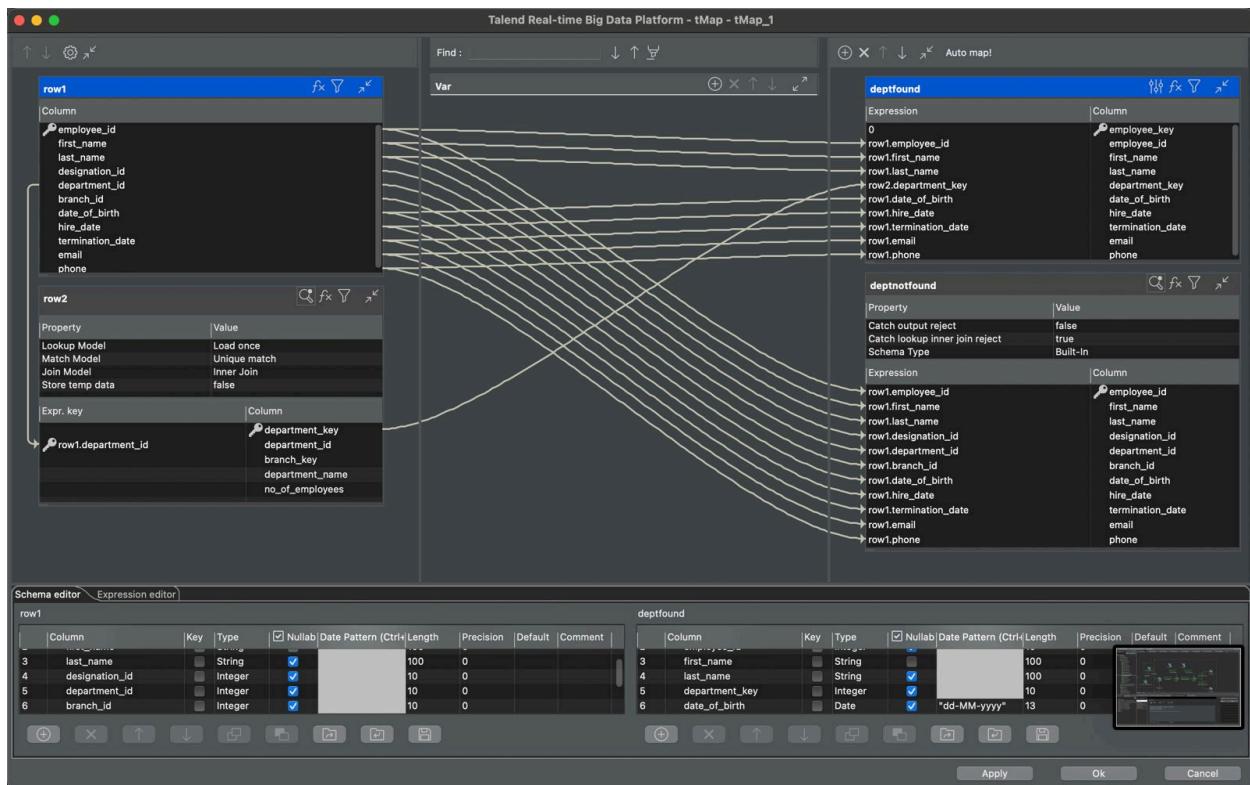
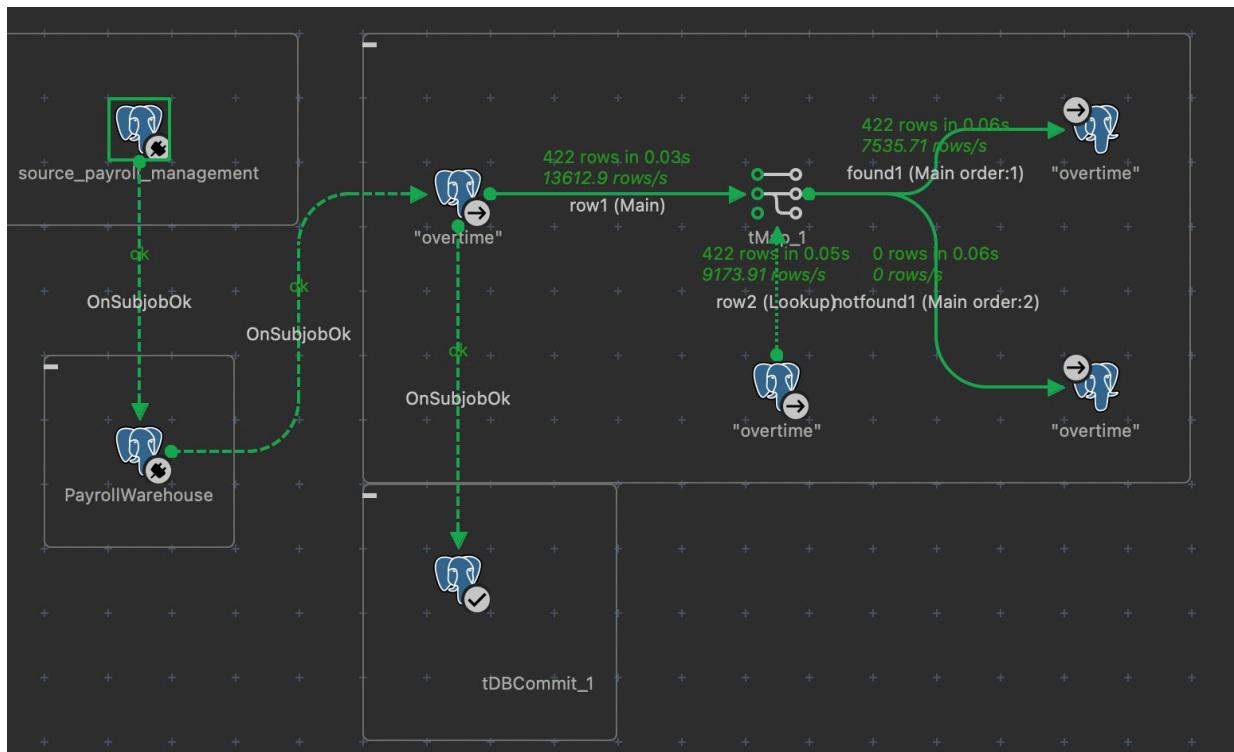
