

# Data Preparation R-1.1



## Contents

1.	Abo	out this Guide			
	1.1.	Document History	. 4		
	1.2.	Overview	. 4		
	1.3.	Target Audience	. 4		
2.	Intro	oduction	. 4		
	2.1.	Introducing the Big Data BizViz Data Preparation	. 4		
	2.2.	Prerequisites and Supported Devices	. 4		
3.	Gett	ing Started with the BDB Data Preparation	. 4		
	3.1.	Accessing the BDB Data Preparation	. 4		
	3.1.:	1. Forgot Password Option	. 6		
4.	Basi	c Features	. 9		
	4.1.	Workflow Editor	. 9		
	4.2.	Extracting Data: Full and Incremental	. 9		
	4.3.	Loading Data	12		
	4.4.	Saving a Workflow	14		
	4.5.	Run Preview	15		
	4.6.	Save and Execute	16		
	4.7.	Schedule a Workflow	16		
	4.8.	Job	17		
	4.9.	Trash	18		
5.	Tran	isform	18		
	5.1.	Constants	18		
	5.2.	Data Type	20		
	5.3.	Date Operations	22		
	5.4.	Filter	23		
	5.5.	Formula Fields	24		
	5.6.	Group By	25		
	5.7.	Mapping	27		
	5.8.	Replace Text	29		
6.	Mer	ge	31		
	6.1.	Append	31		
	6.1.3	1. Append All Columns	31		
	6.2.	Join	35		



	6.2.1	. Join Types:	1
7.	Scheo	duler	L
	7.1. 9	Schedule Configuration Options	)
8.	Signiı	ng Out	ł



## 1. About this Guide

## 1.1. Document History

Product Version	Date (Release date)	Description		
BizViz Data Preparation 1.0	August 31 <sup>st</sup> , 2017	First Release of the document		
BizViz Data Preparation 1.1	December 11 <sup>th</sup> , 2017	Updated document		

#### 1.2. Overview

This guide covers:

- Introduction and steps to use the Big Data BizViz ETL plugin
- Configuration details for the Data Preparation components

#### 1.3. Target Audience

This guide is aimed at business users of all skill levels who deal with vast amounts of data and requires data preparation to be attempted before getting informative insights from the collated business datasets.

## 2. Introduction

## 2.1. Introducing the Big Data BizViz Data Preparation

The BDB Data Preparation is a self-service data preparation tool that empowers data-driven Business users with powerful capabilities to extract, transform, and merge new data sources. The tool offers a range of components to transform and merge the selected dataset. Users can get analytics-ready data faster to generate valuable insights in less time.

## 2.2. Prerequisites and Supported Devices

- A browser that supports HTML5
- Operating System: Windows 7
- Basic understanding of the BizViz Server

## 3. Getting Started with the BDB Data Preparation

## 3.1. Accessing the BDB Data Preparation

This section explains how to access the BizViz Platform and a variety of plugins that it offers:

- i) Open BDB Enterprise Platform Link: <u>http://apps.bdbizviz.com/app/</u>
- ii) Enter your credentials to Login.
- iii) Click 'Login'



	Welcome to Big Data BizViz (BDB)	
	Email	Decision Platform
	Password	Big Data Pipeline Framework
	Forgot password?	Dashboard Designer
	Enterprise •	ETL (Self-Service Data Preparation)
		Geospatial Analysis (Location Intelligence)
		Predictive Analysis
		Play (Beta Release)
	Login	Self-Service BI (Business Story)
		Social Media Browser
		Sentiment Analysis
	Copyright © 2015-2017 BDB (BizViz Technologies Pvt Ltd)	• Survey
$/$		

Users will be redirected to the BDB Platform homepage. iv)

			Search 🗙 🔳
Decision Platform	A complete decision platform for all your business needs. Drive from data to dynamic visuals and derive an actionable insight into your business data. Avail 360° view of your business by assembling, processing, and analyzing the acquired data. Access incomparable analytics from	Dashboard Designer Design, save and publish splendid visual reports as dashboards. Display informative progress report of any business process containing series of stunning visuals to denote informative business data. Drag and drop functionality to access a comprehensive	BizViz Sentiment Analysis BizViz Sentiment Analyzer is a text analyzer that chases words with positive, negative, or neutral connotation. Empowered with Natural Language Processing (NLP) and machine learning algorithms the tool can identify, extract, and exhibit sentiments out of any
Big Data BizViz Release 3.2	anywhere, at any time on any device. Business Story (Self- service BI) Go beyond the classic BI with our ground-breaking tool, BizVtz Business Story. Innovative yet easy, flawless but rapid, and systematic yet spontaneous, this visualization tool will make you go 'Wow for what it offers. Select relevant data, generate immaculate views, and concoct a perlinent insight into your businessAll this on your ownIII	view of relevant KPIs regarding a business objective. Predictive Analysis Plan your next business move based on the reliable information instead of intuition. Let the power of advanced statistical analysis and machine learning technology take care of your business barriers. Access and apply consistently communicative predictive models to reduce the risk factor and maximize the future opportunities.	expressive text. Equipped with ANEVV dictionary ('Affective Norms for English Words'), the tool rates selected set of words in terms of pleasure, activation, and dominance to create a standard terminology for use in studies of emotion and attention. Use of dotter personalized text as subject matter to be analyzed based on measures of "pleasure" along with "activation".
	Data Preparation	🌔 Geospatial Analysis	Interact with a real audience and collect reliable data in real time. Track unlimited
	Experience a secure yet self-driven mode	Access business information for the distant geographic locations at the tip of	questions and responses via BizViz survey collectors. Analyze and Publish

- v) Click 'App Menu' option
- All the available plugins will be listed in the displayed window. vi)
- Click 'Data Preparation' vii)



Home	Admin	User r
	2	<b>E</b>
Data Center	Designer	Survey
<b>⇔⇔</b>	C	
Sentiments	Play	Report
<b>\$</b>	ft in g	
Data Preparation	Social Media	Predictive
6		
GeoSpatial		

- viii) Users will be redirected to the Data preparation landing page.
- ix) Users will find four major modules on the Data Preparation landing page:
  - a. My Workspace (Default Component)
  - b. Job
  - c. Trash
  - d. Scheduler

	Decision Platform		 •		?	+
Data	preperation 1.1.0			N	EW	
	My Workspace	My Workspace (0)				
ů	Job					
Î	Trash					
S	Scheduler					

This document will describe all the major components and the related workflows at details.

