



Addis Ababa University  
አዲስአበባ ዩኒቨርሲቲ

SEEK WISDOM, ELEVATE YOUR INTELLECT AND SERVE HUMANITY !



## Big Data Analysis and Visualization

Name Of Group

1. Ababa Lata
2. Wasihun Geletu
3. Walfana Magarsa

To: Addisu Semie(PHD)



# Agenda

1. Introduction to Talend
2. History of Talend
3. Talend Products
4. Talend Open Studio
5. Benefits Talend for Open Studio
6. Open-Studio Architecture
7. How Talend Open Studio works



# Introduction Talend



## What is Talend





# Introduction of Talend



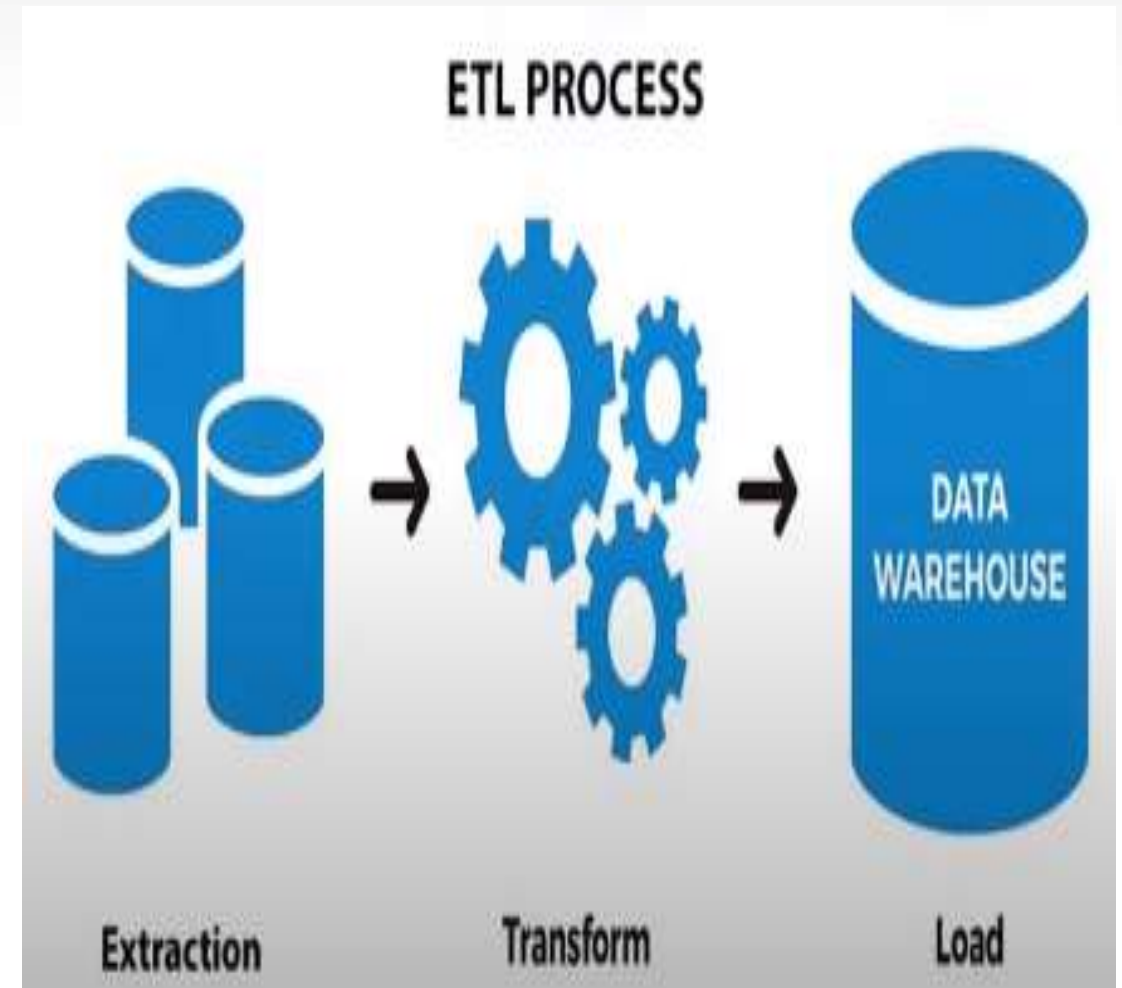
- Talend is an open source software platform which offers data integration and data management solutions.
- Talend specializes in the big data integration. This tool provides features like



