

Big Data Sets

 Big Data Sets are available since BellaDati 2.9

Big Data Sets are special types of data sets that can be used to store a very large amount of data and build pre-calculated cubes. The main differences between standard data sets and big data sets are:

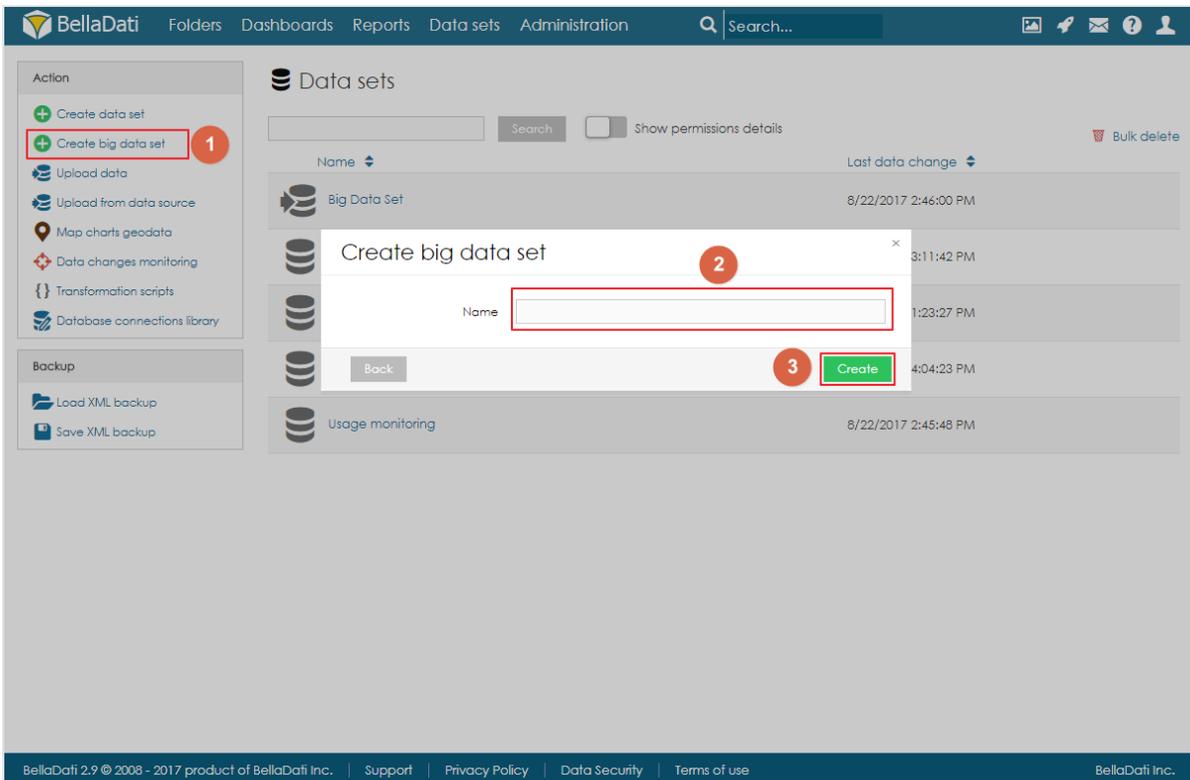
- Reports cannot be built directly on big data sets. Cubes have to be created first. Since **BellaDati 2.9.17**, Reports can be built directly on big data sets.
- In big data sets, it is not possible to browse all data. Only a random data sample is available. Filters, edit function, and delete function are not available in the data sample.
- Big data sets cannot be joined.

 The main advantage of Big Data Sets is the ability to create pre-calculated cubes resulting in a rapid speedup of reports loading times.

Creating Big Data Set

 Please note that Big Data Set functionality needs to be enabled in the license and in the domain.

Big Data Set can be created by clicking on the link **Create big data set** in the Action menu on the Data Sets page, filling in the name of the big data set, and clicking on **Create**.



The screenshot shows the BellaDati web interface. The top navigation bar includes 'BellaDati', 'Folders', 'Dashboards', 'Reports', 'Data sets', and 'Administration'. A search bar is on the right. The left sidebar contains an 'Action' menu with 'Create big data set' highlighted by a red box and a red circle with the number '1'. The main area is titled 'Data sets' and shows a table of existing data sets. A modal dialog box titled 'Create big data set' is open, with a red box and a red circle with the number '2' around the 'Name' input field. At the bottom of the dialog, a 'Create' button is highlighted by a red box and a red circle with the number '3'. The footer of the page contains copyright information and links for 'Support', 'Privacy Policy', 'Data Security', and 'Terms of use'.

