

# **User Guide**

# **ETL R-4.5**



# Contents

1.	Abo	out thi	is Guide	4
	1.1.	Doc	ument History	4
	1.2.	Ove	rview	4
	1.3.	Targ	get Audience	4
2.	Intr	oduct	tion	4
	2.1.	Intro	oducing the BDB ETL	4
	2.2.	Sup	ported Web Browsers	4
3.	Get	ting S	itarted with the BDB ETL	5
	3.1.	Acce	essing the ETL Plugin	5
	3.1.	1.	Forgot Password Option	7
	3.1.	2.	Force Login	9
4.	Bas	ic Fea	itures	10
	4.1.	Wor	rkflow Editor	10
	4.2.	Extr	acting Data: Full and Incremental	11
	4.2.	1.	Create Data Set	17
	4.3.	Load	ding Data	18
	4.3.	1.	Configuring Elastic	19
	4.3.	2.	Configuring RDBMS	20
	4.3.	3.	Configuring Cassandra	21
	4.3.	4.	Configuring HDFS	22
	4.4.	Savi	ng a Workflow	23
	4.5.	Save	e & Run Preview	24
	4.6.	Save	e As	25
	4.7.	Save	e and Execute	25
	4.8.	Sche	edule a Workflow	26
	4.9.	Job.		27
	4.10.	Ті	rash	28
5.	Trai	nsforr	n	31
	5.1.	Con	stants	31
	5.2.	Data	а Туре	34
	5.2.	1.	Inferring Date & Date Time Formats	36
	5.3.	Date	e Operations	38
	5.4.	Filte	r	40



5.5.	. Formula Fields	41
5.6.	Group By	43
5.7.	. Mapping	45
5.8.	. Replace Text	47
5.9.	. Data Preparation	48
6. N	Aerge	51
6.1.	. Append	51
6	5.1.1. Append All Columns	52
6	5.1.2. Append Only Shared Columns	54
6.2.	Join	56
6	5.2.1. Join Types	59
7. S	cheduler	63
7.1.	. Schedule Configuration Options	64
8. S	igning Out	66



# **1. About this Guide**

### 1.1. Document History

Product Version	Release Date	Description
Data Preparation (ETL) 1.0	August 31 <sup>st</sup> , 2017	First Release of the document
Data Preparation (ETL) 1.1	December 11 <sup>th</sup> , 2017	Updated document
Data Preparation (ETL) 1.2	April 15 <sup>th</sup> , 2018	Updated document
ETL 3.8	December 1 <sup>st</sup> , 2018	Updated document
ETL 4.0	December 31 <sup>st</sup> , 2018	Updated document
ETL 4.2	March 25 <sup>th</sup> , 2019	Updated document
ETL 4.3	April 26 <sup>th</sup> , 2019	Updated document
ETL 4.4	June 7 <sup>th</sup> , 2019	Modified document
ETL 4.5	August 5 <sup>th</sup> , 2019	Updated document

Note: ETL 3.8 onwards the plugin Product Version column reflects the BDB Platform release version.

### 1.2. Overview

This guide covers:

- Introduction and steps to use various components of the BDB ETL plugin
- Configuration details of the ETL 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 BDB ETL

The BDB ETL is a self-service data preparation tool that empowers data-driven Business users with powerful capabilities to extract, transform, and load 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. Supported Web Browsers

The BDB Platform is a web browser-based application. The users can run the BDB Platform and its various plugins on the below given versions of the browsers:

Mozilla Firefox/ Firefox ESR	Latest Version
Microsoft Edge	Latest Version
Apple Safari	10
Google Chrome	Latest Version (recommended web browser)



# **3. Getting Started with the BDB ETL**

### 3.1. Accessing the ETL Plugin

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

- i) Open BDB Enterprise Platform Link: https://app.bdb.ai
- ii) Enter your credentials.
- iii) Select an Auth Type from the drop-down menu.
- iv) Click the 'Sign In' option.



- v) BDB Platform homepage opens.
- vi) The user gets redirected to the BDB Platform homepage.





### Note:

- a. The above screen opens only for those newly created users who have not yet created any document using the BDB Platform.
- b. If the user has created some documents previously, then the Platform homepage opens displaying the '**My Documents**' page by default.
- vii) Click the 'Apps' III icon.
- viii) All the available plugin applications get displayed.
- ix) Select the 'ETL' plugin.



- x) The ETL landing page opens.
- xi) The major ETL modules get displayed on the landing page:
  - a. Home (Default Component)
  - b. My Workspace
  - c. Job
  - d. Trash
  - e. Scheduler





This document aims to describe all the significant components and the related workflows at details.

### **3.1.1. Forgot Password Option**

Users are provided with a choice to change the password on the Login page of the platform.

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

Decision Platform
Email *
bdbuser@bdb.ai
Password *
Auth Type
Enterprise
Forgot Password
Sign In

Copyright © 2015-2019 BDB (BizViz Technologies Pvt Ltd)

- iii) A new window opens.
- iv) Provide the email id that is registered with BDB to send the reset password link.
- v) Click the '**Continue**' option.





# Forgot Password?

reset your password.	address to
Email * bdbuser@bdb.ai	
	Sign in
Continue	

Copyright © 2015-2019 BDB (BizViz Technologies Pvt Ltd)

vi) Users may be redirected to select a space in case of multiple spaces under one server link (They need to select a space and click the '**Continue**' option once again). If users do not have multiple spaces then, a message appears to notify the user about the password reset link (The users receive the reset link via their registered email.)



- vii) Click the password reset link from your registered email.
- viii) The user gets redirected to the 'Reset Password' page to set a new password.
- ix) Set a new password.
- x) Confirm the newly set password.
- xi) Click the 'Continue' option.



### Reset Password

Continue

You have confirmed ownership of the BDB
account. Please reset your password to get
access.
New Password *
••••••
Confirm New Password *
••••••

Copyright © 2015-2019 BDB (BizViz Technologies Pvt Ltd)



xii) The password for the selected BDB account gets reset, and the user receives a notification to assure them.

New password has been updated successfully

### 3.1.2. Force Login

The '**Force Login**' functionality has been introduced to control the number of active sessions up to three. The users can access only 3 sessions at a time when they try to access 4<sup>th</sup> session a warning message displays to inform that the user has consumed the permitted sessions and a click on the '**Force Login**' would kill all those active sessions.

- i) Navigate to the BDB Platform Login page.
- ii) Enter the valid credentials to log in.
- iii) Click the 'Sign In' option.



- iv) The user gets the following message if the user already consumes the permitted active sessions (3 sessions at a time).
- v) Click the 'Force Login' option.



← → C Secure	https://app.bdbizviz.com/home/#/max-users	아 Q ☆ 🖲 🗄
>	Welcome	
	to BDB Decision platform	
	Big Data Pipeline Framework	BBB®
	Dashboard Designer	Decision Platform
	ETL (Self-Service Data Preparation)	
	Geospatial Analysis (Location Intelligence)	Permitted sessions are already consumed. Do you want to force login? It will kill all the active
	Predictive and Prescriptive Workbench	sessions.
	Play (Beta Release)	Cancel Force Login
	Self-Service BI (Business Story)	
	Social Media Browser	
	Sentiment Analysis	
	Survey	
		Copyright © 2015-2019 BDB (BizViz Technologies Pvt Ltd)

- vi) A warning message appears that the currently active sessions get killed for the user and the user gets redirected to the SignIn page of the BDB Platform.
- Note: The user can successfully login to the BDB Platform after selecting the '**Force Login'** option to Sign In to the platform.

# **4. Basic Features**

The landing page of ETL launches workspace view. 'My Workspace' gets displayed by default.

### 4.1. Workflow Editor

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

- i) Navigate to the ETL landing page.
- ii) Click the '**New**' option.





- iii) The user gets redirected to the Workflow Editor.
- iv) The Workflow editor exposes the user to the following aspects to autonomously prepare data:
  - a. Data
  - b. Transform
  - c. Merge



# 4.2. Extracting Data: Full and Incremental

- i) Navigate to the Workflow Editor.
- ii) The 'Data' option gets selected by default.
- iii) Drag and drop the 'Input' component onto the workflow editor.



iv) Click on the dragged Input component to get the configuration tab.



<b>[</b>	Input				· · · - · · · · · · · · · · · · · · · ·						
					·						
	- <u>-</u>										
			$\frac{1}{2} - \frac{1}{2} - \frac{1}{2} - \frac{1}{2}$	 							
									;		()
input 💋								•		-	1
CONFIGURATION	DATA PREVIEW										
CONFIGURATION	DATA PREVIEW										
CONFIGURATION Select Source Type	DATA PREVIEW										
CONFIGURATION Select Source Type	DATA PREVIEW										
CONFIGURATION Select Source Type	DATA PREVIEW										

v) Select a data source type using the 'Select Source Type' drop-down menu (At present only MYSQL, MSSQL, Oracle, Google Sheet, and Excel Sheet are supported).

🖯 Input 🖉							
CONFIGURATION	DATA PREVIEW						
MYSQL							
MSSQL							
Oracle							
Google Sheet							
Excel Sheet							

E.g., In this case, the selected data source type is **MySQL**.

- vi) A list of Query Services based on the MySQL database appears.
- vii) Select a Query Service from the list.

### a. BASIC INFO

i. The '**BASIC INFO**' tab opens (by default) describing the basic information of the selected Database and Query Service.

