Tech Mahindra

Talend

Introduction

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend Open Studio

Big Data

Talend Open Studio for Big Data combines big data technologies into a unified open source environment simplifying the loading, extraction, transformation and processing of large and diverse data sets. Users are also presented with a full palette of components for NoSQL connectivity, all under an open source Apache license.

Learn More Download Now

BPM

Bonita Open Solution from Talend's partner Bonitasoft, is a robust, open source, business process management (BPM) solution for modeling, creating and optimizing business processes.

Learn More Download Now

Data Integration

Talend Open Studio for Data Integration provides a set of data integration tools to access, transform and integrate data from any business system in real time or batch to meet both operational and analytical data integration needs.

Learn More Download Now

Data Quality

Talend Open Studio for Data Quality delivers end-to-end profiling and monitoring capabilities with the ability to identify anomalies, find relationships in data, and ensure data quality is fit-for-use within your enterprise application

Learn More Download Now

ESB

Talend ESB is a reliable and scalable enterprise service bus (ESB) for the creation, connection, mediation and management of services and their interactions.

Learn More Download Now

MDM

Talend Open Studio for MDM provides master data management, data stewardship, data integration and data quality in a single platform.

Learn More Download Now

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

What Talend Open Studio is?

- An open source graphical development environment
- It comes over 800 pre-built connectors
- It is quick and easy to connect
  - databases, transform files, load data, move, copy and rename files, and
  - connect individual components

  for complex integration processes.

- It is a code generator
- Is easy to use and reduces the time taken to develop integrations

- Integration jobs are created from components that are configured rather than coded and jobs can be run from within the development environment or executed as standalone scripts.

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Use cases

- Some common use cases for Talend Open Studio for Data Integration include:
  - Data migration from one database to another
  - Regular file exchanges between systems
  - Data synchronization
  - ETL (Extract, Transform, and Load)

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

History of Talend Open Studio

- Talend was founded in 2005 and is an open source software vendor providing solutions for data integration, data quality, master data management, enterprise service bus, and business process management.

- Talend's first product, Talend Open Studio for Data Integration, was launched in 2006, under the name Talend Open Studio, and has since been downloaded over 20 million times.

- Talend has continued to develop its product portfolio and has added complementary tools that provide a single platform for application, data, and process integration.

- The Talend Open Studio brand has since been adopted across the range of Talend's products.

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Benefits of Talend Open Studio

- The Studio is open source, free to download and use, with access to the source code, allowing users to extend the product to their particular needs if required.
- The Studio is a great productivity-booster. It's easy to learn and quick to develop with. Even novice developers will be building complex integrations in no time.
- The Studio's pre-built components handle many common and not-so-common tasks. Developers can focus on the end-to-end process, rather than the low-level technical details.
- Talend has an active and open user community. Practical, problem-solving advice is e

TechMahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Working with Talend Open Studio

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Studio definitions

- Few definitions:
  - A ***repository*** is the storage location Talend Studio uses to gather data related to all of the technical items that you use either to describe business models or to design Jobs
  - A ***workspace*** is a directory on your computer that contains one or more projects
  - A ***project*** is a logical grouping of one or more jobs
  - A ***job*** is a group or one or more components that, when executed, implement a data flow or integration process
  - A ***component*** is a preconfigured connector used to perform a specific data integration operation
  - An ***item*** is the fundamental technical unit in a project. Items are grouped, according to their types, as: Job Design, Business model, Context, Code, Metadata, etc. One item can include other items.

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

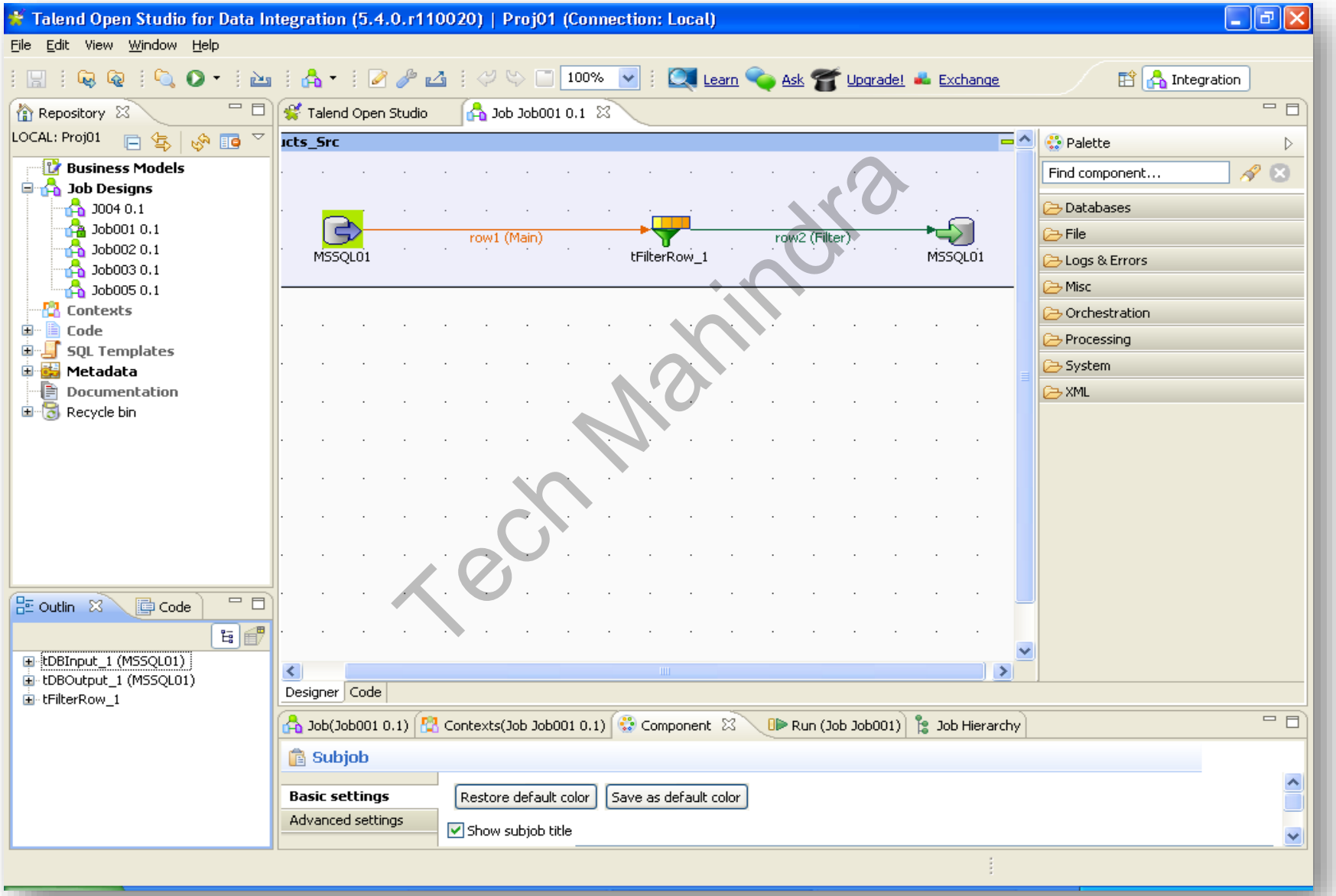
Starting the Studio

1. When you start the Studio for the first time, click on **Accept** license notification to proceed. We can now:
   - Import a demo project
   - Create a new project
   - Change some basic settings
2. We will start by amending some settings. Click on **Advanced**. You will see the following screen:
3. First change the workspace location. Click on the **Change** button and select an appropriate path, such as **C:\Talend\Workspace.**
4. Click on the **Restart** button now.
5. Import the demo project provided with the Studio. Select the default demo project, TALENDDEMOSJAVA and click on Import.
6. Enter name, such as **DEMOPROJECT**, and click on **Finish**
7. Make sure the demo project is highlighted and click on the **Open** button.
8. Click on **Skip** for now. The Studio will start to load.
9. Once the Studio is open, it runs a **Generation Engine Initialization** process. We can choose to have this run in the background if you wish by clicking on the **Run in Background** button as shown in the following screenshot:
10. Click on the **Start now!** button

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

The environment

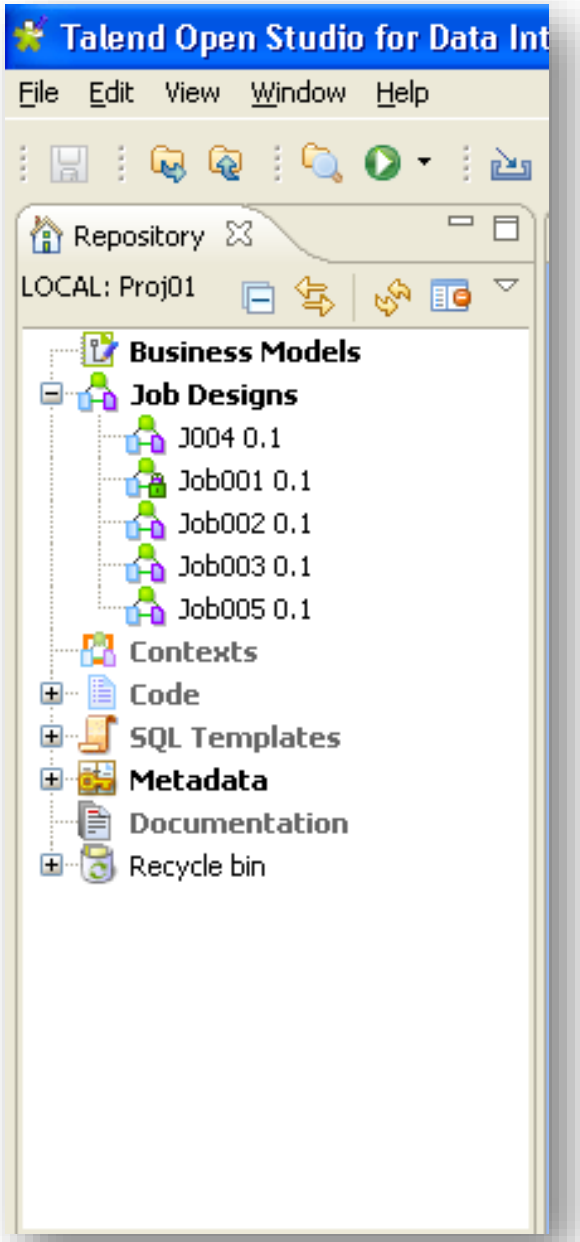


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

The environment

- **Repository Window**
  - Job Designs, Business Models, Metadata,
  - Database Connection details,
  - FTP connection details and File schema definitions
  - Reusable code snippets
  - Contexts – global or job-specific variables

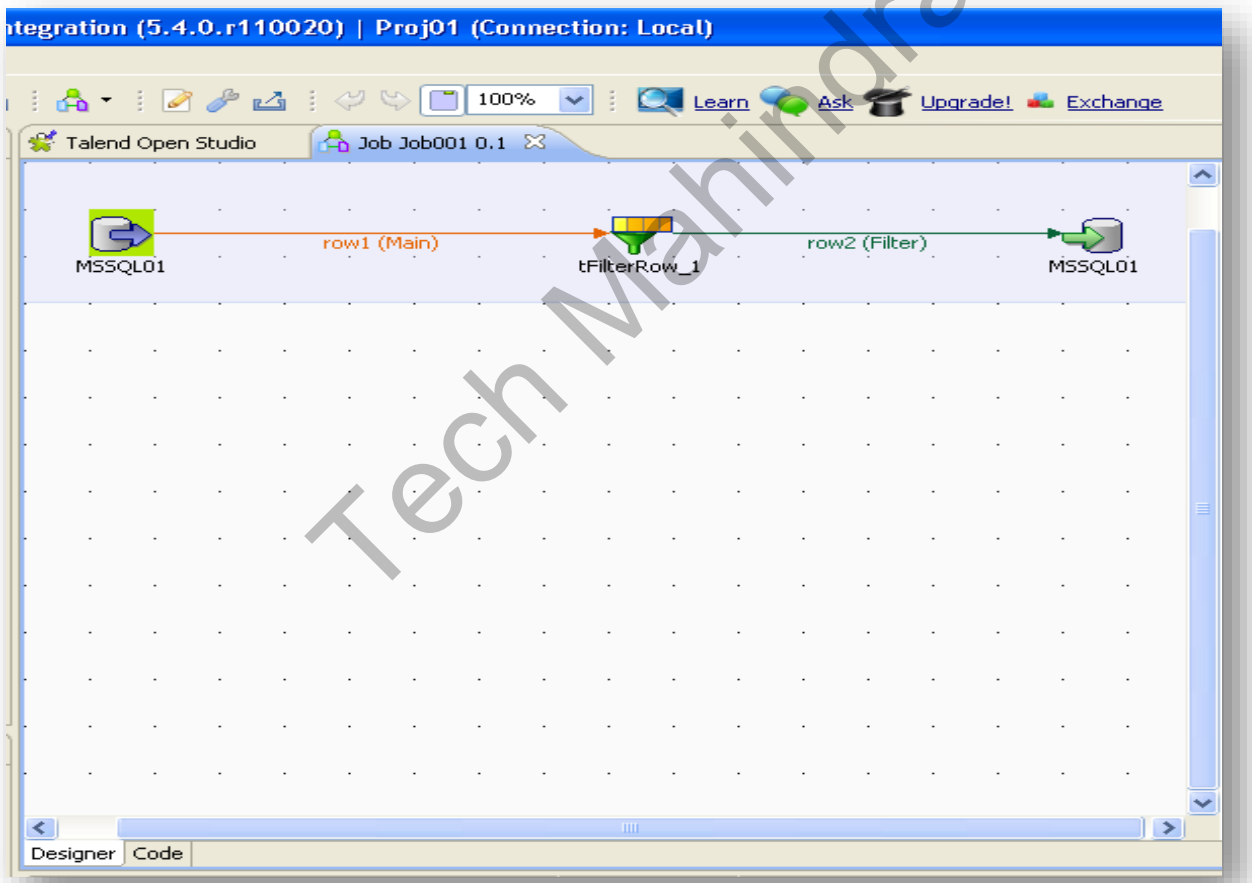


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

The environment

- **Design Workspace**
  - Developers space for designing jobs and configuring components

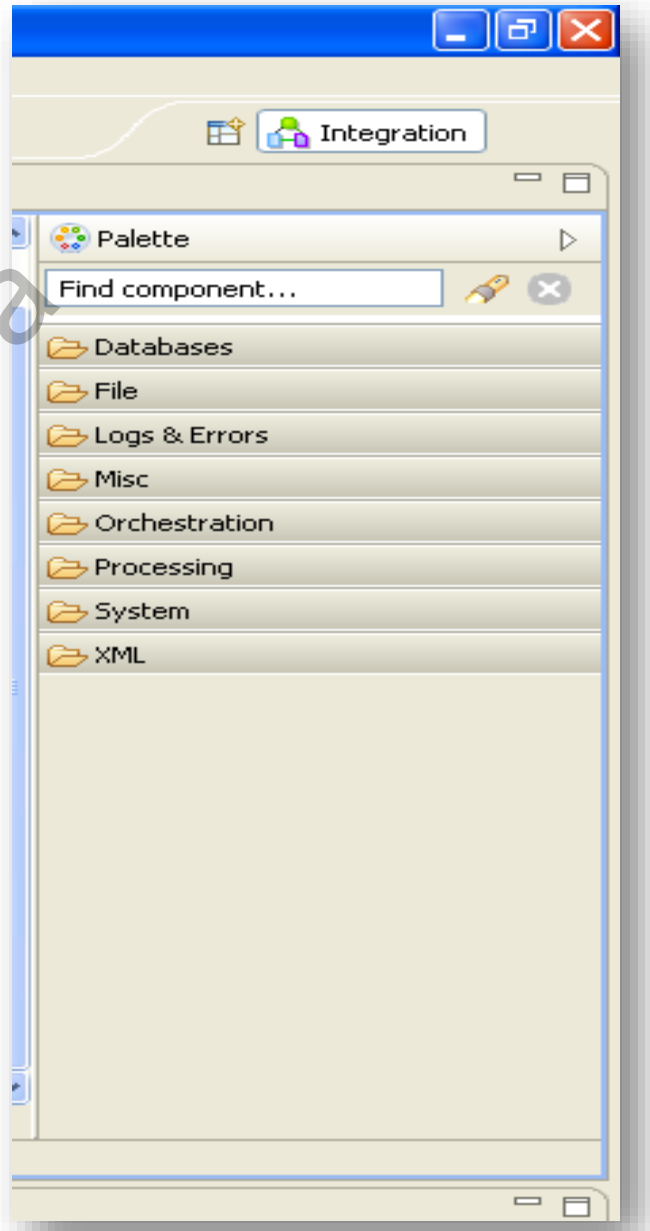


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

The environment

- **The Palette**
  - Collection of components used in data integration jobs

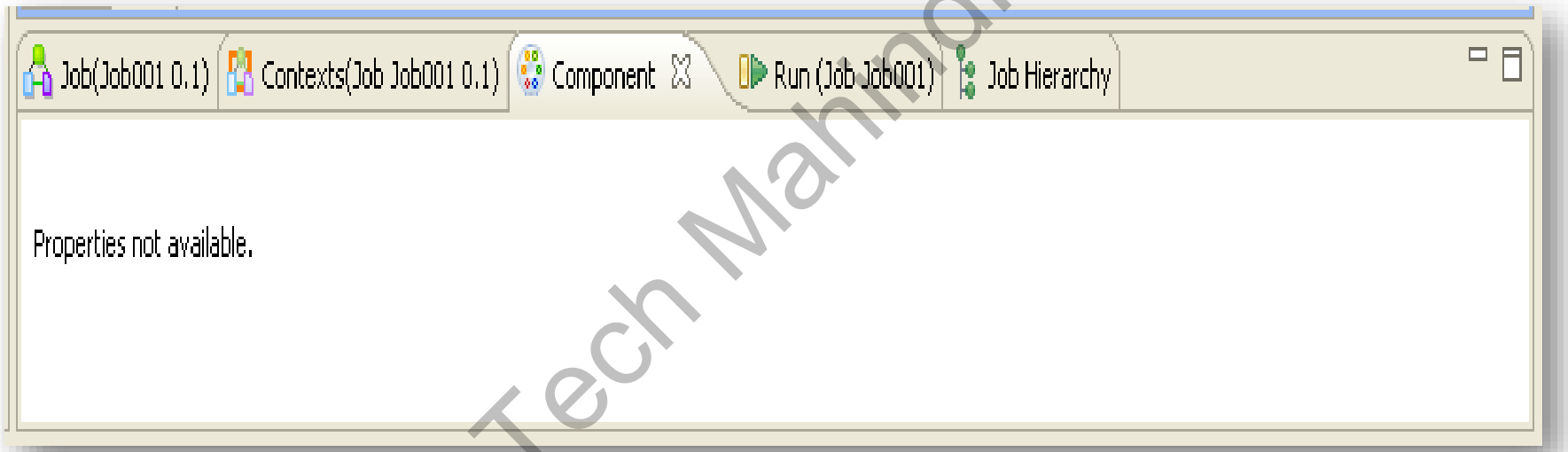


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

The environment

- **Configuration Tabs**
  - Display properties of jobs or specific components

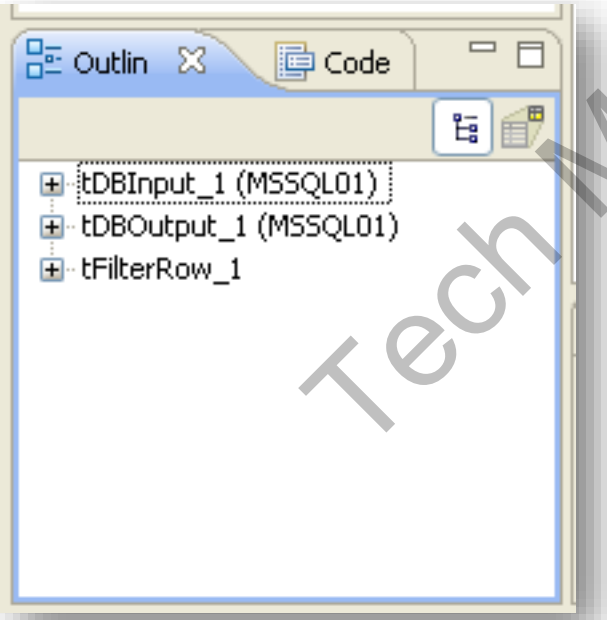


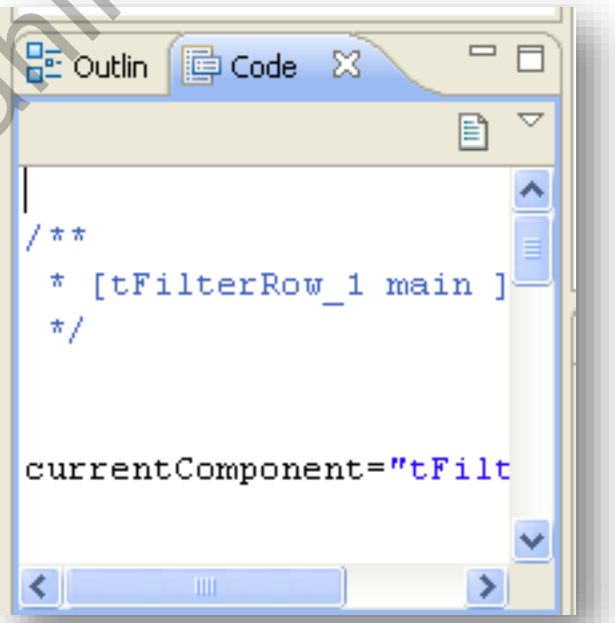
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

The environment

- **Outline and Code panels**
  - Lists the components that have been added to the design workspace
  - Gives quick access to standard variables of each component
  - Displays code associated with each component





Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

First Job – Welcome Msg

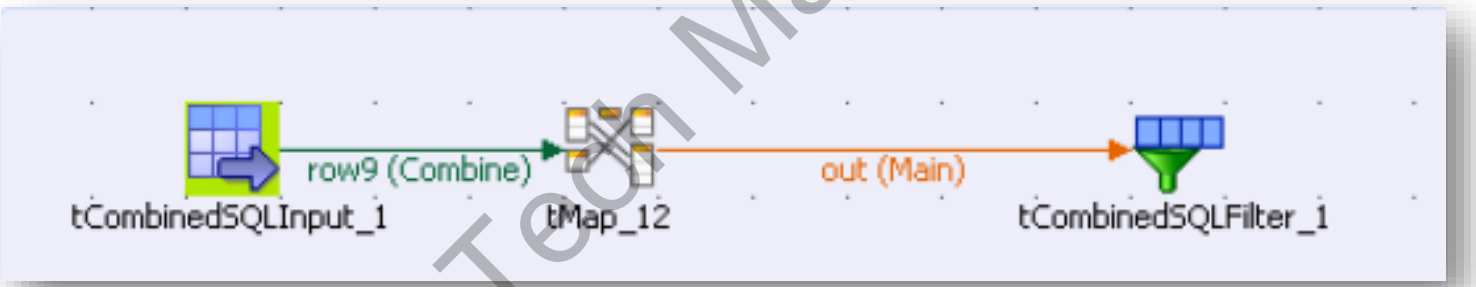
Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

What is a Job design

- A Job Design is the runnable layer of a business model.
- It is a graphical design, of one or more components connected together, that allows you to set up and run dataflow management processes.
- A Job Design translates business needs into code, routines and programs, in other words it technically implements your data flow.