Big Data Set summary page

The landing page (summary page) is very similar to the standard data set summary page. There are a left navigation menu and the main area with basic information about the data set:

- description,
- date of last change,
- records count,
- cubes overview,
- import history.

Importing Data

Data can be imported to a big data set the same way as to standard data set. Users can either import data from a file or from a data source. However, the big data set is not using standard indicators and attributes, but instead, each column is defined as an object. These objects can have various data types:

- text,
- date,
- time,
- DateTime,
- GEO point,
- GEO JSON,
- long text,
- boolean,
- numeric.

After import, users can open the data sample page to see a randomly selected part of the data.

The screenshot shows the BellaDati interface. The top navigation bar includes 'BellaDati', 'Folders', 'Dashboards', 'Reports', 'Data sets', 'Administration', and a search bar. The main content area is titled 'Big Data Set - Data sample'. It features a table with the following columns: ClosedDate, CreationDate, DeletionDate, Id, and Score. The table contains 42 rows of data. The interface also includes a left sidebar with 'Settings' (Data set summary, Data model (7), Cubes (0), Advanced settings), 'Data' (Upload data, Data source, Data sample, Erase data), and 'Action' (Share with users, Duplicate, Save XML backup, Delete Data Set). The footer shows 'BellaDati 2.9 © 2008 - 2017 product of BellaDati Inc.' and 'Records count: 2,000,000'.

Managing Objects

Objects (columns) can be created automatically during the import, or they can be defined on the **data model** page. When adding a new object, users can specify its name, data type, indexation, and whether it can contain empty values or not. Please note that GEO point, GEO JSON, long text, boolean and numeric cannot be indexed.

Objects can also be edited and deleted by clicking on the row.

Cubes

Cube is a data table that contains aggregated data from the big data set. Users can define the aggregation and also limit the data by applying filters. Data from the cube can then be imported to a data set. Each big data set can have more than one cube, and each cube can have different settings.

Creating Cube

To create a cube, users need to follow these steps:

1. Click on Create cube
2. Fill-in the name and optionally the description.
3. Select which columns (attribute elements and data elements). Attribute elements define the aggregation of the cube. For example, if the user selects column Country, the data will be aggregated for each country (one row = one country). Users can also create formula indicators. In real-time, users can also see the preview of the cube on the right side of the screen. Please note that the preview is built on the data sample only, which means that it can be empty, although

some data will be imported to the data set after the execution. It is possible to change the order of attributes and indicators by using the arrows located next to the names.

The screenshot shows the BellaDati interface for configuring a data set. The main area displays a table with columns for date, company, value2, and value1. The table contains 15 rows of data. Below the table, there are buttons for 'Refresh' and 'Estimate results'. The sidebar on the left contains sections for 'Settings' (Data set summary, Data model, Cubes, Advanced settings, Partitioning) and 'Data' (Upload data, Data source, Data collecting forms, Data sample, Erase data). The 'Action' section includes 'Share with users', 'Duplicate', 'Save XML backup', and 'Delete Data Set'.

> date	< company	> value2	< value1
2017-10-12	Ac Incorporated	178	178
2017-10-12	Ac Ipsum Limited	480	480
2017-10-12	Adipiscing Corporation	411	411
2017-10-12	Adipiscing Inc.	590	590
2017-10-12	Aliquam Corporation	322	322
2017-10-12	Aliquam Institute	429	429
2017-10-12	Aliquam Omare PC	563	563
2017-10-12	Aliquam Rutrum Consulting	918	918
2017-10-12	Aliquet Consulting	123	123
2017-10-12	Amet Consectetur Consulting	300	300
2017-10-12	Amet Ltd	1721	1721
2017-10-12	Arcu Company	184	184
2017-10-12	At Iaculis Corp.	800	800
2017-10-12	Auctor Vitae PC	128	128
2017-10-12	Augue Porttitor Interdum Ltd	364	364

4. Optionally, users can also apply filters to work with only part of the data.
 - a. In the filters, users can reference first and last values from different data set by using the following functions:

