INFOGIX

THE POWER TO TRUST

Data3Sixty™ DQ+

Release Notes

Version 3.0

Table of Contents

1. What's New in 3.0?

The *Infogix Data3Sixty DQ+ Release Notes* briefly describes enhancements and modifications included in Infogix Data3Sixty DQ+ 3.0. For more detailed and functional descriptions of newly added features, please consult the Infogix Data3Sixty DQ+ User Guide.

Analysis Improvements

Changes to support better refactoring in the future

Changes to the product's Analysis code have been made so that once any Analysis that was created in a previous version of the product is saved in version 3.0, a migration will be performed.

Convert from Hadoop folder structure to traditional folders

A check box labeled "Use traditional folder structure when used for output" has been added to Data Store definitions and Excel Output nodes. This check box allows the user to output files to a traditional folder structure as opposed to a Hadoop folder structure. For Data Stores, this option is supported for External Data Stores with Delimited, Parquet, ORC, and AVRO layouts; and, it is available within the Details tab. For the Excel Output node in Analyses, the check box is available in the Options tab of the node's properties panel.

Compute Node

This new node allows the user to create a new computed field at any point within an Analysis or change the values of a pre-existing field using the expression editor. While the ability to add new computed fields is a feature that did already exist in most Analysis nodes, in the form of using the New Column button, the ability to change the values of pre-existing fields is completely new functionality.

Change Spark type of date fields to date from timestamp

A routine code review revealed that the product was using the Spark timestamp type to represent date values. To prevent future issues, date fields now use the Spark date type.

Support partitioned delimited file output to external Data Stores

Support for partitioning an Analysis's Data Store output has been added. Such partitioning can be performed on a Data Store Output node when it is of the External Store Type, it uses a file system (S3, HDFS, or another file system) as its Store Repository type, and it has a Delimited or Parquet Layout Type.

When using such a Data Store, this new feature becomes available to the user by way of a new Options tab. Within the tab, the user may select a field or fields to partition by. Partitions will then be created as folders within the file system in use, using the unique values from these selected fields.

For Data Store of the Delimited Layout Type, the Options tab also features an Output Writer drop down, which allows the user to output files in a Basic or Enhanced manner. Here, Basic refers to the deprecated method used in prior releases, where control over partitioning did not exist; and, Enhanced refers to the new method that uses enabled partitioning. Additionally, the Options tab also features a drop down that allows the user to control what type of file compression to apply to the output files.

Improve execution performance of Computed Fields and Script node

In this release, the Select and Script nodes have been modified to improve their performance.

Add support for upsert for internal Data Stores of type database

If Update Data in Data Store has been turned on for a Data Store Output node that represents an Internal Database Data Store, the user may now also enable record insertion when a record to update does not exist, by checking Insert when Record to update doesn't exist. Turning this option on will mean that if the Data Store is set to update, and it encounters a record with new key field values that are not present in the Data Store, then the new record will be inserted into the Data Store.

Delimited/Excel file processing changes

In previous releases, system processing required that data fit Data Store definitions exactly. With this release, a lenient parsing flag has been added to Data Stores of the Delimited/Excel layout types that allows for the use of data that does not fit a Data Store definition exactly.

With lenient parsing:

- Files may contain more fields than are defined in the Data Store definition. The additional fields and their values will be set into the field that is defined as the overflow field in JSON format.
- Files may contain fewer fields than are defined in the Data Store definition. Any missing fields will be assigned the default value, if one is present for the field.
- Field order in a file may be different from that in the Data Store definition.
- Synonyms may be used for field names.

HTTP REST API node

This enhancement introduces a new Analysis node that allows a REST service to be read from or written to. HTTP methods available for use include: GET, PUT, POST, DELETE, and HEAD.

Data Store and Data View Improvements

Don't use concurrent readers for External Database Data Stores when in Analysis Editor

Changes have been made so that when working with an External Database Data Store in the Analysis Editor, concurrent reads are no longer performed and the Data Store's Sample SQL is used. As a result of this change, Sample SQL within a Data Store definition no longer requires an ORDER BY clause.

Improve housekeeping for Redshift

In this release, re-indexing of the Redshift database has been improved to prevent loss of performance over time.

