DMETL PROJECT REPORT

BY-1) DHANWANTARI R. ROKADE 2) SAPTARSHI JANA

3) ANUSHA JADAV.

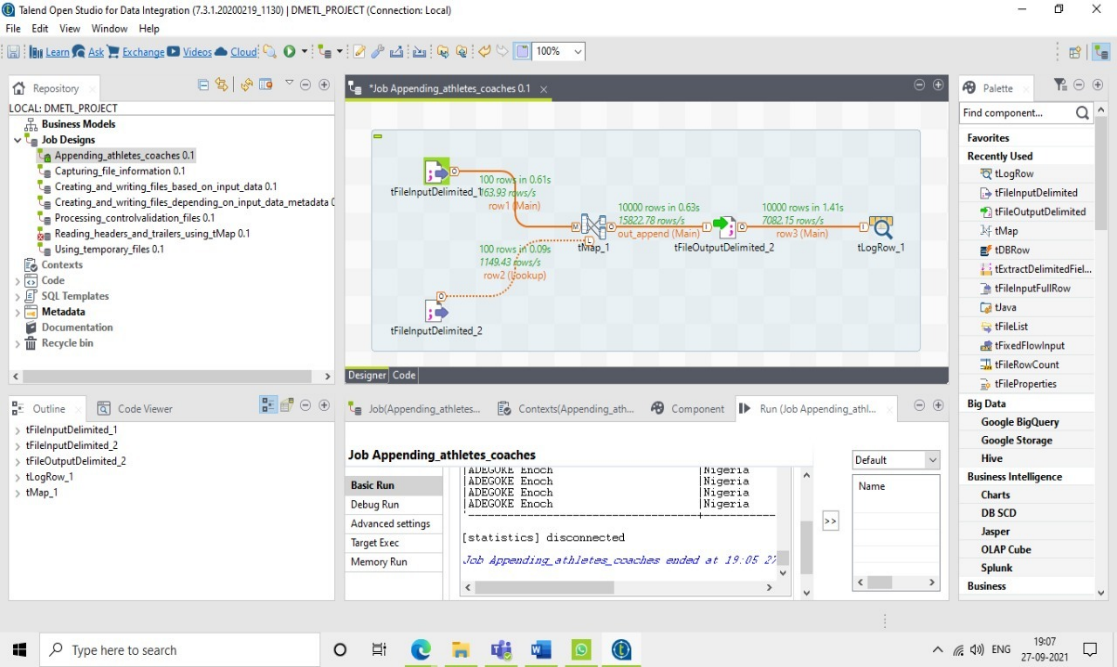
TOPIC-MANAGING FILES

**Introduction-**

It isn't very efficient to process large batches of information via a web service, nor is it particularly desirable to pull data from an application database during peak hours. Thus, many organizations still maintain a file-based overnight batch processes using large extracts in file format.

