



Talend Component tGoogleAnalyticsManagement

Purpose

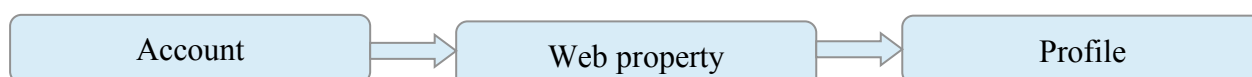
This component addresses the needs of gathering Google Analytics data for a large number of profiles and fine-grained detail data.

To prepare the queries you need to know what account, profiles and so on you have access.

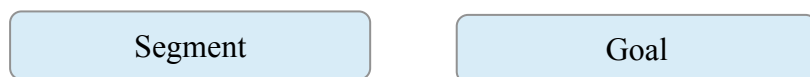
This component collects all elements of the Google analytics model and provides their data as separate output flows.

The component uses the Core Reporting API 3.0 and the Authentication API OAuth 2.0 final.

To provide the ability to run in multiple iterations the component has special capabilities to avoid multiple logins through iterations. Usually automated processes should not use personal accounts. This requirement is addressed by using Service Accounts, which are the only preferred way to login into Google Analytics for automated processes.



An account contains one or more web properties. A web property contains one or more profiles.



A segment is independent; a goal always belongs to a profile.

Additional the component provides an output flow for un-sampled reports. But this requires a premium account. It is recommended to use the flow from the component tGoogleAnalyticsUnsampledReports.

Talend-Integration

This component can be found in the palette under Business->Google

Parameters

Parameters to connect to Google Analytics (setup client)

| Property | Content | Data types |
|-----------------------|--|------------|
| Application Name | Not necessary, but recommended by Google. Simple provide the name of your application gathering data. Required | String |
| Use existing client | Choose here the tGoogleDrive component which client do you want to reuse in this component instance. | Boolean |
| Authentication Method | Choose the method to authenticate: Service Account or Client-ID for native applications | String |

Properties to use the Service Account

| Property | Content | Data types |
|-----------------------|--|------------|
| Service Account Email | The email address of the service account. Google creates this address within the process of creating a service account. Only for service accounts! Required | String |
| Key File (p12) | The Service Account Login works with private key file for authentication. In the process of creating a service account you download this file. | String |

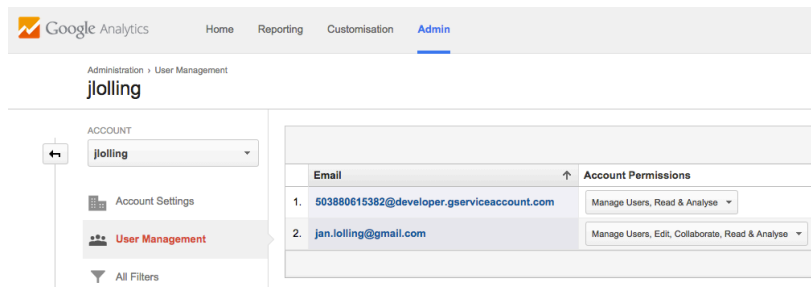
Properties to use the Client-ID for native applications

| Property | Content | Data types |
|---------------------------|---|------------|
| User Account Email | Email of the user account or the Client-ID | String |
| Client secret file (json) | This json file downloaded for the Client-ID | String |

The usage of the “Client-ID for native applications” expects on the first run an user interaction with the Google web page and after finishing the form to approve the access right you need to close the browser to let the component continue, otherwise the authentication process will not complete.

Miscellaneous Properties

| Property | Content | Data types |
|--|---|------------|
| Ignore permission errors while collecting user permissions | If some of the profiles, web properties or accounts are not accessible with “Manage User” rights, with this option the component ignores the problem and continues with the next request. | Boolean |



Example: How to configure an account to allow collecting user permission information.