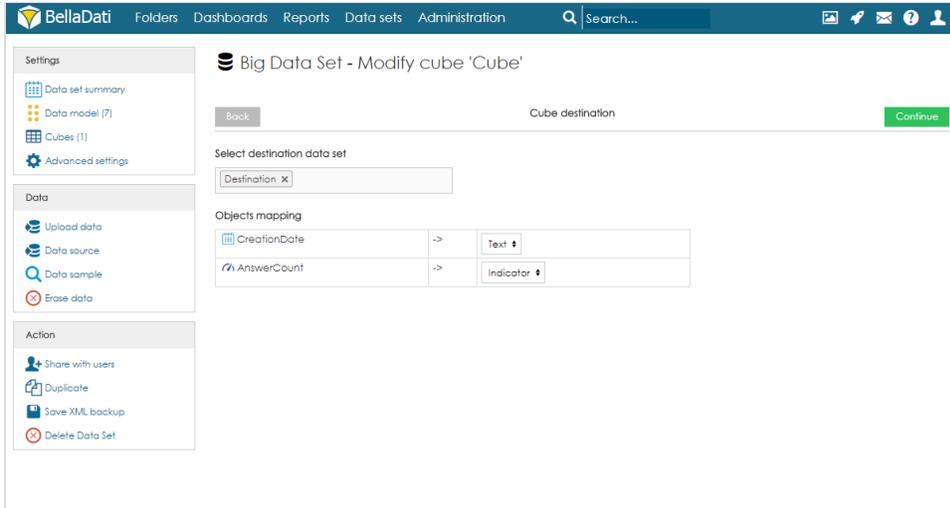
The function has to be added as a custom value to the filter.

The 'Filter settings' dialog box shows the configuration for an attribute filter. Under 'Attributes', there is a logical operation of 'AND'. Two attribute filters are defined: 'Month' with the operator 'contains' and value '10', and 'year' with the operator 'contains' and a custom value: '\$(firstValue(LOOKUP,L_YEAR))'. The custom value is highlighted with a red box. The 'Indicators' section is currently empty.

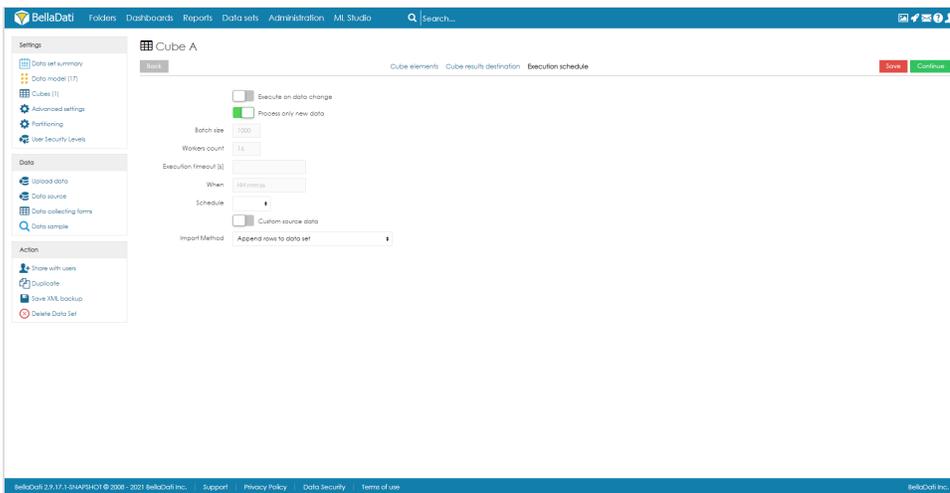
- b. It is also possible to add a filter formula. This allows users to create more complex filter algorithms. It is also possible to use the function `getLastSuccessfulCubeExecution()` or `getCurrentCubeExecution()` to get the date and time of the last successful or current cube execution correspondently.

5. Select destination data set and mapping. By using the search field, users have to select the destination data set. After execution, data will be imported from the cube to this data set. After choosing the data set, users have to specify the mapping. Each column of the cube has to be assigned to an attribute or indicator of the destination data set. Attribute elements can be mapped to attribute columns in the destination data set. Data elements can be mapped to indicator columns and

also attribute columns.



