



Purpose

This component addresses the needs of gathering Google Analytics Multi channel funnels data for an large number of profiles and fine grained detail data.

The component uses the Multi-Channel Funnels Reporting API (v3) and the Authentication API OAuth 2.0 final. This API was introduced by Google in end of July 2012.

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.

Talend-Integration

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

This component provides an output flow and several return values.

Parameters

Parameters to connect to Google Analytics (setup client)

Property	Content
Application Name	Not necessary, but recommended by Google. Simple provide the name of your application gathering data. Required
Service Account Email	The email address of the service account. This address is created by Google within the process of creating a service account. Required
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. Required

Parameters to define the query

Property	Content
Profile-ID	All profiles get an unique integer identifier. This id is needed here. As profile names can be reused within other Google-Accounts and Google-Properties, a profile ID is unique over all accounts. Required!
Start Date	All queries need always a time range (only date, not time). Format must be "yyyy-MM-dd". Required!
End Date	Time range end. If you want gather data for one date, use start date as end date. Format must be "yyyy-MM-dd". Required!
Dimensions	Dimensions are like group clauses. The metrics values will be grouped by these dimensions. See advice for notations below. Not required (since release 1.5)
Metrics	Things you want to measure. Separate all metrics with a comma.

	See advise for notations below. Required!
Filters	Contains all used filters as concatenated string. See advise for notation below
Sorts	Contains all sort criteria as concatenated string. See advise for notation below
Deliver Totals Data Set	The API provides a totals data set. This can be used to calculate percentage values or check results. This data set will be delivered (as first row) if option is checked or will be omitted if option is not checked. Date values remains empty (null) in the totals dataset.

Advice for notation

For dimensions, metrics, filters and sorts you have to use the notation from the Multi-Channel Reporting API:

<https://developers.google.com/analytics/devguides/reporting/mcf/dimsmets/>

All dimensions, metrics, sorts and filters must use the namespace mcf: .

Filters can be concatenated with OR or AND operator.

Separate filter expressions with a comma means OR

Separate filter expressions with a semikolon means AND

Filter comparison operators:

Operator	Meaning
"=="	Exact match to include
"!=="	Exact match to exclude
"=~"	Regex match to include (only for strings)
"!~"	Regex match to exclude (only for strings)
">="	Greater or equals than (only for numbers)
"=@"	Contains string
"!@"	Does not contains string
">"	Greater than (only for numbers)
"<="	Lower or equals than (only for numbers)
"<"	Lower than (only for numbers)

Building Schema for Component

In the schema you need a amount of columns equals to the sum of the number of dimensions and metrics.

Columns in the schema must start at first with dimensions (if provided) and ends with metrics.

Schema column types must match to the data types of the dimensions and metrics. The schema column names can differ from the names of dimensions and metrics. Only the order and there type are important.

Important: For date dimensions you must specify the date pattern as "yyyyMMdd"!

Advanced Option Parameters

Property	Content
Conversion path delimiter	The API provides conversion paths. The path will be delivered in one data field. This delimiter helps to separate the path elements.
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 request 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)
Fetch Size	This is the amount of data the component fetches at once. The value is used to set the max_rows attribute. max_rows means not the absolute amount of data! The component manages setting the start index to get all data. To achieve this, the component iterates as long as the last result set are completely fetched.
Local Number Format	You can get numbers in various formats. Here you can define the locale in which format double or float values are should textual formatted by the API.
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 max. amount of connects per time range.
Distinct Name Extension	The client will be kept with a 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
NB_LINE	Number of delivered lines
CONTAINS_SAMPLED_DATA	True if data are sampled, means not exactly calculated. This can happen if you query to many details.
TOTAL_AFFECTED_ROWS	Number of rows which are collected by Google to calculate the result.

Scenario 1

Simple take all data by fix entered setup for the query.

The screenshot displays the Talend Designer interface. At the top, a data flow is shown from the **tGoogleAnalyticsMCFInput_1** component to the **tLogRow_1** component, connected by a line labeled **row1 (Main)**. Below this, the **tGoogleAnalyticsMCFInput_1** component is selected, and its configuration panel is open. The panel is divided into two main sections: **Client setup** and **Query definition**.

Client setup:

- Application name: "Fetch Analytics MCF Data"
- Service account email: "503880615382@developer.gserviceaccount.com" *
- Key file (*.p12): "/home/jlolling/Documents/cimt/workspace_talend_col" *

Query definition:

- Profile-ID: "59815695" *
- Start date: "2012-05-01" *
- End date: "2012-09-04" *
- Dimensions: "mcf:sourcePath"
- Metrics: "mcf:totalConversions" *
- Filters: (empty field)
- Sort by: (empty field)
- Schema: Built-In (dropdown menu) Edit schema ... ☒ Deliver totals data set

The component works very similar to the **tGoogleAnalyticsInput** component.

You can set profile-ID or service account informations and all other properties from context variables – this is the best approach for a stable and maintainable operation.