Advanced Option Parameters

| Property | Content |
|------------------------------|--|
| Timeout in s | How long should the component wait for getting the first result and fetching all result with one internal iteration |
| Static Time Offset (to past) | Within the process of login, the component requests an access token and use the local time stamp (because these tokens will expire after a couple of seconds) Google rejects all requests to access tokens when the request is in the future compared to the timestamp in Google servers. If you experience such kind of problems, this options let the requests appear to be more in the past (5-10s was recognized as good time shift) |
| Reuse Client for Iterations | If you use this component in iterations it is strongly recommended to set this option. It saves time to authenticate unnecessary often and avoids problems because of maximum amount of connects per time range. |
| Distinct Name Extension | The client will be kept with an automatically created name: Talend-Name-Component name + job name. In case this is not distinct enough, you can specify an additional extension to the name. |

Return values

| Return value | Content |
|--------------------------------------|--|
| ERROR_MESSAGE | Last error message |
| ACCOUNTS_NB_LINE | Number of accounts |
| WEBPROPERTIES_NB_LINE | Number of web properties. |
| PROFILES_NB_LINE | Number of profiles. |
| SEGMENTS_NB_LINE | Number of segments |
| GOALS_NB_LINE | Number of goals |
| GOAL_URL_DEST_STEPS | Number of URL destination steps belongs to goals |
| GOAL_EVENT_CONDITIONS | Number of event conditions belongs to goals |
| REPORT_COLUMNS_NB_LINE | Number of column (dimensions and metrics) metadata records |
| ACCOUNT_USER_PERMISSIONS_NB_LINE | Number users have access the accounts |
| WEBPROPERTY_USER_PERMISSIONS_NB_LINE | Number users have access to the web properties |
| PROFILE_USER_PERMISSIONS_NB_LINE | Number users have access to the profiles (views) |

Output flows

The output flows have fixed schemas. The schema columns are completely defined with key, nullable and length. Because of the well-defined schemas, the flows can direct connected to database output components with the option to create the tables.

Output flow: Accounts

| Schema column | Type | Description |
|------------------|--------|---------------------------|
| ACCOUNT_ID | Long | Primary key |
| ACCOUNT_NAME | String | Name |
| ACCOUNT_CREATED | Date | Created at |
| ACCOUNT_UPDATED | Date | Updated at |
| ACCOUNT_SELFLINK | String | Url to manage the account |

Output flow: Web properties

| Schema column | Type | Description |
|-------------------------------|--------|---|
| WEBPROPERTY_ID | String | Primary key, This is the key you use within your web pages to track them. |
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_INTERNAL_ID | Long | Internal ID |
| WEBPROPERTY_NAME | String | Name |
| WEBPROPERTY_SITE_URL | String | URL of the tracked site |
| WEBPROPERTY_CREATED | Date | Created at |
| WEBPROPERTY_UPDATED | Date | Updated at |
| WEBPROPERTY_SELFLINK | String | Url to manage the web property |
| WEBPROPERTY_LEVEL | String | Level for this web property. Possible values are STANDARD or PREMIUM |
| WEBPROPERTY_INDUSTRY_VERTICAL | String | The industry vertical/category selected for this web property |

Output flow: Profiles

| Schema column | Type | Description |
|-------------------------------------|--------|---|
| PROFILE_ID | Long | ID of the profile |
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_ID | String | Primary key, This is the key you use within your web pages to track them. |
| WEBPROPERTY_INTERNAL_ID | Long | Internal ID |
| PROFILE_NAME | String | Name |
| PROFILE_DEFAULT_PAGE | String | |
| PROFILE_EXCLUDE_QUERY_PARAMS | String | |
| PROFILE_SITE_SEARCH_CATEGORY_PARAMS | String | |
| PROFILE_CURRENCY | String | Currency in 3 digits |
| PROFILE_TIMEZONE | String | Time zone (long description) |
| PROFILE_CREATED | Date | Created at |

| | | |
|---|---------|---|
| PROFILE_UPDATED | Date | Updated at |
| PROFILE_SELFLINK | String | URL to manage the profile |
| PROFILE_ECOMMERCE_TRACKING | Boolean | Indicates whether ecommerce tracking is enabled for this view (profile) |
| PROFILE_STRIP_SITE_SEARCH_QUERY_PARAMS | Boolean | Whether or not Analytics will strip search query parameters from the URLs in your reports. |
| PROFILE_STRIP_SITE_SEARCH_CATEGORY_PARAMS | Boolean | Whether or not Analytics will strip search category parameters from the URLs in your reports. |