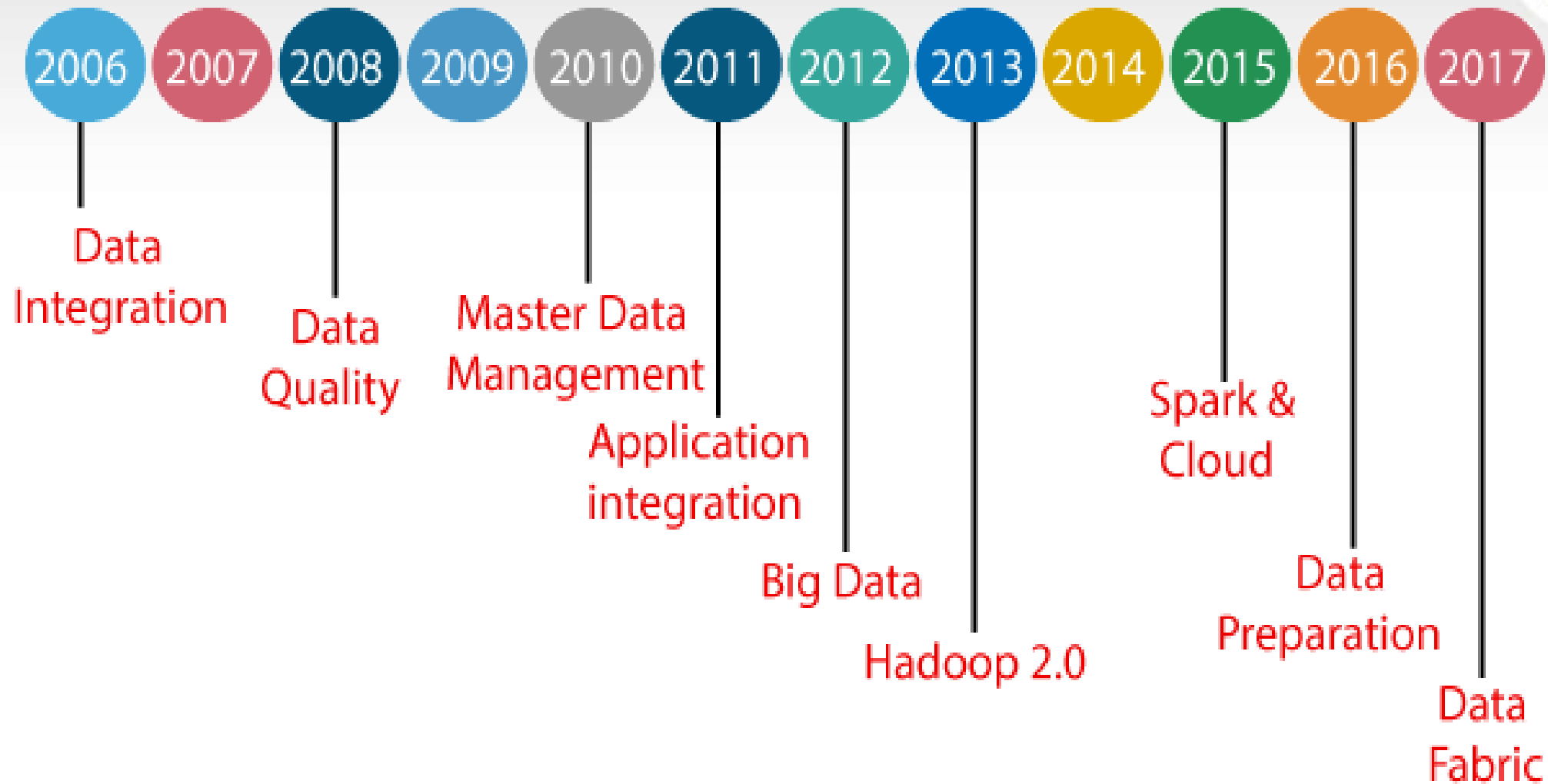
# Introduction of Talend



- ✓ Talend is an ETL tool that provides solutions for
  - ✓ data integration,
  - ✓ data quality,
  - ✓ data preparation,
  - ✓ big data, and application integration.
- ✓ The big data and data integration tools of Talend are extensively used.
- ✓ Talend tool is available in both open-source and premium versions.