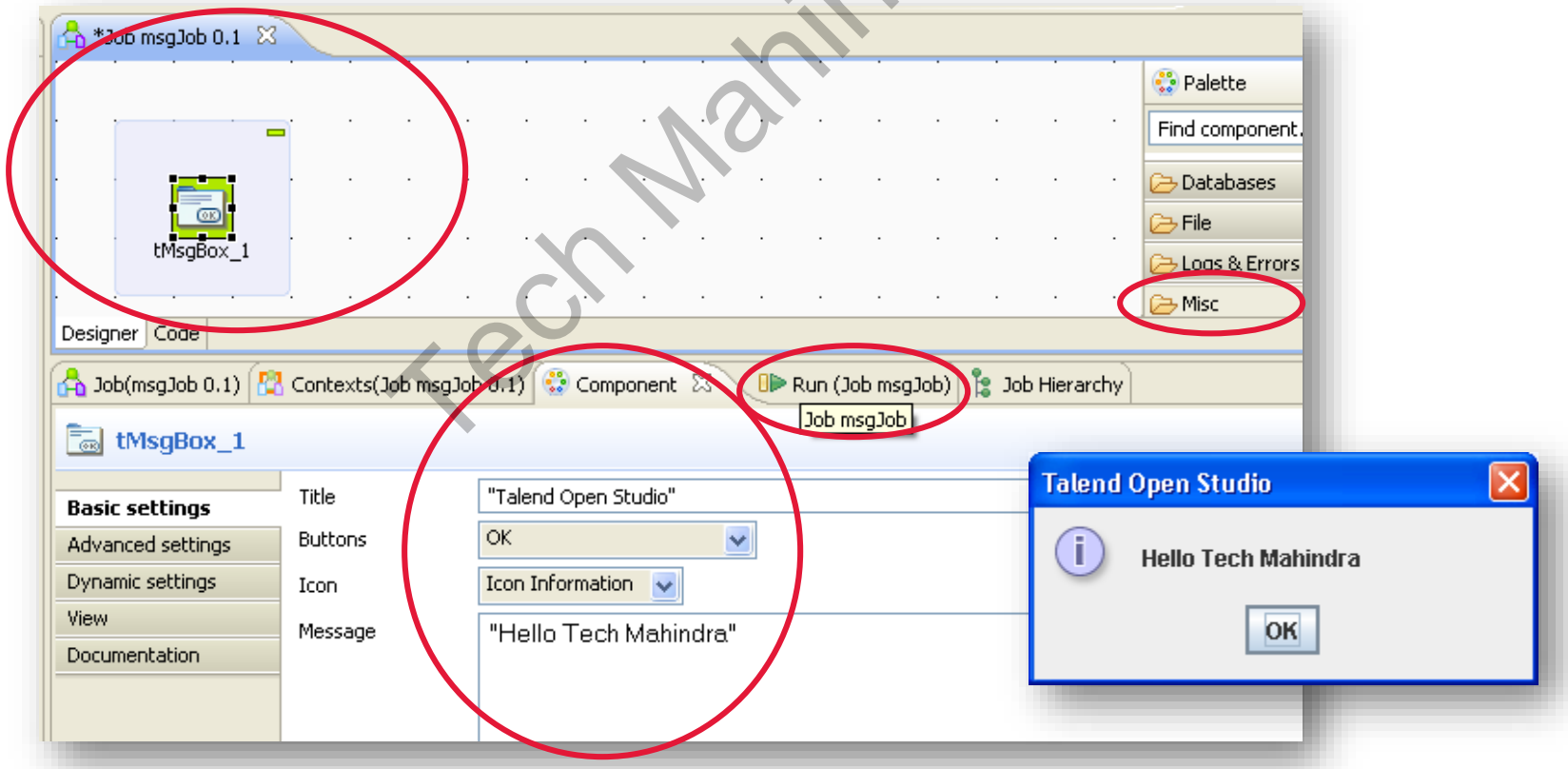


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Creating a simple Job

- In Repository, right-click on **Job Designs** and select **Create Job**.
- Enter details of **New Job** and click **Finish**.
- Drag **tMsgBox** component from **Misl** folder of **Palette** window on to the job
- Click the tMsgBox1 component and click **Components** tab, set properties.
- Save the Job and go to Run tab and click **Run** button. [F6 will also run]



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Demo

Create a simple job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Connections

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

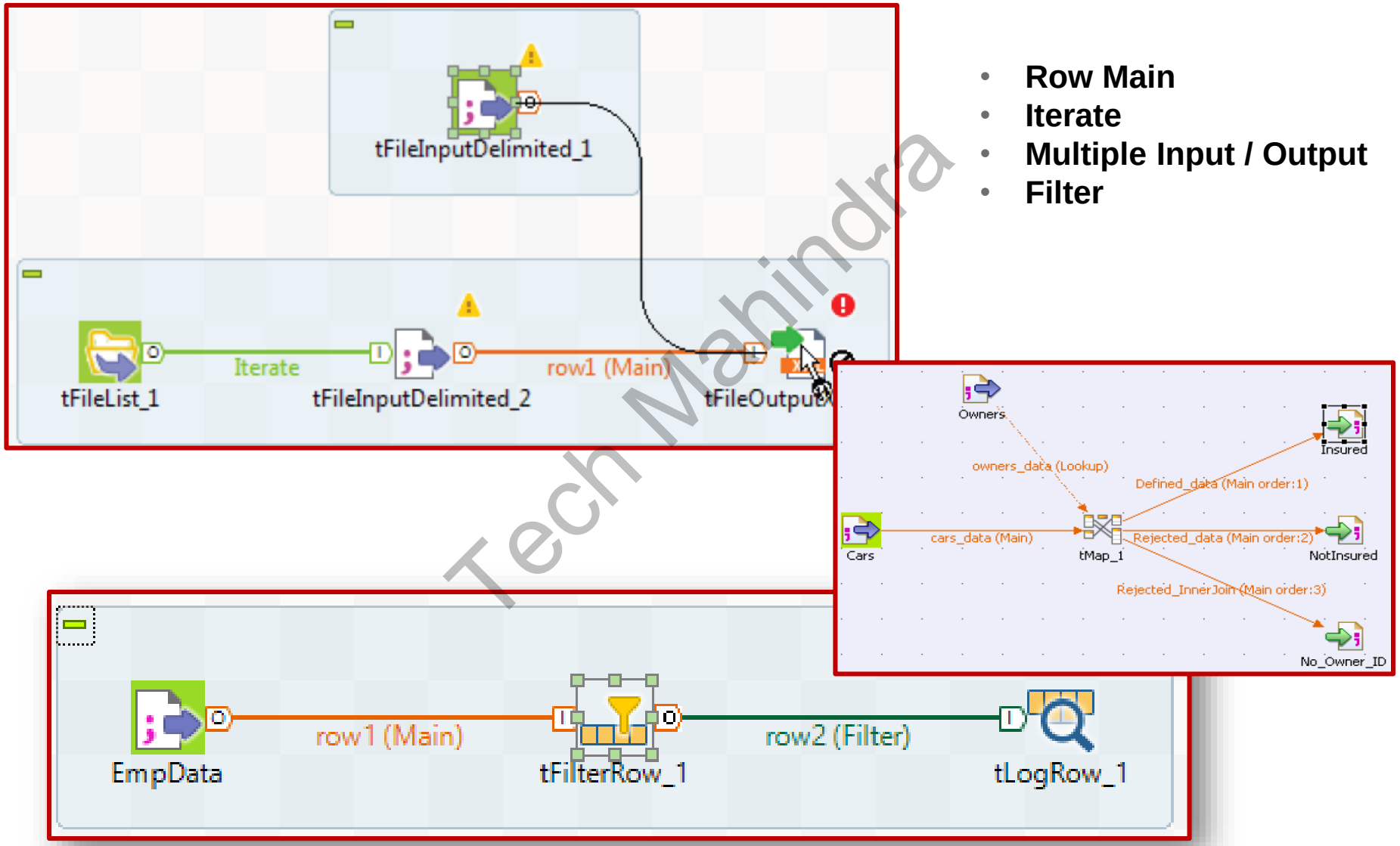
Connections

- A Job or a subjob is composed of a group of components logically linked to one another via connections.
- You need to use the connections to define how the components in use are coordinated.
- **Connection types**
  - **Row connection**
    - A **Row** connection handles the actual data.
      - Main,
      - Filter,
      - ErrorReject,
      - Unique/Duplicates,
      - Lookup,
      - Rejects,
      - Output,
      - Multiple Input/Output
  - **Iterate connection**
    - The Iterate connection can be used to loop on files contained in a directory, on rows contained in a file or on DB entries.
  - **Trigger connections**
    - Trigger connections define the processing sequence, i.e. no data is handled through these connections
  - **Link connection**
    - These links transfer table schema information to the ELT mapper component in order to be used in specific DB query statements.

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Row Connections

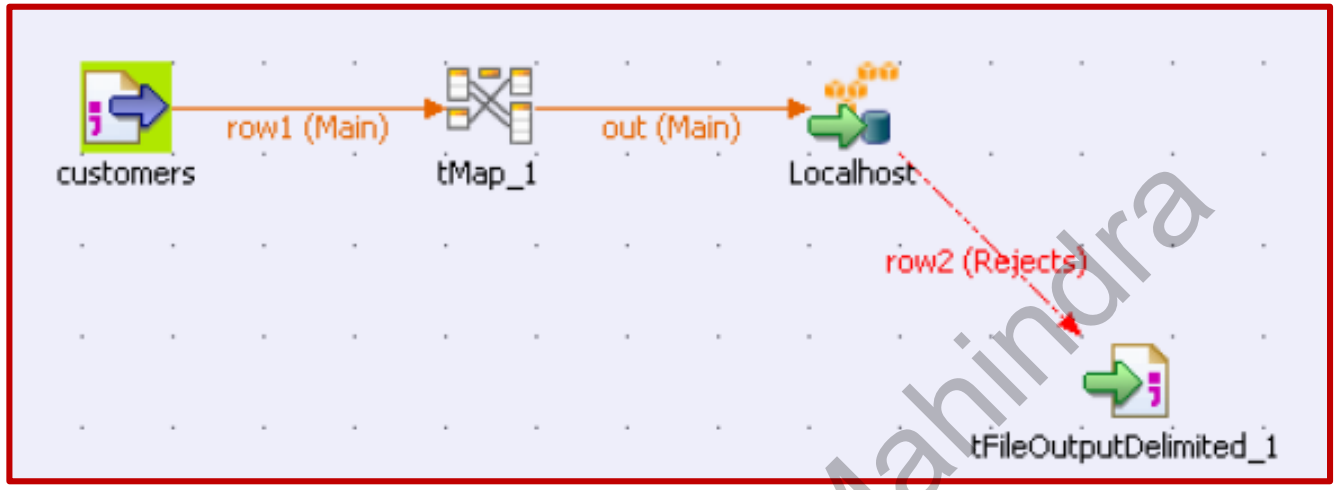


- **Row Main**
- **Iterate**
- **Multiple Input / Output**
- **Filter**

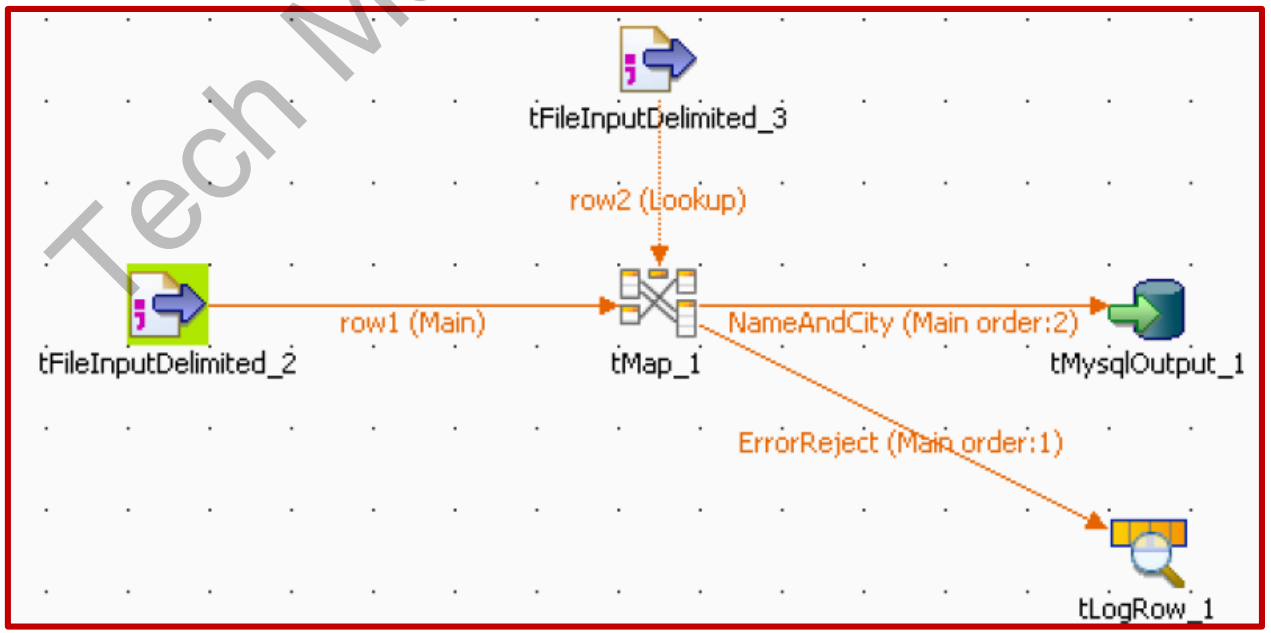
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Lookup, Reject, Error Reject Connections



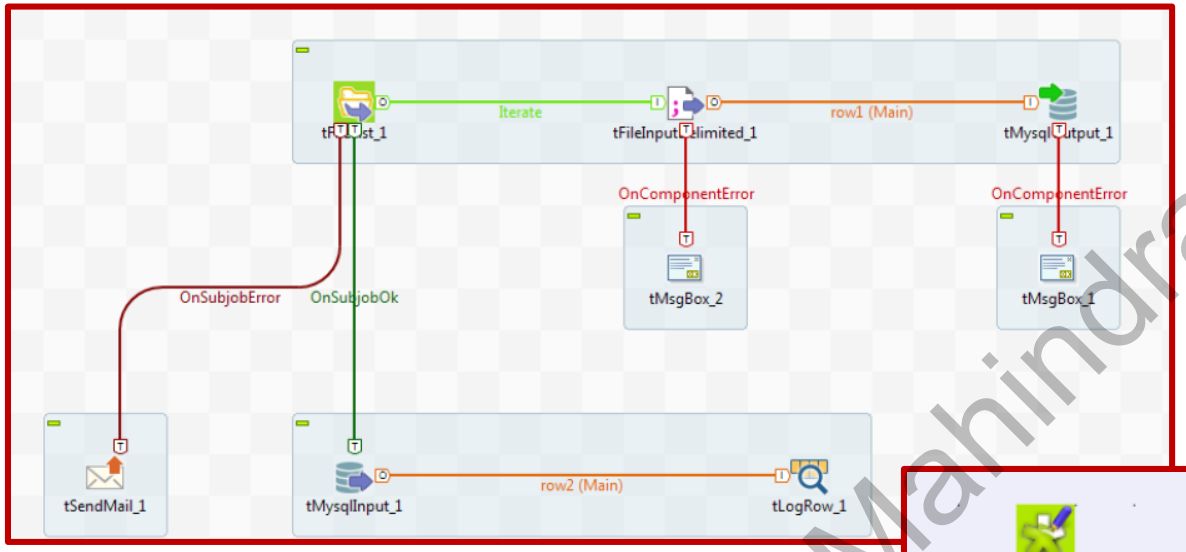
- **Reject**
- **Lookup**
- **Error Reject**



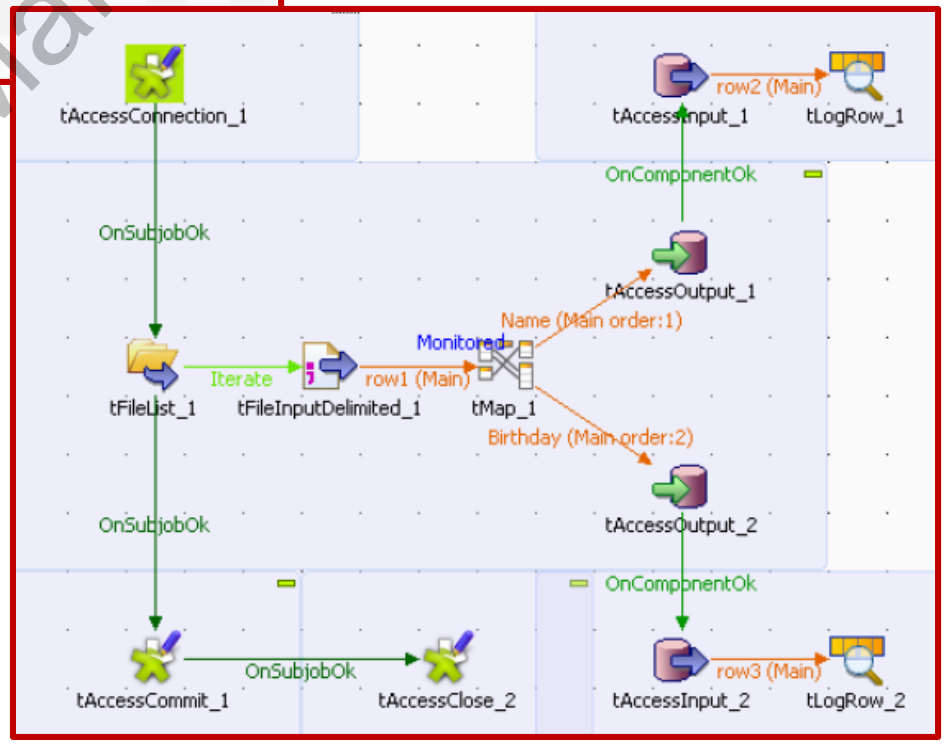
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Trigger Connections



- **OnSubjobOK**
- **OnSubjobError**
- **OnComponentOK**
- **OnComponentError**



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Filter Row

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

Filter Row

- **tFilterRow** filters input rows by setting one or more conditions on the selected columns

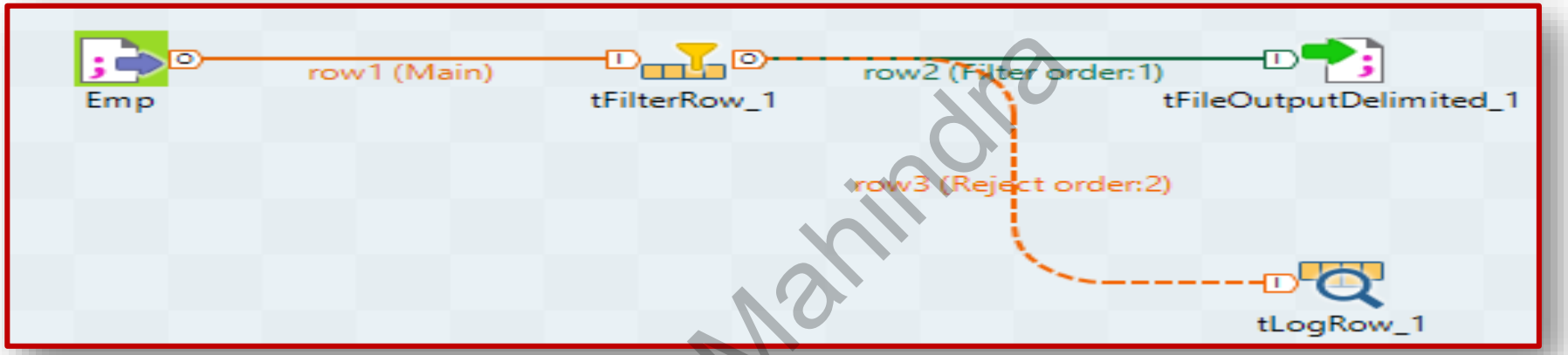
- ***Conditions***
  - **Input column**: Select the column of the schema the function is to be operated on
  - **Function**: Select the function on the list
  - **Operator**: Select the operator to bind the input column with the value
  - **Value**: Type in the filtered value, between quotes if needed

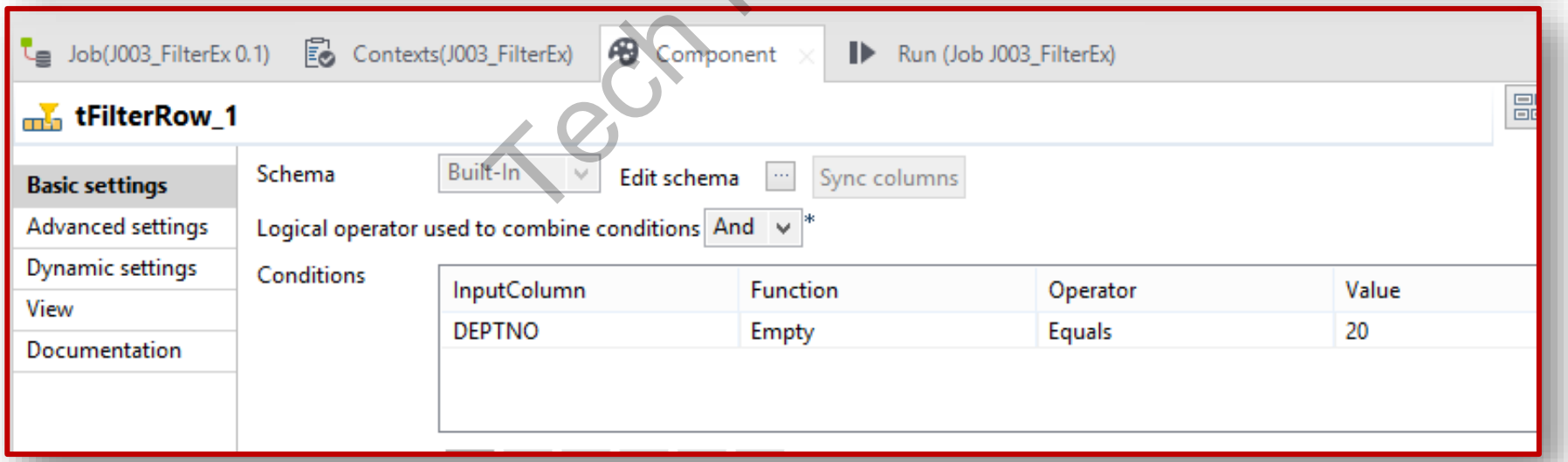
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Filter Row

- Design the job as follows and set properties





Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Demo

Create a simple job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Managing Metadata

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

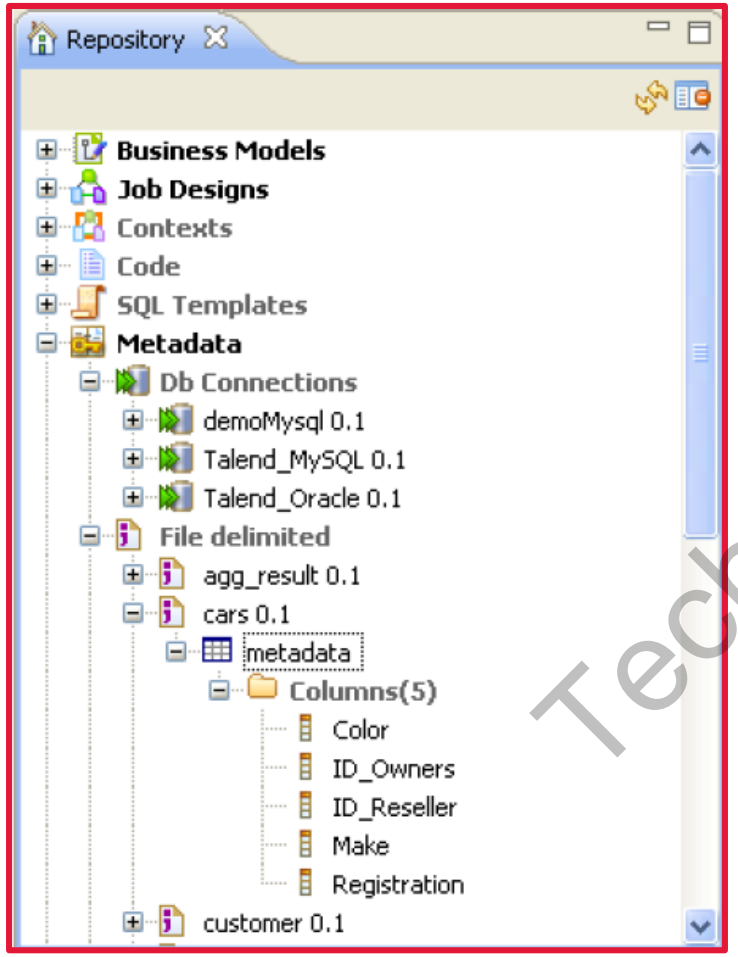
Managing Metadata

- To create and manage various metadata items in the **Repository** that can be used in all the Job designs
- The **Metadata** folder in the **Repository** tree view stores reusable information on files, databases, and/or systems that need to be created in Jobs
- Various corresponding wizards help you store these pieces of information that can be used later to set the connection parameters of the relevant input or output components and the data description called "schemas" in a centralized manner in *Talend Studio*
- The procedures of different wizards slightly differ depending on the type of connection chosen

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Managing Metadata



From *Talend Studio*, one can set up the following :

- a DB Schema
- a JDBC schema
- a SAS connection
- a file schema
- an LDAP schema
- a Salesforce schema
- a generic schema
- a MDM connection
- a FTP connection
- a WSDL schema
- Etc …

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Aggregate Row

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Aggregate Row

- **AggregateRow** receives a flow and aggregates it based on one or more columns. For each output line, are provided the aggregation key and the relevant result of set operations (min, max, sum...).
- ***Group by***
  - Define the aggregation sets, the values of which will be used for calculations.
  - **Output Column**: Select the column label in the list offered based on the schema structure you defined. You can add as many output columns as you wish to make more precise aggregations.
  - **Input Column**: Match the input column label with your output columns, in case the output label of the aggregation set needs to be different.
- ***Operations*** Select the type of operation along with the value to use for the calculation and the output field

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Aggregate Row

- Drag tFileInputDelimited, tAggregatorRow and tLogRow

- Set Properties of **tFileInputDelimited**
  FileName, CSV Options, Header, Footer etc..