🖯 Input 🧭		
CONFIGURATION DATA PREVI	EW	
MYSQL -	BASIC INFO	SETTINGS DATA PREPARATION
Q hiri 🛛	DataBase Name	BDB
· · · ·	QueryName	HiringData
hiring_feeddback_smiley	DataSource Name	BDB
hiring_feeddback_smiley	Query	select candidate_id,name,gender,source,referral_of,designation,tea from hiring_data;
HiringData 💟		
SAMPLE_3_HIRING DATA		



### b. SETTINGS

- i. Click the **'SETTINGS**' tab.
- ii. The 'Incremental Load' option appears.
- iii. Enable the Incremental Load to get the recently updated data.
- iv. The user needs to configure the following options after enabling the Incremental Load:
  - 1. **Primary Key** Select a primary key of the data source by using a checkmark in the given box.
  - 2. **Delta Load**-Select a column of type timestamp or date or long which is updated whenever a new row is inserted or updated in the data source. This column is used to load the **Incremented data.** Use the radio button to select a '**Delta Load**' column.

	BASIC INFO SETTINGS	DATA PREPARATION		
~	Incremental Load		<b>Q</b> Search Colum	IN
	Columns	Туре	Primary Key	Delta Load
	designation	Text		
	expected_joining_date	Date		0
	experience	Decimal		
	expyrsper ctc	Decimal		

Note: The users can choose not to enable the '**Incremental Load**' option. In this case, the following details get displayed, and the complete data gets extracted.

BASIC INFO SETTINGS	DATA PREPARATION	
Incremental Load		Q Search Column
Columns		Туре
designation		Text
expected_joining_date		Date
experience		Decimal
expyrsper ctc		Decimal

### c. DATA PREPARATION

- i. Click the 'DATA PREPARATION' tab.
- ii. Two options get displayed to proceed with the action of data preparation.
  - 1. Launch Data Prep



2. R	efresh Metad	ata
BASIC INFO	SETTINGS	DATA PREPARATION
Seat Launch Dat	a Prep 🗘 R	efresh Metadata
S No.	Actions taken	Affected Column(s)

### iii. A window opens prompting to save the workflow.

- 1. Provide a Workflow Name.
- 2. Describe the workflow (Optional).
- 3. Select a Workspace from the drop-down menu or create a new workspace.
- 4. Click the **'Save'** option.

e Save Workflow	×
- 1 Workflow Name *	
lf you want, you can add a descrip	tion to explain what you changed.
2 Description	
3 Select workspace *	
	Canc 4 Save

- iv. The workflow gets saved by the given name and the provided name of the Workflow displays on the workspace.
- v. The following icons appear after saving the workflow in the workflow editor taskbar:
  - 1. Save & Run Preview
  - 2. Save As
  - 3. Save and Execute
  - 4. Schedule

۲							.1		×	2	4			-	ł	≔	
	Sampl	e WF	0	 	 		 	 		 		 	 				· = -
				 	     		 	     ·		 		 	 	 			·

- vi. The 'Launch Data Prep' gets enabled only after saving the workflow.
- vii. Click the 'Launch Data Prep' option.



K : 🕨 😣	∎ 🛃 🗮	() Execu	ite processing finishe	d.		NEW		:=
🔒 Input 🖉		_				Î	<u>+</u>	<u>+</u> ↑
CONFIGURATION	DATA PREVIEW							
MYSQL	•	BASIC INFO	SETTINGS	DATA PREPARA	TION			
Q hiring	€	🕞 Launch Da	ata Prep 🗘 Ref	resh Metadata				
BDB_Hiring		S No.	Actions taken		Affected Column(s)			
BDB_Hiring_Data	<b>O</b>	No Preparatio	on steps available!					

viii. The Data Preparation plugin gets launched.

a Pr	eparation		Export Se	ttings Export Steps to ETI	L Export Steps to Pipeline 🕤 C
					candidate_id
	candidate_id	name 🗮	gender 📃	source	Profile Transforms Steps: 0
	Integer	sang	string	Sung	Chart Info Pattern
	1.	Emp ID 1	male	internal	Row Count
	2	Emp ID 2	female	internal	
	3	Emp ID 3	female	internal	200
	4	Emp ID 4	male	internal	
5	5	Emp ID 5	female	internal	150
6	6	Emp ID 6	female	agency	> <b></b>
	7	Emp ID 7	female	portal	100
8	8	Emp ID 8	female	portal	
9	9	Emp ID 9	male	portal	
10	10	Emp ID 10	female	portal	50

- ix. The user can perform the desired transform using the '**Transforms**' tab provided on the Data Preparation.
- x. Click the 'Export Steps to ETL' option.

a Pre	paration				E	xport Settings	Exp	ort Steps to ETL	Ехро	ort Steps to Pipeline 🍎 🕑 🕨 🕨 🗙
										cur_monthly_payment
:	usd_billing	double	cur_monthly_paym = double	bill_start_date	≡ date	created_datetime	, string	updated_datetime	tri 🔺	Profile Transforms Steps: 3
1	2200.0		48333.0	12-06-2016						Column Row
2	2600.0			12-06-2016						Find a Function
3	2200.0			12-06-2016						Fill empty cells with text
4	1000.0			12-06-2016						Use with:
5	2000.0			12-06-2016						Value
6	1500.0			12-06-2016						Value:
7	2000.0			12-06-2016						
8	3000.0			12-06-2016		N				> Submit
9	2200.0			12-06-2016		145				
10	1800.0			12-06-2016						Find Anomaly
11	1000.0			12-06-2016						Flag Duplicates In Columns
12	1000.0			12-06-2016						Flag Duplicates In Table
4									•	Remove Duplicates From Column
90/ 90							2 3	4 5 Nevt		Remove Duplicates From Table

xi. A success message appears to inform about the completion of the export action.



xii. Click the '**Refresh Metadata**' option using the 'DATA PREPARATION' tab.



xiii. All the performed transforms get listed at the bottom of the tab.

🗟 Input 💋				Î	<u>+</u>
CONFIGURATION	DATA PREVIEW				
MYSQL	-	BASIC INFO	SETTINGS DATA PREPAR	ATION	
Q, hiring	æ	🞇 Launch Da	ata Prep 🗘 Refresh Metadata	]	
BDB_Hiring		S No.	Actions taken	Affected Column(s)	
BDB_Hiring_Data	<b>S</b>	1	DELETE_COLUMN	["comments"]	1
bdb_hiringstory		2	FILL_EMPTY_WITH_DEFAULT	["bill_start_date"]	
finance details for hirin	g	3	DELETE_EMPTY_ROWS	["cur_monthly_payment"]	

#### Note:

- a. The 'BASIC INFO' tab varies based on the selected Input data source types:
  - i. The '**BASIC INFO**' tab opens the same information about database and query service for the MSSQL and Oracle data sources types as displayed for the MySQL data source type.
  - ii. The '**BASIC INFO**' tab displays the query information and the concerned email address for the Google Sheet data source type.

🔒 Input 🖉				<u>+</u>	+
CONFIGURATION DATA PREVI	EW				
Google Sheet 🗸	BASIC INFO	SETTINGS	DATA PREPARATION		
Q Search Query Services	QueryName	6Feb Hirir	ng Data dataset		
6Feb Hiring Data	Email	dataprepa	aration.bizviz@gmail.com		
7 March GS DS 3	marrie	BIZVIZ DA	IA PREPARATION		

iii. The 'BASIC INFO' tab for the Excel Sheet displays only query name.

🗧 Input 🕜				Î	<u>+</u>	+ +
CONFIGURATION DATA PREVI	EW					
Excel Sheet 🗸	BASIC INFO	SETTINGS	DATA PREPARATION			
QSearch Query Service	QueryName	dataset				
dataset 📀						
datatest						

- b. The 'SETTINGS' and 'DATA PREPARATION' tabs display the same set of information as described above for all the Input data source types.
- c. The user can prepare the input data using the Data Preparation module. Refer the Data Preparation UG for more details.



### 4.2.1. Create Data Set

The 'Create Data Set' functionality helps the user to create a new dataset based on a local or SFTP file.

The 'Create Dataset' icon appears only for the Excel Sheet input data source type.

- i) Select the 'Excel Sheet' as an input data source.
- ii) Click the 'Create Data set' 😌 icon.

🗐 Input 💋	
CONFIGURATION	DATA PREVIE
Excel Sheet	Create Data set
<b>Q</b> Search Query Servic	e

- iii) The Create Dataset window opens.
- iv) The user needs to configure other information based on the selected File source:

#### a. Local

- i. Select the 'Local' file source option using the radio button.
- ii. Provide a Dataset name.
- iii. Select a file using the 'Choose File' option.
- iv. Select a sheet using the 'Sheet List' drop-down menu (In case of the Excel sheets only).
- v. Click the 'CREATE' option.



Note: The 'Local' file source option can be used only if the file size is less than 4mb.

#### b. SFTP

- i. Select the 'Sftp' File source option.
- ii. Provide the Dataset name.
- iii. Select a Data Source using the drop-down (the user gets access to the data sources created using the Data Center under the drop-down).



- iv. Select a file using the 'Files' drop-down menu.
- v. Click the 'CREATE' option.

Create Dataset	×
File Source 🔿 Local Sftp	A .
2 Data Set Name *	
Sample Dataset	1
3 Data Source *	
EtlExcelConnector	-
4 Files *	
_ 1539191442_hiring_data_and_salary.xlsx	•
1	CREATE

v) A success message appears.



vi) The newly created dataset gets added to the available list of the datasets.

EVIEW		
▼ BASIC INFO	SETTINGS DATA PREPARAT	TION
• QueryName	Sample Dataset	
2		
	EVIEW BASIC INFO QueryName	EVIEW         BASIC INFO       SETTINGS       DATA PREPARAT         QueryName       Sample Dataset

### 4.3. Loading Data

The user can load the extracted data into Output Connectors 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.





- iii) Click on the 'Output' component to display the 'CONFIGURATION' option.
- iv) The following options get displayed:
  - a. Elastic
  - b. RDBMS
  - c. Cassandra
  - d. HDFS
- v) Select any one option to access the CONFIGURATION tab.

😝 Output 🖉				
CONFIGURATION DATA PREVIEW				
Connector	Elastic	•		
	Elastic			
Select resource	RDBMS	-		
	Cassandra			
Select Mappii	HDFS			

# **4.3.1. Configuring Elastic**

The following configuration fields open when the selected Output option is Elastic.