# History of Talend

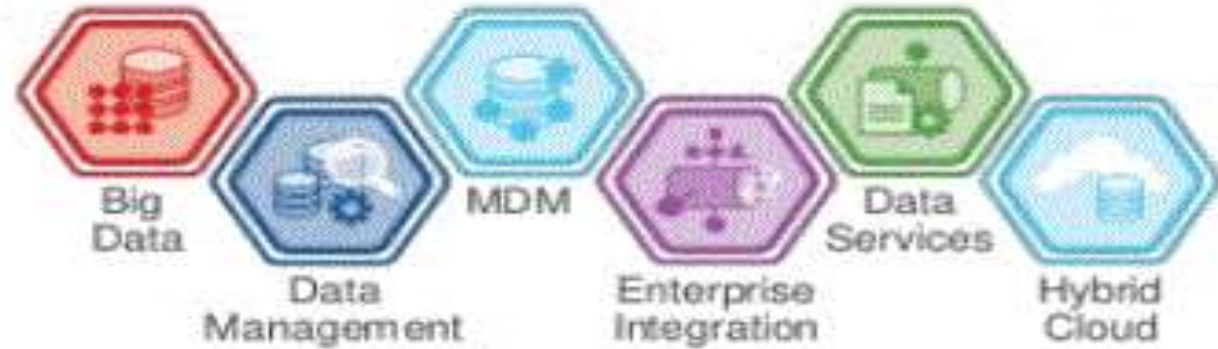


# Talend Products

There are various products of Talend and Their features



## Talend Platforms



## Talend Enterprise



## Talend Open Studio



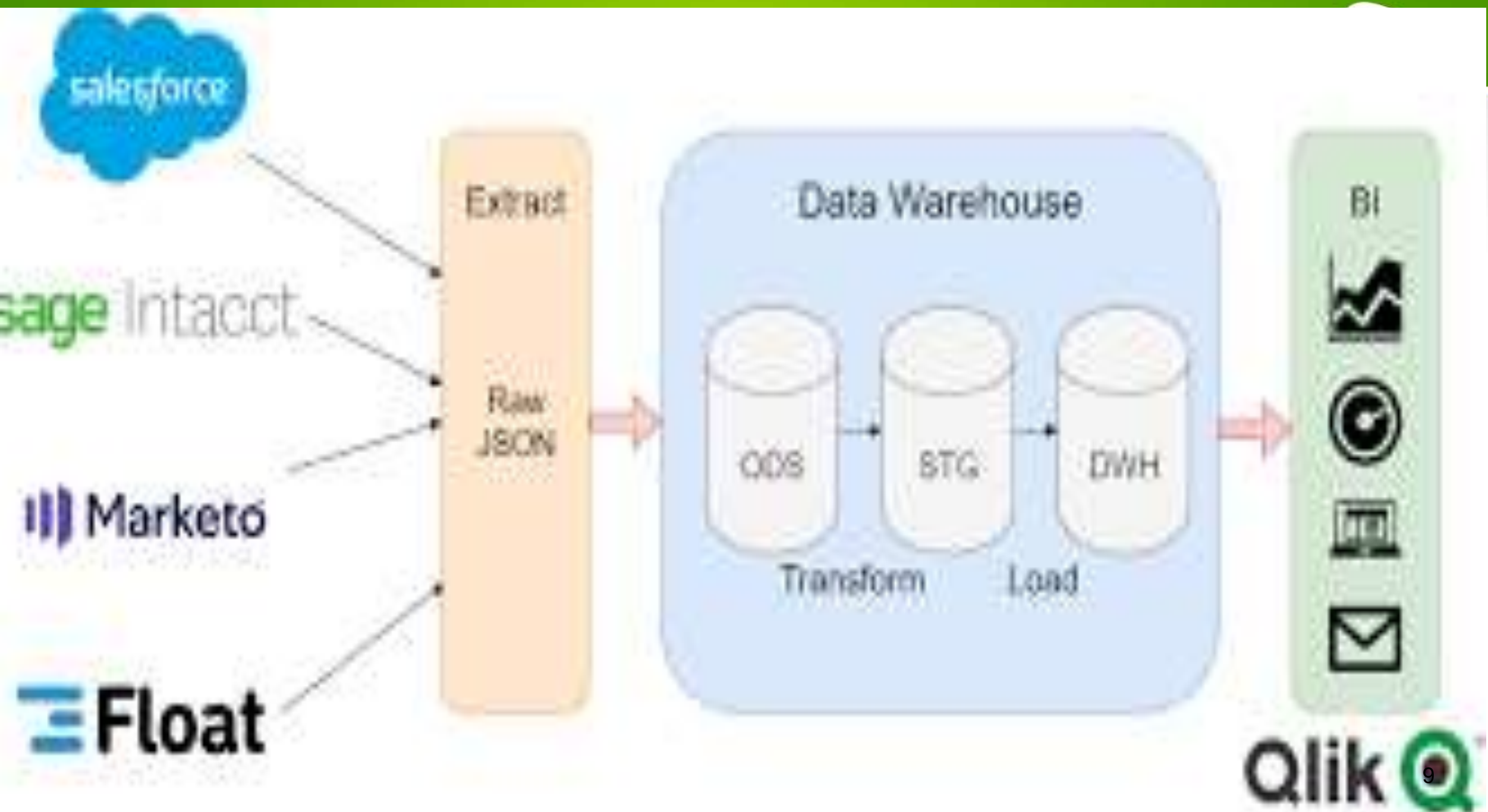


# Talend Open Studio



- Talend Open Studio is a free open source ETL tool for
  - ✓ Data Integration and
  - ✓ Big Data.
- It is an Eclipse based developer tool and job designer.
- The tool will create the Java code for the job automatically and you need not write a single line of code





