

Arvind Pal

Arvind.pal@itcinfotech.com

ZEAS

Ingestion Scheduler

**ZEAS – Ingestion Scheduler**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version | Date | Created by | Reviewed by | Changes |
| 1.0.0 | 7/12/2016 | Arvind Singh Pal |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Table of Contents

[Introduction 3](#_Toc468906463)

[Pre-requisites 3](#_Toc468906464)

[Business Requirements 3](#_Toc468906465)

[Existing functionality 3](#_Toc468906466)

[Proposed functionality 3](#_Toc468906467)

[Option 1 4](#_Toc468906468)

[Detailed functionality 4](#_Toc468906469)

[Option 2 6](#_Toc468906470)

[Detailed functionality 7](#_Toc468906471)

[Assumptions and Risks 15](#_Toc468906472)

# Introduction

ZEAS is ETL tool which provide single platform for data lack, Ingestion and data transformation, it provides data connection which supports large range of data-sources formats such as

RDBMS – MYSQL, Oracle and DB2

Files - CSV, XLS, JSON, XML, Fixed width delimiter etc.

# Pre-requisites

ZEAS should be running in any one Hadoop distributed enlivenments (Apache Hadoop or HDP) with below Hadoop services-

* Hive
* Sqoop
* Oozie

# Business Requirements

User has created dataset for “Insurance claim data” for some region and want to analysis data every months or every week.

He will schedule this dataset based on his requirement which will keep giving him output without running dataset manually.

# Existing functionality

ZEAS allows to create profile for Ingestion, user can apply filters or data cleansing rules when creating profile using data source.

Once dataset is ready user can kick start for Ingestion, currently Ingestion for dataset in one time.

# Proposed functionality

I am proposing two ways to schedule dataset Ingestions.

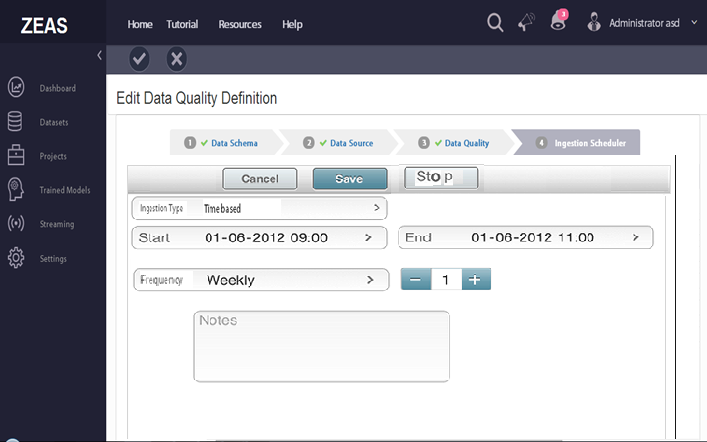
# Option 1

While creating dataset, give choice to schedule dataset by choice and allow user to select start date, end date and frequency-

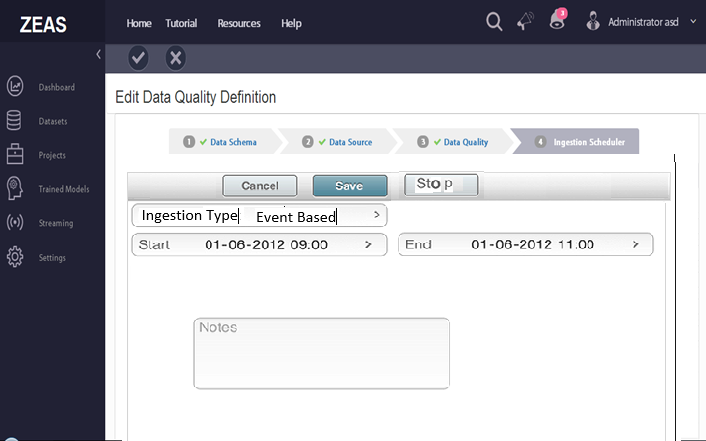
* Scheduled dataset name will display clock icon next to dataset name.
* On clock click, will open popup where user can kill scheduler, show scheduler history and reschedule.
* If dataset is not scheduled no clock icon will appear next to dataset name.
* User can schedule dataset Ingestion any time, Edit dataset will allow user to schedule.

## Detailed functionality

1. While creating profile provide one more step to schedule Ingestion, save scheduler and stop scheduler can be handle from here.