- i) Select a resource using the drop-down menu for the Elastic writer.
- ii) Enable 'Select Mapping ID' option.
- iii) After enabling the 'Select Mapping ID' option, the user gets redirected to select a mapping id from the 'Mapping id' drop-down menu.



	E Output	Î	<u> </u>	<u>↓</u> ↑
	CONFIGURATION DATA PREVIEW			
	Connector Elastic -			
•	Select resource Hiring_Data			
•	2 Select Mapping ID name			

Note:

- a. The 'Mapping Id' drop-down menu appears when the user enables the 'Select Mapping ID' option with a checkmark.
- b. Click the **'Create Meta Data**' 🙂 icon to open the configuration window for creating metadata.

### 4.3.2. Configuring RDBMS

The following configuration fields appear for the RDBMS output connector.

- i) Select a Data Source Type from the drop-down menu.
- ii) Select a Data Source Name from the drop-down menu.
- iii) Select a Database Name from the drop-down menu.
- iv) Select a Table Name from the drop-down menu or click the 'Add New Table'  $\stackrel{ullet}{\bullet}$  icon to Create a New Table.
- v) Choose a Table Operation from the given choices:
  - 1. Overwrite: By choosing Overwrite as operation, the existing records gets overwritten in the selected table.
  - 2. Append: By choosing Append as operation, the extracted records get added at the end of the existing records in the columns of the selected file or table.
  - 3. Upsert: By choosing Upsert as operation, only new records get added to the file or selected table.
- vi) Click the '**Apply**' option.

Output 🖉		
ONFIGURATION DATA PREVIEW		
Connector RDBMS -		
CONFIGURATION PROJECTION		
Select Data Source Type *	Select Data Source Name *	
MYSQL -	BDB_hiring	•
Select Database Name *	Select Table Name *	
BDB_Hiring_Data	Hiring_FromETL	
Choose Table Operation		
🔿 Overwrite 🔿 Append 🧿 Upsert		
Batch Size	No. of Dataset Partition	
	10	



Note: The user gets 'Batch Size' and 'No. of Dataset Partition' options only when 'Upsert' is selected as the table operation.

### 4.3.3. Configuring Cassandra

The following configuration fields appear for the Cassandra output connector.

- i) Select a Data Connector from the drop-down menu.
- ii) Host Name: Displays a preselected Host Name based on the selected data connector.
- iii) Port Number: Displays a preselected Port Number based on the selected data connector.
- iv) Username: Displays a preselected username based on the selected data connector.
- v) Password: Enter the Password
- vi) No. of Rows in Batch: Set number of Rows that you wish to allow in one Batch (the default value for this field is 1000)
- vii) Select Key Space: Select a Key Space from the drop-down menu
- viii) Replication Factor: Enter the Replication Factor
- ix) Select Columns: Select Columns using the drop-down menu
- x) Select Table: Select an existing table from the drop-down menu or choose the '**Create New Table**' option to create a new table.
- xi) Consistency: Select a Consistency option from the drop-down menu
- xii) New Table: Provide a title to the newly created table using the 'New Table' field
- xiii) New time uuid column name: Provide a name for the new Time UUID Column

😝 Output 🖉			Î	<u>+</u>	<u>+</u> ↑
CONFIGURATION DATA PREVIEW					
Connector Cassandra -					
Select Data Connector cassandra_4.5_etl	Host Name		Port Number		^
User Name cassandra	Password		No Of Rows In Batch		Ŀ
Select Key Space	Replication Factor		Select Columns Salary, gender, Dept, Age, Name	•	ł.
Select Table		Consistency			
New Table		New time uuid column name			•

- xiv) Headers: All the columns from the data set get listed.
- xv) **Partition Key**: The Partition Key determines which node stores the data. It is responsible for data distribution across the nodes.
  - a. The UUID Column name gets displayed under the 'Partition Key' window.
  - b. The user can select and move any column from 'Header' (Select Column) to 'Partition Key' space.
  - c. The sequence of the columns listed under Partition Key can be arranged by using '**Up**' or '**Down**' options.
- xvi) **Clustering Key**: The Clustering Key is a storage engine process that sorts data within the partition. It determines per-partition clustering.



- a. The items listed under the Clustering Key box can be arranged by using 'Up' or 'Down' options.
- b. Users can select any column from 'Headers' (Select Column) to 'Clustering Key' space.

Headers		Partition Key
Salary	>	eid
gender	<	
Dept		
Age		Clustering Key
Name	>	
	<	

### 4.3.4. Configuring HDFS

The following fields appear for the HDFS output connector.

- i) Provide file path
- ii) Select a File Format from the below given choices in the drop-down menu.
  - 1. Parquet
  - 2. Json
  - 3. Avro
  - 4. CSV
- iii) Select a Save Mode from the below given choices in the drop-down menu.
  - 1. Append
  - 2. Overwrite
  - 3. Error
  - 4. Ignore
- iv) Select a Compression Method from the below given options in the drop-down menu.
  - 1. Gzip
  - 2. Snappy
  - 3. None

PREVIEW	
S <del>-</del>	
File Format rt>/ <dir> parquet</dir>	÷
r	PREVIEW 5  File Format parquet

Note: The user should run or execute the workflow after configuring the Cassandra and HDFS output connectors.



# 4.4. Saving a Workflow

Users are provided with two options to save a workflow.

- i) Click the 'Save' 🗖 icon from the workflow editor header.
- ii) A new window appears to redirect the users to save the workflow.
  - a. Workflow Name: Provide a name for the Workflow (mandatory field)
  - b. Description: Enter Description for the Workflow (Optional)
  - c. Select Workspace: Select a workspace from the drop-down menu or Add a new Workspace by clicking the '**Add'**  $\stackrel{\bullet}{\bullet}$  icon.
- iii) Click the 'Save' option.

Save Workflow		×
Workflow Name *		
Sample WF		
lf you want, you can add	l a description to explain what you ch	langed.
Description		
Description Select workspace *		

iv) A success message appears to assure the action.



- v) The Workflow gets saved in the specific Workspace.
- vi) The Workspace can be opened by using the 'My Workspace' option.

ETL					New 😑	53
<u></u>	Home	Sample Works	space ( 12)		Q Search	
8	My Workspace	filter	Sample WF (old	Date Operation	Filter	
•	Job	ī	Î	Î		Ŧ
Ū	Trash	Formula Fields	Group By	Mapping Transf	Replace Text	
J	Scheduler	Î	Ť	i		î
		Append	Data Preparati	Input Data Pre	Sample WF	
		Û	Î	Û		Û

Note: The user can use the 'Save As'  $\stackrel{{}_{\scriptstyle \mbox{\footnotesize base}}}{=}$  icon if they wish to save it in another workspace.



### 4.5. Save & Run Preview

The 'Save & Run Preview' icon appears only for the saved workflows. The users can get the data preview only after running the workflow using this option.

- i) After saving a workflow, Users can access the 'Save & Run Preview' option on the workflow editor toolbar.
- ii) Click the 'Save & Run Preview' 🏲 option.
- iii) The ongoing execution process gets displayed through a continuous blue line.
- iv) The user gets notified about the beginning and end of the execution process by pop-up messages.
- v) After the execution gets completed, a green tick mark appears on the top of each component in the workflow. The input data with a green checkmark is ready to preview.
- vi) A pop-up message appears asking permission for showing the notifications.
- vii) Click the 'Allow' option to display the notification.

Show notifications		
	Allow	Block
 ponent	_	

viii) A notification appears to inform about the preview process status.



ix) Open the '**DATA PREVIEW**' tab by clicking the input or output component to view the preview of the extracted or loaded data.





Note: The users get notifications on the screen for success or failure of the preview processing.

### 4.6. Save As

By using the 'Save As' rightarrow icon, the user can save the workflow to a different Workspace.

- i) Click the 'Save As' 📥 icon.
- ii) Modify the workflow name.
- iii) Modify the workflow description.
- iv) Change the Workspace or create new workspace using the 'Add'  $\stackrel{\bullet}{\longrightarrow}$  icon.

ave Workflow				×		
Workflow Name *					ъ.	
Sample WF						
lf you want, you can add a de	escription to expla	ain what yo	ou chai	nged.	L	
lf you want, you can add a de Description	escription to expla	ain what yo	ou chai	nged.	L	
If you want, you can add a de Description	escription to expla	ain what yo	ou chai	nged.		

### 4.7. 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 and Execute'** = icon from the workflow editor taskbar.
- ii) A message appears to inform the user that the execution process has been started.



iii) The user gets another message to obtain confirmation about the completion of the execution process.



iv) Another notification appears at the bottom of the page displaying the status of the execution process.



### 4.8. Schedule a Workflow

Users can schedule a created workflow for data refresh. The Schedule option appears only for the saved workflows.

- i) Create a workflow.
- ii) Save and run the workflow.
- iii) Click the 'Schedule' 🕓 icon.
- iv) Click a range of time.
- v) Set the required information asked for the selected time range. E.g., The below-given image displays Schedule Workflow configuration details for the '**DAILY**' option.
- vi) Click the 'Schedule' option.



Schedule Workflow				×
DAILY WEEKLY	MONTHLY	YEARLY		
Every 1	day(s)			
Start : time 12 - 00	•			
			Cancel	Schedule

- vii) The selected workflow gets scheduled for the data refresh as per the configured information.
  - Note: The user can access the 'Scheduler' option from the ETL homepage or by clicking the  $\stackrel{i=}{=}$  icon from the workflow editor.



### 4.9. Job

Users can see the job status for the saved workflows.

i) Navigate to the ETL homepage.

or

Click the iii icon from the workflow editor.

ii) Select the 'Job' option from the menu list.



ETL	
<u></u>	Home
3	My Workspace
ô	Job
Ŵ	Trash
	Scheduler

iii) A table appears displaying a list of jobs.