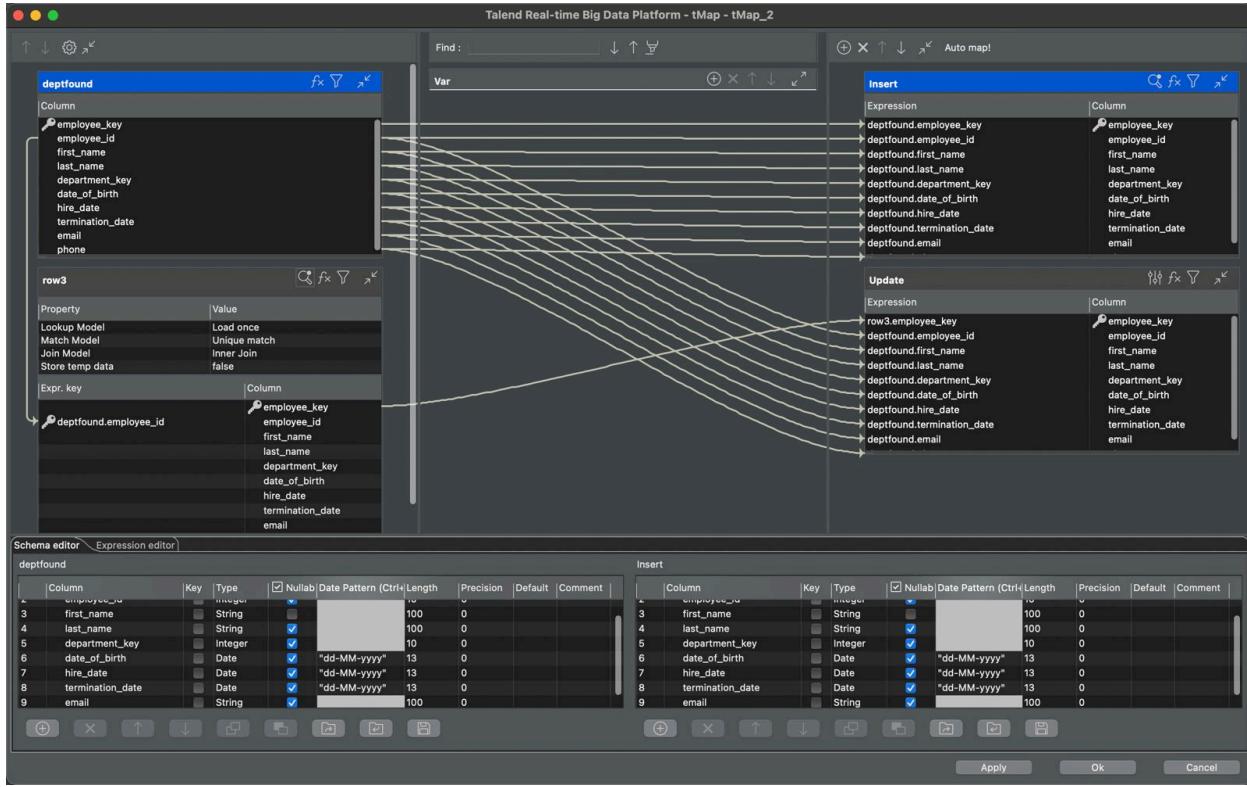


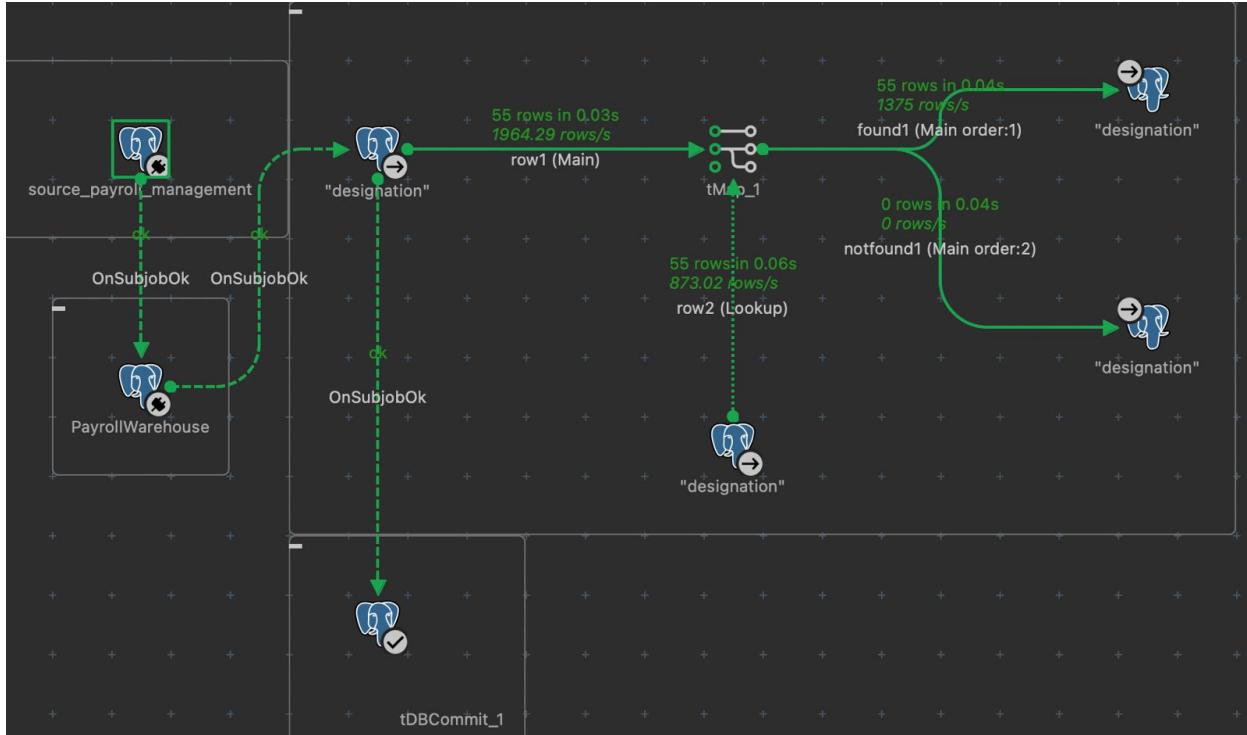
# 1. ETL Execution

