Tableau Online
Understanding Data Updates

Author: Francois Ajenstat
July 2013

Whether your data is in an on-premise database, a database, a data warehouse, a cloud application or an Excel file, you can analyze it with Tableau. You can create views of your data and share it with colleagues, customers, and partners. You can use Tableau to blend it with other data. And you can keep your data up to date automatically.

For data in the cloud, Tableau Online provides a number of approaches to keeping your data up-to-date. For cloud applications like Google Analytics or Salesforce.com, Tableau Online can automatically update snapshots of your data (Tableau Data Extracts) on a scheduled-basis. For data warehouses based in the cloud, such as Google BigQuery and Amazon Redshift, Tableau Online can connect directly. This live connection means you do not have to create creating extracts since BigQuery and Redshift support fast query for large amounts of data over the Internet.

Of course, not all data lives in the cloud. Data sources that are within your firewall can be refreshed by pushing data to Tableau Online using desktop tools. You schedule and manage these updates within your own environment. This approach minimizes any security risk while providing you with flexibility to keep your data up-to-date on an appropriate schedule.

Tableau Online Architecture

Tableau Online is a secure cloud-based solution to share, collaborate and distribute visualizations and dashboards. Tableau Online is built on the Tableau Server platform and provides ease-of-use, speed, availability and security without having to manage physical infrastructure. Tableau Online runs in the cloud and therefore does not have access to your on-premise data.

Tableau Online runs in a secure data center and can be accessed by clients remotely using Tableau Desktop, a browser and a mobile device. All communication with Tableau Online is performed using SSL.
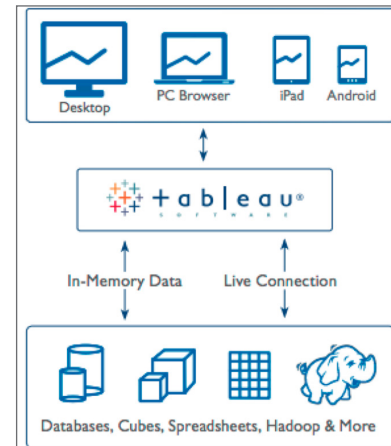
Connecting to Google Analytics and Salesforce.com

Tableau supports native connectivity to Google Analytics and Salesforce.com. You can connect to these data sources directly from Tableau Desktop and download the data as an extract on the desktop. Once the initial connection has been made and the data has been extracted, you can publish the workbook or data source to Tableau Online. Tableau will provide you with an option to refresh the extract on a specific schedule. You can choose an incremental refresh in which case Tableau will only download the latest rows from the data source. Or, you can select to download all of the rows and perform a full refresh.

In order to schedule these refreshes you will be asked to provide the credentials for the user that will be used to connect and download the data. This will establish a secure connection to the data source and create an Authentication key that will be used during the refresh ensuring the most secure connection to your data.

Connecting to Google BigQuery and Amazon Redshift

Tableau Online provides live connectivity to Google BigQuery and Amazon Redshift. Since these cloud data warehouses were designed to be scalable platforms for interactive analysis of massive datasets, Tableau can support a live, interactive connection to each of these data sources. As a result, you will not need to create an extract when publishing the data source or workbook to Tableau Online. This ensures that you always get the latest data from the Server and do not need to replicate the data in Tableau Online.

Connecting to On-Premise Data

Since Tableau Online runs in the cloud, it does not have access to the data sources that reside within your corporate network. In order to access this data, you will need to 'push' this data to Tableau Online that can be done manually or automatically on a given schedule.

To publish this data, you must first extract the data into Tableau's fast data engine. Once the extract has been created, you can then publish the workbook or the data source to Tableau Online. As a best practice, we recommend publishing the data source separately from the workbook in order to simplify data updates and ensure consistency across workbooks.

Once the data has been published to Tableau Online, there are 3 ways to update your data:

1. Republish the Data Source or Workbook

Once you refresh your local data extract, you can manually republish the data source or workbook to Tableau Online to update the data and metadata. This is done via Tableau Desktop and will overwrite the data in Tableau Online.

We recommend publishing commonly used data sources separately from workbooks. Once the data is published, it can be accessed in your workbooks by connecting to Tableau Server data sources. This ensures consistency across workbooks and minimizes

the amount of data transfer to the Server. Workbooks that leverage the centralized data sources will always have the latest data and will provide tighter security controls on your data.