- Set Properties of **tAggregateRow**
  GroupBy : DeptNo
  Operations :

Group by

Output column	Input column position
DEPTNO	DEPTNO

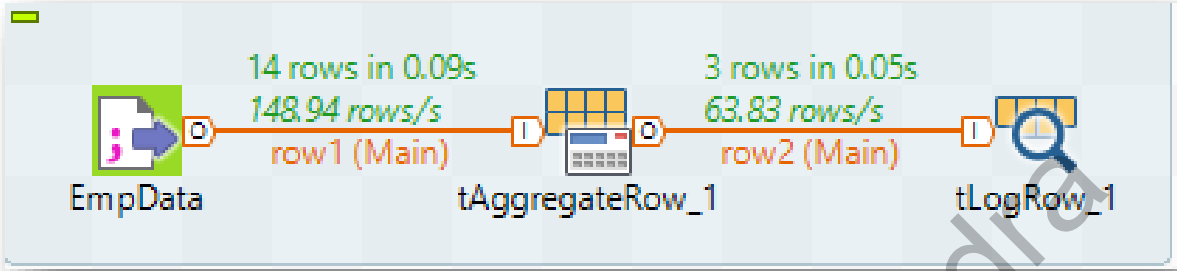
Operations

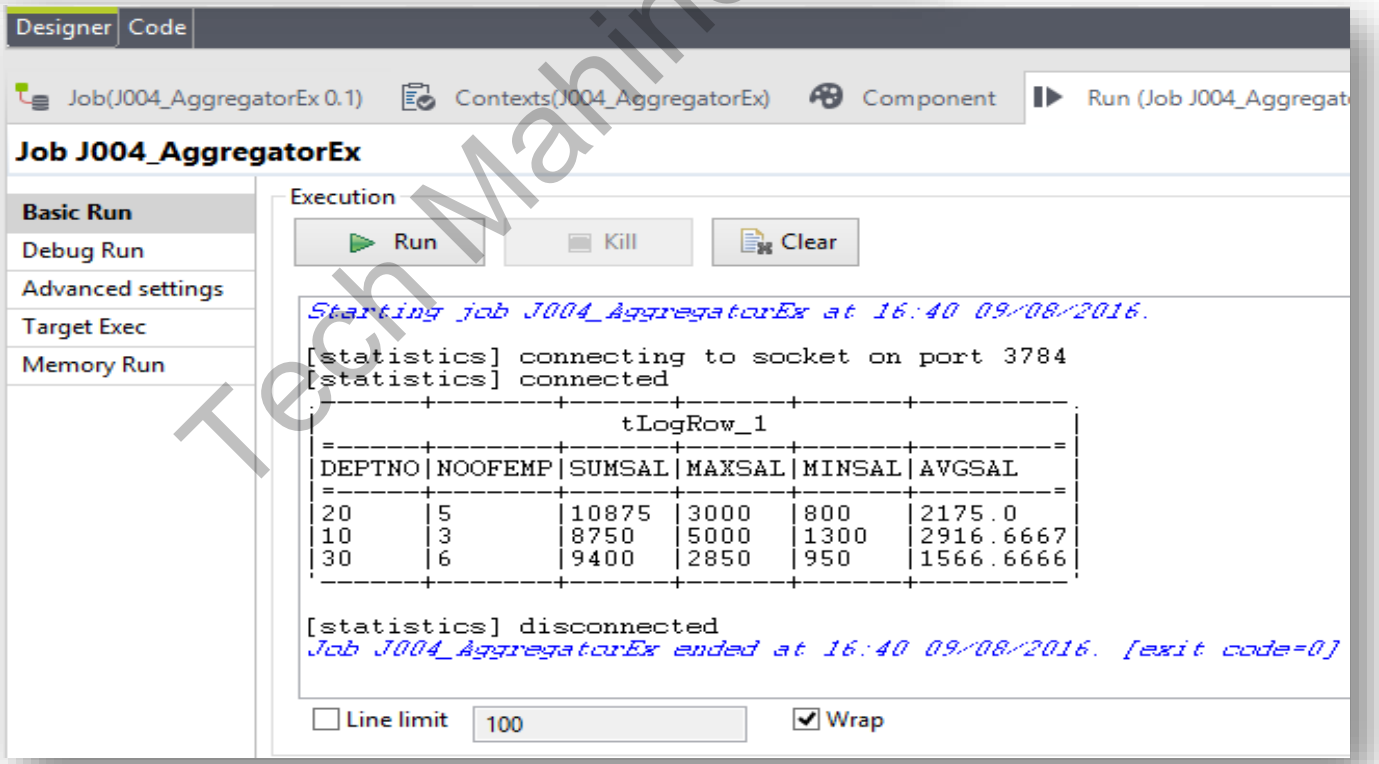
Output column	Function	Input column position	Ignore null values
NOOFEMP	count	EMPNO	☐
SUMSAL	sum	SAL	☐
MAXSAL	max	SAL	☐
MINSAL	min	SAL	☐
AVGSAL	avg	SAL	☐

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Aggregate Row





Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Demo

Create a simple job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

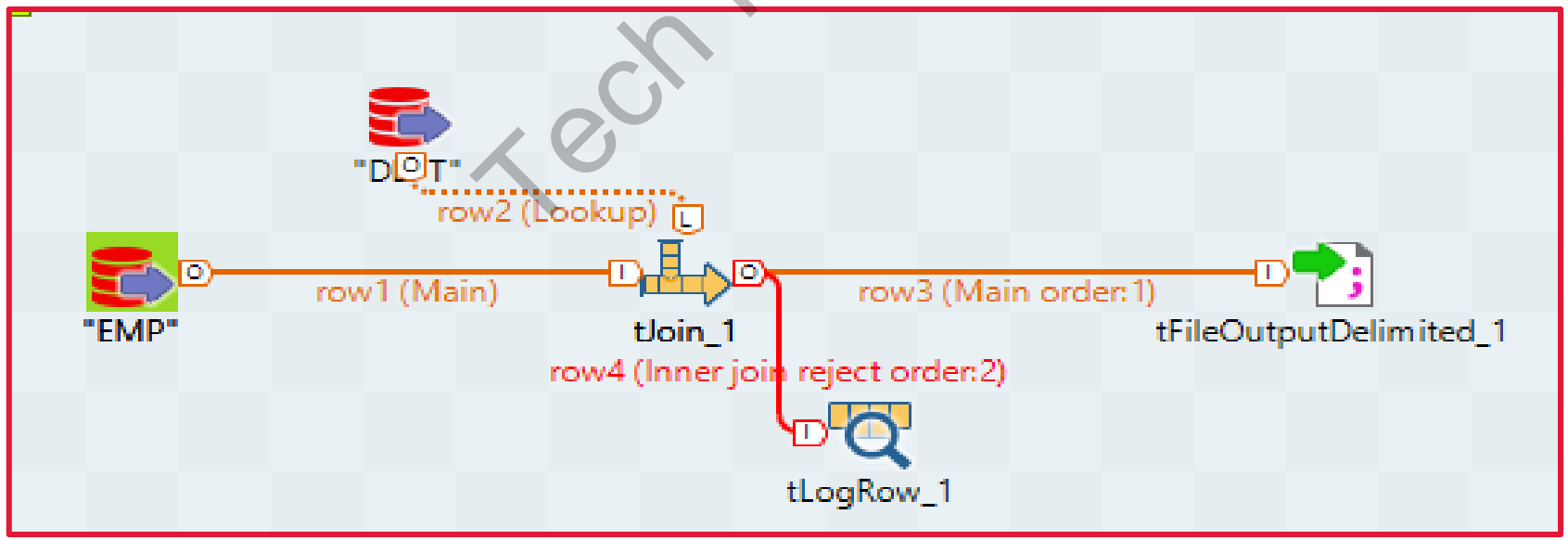
Joiner

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Join

- Drag **Emp** and **Dept** – **OracleInput** Components, **tJoin**, **tFileOutputDelimited** and **tLogRow** components and connect as shown in the diagram.
- Set Properties of **tFileInputDelimited**
  FileName, CSV Options, Header, Footer etc..



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Demo

Create a simple job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Global Variables and Contexts

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

Global Variables and Contexts

- Variables
  - Placeholder for values used during execution
    - Variable's values changes at different executions
    - values also change within a single execution

- Types of Variables:
  - Studio Global Variables
  - User defined global variables
  - Job Contexts

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

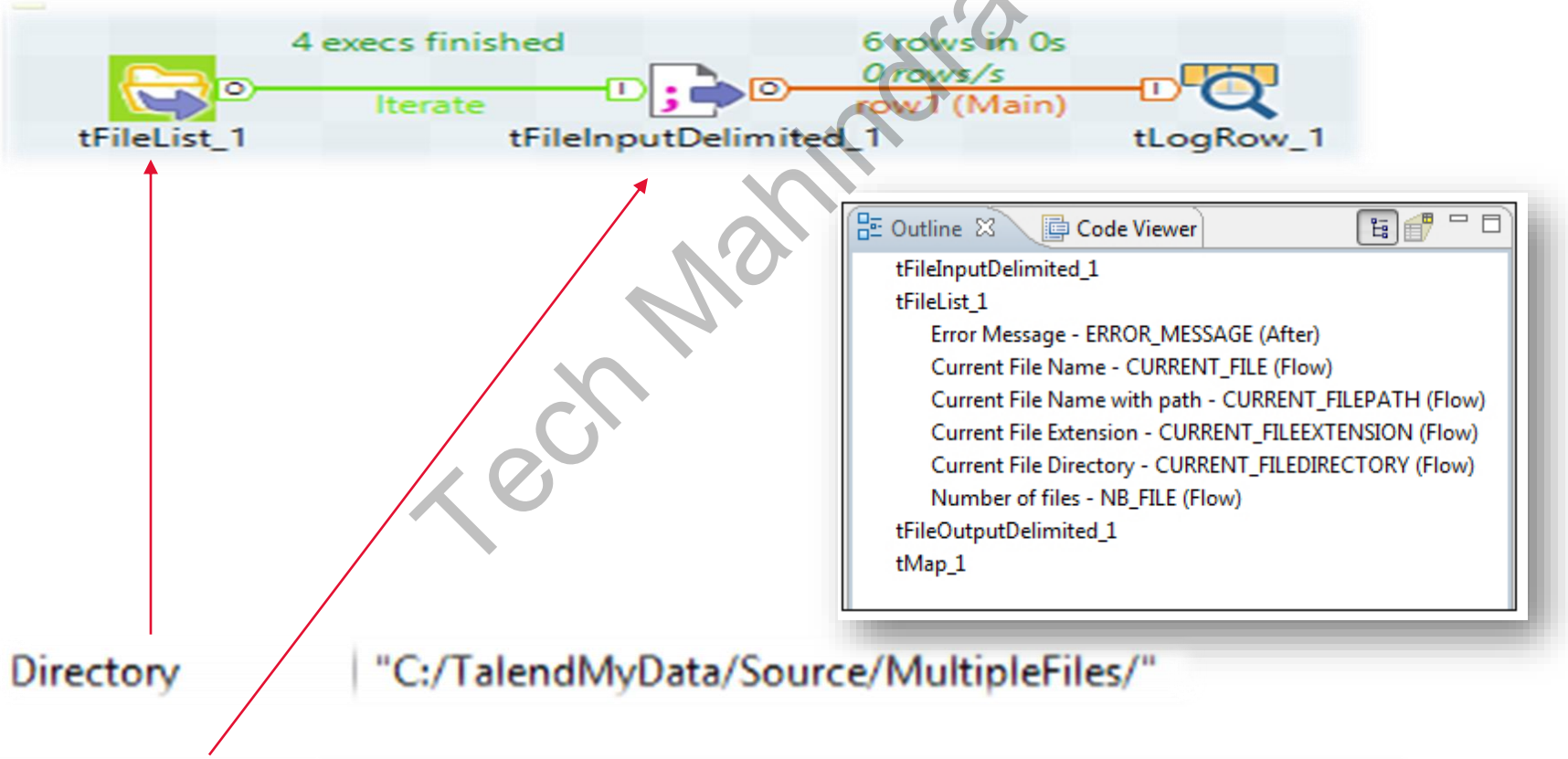
Studio Global Variables

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Studio Global Variables

- Available to all components, modules or functions within a Job
- **Ex**: Usage of Studio Global Variables -



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

User defined Global Variables

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

User defined global variables

- Ad-hoc variables that can be configured in your jobs
- **Ex:**
  - Fixed" values that you need to access in your integration job—client = "Acme Corporation" or tax_rate = 20%, for example
  - Variables that you create from standard Java functions—day, month, year values, for example
  - Variables that you set based on the data accessed by the integration job

- **globalMap.put("varName", varValue)** function to sets the variable value
- **globalMap.get("varName")** function to gets the variable value

- **Ex:** Using **globalMap** function

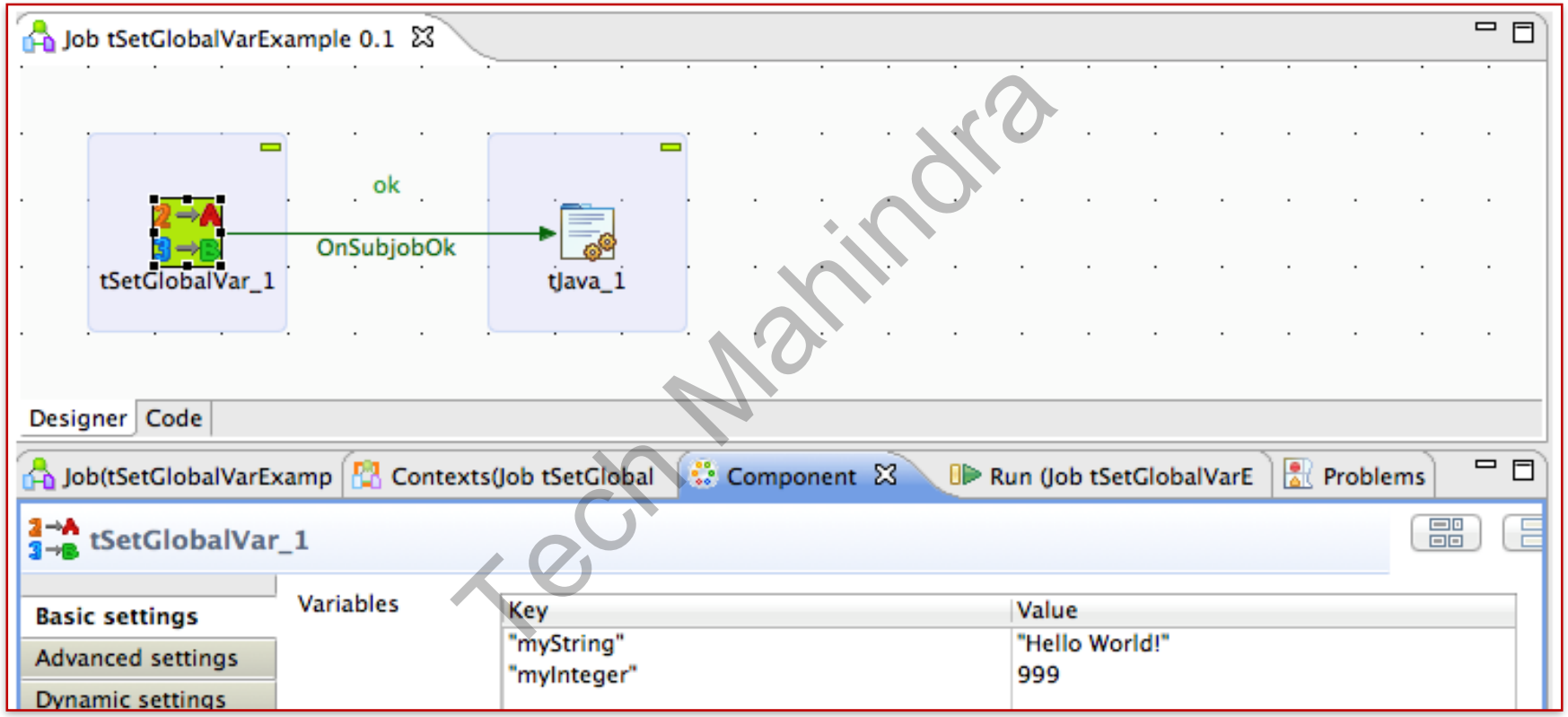
```
System.out.println("myString=" + globalMap.get("myString"));
System.out.println("myInteger=" + globalMap.get("myInteger"));
```

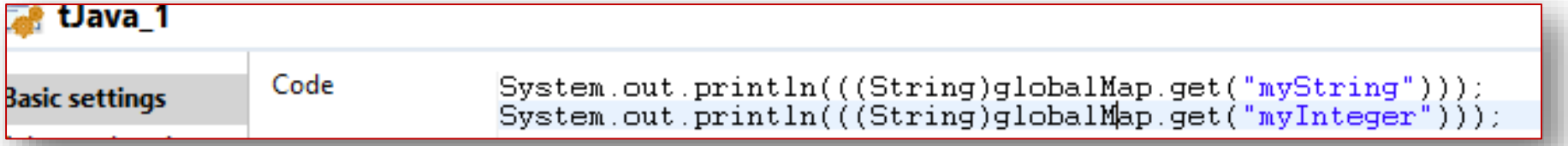
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

User defined global variables

- Ex: Using **tSetGlobalVar** component





Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

JOB Contexts

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Job contexts

- The variables we can create to execute jobs with different parameters for different environments or scenarios
- **Ex**: Connection info in 3-different environments – using context variables.
- Invoked at runtime
- Different ways of implementing
  - **embedded variables**,
  - **repository variables**, or
  - **external variables**.

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

JOB Contexts

Embedded Context Variables

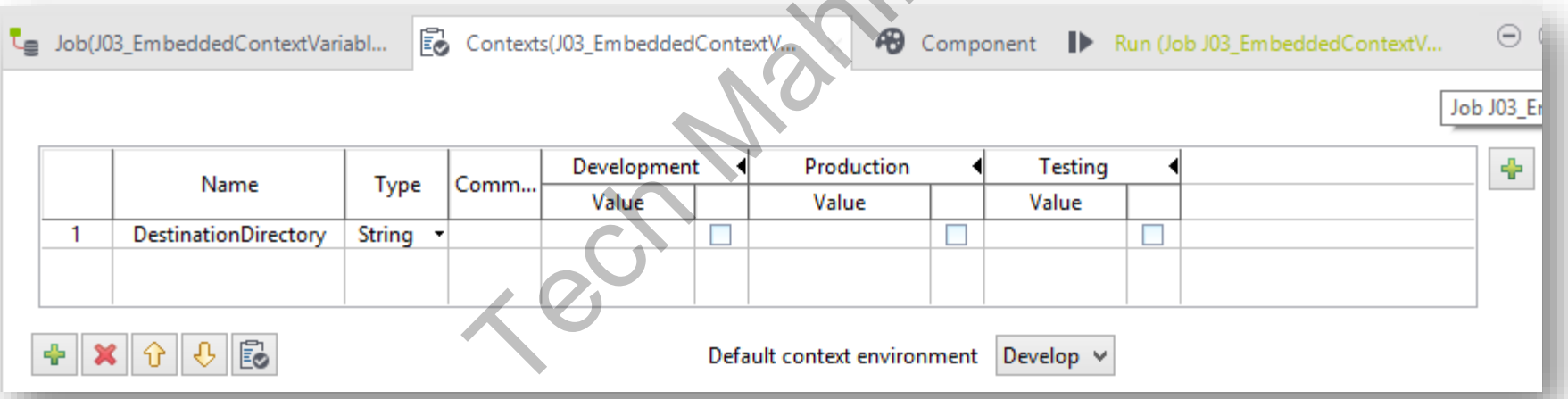
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : Embedded Variables

- Embedded in the job and are configured much like any other component parameters

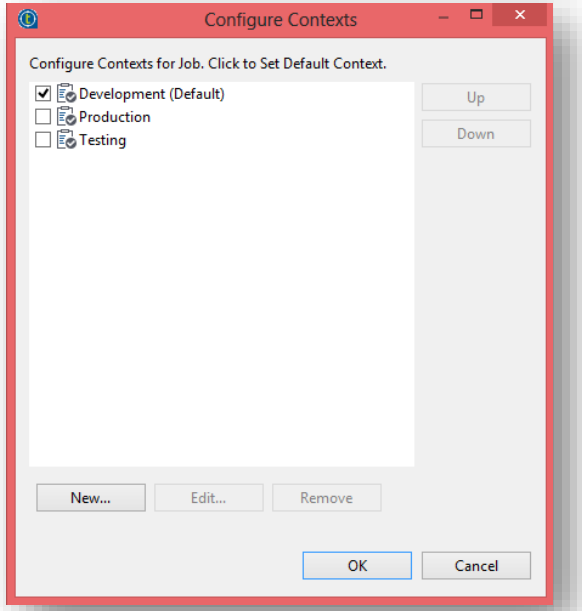
- **Ex:** How to access different file locations with different context ("test" context or a "production" context)



Tech Mahindra

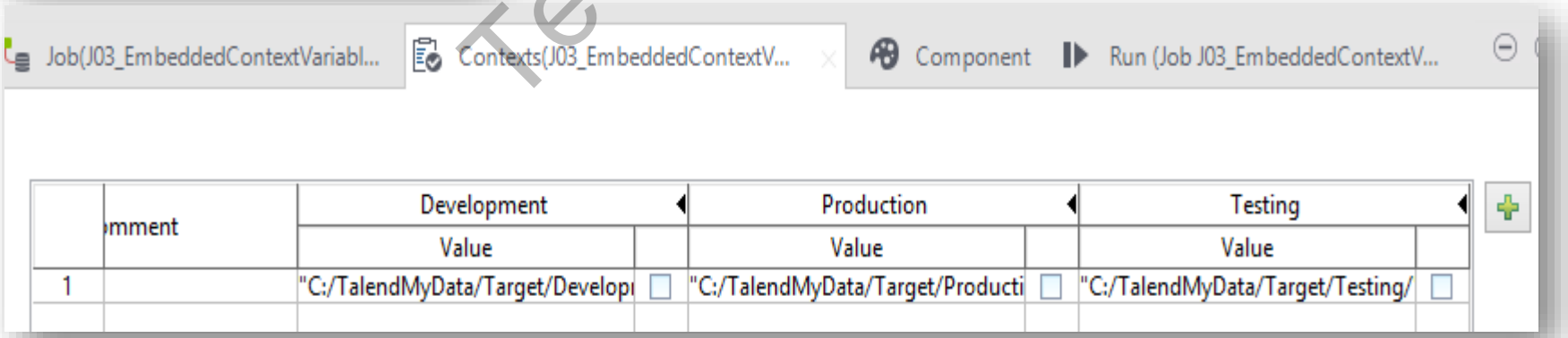
Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : Embedded Variables



Create context variables using Context tab

- **Setting multiple contexts [Development / Production / Testing] of the same variable (DestinationDirectory)**

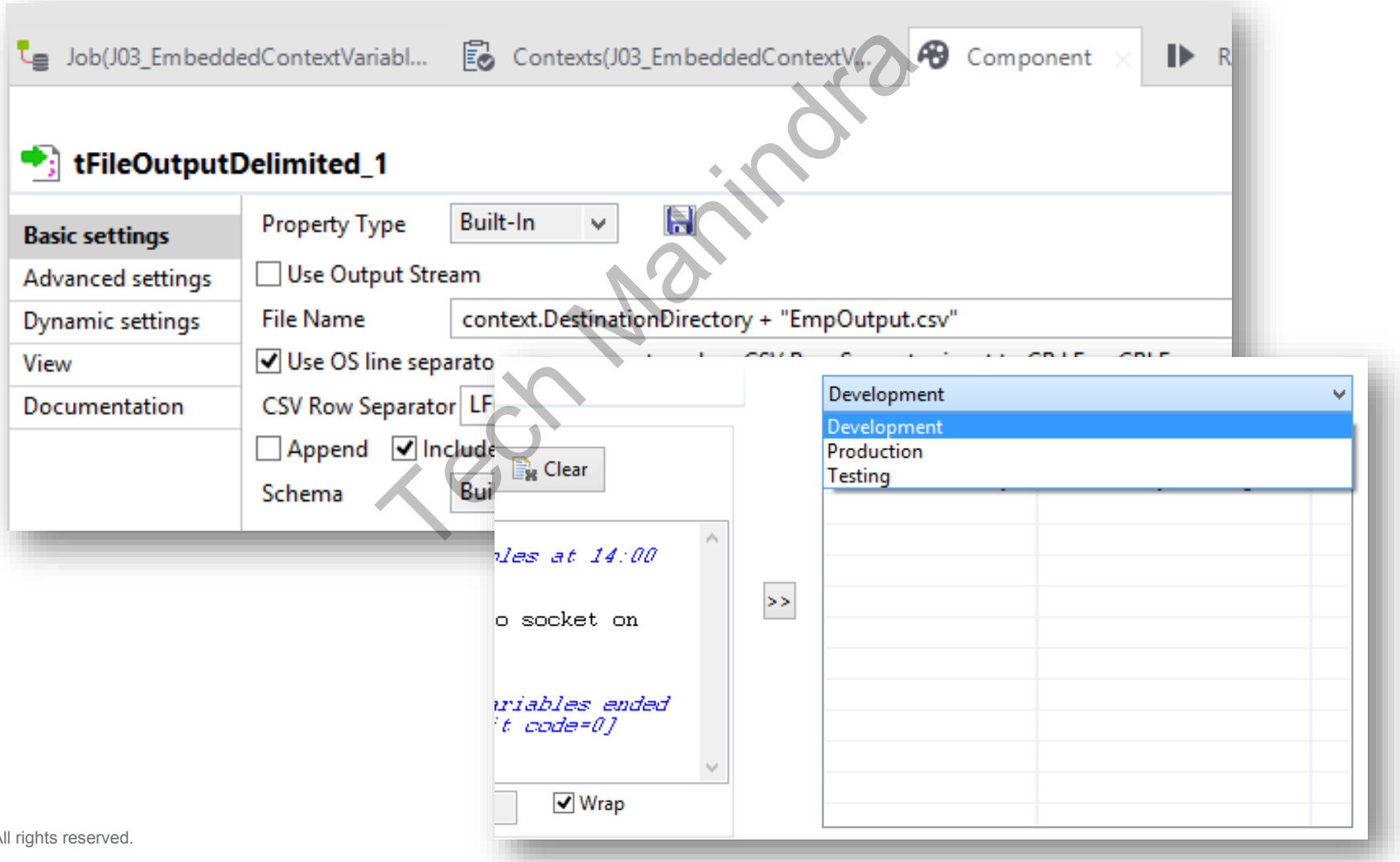


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : Embedded Variables

Using the context variable in the job



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

JOB Contexts

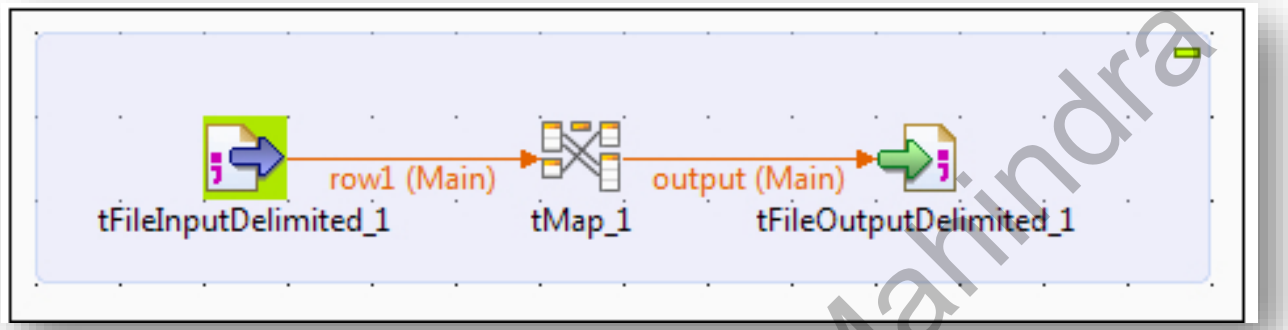
Repository Context Variables

Tech Mahindra

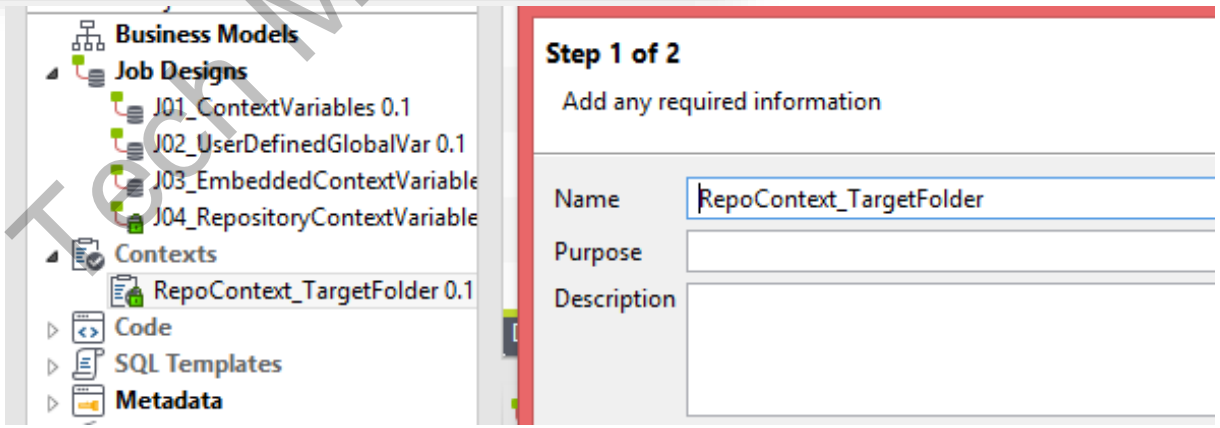
Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : Repository context variables

