

**BDB ETL – High Level Introduction**

**AI, Data Science & Advanced Analytics Platform**

**Unique Disruptive Platform by Ex-SAP BO R&D Team**

**One Platform, End to End Solution**

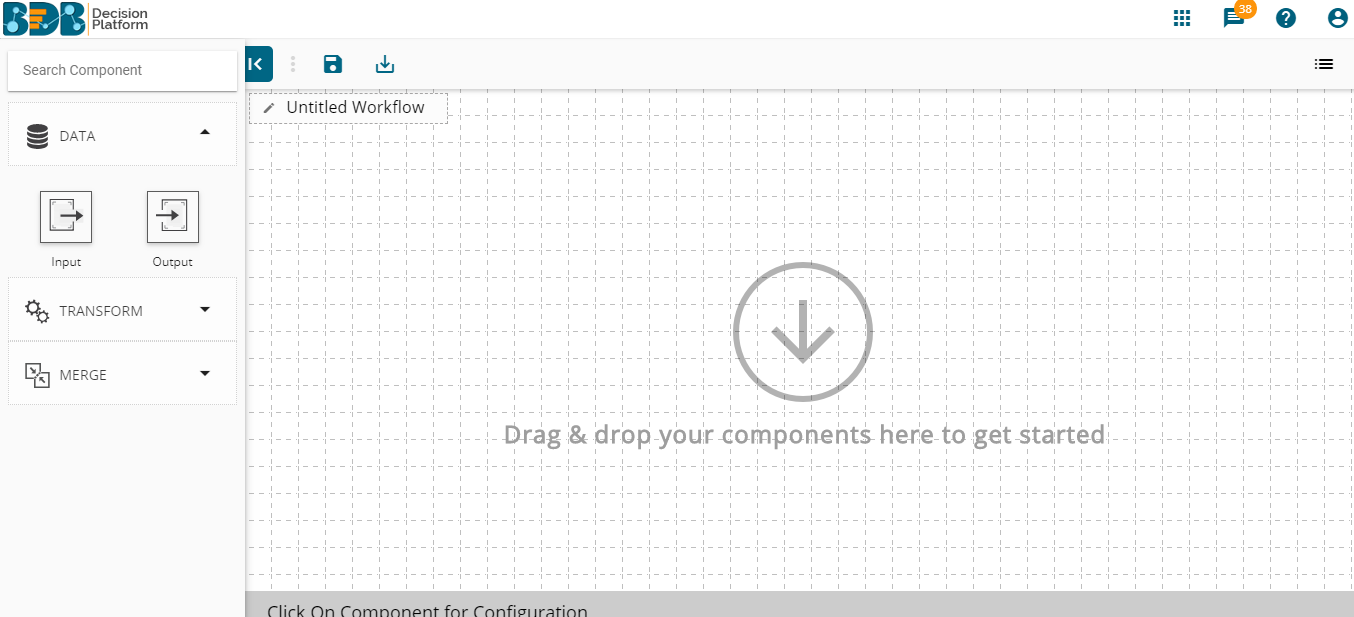
**Removes industry pain points**

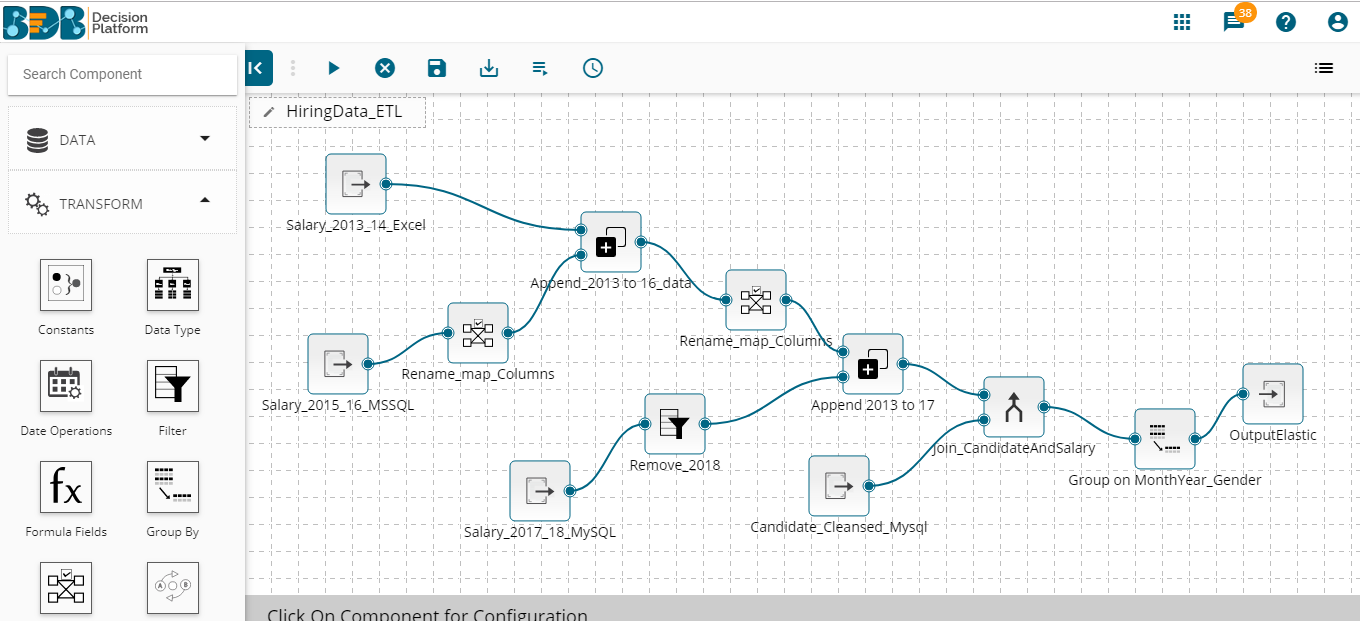
**Leads to exponential benefits realization**

BDB ETL-

**BDB ETL is a drag and drop based secure yet self-driven module for Data Wrangling. It seamlessly pulls, cleanse, transform and load data from diverse and prominent data sources like MySQL, CSV, Excel etc. without any previous ETL or SQL skills.**

BDB ETL lets you visually inspect your data flow in a flow chart. You can configure actions, schedule ETL workflows and preview result. BDB ETL empowers all skill level users to extract and transform data easily by enforcing data quality and consistency standards, merging various data sources and store quality data output in data lake, RDBMS, Elastic etc.





**BDB ETL Purpose-**

##### **Extract data from multiple data sources to cleanse & transform your raw data without SQL skills**

##### **BDB ETL workflows prepare data to be analysis ready**

##### **Drag & drop based interface - designed as per ease of use**

##### **Integrated BDB Data Prep module inside BDB ETL for deeper cleaning of data**

**High-Level Features of BDB ETL**

* Drag & Drop based interface for designing workflow
* Major data transform components to clean raw data
* Combine multiple data source through Append & Join
* Embedded BDB data prep inside ETL to clean data at the column level.
* Quick run the ETL workflows or schedule transformation jobs for any time.
* Monitor the status of scheduled ETL Workflows, success & failure of jobs.