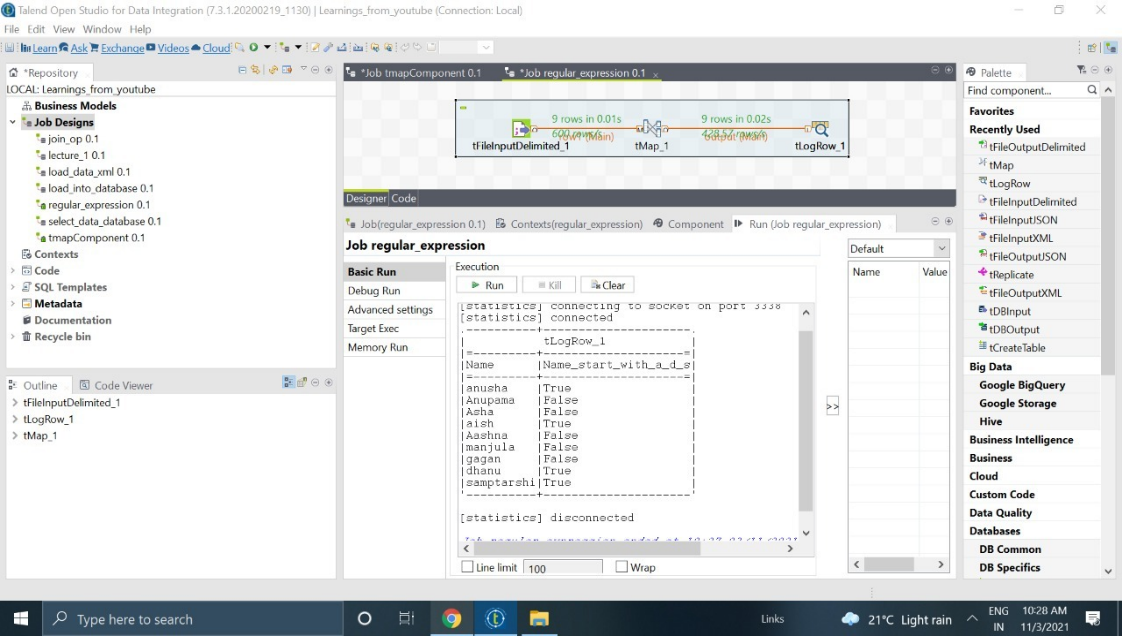
In addition, many older, legacy applications rely solely on file-based data for communicating with the outside world.

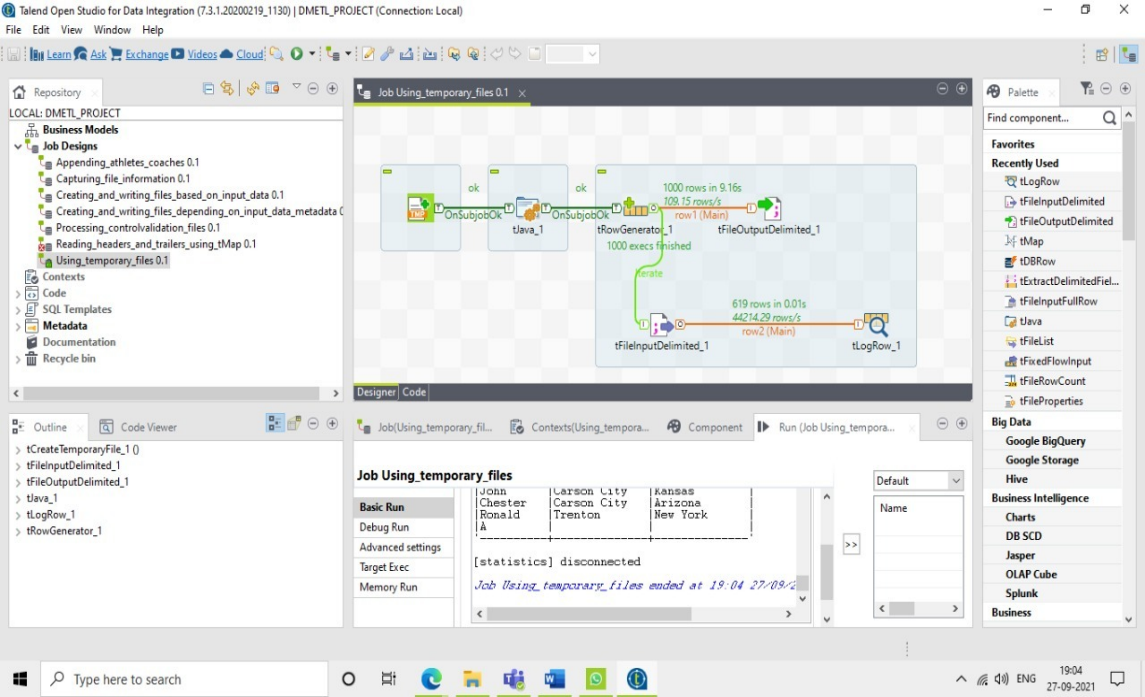
It is therefore very important for the data integration developer to understand many file types and be able to manage them efficiently and effectively.