ETL							New	i	£3
<u></u>	Home	🖻 Jobs (81) C			Q Search		Job		
6	My Workspace	NAME 🛧	EXEC START TIME ↑	LAST UPDATED DATE 🗸	STATUS ↑	Ŧ	DETAILS		
Ô	Job	SAMPLE WF	8/9/2019, 12:11:16 PM	8/9/2019, 12:11:17 PM	FINISHED				
Ŵ	Trash	SAMPLE WF	8/9/2019, 12:11:03 PM	8/9/2019, 12:11:04 PM	FINISHED				
<li></li>	Scheduler	SAMPLE WF	8/9/2019, 12:08:24 PM	8/9/2019, 12:08:24 PM	FAILED				
		SAMPLE WF	8/8/2019, 11:45:05 AM	8/8/2019, 11:45:06 AM	FINISHED				
		SAMPLE WF	8/8/2019, 11:42:47 AM	8/8/2019, 11:42:47 AM	FAILED				
					EAU ED				
		<	< 1 2 3	4 5 >>>					

- iv) Click on a job from the list.
- v) The '**DETAILS**' tab opens, displaying the execution details for the selected job on the right-hand side.

🖻 Jobs (81) C			Q Search		Job	
NAME 🛧	EXEC START TIME $\Lambda$	LAST UPDATED DATE $\downarrow$	STATUS ↑	Ŧ	DETAILS	
➡ SAMPLE WF	8/9/2019, 12:11:16 PM	8/9/2019, 12:11:17 PM	FINISHED		ExecutionId	405274813
SAMPLE WF	8/9/2019, 12:11:03 PM	8/9/2019, 12:11:04 PM	FINISHED		Execution Mode Elapsed Duration	00:00:01
SAMPLE WF	8/9/2019, 12:08:24 PM	8/9/2019, 12:08:24 PM	FAILED			

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

### 4.10. 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) Navigate to the 'My Workspace' page.
- ii) Select a workflow and click the 'Delete' icon.
- iii) A pop-up window opens to assure about the action of deletion. Click the **'OK'** option from the pop-up window.

	ETL					New	:=	0
	<u></u>	Home	← 🙆 Sample Work	space ( 12)		Q Search		
1	Ô	My Workspace	filter	Sample WE (old	Date Operation	Filter		
	6	Job	2			The		
	Ū	Trash 📄 Formula Fi	Formula Fields	Would you like to delete Workflow ?	Replace Text			
	(	Scheduler		Clicking "OK" will remove filter from Workflow	ws.			T
			Append	Cancel Ok	Input Data Pre	Sample WF		
				T T	1			II.

iv) A success message appears to assure the action. The Workflow gets deleted and moved to the Trash folder.



- v) Click on the '**Trash**' option.
- vi) The Trash folder displays the deleted workflows.
- vii) The 'DETAILS' tab displays information about the selected workflow.

	ETL													New	:=	()
	6 Home C					Q Search			Trash							
	٥	My Workspace		Test			filter		sample		Data Preparati		7	DETAILS		
	•	Job				-6 -2		-9 🖹		-9 î		-9 i		Name		filter
5	Û	Trash		💼 Join			Inner Join							Name		Inter
	(	Scheduler				ூ ∎		-9 🗎								

viii) Click the 'Restore' icon to restore the selected workflow/workspace.

ETL				
<u></u>	Home	← 🙆	Trash ( 6 )	
٥	My Workspace	Test		<b>filter</b>
•	Job		-9 î	<u>@</u> =
Ŵ	Trash	🖨 Join		filter Inner Join
J	Scheduler		-9 î	٩ 🗈



ix) Click 'delete permanently' <sup>i</sup> icon to permanently delete the selected workflow/workspace.

ETL			
<u></u>	Home	🤆 🙆 Trash ( 6 )	
	My Workspace	Test	filter
•	Job		
Ŵ	Trash	join	filter delete permanently
()	Scheduler	Ð 🖡	ூ ∎

Note:

- a. Users can check out all the essential features of the ETL module on a relevant input dataset.
- b. Other options provided on the workflow editor are as described below:

Icons	Name	Description
К	Hide and Show	Hides or shows the components on the left-
or or	Components	hand side.
×	Clear Workflow	Clears the current workflow from the
		workflow editor.
:=	Navigator	Redirects the users to the following
		hyperlinks:
		1. Workspace
		2. Job
		3. Trash
		4. Scheduler

- c. The user can also delete a connection or component from an ETL workflow by selecting the 'Delete' option as displayed below:
  - i. Deleting a Connection
    - 1. Use right-click on the connection to get the delete option.
    - 2. Click the '**Delete**' option.



- 3. A window opens asking the user to confirm the connection deletion.
- 4. Click the 'Continue' option.



Would you like to delete selec	ted connection?	
Click on "Continue" will remove the	selected component f	rom the editor.
	Cancel	Continue

- 5. The selected connection gets deleted.
- ii. Deleting a Component
  - 1. Use right-click on the component to get the delete option.
  - 2. Click the 'Delete' option.



- 3. A window opens asking the user to confirm the component deletion.
- 4. Click the 'Continue' option.

Would you like to delete selected component?						
Click on "Continue" will remove the selected component from the editor.						
	Cancel	Continue				

5. The selected component gets deleted from the workflow.

# 5. Transform

The Transformation components have both input ports and output ports when used in a workflow, and they apply specific transformations to the input data in the transformation stream.

### 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. Connect the output node of the dragged Constants component with the Output component and create a workflow.





- iii) Click on the 'Constants' component to open the 'CONFIGURATION' fields.
  - Configure the required details for the 'Constants' component:
    - a. Click the 'ADD NEW COLUMN' option
    - b. Column Name: Add new columns to the input dataset
    - c. Column Type: Set column type from the drop-down menu
    - d. Constant: Set constant value using the drop-down calendar.
    - e. Click the 'Apply' option.

CONFIGURATION DATA PREVIEW				
Column Name* (Add new columns to input data)	Column Type* (Set column type)	Constant (Set constant value)	•	Apply
Date 3	Date	25-04-2019 🔽	\$	×

- v) A message appears stating that the Constant fields are successfully configured.
- vi) Save the workflow.

iv)

vii) Run/Execute the workflow.

ADD NEW COLUMN

- viii) A notification appears at the bottom of the screen to inform about the process execution status.
- ix) The success of the data preview process gets indicated through the green checkmarks at the top of the components in a workflow.



Copyright © 2017-2019 BDB



Note: The user can also use the 'DATA PREVIEW' tab given in the 'Constants' component.

x) The set constant value gets applied to the selected column in the output dataset.

😝 Output 🖉				<b>i</b> <u>1</u>	+ +
CONFIGURATION DATA PRE	VIEW				
monthly_salary	comments	usd_billing	Date		
87556.33		4000.0	2019-04-25		<b>^</b>
28155.67		2400.0	2019-04-25		
29673.58		2400.0	2019-04-25		
63824.17		3000.0	2019-04-25		
25603.75		2400.0	2019-04-25		
25718.58		2400.0	2019-04-25		
56575.33	Relocating	3500.0	2019-04-25		

### Note:

- a. Click the '**Remove**' imes icon to remove the added constant information.
- b. The Constants component can take current date by clicking the  $^{42}$  icon.
  - i. Click the icon.
  - ii. The 'current date' option appears for the constant value field.
  - iii. Click the 'Apply' option.

CONFIGURATION DATA PREVIEW			
Column Name* (Add new columns to input data)	Column Type* (Set column type)	Constant <b>(Set constant value)</b>	3 Apply
Date	Date	2 [current date]	1 🔅 ×

- iv. Run the workflow and open the data preview.
- v. The set data gets added to the column created using the 'Constants' component.

🔁 Output 🥥			
CONFIGURATION DATA PREVIEW			
monthly_salary	comments	usd_billing	Date
63824.17		3000.0	2019-05-31
25603.75		2400.0	2019-05-31
25718.58		2400.0	2019-05-31
56575.33	Relocating	3500.0	2019-05-31
33565.75		2400.0	2019-05-31
37670.42	Not happy with the CTC	2400.0	2019-05-31
33565.75		2400.0	2019-05-31



### 5.2. Data Type

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

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



- iii) Click on the Data Type component to open the CONFIGURATION tab.
- iv) Click the 'ADD COLUMN' option to add a new column.
- v) 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 the following instance, the column data type for the selected columns has been changed from 'Decimal' to 'Decimal (Fixed)' and 'Date' to 'Date & Time.'

d. Click the 'Apply' option.

👬 Data Type 🖉				
CONFIGURATION DATA PREVIEW				
Column Name* (Select columns from input data)	Data Type (Change column data type)	Date Format/Infer Format (Select source date format)		Apply
previous_ctc [Decimal]	Decimal (Fixed)	•	×	
expected_joining_date [Date] -	Date & Time	•	×	

ADD COLUMN
------------

vi) A success message appears to confirm the Data Type field configuration.



# Datatype fields are configured successfully!

- vii) Save the workflow.
- viii) Run/Execute the workflow.
- ix) The success notification appears, and the components in the workflow get green checkmarks at the top.



