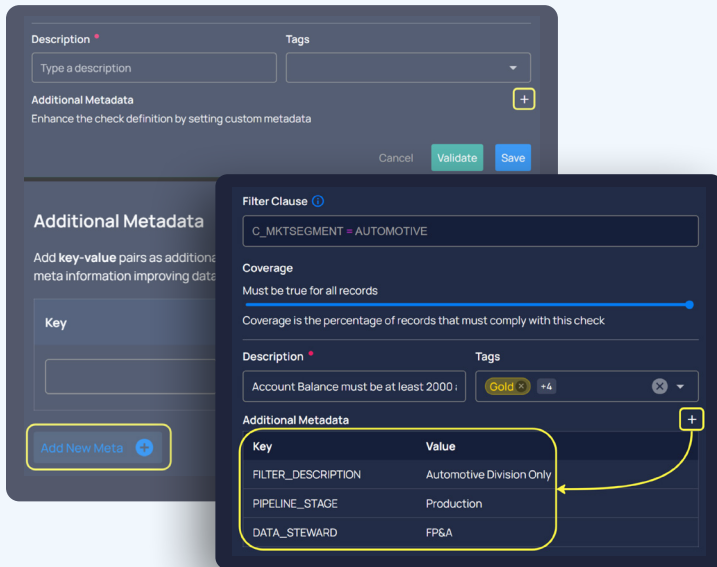
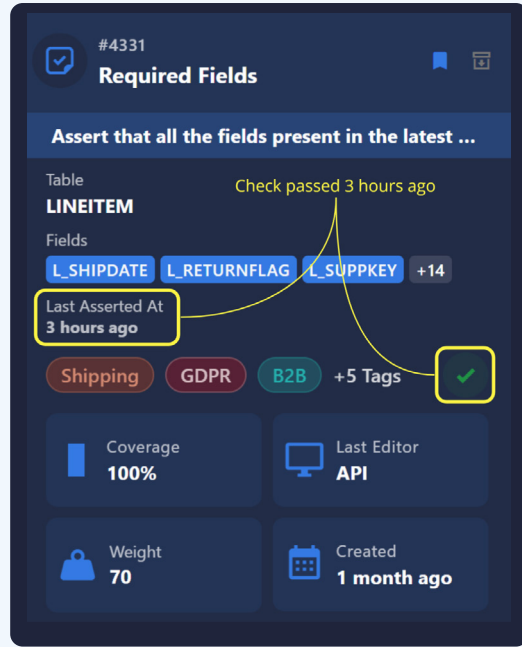


- **Important:** This displays all authored checks, and all automatically inferred checks that currently have an active anomaly. It also includes all checks with a sufficient associated weighting (currently set at 7. We plan to give users the ability to define this threshold in a future release).
- **Favorite:** Includes only checks that have been marked as "Favorite." Since Favorites are not global designations, this view will be unique for each individual user. In case you haven't used Favorites in the past, this designation can be toggled in the upper right of the check card.
- **Metrics:** Displays the time-series charts of all metric checks.
- **All:** Includes the entire population of checks.

We have added a **navigation bar** to the Anomalies tab as well, allowing users to easily navigate to anomalies by current status.

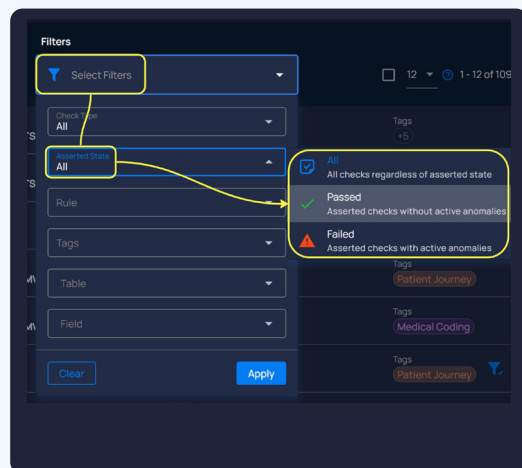
## Advanced Check Surveillance

Check cards now display the most **recent time** that the check was asserted, and affirmatively indicate, via a **green check mark**, that a check passed at the time of the most recent scan operation. This icon will only appear when the check does not have any active anomalies.



The Check management interface now includes the ability to **add custom metadata fields**. This enhancement allows users to add key-value pairs to track important additional details, significantly increasing the flexibility and customization of the Check definition. These key-value pairs are not only visible through the Qualytics interface, but are also written downstream to your Enrichment store.

Talking about **Passing** and **Failing** checks, we have introduced a new filter in the Check listing, allowing users to easily interact with checks that **have or have not** generated active anomalies.



## Source Records Display Enhancement

Columns in the anomaly dialog **source records** can now be **manually resized**, allowing users to view lengthy fields In-App.