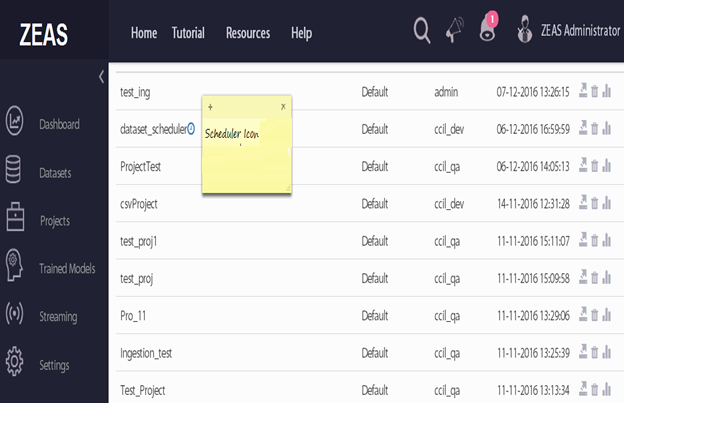


Screen 1 - Time Based scheduler



Screen-2 Event Based scheduler

1. Once dataset created with scheduler clock icon should display next to scheduler name.



Screen- 3

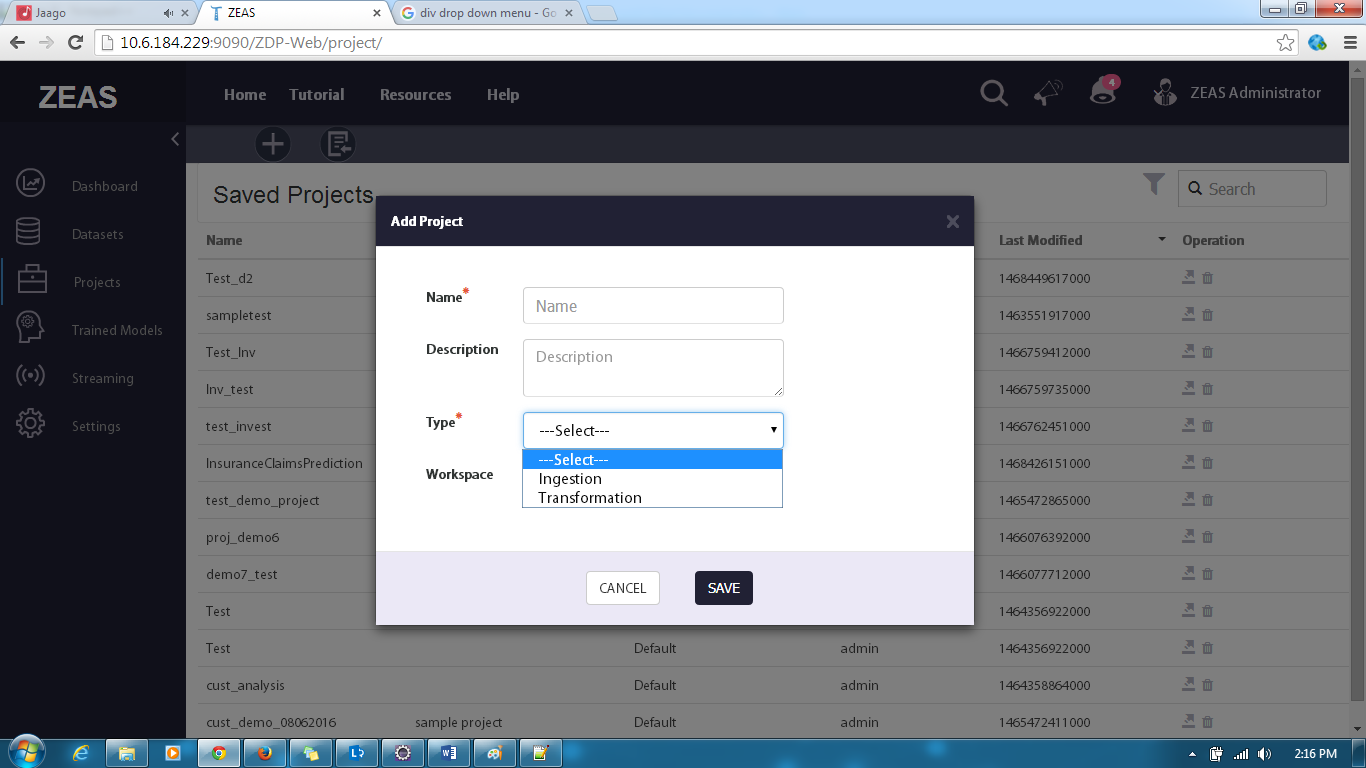
1. Display scheduler details on mouse hover on clock (start, end and frequency).
2. On clock click display run history and stop scheduler option.
3. User can schedule dataset any time on edit dataset option.