### x) Click the 'DATA PREVIEW' tab for the Output component to see the transform result.

😝 Output 🖉			± ±     ±
CONFIGURATION DATA PREVIEW			
skills	expected_joining_date	joining_status	current_status
Dot Net Manager	2013-01-04T00:00:00.000+0000	joined	joined
Java	2013-01-04T00:00:00.000+0000	joined	joined
Dot Net	2013-01-04T00:00:00.000+0000	joined	joined
Java + Sql	2013-01-04T00:00:00.000+0000	joined	joined
Java	2013-01-04T00:00:00.000+0000	joined	joined
Java	2013-01-04T00:00:00.000+0000	joined	joined
Dot Net	2013-01-04T00:00:00.000+0000	joined	resigned
Dot Net	2013-01-04T00:00:00.000+0000	joined	joined

xi) Users can compare the data previews of the Input and Data Type modules (E.g., the selected input, in this case, contains the following column types)

🔒 Input 🖉			
CONFIGURATION DATA PREVIEW			
skills	expected_joining_date	joining_status	current_status
Dot Net Manager	2013-01-04	joined	joined
Java	2013-01-04	joined	joined
Dot Net	2013-01-04	joined	joined
Java + Sql	2013-01-04	joined	joined
Java	2013-01-04	joined	joined
Java	2013-01-04	joined	joined
Dot Net	2013-01-04	joined	resigned
Dot Net	2013-01-04	joined	joined

#### Note:

a. Users can get the same Data Preview as Output dataset while opening the '**DATA PREVIEW**' tab from any selected transform component. E.g., The '**DATA PREVIEW**' tab for the '**Data Type**' Transform component is as displayed below.

Data Type 🕗			
CONFIGURATION DATA PREVIEW			
skills	expected_joining_date	joining_status	current_status
Dot Net Manager	2013-01-04T00:00:00.000+0000	joined	joined
Java	2013-01-04T00:00:00.000+0000	joined	joined
Dot Net	2013-01-04T00:00:00.000+0000	joined	joined
Java + Sql	2013-01-04T00:00:00.000+0000	joined	joined
Java	2013-01-04T00:00:00.000+0000	joined	joined
Java	2013-01-04T00:00:00.000+0000	joined	joined
Dot Net	2013-01-04T00:00:00.000+0000	joined	resigned
Dot Net	2013-01-04T00:00:00.000+0000	joined	joined

### 5.2.1. Inferring Date & Date Time Formats

The Infer Date/Data Time functionality is provided for users to include various Date/Date Time formats which are not provided by ETL plugin. This functionality works only on the Text type of columns.

i) Select an input data set with some columns in the text data type (preferably the Input data set should have a column displaying Date in the Text data type).



🔒 Input 🖉		
CONFIGURATION DATA PREVIEW		
Excel Sheet 🗸	BASIC INFO SETTINGS DATA PREPARATION	
Q Search Query Services		Q Search Column
	Incremental Load is disabled because the input dataset you have chosen does not $\left( timestamp \right)$	contain a Delta Column of type long / date / datetime
0_jamal_sample1	Columns	Туре
0_jamal_sample2	Country	Text
0_jamal_sample3	Item Type	Text
0jamal_sample1	Order Date	Text
0jamal_sample2	Order ID	Text
0jamal_sample3	Order Priority	Text
1.excel	Region	Text

- ii) Connect the configured dataset with the Data Type transform component.
- iii) Provide the required information to configure the 'Data Type' component.
  - a. Pass the Date column(s) that is in Text data type from the input dataset.
  - b. Change the column data type using the drop-down icon provided for the 'Data Type' field.
  - c. The Date Format/Infer Format displays a drop-down icon to select an option for the source data format.
  - d. Use a checkmark in the given box to enable the Infer Format.
  - e. Click the '**Apply**' icon.
- iv) Run and execute the workflow.

🚟 Data Type 🖉							i	<b>i</b> ⊥	Ť
CONFIGURATION DATA PREVIEW									
Column Name* (Select columns from input data)	Data Type (Change column data type)		Date Format/Infer Format (Select source date format)				5	Apply	
Order Date [Text] 🔻	Date	•	day first	3 -	4	×			
ADD COLUMN REMOVE ALL COLUMNS									

- v) Open the 'DATA PREVIEW' tab of the Data Type component.
- vi) Check the column provided for the Data Infer Format function. It displays the selected date column in the original order (E.g., In this case, the selected column is '**Order Date'**).

🚠 Data Type 🕗			Ĩ
CONFIGURATION DATA PREVIEW			
Total Revenue	Total Cost	Total Profit	Order Date
694868.46	481808.34	213060.12	
1980685.3	1127652.19	853033.11	
49253.07	36530.68	12722.39	2014-12-07
1569766.23	1180466.46	389299.77	2016-01-05
370822.42	218868.51	151953.91	
4608613.17	3715141.92	893471.25	2016-10-04
1111240.14	709655.52	401584.62	2016-04-11
25545.54	18946.96	6598.58	



- vii) Check the selected Date column from the Input 'DATA PREVIEW' tab.
- viii) The data gets displayed in the selected Date Format.

CONFIGURATION DATA PREVIEW		
Order Priority	Order Date	Order ID
Μ	8/31/15	897751939
н	11/20/10	599480426
L	6/22/17	538911855
L	2/28/12	459845054
м	12/8/10	626391351
н	8/20/10	472974574
М	3/2/11	854331052
L	11/9/15	895509612

Note:

- a. The functionality only works for the '**Text**' type of column.
- b. If the source data format does not befit in the selected infer format, then those entries do not list in the output.
- c. Use the 'ADD COLUMN' option to add a new column for the Data Type change

### 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 times 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. Click the 'ADD NEW COLUMN' option.
  - b. Column Name: Provide a name for the New Column.
  - c. Operation: Select a data operation option from the drop-down menu.
  - d. Column/Value: Select a column or value for operations.
    - i. By selecting the '**column**' option, a drop-down menu appears.
    - ii. By selecting the 'value' option, users get redirected to enter a value.
      - E.g., In the following image, the selected option is Value for the operation, and three as a constant value gets added to the selected date.
  - e. Click the '**Apply**' option.

	📅 Date Operatio	ns 🖉		Î	<u>+</u>	+
	CONFIGURATION	DATA PREVIEW				
2	Column Name* (New column name)	3 Operation* (Select date operation)	Column / Value* (Select column/value for operation)	5	Apply	
	Next Date	Add days to date	▼ Value ▼ 🖶 03-06-2019 ▼ 🕀 Value ▼ 3		×	



- iv) Save the workflow.
- v) Run/Execute the workflow.
- vi) Open the 'DATA PREVIEW' tab of Date Operation or Output component.
- vii) The newly configured Date Column 'Next Date' gets added to the data with the set value.



🛃 Output 🧭			
CONFIGURATION DATA PREVIEW			
Source	Team	USDBilling	Next Date
Indeed	BU 6	4000.0	2019-06-06
Orgspire	BU 6	4000.0	2019-06-06
Orgspire	BU 11	2600.0	2019-06-06
Referral	BU 6	2300.0	2019-06-06
Referral	BU 6	1750.0	2019-06-06
BMS Innolabs	BU 7	0.0	2019-06-06
Orgspire	BU 7	0.0	2019-06-06

### 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) Drag the '**Filter**' component.
- iii) Connect the 'Filter' component to the configured input dataset and output component.



- iv) Configure the 'Filter' Component as described below:
  - a. Select a filter rule from the drop-down
    - i. ALL: By selecting this option filter gets applied only if all the added conditions are true
    - ii. ANY: By choosing this option filter gets applied even if any one condition is true

E.g., in this case, the selected Filter Rule is 'ANY' so the data preview displays data satisfying either of the filter conditions.

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



- f. Click the 'ADD NEW COLUMN' option to insert fields for the new column
- g. Click the 'Apply' option.

Image: Include rows that meet     ANY     of the following rules	NEV	v <u>1</u>	≔ + †
Filter       CONFIGURATION         DATA PREVIEW         Filter rules: include rows that meet       ANY         Include rows that meet       ANY	Ĩ	<u>+</u>	+
CONFIGURATION     DATA PREVIEW       Filter rules : Include rows that meet     ANY     of the following rules			
Filter rules : Include rows that meet ANY - of the following rules			
Column Name* 2 Operation* 3 Type* 4 Compare* (Enter/Select value to compare)	6	Apply	
candidate_id [Whole Number] - Equals - Value - 104		×	
name [Text] - Starts with - Value - P		×	



- v) Save & Run the workflow.
- vi) The input data gets filtered as per the applied conditions.

🟺 Filter ⊘			
CONFIGURATION DATA PREVIE	w		
id	candidate_id	name	gender
196	196	Poorvi	Female
148	148	Pankaj Kumar	Male
216	216	Pavan	Male
169	169	Pradeep	Male
147	147	Pragati	Female
144	144	Praful Managoli	Male
104	104	Virat Mishra	Male

### 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 Fields' 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. The 'Value	' option is c	hosen in the	below	given e	example.
-----------------	---------------	--------------	-------	---------	----------

	<b>f</b> <sub>X</sub> Formula Fields	0		Î	<u>+</u>	+
	CONFIGURATION	DATA PREVIEW				
1	Column Name* (Name of new column)	2 Calculation Type* (Set column type)	3 Select Columns for Calculation (Select columns to use in the calculation)	5	Apply	/
	Formula Field Column	Subtraction	Column      offered_ctc [Decimal]     Column      previous_ctc [Decimal]	] •	×	

#### ADD NEW COLUMN

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

🔁 Output 🖉			
CONFIGURATION	DATA PREVIEW		
previous_ctc	offered_ctc	monthly_salary	Formula Field Column
200000.0	180000.0	150000.0	-200000.0
200000.0	150000.0	125000.0	-50000.0
650000.0	1024000.0	85333.0	374000.0
580000.0	650000.0	54167.0	70000.0
500000.0	520000.0	43333.0	20000.0
730000.0	980000.0	81667.0	250000.0
510000.0	650000.0	54167.0	140000 0

### 5.6. Group By

The '**Group By**' feature allows multiple aggregations on the same or different columns. Users can obtain numerous 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: Select 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 🖉				Î	<u> </u>
CONFIGURATION	DATA PREVIEW				
1 Column Name*	Field Name				-
(Choose column)	candidate_id 😒				
2 New Column* (Aggregate column)		Column Aggregate* (Select column to aggregate)	4 Aggregate Type* (Select aggregate operation)	5	Apply
Aggregate Column		id	[Whole Number]	•	×

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

🔁 Output 🖉	
CONFIGURATION DATA PREVIEW	
candidate_id	Aggregate Column
191	191
222	222
112	112
113	113
155	155
167	167
198	198

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

Data Type	Aggregate
Text	Count
Date	First Value
Date Time	Last Value
Whole Number	Sum
Decimal	Average
Decimal (Fixed)	Minimum
	Maximum
	Standard Deviation
	Count
	First Value
	Last Value



# 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. Choose either of the options from the below given choices:
    - i. ADD COLUMN: Click this option to add one more column from the input dataset
    - ii. ADD ALL COLUMNS: Click this option to map all the columns from the input dataset
  - d. Click the 'Apply' option.

