Student Exercise: Real-Time Micro-Batch Processing of Temperature Readings

**Module:** 3.1 – Real-Time Data Integration Techniques
Edition: Talend Open Studio

Objective

In this exercise, you will simulate a real-time temperature sensor stream. The goal is to process each sensor reading (device ID, timestamp, temperature) one-by-one from a batch file using the components **tFlowToIterate** and **tIterateToFlow**. Each row will:

1. Be read from a batch file

2. Be streamed individually (converted to variables)

3. Be transformed to generate alerts if temperature > 50°C

4. Be logged and stored in the final output

Workflow:
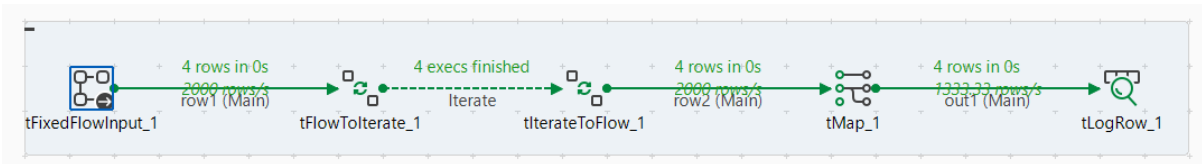
Components Used

Component	Purpose
tFixedFlowInput	Simulates incoming sensor batch (CSV-style)
tFlowToIterate	Converts each row into a variable for streaming
tIterateToFlow	Converts each iteration back to a row
tMap	Adds transformation logic (e.g., alert level)
tLogRow	Prints streaming output to console
(Optional) tSleep	Adds delay to simulate continuous ingestion

Simulated Input Data

The input data simulates temperature readings from different sensors.

Columns:

- device_id (String)

- timestamp (String)

- temperature (Float)

Example rows in **temperature_readings.csv**:

device_id	timestamp	temperature
D001	2025-06-18 10:00:01	44.5
D002	2025-06-18 10:00:05	52.0
D003	2025-06-18 10:00:10	47.3
D004	2025-06-18 10:00:15	61.9

Job Flow Configuration

1. **tFixedFlowInput**: Reads all the sensor rows (4–5 for demo).

2. **tFlowToIterate**: Streams each row one-by-one as globalMap variables.

3. **tIterateToFlow**: Reassembles each row from globalMap variables.

4. **tMap**:

   o   Adds alert level:

      ▪   If temperature > 50°C → Alert as "HIGH"

      ▪   If temperature ≤ 50°C → Alert as "NORMAL"

5. **tLogRow**: Prints each alert row live in the console.

6. (Optional) Add **tSleep** inside iteration to simulate delay.

Step-by-Step Configuration

Step 1: Setting Up tFixedFlowInput

- Drag and drop the **tFixedFlowInput** component.

- Define the schema to match the input data (device_id, timestamp, and temperature).

- Add a CSV file path (e.g., temperature_readings.csv), simulating batch sensor data.

Schema	Built-In ▼	Edit schema ⋯	

Number of rows: 1

Mode
- ○ Use Single Table
- ● Use Inline Table

Inline Table

device_id	timestamp	temperature
"D001"	"2025-06-18 10:00:01"	44.5
"D002"	"2025-06-18 10:00:05"	52.0
"D003"	"2025-06-18 10:00:10"	47.3
"D004"	"2025-06-18 10:00:15"	61.9

-

Step 2: Setting Up tFlowToIterate

- Drag and drop **tFlowToIterate**.

- **Link** the output of **tFixedFlowInput** to **tFlowToIterate**.

- Configure **tFlowToIterate** to convert each row to a globalMap variable. The column names will become variables:

   - globalMap.get("device_id")

   - globalMap.get("timestamp")

   - globalMap.get("temperature")

Step 3: Setting Up tIterateToFlow

- Drag and drop **tIterateToFlow**.

- **Link** the output of **tFlowToIterate** to **tIterateToFlow**.

- Define the output schema:

   - device_id (String)

   - timestamp (String)

   - temperature (Float)

- Assign the values from **globalMap**:

   - ((String)globalMap.get("device_id"))

   - ((String)globalMap.get("timestamp"))

- ((Float)globalMap.get("temperature"))

Step 4: Setting Up tMap (Transformation Logic)

- Drag and drop **tMap**.

- **Link** the output of **tIterateToFlow** to **tMap**.

- Define the mapping for the output schema.

- Add a new column: **alert_level**

- **Expression** for alert level:

- row1.temperature > 50 ? "HIGH" : "NORMAL"

Step 5: Setting Up tLogRow

- Drag and drop **tLogRow** to print the results.

- **Link** the output of **tMap** to **tLogRow**.

- In **tLogRow** configuration, select the **Table** mode to view the results as a table.

Step 6: Optional Step - Simulate Delay with tSleep

- Drag and drop **tSleep**.

- Place **tSleep** inside the iteration to simulate a 1-second delay for each row, mimicking a real-time sensor stream.

- Set **tSleep** delay to **1000 ms** (1 second).

Sample Output

device_id	timestamp	temperature	alert_level
D001	2025-06-18 10:00:01	44.5	NORMAL
D002	2025-06-18 10:00:05	52.0	HIGH
D003	2025-06-18 10:00:10	47.3	NORMAL
D004	2025-06-18 10:00:15	61.9	HIGH

Learning Outcomes

Concept	Demonstrated?
Micro-batch simulation	tFixedFlowInput
Row-level streaming logic	tFlowToIterate / tIterateToFlow
Conditional transformation	tMap
Real-time simulation	Optional delay via tSleep