**Deep-Dive into Features**

BDB ETL module has below mentioned major features –

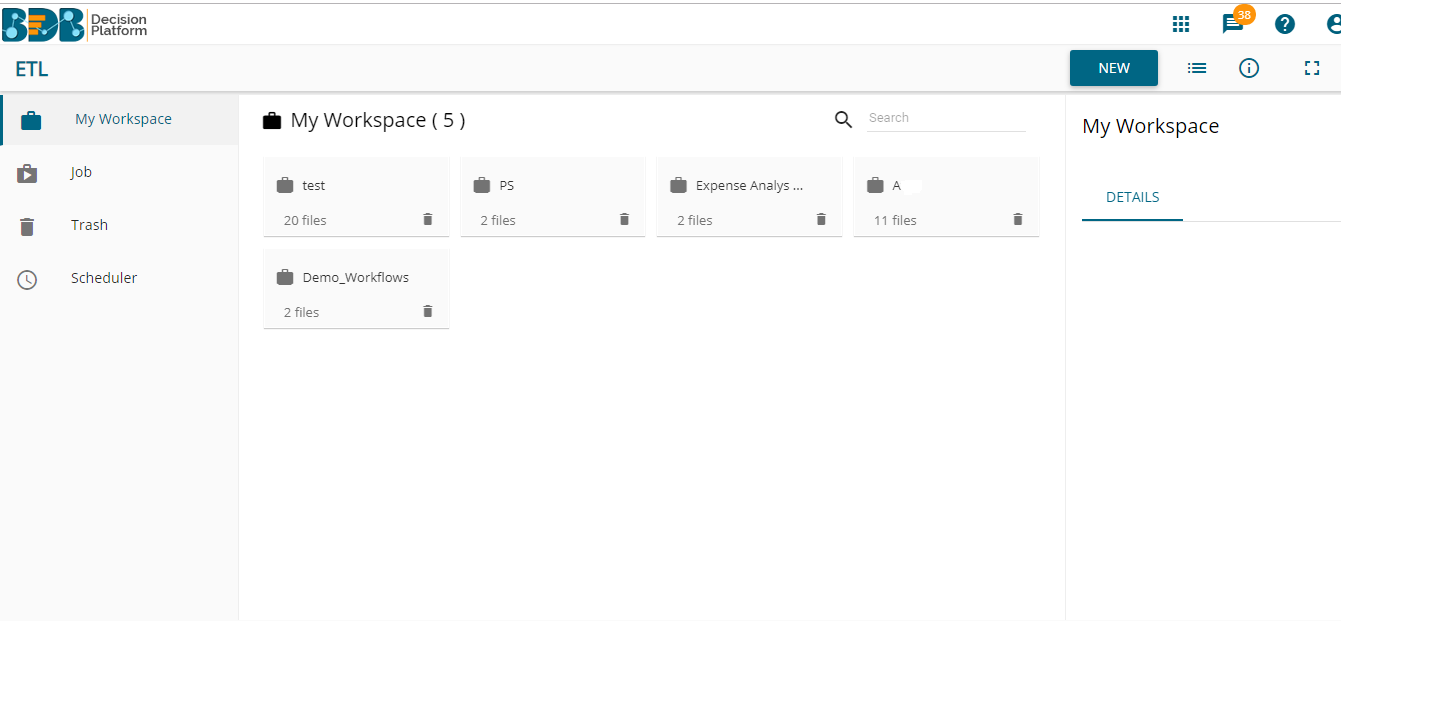
* My workspace – For creating, new ETL workflows & accessing the older ones
* Job – To monitor the status of any ETL workflow and look up the details around workflows
* Trash – Trash space for deleted workflows
* Scheduler – It provides option to schedule a workflow and manage schedule related configurations