## 3.1.1. Forgot Password Option

Users are provided with an option to change the password.

- i) Navigate to the Login page.
- ii) Click 'Forgot Your Password?' option.



We	elcome to	Big Dat	a BizViz	(BDB)
	Email			
	Password			
	Enterprise		Forgot pass	sword?
		Login		

- iii) Users will be redirected to a new window.
- iv) Provide the email id that is registered with BDB to send the reset password link.
- v) Click 'Continue'

Having trouble signing in	n?
To reset your password, enter the email address This can be your email address associated with	you use to sign in to BizViz. your account.
Email address	
Email address	
CONTINUE	

vi) Users will be directed to select a space and continue.





- vii) A reset password link will be sent through email.
- viii) Click on the link.
- ix) Users will be redirected to the 'Reset Password' page to set a new password.
  - a. Set a new password.
    - b. Confirm the newly set password.
    - c. Click 'RESET PASSWORD'



x) The password will be successfully reset.



## 4. Basic Features

The landing page of Data Preparation launches workspace view. 'My Workspace' will be displayed by default.

#### 4.1. Workflow Editor

'My Workspace' is a placeholder for the workflows which are created using various data preparation components. Users can create the workflows using the workflow editor.

- i) Navigate to the 'Workspace' page.
- ii) Click 'New'

	Decision Platform		 1	?	:
Data	preperation 1.1.0			NEW	
	My Workspace	My Workspace (0)			
ů	dof				
Î	Trash				
C	Scheduler				

- iii) Users will be redirected to the 'Workflow Editor.'
- iv) The Workflow editor exposes users to 3 main aspects to autonomously prepare data:
  - a. Data
  - b. Transform
  - c. Merge

Search Component	← : ⊗ ⊡ ≔
	Untitled Workflow
Input Output	
TRANSFORM T	
MERGE -	Drad & drap your components here to get started
	Diag & diop your components here to get started

- 4.2. Extracting Data: Full and Incremental
  - i) Navigate to the Workflow Editor.
  - ii) The 'Data' option will be selected by default.



#### iii) Drag and Drop 'Input' component onto the workflow editor.

Search Component	← : ⊗ ⊟
DATA •	Untitled Workflow
Input Output	
TRANSFORM -	
MERGE -	

- iv) Use right-click on the dragged input component
- v) A new window will be displayed to configure the input data.

Search Component	← : ⊗ ∎ :=
DATA •	Untitled Workflow
Input Output	CONFIGURATION DATA PREVIEW
C TRANSFORM	Select DataBase 👻
MERGE -	
	No DataSet Selected

vi) Select a database using the drop-down menu (At present only MYSQL, MSSQL, Oracle, and Google Sheet are supported).



P	Input 🖍	
C	ONFIGURATION	DATA PREVIEW
	MYSQL	
	MSSQL	
	Oracle	
	Google Sheet	

- vii) Selecting a database will redirect users to the list of query services based on the selected database.
- viii) Select a query service from the list.
- ix) The basic information of the database and query service will be displayed (By Default).

Ð	Input 🖌				0	Î	^	×
со	NFIGURATION	DATA PREVIEV	V					
	MYSQL	•	BASIC INFO	SETTINGS				
	Q, Sales	☑	DataBase Name	etl				
		A	QueryName	ETL_SalesReport				
		_	DataSource Nam	eetl				
	ETL_SALESREPORT		Query	SELECT * FROM etl.	Sales	Report		

- x) Click the 'Settings' tab.
- xi) Users will be redirected to enable 'Increment Load' to access the recently updated data.
- xii) By enabling the 'Increment Load,' Users need to configure the following options:
  - a. 'Primary Key'- Select a primary key of the data source.
  - b. 'Delta Load'-Select a column of type timestamp or date or long which is updated whenever a a new row is inserted or updated in the data source. This column will be used to perform the 'Incremented Load'



□ Input ✓			(	9 î	^
CONFIGURATION DATA PREVIEW					
MYSQL 🗸	BASIC INFO	s			
Q Search Query Services	Incremental Load		<b>Q</b> Search Co	lumn	
ETL-FILTER-INPU	Columns	Туре	Primary Key	Delta Loa	id
ETL-GROUPBY-TES	Data	Data & Tima	$\checkmark$	$\bigcirc$	*
ETL-LOCATION	Date	Date & Time		Ū	- 1
	LocationId	Whole Number			- 1
ETL-SALESREPORT	ProductId	Whole			•

Note: Users can choose not to enable the increment load. In this case, the following details will be displayed, and the full data will be extracted.

□→ Input 🖍		0 î ^
CONFIGURATION DATA PREVIEW		
MYSQL -	BASIC INFO	
Q Search Query Services	Incremental Load	Q Search Column
1	Columns	Туре
1	Date	Date & Time
10KDATA	LocationId	Whole Number
10KDATA		

## 4.3. Loading Data

Users can load the extracted data into an elastic for visualization via the output component.

- i) Drag and drop the 'Output' component on the Workflow editor.
- ii) Connect it with the configured '**Input**' component.



DATA •	Sample Data Preparation WF
Input Output	
😓 TRANSFORM 🗸	
Merge •	

- iii) Click on the 'Output' component to display the 'CONFIGURATION' option.
- iv) Users will get the following options:
  - a. Elastic
  - b. RDBMS
- v) Select an option and configure it

Ð	Output 🖍		
CO	NFIGURATION	DATA PREVIEW	/
	Connector	Elastic	•
	Connector	Elastic Elastic	•

#### a. Configuring Elastic

- i. Select a resource using the drop-down menu (for the Elastic writer)
- ii. Enable 'Select Mapping ID' option-By enabling this choice users will be redirected to select a mapping id from the 'Mapping id' drop-down menu.