RICE	L_LINENUMBER	L_LINESTATUS	L_ORDERKEY	L_ITEMKEY	L_QUANTITY
2363	?		112679834	643865700	682.83
2342	S		550076305	204729846	-16.88
2977	O		381025905	968020292	-42.07
4281	-		17176399	433645263	711.25
9237	nuF		656172261	230199440	6.0031
3678	P		522580044	108541563	-78.75
3630	F		676476325	465592386	443.59
306	nuF		132264401	545616986	924.48
4152	P		367878910	109944492	-40.54
6726	nuF		29077013	254512994	-14.63

Failed Checks	Violation	Percentage
#101775 Not Future	In O_ORDERDATE, 0.001% of 1500044 filtered records (16) are future times	100%
#101772 Between Times	In O_ORDERDATE, 74.984% of 1500044 filtered records (1124791) are not between 2007-01...	100%
#101944 Less Than Field	In O_ORDERDATE, 0.001% of 1500044 filtered records (16) are not less than LAST_MODIF...	100%
#101772 Not Null	In O_ORDERDATE, 25.016% of 1500044 filtered records (375246) are not assigned values	100%

Additionally, **read-only versions** of archived checks can now be viewed from the anomaly interface, providing users with details about why an anomaly was produced, even after a check has been removed.

## Overhauled In-App Notifications Panel

Introduced **a new side panel** for **In-App Notifications**, which categorizes alerts by type (Operations, Anomalies, SLA) and includes additional information such as associated tags and channels notified. Users now have the ability to navigate directly to the Operation, Anomaly, or SLA that produced the notification, as well as take other actions such as muting or editing the notification itself directly from the panel.

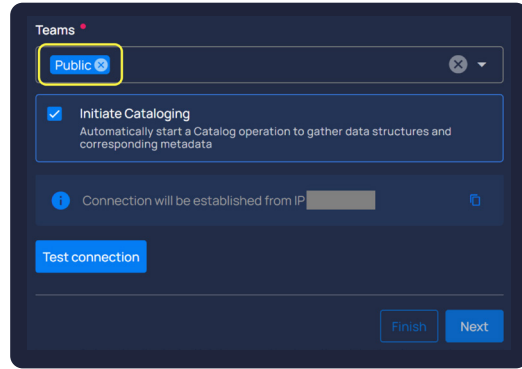
## Dynamic Quality Score Based on Anomaly Resolution

Marking an anomaly **as resolved** now has an **instant positive impact** on **Quality Score**.

# OPERATIONAL EFFICIENCIES

## Specify Teams Upon Datastore Creation

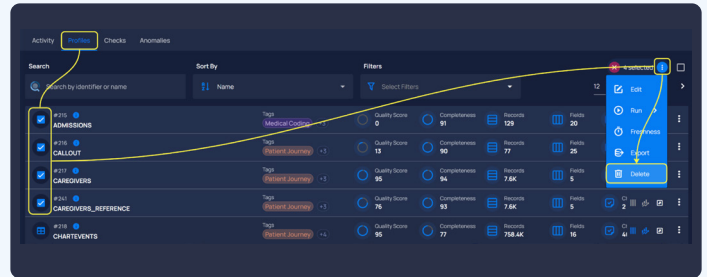
The **Datastore Connection Form** now allows Admins to **control which Teams have access** at setup, ensuring that only the appropriate users can interact with potentially sensitive information from the very moment that the connection is established.



## New Bulk Actions

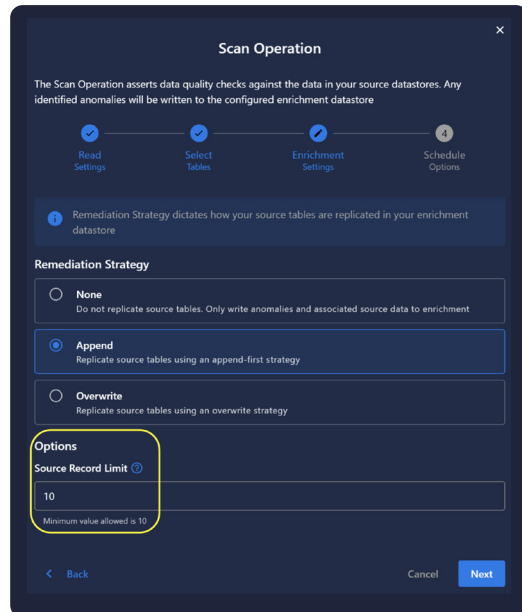
Selecting **multiple profiles** within a datastore will now allow users to adjust freshness and SLA settings in bulk, in addition to the preexisting abilities to edit, run operations, and export metadata.

Users now have the option of **deleting multiple profiles** in one action, rather than one by one.