# My workspace

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

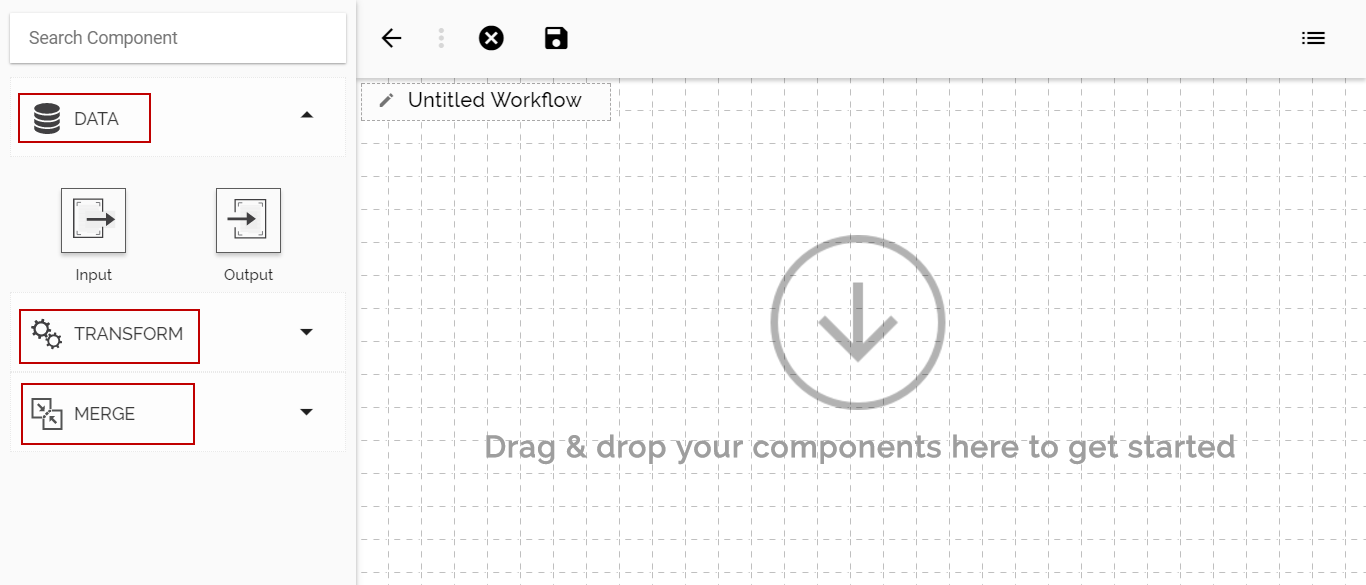
## Workflow Editor

‘**My Workspace**’ is a placeholder for the previously created ETL workflows. One can start creating new workflows from here



The “New Workspace” exposes users to 3 main aspects to autonomously prepare data:

* 1. Data
     + Input Data
     + Output Data
  2. Transform
     + Input Data
     + Output Data
     + Constant
     + Data Type
     + Date Operations
     + Filter
     + Formula Fields
     + Group By
     + Mapping
     + Replacing Text
  3. Merge-
     + Append
     + Join



## Data

## Extracting Data (Input Data)

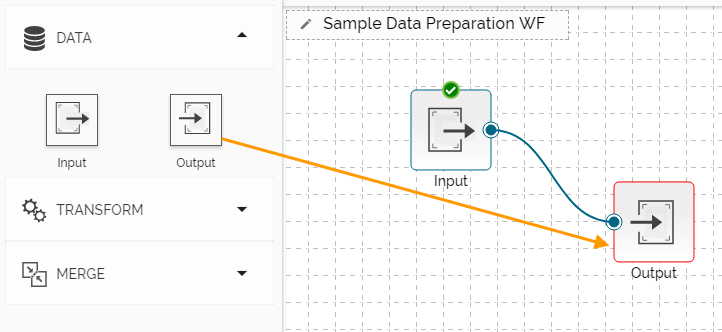
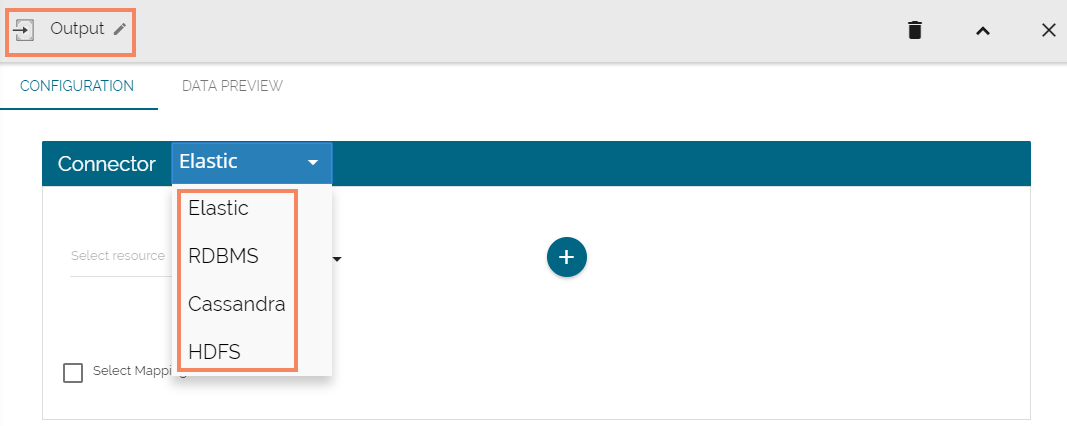
One can start extracting data by data input component, configuring it.  
(It can extract data from MYSQL, MSSQL, Oracle, Google Sheet & Excel Sheet)