******Appending records to a file**

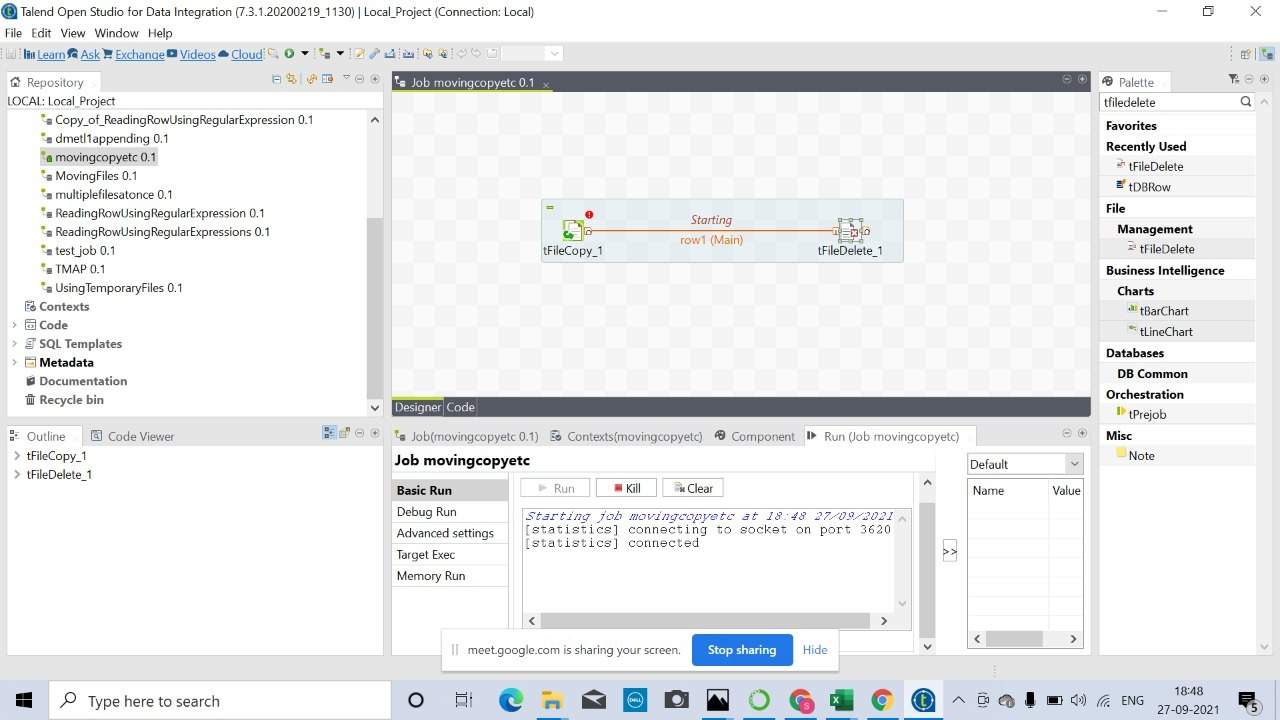
This shows how a file can be built in within different sub jobs by appending data to an existing file. The append method is one way of building complex files.

**Reading rows using a regular expression**

Regular expression (regex) is a powerful method for pattern matching and replacement within many programming languages. One interesting use for regular expressions is when dealing with unusual input formats that are difficult to describe using normal delimited or fixed-width file formatting.