Note: The '**REMOVE ALL COLUMNS**' gets enabled after clicking on the '**ADD ALL COLUMNS**' option. Click this option to remove all the added columns for mapping.

	! : ⊗ ∎			NEW		=
	🖗 Mapping 🥏			Î	<u>+</u>	+
	CONFIGURATION DATA PREVIEW					
1	Column Name* (Select columns from input data)	2	Rename (Set new name)	Apply		
	candidate_id	[Whole Number] 🖍	Candidate ID	×		
	name	[Text] 🖍	Name of Candidates	×		





- iv) A success message appears to confirm the Mapping Component configuration.
- v) Save the workflow.
- vi) Run the workflow.
- vii) The notification messages appear to display the update of the data preview process.
- viii) A notification appears to confirm the successful completion of the Data Preview process. All the components in the workflow get green checkmark at the top.



ix) The aggregated column gets displayed in the '**DATA PREVIEW**' tab of the Mapping or Output component.

🔁 Output 🖉	
CONFIGURATION DATA PREVIEW	
Candidate ID	Name of Candidates
1	Ahsan R
2	Rajive Raveendra Pai
3	Amit Kumar Soni
4	Ritu
5	Vedprakash
6	Shiv Narayan Sahu
7	Animesh Srivastava
8	Vikram Bharti
9	Sudharshan Reddy
10	Ajish.T.Thomas



# 5.8. Replace Text

Users can search by whole word, sensitive to case, search for particular 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) Configure the '**Replace Text**' component as described below:
  - a. Click the 'ADD NEW COLUMN' option to get the configuration fields.
  - b. Column Name: Select a column from the input data set.
  - c. Search For: Enter a term to search from the selected column.
  - d. Replace With: Enter a term to replace the searched term in the input data.
  - e. Click the 'Apply' option.

Column Name* (Select columns to Search)	3	Search For* (Enter a term to search for)	Replace With (Replace found term with)	5 Apr
designation	[Text] 🔻	QA	Quality Assurance	¢ ×

- iv) Run the workflow.
- v) Save the workflow.
- vi) The notification messages appear to display the update of the data preview process.



- vii) A notification appears to confirm the successful completion of the Data Preview process. All the components in the workflow get green checkmark at the top.
- viii) Open the '**DATA PREVIEW**' tab from Replace Text or Output to see the replacement of the selected text in the column.

🤹 Replace Text 🥥			1 ÷
CONFIGURATION DATA PREVIEW			
name	gender	source	designation
Ahsan R	Male	Indeed	QE Manager
Rajive Raveendra Pai	Male	Orgspire	QE Architect
Amit Kumar Soni	Male	Orgspire	Senior Software Engineer
Ritu	Female	Referral	QE Engineer
Vedprakash	Male	Referral	QE Engineer
Shiv Narayan Sahu	Male	BMS Innolabs	Senior Software Engineer
Animesh Srivastava	Male	Orgspire	AWS Consultant
Vikram Bharti	Male	BMS Innolabs	Senior Software Engineer

#### Note:

- a. The users can click on the 'ADD NEW COLUMN' option to add multiple columns for any transform component.
- b. The users can also see data preview of the various transform components by clicking the 'DATA PREVIEW' tab of the respective options.

### 5.9. Data Preparation

The Data Preparation option comes pre-loaded with a series of transforms to be applied to the selected input data set.

- i) Navigate to the Workflow editor.
- ii) Connect the 'Data Preparation' component to the configured Input dataset and Output component.
- iii) Click the dragged Data Preparation component.





- iv) The Save Workflow window pops-up.
- v) Provide the Workflow name and select a workspace to save the workflow.

Jt	Save Workflow		×	
	Workflow Name * Data Preparation			1
	If you want, you can add a description to	explain what yo	u changed.	
3	Select workspace *	•	+	ľ
		CANCEL	SAVE	a

vi) After successful execution, the 'Launch Data Prep' option gets enabled for the Data Preparation component.

K : ► ⊗	8	৶	₽	0	Execute processing finished.												NE	w		=				
Data Preparation																								
	Input				Data Pr.																			
	<u>+</u> ]	J															 							-
翳 Data Preparat	tion 💋																			Î		Ţ	Ť	
CONFIGURATION	DATA	PREVIEV	V																					
器 Launch Data F	Prep							9	🎝 Re	efresh	Met	adat	а									Apply	/	
S No.	A	Actions ta	aken									Affe	ted	Colui	mn(s	)								

vii) Click the 'Launch Data Prep' option to launch it and perform various transforms on the input data set.



ta Pre	eparation			Exp	ort Settings Export			ne "D	C' 🕨	►I	×
							candidate_	_id			
		name == string				team	P 1	Transforms Row	Steps: 0		
1	1	Ahsan R	Male	Indeed	QA Manager	BU 6	Find a Funct				C
2	2	Rajiv		• ·		200					
3	3	Amit Rename Colu	mn				×				
4	4	Ritu						IL			
5		3 Rename this colu	mn					ditor			
6	6	Shiv Candidate IL						rm			
7	7	Anim									
8	8	Vikra						is			
9	9	Sudh					Submit				
10	10	Ajish.T.Thomas	Male	SkillRecruit	Business Analyst	BU 11	Concatenate	with			
11	11	Renuka	Female	Orgspire	Senior QA Engineer	BU 6	Delete colum	ın			
12	12	Arunjunai Sathis R	Male	Orgspire	QA Engineer	BU 6	Duplicate Co	lumns			
13	13	Jalavathi Batchu	Female	CareerNet	Senior QA Engineer	BU 6 🚽	Generate Pri	imary Key			
4						•	2 Rename Coli	umn			
223/ 2	23			« Previous 1	2 3 4 5	. 12 Next »	Poturn Non I		luor		

viii) Click the 'Export Steps to ETL' option

Dat	a Prej	paration			Ex	port Settings Export	Steps to ETL	Export S	teps to Pipeli	ne D	C 🕨	١	×
									Candidate	D			
		126	Adarsh Srivastava	Male	CareerNet	Software Engineer	BU 4	*	Profile	Transforms	Steps: 1		
			Roshan	Male	CareerNet	Senior Software Engin	BU 11		Column	Row			
	128		Sudhakar Magudeshw	Male	Referral	Associate Software En	BU 4		Find a Fund	tion			0
	129	129	Vandita H	Female	CareerNet	Senior QA Engineer	BU 6		T ING & F UNC				~

ix) A success message appears to assure the action of export.

Steps exported to ETL successfully!

- x) Refresh Metadata using the Data Prep Configuration tab.
- xi) Click the '**Apply**' option.

😂 Data Preparation 🧭				i.	<u>+</u>	Ť
CONFIGURATION DATA PREVI	IEW					
Saunch Data Prep		$\diamondsuit$ Refresh Metadata			Apply	
S No. Actions	s taken	Affect	ed Column(s)			
No Preparation steps available!						

- xii) A success message appears.
- xiii) The applied transforms get listed as displayed in the below image:



K : ► ⊗	8	৶	≡,	Q	DataPreparation fields are configured successfully!								N	IEW		=		
Data Preparation				Data Pr														
😂 Data Preparati	on ⊘																	
CONFIGURATION	DATA	PREVIE\	N			<i>A</i> -												
S No.	A	ctions ta	aken			Q Re	tresh M	Affe	a cted Co	olumn(	(s)				1	Арр	ly	
1	R	ename <u>.</u>	_COLUN	٨N				["	candid	ate_id'	ני							

xiv) Open the '**DATA PREVIEW**' tab using Data Preparation or Output component to see the transformed data.

😝 Output 🖉		ĩ	<u>+</u> +
CONFIGURATION DATA PREVIEW			
Candidate ID	name	gender	source
1	Ahsan R	Male	Indeed
2	Rajive Raveendra Pai	Male	Orgspire
3	Amit Kumar Soni	Male	Orgspire
4	Ritu	Female	Referral
5	Vedprakash	Male	Referral
6	Shiv Narayan Sahu	Male	BMS Innolabs
7	Animesh Srivastava	Male	Orgspire

# 6. Merge

The 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 becomes 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) Open the '**SETTINGS**' tab to see the available columns in the respective input dataset.
  - a. Input 1