- Output 🖌	
CONFIGURATION	DATA PREVIEW
Select resource ETL-Sales Report	✓ Select Mapping ID

Note: If the 'Select Mapping Id' option is enabled, users will be asked to configure the mapping id using the drop-down menu:



- Output 🖍	0	Î	^	×
CONFIGURATION DATA PREVIEW				
Select resource Mapping Id etLincdemo17aug Select Mapping ID dob				
Or b. Configuring RDBMS i. Select a Data Source Type ii. Select Data Source Name iii. Select Database Name iv. Select Table Name v. Select 'ADD' + option to Create a New Table				
- Output 🖍				
CONFIGURATION DATA PREVIEW				
Connector RDBMS -				
CONFIGURATION PROJECTION				
Select Data Source Type '		•		
Select Database Name '		•	÷	
<ul> <li>vi. Choose Table Operation</li> <li>1. Overwrite</li> <li>2. Append</li> <li>3. Upsert</li> <li>vii. Click 'APPLY'</li> </ul>				
Choose Table Operation Overwrite Operation Append OUpsert				

## 4.4. Saving a Workflow

Users are provided with two options to save a workflow.

- i)
- Click the 'Save' option **a** A new window pops-up to redirect the user to save the workflow. ii)



- a. Enter a Workflow name
- b. Enter Description (Optional)
- c. Select or Add a Workspace

iii) Click 'Save'

Save Workflow	×
Workflow Name * Sales Report	
If you want, you can add a description to explain what you changed	
Description	
Workspace       Sample Workspace	
CANCEL SAVI	E

#### 4.5. Run Preview

Users can run the created workflow without affecting their production system through '**Run Preview**' option. Users need to save the workflows to get the 'Run Preview' option.

- i) After saving a workflow, Users will be able to access more options on the workflow editor toolbar.
- ii) Click 'Run Preview' option 🕨
- iii) The ongoing execution process will be displayed through a continuous blue line.
- iv) Users will get notified about the beginning and end of the execution process by pop-up messages.
- v) After the execution gets completed a green tick mark will be displayed. The input data with a green mark is ready to preview.



vi) Open 'Data Preview' by clicking the input component to view the preview of the extracted data.



🕞 Input 🖍			@ ī	<b>^</b> ×
CONFIGURATION	DATA PREVIEW			
dob	age	sal	joiningdateandtime	delta_status
1994-05-05	23	3000.92	2017-05-31T15:23:12.000+0530	insert
1993-09-23	24	3900.92	2017-03-21T15:43:12.000+0530	insert
1994-09-23	23	3000.92	2016-04-21T17:43:12.000+0530	insert
1992-07-23	27	4900.92	2014-05-21T16:43:12.000+0530	insert
1980-09-23	40	2300.92	2017-02-21T23:13:12.000+0530	insert

## 4.6. Save and Execute

By using the 'Save and Execute' option = users can save and write a workflow in the metadata to create a datastore out of it.

- i) Click the 'Save' option.
- ii) A new window pops-up to redirect the user to save the workflow.
  - a. Enter a Workflow name
  - b. Enter Description (Optional)
  - c. Select or Add a Workspace
- iii) Click 'Save.'

Save Workflow	>
Workflow Name '	
Sales Report	
If you want, you can add a descrip	otion to explain what you changed.
Description	
Workspace '	
Workspace ' Sample Workspace	• <b>+</b>
Workspace <sup>-</sup> Sample Workspace	- +

## 4.7. Schedule a Workflow

Users can schedule a created workflow for data refresh.



- i) Create a workflow.
- ii) Save and run the workflow.
- iii) Click the 'Scheduler' icon.
- iv) Click a range of time.
- v) Fill in the required details for the selected time range. E.g. The below-given image displays scheduler configuration details for the 'Daily' option.
- vi) Click 'Schedule.'

← : ▶ ⊗ ₪ =, ⊙	
Sample Data Preparation Workflow	
Schedule Workflow	×
DAILY WEEKLY MONTHLY YEARLY	
Every 1 day(s)	
O Every Week Day	
Start time $_{12}$ $\checkmark$ : 00 $\checkmark$	
CANCEL SCH	EDULE

vii) The selected workflow will be scheduled.

#### 4.8. Job

Users can see the job status for the saved workflows.

- i) Navigate to the Data Preparation landing page.
- ii)  $Click \equiv$  icon from the workflow editor.
- iii) Click 'Job'
- iv) Users will be displayed the job details in a table.

	My Workspace	Jobs (78)			Q Search		Job
â	Job	∱ JOB↑	EXEC START TIME ↑	LAST UPDATED DATE $\downarrow$	STATUS 🛧	÷	DETAILS
The second se	Trash Scheduler	SALES REPORT	9/28/2017, 3:31:42 PM	9/28/2017. 3:31:46 PM	FINISHED	•	ExecutionId 135397376 Execution Mode Execute Elapsed Duration 00:00:04
		➡ SALES REPORT	9/28/2017, 2:48:42 PM	9/28/2017. 2:48:46 PM	FINISHED	0 0 0	
		▶ SN	9/13/2017, 4:01:21 PM	9/13/2017, 4:01:22 PM	FAILED	0 0 0 0 0 0	
		WORKFLOW1	8/25/2017. 5:24:53 PM	8/25/2017. 5:24:58 PM	FINISHED	0 0 0 0	
		WORKFLOW1	8/25/2017. 5:22:40 PM	8/25/2017. 5:22:47 PM	FINISHED	0 0 0 0 0	
			R/34/3047 7:38:50 DM	R /34 /3047 7-38-04 DM		0	
			< 1 2	3 4 >>>			

Note: The execution details will be displayed on the right-hand side of the 'Job' page. Users need to click on the 'STATUS' of a job using the list of the jobs.



## 4.9. Trash

The '**Trash**' folder is provided to store all the deleted workflows and workspaces. Users can restore the deleted workflows and workspaces using this folder.