- Contexts can be stored in the Studio repository, allowing them to be centrally maintained, available to other developers and reused on other jobs



Design the Job

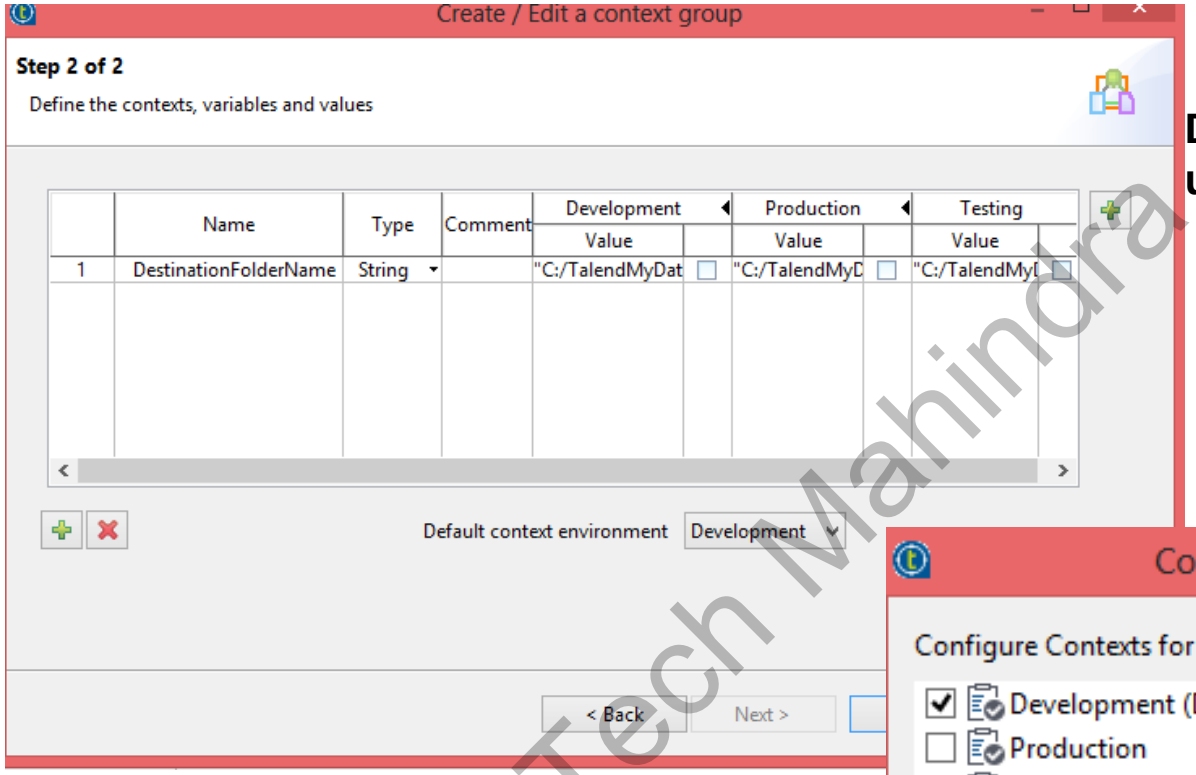


Create Context Group and variables using Repository → Contexts → Create Context Group

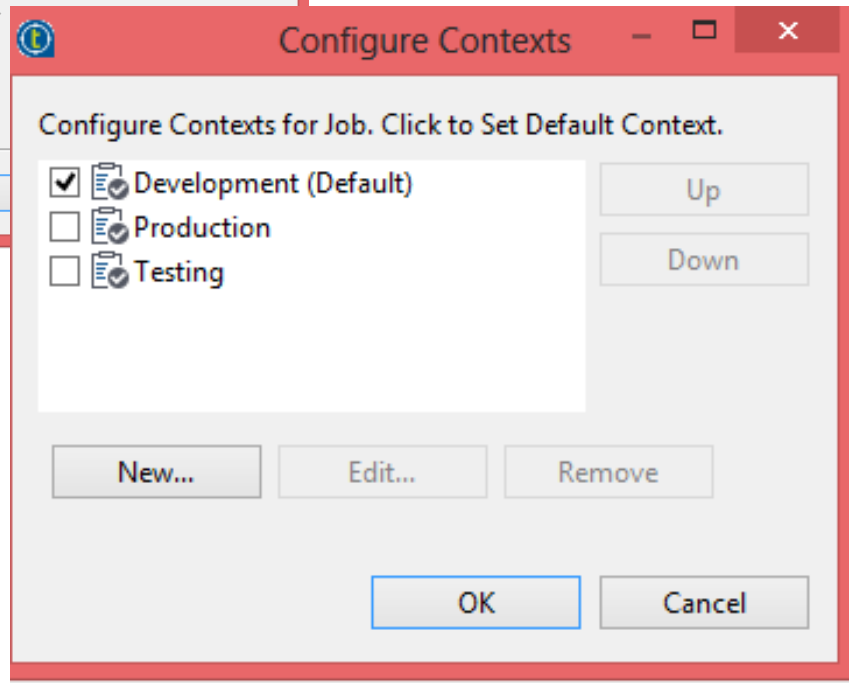
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : Repository context variables



Defining different Contexts using Contexts Tab

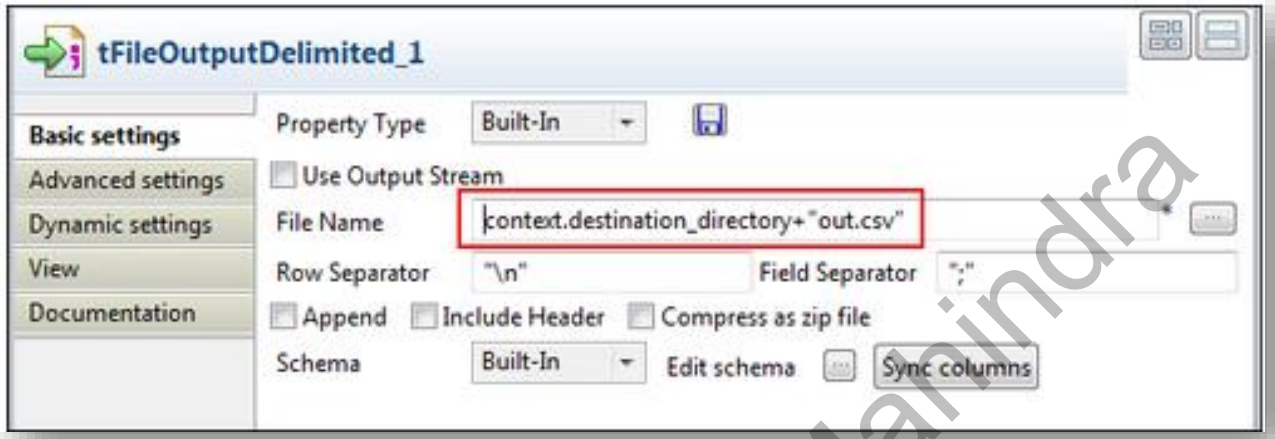


Select the Default Context

Tech Mahindra

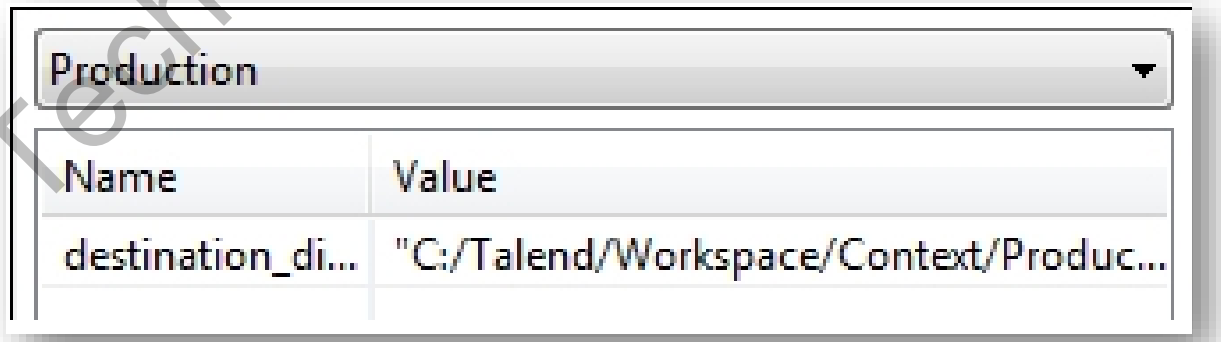
Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : Repository context variables



Setting default Context

Using the required context at run-time



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

JOB Contexts

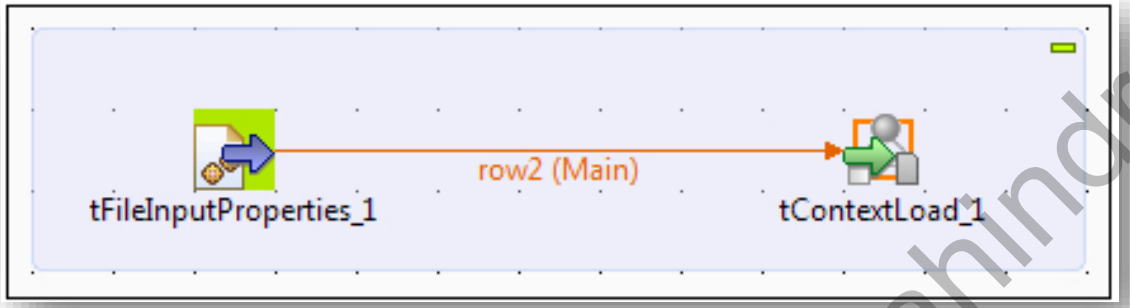
External Context Variables

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

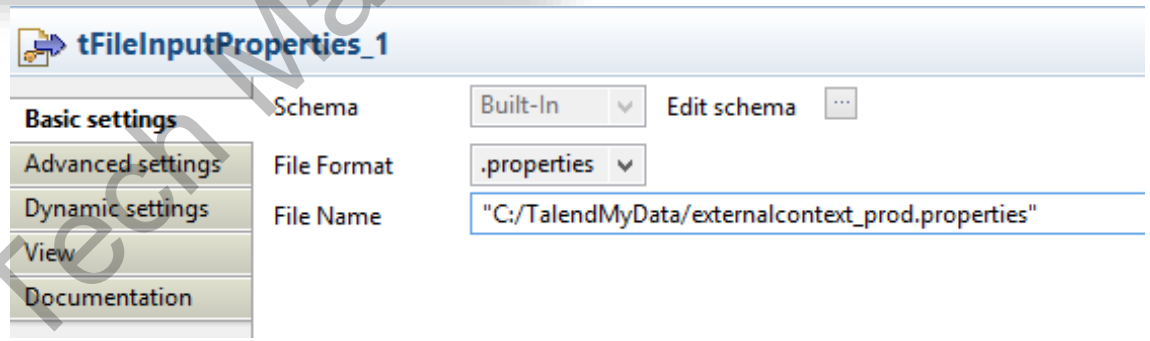
Job Contexts : External context variables

- External context variables are variables held in a file and loaded into the Studio job at runtime

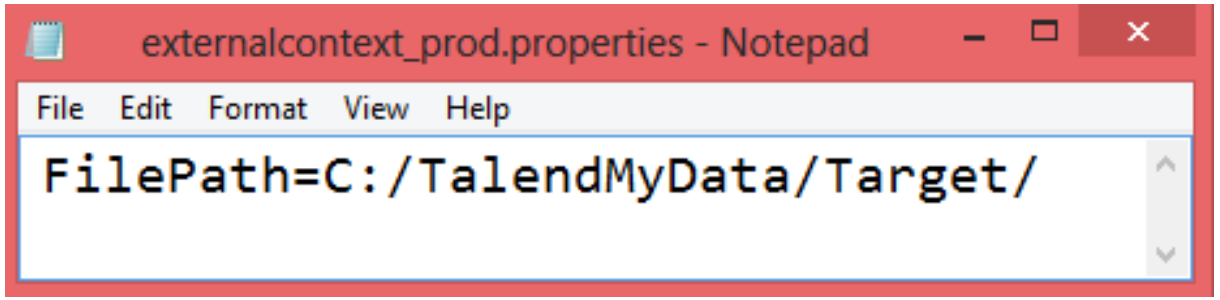


Create External Properties file and use it with File properties component

Create External Properties file and use it with File properties component



Sample properties file (External Context)



Tech Mahindra

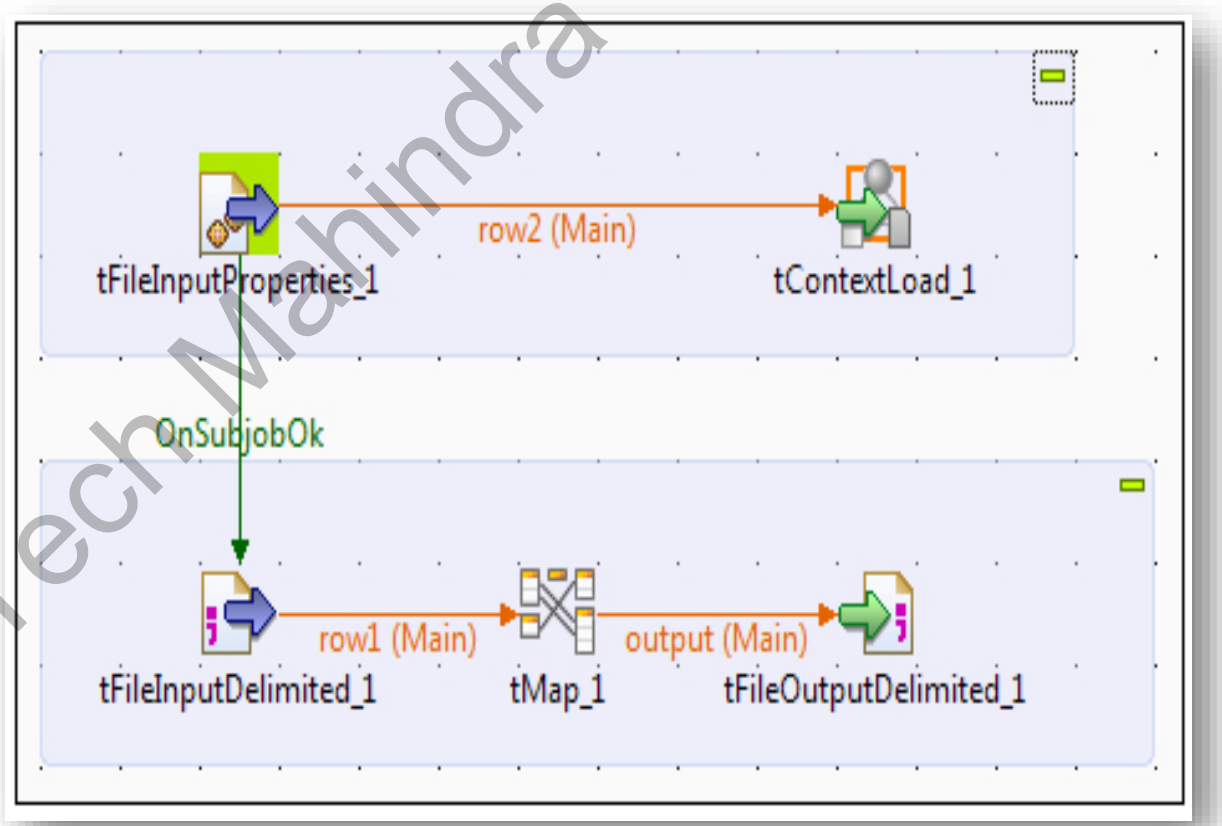
Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : External context variables

- External context variables are variables held in a file and loaded into the Studio job at runtime

Create context variable and map to external variable file path to the context variable

Use it in file out

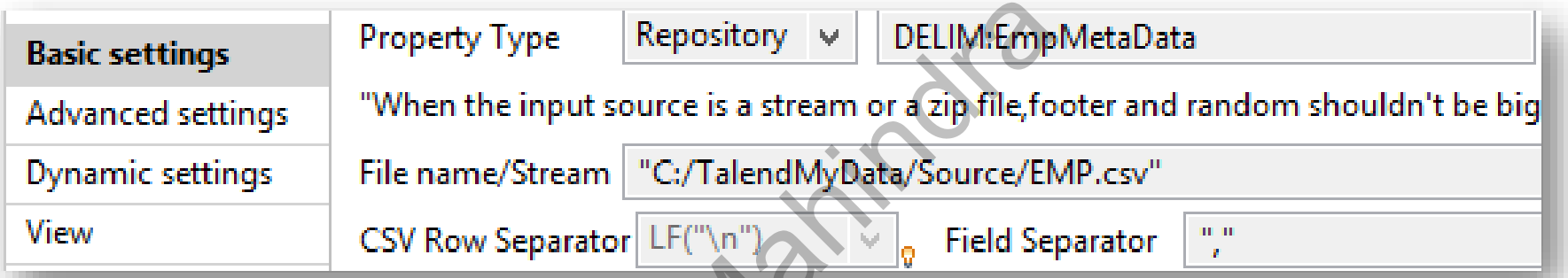


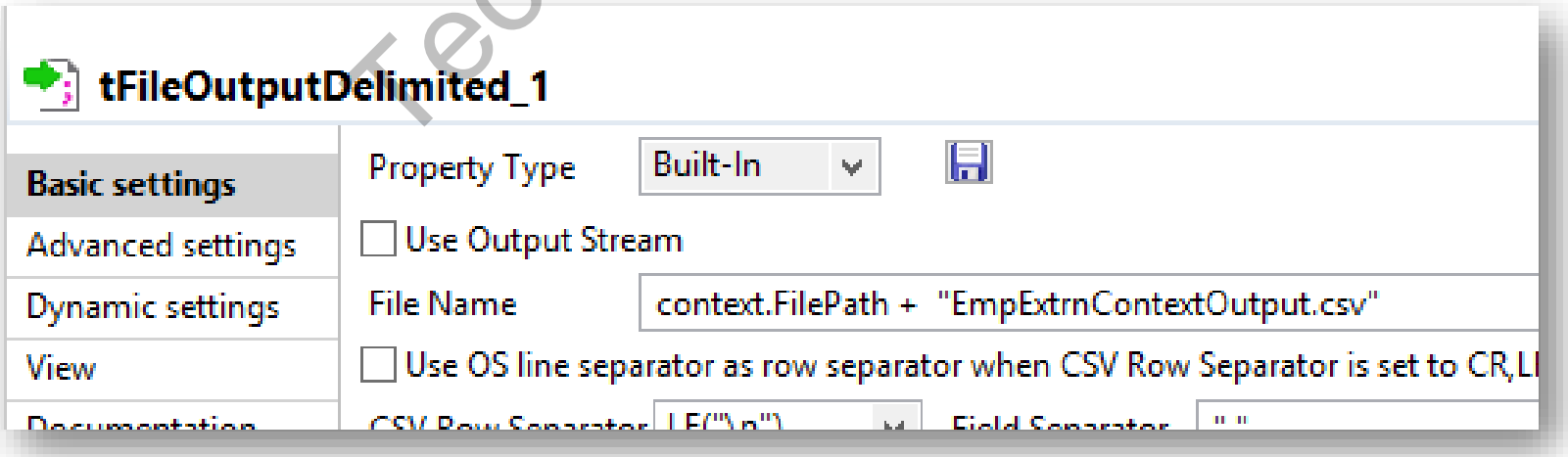
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Job Contexts : External context variables

- Properties of tInputFileDelimited, tOutputFileDelimited





Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

tMap - Router

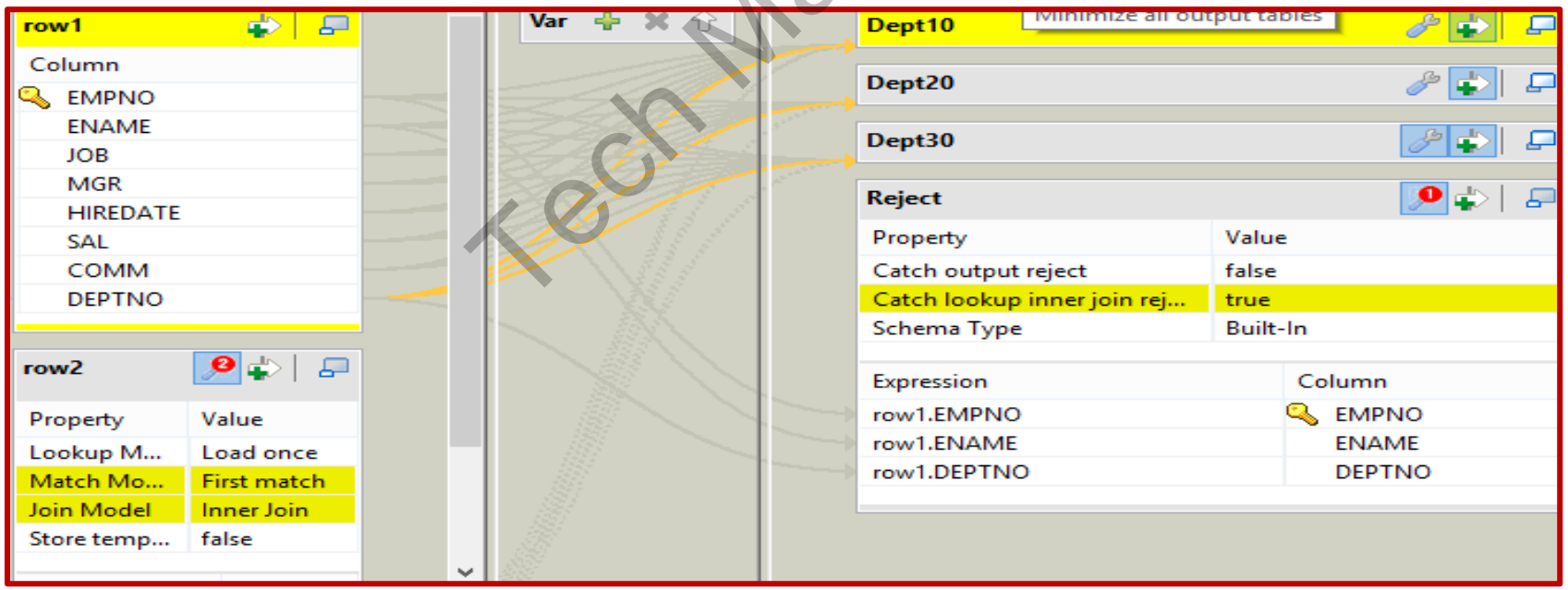
Tech Mahindra

Sreenivas_Ram

Copyright © 2016 Tech Mahindra. All rights reserved.

tMap – as router

- Drag EMP and DEPT tables from Oracle Metdata
- Drag **tMap**, 3 copies of **tFileOutputDelimited** and connect as shown in the pic below
- In Mapping editor of **tMap** :
  - Connect **DeptNo** from **EMP** to **DEPT** columns to join them
  - Let the Join conidition be **InnerJoin**
  - Drag both **EMP** and **DEPT** columns on two 3-Ouput components
  - Select **Loop Inner Join Reject option** to **TRUE** and map related rows