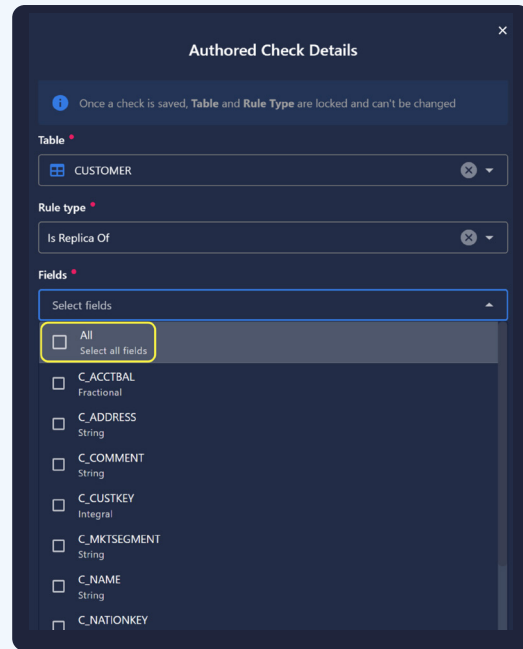
## User-Defined Source Record Limits

Previously, we capped the number of anomalous records written to the Enrichment datastore at 10. Based on customer feedback, we now **allow users to define this limit** based on their individual operational and business needs. This can be combined with a separate parameter around governance of rows required to roll up Record anomalies to Shape anomalies, allowing users to have full control over the aforementioned thresholds.



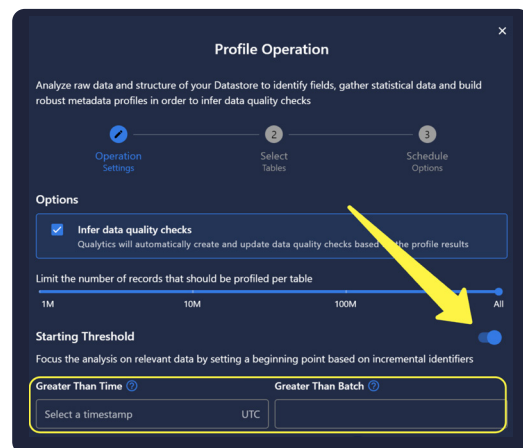
### “Select All” Fields for Checks with Multi-Select Picklist

When authoring a check with a multi-select picklist of fields, we now have an option to **select all** of them at once. This eliminates the tedious process of clicking each field individually when, for example, setting up an IsReplicaOf check on an entire table with a **large number of fields**.



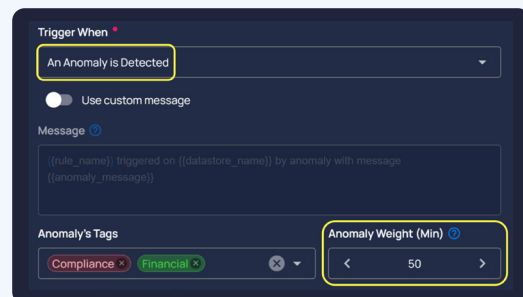
### Profile Based on Incremental Identifiers

Users can now **specify a value** for an **incremental identifier** to target a specific starting point for profiling operations, ensuring that only the most **relevant data** is included for analysis. **Greater Than Time** can be used for time stamped data - for example, the Profile can be trained on only the last year's data via this functionality. For datasets with other incremental identifiers, we also support Greater Than Batch as a threshold variable.



### Conditional Notifications Based on Weight

For notifications that trigger when an anomaly is detected, users can now specify a **minimum weight requirement** for the associated anomaly, ensuring that stakeholders are only **notified of anomalies of substantial severity**.



# PRODUCT ROADMAP 2024

	Q4 2023	Q1 2024	Q2 2024	Q3 2024	Q4 2024
User Interface	Tree View Contextual Navigation	Custom Weightings	List Views	Root Cause Analysis	Sector Specific Libraries
Data Analysis	Time Series Metric Checks NLP Street Address Validation	Top 3 Influences of Quality Score Quality Score Factors	Anomaly Root / Common Factors Analysis Profile Change Analysis	Deduplication Autotuned Operations	Realtime Scanning Nested Datatypes (JSON)
Data Management	Exportable Metadata Enhanced Insights	Check Templates / Library Data Catalog / MDM Integrations	Entity Mgmt / Resolution Predictive SLAs	Custom Field Management / Masking Automated Groupings	Qualytics Reference Sources
Compliance & Security	SOC2 Type 2		IdP Groups to Teams Mappings		
Developer Tools	Public CLI & PyPi Library	Native Slack & Teams Integrations		Adhoc Check Assertions	Enterprise LLM Integrations
Deployments	ML Powered Active Monitoring	Self-managed IdP Authentication	Deployment Telemetry	Dynamic Resource Management	Marketplace Deployments

## CONTACT

Contact our team for a demo of new features at [hello@qualytics.ai](mailto:hello@qualytics.ai)



# PRODUCT UPDATE

End of Year: 2023 Edition

Take a **proactive** approach to **data quality** with **Qualytics**.