## Loading Data (Output Data)

Users can load the transformed data into an elastic for visualization via the output component. One can load data into -

* 1. Elastic
  2. RDBMS
  3. Cassandra
  4. HDFS

## Transform

## Constants

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

## Data Type

Users can change the data type of the selected columns by using the ‘**Date Type.**’

## 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.

## 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.

## Formula Fields

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

## 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.

## Mapping

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

## 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.

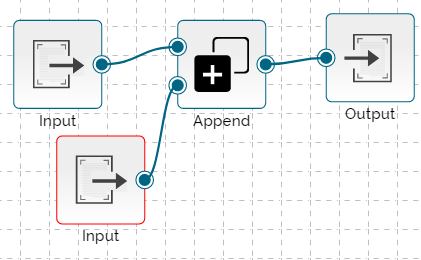
## Merge

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

## Append

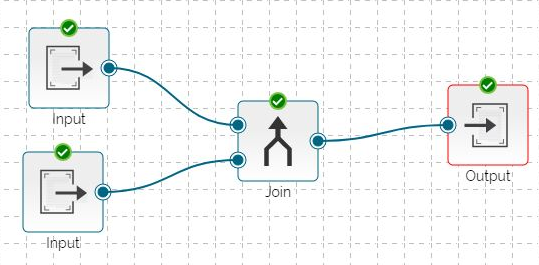
The ‘**Append**’ feature combines one dataset on top of another. If the datasets are of different structures, still the union is possible, and the output will be a unified more massive structure with NULL values populated wherever data is missing. User can apply append in following ways

* Append All Columns
* Append Only Shared Columns



## Join

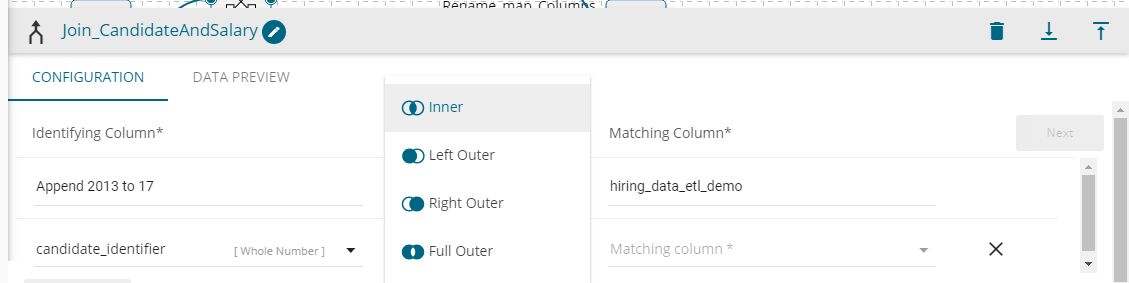
Users can join two datasets and use the merged output to write the workflow in the selected metadata.



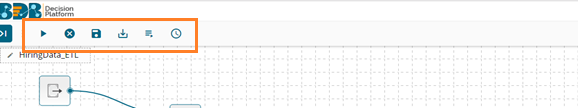
### Join Types:

The ‘**Join**’ feature offers four types of join to merge datasets.