🖯 Input 1 🖉		
CONFIGURATION DATA PREVIEW		
MYSQL	BASIC INFO     SETTINGS     DATA PREPARA	ATION
Q Search Query Services	Incremental Load	Q Search Column
12weeks	Columns	Туре
22datatypes	actual_joining_date	Text
a_20trending_tweet	candidate_id	Whole Number
a_all_social_influencer	comments	Text
a_c_most_recent_TopEngaging_tweet	current_status	Text
a_mention_fil2	designation	Text
a_mention_filter	expected_joining_date	Date
a_mention_trend	• experience	Decimal
a_mention_trend	• experience	Decimal NEW
a_mention_trend	• experience	Decimal NEW
a_mention_trend	exderience      BASIC INFO     SETTINGS     DATA PREPAR	
a_mention_trend	BASIC INFO     SETTINGS     DATA PREPAR     Data prepar	Decimal TEW TEW TEW TEW TEW TEW TEW TEW
a_mention_trend	BASIC INFO     SETTINGS     DATA PREPAR     Discremental Load     Incremental Load     Incremental Load     Incremental Load     Incremental Load	Decimal NEW I L VATION Q Search Column you have chosen does not contain a Delta Column of type long / date / date
a_mention_trend	EXPERIENCE  EXPERIENCE  BASIC INFO SETTINGS DATA PREPAR D Incremental Load Incrementat Incrementat Incrementat Incremental Incrementat In	Decimal  NEW  INTON  Q Search Column  you have chosen does not contain a Delta Column Type
a_mention_trend	Experience      BASIC INFO SETTINGS DATA PREPAR      Discremental Load      Incremental Load      Increme	Decimal  NEW  ATION  Q Search Column  you have chosen does not contain a Deta Column of type long / date / da  Type  Decimal
a_mention_trend	Experience      BASIC INFO SETTINGS DATA PREPAR      Differemental Load      Incremental Load      Increm	Decimal  NEW  NEW  NEW  NEW  NEW  NEW  Type  Decimal  Text
a_mention_trend	Experience      BASIC INFO SETTINGS DATA PREPAR      Discremental Load      Incremental Load      Incremental Load      Columns     candidate_id     comments     cur_monthly_payment	Decimal  TERM  ATTON  Q Search Column  you have chosen does not contain a Delta Column of type long / date / da  Type  Decimal  Text  Decimal
a_mention_trend	Experience      BASIC INFO SETTINGS DATA PREPAR      Differemental Load      Incremental Load      Increm	Decimal  NEW  NEW  NEW  NEW  NEW  NEW  NETION  Q Search Column  you have chosen does not contain a Delta Column of type long / date / da  Type  Type  Type  Decimal  Text  Decimal  Text
a_mention_trend	BASIC INFO SETTINGS DATA PREPAR     BASIC INFO SETTINGS DATA PREPAR     Incremental Load is disabled because the input dataset y     (comesman)     Columns     candidate_id     comments     current_status     designation	Decimal  TERM  Decimal  Decima
a_mention_trend	Experience      BASIC INFO SETTINGS DATA PREPAR     Dirremental Load      Incremental Load      Increment	Decimal

### iv) Connect the '**Append**' component with the configured Input datasets and an Output component.

Contransform	Input 1	Output
Merge	Append	
Append Join		



- v) Click on the dragged Append component to get the configuration fields.
- vi) Select 'INCLUDE ALL COLUMNS' option using the 'Select Columns' drop-down menu.

Pend 🖉	Î	<u>+</u>	Ť
CONFIGURATION DATA PREVIEW			
Select columns*         Review changes           (Which columns should be included in this append?)         (Review the changes that will made to each DataSet)		N	ext
INCLUDE ALL COLUMNS (Review description]			

- vii) Click the 'Next' option.
- viii) The columns from both the selected input data sets get displayed.

🔁 Output 🖉		i	<u>+</u> +
CONFIGURATION DATA PREVIEW			
name	source	monthly_salary	joining_statu
Emp ID 1	internal	87556.33	
Emp ID 2	internal	28155.67	
Emp ID 3	internal	29673.58	
Emp ID 4	internal	63824.17	
Emp ID 5	internal	25603.75	
Emp ID 6	agency	25718.58	
Emp ID 7	portal	56575.33	

- ix) Save the workflow.
- x) Run the workflow.
- xi) The notification messages appear displaying update of the data preview process.
- xii) A notification appears to confirm the successful completion of the Data Preview process. All the components in the workflow get green checkmark at the top.





xiii) The selected columns of both the input data sets get appended in the '**DATA PREVIEW**' tab of the Append and Output components.

😂 Output 🖉		Î	<u>↓</u>
CONFIGURATION DATA PREVIEW			
name	source	monthly_salary	joining_status
Emp ID 1	internal	87556.33	
Emp ID 2	internal	28155.67	
Emp ID 3	internal	29673.58	
Emp ID 4	internal	63824.17	
Emp ID 5	internal	25603.75	
Emp ID 6	agency	25718.58	
Emp ID 7	portal	56575.33	

# 6.1.2. Append Only Shared Columns

- i) Connect the '**Append**' component to the configures input datasets and an output component (As described in section 6.1.1.).
- ii) Click on the 'Append' component from the workflow.
- iii) The 'CONFIGURATION' tab opens.
- iv) Choose 'ONLY INCLUDE SHARED COLUMNS' as an option to append the datasets.
- v) Click the 'Next' option.

🗗 Append 🥥		i i	<u>+</u>	Ť
CONFIGURATION DATA PREVIEW				
Select columns* Review changes (Which columns should be included in this append?) (Review the changes	s that will made to each DataSet)		Ne	ext
ONLY INCLUDE SHARED CO	ption]			

- vi) All the columns from both the input data sets get displayed for the user to select the columns from both the datasets. If the user does not select any column, it automatically takes all the shared columns between both the selected datasets.
- vii) Click the 'Apply' option to configure the Append component.
- viii) A success message appears if the configuration is successful.



>  : ▶ ⊗ 🖬 🗄 ≒ 🕓		Append fields are configured successfully!	NEW
🗗 Append 🥥			i <u>i</u> +
CONFIGURATION DATA PREVIEW			
hiring_data_etl_demo (Select column)	ETL_cube (Select column)	Rename column (Set new name)	Previous Apply
gender [	Text] 🖍 gender	[Text] 🖉 gender	×
source	Text] 🖍 Source	[Text] 🖌 Source	×
designation	Text] 🖍 designation	[Text] 🖌 designation	×
team	Text] 🖍 team	[Text] 🖌 team	×
skills	Text] 🖍 Skills	[Text] 🖌 skills	×

### ADD COLUMN

- ix) Save the Workflow.
- x) Run the Workflow.
- xi) After successful execution of the process, all the components in the workflow get green checkmarks, and a success message appears to confirm the successful completion of the process.
   Note: In case of the failure in the execution process, a notification comes to inform the users about the same. The user can see the reason behind the collapse of the process by opening the 'Job' UI.



xii) Open the '**DATA PREVIEW**' tab from the Append or Output components to see the appended columns.



🔁 Output 🖉				Î	<u>+</u>	*
CONFIGURATION DATA PREVIEW						
source	designation	team	skills	gender		
internal	engineering manager	bu 1 engineering	Dot Net Manager	male		
internal	senior software engineer	bu 1 engineering	Java	female		
internal	senior software engineer	bu 1 engineering	Dot Net	female		
internal	team lead	bu 1 engineering	Java + Sql	0		
internal	senior software engineer	bu 2 engineering	Java	1		
agency	senior software engineer	bu 2 engineering	Java	1		
portal	technical architect	bu 1 engineering	Dot Net	female		
portal	senior software engineer	bu 2 engineering	Dot Net	1		

# 6.2. Join

The 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 🕗		
CONFIGURATION DATA PREVIEW		
MYSQL 🗸	BASIC INFO SETTINGS DATA PREPARATIO	ИС
Q Search Query Services	Incremental Load	Q Search Column
19feb_1	Columns	Туре
19feb_1	add new1	Decimal
19feb_1	date	Date
19feb_2	doj	Date & Time
21feb	id	Whole Number
22Datatypes_21Nov18	longdata	Whole Number
22feb 👻	name	Text -

Input Data Set 2



🖯 Input 2 🖉			i <u>+</u>	+
CONFIGURATION	DATA PREVIEW			
Excel Sheet	•	BASIC INFO	DATA PREPARATION	
Q Search Query Service	s 🖪		Q Search Column	
1.excel	Î	Incremental Load is disabled beca type long / date / datetime (timest	use the input dataset you have chosen does not contain a Delta Column amp)	of
100CSV		Columns	Туре	
10k test		Country	Text	<b>^</b>
10k_Excel		Item Type	Text	
10kCSV		Order Date	Text	
10kexcel		Order ID	Decimal	
		Order Priority	Text	

- ii) Drag the '**Join**' component to the workspace.
- iii) Connect the '**Join**' component to the above-given input datasets and one output component to complete the workflow.

Q Search Component	
DATA •	Untitled Workflow
Co TRANSFORM	Join Output
MERGE	

- iv) Configure the 'Join' component as described below:
  - a. Identify Column: Identify a column from the first Input dataset.
  - 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 second Input dataset.
  - d. Click the '**Next**' option.



[Whole Number] • Id [Whole Number] • X

- e. The next screen appears prompting the user to select columns from both the input datasets. The users can use the '**Select all column'** option to select all the columns. However, they can also deselect any column by removing the checkmark from the box given next to the column name.
- f. Click the 'Apply' option to configure the selection of columns.

	🛧 Join 🕗					■ ±	+
	CONFIGURATION DATA PREVIE	W					
	customer			cust	Previot	7 Apply	
5	Select all column	Set Alias	6	Select all column	Set Alias		Î
	id [Whole Number] 🖍	id	×	🔽 id [Whole Number] 🖌	ID	×	
	name [Text] 🖍	name	×	🗸 name [Text] 🖌	Name	×	
	✓ date [Date] ✓	date	×	🗸 date [Date] 🖌	Date	×	
	🗸 doj [Date & Time] 🖌	doj	×	🗸 doj [Date & Time ] 🖌	DOJ	×	*
_	ADD COLUMN			ADD COLUMN			

g. A notification appears to confirm the status of the configuration process. (Successor error status is communicated through the color-coded pop-up messages.)

Note:

- a. The 'Inner' is the default Join Type.
- b. Users can apply multiple joins by using the 'ADD COLUMN' tab.
- v) Save the workflow.
- vi) Run/Execute the workflow for data preview.
- vii) The notification messages appear to display the update of the data preview process.
- viii) A notification appears to confirm the successful completion of the Data Preview process. All the components in the workflow get green checkmark at the top.
   Note: The notification message also gets displayed for an error in the data preview process.