# Form Any source of data to Data Warehouse

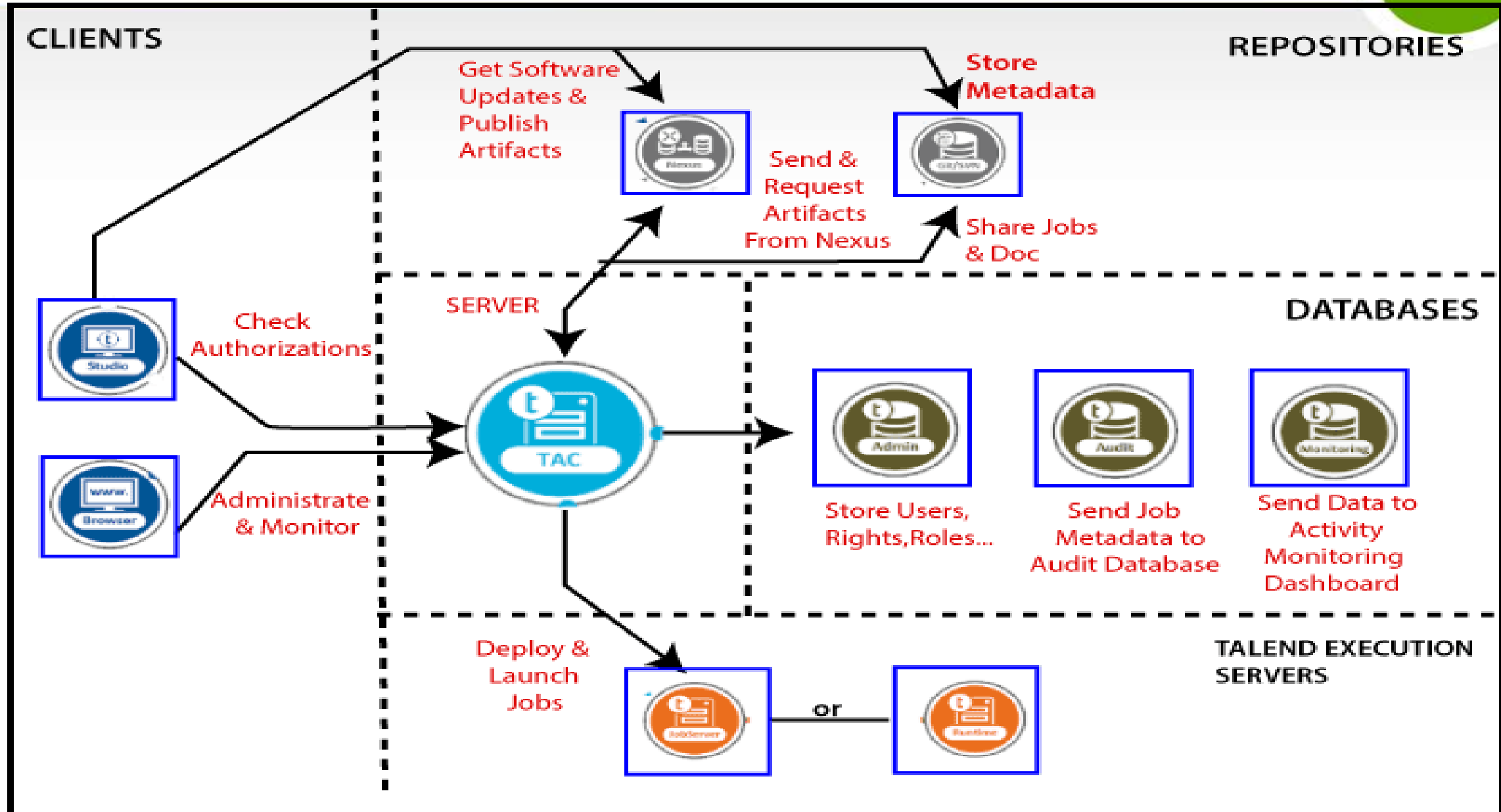


# Some important benefits which Talend Open Studio offers are as below



- Provides all features needed for data integration and synchronization with:
  - ✓ 900 components,
  - ✓ built-in connectors,
  - ✓ converting jobs to Java code automatically.
- The tool is completely free, hence there are big cost savings
- the documents are well Structured
- very easy to follow
- The talend community for data integration is very active.

# Talend Data Integration Architecture





# Functional Blocks of Data Integration



- Functional block of data Integration Architecture are used for managing data Integration activities. like

## Clients:

- ✓ is used for building and monitoring Talend Jobs.
- ✓ have one or more Talend studio(s), and the web browser that could be on the same or different machines.
- ✓ The Talend studio allows us to work on any project if we have the authorization.



- ✓ We can connect the remotely based Talend administration center through a secured HTTP protocol with the help of a web browser.
- ✓ We can also carry out the data integration process regardless of the level of data volumes and process the complexity from the studio.