Output flow: Segments

| Schema column | Type | Description |
|--------------------|--------|---|
| SEGMENT_ID | String | Primary key (only unique within your account) |
| SEGMENT_NAME | String | Name |
| SEGMENT_DEFINITION | String | The filter definition |
| SEGMENT_CREATED | Date | Created at, can be null! |
| SEGMENT_UPDATED | Date | Updated at, can be null! |

Output flow: Goals

| Schema column | Type | Description |
|---------------------------------------|---------|---|
| GOAL_ID | Integer | ID of the goal, the ID is only unique with the profile |
| PROFILE_ID | Long | ID of the profile |
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_ID | String | Primary key, This is the key you use within your web pages to track them. |
| WEBPROPERTY_INTERNAL_ID | Long | Internal ID |
| GOAL_NAME | String | Name |
| GOAL_CREATED | String | |
| GOAL_UPDATED | String | |
| GOAL_ACTIVE | Boolean | |
| GOAL_TYPE | String | Goal type. Possible values are URL_DESTINATION, VISIT_TIME_ON_SITE, VISIT_NUM_PAGES, and EVENT. |
| GOAL_VALUE | Long | |
| VISIT_TIME_ON_SITE_DETAILS_COMP_TYPE | String | Type of comparison. Possible values are LESS_THAN or GREATER_THAN. |
| VISIT_TIME_ON_SITE_DETAILS_COMP_VALUE | Long | |
| VISIT_NUM_PAGES_DETAILS_COMP_TYPE | String | |
| VISIT_NUM_PAGES_DETAILS_COMP_VALUE | Long | |
| URL_DEST_DETAILS_URL | String | URL of the destination (can be relative) |
| URL_DEST_DETAILS_CASE_SENSITIVE | Boolean | Whether or not the URL is case sensitive |
| URL_DEST_DETAILS_MATCH_TYPE | String | Method to match the URL (e.g. REGEX) |

| | | |
|---------------------------------------|---------|---|
| URL_DEST_DETAILS_FIRST_STEP_REQUIERED | Boolean | If the first step is required. See next flow Goal Url Destination Steps |
| GOAL_SELFLINK | String | URL to the Google server to manage the goal |

Output flow: Goal Url Destination Steps

| Schema column | Type | Description |
|-------------------------|---------|---|
| PROFILE_ID | Long | ID of the profile |
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_ID | String | Primary key, This is the key you use within your web pages to track them. |
| WEBPROPERTY_INTERNAL_ID | Long | Internal ID |
| URL_DEST_STEP_INDEX | Integer | Index of the step within the goal |
| URL_DEST_STEP_NAME | String | Name of the step |
| URL_DEST_STEP_NUMBER | Integer | |
| URL_DEST_STEP_URL | String | URL to reach this step |

Output flow: Goal Event Conditions

| Schema column | Type | Description |
|----------------------------|---------|--|
| PROFILE_ID | Long | ID of the profile |
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_ID | String | |
| WEBPROPERTY_INTERNAL_ID | Long | |
| EVENT_CONDITION_INDEX | Integer | Index of the condition within the goal |
| EVENT_CONDITION_COMP_TYPE | String | Condition type |
| EVENT_CONDITION_COMP_VALUE | Long | Value to evaluate the condition |
| EVENT_CONDITION_EXPRESSION | String | Expression to evaluate the condition |
| EVENT_CONDITION_MATCH_TYPE | String | Method to check the condition (e.g. REGEX) |
| EVENT_CONDITION_TYPE | String | |

Output flow: Dimensions and Metrics

| Schema column | Type | Description |
|-----------------|--------|--|
| COL_TYPE | String | Type of the column METRIC or DIMENSION |
| COL_API_NAME | String | Identifier for the usage in the API requests |
| COL_UI_NAME | String | Human readable name in the dashboard |
| COL_DESCRIPTION | String | Description of the dimension or metric |
| COL_DATA_TYPE | String | The data type in capital letters |
| COL_GROUP | String | The category of the column |
| COL_STATUS | String | The status like PUBLIC=use it, DEPRECATED=using the replacement soon |
| COL_REPLACED_BY | String | The replacement column for deprecated columns |
| COL_CALCULATION | String | For metrics the calculation formula if the metric is the result of a calculation, otherwise empty for direct measured metrics. |