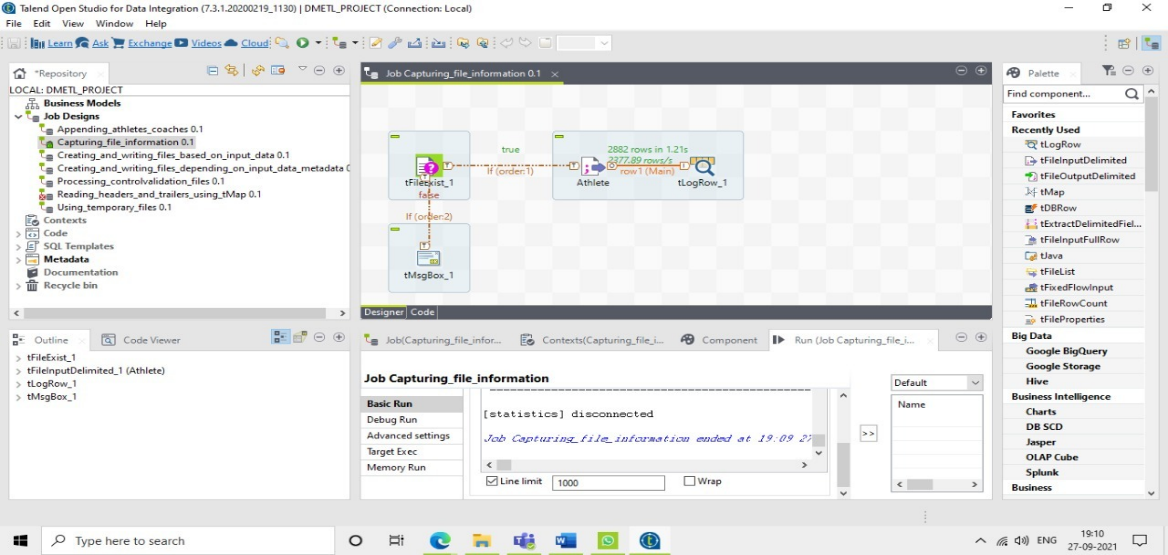
**Using temporary files**

Occasionally, it is necessary to create intermediate files within a job that are only used during the lifetime of the job.

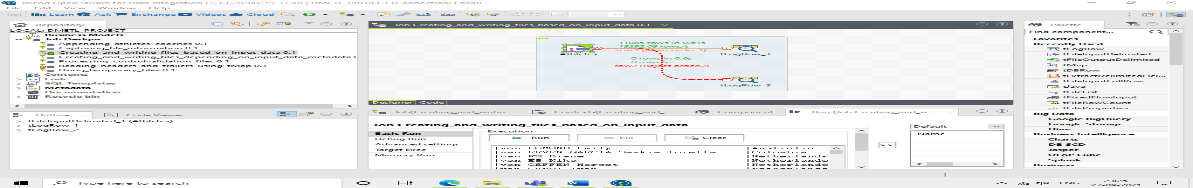
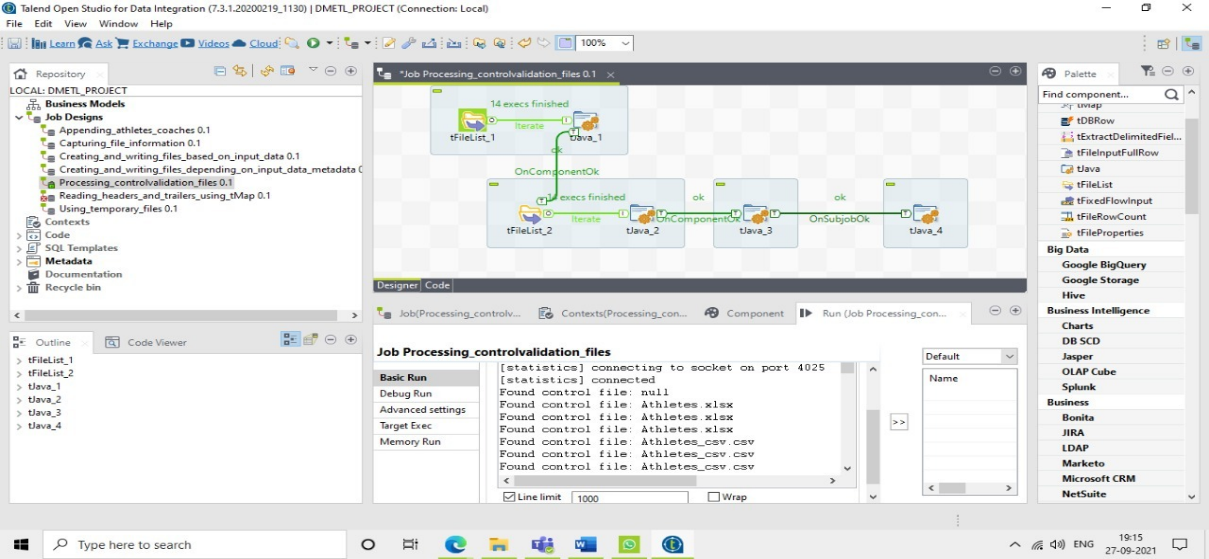
**Moving, copying, renaming, and deleting files and folders**

