Datasets confidentiality configuration System

Each business has confidentiality rules to anonymous its data to prevent the identification of any dataset instances and keep all data confidential.

One way.

You can set many dataset views in your system and set users to each view based on each user's confidentiality level. But this needs maintenance to make sure all dataset views have one ground truth in every aspect of your system.

This needs fancy tools to manage your dataset views and your user's management.

Is this the only way?

There is another way to do this.

You can create a confidentiality system to configure your dataset user visibility configuration by creating a dynamic view configuration for each user instead of creating different dataset views. We only have one dataset view but many user view configurations.

Let us set the requirements first.

1. We only have one ground truth for any dataset for any user.

(Analytics, stream, or ….) [ it is better if all datasets have a common column name]

For example: - the [DateOfBirth] column will be named the same in any dataset.

1. Each user at run time sees only his own set of columns from any dataset.
2. We can hide a full column from some users.
3. We can hide specific cell values from some users in any column in any dataset.
4. We can exclude some values from being hidden in any column from some users.
5. We can mask any cell value by any pattern or mask text.
6. For any user, we can hide or mask one cell value in one column based on another column value.
7. No confidentiality rule execution prevents or rolls back another rule execution (execution order matters).

All these rules can be applied by setting the [converters] parameter [read\_csv()] function

Only point number [7] and point [8]

If we used converters; we can run into an issue when we try to hide one column cell value based on another column value. Because converters work at file load time, when data is loaded it will be already converted and we can not apply confidentiality rules on converted data.

Also, if we used converters, we do not have any control over the order of execution for these converters.

Therefore, we will create dataset confidentiality rules by setting up a user configuration system that creates a runtime dataset confidentiality view for each user.

This user runtime view configuration can be [in-memory] configuration or [saved] configuration into a file or DB.

We have [9] runtime confidentiality rules for any dataset.

1. [names] for column names
2. [use\_cols] column index in the original dataset (the order of the columns in user view)
3. [ColDataType] must be a string for all columns to be able to mask or hide any cell values by mask text.
4. [DynamicColumnsSearch] Specify columns that will be hidden based on other column values.
5. [HideFullcolumn] for columns that need to be fully hidden (all column cells are hidden)
6. [HideDynamicTerms] list of values that we check for to hide one column based on another column.
7. [HideALLExceptThisValue] list of excluded values from being hidden in any column.
8. [MaskPattern] Mask any column using a mask pattern and mask character.
9. [HideValue] Hide specific values in any column.

All these rules can be populated by lists or dictionaries that can be created based on filter rules or business rules.

A screenshot of a computer

Description automatically generated

For each newly added [USERX] to our system, we initiate a new dictionary with these [9] confidentiality rules by adding this new [USERX] into an existing configuration policy for some confidentiality rules. (\*\*only one rule is mandatory the [GroupPlicy.set\_DataSetColumnNames()])

\*\*The rest of the other 8 confidentiality rules are optional

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

This confidentiality [GroupPolicy] [DataFrame] can be saved into a file or DB and applied in any dataset at runtime in memory.

You can have one global confidentiality [GroupPloicy] for any dataset in your system if all your system datasets use the same column names.

Or you can have one confidentiality [GroupPloicy] for each dataset if each dataset in your system has its column names and its own business rules.

A screenshot of a computer

Description automatically generated

These confidentiality rules generate a unique dynamic dataset view at runtime on the fly for each user based on his configuration.

Next, we will see some examples for each confidentiality rule which shows how one dataset ground truth can be viewed by different users based on their configuration on the fly.

Just by selecting a user from the dropdown, the dataset view will be different based on selected user configurations.

A screenshot of a computer

Description automatically generated

\*\*The user on the left has some confidentiality configuration and the user on the right does not have some confidentiality rule configured\*\*

Rule (1) \*\*set\_DataSetColumnNames

This rule sets three rules [names, use\_cols, ColDataType]

A screenshot of a computer

Description automatically generated

Rule (2) set\_HideFullColumn

A screenshot of a computer

Description automatically generated

Rule (3) set\_HideValue

A screenshot of a computer

Description automatically generated

Rule (4) set\_MaskPattern

A screenshot of a computer

Description automatically generated

Rule (5) set\_HideExceptValue

A screenshot of a computer

Description automatically generated

Rule (6.1) set\_HideDynamicTerms and set\_DynamicColumnsSearch

Terms that need to be looked up and columns that are hidden based on each other values.

A screenshot of a computer

Description automatically generated

Rule (6.2) set\_HideDynamicTerms and set\_DynamicColumnsSearch

A screenshot of a computer program

Description automatically generated

Rule (6.3) set\_HideDynamicTerms and set\_DynamicColumnsSearch

A screenshot of a computer

Description automatically generated

Rule (6.4) set\_HideDynamicTerms and set\_DynamicColumnsSearch

A screenshot of a computer screen

Description automatically generated

\*\*\* All these Hide texts like [[HideCellDynamic], [HideExceptValue],[HideFullColumn],[HideValueText]] are configurable and can be replaced by any text that matches your system business requirements.