1. **Inner Join**
2. **Left Outer Join**
3. **Right Outer Join**
4. **Full Outer**



**Additional Features around ETL workflows**

****

## Saving & Run Preview

## Clear Workflow

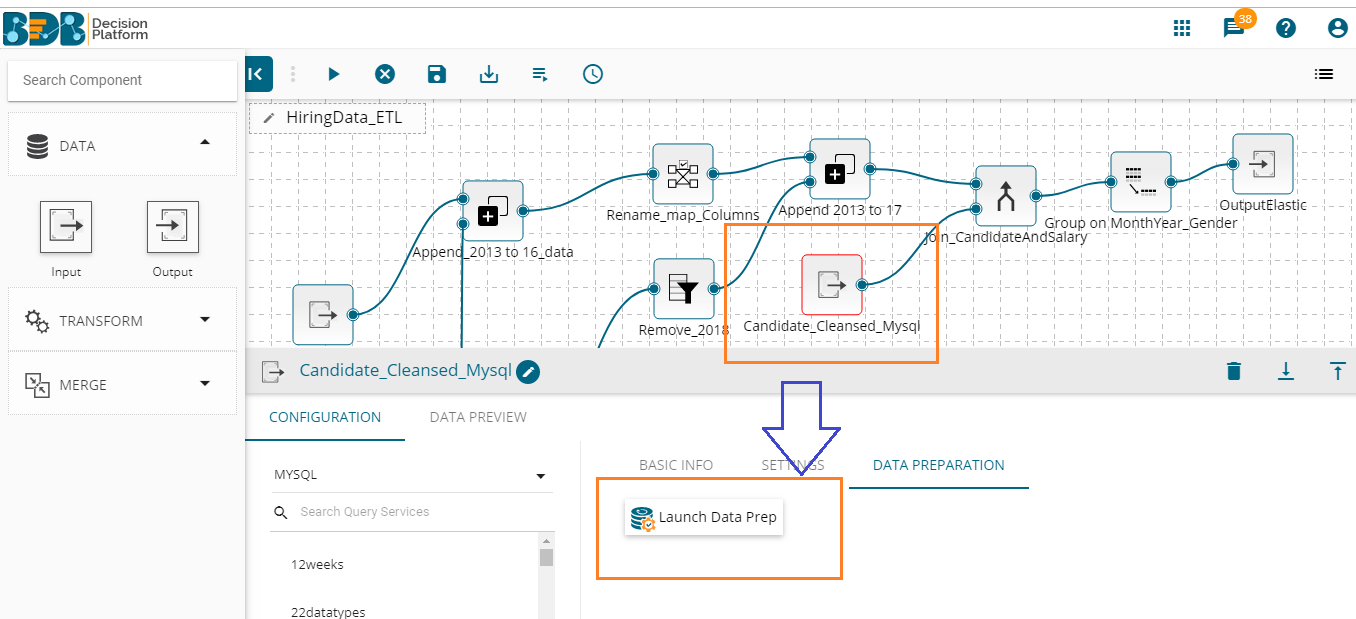
## Save

## Save & Execute

## Schedule

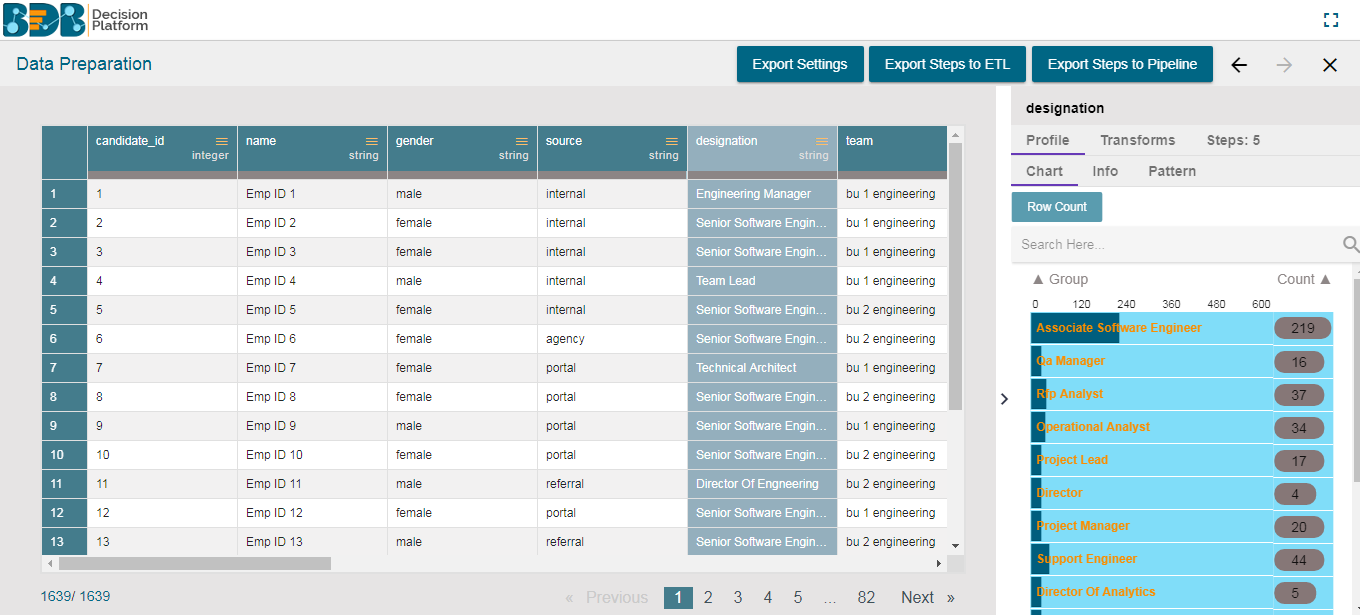
***Note: Using all these components, one can create complex ETL workflows (example shown in image), save & run/ schedule it & clean data easily for analysis purpose***

***For a