6. Set up execution schedule. The execution can be run manually, on data change, or by schedule. When scheduling the execution, users can specify the following parameters:
 - a. **Process only new data** - an option to automatically filter new data
 - b. **Batch size** (default 1000) - the number of rows that will be executed in one batch. In special cases, it might be beneficial to increase or decrease the value. However, in most cases, we strongly suggest leaving it on default.
 - c. **Workers count** (default 8) - the number of workers that should be used for parallel execution.
 - d. **Execution timeout [s]** - sets the maximum duration of the execution.
 - e. **When** - time of the first execution.
 - f. **Schedule** - how often should be executed.
 - g. **Import Method** - what should happen with data in the destination data set. See [Data overwriting policy](#) for more information.



Cubes summary

On the Cubes page, users can see a table with all cubes associated with the big data set. For each cube, information about the schedule and last event are available. Users can also edit the cube by clicking anywhere on the row or on the name of the cube. Several actions are also available for each cube:

- History - see a list of previous executions and their result.
- Run - manually run the execution.
- Schedule - reschedule the execution. Previous settings will be overwritten.
- Delete - delete the cube.

The screenshot shows the BellaDati interface for managing cubes. The top navigation bar includes 'Folders', 'Dashboards', 'Reports', 'Data sets', and 'Administration'. The main content area is titled 'Big Data Set - Cubes' and features a 'Create cube' button. Below this is a table with the following columns: Name, Destination data set, When, Schedule, Execute on data change, Last Event, and Action. A single row is visible with the following data: Name: Cube, Destination data set: Destination, When: 10:25:00 AM, Schedule: Month 8/23/2017 10:25:00 AM, Execute on data change: (indicated by a red 'x' icon), Last Event: (empty), and Action: History | Run | Schedule | Delete.

Name	Destination data set	When	Schedule	Execute on data change	Last Event	Action
Cube	Destination	10:25:00 AM	Month 8/23/2017 10:25:00 AM			History Run Schedule Delete

Cube execution

As mentioned above, execution can be run manually, on data change, or by schedule.

- Manual execution - by clicking on Run in the Action column, users can start the execution manually. They can also select the import method, which can be different than the one used for scheduled execution.
- On data change - every time there is a change in the big data set, the execution will be started. If there is an already running cube execution, the new execution is placed in a queue and automatically triggered after the previous cube execution is completed.
- Scheduled execution - execution will be run periodically after a specified amount of time.

Users can also cancel the next scheduled execution by clicking on the date in the column Schedule and confirming the cancellation. Please note this will only cancel the execution and it won't delete it. After running the execution manually, the schedule will be restored. To delete the scheduled execution completely, users have to edit the cube and delete the Execution schedule.



Limits for concurrent cube executions are available since BellaDati 2.9.18

Since the number of cubes could be large, there is a limit to the number of cubes that will be executed at the same time. By default, the number of concurrent cube executions is 5. To change it, open **Configuration in Administration**. The setting **Import executor pool size** defines this limit.



This setting is available for users with the role **domain admin** only

BellaDati Folders Dashboards Reports Data sets Administration Search...

Configuration

Domains

Users

Monitoring

System Administration

Configuration

General

Property Name	Property Value
Application name	BellaDati
Company name	
Company website	
Enable autoregistration	⊖
Max failed logins count	
Suspended materialization interval (HH:mm:ss)	
Email sender/recipient	support@belladati.com
Support email	
Application URL	https://test.belladati.com
External application URL	
ODRS origin	
Global usage statistics	⊖
Maximum upload file size (in bytes)	30000000
Server file system access enabled	⊖
Backup directory	
Disable login alert	⊖
Disable cross joins	⊖
Billy Access Token	
Import executor pool size	5

SalesForce

Property Name	Property Value
ConsumerKey	
ConsumerSecret	

Look & Feel

Property Name	Property Value
StyleSheet	
Custom HTML	
Login stylesheet	
Email template - end of head	
Email template - beginning of body	
Email template - end of body	
Email template - links to documentation	
Custom login alert	
Favicon	
Application logo	
Application logo dark	
Login logo	
Report view timeout[s]	

SMP server

Property Name	Property Value
Host	email-smp.us-east-1.amazonaws.com
Port	465
SSL/TLS	✓
Login	AKIAJ6RRVWNTYOYNKA
Password	*****

Facebook

Property Name	Property Value
Facebook login enabled	⊖
ApplicationId	

BellaDati 3.0.0-SNAPSHOT © 2008 - 2021 BellaDati inc. Support Privacy Policy Data Security Terms of use BellaDati inc.

Backup of Big Data Set

When using the XML backup of the big data set, the target data set and mapping in the cube are not stored. After restoring, they have to be set up again.