- i) Click on the **'Trash'** option.
- ii) Users will be redirected to see all the deleted files and folders under the trash folder.
- iii) Click 'Restore' to restore the selected workflow/workspace.
- iv) Click 'Delete' to permanently delete the selected workflow/workspace.

Note: Other options provided on the workflow editor are as described below:

Icons	Name	Description
← or ■	Hide and Show Components	Hides or shows the components on the left-hand side.
8	Clear Workflow	Clears the current workflow from the workflow editor.
8	Save	Saves a workflow
	Navigator	Redirects Users to the following hyperlinks: 1. Workspace 2. Job 3. Trash 4. Scheduler

## 5. Transform

### 5.1. Constants

Users can give a corresponding valid constant value for each type of column.

- i) Navigate to the Workflow editor.
- ii) Connect the 'Constants' component to the configured input dataset and output component.





iii) Configure the required details for the 'Constants' component:

- a. Column Name: Select columns from input data
- b. Column Type: Set column type using the drop-down menu
- c. Constant: Set a constant value
- d. Remove: Click the 'Remove' icon to remove the added constant information.

•}•	Constants 🖌				0	Î	^	×
CO	NFIGURATION DATA PREVIE	W						
	Column Name* (Select columns from input data)	Column Type* (Set column type)	Constant (Set constant value)					
	Date	Date 🗸	[current date]	\$ ×				
	ADD NEW COLUMN							

- iv) Save the workflow.
- v) Run/Execute the workflow.



vi) The set constant value will be applied to the selected column in the output dataset.



- Output 🎤				0 i ^
CONFIGURATION	DATA PREVIEW			
Loca	tionId	ProductId	Quantity	Date
25		13	7536	2017-09-28
30		17	6786	2017-09-28
58		5	9315	2017-09-28
26		2	2157	2017-09-28
40		10	6000	2017-09-28

## 5.2. Data Type

Users can change the data type of the selected columns by using the 'Date Type.'

- i) Navigate to the Workflow editor.
- ii) Connect the 'Data Type' component to the configured input dataset and output component.



iii) Select the columns and change the column data type using the drop-down menu.

- a. Column Name: Select columns from input data
- b. Data Type: Change column data type
- c. Date Format: Select source date format

E.g. In this case, the column data type has been changed from 'Date & Time' to 'Date.'



	Data Type 🖌			0	Î	^	×
со	NFIGURATION DATA PREVI	EW					
	Column Name* (Select columns from input data)	Data Type (Change column data type)	Date Format (Select source date format)				
	SalesId [Whole Number] -	Text 👻		×			
	Date [Date & Time]	Date 🗸		×			
	ADD COLUMN REMOVE ALL	COLUMNS					

- Save the workflow. iv)
- Run/Execute the workflow. V)
- vi)
- Compare the data previews of the input and output datasets. a. Input Data Preview: The 'Date' column is displayed in the original data type.

🕞 Input 🗸				<b>@ î ∧</b> ×
CONFIGURATION	DATA PREVIEW			
SalesId	LocationId	Productid	Quantity	Date
1535978	25	13	7536	2017-09- 14T17:47:04.000*0530
1535979	30	17	6786	2017-09- 14T17:47:04.000*0530
1535980	58	5	9315	2017-09- 14T17:47:04.000*0530
1535981	26	2	2157	2017-09- 14T17:47:04.000+0530
1535982	40	10	6000	2017-09- 14T17:54:04.000*0530
1535983	40	9	6000	2017-09- 14T17:47:04.000*0530

b. Output Data Preview: The data type of the 'Date' column will be changed.

⊖ Output 🗸				<b>9</b> î ^
CONFIGURATION	DATA PREVIEW			
Salesid	LocationId	Productid	Quantity	Date
1535978	25	13	7536	2017-09-28
1535979	30	17	6786	2017-09-28
1535980	58	5	9315	2017-09-28
1535981	26	2	2157	2017-09-28
1535982	40	10	6000	2017-09-28
1535983	40	9	6000	2017-09-28



### 5.3. Date Operations

Users can perform various operations of dates addition/subtraction with integers or other dates. It also allows extraction of parts of dates like day-part, month part, etc.

- i) Navigate to the Workflow editor.
- ii) Connect the 'Date Operations' component to the configured input dataset and output component.



- iii) Configure the 'Date Operations' component as described below:
  - a. Column Name: Enter the New Column Name
  - b. Operations: Select one operation using the drop-down menu.
  - c. Column/Value: Select a column or value for operations.
    - i. By selecting 'column' option, the column drop-down menu will be displayed.
    - ii. By selecting the 'value' option, users will be redirected to enter a value.

t	Date Operation	ns 🖊							•	2	Î	^	×
co	ONFIGURATION	DAT	A PREVIEW										
	Column Name" (New column name)		Operation* (Select date operation)	Colum (Select o	n / Vi colum	'alue" m/value for operation)							
	Next Date		Add months to d 👻	Colu	•	Date [Date & Time]	▪ ⊕ Colu	•	SalesId [ Whole	Numb	er] 🔻	×	
	ADD NEW COLUM	ИN											

- iv) Save the workflow.
- v) Run/Execute the workflow.
- vi) The new column, 'Next Date' will be added in the output dataset. Users can view it in the output data preview.



🔁 Output 🖍				0 î	^
CONFIGURATION DA	TA PREVIEW				
LocationId	ProductId	Quantity	Date	Next Date	
25	13	7536	2017-09- 14T17:47:04.000*0530	□015-11-14	
30	17	6786	2017-09- 14T17:47:04.000*0530	□015-12-14	
58	5	9315	2017-09- 14T17:47:04.000*0530	□016-01-14	
26	2	2157	2017-09- 14T17:47:04.000*0530	□016-02-14	
40	10	6000	2017-09- 14T17:54:04.000*0530	□016-03-14	
40	9	6000	2017-09- 14T17:47:04.000*0530	□016-04-14	

## 5.4. Filter

Users can filter the input dataset by specifying conditional expressions using the '**Filter**' transform. Multiple filter conditions can be imposed in the same transform. The following table lists the map of data types and permissible filter conditions.