Output flow: User Permissions for Accounts

| Schema column | Type | Description |
|-----------------------|--------|--|
| ACCOUNT_ID | Long | ID of the account |
| EMAIL | String | User email address (also in case of using service accounts) |
| PERMISSIONS_LOCAL | String | Comma separated list of the permissions explicit set for the account |
| PERMISSIONS_EFFECTIVE | String | Comma separated list of the permissions implicit given for the account |

Output flow: User Permissions for Web Properties

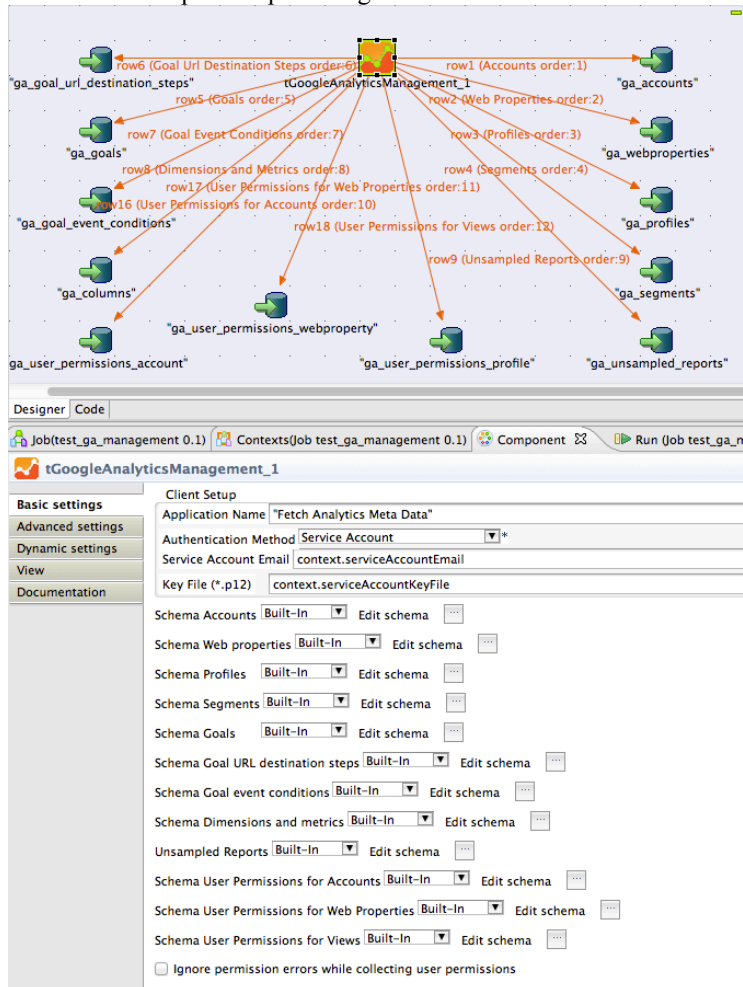
| Schema column | Type | Description |
|-----------------------|--------|---|
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_ID | String | ID of the web property |
| EMAIL | String | User email address (also in case of using service accounts) |
| PERMISSIONS_LOCAL | String | Comma separated list of the permissions explicit set for the web property |
| PERMISSIONS_EFFECTIVE | String | Comma separated list of the permissions implicit given for the web property |

Output flow: User Permissions for Views

| Schema column | Type | Description |
|-----------------------|--------|---|
| PROFILE_ID | Long | ID of the profile |
| ACCOUNT_ID | Long | ID of the account |
| WEBPROPERTY_ID | String | ID of the web property |
| EMAIL | String | User email address (also in case of using service accounts) |
| PERMISSIONS_LOCAL | String | Comma separated list of the permissions explicit set for the view |
| PERMISSIONS_EFFECTIVE | String | Comma separated list of the permissions implicit given for the view |

Scenario: Fetch all meta-data

In this scenario you fetch all data from accounts, web properties, profiles, segments, goals and goal-url-destination-steps and goal-event-conditions, unsampled reports and 3 times user permissions and save them into separate database tables. All database output components get their schema from the flow.



To create this scenario you have to do:

1. Drop the tGoogleAnalyticsManagement component and 12 database output components into the job.
2. Give all 12 database output components its table names and configure them to "Create table if not exists"
3. Than create the 12 flows.
4. Sync all schemas of the database components.