# Option 2

* We can create two types of projects – Ingestion and Transformation.
* When user selects Ingestion Type project to schedule created dataset.
* Once Ingestion project type is created and scheduled dataset, clock icon will appear next to project name.
* On Clock click, basic scheduler information will pop-ups (start date, end date, frequency).
* User can see scheduler history and kill scheduler after opening.

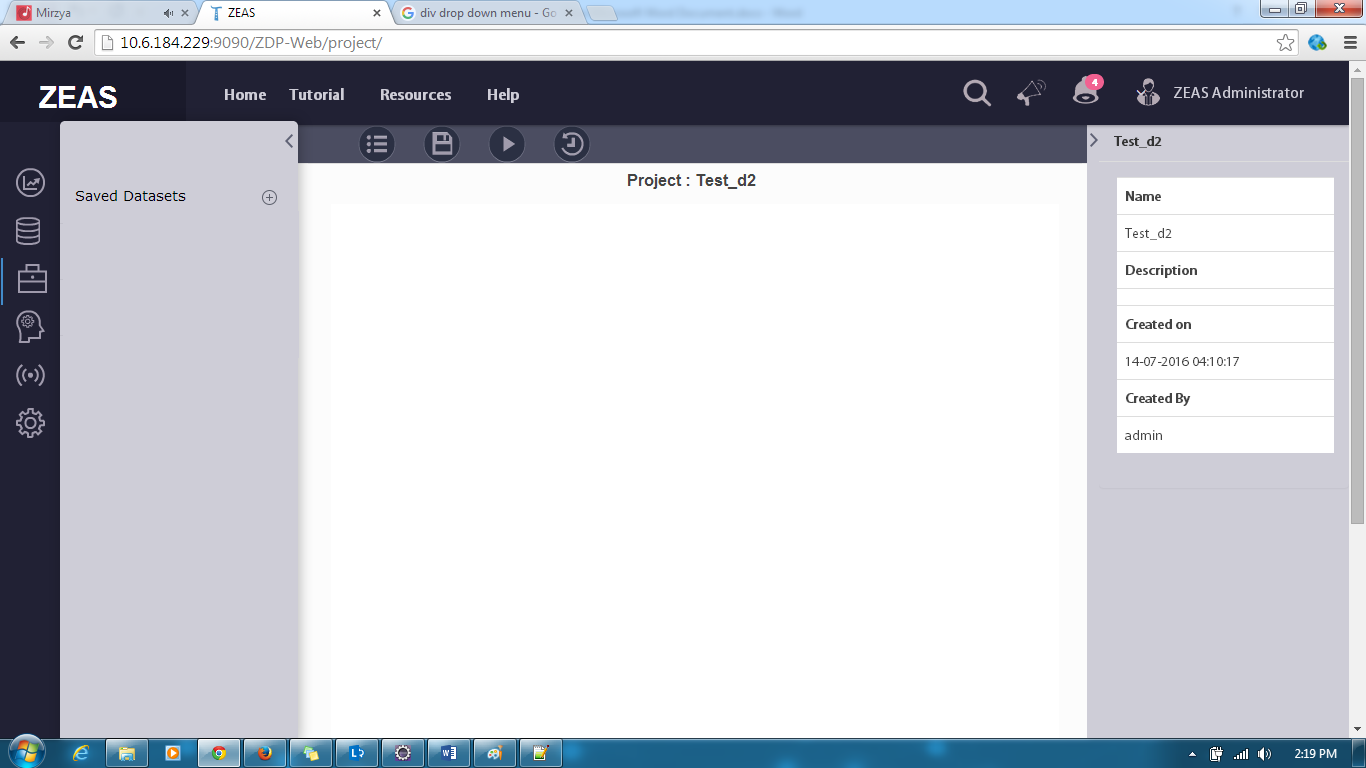
## Detailed functionality

1. Select project type “Ingestion” or “Transformation”. Ingestion will navigate to new page which will give scheduler feature and Transformation will continue to old Page.