deeper level of data cleansing, BDB data Preparation module is embedded inside BDB ETL (which can be accessed through configured input data component)***

******

***Significance of BDB Data Prep in BDB ETL***

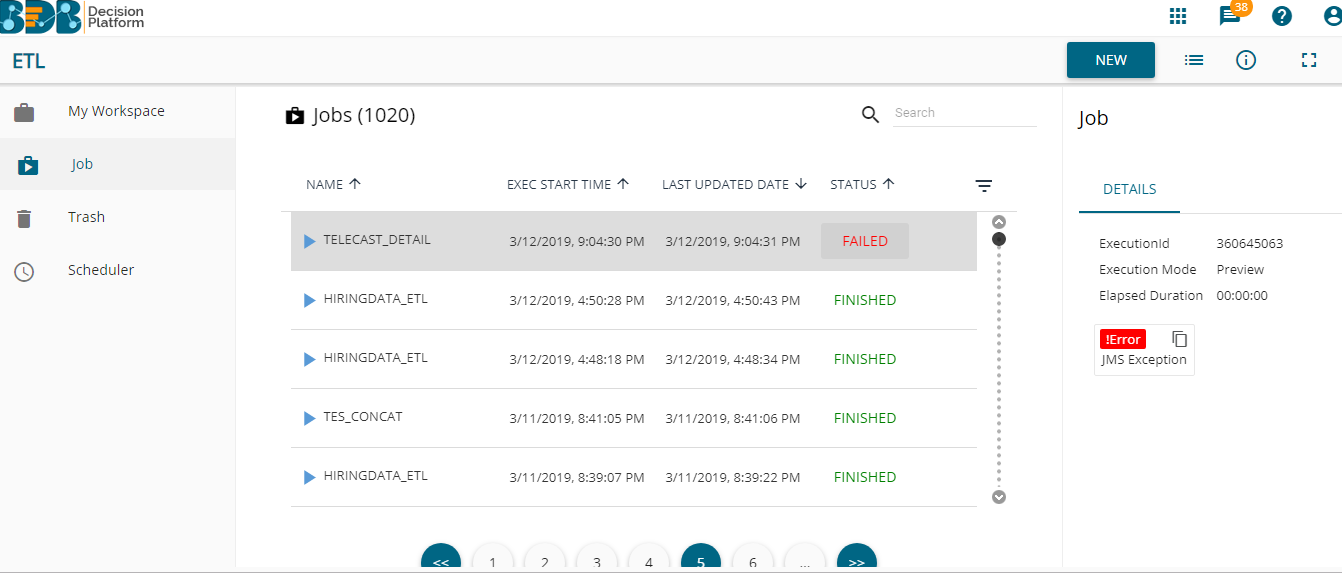
BDB Data Prep lets a user transform data at column level. It visually profiles the data into grid format which lets user perform transformations from across 50+ available transforms. (For more details, refer BDB data prep document)