That's it.

Advise to install and use the component

Please ensure, that your service account email address is added as user to all relevant profiles!

You will only retrieve account, web properties, profiles where your service account email is added (to the profiles, that enough).

The best way is to download and install it from Talend Exchange.

In case of missing libraries:

1. Close Studio
2. Delete the file: %Studio install dir%/configuration/ComponentsCache.javacache
3. Start Studio

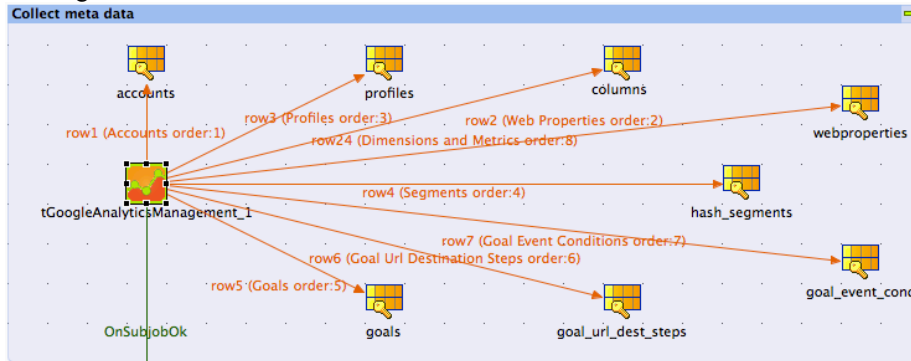
In the Exchange panel you can update or delete your downloaded components.

Next a real live scenario:

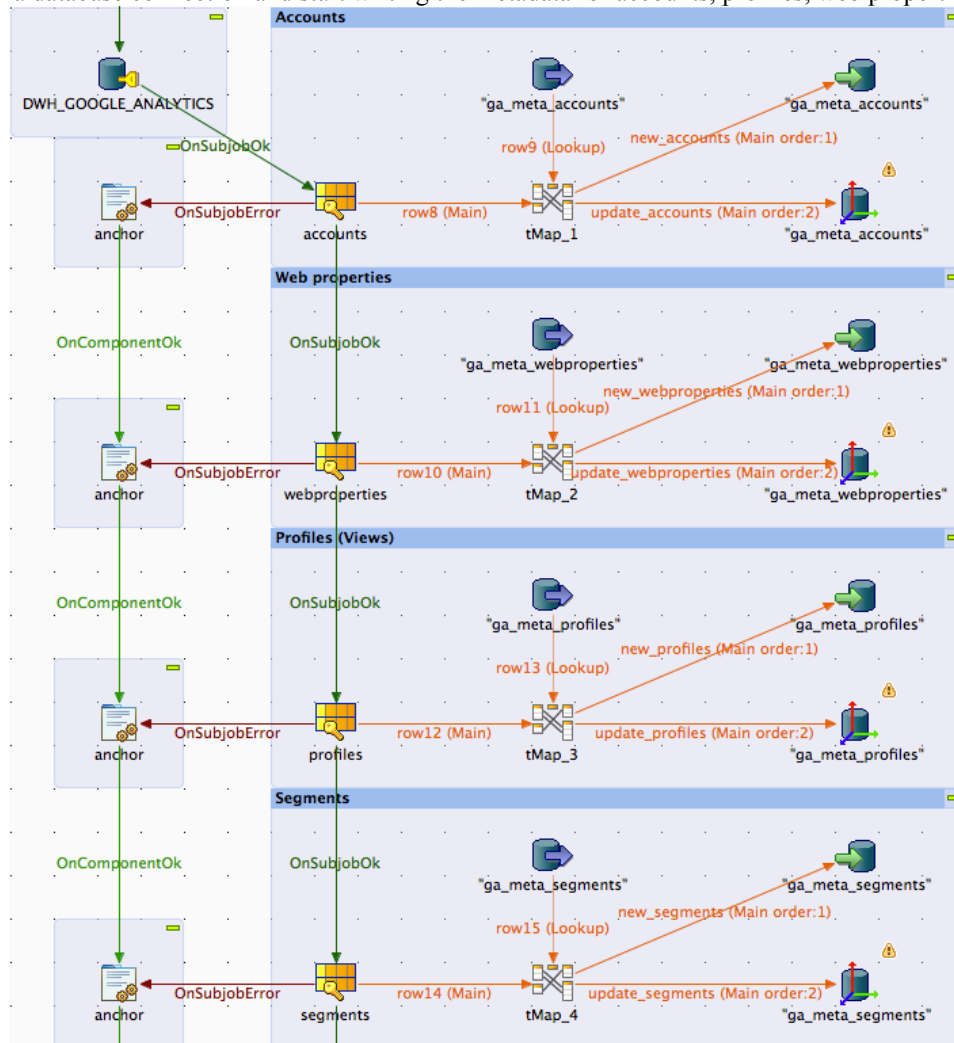
In this scenario the job recognises when some changes in the metadata happens using SCD logic.