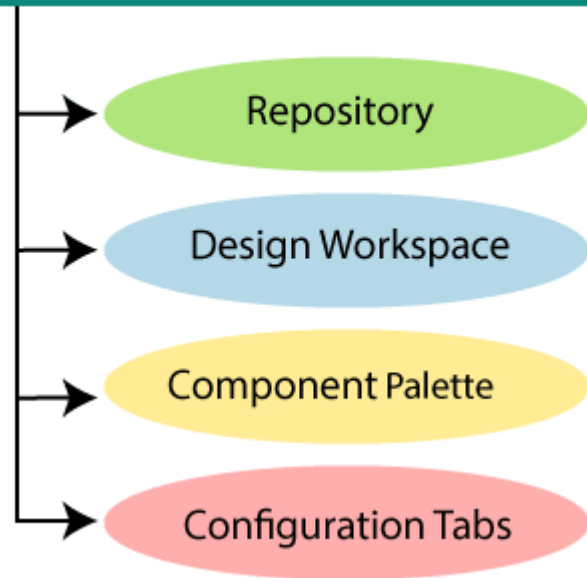
- **Server:**

- ✓ is used for administration, management, and monitoring.
- ✓ contains the web-based application server that connects to
  - ✓ Two shared repositories
  - ✓ Three databases
  - ✓ Excursion Servers for Talend
- ✓ Talend Administration Center is used to enable the management and administration for all the projects
- ✓ The administration metadata is stored in the administration database.
- ✓ For Example: the user accounts, access rights, and project authorization.
- ✓ The data of project items like jobs, business models, and routines are stored in the SVN or a Git server.

# How Talend OpenStudio works



## Features of Talend open Studio



- **Repository:-** used to design jobs, create jobs or manage metadata. Metadata contains many sources, for example, DB connections, different kinds of files like Azure, LDAP, Marketo, Salesforce, web services, Hadoop cluster, and FTP.

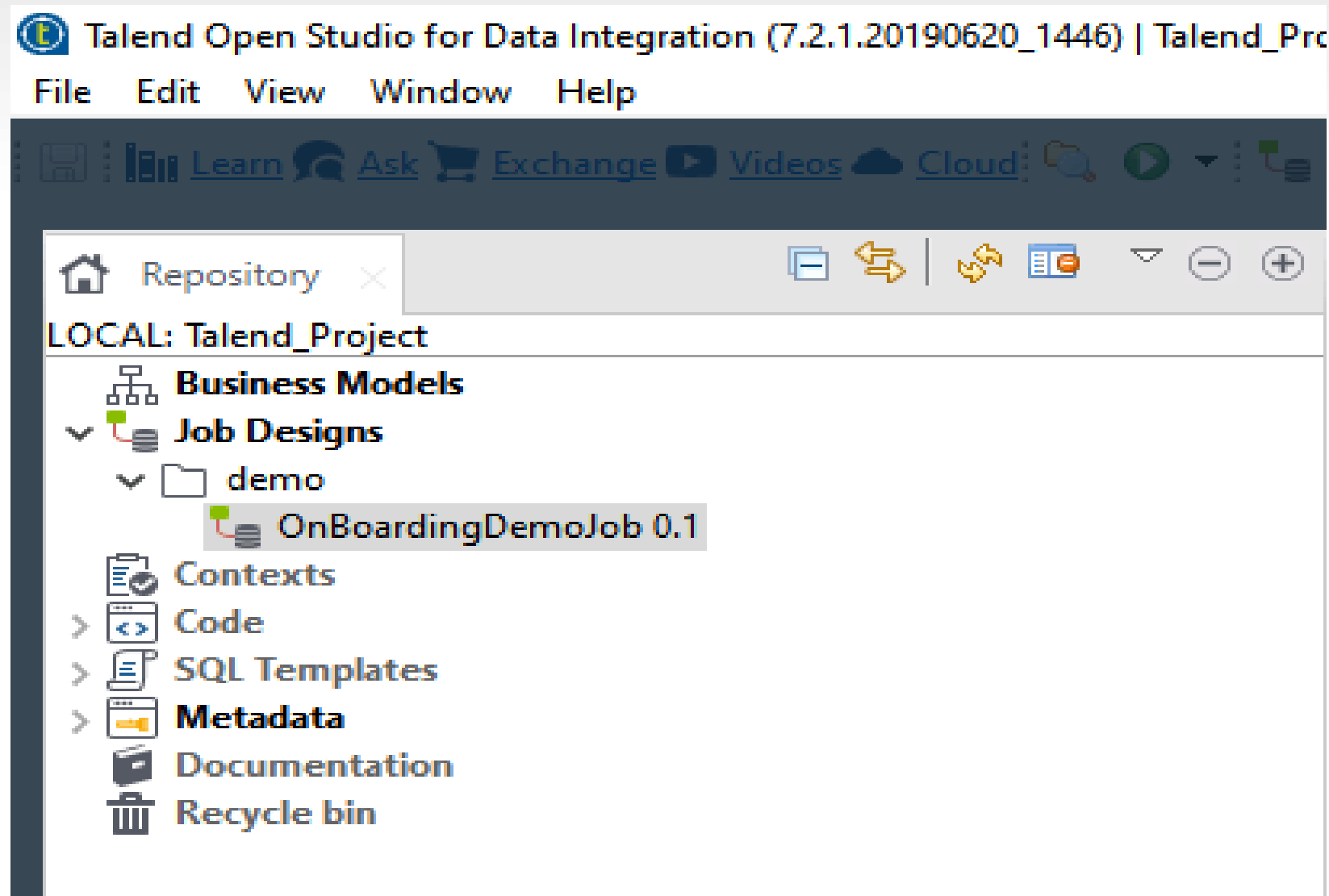
**Design Workspace:-** Contains two fields designer tab and code tab .

