Talend ETL Workflow - Student Handout

Goal:

To create an ETL job in Talend that:

- Reads employee data from a CSV file using tFileInputDelimited.

- Filters out employees with salary <= 45000 using tFilterRow.

- Adds a new derived column salary_grade using tMap.

- Displays the final data using tLogRow.

Step 1: tFileInputDelimited

- Drag tFileInputDelimited to canvas.

- Set file path to employees.csv

- Field Separator: , | Header: 1

- Edit Schema:

    empno (Integer)

    empname (String)

    sal (Integer)

    designation (String)

    dept (String)

Step 2: tFilterRow

- Drag tFilterRow and connect from tFileInputDelimited.

- In Basic settings, set condition:

    InputColumn: sal

    Function: Empty

    Operator: Greater than

    Value: 45000

Step 3: tMap

- Connect from tFilterRow to tMap.

- Drag all input columns to output.

- Add new column: salary_grade (String).

- Set expression:

    row2.sal >= 60000 ? "A" : (row2.sal >= 50000 ? "B" : "C")

Step 4: tLogRow

- Connect tMap to tLogRow.

- Set Mode: Table

- Enable: Print content with log4j

Sample Output

empno	empname	sal	designation	dept	salary_grade
101	Alice	50000	Engineer	IT	B
102	Bob	60000	Manager	Sales	A
104	David	70000	Lead	IT	A
105	Eva	62000	HR	HR	A

Bonus Tasks

- Change filter condition to sal >= 50000 and dept == 'IT'.

- Replace tLogRow with tFileOutputDelimited to write to CSV.