Transformations performed in this can be extracted to ETL workflow.

# Jobs

Users can monitor the job status from this space and check which ETL workflow is failed, finished, killed or running.

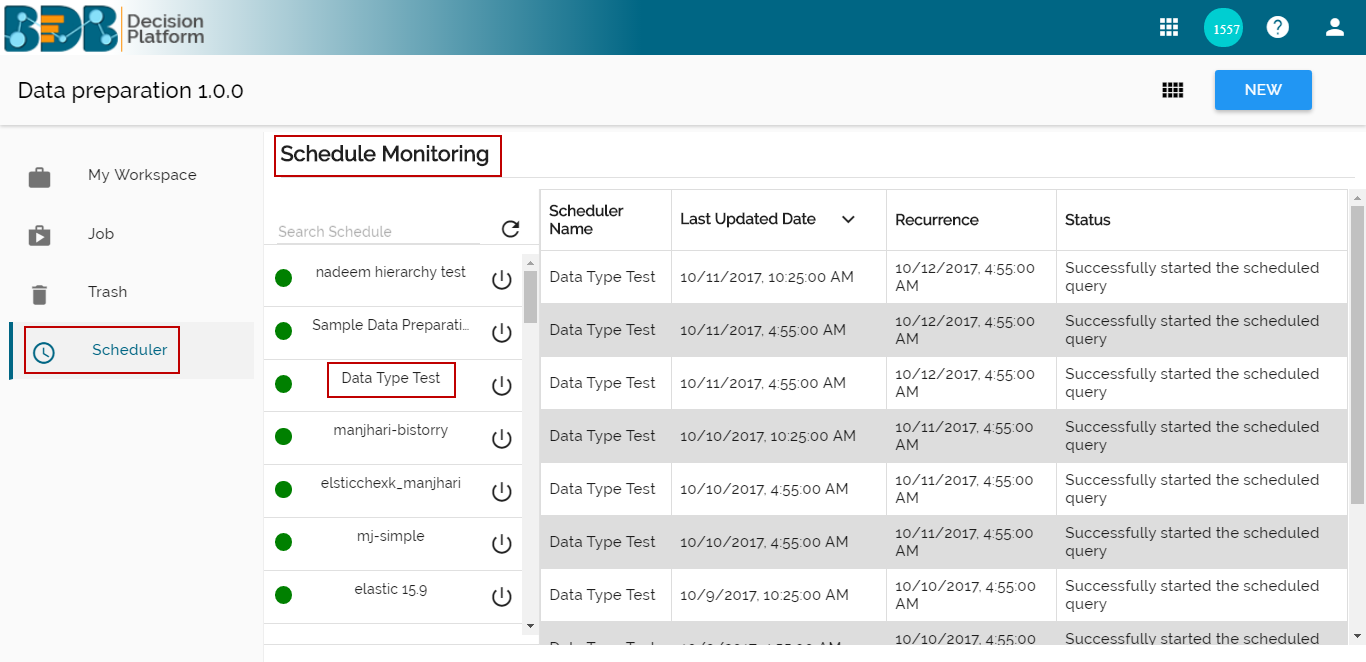


# Trash

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

# Scheduler

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



## Schedule Configuration Options

These options are provided to configure a range of time for a scheduled workflow – daily, weekly, monthly & yearly.