- i) Navigate to the Workflow editor.
- ii) Connect the 'Filter' component to the configured input dataset and output component.



iii) Configure the 'Filter' Component as described below:

- a. Column Name: Select a column from the drop-down menu
- b. Operation: Select an operation from the drop-down menu
- c. Type: Select one option out of 'Column' or 'Value.'
- d. Compare: Enter/Select a value to compare with



Filter 🖌				0	Î	^
NFIGURATION DATA	A PREVIEW					
Column Name* (Choose column)	Operation* (Select operation)	Type* (Select column/value type)	Compare* (Enter/Select value to compare)			
Locational Number]	Equals	▼ Value ▼	25			×

- iv) Save the workflow.
- v) Run the workflow.
- vi) The output data will be filtered as per the applied conditions.

Output 🖍				0 î ^
CONFIGURATION DATA PREV	VIEW			
SalesId	LocationId	ProductId	Quantity	Date
1535978	25	13	7536	2017-09-14T17:47:04.000*0530
1536072	25	3	9687	2017-09-14T17:48:42.000+0530

#### 5.5. Formula Fields

Users can perform most common arithmetic operations (add, subtract, multiply and divide) on constants and columns.

- i) Navigate to the Workflow editor.
- ii) Connect the 'Formula Fields' to the configured input dataset and output component.





- iii) Configure the 'Formula' component as described below:
  - a. Column Name: Enter a name for the formula column
  - b. Calculation Type: Select a calculation type using the drop-down menu
  - c. Select Columns for Calculation: Select columns to be used in the calculation. Users can choose either a column or enter a value to complete the calculation process. E.g. In this case, the value option is chosen.

fx	Formula Fields 🖌				0	Î
C	ONFIGURATION DATA	PREVIEW				
	Column Name* (Name of new formula column)	Calculation Type' (Set column type)		Select Columns for Calculation (Select columns to use in the calculation)		
	Formula col	Addition	•	Colu Salesid (Whole Number) - Value - 25		×
	ADD NEW COLUMN					

- iv) Save the workflow.
- v) Run the workflow.
- vi) The calculated column will be added in the output dataset.

CONFIGURATION	ATA PREVIEW				0 î
Salesid	LocationId	Productid	Quantity	Date	Formula column
1535978	25	13	7536	2017-09-14T17:47:04.000+0530	1536003
1535979	30	17	6786	2017-09-14T17:47:04.000+0530	1536009
1535980	58	5	9315	2017-09-14T17:47:04.000+0530	1536038
1535981	26	2	2157	2017-09-14T17:47:04.000+0530	1536007
1535982	40	10	6000	2017-09-14T17:54:04.000+0530	1536022
1535983	40	9	6000	2017-09-14T17:47:04.000+0530	1536023
1535984	52	5	7346	2017-09-14T17:47:04.000+0530	1536036

#### 5.6. Group By

The '**Group By**' feature allows multiple aggregations on the same or different columns. Users can obtain multiple aggregations in the same transform. The aggregated values are added to a new column.

- i) Navigate to the Workflow editor.
- ii) Connect the 'Group By' component to the configured input dataset and output component.





- iii) Configure the 'Group By' component as described below:
  - a. Column Name: Choose a column from the drop-down menu.
  - b. New Column: Enter a title for the aggregate column.
  - c. Column Aggregate: Select a column from the drop-down menu to apply aggregation.
  - d. Aggregate Type: Select an aggregation operation from the drop-down menu.

Group By 🖍				01	^
CONFIGURATION	DATA PREVIEW				
Column Name* (Choose column)	Field Name				×
	LocationId				
New Column* (Aggregate column)		Column Aggregate (Select column to agg	regate)	Aggregate Type* (Select aggregate operation)	
Max		ProductId	[Whole Number]	Maximum 🗸	×
ADD NEW COLUM	Ν				

- iv) Save the workflow.
- v) Run the workflow.
- vi) The aggregated column will be displayed in the output data preview.



-	→ Output 🖍	
	CONFIGURATION DATA PREVIEW	
	LocationId	Max
	46	21
	18	8
	38	13
	58	5
	77	21

Note: The supported data types and aggregate operations are displayed in the following table:

Data Type	Aggregate
Text	Count
	Count Including NULLs
	Count Distinct Values
	First Non-Null Value
	Last Non-Null Value
	First Value
	Last Value
	Combine Strings Separated by Comma
Date	Minimum
Date Time	Maximum
	Count
	Count Including Nulls
	Count Distinct Values
	First Non-Null Value
	Last Non-Null Value
	First Value
	Last Value
Whole Number	Sum
Decimal	Average
Decimal (Fixed)	Minimum
	Maximum
	Standard Deviation
	Count
	Count Including NULLs

## 5.7. Mapping

Users should be able to select, remove or rename columns in the input dataset to fit the structure of the sink.

- i) Navigate to the Workflow editor
- ii) Connect the 'Mapping' component to the configured input dataset and output component





- iii) Configure the '**Mapping**' component:
  - a. Column Name: Select a Column from the input data using the drop-down menu
  - b. Rename: Rename the selected column of the input data
  - c. ADD Column: Click this option to add one more column from the input dataset
  - d. ADD ALL COLUMNS: Click this option to map all the columns from the input dataset
  - e. REMOVE ALL COLUMNS: Click this option to remove all the added columns for mapping

🗄 Mapping 🖍			
CONFIGURATION	TA PREVIEW		
Column Name* (Select columns from inpu	t data)	Rename (Set new name)	
LocationId	[Whole Number]	Location Name	×
ADD COLUMN AD	DALL COLUMNS REMOV	E ALL COLUMNS	

- iv) Save the workflow
- v) Run the workflow
- vi) The aggregated column will be displayed in the output data preview





#### vii) The aggregated column will be displayed in the output data preview

Output 🖍	
CONFIGURATION	DATA PREVIEW
Location Name	
25	
30	
58	
26	
40	

## 5.8. Replace Text

Users can search by whole word, sensitive to case, search for special values like NULL or empty strings, or use regular expressions, and then replace with any given constant values or even empty strings. Only text columns can be transformed using this component. Users can replace text for the multiple text columns.