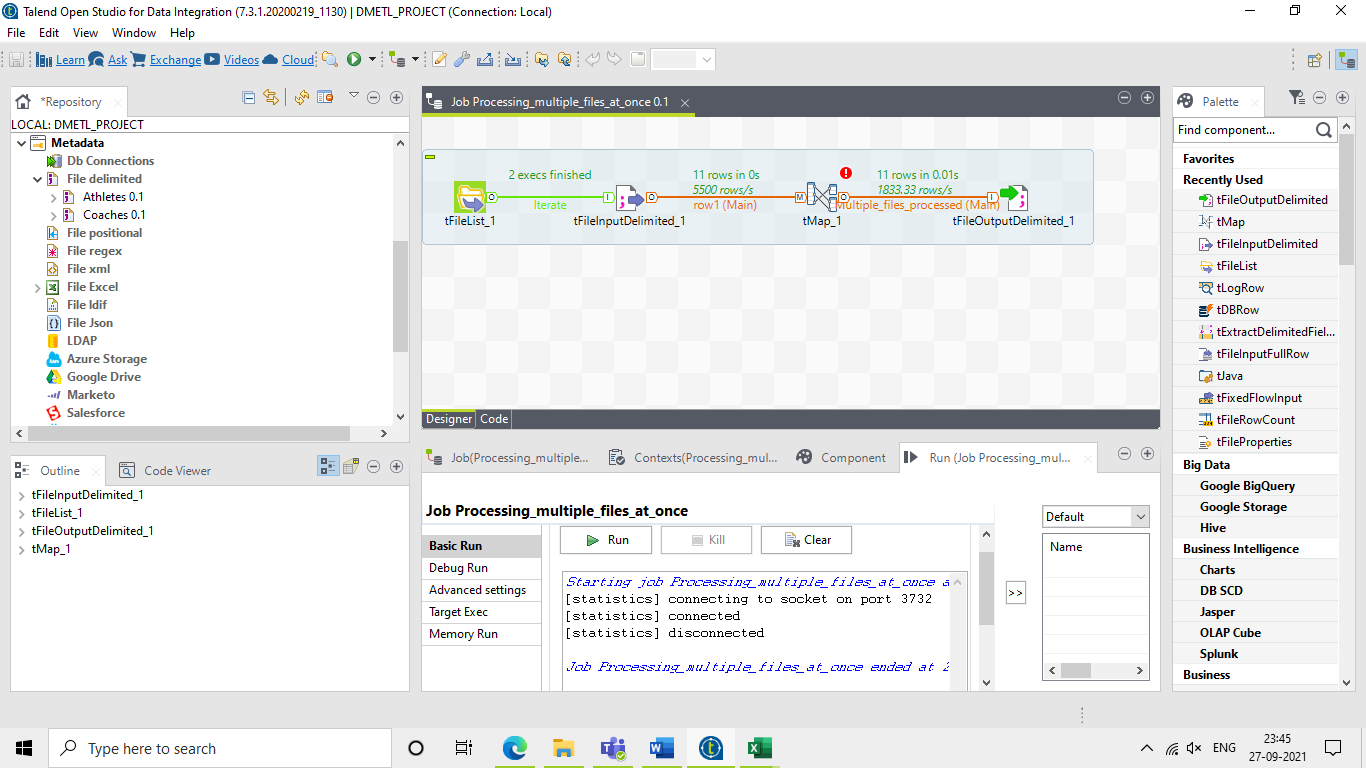
As well as reading from and writing to files, Talend has a set of components that allow developers to perform file functions without the need to call native operating system commands. tFileCopy is used to copy, move, and rename files, depending upon the options selected.The tFileDelete component is simply used to delete files.