Add support for deleting duplicate records from Data Views

To support the updating of records in a Data View, this release adds the ability to delete duplicate, older versions of a record from a Data View. To achieve this, a new field set called **Delete duplicate data after load** has been added to the Data View Other tab.

For this release, this feature is only available on the cloud version of the product.

Add Support for Synonyms for Data Store Fields

Data Store definitions now support use cases where the name of a field in a file may vary from file to file but is expected to be treated as the same field. To support such use cases, Data Store field Synonyms have been added, which allow the user to specify words that will be interpreted as the field's name, when the word is encountered as a field name in a data source. Synonyms may either be specified in the Data Store UI or retrieved from Govern.

Query and Cache Synonyms for Data Store fields from Govern

As mentioned above, Data Store field synonyms may be retrieved from Govern. This capability has been enabled by a change which allows Data3Sixty DQ+ to query Govern for business term artifacts and cache the results of these queries.

Allow any field type for Data View partition field in cloud version

In previous releases of the product, only date fields could be used as partition fields in a Data View. With this release, any type of field may be used.

Change organization for S3 file storage

Previously, the storage for S3 Internal Data Stores did not follow Hadoop folder naming conventions. This prevented Data Stores from being used efficiently with Hadoop compatible services, like Amazon Athena. With this release, the organization of newly created Data Stores will follow Hadoop folder naming conventions. Older Data Stores created before this release will continue to use the old organization.

Show data sample when Test is run in Data Store Fields tab

After successful testing of a data store within its Fields tab, a user can now use a Show Sample Data button to view a sample of the Data Store's contents.

Provide XML reading support in Data Stores

XML files may now be used to create external Data Stores of the S3/HDFS/file system store repository type.

Add support for reading Excel files in Data Stores

When creating Data Stores of the External Store Type, a new Layout Type, Microsoft Excel, is now available. This option can be used to create Data Stores from Microsoft Excel .xlsx files.

When using this option, a new Microsoft Excel Layout form becomes available, which allows the user to customize how their Excel file is read. The user may: skip empty columns, skip empty rows, define which row within the file's sheets contains column names, and define how to select worksheets within the file, using numbers, names, or a regular expression.

Mass edit of search results

Users now have the capability to mass edit result records in a search results screen. This can be done when the user selects multiple records in a results grid and uses the Mass Edit button. The Mass Edit button will then bring up a screen that allows the user to work with all editable fields.

Reusable Rules, a.k.a. the "Rule Library"

With this release, a new Data Stage type has been added: the Rule Library. A Rule Library allows users to create reusable, data quality rule sets that can be applied to data in an Analysis. Reusable rules created with a Rule Library are used in an Analysis via the Execute Rule Library node.

When creating a Rule Library, the user can create the types of rules outlined below.

Data Quality Rule Group

A Data Quality Rule Group is a top level folder, used to hold reusable rules within the Rule Library. The Data Quality Rule Group is also where users define placeholder input fields to be used by the reusable rules that the Data Quality Rule Group contains.

Completeness Check, Consistency Check, Value Conformity, Type Conformity, and Timeliness Check Rules

These reusable rules reflect what can already be done with Data Quality Check nodes in an Analysis; however, within the new Rule Library functionality they are intended to be created by the user as generic rules that can be reused in multiple Analyses. These checks are made reusable by the user selecting a placeholder field from a check's parent Data Quality Rule Group to perform the check on. The check can then be applied to any field with the same data type as the selected placeholder field, by using an Execute Rule Library node in an Analysis.

Regular Expression Rule

This reusable rule allows a user to create a regular expression check that will be run on a placeholder field from the parent Data Quality Rule Group. The UI for creating regular expressions also provides a way for the user to test the regular expression.

General Expression Rule

This reusable rule allows a user to create an expression check in the product's general formula language that will be run on a placeholder field from the parent Data Quality Rule Group.

Data Quality Reusable Rule node in Analysis

To support reusable rules, a new node called the Execute Rule Library node has been added to the Data Quality Check palette within Analyses. This node allows the user to select reusable rules from a Rule Library and apply the rules to fields within their Analysis.

Visualization Improvements

Allow table related HTML tags in Text Dashlet

With this release, text dashlets now allow for table related HTML tags, to enable the creation of tables within the dashlet.