- i) Navigate to the Workflow editor.
- ii) Connect the '**Replace Text**' component with the configured Input dataset and Output component.





#### iii) Run the workflow to preview the input data.



- iv) Configure the 'Replace Text' component as described below:
  - a. Column Name: Select a column from the input data set.
  - b. Search for: Enter a term from the selected column to search for.
  - c. Replace with: Enter a term to replace the searched term in the input data.

000	ိေ့ာ Replace Text 🖌							
СС	CONFIGURATION DATA PREVIEW							
_	Column Name" (Select columns to Search)	Search For* (Enter a term to search for)	Replace With (Replace found term with)					
-	LocationCode [Text] -	A	¢	<b>☆</b> ×				
	ADD NEW COLUMN							

- v) Run the workflow.
- vi) Save the workflow.
- vii) Open the Output data preview to see the replacement of the selected text in the column.

Replace Text 🖍						
CONFIGURATION DATA PREVIEW						
LocationId	LocationCode	City	State	Country		
1	~L	Montgomery	Alabama	USA		
2	~К	Juneau	Alaska	USA		
3	~Z	Phoenix	Arizona	USA		
4	~R	Little Rock	Arkansas	USA		
5	C~	Sacramento	California	USA		

Note:



- Users can click on the 'ADD NEW COLUMN' option to configure the multiple columns for any transform component.
- b. Users can also see data preview of the various transform components.

## 6. Merge

Users can use the 'Merge' components to combine input data sets and get the required output.

### 6.1. Append

The '**Append**' feature combines one dataset on top of another. If the datasets are of different structures, still the union is possible, and the output will be a unified more massive structure with NULL values populated wherever data is missing. Users can choose whether to include only shared columns or all columns to append.

## 6.1.1. Append All Columns

- i) Navigate to the Workflow editor.
- ii) Configure two input datasets.
- iii) Connect the 'Append' component with the configured Input datasets and an Output component.



iv) Select 'Include All Columns' option using the 'Select Columns' drop-down menu.

٦	Append 🖌		0	Î	^	×
СС	DATA PR	REVIEW				
	Select columns* (Which columns should be included in this append?)	Review changes (Review the changes that will made to each DataSet)				
	INCLUDE ALL COL 🗸	[Review description]				

- v) Save the workflow.
- vi) Run the workflow.





vii) The entire data of both the input data sets will be appended in the output data preview.

- Output 🖍				0	Î	~	×
CONFIGURATION	DATA PREVIEW						
empno1	bonousi	dojı	dob1			salı	
1	23.43453	2016-11-11T23:59:59.000+0530	1992-08-23			3490.65	
2	25.45457	2017-12-12T22:59:59.000+0530	1993-09-22			3596.66	
3	22.42457	2014-11-13T23:59:59:000+0530	1992-03-25			3495.67	
1	23.43453	2016-11-11T23:59:59.000+0530	1992-08-23			3490.65	
2	25.45457	2017-12-12T22:59:59.000+0530	1993-09-22			3596.66	
3	22.42457	2014-11-13T23:59:59.000+0530	1992-03-25			3495.67	

#### **Append Only Shared Columns**

- i) Connect the 'Append' component to the configures input datasets and an output component.
- ii) Choose 'ONLY INCLUDE SHARED COLUMNS' as an option to append the datasets.
- iii) The entire data of both the input data sets will be appended in the output data preview.

٥	Append 🖌		0	Î	^	×
co	DATA PREVIEW					
	Select columns* (Which columns should be included in this append?)	Review changes (Review the changes that will made to each DataSet)				
	ONLY INCLUDE SHARED COL •	[Review description]				



iv) Save the Workflow.

Save WORKILOW		×
Workflow Name '		<b>A</b> = -
Append only shared columns		
If you want, you can add a de	escription to explain what you chang	ed. ut
Workspace '		
Append		

v) Run the Workflow.



- vi) The shared column(s) will be appended in the output data set.
- E.g. The following images illustrate that the shared column '**Location**' has been displayed under the data preview of Append and Output components.
- a. Input Dataset-1



🕞 Input 🗸	*		0 î v
CONFIGURAT	DATA PREVIEW		
LocationId	LocationCode	City	State
1	AL	Montgomery	Alabama
2	AK	Juneau	Alaska
3	AZ	Phoenix	Arizona
4	AR	Little Rock	Arkansas
5	CA	Sacramento	California

## b. Input Dataset-2

🕞 Input 🖍			0 🖬 🗸
CONFIGURATION	DATA PREVIEW		
SalesId	LocationId	Productid	Quantity
1535978	25	13	7536
1535979	30	17	6786
1535980	58	5	9315
1535981	26	2	2157
1535982	40	10	6000

## c. Append Data Preview

+ Append /	
CONFIGURATION	DATA PREVIEW
LocationId	
1	
2	
3	
4	
5	
6	

d. Output Data Preview



CONFIGURATION	DATA PREVIEW
LocationId	
40	
52	
41	
7	
48	

## 6.2. Join

Users can join two datasets and use the merged output to write the workflow in the selected metadata.

i) Drag two input datasets and configure them to see the dataset preview.

## Input Data Set 1

🕞 Input 1 🖍					0 i ^ ×
CONFIGURATION	PREVIEW				
empno	name	dob	age	sal	joiningdateandtime
1	David	1994-05-05	23	3000.92	2017-05-31T15:23:12.000+0530
2	Louie	1993-09-23	24	3900.92	2017-03-21T15:43:12.000+0530
3	Jake	1994-09-23	23	3000.92	2016-04-21T17:43:12.000+0530
4	Harvey	1992-07-23	27	4900.92	2014-05-21T16:43:12.000+0530
5	Matthew	1980-09-23	40	2300.92	2017-02-21T23:13:12.000+0530

## Input Data Set 2

🕞 Input 2 🖊				0 î ^
CONFIGURATION DATA F	PREVIEW			
Salesid	LocationId	Productid	Quantity	Date
1535978	25	13	7536	2017-09-14T17:47:04.000*0530
1535979	30	17	6786	2017-09-14T17:47:04.000+0530
1535980	58	5	9315	2017-09-14T17:47:04.000+0530
1535981	26	2	2157	2017-09-14T17:47:04.000*0530
1535982	40	10	6000	2017-09-14T17:54:04.000+0530
1535983	40	9	6000	2017-09-14T17:47:04.000+0530