**Capturing file information**

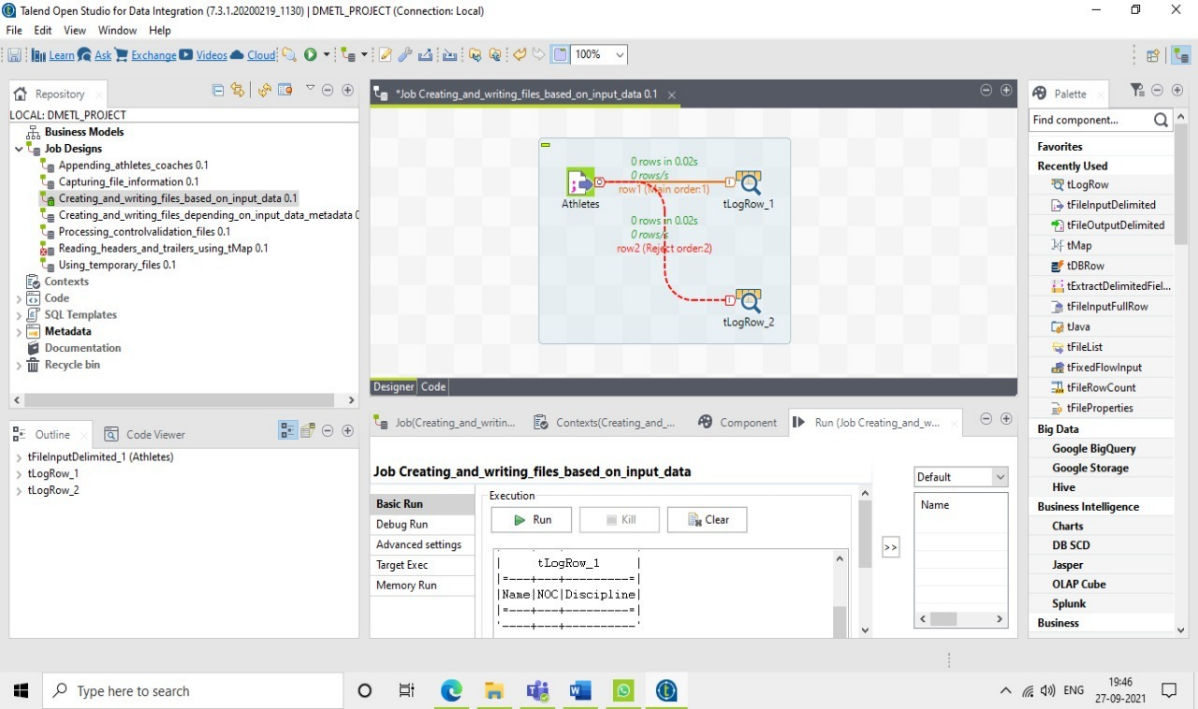
 Talend has a set of components that allow developers to perform file functions without the need to call native operating system commands.

**Processing control/validation files**

Some organizations prefer to use a companion (control/validation) file containing file information instead of storing the information in the file header or trailer. This means that the detail file is much simpler to process, because it is a normal flat file.

**Processing multiple files at once**

Often, with batch processes, it is required that multiple files are processed by the same job in a single tranche. This example shows how this can be achieved by merging a group of input files into a single output.

**Creating and writing files depending on the input data**

Creating and writing files depending on the input data Sometimes it is required that multiple files are written from a single data source where the file name is dependent upon the data held within the row.