Allow filters to be passed into Dashboards in portal mode

Users of Infogix Govern may now pass filters to Dashboards from Infogix Data3Sixty DQ+ displayed in Govern in portal mode. This is done through a request parameter using an HTTP POST request. The parameter must be named dashboardFilterValues and its value must be a JSON string of an array of filter values.

Add refresh capability to Dashboards

Capability has been added to refresh a Dashboard's data using a Refresh button, within the View Dashboard and Home screens. This has been added to allow the Dashboard user to refresh the data that the Dashboard is displaying on demand.

Allow Current Filters component column widths to be resized and save setting

Users may now resize the column widths of a Current Filters dashlet component and save the resized width. Previously, saving of such widths was not available.

Populate dashlet only when filtering on specific fields

Within a dashlet's Query Criteria, an option is now provided to only populate the dashlet when there is a global filter on one or more fields specified by the user. Enabling this option allows better performance when there is a large amount of data when no filter is active.

Undo/Redo for filters and drill downs in Dashboards

When viewing a Dashboard, a user may now undo or redo filtering and drilling actions one by one by way of new Undo/Redo buttons. As part of this improvement, the separate back button that appears when drilling to a Dashboard has been removed.

Filters that change state based on global filters

Filter dashlets now feature a cascading option that allows you to enable filtering of the contents of the filter dashlet itself, based on the filter state of your entire Dashboard sheet. The filter state of a Dashboard sheet is comprised of all filters applied to the Dashboard sheet by drilling, brush selection, or filter dashlets.

Support for turning Visualizer state into a Dashboard

When using the Visualizer, a user may now copy their chart to a new Dashboard or to an existing Dashboard, by way of a newly added Copy To button. This feature effectively allows users to save charts they create in the Visualizer, something that could not previously be done.

General Usability Improvements

Customize Process Model failure notifications

With this release, a new field set has been added to Process Model settings that allows the user to customize the message that is sent to users when a Process Model execution fails. As before, this functionality includes a 'Groups to Notify' parameter, but now the user may customize the message subject and message details that are sent to these groups.

Improvements to Executions screen

Improvements to the Executions screen that have been made in this release include the following:

- The ability to view an execution by its Work Id, at its own url.
- The ability to view an execution by its Process Id, at its own url.
- The ability to filter executions by a Process Id, showing the user all Data Stages that were executed during a specific run of a Process Model.
- The ability to sort executions by Stage Type, Start Time, End Time, Status, Work Id, and Process Id.
- The ability to filter executions by Status.

Navigate to Stage Functionality

Throughout the product, functionality has been added to "Go to a selected stage". This functionality is a way to quickly move from one Data Stage to another, and it is available in places where one Data Stage uses or references another. When clicking "Go to Selected Stage" you will simply be transitioned to the View mode of the Data Stage that you have selected.

The Go to Selected Stage feature is now available when:

- Viewing or Editing an Analysis or Process Model.
- Viewing the Executions History of an Analysis or a Process Model.
- Using the Show References or Find Usages functionality.

Add Test Capability to Expression Builder

With this release, expressions may now be tested within the Expression Builder. To test expressions, users may enter input data into their expression and the tester will evaluate the result.

Secure Value Enhancements

With this release, three significant changes to field level security have been introduced.

- All field types may be made secure, whereas previously only String fields could be made secure.

- Masking and encryption may now be independently controlled when applying field security. Previously, if a field was marked secure it was always masked and encrypted.

- An Unmask All secure values feature has been introduced, to Analyses, Dashboards, and Case Management/Search Screens.

Changes to secure values support in Data Stores UI

With this release, masking and encryption are decoupled in Data Store field security. That is, while previously marking a field as secure meant that it would be both masked and encrypted, in this release a user may mark a field as masked or encrypted or both. Additionally, masking and encryption may be applied to any field type, whereas previously field security could only be applied to String fields.

Due to the fact that data residing in an external source cannot be encrypted by the product, field encryption is only supported for internal type Data Stores. External type Data Store fields may however be masked.

Changes to secure values support in Data View UI

With this release, masking and encryption are decoupled in Data View field security. That is, while previously marking a field as secure meant that it would be both masked and encrypted, in this release a user may mark a field as masked or encrypted or both. Additionally, masking and encryption may be applied to any field type, whereas previously field security could only be applied to String fields.