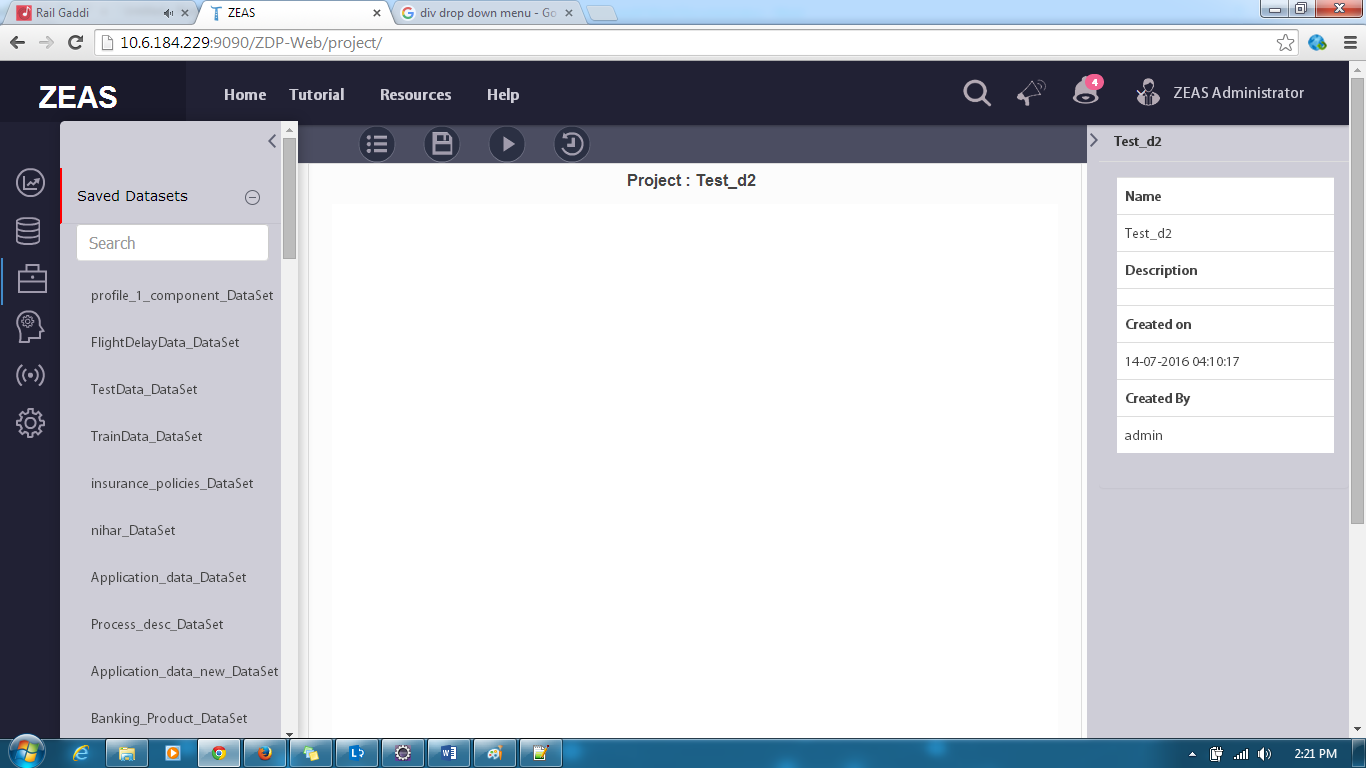
Screen-4

1. If Type of Project is “Ingestion” display only saved dataset.



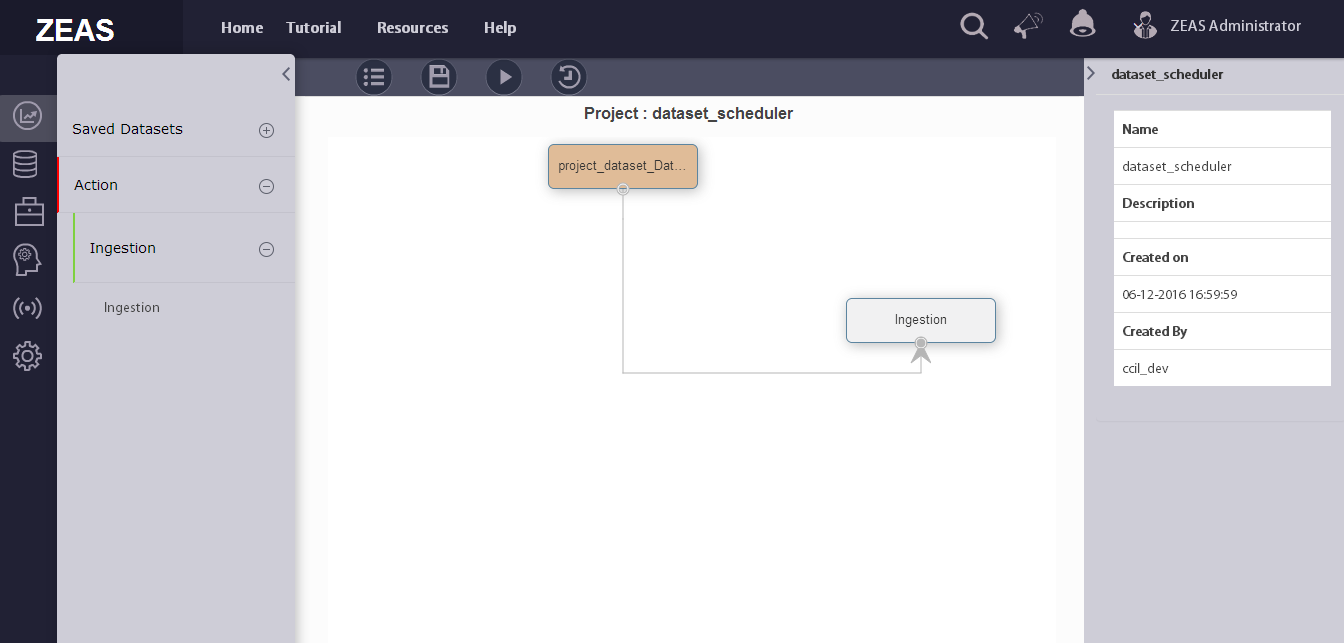
Screen-5

1. List all saved dataset –



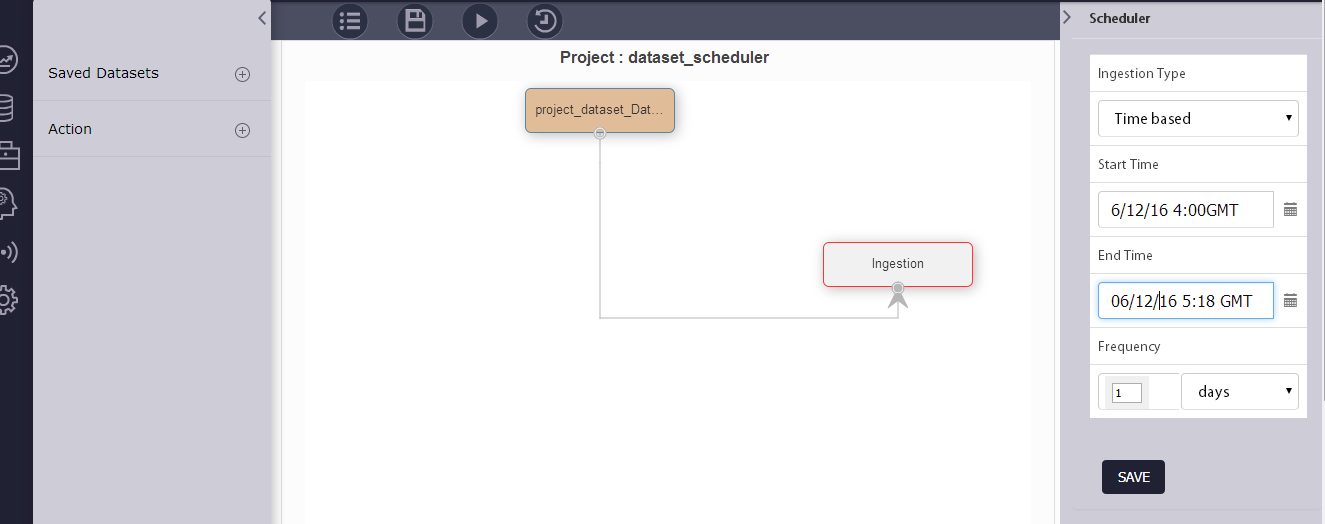
Screen -6

1. Single action “Ingestion” is available