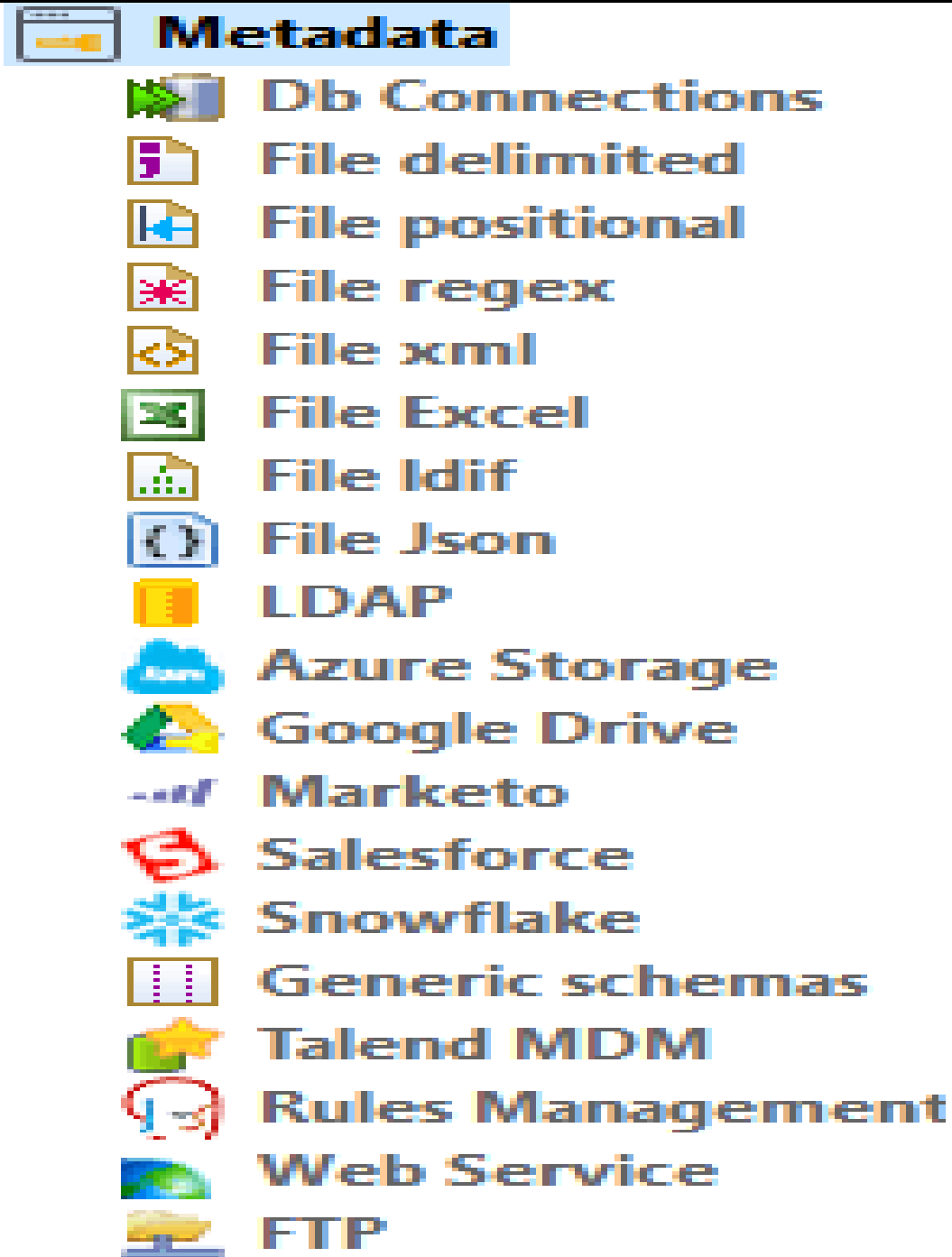
**The component palette :-** is used to perform the specific data integration operation because it is a preconfigured connector.