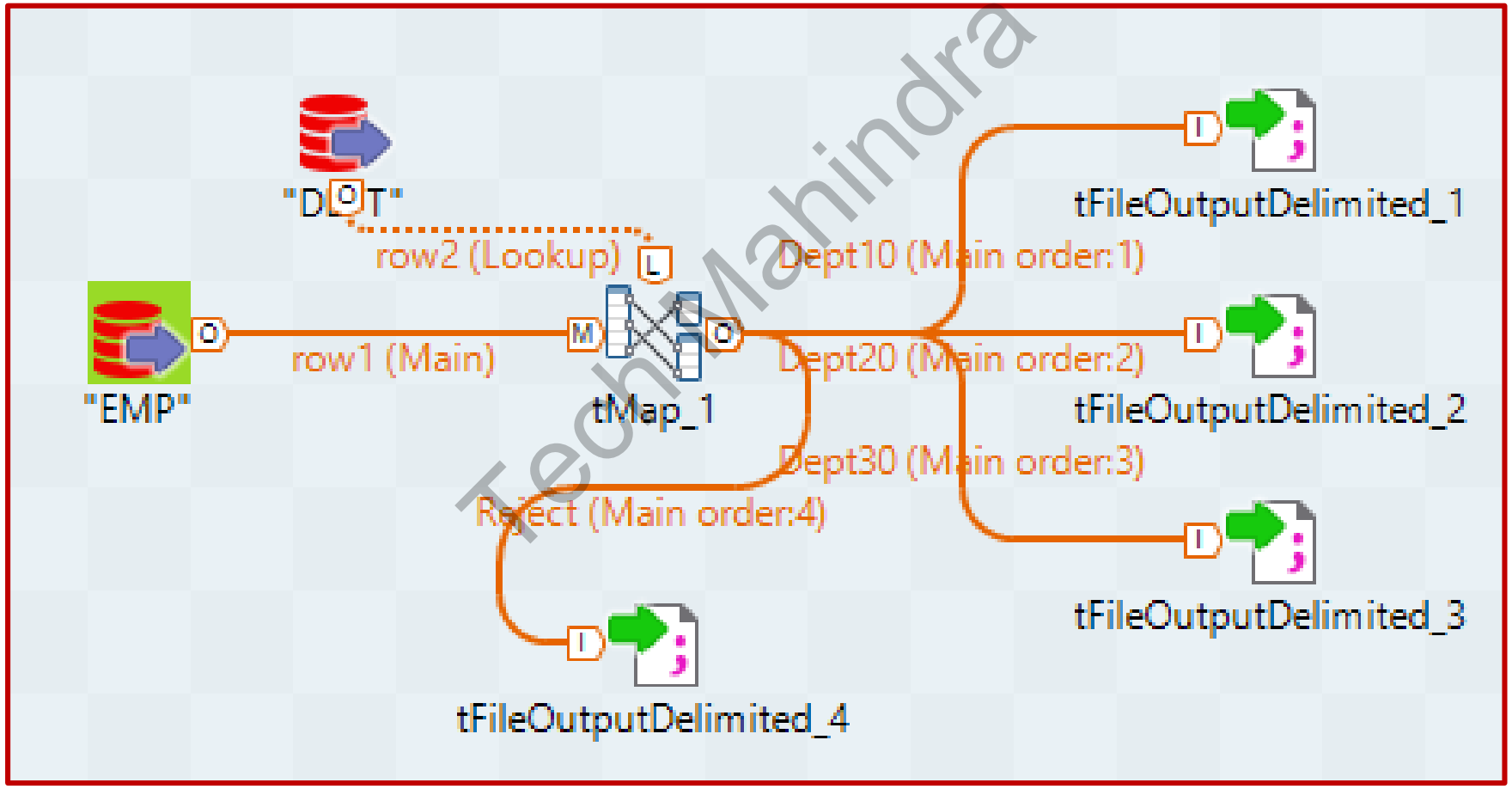


Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

tMap – as router contd …

- **Save** and **Run** the job



Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Talend

tReplace

Tech Mahindra

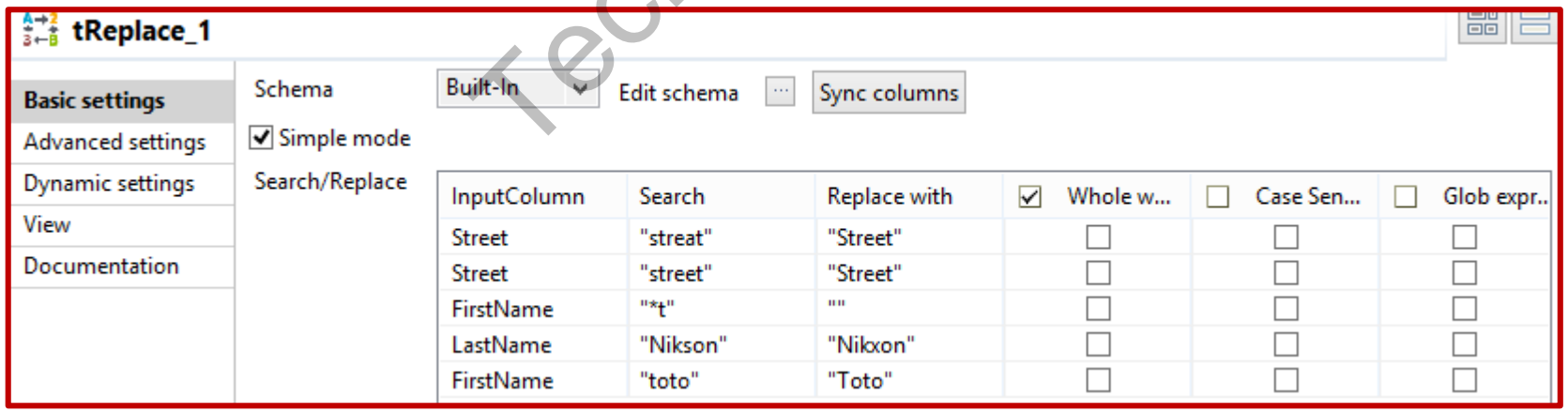
Sreenivas_Ram

Copyright © 2016 Tech Mahindra. All rights reserved.

tReplace

- This Job searches and replaces various typos and defects in a csv file then operates a column filtering before producing a new csv file with the final output
- Drag **tFileInputDelimited**, **tReplace**, **tFilterColumns, tFileOutputDelimited** on to the job.

Street	FirstName	LastName	Amount
stream	John	Kennedy	98.3
stream	Richad	Nikson	78.23
stream	Richard	Nikson	78.2
stream	toto	Nikson	87.23
stream	Richard	Nikson	87.23
street	George*t	bush	99.98



Tech Mahindra

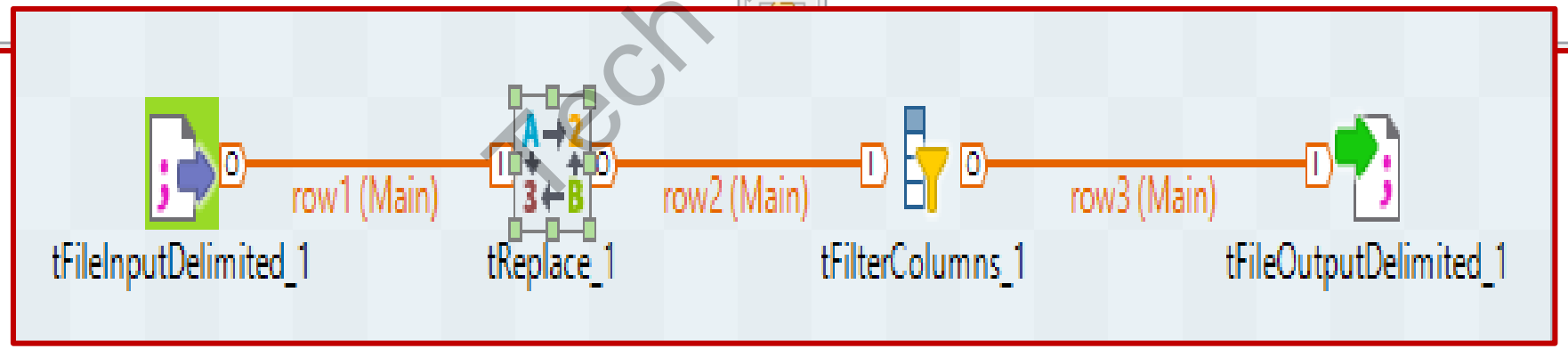
Copyright © 2016 Tech Mahindra. All rights reserved.

tReplace … contd.

- Add an additional column in **tFilterColumn** component
- **Save** and **Run** the job

Column	Key	Ty...	N..	Date P...	Le...	Pr...	D...	C...
Street	☐	St...	☑		6			
FirstName	☐	St...	☑		10			
LastName	☐	St...	☑		10			
Amount	☐	Fl...	☑		5	2		

Column	Key	Ty...	N..	Date P...	Le...	Pr...	D...	C...
EmpNo	☐	In...	☑		5			
Street	☐	St...	☑		6			
FirstName	☐	St...	☑		10			
LastName	☐	St...	☑		10			
Amount	☐	Fl...	☑		5	2		



Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Talend

Execute Job in Loop

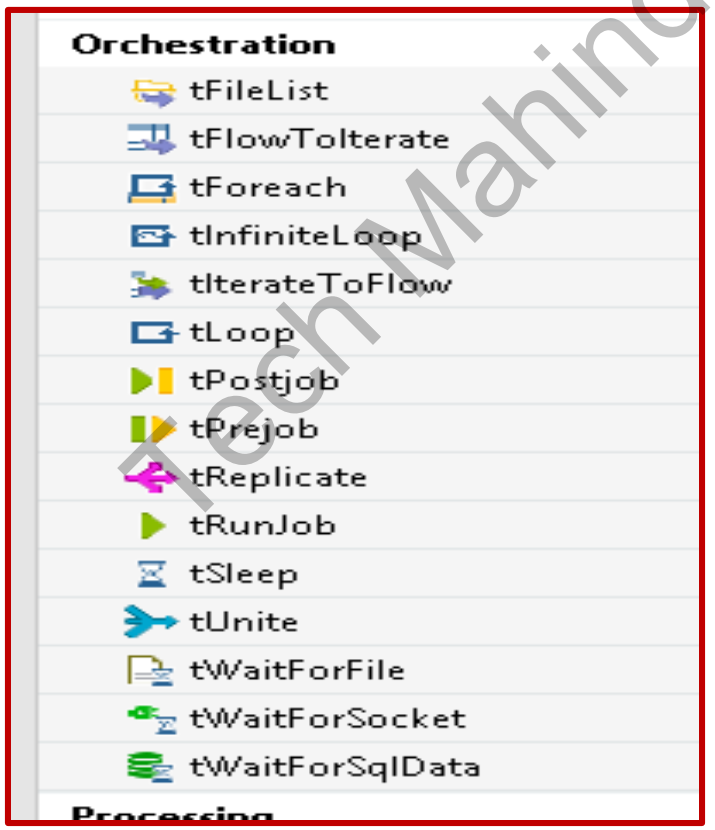
Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

Orchestration components

- The Orchestration family groups together components that help you to **sequence** or **orchestrate** tasks or processing in your Jobs or subjobs and so on

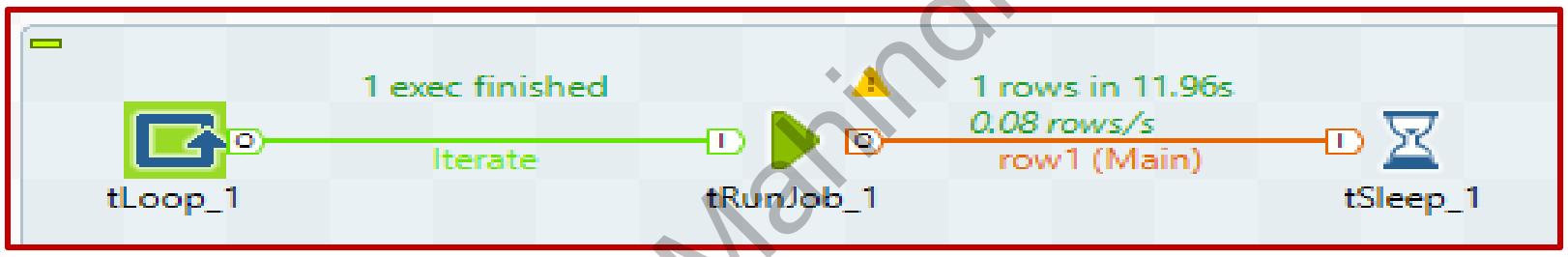


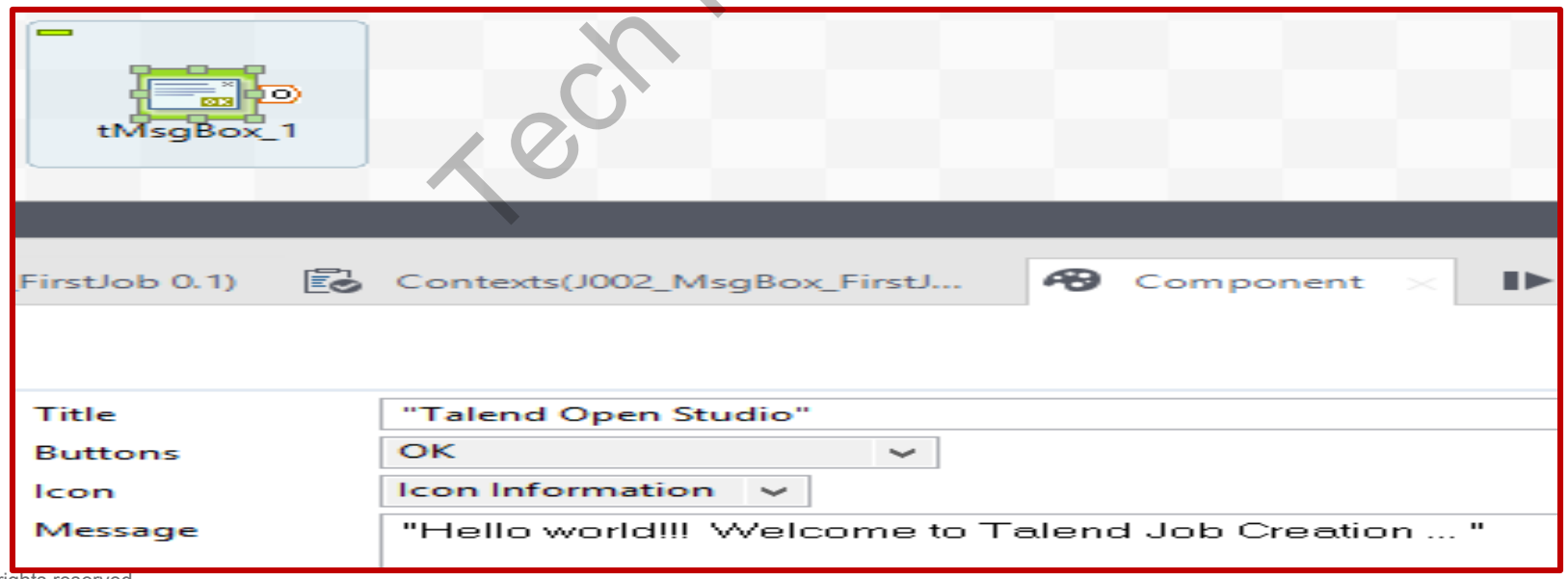
Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Execution of a Job in Loop

- This scenario describes a Job composed of a parent Job and a child Job. The parent Job implements a loop which executes n times a child Job, with a pause between each execution





Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

In the parent Job

- Drop a **tLoop**, a **tRunJob** and a **tSleep** component from the **Palette** to the design workspace.

- Connect the **tLoop** to the **tRunJob** using an **Iterate** connection

- Then connect the **tRunJob** to a **tSleep** component using a **Row** connection.

- **tLoop** Component properties :
  From : **1**
  To : **2**
  Step : **1**
- **tRunJob** component Properties :

  Job : **J002_MsgBox_FirstJob**

- **tSleep** component properties
  pause (in seconds) : **10**

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

On the Child Job

- Drop the following components:
  tMsgBox

- On **MsgBox** properties:
  Message**: "Hello world!!! Welcome to Talend Job Creation ... "**

- **Run** the job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

tPreJob
tPostJob

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

tPreJob & tPostJob Components

- The **tPreJob** and **tPostJob** components are part of the Orchestration family of components
- These components help to make your Jobs flow with more elegance, and provide the ability to perform some clean-up, should your Job failure

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

tPreJob

- The **tPreJob** component starts the first series of Subjobs that will be executed, when your Job starts
- Functionally, it performs no action that you could not achieve by connection your components in the correct order; however, it does provide a clear demarcation between your Job's initialisation sequence, and the main tasks that the Job is intended to perform

- The following tasks could be considered as pre-Job tasks: -
  - Load Context
  - Validate external environment
  - Establish database connections
  - Test connectivity to external services
  - Check input files exist
  - logging the start of your Job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

tPostJob

- The **tPostJob** component starts the final series of Subjobs that will be executed, when your Job finishes; that is, after **tPreJob** has completed and any other Subjobs that you have defined

- A key aspect to the execution of this component is that it is always executed, even when an Exception is thrown by any previous processing

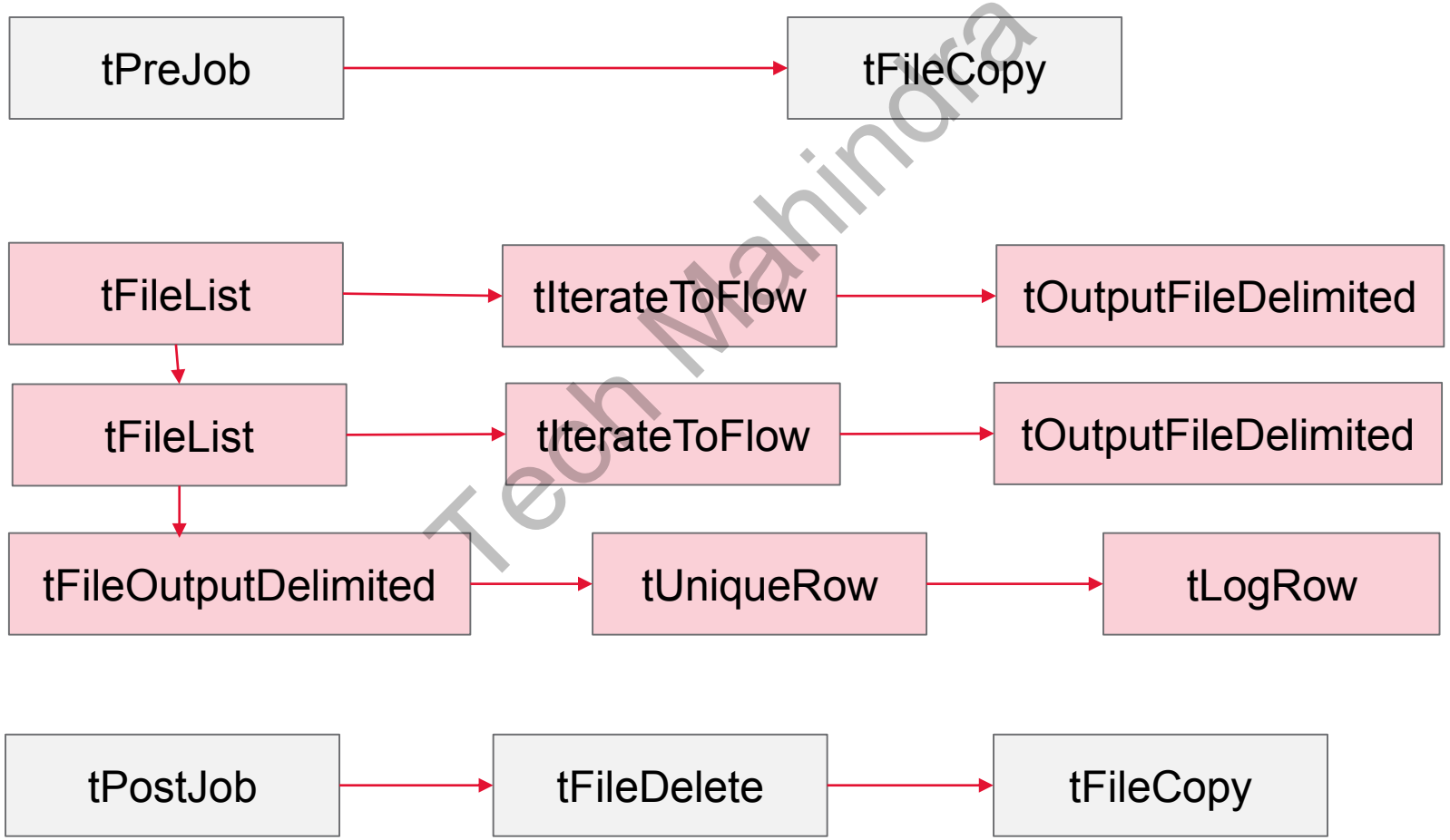
- This makes it ideal for controlling de-initialisation and clean-up.
  - The following tasks could be considered as post-Job tasks: -
  - Deletion of temporary files
  - Database disconnection
  - logging the finish of your Job

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Job Design

- Design a Job with the following Components

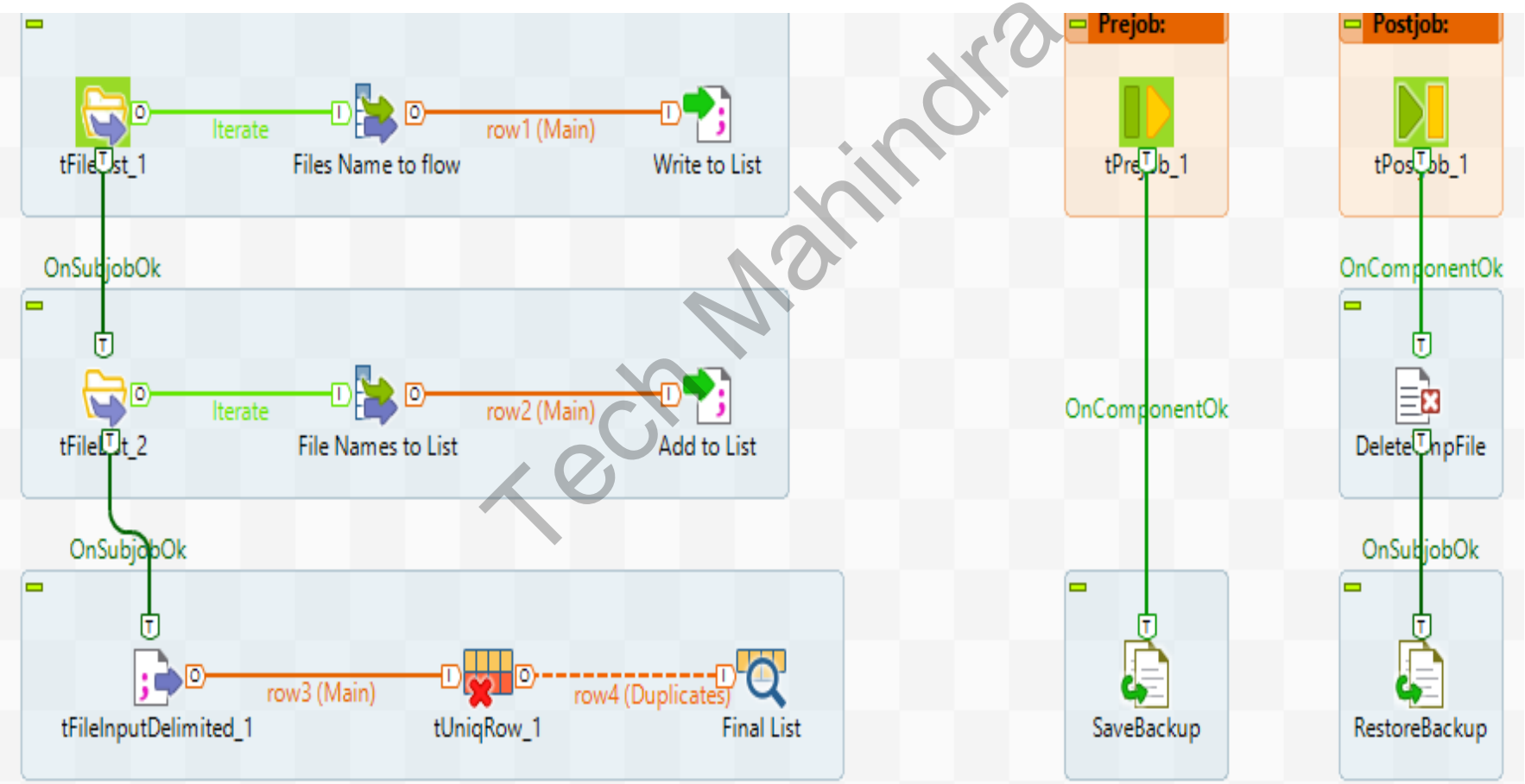


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Connectors

- These two components only support an *out* connector of **OnComponentOK**.



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Configurations (Properties)

- **InPreJob > Save Backup (tFileCopy)**

SaveBackup (tFileDelimited)	
FileName	"C:/TalendMyData/Source/MultipleFiles/EMP10.csv"
Destination Directory	"C:/TalendMyData"
Destination Name	"NewEmp10.csv"

- **InPostJob > Delete Temp File (tFileCopy)**

Delete Tmp File (tFileDelete)	
FileName	"C:/TalendMyData/Source/MultipleFiles/EMP10.csv"

Restore Backup (tFileCopy2)	
FileName	"C:/TalendMyData/Source/MultipleFiles/NewEmp10.csv"
Destination Directory	"C:/TalendMyData/Source/MultipleFiles/"
Rename Destination Name	"Emp10.csv"
Remove Source File	Select
Replace existing File	Select

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Configurations (Properties) ...contd.,

- **Main job**

tFileList	
Directory	"C:/TalendMyData/Source/MultipleFiles/"
File Names to flow (tIterateToFow)	
FileName	“((String)globalMap.get("tFileList_1_CURRENT_FILE"))”
tFileOutputDelimited (Write to List)	
FileName	"C:/TalendMyData/FileList1.csv”

tFileList	
Directory	""C:/TalendMyData/Source/MultipleFiles2/"
File Names to flow (tIterateToFow)	
FileName	“((String)globalMap.get("tFileList_1_CURRENT_FILE"))”
tFileOutputDelimited (Write to List)	
FileName	"C:/TalendMyData/FileList1.csv”
Select	Append Option

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Configurations (Properties) …contd.,

- **Main job … contd…**

tFileInputDelimited	
FileName/stream	"C:/TalendMyData/FileList1.csv"
tUniqueRow	
UniqueKey – Add FileName	Key attribute
tLogRow	
Select	Table View

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Information about Jobs Executions

Sreenivas_Ram

Tech Mahindra

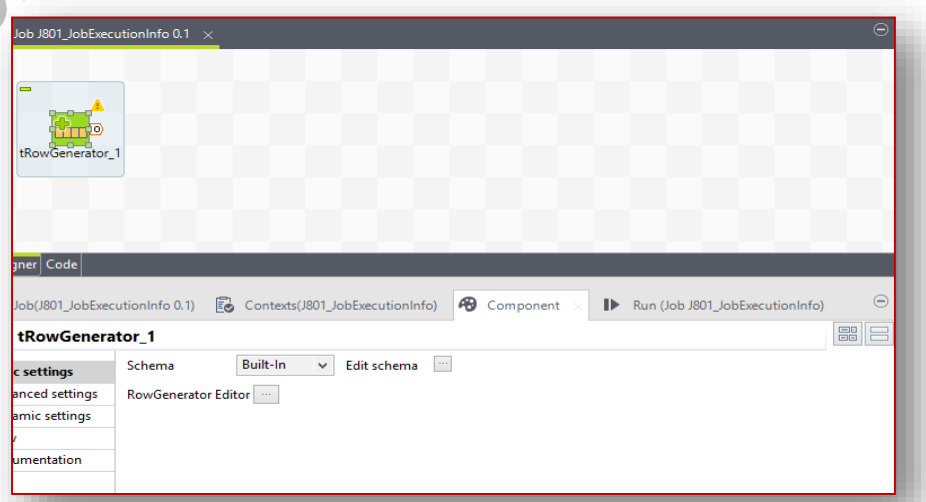
Copyright © 2013 Tech Mahindra. All rights reserved.