## Insert/Update flow for Employee Dimension



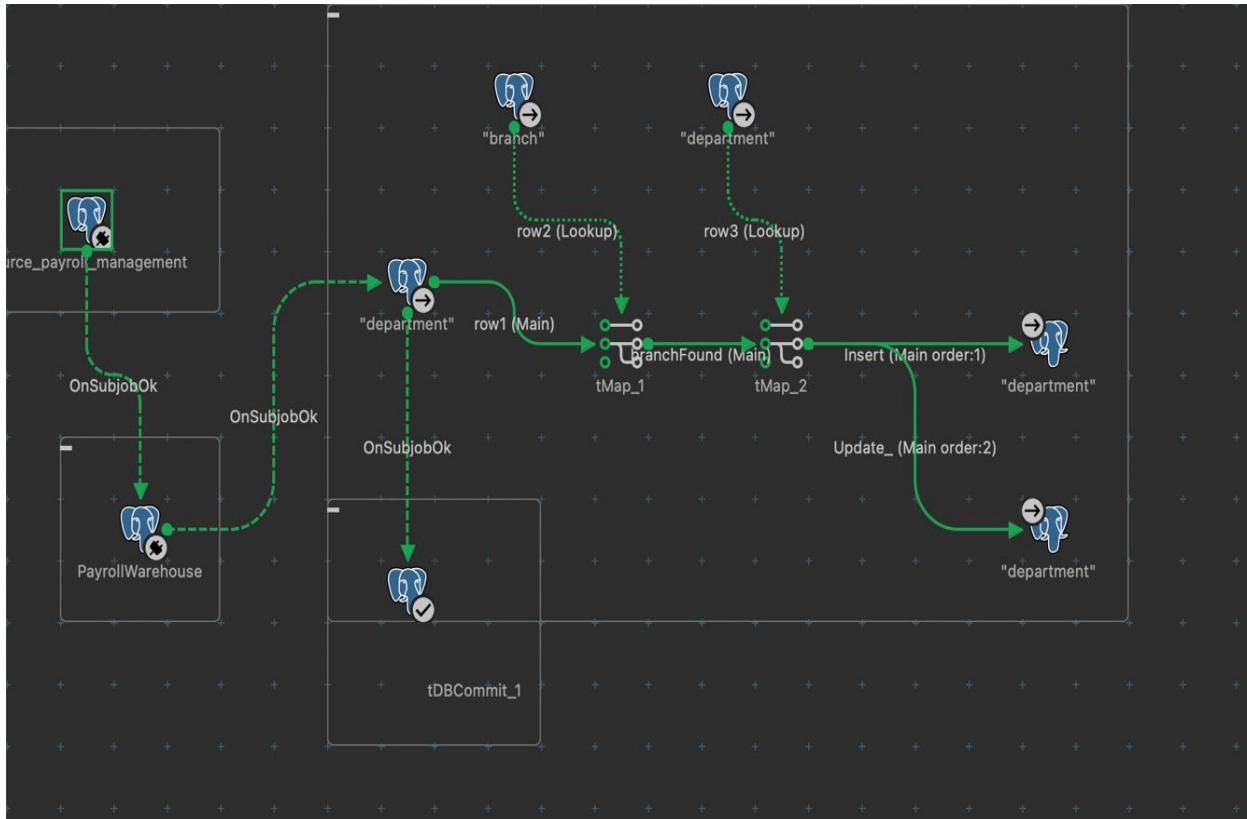