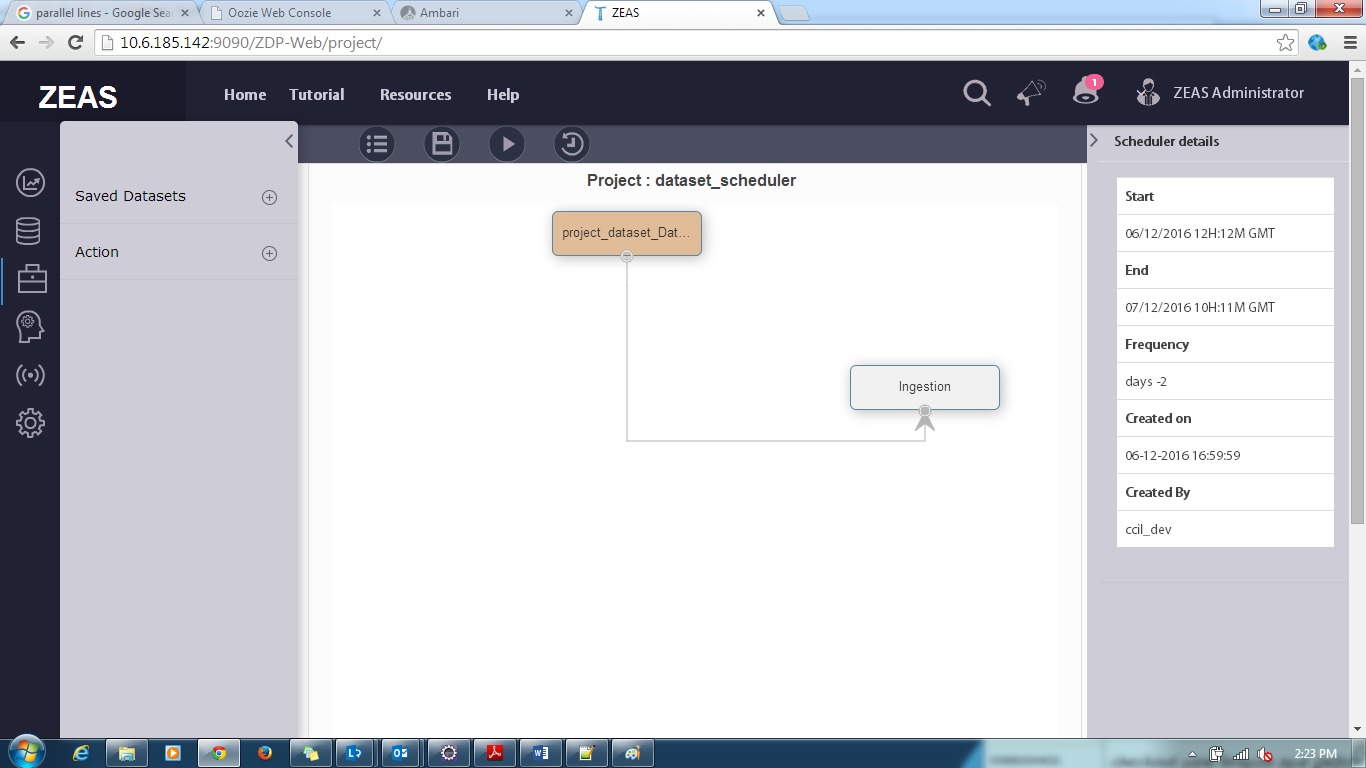
Screen -7

1. After clicking on Ingestion it will open scheduler page at right panel.



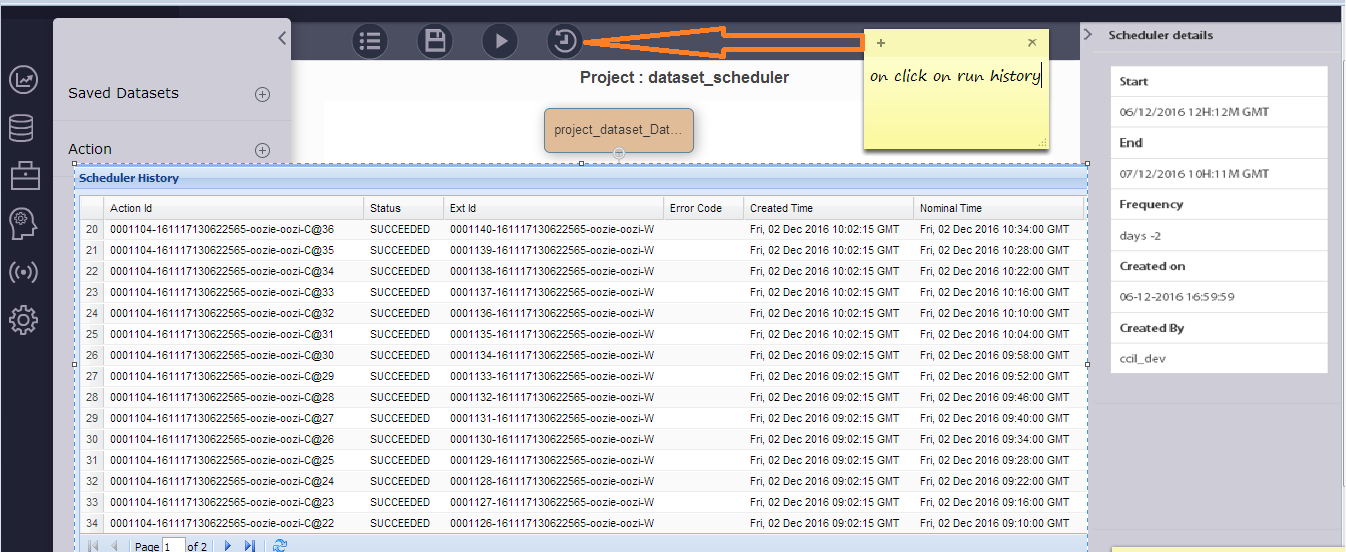
Screen -8

1. Click on ‘Save’ button scheduler details will display on right side panel.