- ii) Connect the 'Join' component with the above-given input datasets and one output component to complete the workflow.
- iii) Configure the 'Join' component as described below:
  - a. Identify Column: Identify a column from the input dataset 1
  - b. Join Type: Choose a join type to merge the selected datasets out of the given choices
    - i. Inner
    - ii. Left Outer
    - iii. Right Outer
    - iv. Full Outer
  - c. Matching Column: Select a column from the input dataset 2

Join 🖍	0	Î	^	×
CONFIGURATION DATA PREVIEW				
Identifying Column* Join Type* Matching Column*				
productApplication Inner - test_db1				
Identifying Column ' 👻 Matching column ' 👻 🗙				
ADD COLUMN SWAP SOURCE				

Note:

- a. By default, the 'Inner' join type will be selected. Users can apply multiple inner joins by using the 'ADD COLUMN' tab.
- b. Click 'SWAP SOURCE' to interchange the input datasets and the selected columns from the data sets.
- iv) Save the workflow.

Save Workflow		
Worlflow Name *		
Join Inner		
f you want, you can add a descrip	ption to explain what you chang	ed.
f you want, you can add a descrip Description	otion to explain what you chang	ed.
f you want, you can add a descrip Description New Workspace	otion to explain what you chang	ed.
f you want, you can add a descrip Description New Workspace Join	otion to explain what you chang	ed.
If you want, you can add a descrip Description New Workspace Join	otion to explain what you chang	ed.



#### v) Run the workflow.



#### vi) Click the 'Data Preview' tab from the Join component to view data preview of the merged data.

Join 🖍			0 î	• >
CONFIGURATION	DATA PREVIEW			
SalesId	LocationId	CategoryId	Date	Amount
148	1	1	2016-05- 27T00:00:00.000+0530	2331
463	1	1	2017-04- 07T00:00:00.000+0530	3226
471	1	2	2016-01- 04T00:00:00.000+0530	1409
496	1	2	2016-01- 29T00:00:00.000+0530	1239
833	1	2	2016-12- 31T00:00:00.000+0530	4728
65	1	1	2016-03-	3481

## vii) Users can preview data under the 'Data Preview' tab of the selected output component.

🕂 Output 🖊			0 î	<b>^</b> >
CONFIGURATION DATA PREV	VIEW			
SalesId	LocationId	Categoryld	Date	Amount
243	1	i.	2016-08- 30T00:00:00.000+0530	1280
392	1	1	2017-01- 26T00:00:00.000+0530	5115
540	1	2	2016-03- 13T00:00:00.000+0530	2027
623	1	2	2016-06- 04T00:00:00.000+0530	5491
737	1	2	2016-09- 26T00:00:00.000+0530	5144

## 6.2.1. Join Types:

The 'Join' feature offers four types of join to merge datasets.

The sample data sets used to describe the supported join types are:



#### 1. Input Dataset 1

🕞 Input 1 🖊		
CONFIGURATION DATA PREVIEW		
empno	name	age
1	David	23
2	Louie	24
3	Jake	23
4	Harvey	27
5	Matthew	40

#### 2. Input Dataset 2

☐ Input 2 /		
CONFIGURATION DATA PREVIEW		
Salesid	LocationId	Productid
1535978	25	13
1535979	30	17
1535980	58	5
1535981	26	2
1535982	40	10
1535983	40	9

#### a) Inner Join

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Specify a join type from the 'Configuration' tab of the join component.

Inner Join 🖍	0	Î	^
CONFIGURATION DATA PREVIEW			
Identifying Column" Join Type" Matching Column"			
Mapping Inner - Mapping			
empno [Whole Number] - Location	NdWhole Number]	×	
ADD COLUMN SWAP SOURCE			

iii. Save and run the workflow.



#### iv. Click the 'Data Preview' tab using the join component to view the merged datasets.

Inner Join 🖍					0	Î	^
CONFIGURATION	DATA PREVIEW						
empno	name	age	SalesId	LocationId		ProductId	
3	Jake	23	1536027	3		18	
3	Jake	23	1536059	3		1	
5	Matthew	40	1536041	5		15	

#### b) Left Outer Join

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Specify a join type from the 'Configuration' tab of the join component.

Left Outer Join 🖌				0	Ī
CONFIGURATION DATA PREVIEW					
Identifying Column* Join Typ	e* Matching Colu	ımn*			
Mapping	Left Outer 🗸	Mapping			
empno [Whole Number] 🔻		LocationId	[Whole Number]	×	
ADD COLUMN SWAP SOURCE					

- iii. Save and run the workflow.
- iv. Click the 'Data Preview' tab using the join component to view the merged datasets.

Left Outer Join	/				0	<b>i</b> ~	
CONFIGURATION	DATA PREVIEW						
empno	name	age	SalesId	LocationId	I	ProductId	
3	Jake	23	1536027	3		18	
3	Jake	23	1536059	3		1	
1	David	23					
2	Louie	24					
4	Harvey	27					
5	Matthew	40	1536041	5		15	

Note: The output data preview will be aligned with the selected left input dataset.



#### c) Right Outer Join

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Specify a join type from the 'Configuration' tab of the join component.

Right Outer Join 💉	G	
CONFIGURATION DATA PREVIEW		
Identifying Column" Join Type" Matching Colu	umn'	
Mapping Right Outer -	Mapping	
empno [Whole Number] -	LocationId [Whole Number]	<
ADD COLUMN SWAP SOURCE		

- iii. Save and run the workflow.
- iv. Click the 'Data Preview' tab using the join component to view the merged datasets.

Right Outer Jo	in 🌶				0	Î	*
CONFIGURATION	DATA PREVIEW						
empno	name	age	SalesId	LocationId		Productl	d
			1535979	30		17	
			1535982	40		10	
			1535983	40		9	
			1535986	7		15	
			1535981	26		2	
			1535985	41		17	

#### d) Full Outer

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Specify a join type from the 'Configuration' tab of the join component.