K 🛛 🕨 🗟 🐱 🗮 🕓	Preview processing is finished.		NEW :=
Inner Join			
			$ \frac{1}{2} = -\frac{1}{2} $
	Output		
Input 2			
🔁 Output 🖉			
CONFIGURATION DATA PREVIEW			
Connector Elastic -			
Select resource	•	ETL - Execution Preview processing Sur	¢×
Select Mapping ID		qa.bdbizviz.com	

ix) Open the '**DATA PREVIEW**' tab using the Join or Output component to see the merged data based on the selected Join type.

★ Join 🕗		Î	<u>↓</u> +
CONFIGURATION DATA PREVIEW			
id	name	date	doj
1	naincy	2019-06-10	2019-06-06T12
2	nai	2019-06-10	2019-06-06T13

# 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 💋		i <u>+</u>	· +
CONFIGURATION DATA PREVIE	EW		
MYSQL	BASIC INFO SETTINGS DATA PREPARATION		
Q Search Query Services	Incremental Load	Q Search Column	
19feb_1	Columns	Туре	
19feb_1	add new1	Decimal	<b>^</b>
19feb_1	date	Date	
19feb_2	doj	Date & Time	- 1
21feb	id	Whole Number	1
22Datatypes_21Nov18	longdata	Whole Number	
22feb	• name	Text	-



### 2. Input Dataset 2

🔒 Input 2 🖉		i <u>+</u>	+ +
CONFIGURATION DATA PREVIEW			
Excel Sheet 🗸	BASIC INFO SETTINGS	DATA PREPARATION	
🔍 Search Query Services 🕢 🔒		Q Search Column	
1.excel	Incremental Load is disabled becau type long / date / datetime (timesta	use the input dataset you have chosen does not contain a Delta Colun amp)	nn of
100CSV	Columns	Туре	
10k test	Country	Text	^
10k_Excel	Item Type	Text	
10kCSV	Order Date	Text	
10kexcel	Order ID	Decimal	
20kdata col	Order Priority	Text	

#### a) Inner Join

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Configure the 'Join' component for the 'Inner' join type.

	DATA PREVIEW				
lentifying Column*	2	Join Type*	3 Matching Column*	4	N
customer		() Inner	Cust		
d	[Whole Number]		id	 ~	

- iii. Save and run the workflow to get data preview of the merged data.
- iv. Click the '**DATA PREVIEW**' tab using the Join or Output component to view the merged datasets.
- v. The DATA PREVIEW displays only matching columns.

🛧 Join 🖉			Î	<u>+</u>	+
CONFIGURATION DATA PREVIEW					
id	name	date		doj	
1	naincy	2019-06-10		2019-0	6-06T12
2	nai	2019-06-10		2019-0	6-06T13

b) Left Outer Join



- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Configure the 'Join' component for the 'Left Outer' join type.

🛧 Join 🥥				Î	<u>+</u>	+
CONFIGURATION DATA PREVIEW						
Identifying Column*	Join Type*	Matching Column*			Next	
customer	Left Outer	cust				
id [Whole Number] 👻		id	[Whole Number]	×		

ADD	COLUMN
ADD	COLOIVIN

- iii. Save and run the workflow to get data preview of the merged data.
- iv. Click the '**DATA PREVIEW**' tab using the Join or Output component to view the merged datasets.
- v. The '**DATA PREVIEW**' tab displays the complete data from the left input dataset and matching columns from the other input data set.

🛧 Join 🧭		Î	<u>↓</u> +
CONFIGURATION DATA PREVIEW			
id	name	date	doj
1	naincy	2019-06-10	2019-06-06T12
6	ranjeet	2019-06-06	2018-08-20T12
5	update	2019-06-07	2018-07-09T15
7	manjhari	2019-06-08	2018-08-31T16
2	nai	2019-06-10	2019-06-06T13

### c) Right Outer Join

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Configure the 'Join' component for the 'Right Outer' join type.



★ Join ⊘				Î	Ţ	<b>↓</b> ↑
CONFIGURATION DATA PREVIEW						
Identifying Column*	Join Type*	Matching Column*			Next	
customer	C Right Outer	cust				
id [Whole Number] 🔻		id	[Whole Number] 🗧	×		

#### ADD COLUMN

- iii. Save and run the workflow to get data preview of the merged data.
- iv. Click the 'DATA PREVIEW' tab using the Join or Output component to view the merged datasets.
- v. The '**DATA PREVIEW**' tab displays the complete data from the left input dataset and matching columns from the other input data set.

🛧 Join 🥏			■ <u>↓</u> + +
CONFIGURATION DATA PREVIEW	_		
new_column	add new1	ID	Name
bizviz	121212	1	d
		16	fff
bizviz	121212	2	entry
		1212	ffw223
		44	4erfe
		12	grgf
		19	e2

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

### d) Full Outer

- i. Connect the join component to the configured input datasets and output component to create a workflow.
- ii. Configure the 'Join' component for the 'Full Outer' join type.



🛧 🛛 Join 🖉						Î	<u>+</u>	-
CONFIGURATION DATA F	PREVIEW							
Identifying Column*		Join Type*		Matching Column*		[	Next	
customer		Full Outer	$\odot$	cust				
id	[Whole Number]			id	[Whole Number]	×		

ADD	COLUMN

- iii. Save and run the workflow to get data preview of the merged data.
- iv. Click the '**DATA PREVIEW**' tab using the Join or Output component to view the merged datasets.

🛧 Join 🖉			
CONFIGURATION DATA PREVIEW			
new_column	add new1	ID	Name
bizviz	121212	1	d
bizviz	121212		
		16	fff
bizviz	121212		
bizviz	121212		
bizviz	121212	2	entry
		1212	ffw223

# 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 the 'Scheduler' option from the drop-down menu.
- iii) The 'Schedule Monitoring' page opens.
- iv) The scheduled workflow gets listed on the Schedule Monitoring page.
- v) By clicking a scheduled workflow from the list, the following schedule details appear:
  - a. Scheduler Name
  - b. Last Updated Date
  - c. Recurrence (Date and Time)
  - d. Status



ETL	]						NEW	:=	0
7	Home	Schedule Monitoring							
6	My Workspace	Search Schedule	С	Scheduler Name	Last Updated Date 🧄 🔨	Recurrence	Status		
Ô	Job	<ul> <li>Incremental_laodtestQA</li> </ul>	Û	Incremental_laodtestQA	6/10/2019, 10:58:00 PM	6/11/2019, 10:58:00 PM	Scheduler success		
Û	Trash	mysql_dataprep	Ċ	Incremental_laodtestQA	6/7/2019, 10:58:00 PM	6/8/2019, 10:58:00 PM	Scheduler success		
0	Scheduler	<ul> <li>mssql</li> </ul>	Ċ	Incremental_laodtestQA	6/6/2019, 10:58:01 PM	6/7/2019, 10:58:00 PM	Scheduler success		
		9 37k	Ċ	Incremental_laodtestQA	6/5/2019, 10:58:00 PM	6/6/2019, 10:58:00 PM	Scheduler success		
		• gssch	Ċ	Incremental_laodtestQA	6/4/2019, 10:58:01 PM	6/5/2019, 10:58:00 PM	Scheduler success		
		age	С U	Incremental_laodtestQA	6/3/2019, 10:58:00 PM	6/4/2019, 10:58:00 PM	Scheduler success		

# 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: The user can schedule the job daily by using this option.
  - a. Click the 'Scheduler' 🕓 icon on the workflow editor.
  - b. Choose the '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 Weekday
      - 3. Set the start time using the drop-down
  - c. Click the 'Schedule' option.

K			⊗	8 4	=,	
Schedule Workflow					×	
DAILY WEEKLY M	ONTHLY	YEARLY			-	
Every 1 Every Week Day Start : time 12 • 00 •	day(s)				-	
ONLY INCLUDE SHARE	[Review des	cription]	Cancel	Sched	ule	

- 2. Weekly: The user can schedule the job weekly by using this option.
  - a. Click the '**Scheduler**' Oicon on the workflow editor.
  - b. Choose the '**WEEKLY**' 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 on the box.



- 2. Set the start time using the drop-down.
- c. Click the 'Schedule' option.



- **3.** Monthly: User can schedule the job monthly by using this option.
  - a. Click the '**Scheduler**' O icon on the workflow editor.
  - b. Choose the 'MONTHLY' 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 menu.
  - c. Click the 'Schedule' option.

K	▶ ⊗ 🖬 坐 🗮 🕓
Schedule Workflow	×
DAILY WEEKLY MONTHLY YEARLY	
O Day 1 of every 1	month(s)
O The First  ✔ Monday  ✔ of every 1	month(s)
Start : time 12 • 00 •	)
	Cancel Schedule

- 4. Yearly: User can schedule the job yearly by using this option.
  - a. Click the '**Scheduler**' O icon on the workflow editor
  - b. Choose the 'YEARLY' 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 the 'Schedule' option.



۲				:		8	8		₹,	
1	Schedule W	orkflow						×	-	
	DAILY	WEEKLY	MONTHLY	YEARLY	_					
	O Every J	anuary 🔻 1								
	O The Fir	rst 👻 Monda	y 👻 of January	•					l	Ĩ
	Start time 12	: • 00 ·	-						)	
						Cancel	9	Schedule		

# 8. Signing Out

The user can log out from the BDB ETL plugin at any given stage. Users need to click on the '**Close**' × option to close the ETL page.



The following steps describe how to Sing-off from the BDB Platform.

- i) Click the 'User Profile' icon on the Platform homepage.
- ii) Click the 'Sign Out' option.



		Ļ <sup>12</sup>	?	A			
- A adm admi	nin n@bdbizv	iz.com					
e My Account							
i Abour	t						
U Sign (	Dut						

iii) The user successfully signs off from the BDB Platform.

### Note:

- a. By clicking the 'Sign Out' option, the user gets back to the Sign-in page of the BDB platform.
- b. Click the 'About' option to open the default homepage for the BDB Platform.