Job Execution Info

- Create a New job - Name it as **JobExecutionInfo**
- Drag a RowGenerator Component and create row structure as :

  EmpNo Integer Numeric.random(1101-1900)
  EName String RowGenerator.getLastName()
  SAL Integer Numeric.random(3000,99000)

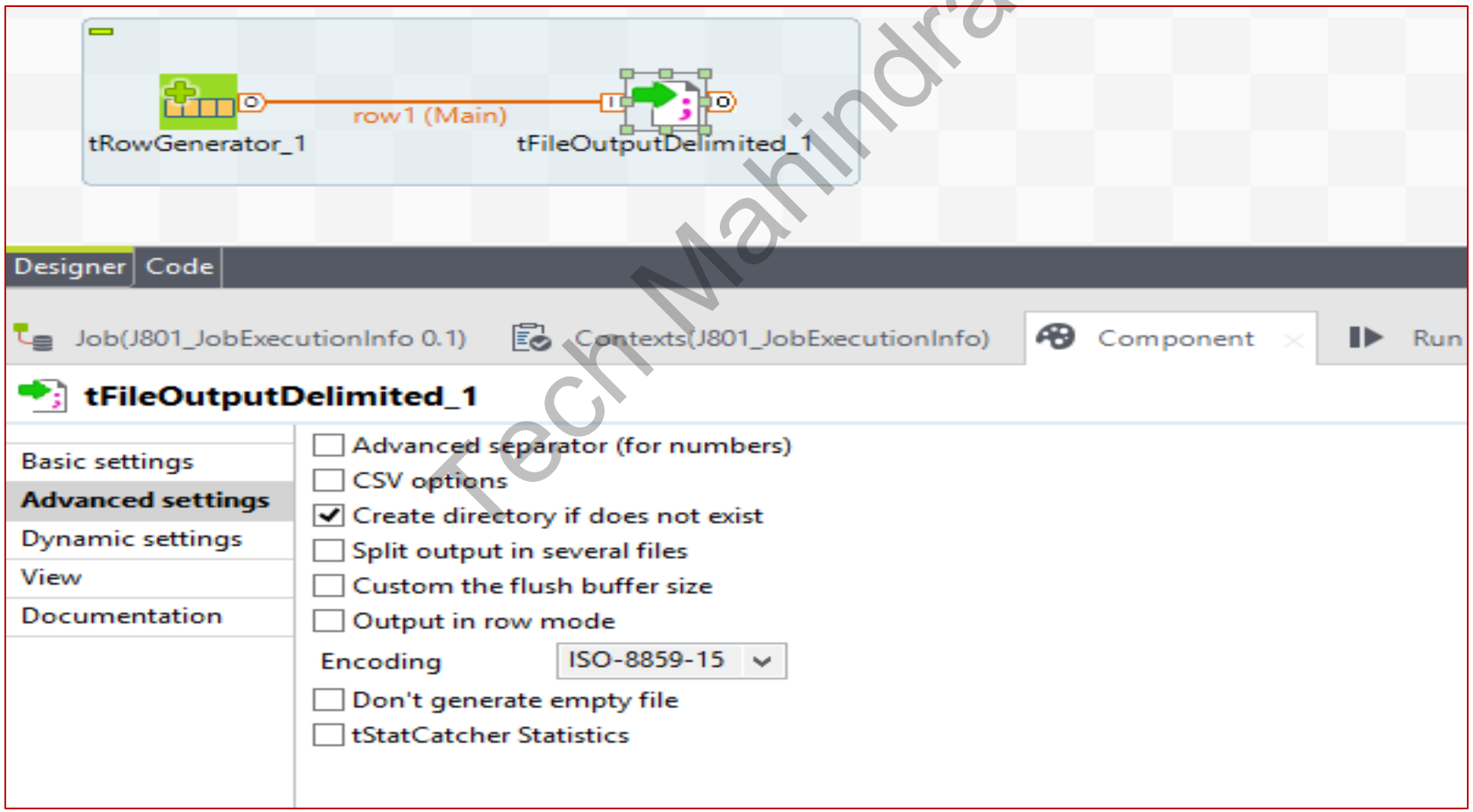
Schema		Functions		Previe
Column	Type	Functions	Environment variables	Preview
EMPNO	Integer	Numeric.random(int,int)	min value=>1101 ; max value=>1990 ;	
ENAME	String	TalendDataGenerator.getLastName()		
SAL	Integer	Numeric.random(int,int)	min value=>3000 ; max value=>99000 ;	



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

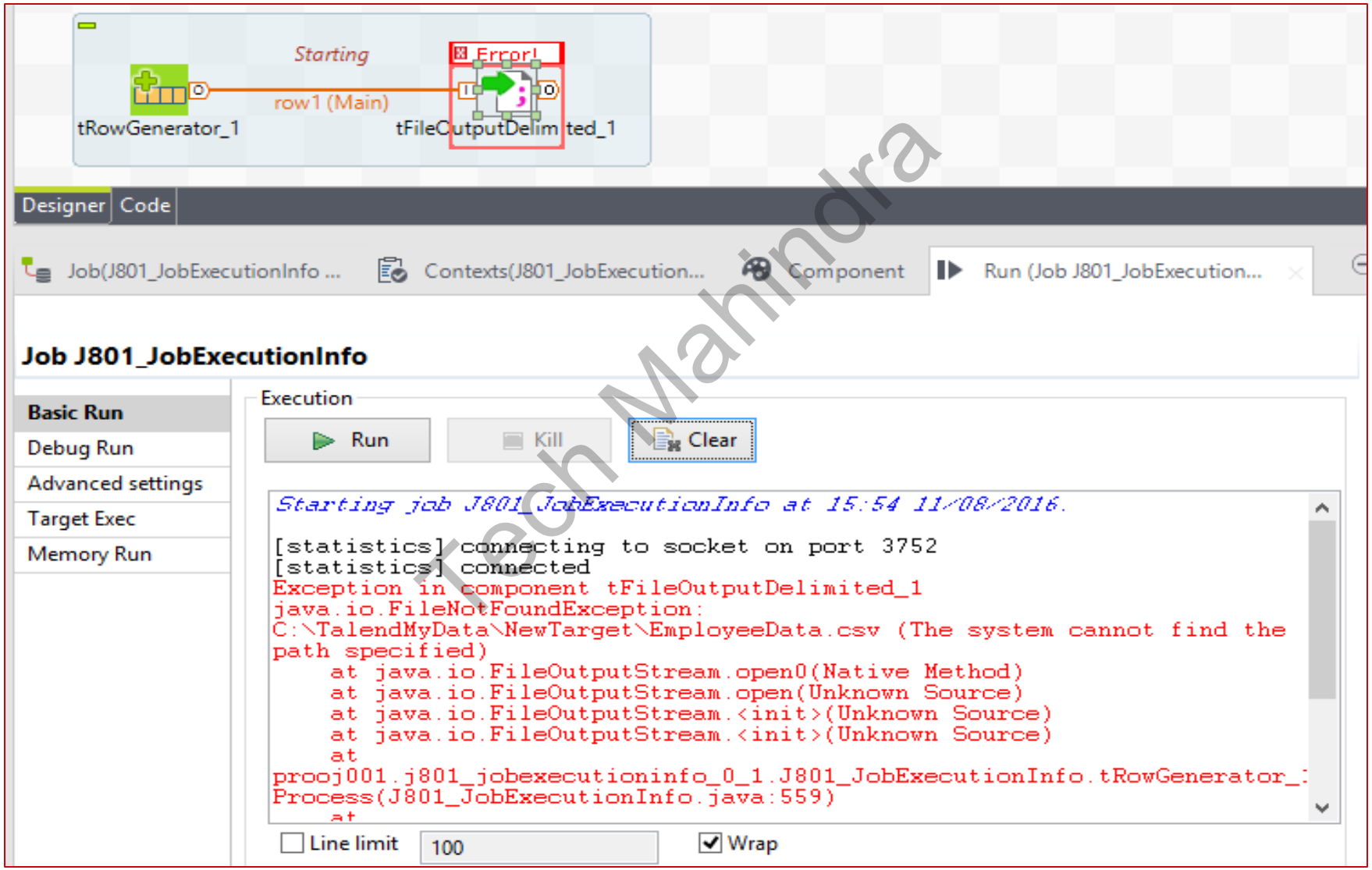
- Add **tFileOutputDelimited** to the Job
- Connect both RowGenerator and FileOutputDelimited
- Create an error – say : **uncheck Create Directory if not exists** in Advanced settings



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

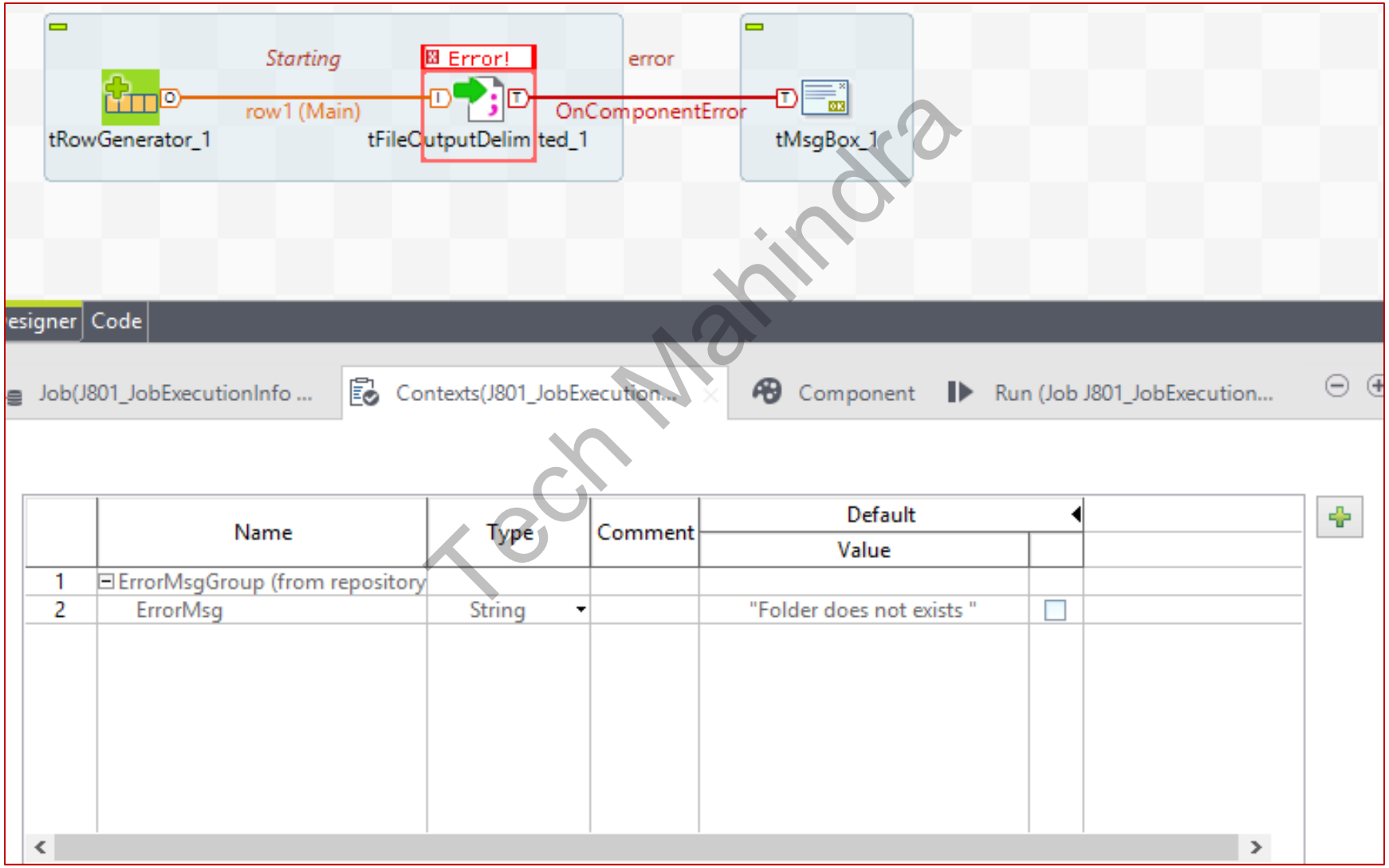
- Run the Job by clicking – **F6**



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

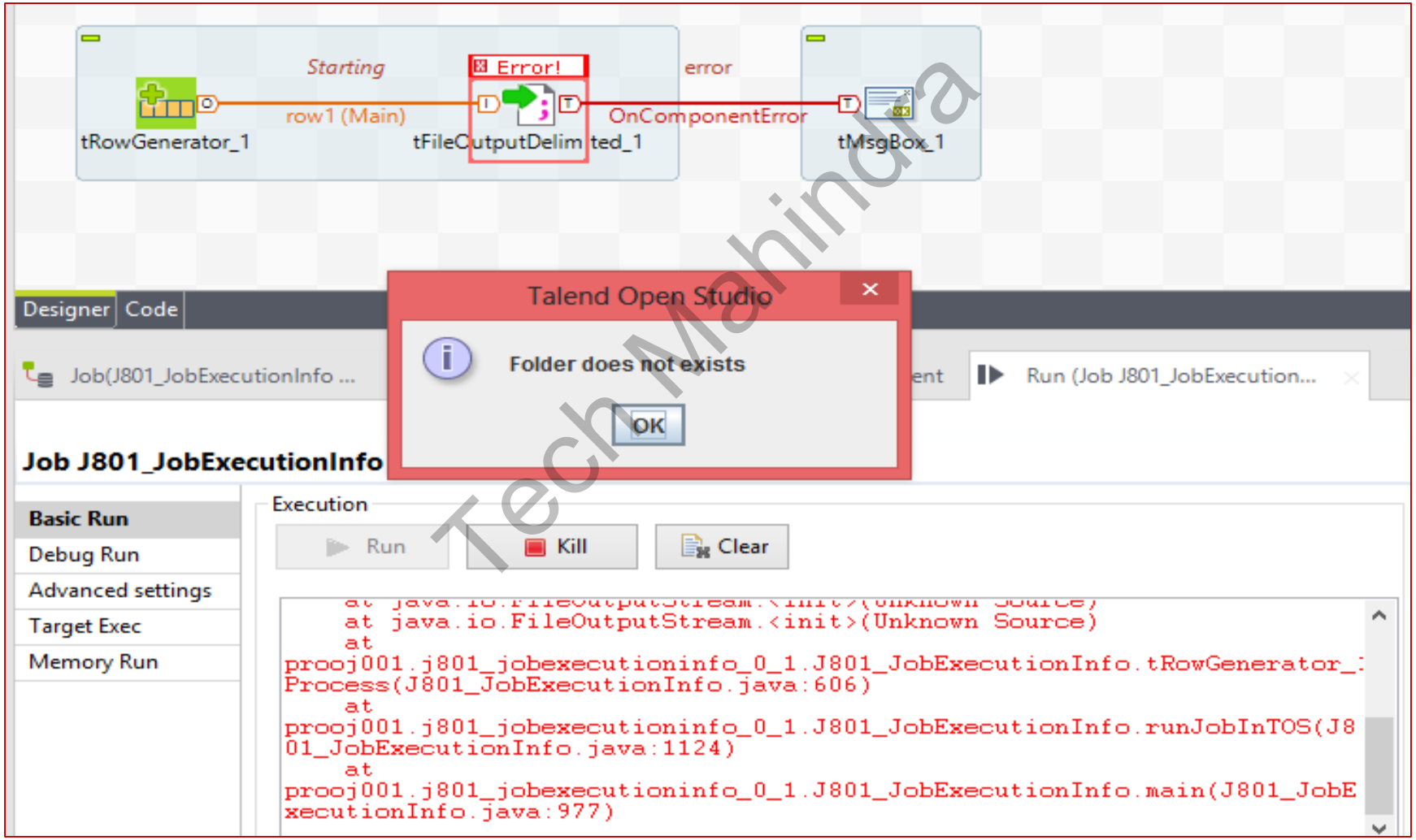
- Add an ErrorMsg in Context and use it in ErrorMsg box



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

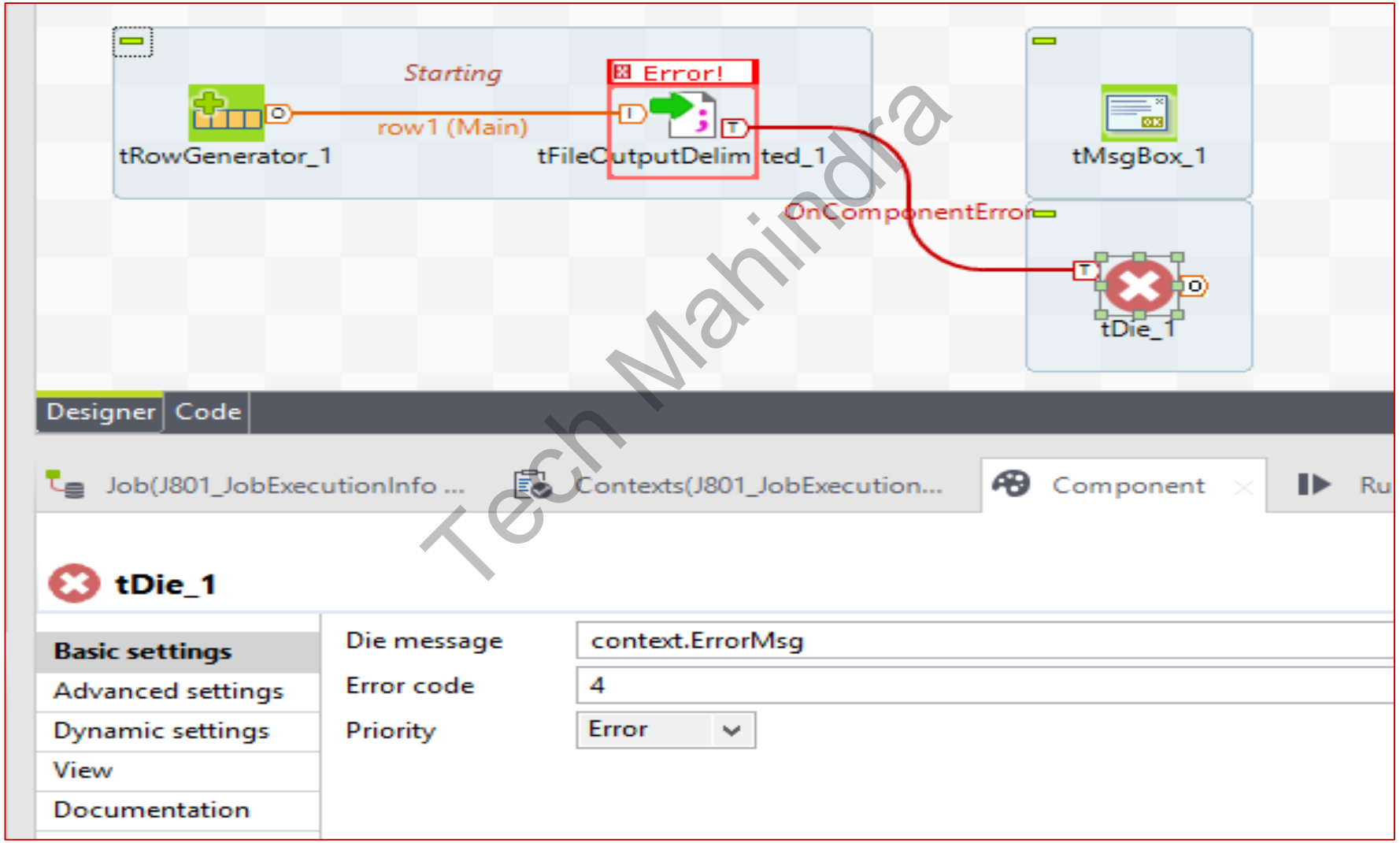
- Add **tMsgBox** to the job
- Connect using **OnComponentError** to FileOutputDelimited



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

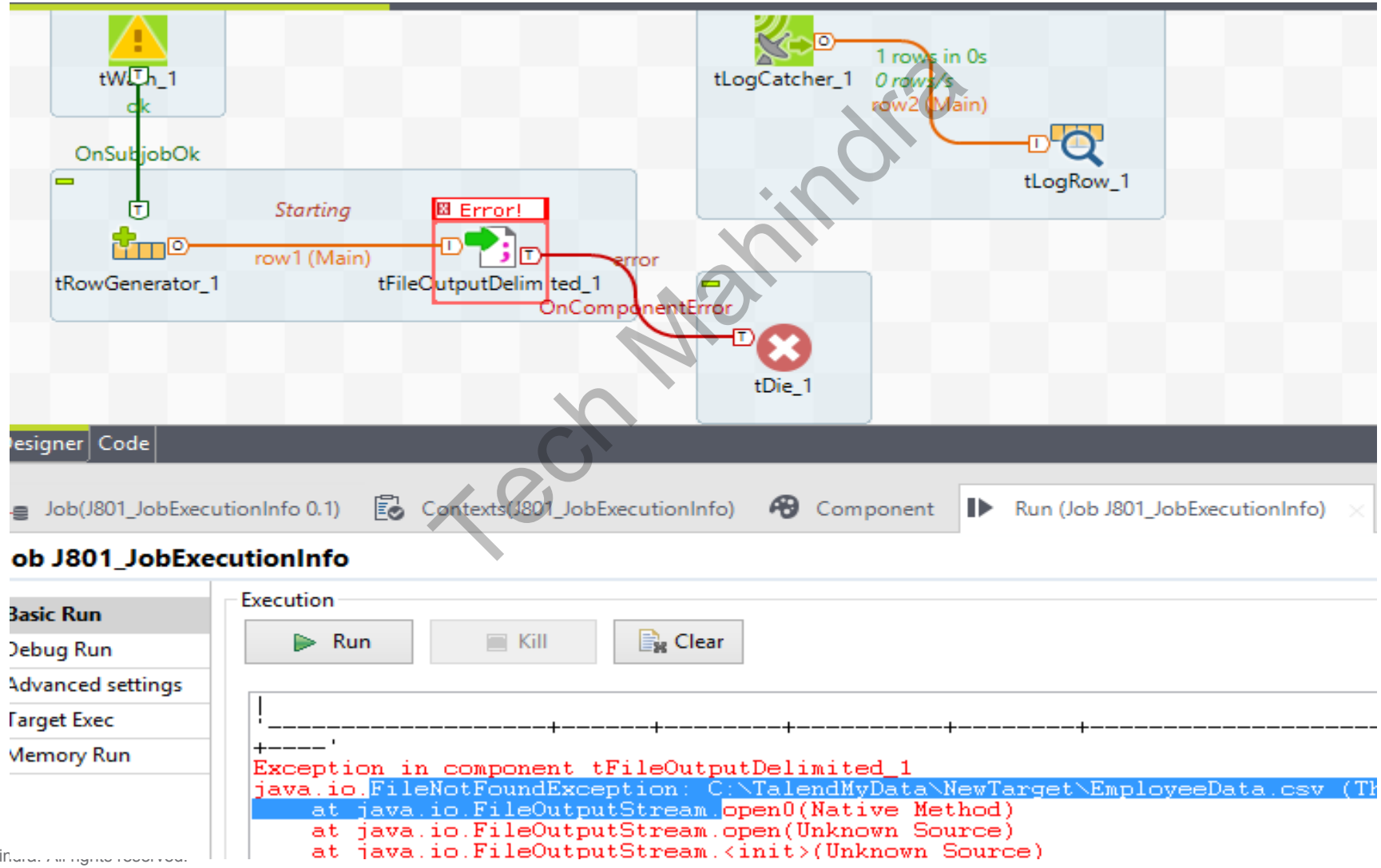
- Drag **tDie** component and connect in place of **tMsgBox**, then delete tMsgBox



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

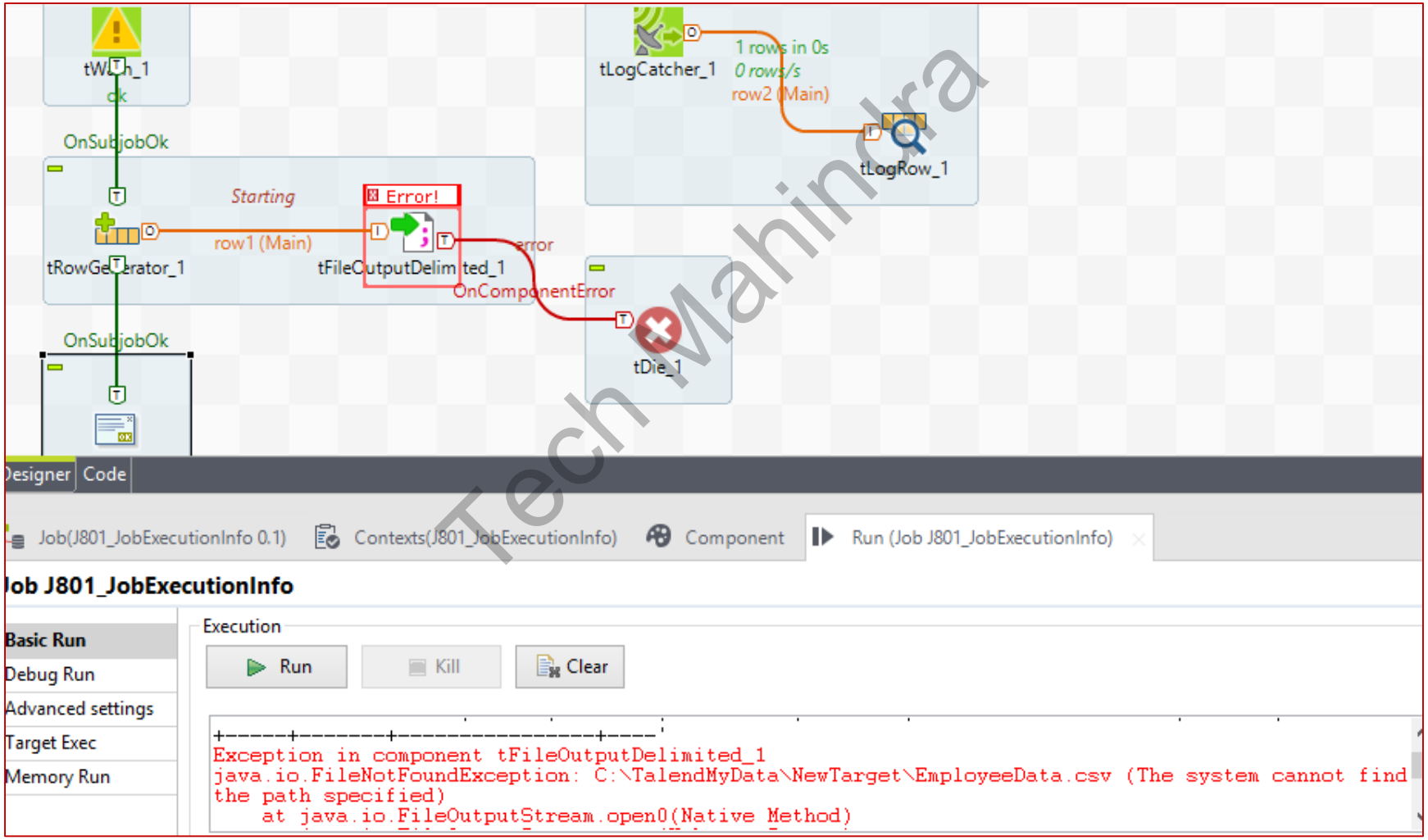
- Add **tWarn** and connect to **tRowGenerator** using **onSubjectOK**
- Add **tLogCatcher** and **tLogRow** as shown
- Run the Job – **F6**