Full Outer Join 🖌				0	Î
CONFIGURATION DATA PREVIEW					
Identifying Column* Join Type*	Matching C	olumn*			
Mapping	Full Outer 🗸	Mapping			
empno [Whole Number] -		LocationId	[Whole Number]	×	
ADD COLUMN SWAP SOURCE					

- iii. Save and run the workflow.
- iv. Click the 'Data Preview' tab using the join component to view the merged datasets.



Full Outer Join	/				0	۱ ×
CONFIGURATION	DATA PREVIEW					
empno	name	age	SalesId	LocationId		ProductId
			1536043	6		1
			1536077	6		3
			1535998	39		9
			1536036	39		8
3	Jake	23	1536027	3		18
3	Jake	23	1536059	3		1

## 7. Scheduler

The 'Scheduler' section displays the schedule monitoring details. Users can see a list containing all the scheduled workflows.

- i) Click the 'Navigator' icon 🔚
- ii) Select 'Scheduler' from the drop-down menu.
- iii) Users will be redirected to the 'Schedule Monitoring' page.
- iv) The scheduled workflow will be added to the list of all the schedules.
- v) Click on a scheduled workflow will display the following schedule details:
  - a. Scheduler Name
  - b. Last Updated Date
  - c. Recurrence date and time
  - d. Status

Decision Platform						III 🔢 ?	2
Data preparation 1.0.0						NEW	
My Workspace	Schedule Monitorin	g					
dol 🕄	Search Schedule	C	Scheduler Name	Last Updated Date 🗸 🗸	Recurrence	Status	Â
Trash	nadeem hierarchy test	Û	Data Type Test	10/11/2017, 10:25:00 AM	10/12/2017, 4:55:00 AM	Successfully started the scheduled query	
Scheduler	Sample Data Preparati	Ċ	Data Type Test	10/11/2017, 4:55:00 AM	10/12/2017, 4:55:00 AM	Successfully started the scheduled query	
	<ul> <li>Data Type Test</li> </ul>	Ċ	Data Type Test	10/11/2017, 4:55:00 AM	10/12/2017, 4:55:00 AM	Successfully started the scheduled query	
	<ul> <li>manjhari-bistorry</li> </ul>	Ċ	Data Type Test	10/10/2017, 10:25:00 AM	10/11/2017, 4:55:00 AM	Successfully started the scheduled query	
	elsticchexk_manjhari	Ċ	Data Type Test	10/10/2017, 4:55:00 AM	10/11/2017, 4:55:00 AM	Successfully started the scheduled query	J
	• mj-simple	Ċ	Data Type Test	10/10/2017, 4:55:00 AM	10/11/2017, 4:55:00 AM	Successfully started the scheduled query	
	elastic 15.9	Ċ	Data Type Test	10/9/2017, 10:25:00 AM	10/10/2017, 4:55:00 AM	Successfully started the scheduled query	
		Ť			10/10/2017. 4:55:00	Successfully started the scheduled	Ŧ



## 7.1. Schedule Configuration Options

These options are provided to configure a range of time for a scheduled workflow. The user can select only one option at a time from the given menu.

- **1.** Daily: User can schedule the job on a daily basis by using this option.
  - a. Click the 'Scheduler' icon on the workflow editor
  - b. Choose 'Daily' option from the 'Schedule Workflow' window (It is a default option).
    - i. Select an option out of the given choices
      - 1. Every \_\_\_\_ day(s)
      - 2. Every Week Day
      - 3. Set the start time using the drop-down
  - c. Click 'SCHEDULE'

← : ▶ ⊗ 🖬 ≒ 🕓		
Inner Join		
Schedule Workflow		×
DAILY WEEKLY MONTHLY YEARLY		
Every 1 day(s)		
O Every Week Day		
Start time $12$ $\checkmark$ : 00 $\checkmark$		
	CANCEL	SCHEDULE

- **2.** Weekly: User can schedule the job on a weekly basis by using this option.
  - a. Click the 'Scheduler' icon on the workflow editor
  - b. Choose the 'Daily' option from the 'Schedule Workflow' window.
    - i. Select an option out of the given choices
      - 1. Choose the days of the week by check marking in the box
      - 2. Set the start time using the drop-down
  - c. Click 'SCHEDULE'



Inner Join
Schedule Workflow X
DAILY WEEKLY MONTHLY YEARLY
Monday Tuesday Wednesday Thursday Friday Saturday Sunday
CANCEL SCHEDULE

- **3.** Monthly: User can schedule the job on the Monthly basis by using this option.
  - a. Click the 'Scheduler' icon on the workflow editor
  - b. Choose the 'Daily' option from the 'Schedule Workflow' window.
    - i. Select an option out of the given choices to choose a day for each month.
      - ii. Set the start time using the drop-down
  - c. Click 'SCHEDULE'

← : ▶ ⊗ 🖬 =, 🕓	
Inner Join	
Schedule Workflow ×	-
DAILY WEEKLY MONTHLY YEARLY	-
Day 1 of every 1 month(s)	
O The First ▼ Monday ▼ of every 1 month(s)	
Start time 12 • : 00 •	
CANCEL SCHEDULI	

- **4.** Yearly: User can schedule the job on a yearly basis by using this option.
  - a. Click the 'Scheduler' icon on the workflow editor
  - b. Choose the 'Daily' option from the 'Schedule Workflow' window.
    - i. Select an option out of the given choices
      - 1. Specify either a day or date of a specific month in a year
      - 2. Set the start time using the drop-down



c. Click 'SCHEDULE'



## 8. Signing Out

Follow the below given steps to sign out from the BizViz Platform:

- i) Click the 'User' icon 🛃 on the Platform home page.
- ii) A menu appears with the logged in user details.
- iii) Click 'Sign Out.'
- iv) Users will be successfully logged out from the **BizViz Platform**.

Note: Clicking on 'Sign Out' will redirect the user back to the login page of the BizViz platform.