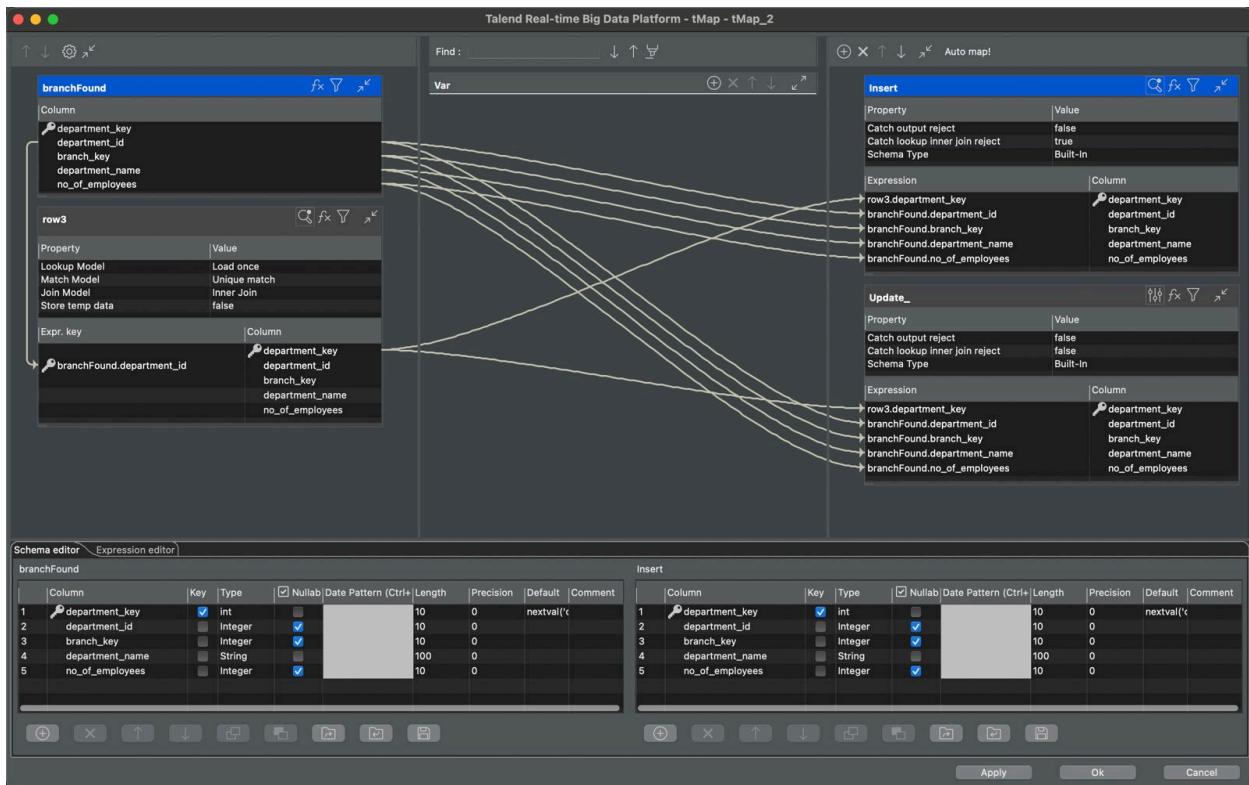
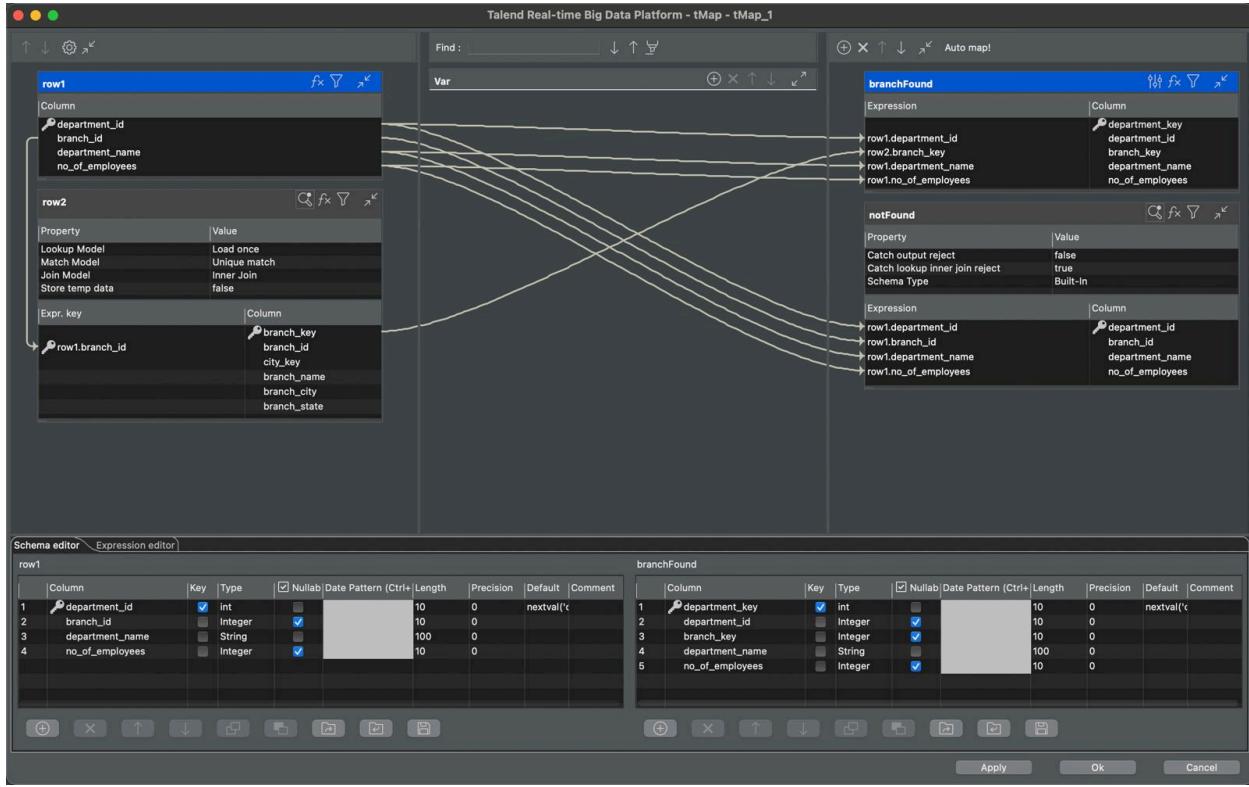