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

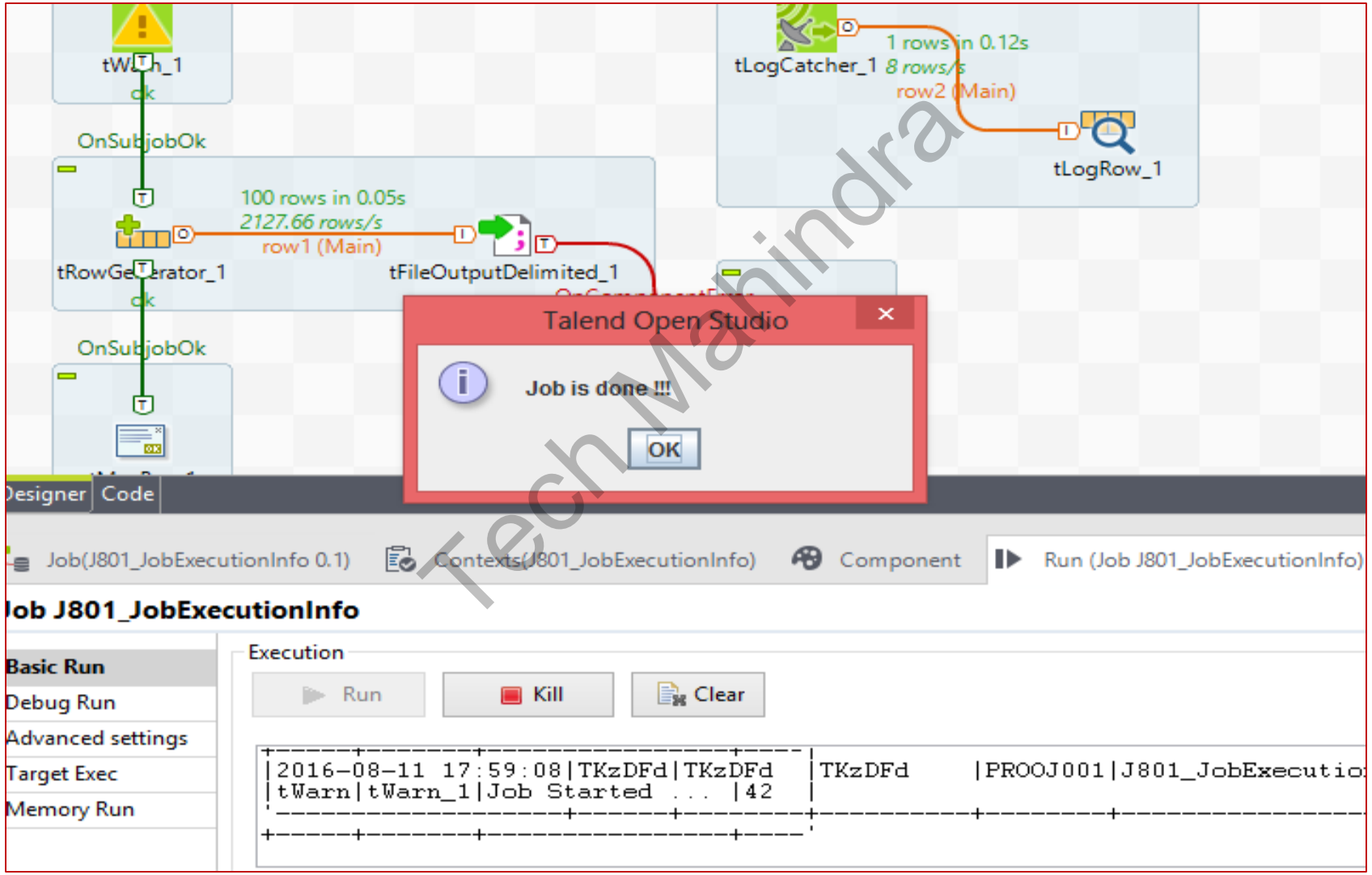
- Add another **tMsgBox** and connect to **tLogRowGenerator** using **OnSubjobOK** – add "**Job is done !!!**" in the MsgBox text
- **Run** the job – **F6**.



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

- Correct the error [Create the directory in **tFileOutputDelimited**]
- Run the Job. You will see the Msg – **Job is done !!!**



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Talend

Oracle database

Tech Mahindra

Sreenivas_Ram

Copyright © 2013 Tech Mahindra. All rights reserved.

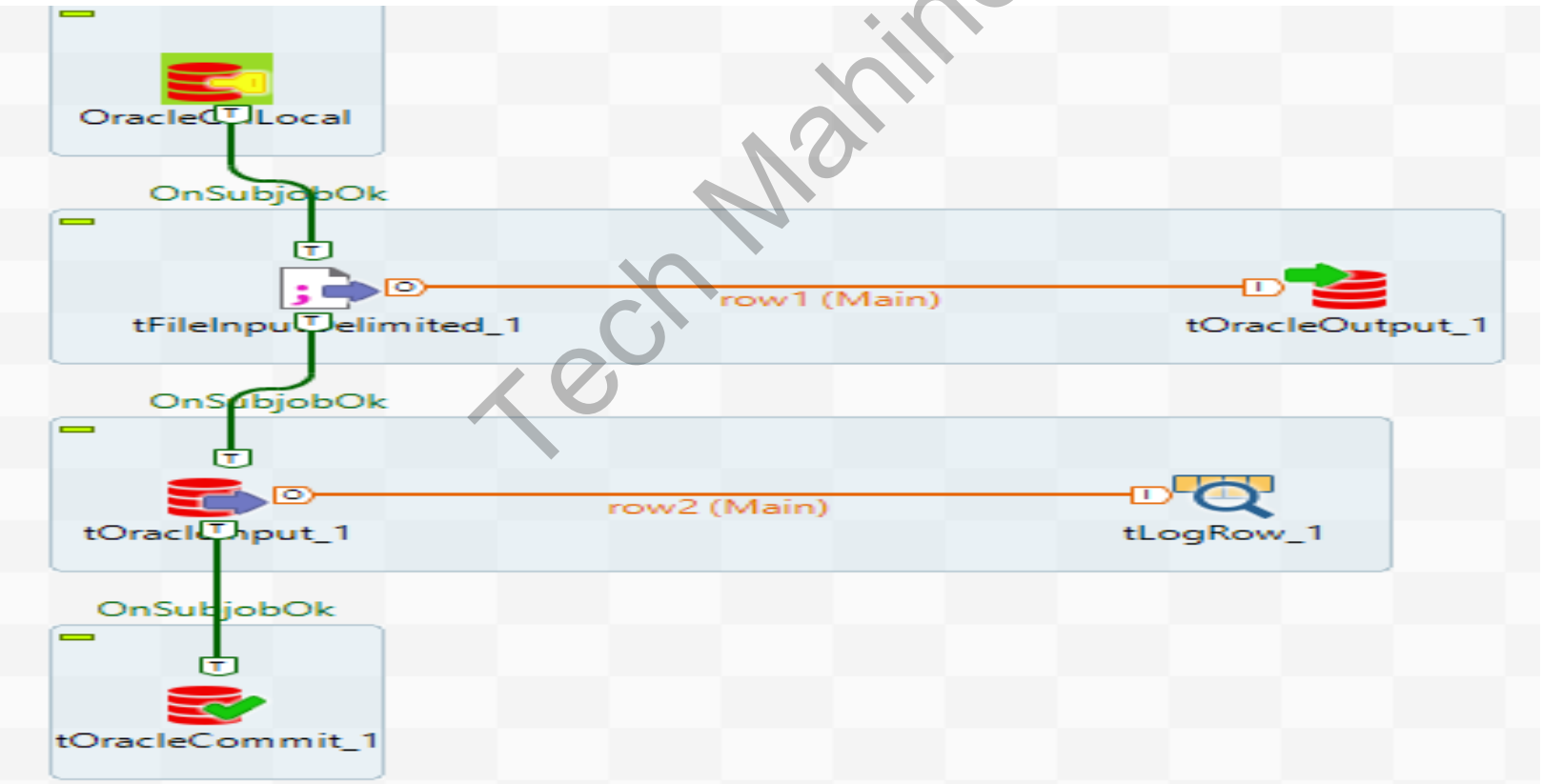
Inserting data into a Oracle table and extracting useful information from it

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Writing and Reading data - Oracle

- Design the job with **tOracleConnection**, **tFileInputDelimited**, **tOracleOutput**, **tOracleInput**, **tLogRow** and **tOrcleCommit** components as follows:

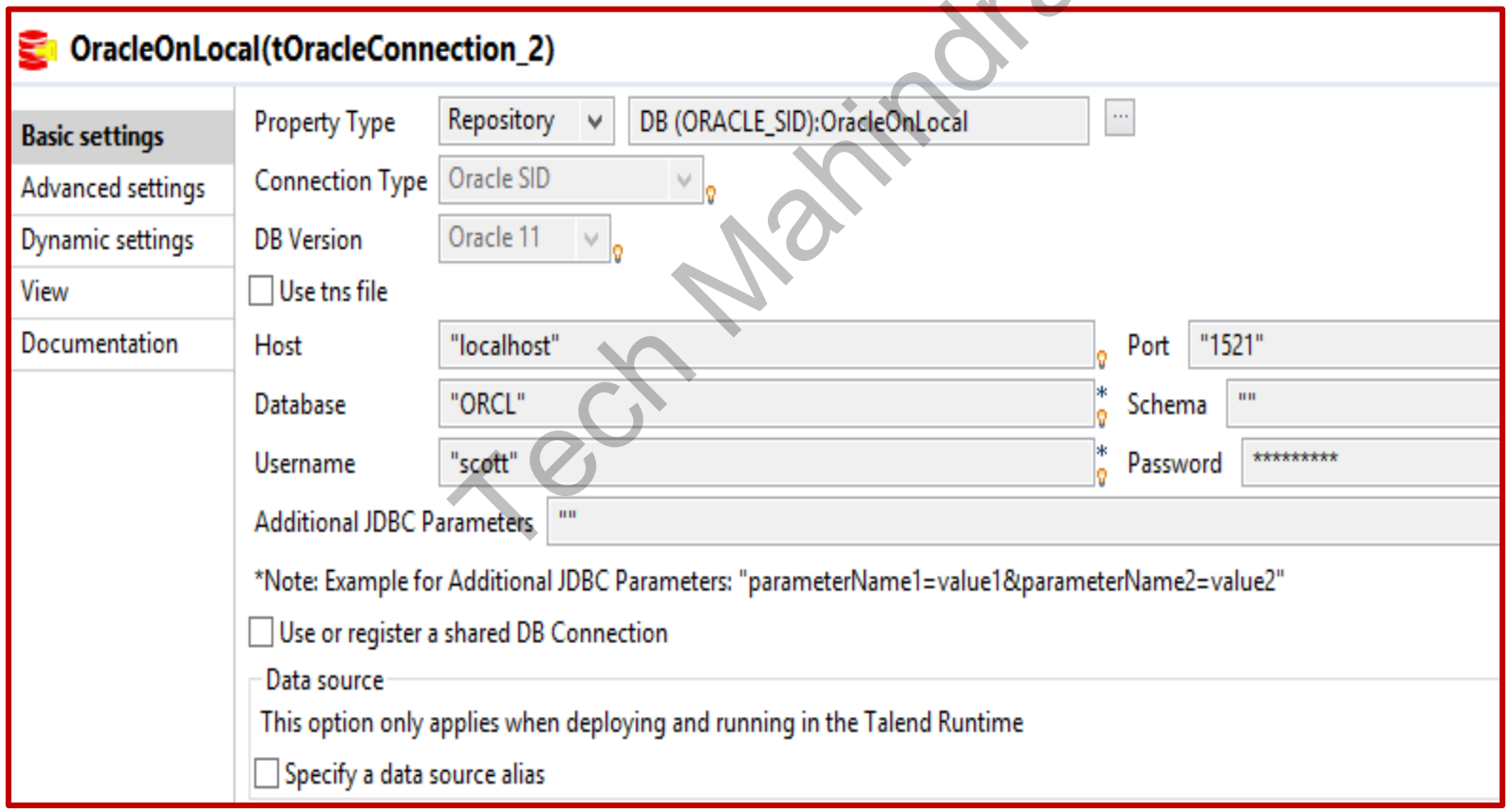


Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Configure the components

- **Oracle Connection** : or : Configure **Metadata** to set **Oracle Connection**



Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Reading data – Writing to ORACLE table

- Sample data - **tFileInputDelimited**

tFileInputDelimited_1

Basic settings | Advanced settings | Dynamic settings | View | Documentation

Property Type: Repository — DELIM:Ora-EmpSample

"When the input source is a stream or a zip file,footer and random shouldn't be bigger than 0."

File name/Stream: "C:/TalendMyData/Source/Ora-SampleEmp.csv"

CSV Row Separator: LF("\n") — Field Separator: ","

CSV options — Escape char: """ — Text enclosure: """

Header: 1 — Footer: 0 — Limit:

Schema: Repository — DELIM:Ora-EmpSample - metadata — Edit schema

Skip empty rows — Uncompress as zip file — Die on error

id	name	wage
51	Harry	2300
40	Ronald	3796
17	Theodore	2174
21	James	1986
2	George	2591
89	Calvin	2362
84	Ulysses	3383
4	Lyndon	2264
17	Franklin	1780
86	Lyndon	3999

tOracleOutput_1

Basic settings | Advanced settings | Dynamic settings | View | Documentation

Use an existing connection — Component List: tOracleConnection_2 - OracleOnLocal

Table: "WagesInfo"

Action on table: Create table if does not exist — Action on data: Insert

Warning : this component configuration will automatically generate a commit before data insert/update/de

Schema: Built-In — Edit schema — Sync columns

Die on error

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Read data - Oracle Input - write to Log

tOracleInput_1

Basic settings | Advanced settings | Dynamic settings | View | Documentation

☑ Use an existing connection — Component List: tOracleConnection_2 - OracleOnLocal *

Schema: Repository — DB (ORACLE_SID):OracleOnLocal - WAGESIN — Edit schema

Table Name: "WAGESINFO"

Query Type: Built-In — Guess Query — Guess schema

Query:

```
"SELECT
 WAGESINFO.\"ID\",
 WAGESINFO.NAME,
 WAGESINFO.WAGE
FROM WAGESINFO
WHERE WAGE >
(SELECT avg(WAGE)
FROM WAGESINFO)
ORDER BY \"ID\"
"
```

tLogRow_1

Basic settings | Advanced settings | Dynamic settings | View | Documentation

Schema: Built-In — Edit schema — Sync columns

Mode
- ○ Basic
- ◉ Table (print values in cells of a table)
- ○ Vertical (each row is a key/value list)

ontent with log4j

tOracleCommit_1

Basic settings | Advanced settings | Dynamic settings

Component List: tOracleConnection_2 - OracleOnLocal *

☑ Close Connection

Tech Mahindra

Copyright © 2013 Tech Mahindra. All rights reserved.

Tech Mahindra

Talend

ORACLE-SCD

Tech Mahindra

Sreenivas_Ram

Copyright © 2016 Tech Mahindra. All rights reserved.

Types of SCD

TYPE – I

- Overwrites old with new data.
- No history is maintained
- No need to change anything

TYPE – II

- Multiple records created
- Data history is maintained
- Can be use like start-date, end-date, or versioning or maintaining a flag

TYPE – III

- Multiple columns are created
- Limited data history is maintained (one level)
- Creating column like – Previous-Col, Current-Col

TYPE – IV

- Separate history table is maintained

TYPE – VI

- Hybrid (1 + 2 + 3)

Tech Mahindra

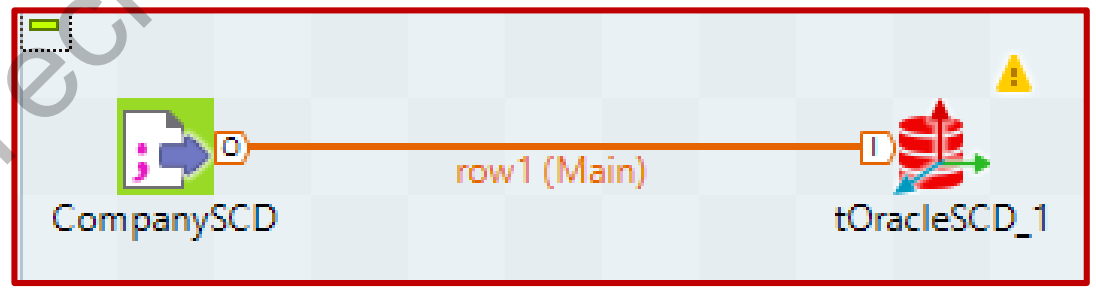
Copyright © 2016 Tech Mahindra. All rights reserved.

SCD-Type-I

- Change the data in the current record itself by overwriting

- Drag and drop CSV Data as source to CompanySCD
- Select InputFileDelimited details with **Compnay** Table in **ORCLESCD**
- In the SCD Editor – Map **CompanyID** as **SourceKey** and **CompanyName** as **type-1 field** changes

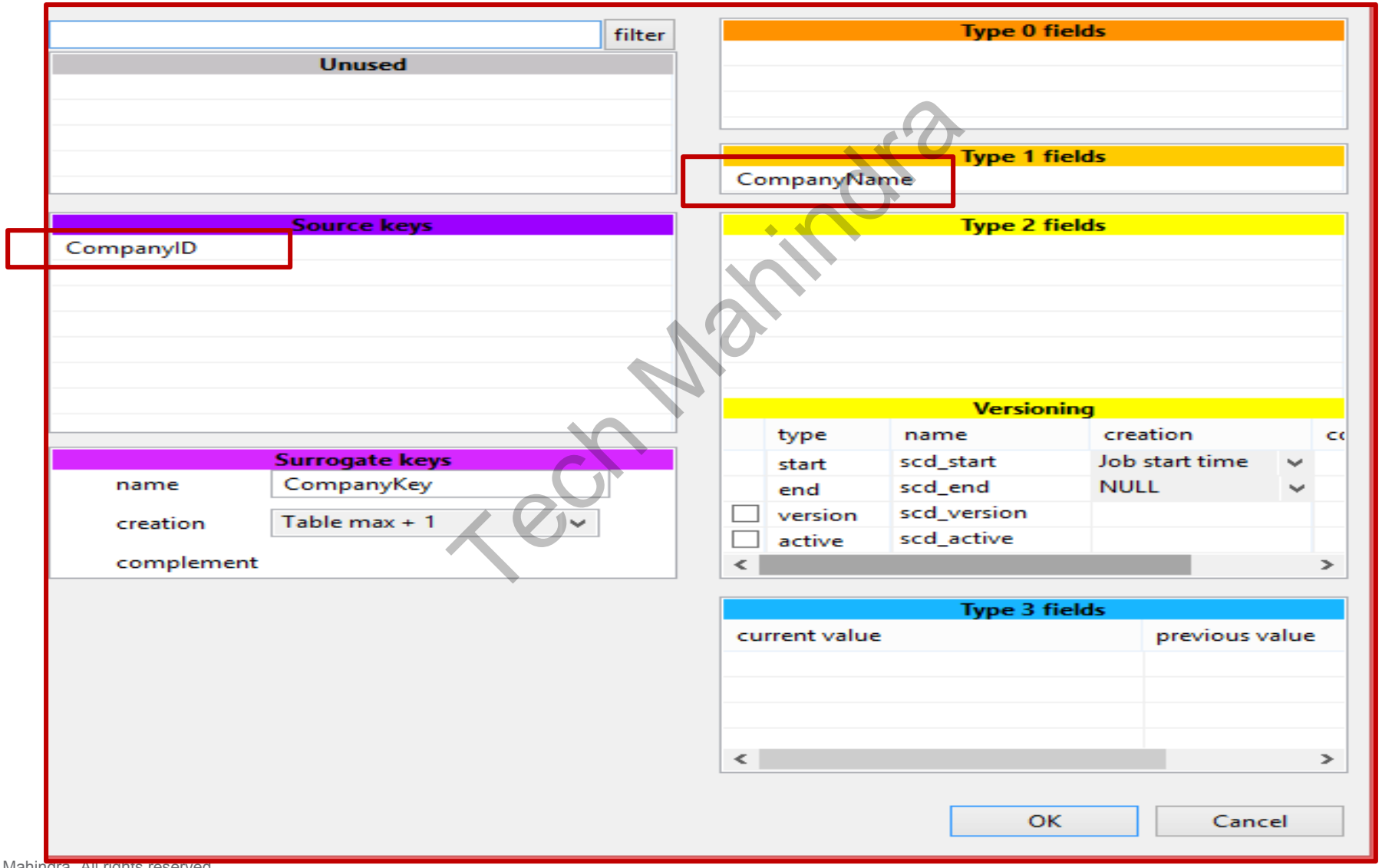
CompanyID	CompanyName
COM001	TCS
COM002	ONGC
COM003	Reliance
COM004	ITC
COM005	Coal India
COM006	Infosys
COM007	HDFC Ltd
COM008	SBI
COM009	Tata
COM010	Wipro



Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD Editor for Type-I Changes



Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Changes made in Company Table

- Before and after execution

Compa...	CompanyID	COMPANYNAME
1	COM001	TCS
2	COM002	ONGC
3	COM003	Reliance
4	COM004	ITC
5	COM005	Coal India
6	COM006	Infosys
7	COM007	HDFC Ltd
8	COM008	SBI
9	COM009	Tata
10	COM010	Wipro

CompanyKey	CompanyID	COMPANYNAME
1	COM001	TCS
2	COM002	ONGC
3	COM003	Reliance
4	COM004	ITC Ltd
5	COM005	Coal India
6	COM006	Infosys
7	COM007	HDFC Ltd
8	COM008	SBI
9	COM009	Tata Co Ltd
10	COM010	Wipro

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

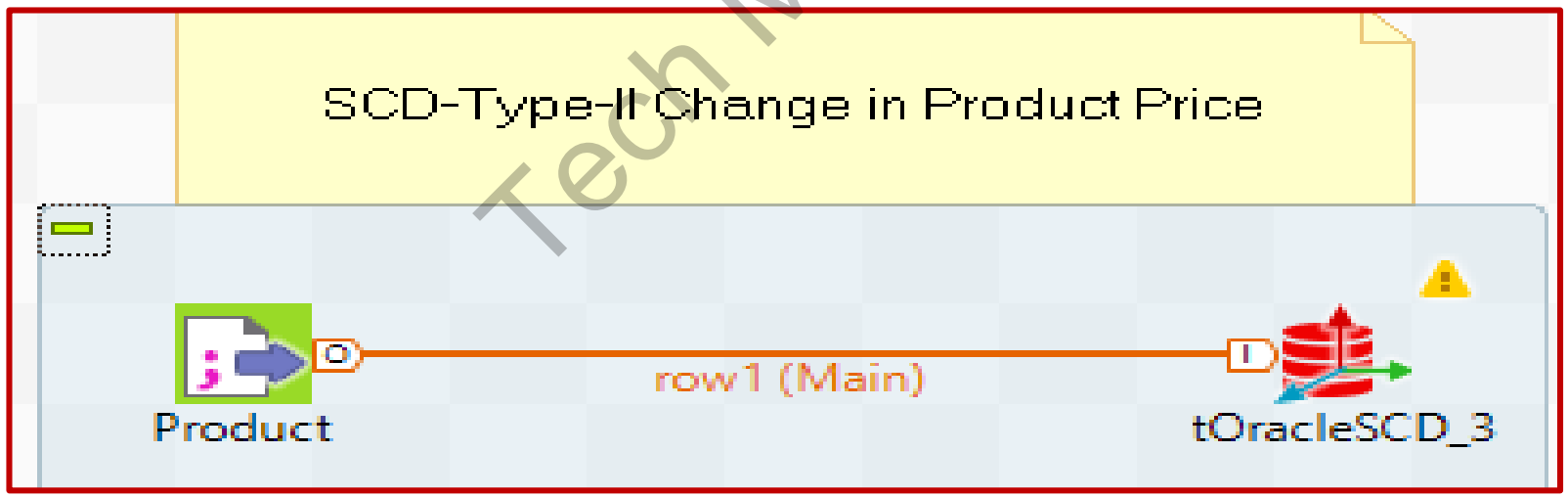
SCD – Type-II

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD-Type-II

- Multiple records created and the data history is maintained
- Drag and drop CSV Data as source - **Products**
- Select InputFileDelimited details with **DimProducts** Table in **ORACLESCD**
- In the SCD Editor – Map **ProductID** as **SourceKey** and ProductName**,** as type-0 field changes, **Price** as **Type-2 field changes** add **Productkey** as **Surrogate Key** Leave Start_Date and En_Date as it is.