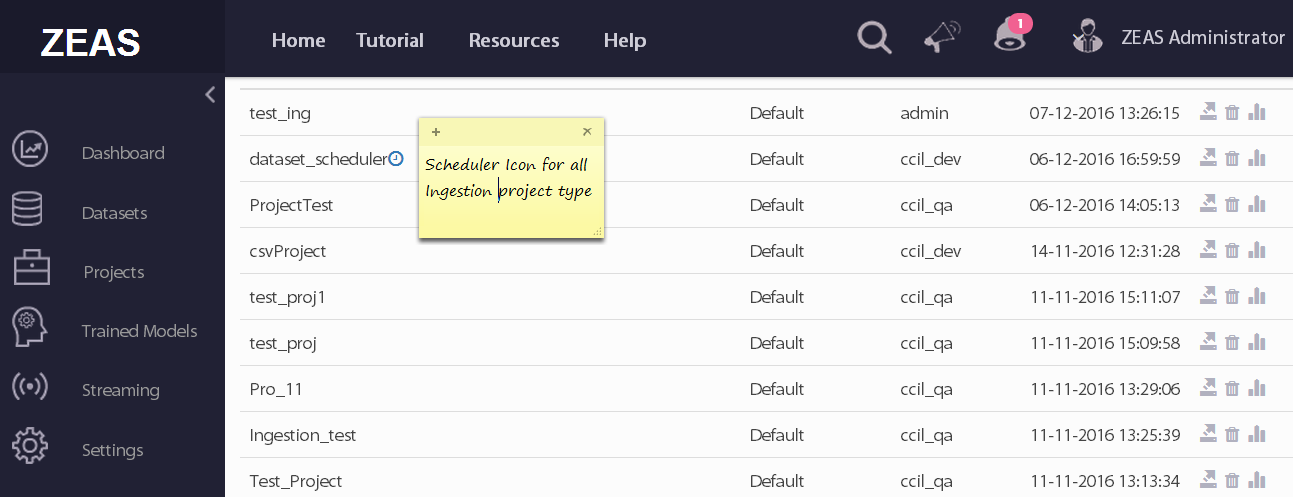
Screen -9

1. To start scheduler click on run project play icon  which will turn to square icon  and it will remain same till scheduler is active.
2. To see scheduler run history click on  it will display below page



Screen -10

1. To identify Ingestion (scheduled) and transformation project will add clock icon next to project name with hover scheduler details (active, expired or killed)



Screen -11

# Assumptions and Risks

1. Lack of dedicated resources.
2. No UI Developer
3. In case any requirement change the delivery time line will be impacted.