## Insert/Update flow for Designation Dimension



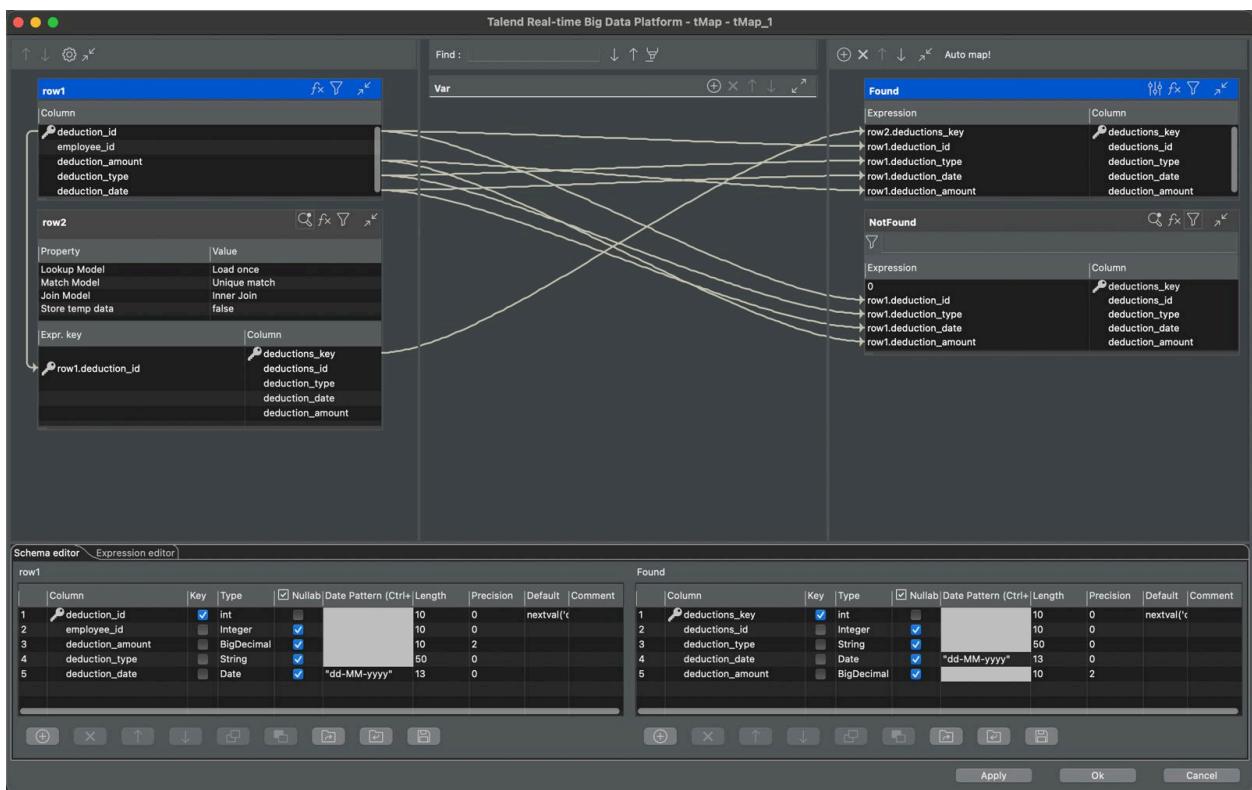