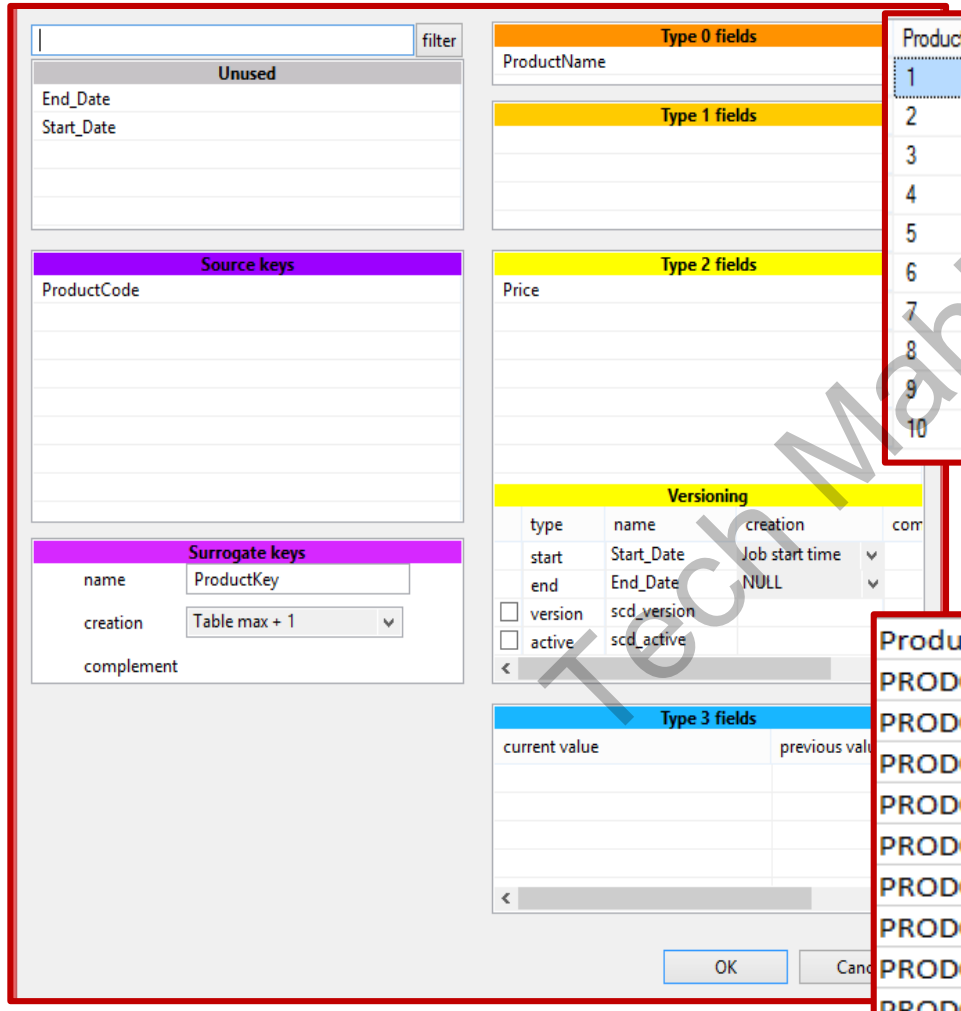


Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD-Type-II … contd…

- The original data of **DimProducts**



ProductKey	ProductCode	ProductName	Price	Start_Date	End_Date
1	PROD001	LG Nexus 5	22000	2016-08-18	NULL
2	PROD002	Samsung Galaxy Note 2	15000	2016-08-18	NULL
3	PROD003	Nokia Lumia 1111	9999	2016-08-18	NULL
4	PROD004	Sony Xoeria Z	20000	2016-08-18	NULL
5	PROD005	HTC One Mini	9999	2016-08-18	NULL
6	PROD006	Samsung Galaxy Core Duos	9999	2016-08-18	NULL
7	PROD007	Nokia Lumia 625	9999	2016-08-18	NULL
8	PROD008	Sony Xeria L	9888	2016-08-18	NULL
9	PROD009	Micromax Canvas	5555	2016-08-18	NULL
10	PROD010	HTC One V	8888	2016-08-18	NULL

- Changes in **Price** (**csv**)

ProductCc	ProductName	Price	Start_Date	End_Date
PROD001	LG Nexus 5	20000	10/15/2013	NULL
PROD002	Samsung Galax	12000	10/15/2013	NULL
PROD003	Nokia Lumia 11	9999	10/15/2013	NULL
PROD004	Sony Xoeria Z	20000	10/15/2013	NULL
PROD005	HTC One Mini	9000	10/15/2013	NULL
PROD006	Samsung Galax	9999	10/15/2013	NULL
PROD007	Nokia Lumia 62	9000	10/15/2013	NULL
PROD008	Sony Xeria L	8000	10/15/2013	NULL
PROD009	Micromax Canv	5555	10/15/2013	NULL
PROD010	HTC One V	8888	10/15/2013	NULL

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD-Type-II … contd…

- Changes reflected in **DimProducts** table [**MSMSQL**]

ProductKey	ProductCode	ProductName	Price	Start_Date	End_Date
1	PROD001	LG Nexus 5	22000	2016-08-18	2016-08-18
2	PROD002	Samsung Galaxy Note 2	15000	2016-08-18	2016-08-18
3	PROD003	Nokia Lumia 1111	9999	2016-08-18	NULL
4	PROD004	Sony Xoeria Z	20000	2016-08-18	NULL
5	PROD005	HTC One Mini	9999	2016-08-18	2016-08-18
6	PROD006	Samsung Galaxy Core Duos	9999	2016-08-18	NULL
7	PROD007	Nokia Lumia 625	9999	2016-08-18	2016-08-18
8	PROD008	Sony Xeria L	9888	2016-08-18	2016-08-18
9	PROD009	Micromax Canvas	5555	2016-08-18	NULL
10	PROD010	HTC One V	8888	2016-08-18	NULL
11	PROD001	LG Nexus 5	20000	2016-08-18	NULL
12	PROD002	Samsung Galaxy Note 2	12000	2016-08-18	NULL
13	PROD005	HTC One Mini	9000	2016-08-18	NULL
14	PROD007	Nokia Lumia 625	9000	2016-08-18	NULL
15	PROD008	Sony Xeria L	8000	2016-08-18	NULL

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD – Type-III

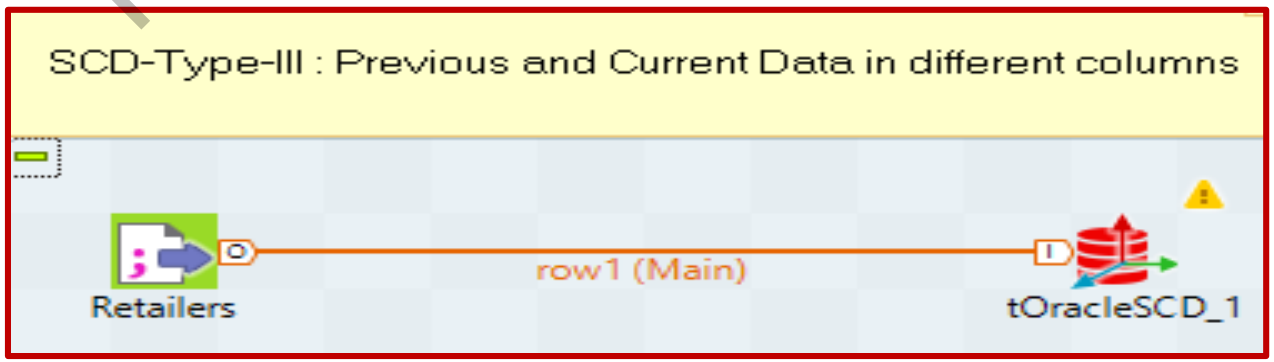
Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD – Type-III

- To maintain history
  - Using a date (active date)
  - Versioning
  - IsActive (flag)

- Design the job as follows:
- Drag Retailers (csv) – **tFileInputDelimited** and t**ORACLESCD**
- Make the following changes in SCD editor
- **RetailerID** – SourceKey
- **Company** and **RName** in Type-0 – No changes
- **RetailerKey** – Surrogate Key
- Add **Start_Date** and **End_Date** in Versioning in place of original columns



Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD – Type-III … contd.

- SCD Designer

SCD component editor

filter

Unused

Source keys
RetailerID

Surrogate keys
name RetailerKey
creation Table max + 1
complement

Type 0 fields
Country
RName

Type 1 fields

Type 2 fields

Versioning

type	name	creation	complement
start	scd_start	Job start time	
end	scd_end	NULL	
version	scd_version		
active	scd_active		

Type 3 fields

current value	previous value
State	Prev_State

OK Cancel

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

SCD – Type-III … contd.

- Original Source Data

RetailerID	Name	State	Country
RET0001	Reliance	Maharastra	India
RET0002	Pantaloons	Maharastra	India
RET0003	Provogue	Maharastra	India
RET0004	Shooper Stp	Maharastra	India
RET0005	ITC	West Bengal	India
RET0006	Trent	Maharastra	India
RET0007	McDonalds	Illinois	
RET0008	Aditya Birla	Maharast	
RET0009	Titan	Karnataka	
RET0010	Kewel Kiran	Maharast	

- **Updated data**

RetailerKey	RetailerID	RName	State	Country	Prev_State
1	RET0001	Reliance	Maharastra	India	NULL
2	RET0002	Pantaloons	Telangana	India	Maharastra
3	RET0003	Provogue	Maharastra	India	NULL
4	RET0004	Shooper Stp	Andhra Pradesh	India	Maharastra
5	RET0005	ITC	West Bengal	India	NULL
6	RET0006	Trent	Maharastra	India	NULL
7	RET0007	McDonalds	Illinois	USA	NULL
8	RET0008	Aditya Birla	Maharastra	India	NULL
9	RET0009	Titan	Karnataka	India	NULL
10	RET0010	Kewel Kiran	Maharastra	India	NULL

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Talend

Printing out a variable content using tJava component

Tech Mahindra

Sreenivas_Ram

Copyright © 2016 Tech Mahindra. All rights reserved.

tJava Ex:

- The Job aims at printing out the number of lines being processed using a Java command and the global variable
- Design the job as follows: - using **tFileInputDelimited**, **tFileOutputExcel** and **tJava** components.

- Connect the **tFileInputDelimited** to the **tFileOutputExcel** using a **Row Main** connection. The content from a delimited txt file will be passed on through the connection to an xls-type of file without further transformation

- Then connect the **tFileInputDelimited** component to the **tJava** component using a **Trigger > On Subjob Ok** link

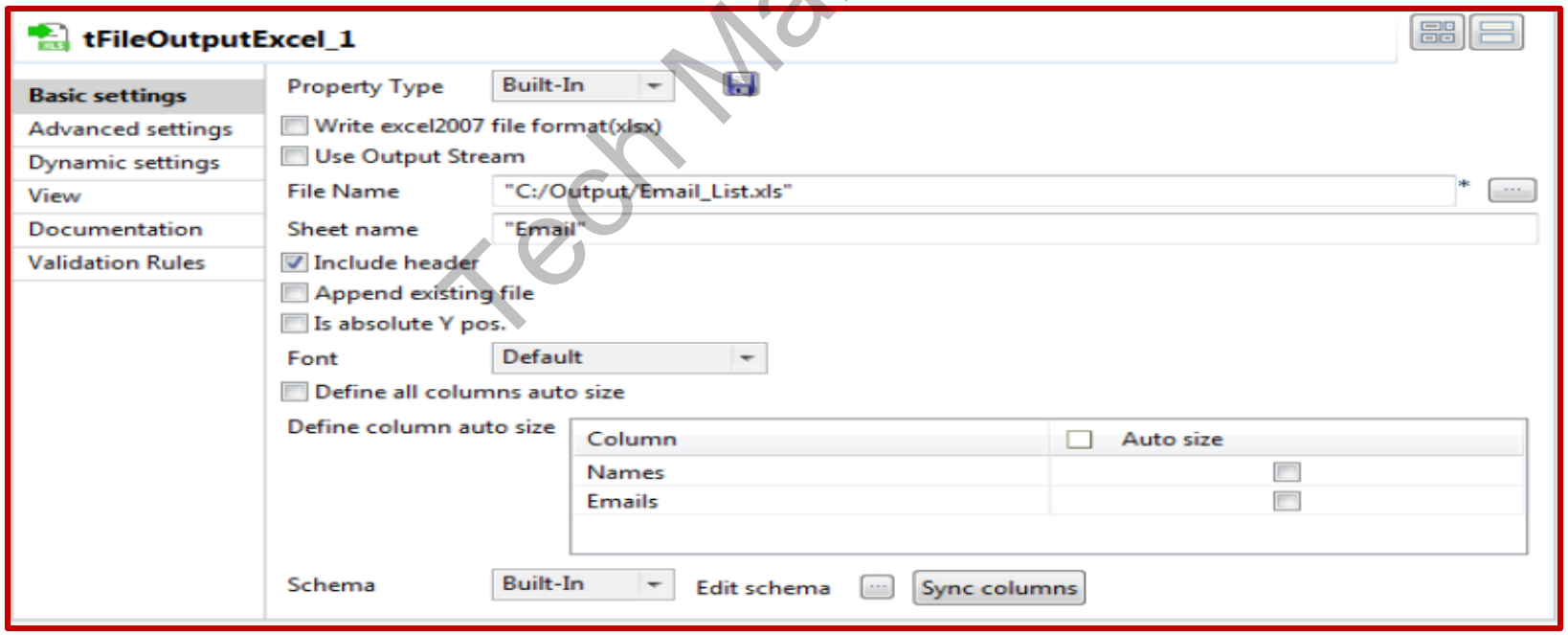
Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Properties of Components

- **tFileInputDelimited** : **1.** In File name field, enter the path and file name
- **2.** Asimple text file made of **two columns**: ***Names*** and their ***Emails***
- Click the **Edit Schema** button, and set schema accordingly

- **tFileOutputExcel :**
- Enter File Name an define schema.

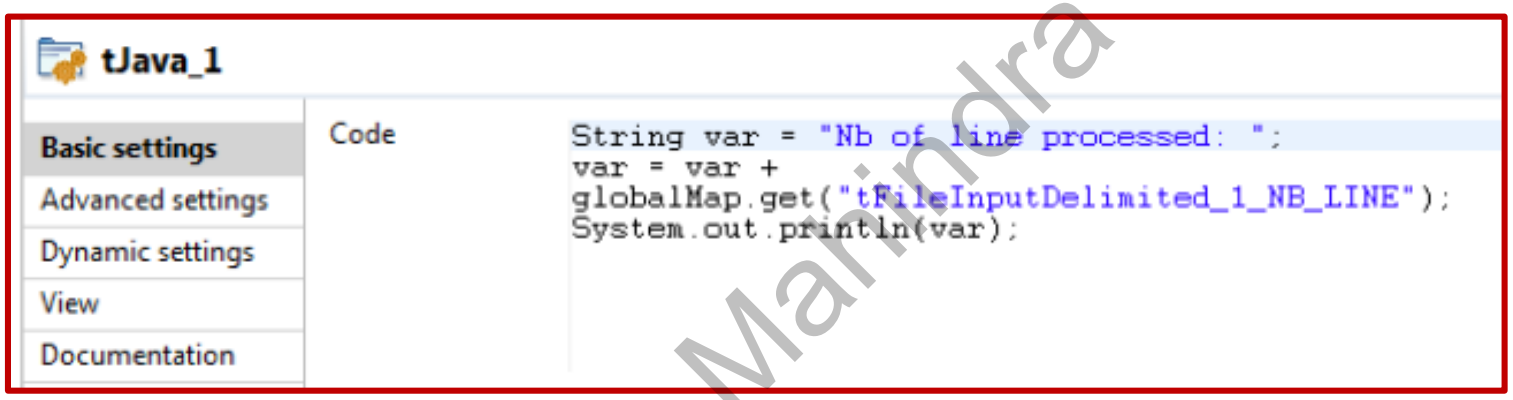


Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Properties of Components … contd …

- In the code area :



- Execute the job

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Tech Mahindra

Talend
Row-wise processing using
tJavaRow Component

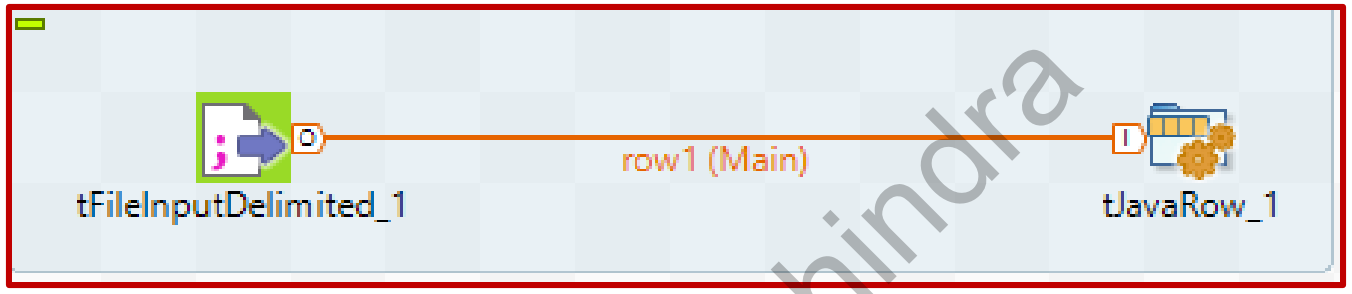
Tech Mahindra

Sreenivas_Ram

Copyright © 2016 Tech Mahindra. All rights reserved.

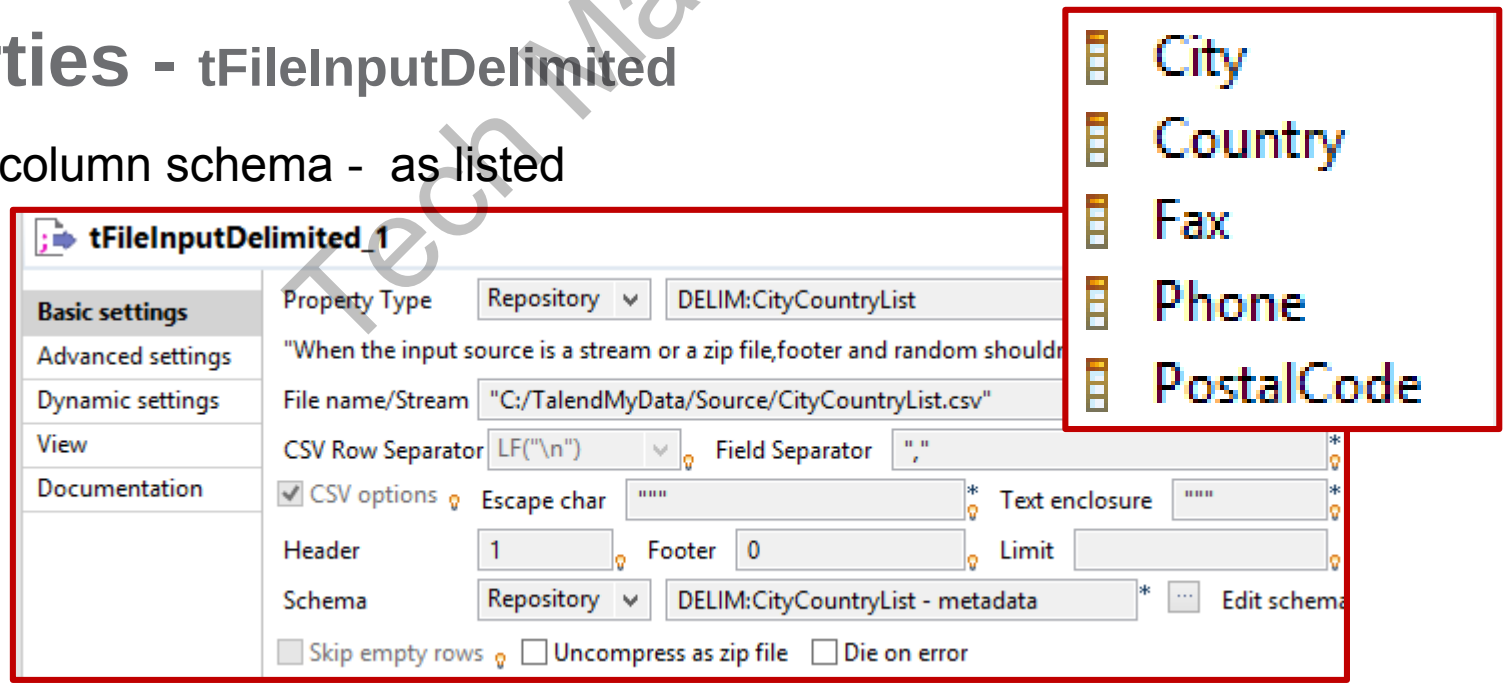
tJavaRow

- We can process each row of the flow using Java code



Properties - tFileInputDelimited

- We use column schema - as listed

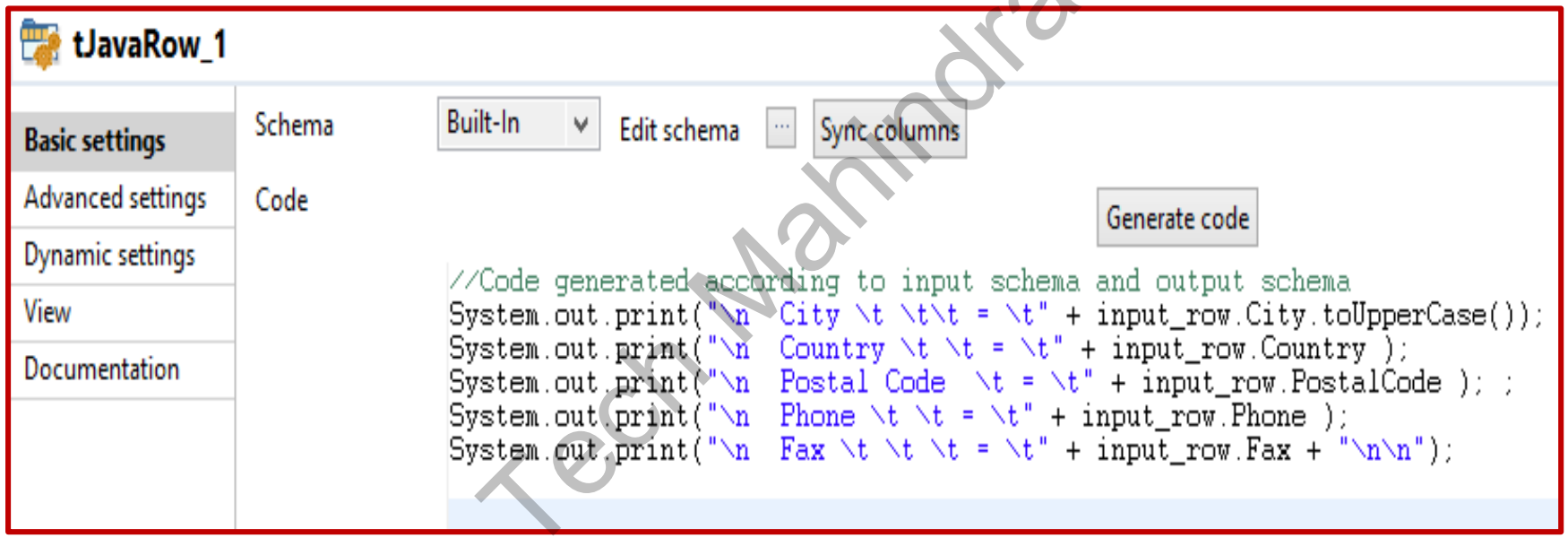


Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Properties

- **tJavaRow :** Enter the following **Java Code** - to read multiple lines of input and print the same



- Execute the Job

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Tech Mahindra

Talend

Multi-part Java Code processing using

tJavaFlex component

Tech Mahindra

Sreenivas_Ram

Copyright © 2016 Tech Mahindra. All rights reserved.

Code to process in three setps

- Sample code :

tJavaFlex_1

Basic settings
Advanced settings
Dynamic settings
View
Documentation

Schema: Built-In — Edit schema

☑ Data Auto Propagate

Start code

```
// start part of your Java code

System.out.println("## Start ## \n");
String [] nameArray = {"Berlin", "Mexico",
"Strasbourg", "Buenos Aires", "London" };

for(int i = 0; i<nameArray.length; i++)
{
```

Main code

```
// here is the main part of the component,
// a piece of code executed in the row
// loop

row1.key = i + 1;
row1.value = nameArray[i];
```

End code

```
// end of the component, outside/closing the

}

System.out.println("\n## End ## \n ");
```

Tech Mahindra

Copyright © 2016 Tech Mahindra. All rights reserved.

Thank you

Tech Mahindra

Disclaimer

Tech Mahindra Limited, herein referred to as TechM provide a wide array of presentations and reports, with the contributions of various professionals. These presentations and reports are for informational purposes and private circulation only and do not constitute an offer to buy or sell any securities mentioned therein. They do not purport to be a complete description of the markets conditions or developments referred to in the material. While utmost care has been taken in preparing the above, we claim no responsibility for their accuracy. We shall not be liable for any direct or indirect losses arising from the use thereof and the viewers are requested to use the information contained herein at their own risk. These presentations and reports should not be reproduced, re-circulated, published in any media, website or otherwise, in any form or manner, in part or as a whole, without the express consent in writing of TechM or its subsidiaries. Any unauthorized use, disclosure or public dissemination of information contained herein is prohibited. Unless specifically noted, TechM is not responsible for the content of these presentations and/or the opinions of the presenters. Individual situations and local practices and standards may vary, so viewers and others utilizing information contained within a presentation are free to adopt differing standards and approaches as they see fit. You may not repackage or sell the presentation. Products and names mentioned in materials or presentations are the property of their respective owners and the mention of them does not constitute an endorsement by TechM. Information contained in a presentation hosted or promoted by TechM is provided “as is” without warranty of any kind, either expressed or implied, including any warranty of merchantability or fitness for a particular purpose. TechM assumes no liability or responsibility for the contents of a presentation or the opinions expressed by the presenters. All expressions of opinion are subject to change without notice.

Copyright © 2016 Tech Mahindra. All rights reserved.

Tech Mahindra

Tech Mahindra