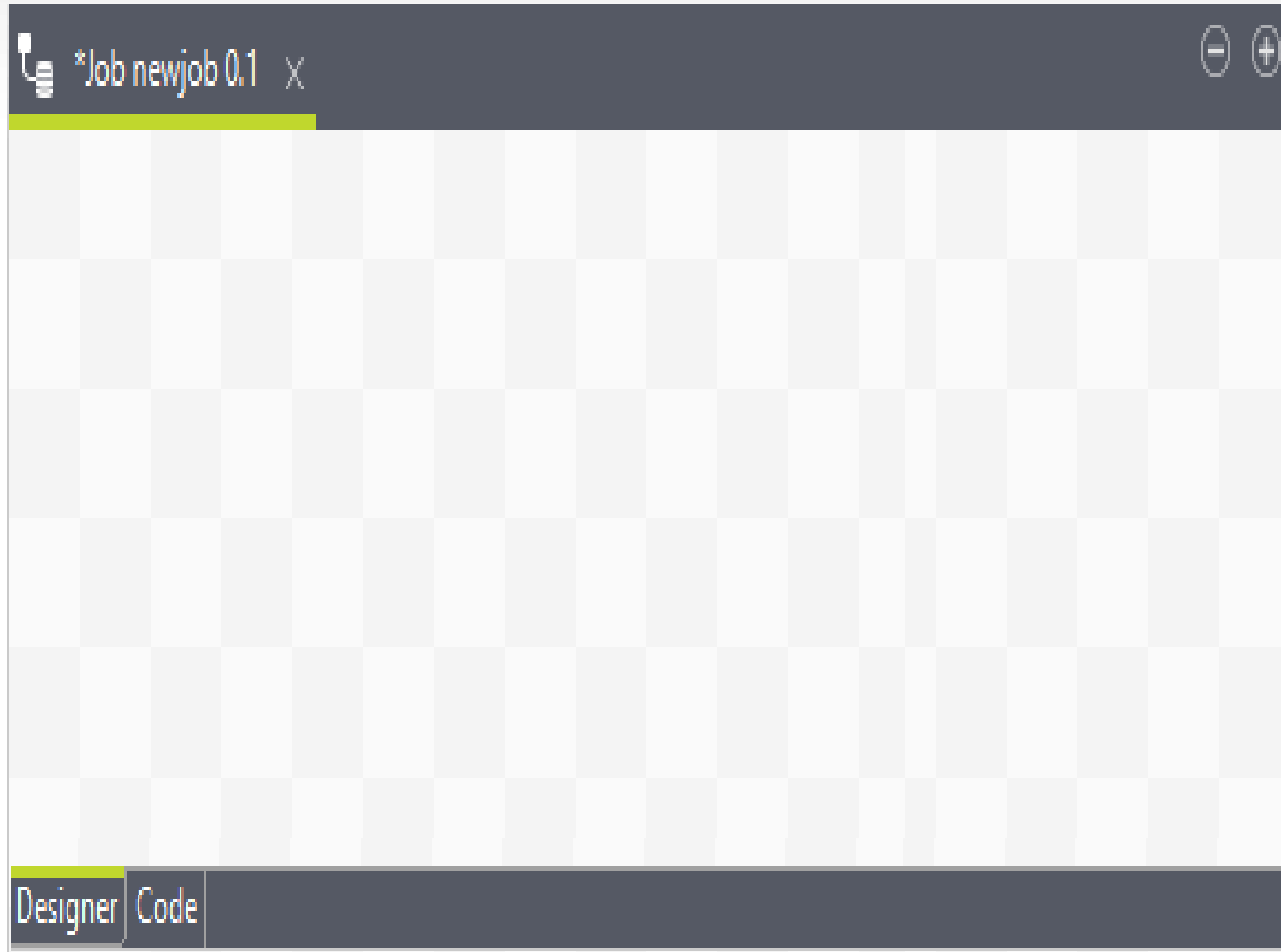
- **The configuration tab::-** is used to display and edit the properties of specific components within the design workspace.