Secure value changes in Analysis UI

Due to the changes to Data Store field security and encryption mentioned above, functionality has been added to the Analysis UI to allow users to unmask secure (but not encrypted) values. This functionality is available by way of a new Unmask All button.

Additionally, the Sqoop In, Kinesis, and Plugin nodes that previously only supported encryption of String fields now support encryption of all field types.

Add new configuration related to secure values for Environments

To accommodate the secure value changes made to Data Stores and Data Views, two new check boxes have been added to the Environments screen. The first check box is labeled 'Allow Unmask All for Secure Fields' and it controls whether users with the appropriate permissions can unmask all secure values at once wherever they appear throughout the product - except for in Analyses. The second check box is labeled 'Allow

Unmask of Secure Fields in Analysis' and it controls whether users with the appropriate permissions can unmask secure values within an Analysis.

Secure value changes in Data Store/View back end

With the newly supported functionality available when making Data Store and Data View fields masked and encrypted mentioned above, changes were also required within the product's back end processing.

These changes include:

- Changes to the code that creates/modifies database tables to make it use string fields for encrypted fields of types other than string.
- Changes to handle encrypted values and not choke on attempts to convert encrypted values to expected data types.
- New limitations put in place on queries made on fields containing secure values, to prevent breaches in security.

Secure value changes in Analysis back end

With the newly supported functionality available when making Data Store fields masked and encrypted mentioned above, changes were also required within the product's back end processing.

These changes include:

- Enhancements to the converters for each field type to handle encryption of that field type. Previously, only the String field type could handle encryption.
- Changes to the Sqoop in, Kinesis, and Plugin nodes to handle encryption of types other than String.

Secure value changes to Search/View/Edit screens in Case Management and Data Entry

With the changes made to secure values, the search/view/edit screens available in Case Management and Data Entry have undergone the following changes.

- No user may perform a query with a filter operation on one of these screens using an encrypted field.
- Users without permission to unmask a secure field may not use the field in a query with a filter operation.
- An Unmask All toggle button is now available in these screens that unmasks all secure field values the user has permission to unmask and that generates an audit trail entry when they do so.

Secure value changes for Dashboards

With the changes made to secure values, Dashboards have undergone the following changes.

- No user may perform a query where an encrypted field is used in anything other

than the projection.

- Users without permission to unmask a secure field may not use the field in queries for anything other than the projection.
- An Unmask All toggle button is now available in Dashboards that unmasks all secure field values the user has permission to unmask and that generates an audit trail entry when they do so.

Migration of Data Stores and Data Views

With the introduction of the separation of masking and encryption, the user may now switch a Data Store or Data View field from encrypted to non-encrypted and vice versa. Normally, such a switch would require that any pre-existing data in the Data Stage be deleted; however, data may now be retained using data migration.

Data migration is now performed automatically whenever field encrypted is changed and the change is saved to the Data Stage's definition. A separate Migrate Data option is also available to administrators within Pipelines.

For this release, migration is supported for Internal Database based Data Stores and for all Data Views except regular performance Data Views in the product's Enterprise mode.

Library Upgrades

Upgrade Hadoop
With this release, the product now uses Hadoop 2.7.3.

Upgrade Spark

With this release, the product now uses Spark 2.3.1.

Upgrade Highstock
With this release, the product now uses Highstock 6.0.7.

On-premises Installation and Infrastructure Improvements

Include updates to installer and 3rd party components as part of patches
In order to address concerns about patching third party components used in the product, an RPM based distribution of third party components has been adopted.

Support for HA name node and resource manager
With this release, the HA name node and resource manager in the installed Hortonworks cluster are now turned on.

Installation without requiring passwordless ssh

With this release, the Enterprise version of the product may be installed using ssh access with a password. Previously, only passwordless ssh was supported.

Replace Nginx with Apache

With this release, Nginx has been replaced by Apache in the Enterprise version of the product.

Generate a Vertica self-signed certificate automatically

With this release, Enterprise installations are now able to generate a Vertica self-signed certificate automatically.

Add generated self-signed certificate for Postgres

With this release, the *generateApplicationDbSecurityCertificates* task has been added to Enterprise installations of the product, which generates a self-signed SSL certificate for Postgres.

