Talend Exercise: Data Exception Handling and Stewardship Simulation

This exercise demonstrates how to build a data exception handling and governance workflow in Talend Studio Trial. It includes reading data, validating it, segregating invalid records, and logging them into a stewardship review file.

Component Architecture Overview

tFileInputDelimited → tMap (split valid/invalid)
└── ValidOutput → tFileOutputDelimited (valid_employees.csv)
└── InvalidOutput → tLogRow (console error preview)
→ tMap (Add timestamp + status)
└── tFileOutputDelimited (data_steward_review.csv)
→ tFileOutputDelimited (error_report.csv)

Step-by-Step Implementation

Step 1: Read Employee Data
Component: tFileInputDelimited
File Name: employee_data.csv
Separator: ,
Header: 1
Schema:
emp_id: String, emp_name: String, dept: String, email: String, salary: String

Step 2: Validate Data and Split Outputs
Component: tMap_1
Create two outputs: ValidOutput and InvalidOutput

ValidOutput Condition:

row1.salary != null && !row1.salary.trim().isEmpty() &&
row1.salary.matches("\\d+(\\.\\d+)?") && Double.parseDouble(row1.salary) > 0

Add new column: alert_flag = "NO"

InvalidOutput Condition:

row1.salary == null || row1.salary.trim().isEmpty() ||
!row1.salary.matches("\\d+(\\.\\d+)?") || Double.parseDouble(row1.salary) <= 0

Add new column: alert_flag = "YES"

Step 3: Write Valid Records
Component: tFileOutputDelimited
Output: valid_employees.csv
Include Header: Yes

Step 4: Write Invalid Records
Component: tFileOutputDelimited
Output: error_report.csv
Include Header: Yes

Step 5: Print Invalid Records to Console
Component: tLogRow (optional)
Mode: Table

Step 6: Add Governance Info
Component: tMap_2
Add output schema: timestamp (String), resolution_status (String)
Expressions:
timestamp = TalendDate.formatDateInUTC("yyyy-MM-dd'T'HH:mm:ss'Z'",
TalendDate.getCurrentDate())
resolution_status = "Pending"

Step 7: Output to Stewardship Review File
Component: tFileOutputDelimited
Output: data_steward_review.csv
Include Header: Yes

Sample Output Files

valid_employees.csv
emp_id,emp_name,dept,email,salary,alert_flag
E101,Alice,HR,alice@company.com,55000,NO
E104,David,IT,david@company.com,70000,NO
E106,Frank,Admin,frank@company.com,62000,NO

error_report.csv

emp_id,emp_name,dept,email,salary,alert_flag
E102,Bob,Finance,bob@company.com,,YES
E103,Charlie,IT,charlie@company.com,-4000,YES
E105,Eve,HR,eve@company.com,abc,YES

data_steward_review.csv

emp_id,emp_name,dept,email,salary,alert_flag,timestamp,resolution_status
E102,Bob,Finance,bob@company.com,,YES,2025-07-16T11:06:00Z,Pending
E103,Charlie,IT,charlie@company.com,-4000,YES,2025-07-16T11:06:00Z,Pending
E105,Eve,HR,eve@company.com,abc,YES,2025-07-16T11:06:00Z,Pending