The job starts with collecting all metadata into the memory and next starts writing the data within a transaction.

Collecting the metadata....

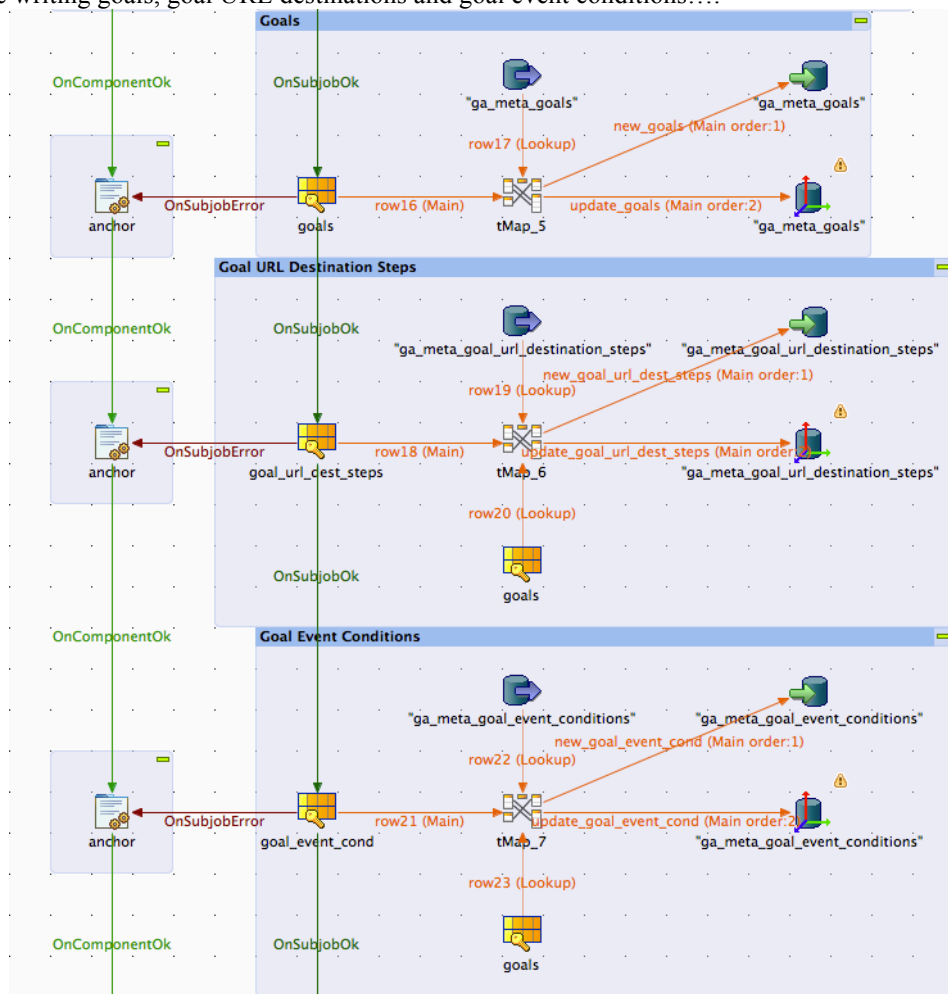


Open a database connection and start writing the metadata for accounts, profiles, web properties and segments...



Continued next page...

Continue writing goals, goal URL destinations and goal event conditions....



And finally the metadata about dimensions and metrics and commit the transaction...