Miscellaneous Enhancements

Upgrade Java version used in cloud deployments

With this release, cloud deployments of the product now use Java 8 Update 181.

Add authorized users message to login screen

A new message has been added below the log in form of the product, which communicates the following: "This service is restricted to authorized users and purposes only. All user and network activities on this system are fully logged. Unauthorized access or actions will be fully investigated and reported to the appropriate law enforcement agencies."

Retry Analysis and Data View when Application Server instance dies

Previously, when an app server instance would die, the system would mark any executions on that app server as having failed, when the lease acquired by the execution expired. This would happen even if the underlying execution that was submitted to the Hadoop cluster completed successfully.

To make the system more reliable, the system will now attempt to reconnect with Hadoop when a lease expiration is encountered for an Analysis or Data View execution, to see if the execution succeeded. The system will then update the status of the Data Stage accordingly.

Retry Data View loads if Redshift/Vertica is down

To make the system more reliable, Data View executions will now be reattempted if an execution fails due to Redshift or Vertica being down.

Reduce space taken by JSON elements in Custom Completion results

With this release, the amount of space taken up by the JSON structure used to represent Custom Completion results has been reduced. This has given the user more space to store actual data.

Retention Execution Viewer

A debugging screen that can be used to view and terminate system retention runs has been added to the product. This screen is only available to admin users who have added the *dev=true* query parameter to the product's url.

Use Hadoop vendor specific Spark jars during yarn execution

With this release, the product now uploads vendor provided Spark jars to the Cloudera/Hortonworks cluster and uses these jars when executing in the cluster. Previously, open source versions of the Spark jars were used.

Add additional security related headers

Additional security related headers, including Strict Transport Security, Content-Security-Policy, and Cache-Control and Pragma, were added to some of the product's responses.

Mark session related cookies as secure

With this release, code changes have been made to ensure that session related cookies are being marked as secure when the product is being rendered using https.

Move housekeeping related processing to a separate scheduled job

With this release, housekeeping related processing has been decoupled from system retention processing. This provides more predictability and also allows housekeeping related processing to be scheduled less frequently than retention, which usually runs every day.

REST API to load/sync users with an external system

This new API allows users of type member to be bulk loaded or bulk updated.

2. Technical Preview Features

The following features are now available within Infogix Data3Sixty DQ+ as "Tech Previews". A feature classified as Tech Preview is not guaranteed to be 100% functional but has been included in this release to enable customer exploration. Infogix encourages you to explore these "Tech Preview" features and provide feedback on your experiences. All "Tech Preview" functionality is marked as such within the Infogix Data3Sixty DQ+ UI.

Excel Output Node

This new node allows the user to output an Analysis to a multi-tab Excel file. It also supports file partitioning and the use of a traditional folder structure or a Hadoop file structure.

Allowing nesting of an Analysis

With this release, one Analysis may be nested within another Analysis, effectively allowing the user to created reusable Analyses. This functionality has been made possible by the addition of five new nodes: Nested Input Analysis, Nested Output Analysis, Nested Analysis, Placeholder Input, and Placeholder Output. Different combinations of nodes may be used for different applications; however, in general a nested node will be combined with a placeholder node.

Execution of an Analysis using Livy

In the cloud version of the product, there is now an option within Analysis settings to execute an Analysis with a Livy cluster type.

3. Customer Reported Issues

This section of the release notes lists Customer Reported Issues that were resolved in Infogix Data3Sixty DQ+ 3.0

Unable to Select Users in Google Chrome for Data3Sixty Foundation

In previous releases, Infogix services users acting as administrators on customer systems would intermittently report a problem with the Users Grid, where the grid would fall into a bad state, making all user records unselectable until the grid was refreshed. With this release, this issue has been fixed.

Incorrect syntax near the keyword 'TOP' when using concurrent reads/ORDER BY

During Infogix services use of SQL layouts for Data Stores, a java.lang.StringIndexOutOfBoundsException was discovered in a case where there was a newline and no space following an SQL SELECT statement. With this release, this issue has been fixed.

Contact us

If you need assistance, submit a case online at support.infogix.com, or contact Infogix Customer Support at +1-630-505-1890.

If you are experiencing issues contacting us by phone or by our website, please email support at support@infogix.com

Visit our website at www.infogix.com