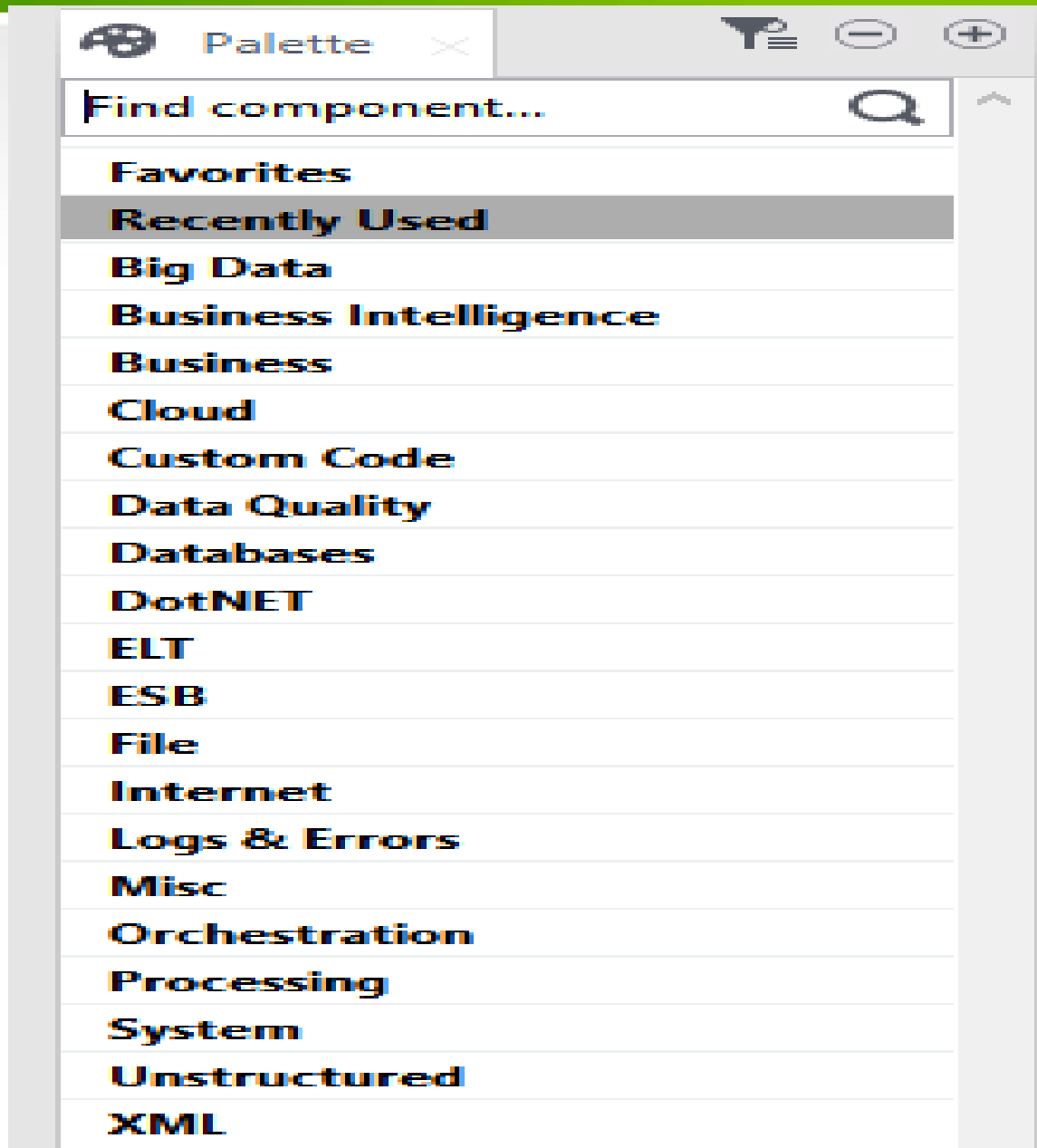
we can see in the below screenshot







# Components of Palette





# ConfigurationTab



Designer

Code

Job(OnBoardingDemoJob... x

Contexts(OnBoardingDe...

Component

Run (Job OnBoardingDe...

-

+

OnBoardingDemoJob 0.1

Main	Name	OnBoardingDemoJob		
Extra	Author	user@talend.com	Version	0.1
Stats & Logs				M m
Version	Creation	11/21/19 3:59 PM	Modification	11/21/19 6:29 PM
	Purpose	Used for on-boarding presentation	Status	
	Description	A simple row		

Activate W

Go to Settings

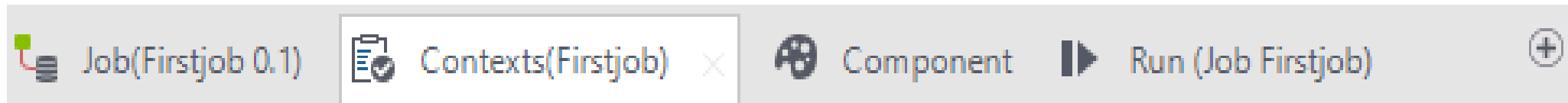
# Description of Configuration Tab



- **The job tab:** is used to give information about the current job in the design workspace window, which has
  - ✓ name,
  - ✓ version,
  - ✓ creation,
  - ✓ purpose,
  - ✓ description,
  - ✓ status, and so on.
- **Context variables:** is used to set the different values in the various platforms, and we can also create a context group where we can hold multiple context variables.



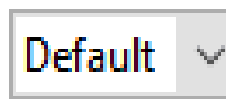
- It is also used to move the code into the development
- For any job, we can go to the context tab, and add the context variables, as shown in the below snapshot.



	Name	Type	Comment	Default	
				Value	



Default context environment



# Cont>>>



- **Component tab:** is displays all the parameters that are requested to configure the components



## Subjob

<b>Basic settings</b>	<div data-bbox="868 858 1653 1011"><b>Restore default color</b></div> <div data-bbox="1696 858 2479 1011"><b>Save as default color</b></div>
<b>Advanced settings</b>	<div data-bbox="828 1086 1538 1158"><input type="checkbox"/> <b>Show subjob title</b></div>
	<div data-bbox="848 1243 1294 1329"><b>Subjob color</b></div> <div data-bbox="1493 1218 1860 1379"></div>





- **Run tab:** shows the progress of the execution of a job, and the log pane displays the starting, and ending of the error messages as we can see in below image.

The screenshot shows the 'Run' tab for 'Job Firstjob'. The toolbar includes a 'Run' button (highlighted with a dashed border), a 'Kill' button, and a 'Clear' button. The log pane displays the following text:

```
Starting job Firstjob at 16:30 23/11/2023
[statistics] connecting to socket on p
[statistics] connected
101|ram|Ny|It
102|john|Rw|Sales
103|harvey|usa|accounts
104|wikel|usa|accounts
```

To the right of the log pane is a table with a 'Name' column and a 'Default' dropdown menu. The table is currently empty.



**Thank You**