2. Refresh the Tableau Server Data Source

Published data sources are referred to as Tableau Server data sources. When you are using a Tableau Server data source, you will have the ability to manually refresh the Server data source. This method will use your local Tableau Desktop as a proxy between your data source and Tableau Online. All of the metadata to connect to the database are embedded in the data source and will be used for the update. Tableau will then connect to your on-premise data source, extract the data (either full or incremental) and then send the updates to Tableau Online to refresh the extract.
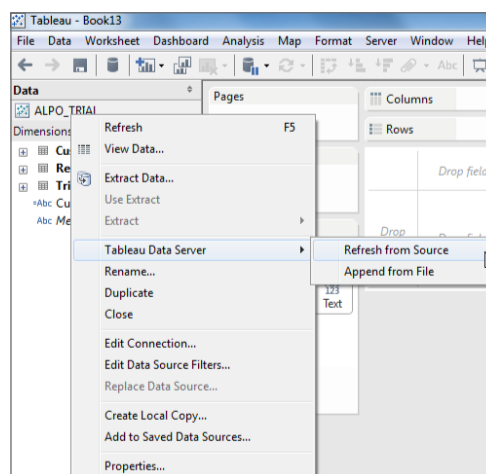
3. Use Command-Line Tools to Schedule Batch Updates Locally

The previous 2 methods are manual operations that need to be performed using Tableau Desktop. If you want to automate the updates to your Server data sources than you can accomplish this using some straightforward command-line tools. These tools are installed with Tableau Desktop.

You can add these operations to a Batch script and automate the process using Windows Task Scheduler. This can be done on a particular schedule or after an event has occurred on the system. You have full flexibility on the frequency of these updates.

For more information on the Data Extract tools, *click here*.

Here are some basic examples of the commands to use to update your Tableau Online data sources.

If your data source is a file (Excel, CSV, Access):

```
tableau refreshextract -s https://online.
tableausoftware.com -t "SITENAME"
--datasource TEST -u myUSERNAME -p
myPASSWORD --original-file "c:\Downloads\
some content.xlsx"
```

If your data source is a database with Windows Authentication – you connect as the local desktop user and are not prompted for credentials:

```
tableau refreshextract -s https://online.
tableausoftware.com -t "SITENAME"
--datasource TESTV1 -u myUSERNAME -p
myPassword
```

If your data source is a database that requires authentication:

```
tableau refreshextract -s https://online.
tableausoftware.com -t "SITENAME"
--datasource TESTV1 -u myUSERNAME -p
myPassword --source-username test
--source-password password
```

You can also define these parameters via a configuration file. A configuration file can be useful if you expect to be updating the same data source multiple times. It also has the advantage of not exposing usernames and passwords on the command line.

```
c:\Program Files (x86)\Tableau\Tableau
8.0\bin>
tableau refreshextract -c
UploadFromFile.txt
server=https://online.tableausoftware.com
site=SITENAME
datasource=TEST
username=myUSERNAME
password=myPASSWORD
source-username=test
source-password=password
```

Best Practice For Centralized Management of Data Sources

A best practice when using Tableau Online is to leverage the Data Server to centrally manage Data

Extracts and database and shared those extracts across workbooks. This ensures consistency across workbooks and minimizes the amount of data transfer to the Server. Workbooks that leverage the centralized data sources will always have the latest data and will provide tighter security controls on your data.

The published data sources include all calculations, user filters, sets and groups created in Tableau, so both definitions and data can be easily shared. The extracts are stored in Tableau Online rather than downloaded/uploaded for each workbook. Changes to the master data source can be automatically propagated to workbooks that use the data source. At any stage of creating a data source, the metadata can be published to Tableau Online reducing complexity and administrative overhead.

To publish a data source to Tableau Online, right-click on the data source and select the 'Publish to Server' option. Publishing a data source is very similar to publishing a workbook to Tableau Online. A user with the appropriate permissions can publish the data source including all of its metadata to Tableau Online.

You can also define permissions on the data source to specify which users are allowed to connect, modify, or download the data source. This ensures that only authorized users can access this data source.

Once the data source has been published, you can use it by choosing the 'Tableau Server' option in the Connect to Data screen. All of your published data sources that you have access to will be available to use. Once you are connected to the Tableau Server data source, the workbook will always contain the latest data and metadata published to Tableau Online.

About Tableau

Tableau Software helps people see and understand data. Tableau helps anyone quickly analyze, visualize and share information. More than 12,000 customer accounts get rapid results with Tableau in the office and on-the-go. And tens of thousands of people use Tableau Public to share data in their blogs and websites. See how Tableau can help you by downloading the free trial at *www.tableausoftware.com/trial*.

Tableau and Tableau Software are trademarks of Tableau Software, Inc. All other company and product names may be trademarks of the respective companies with which they are associated.