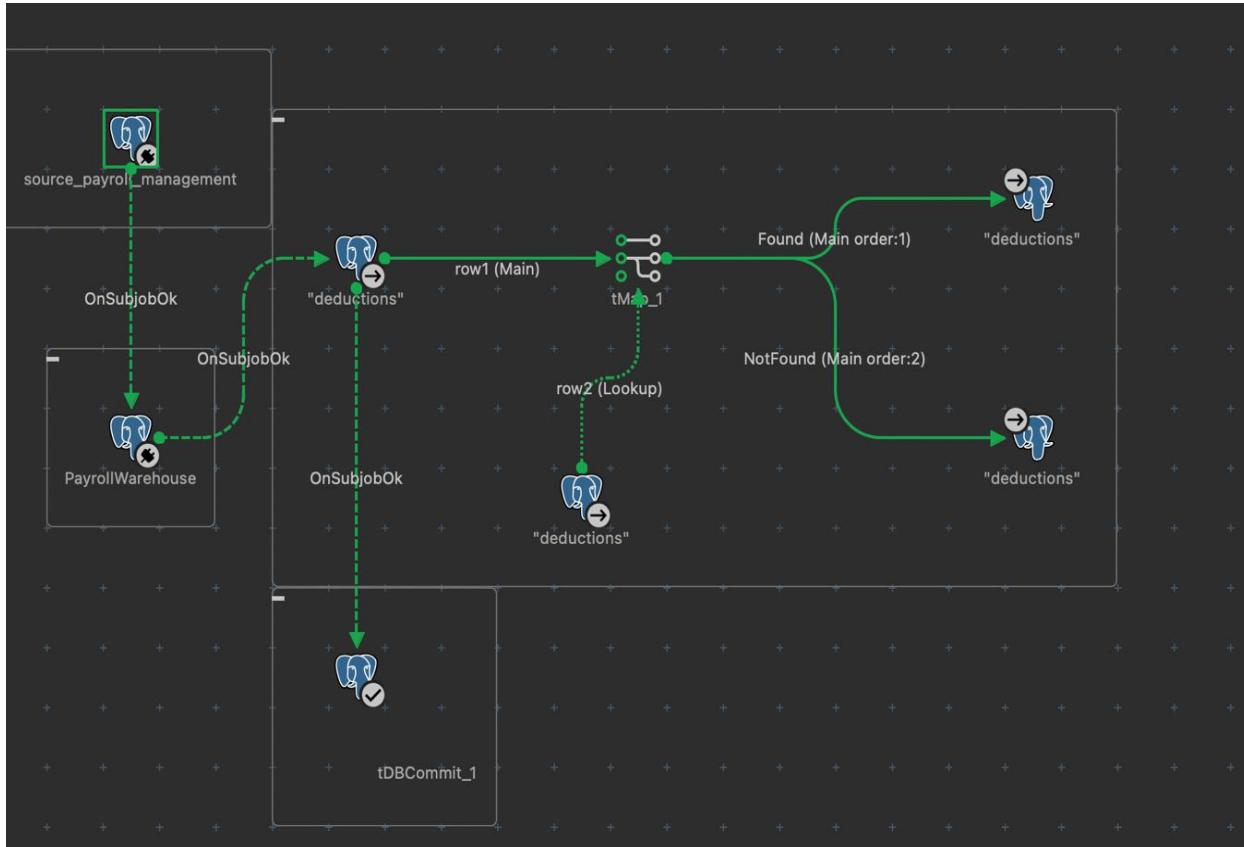


## Insert/Update flow for Department Dimension





## Insert/Update flow for Deductions Dimension



## Insert/Update flow for Overtime Dimension

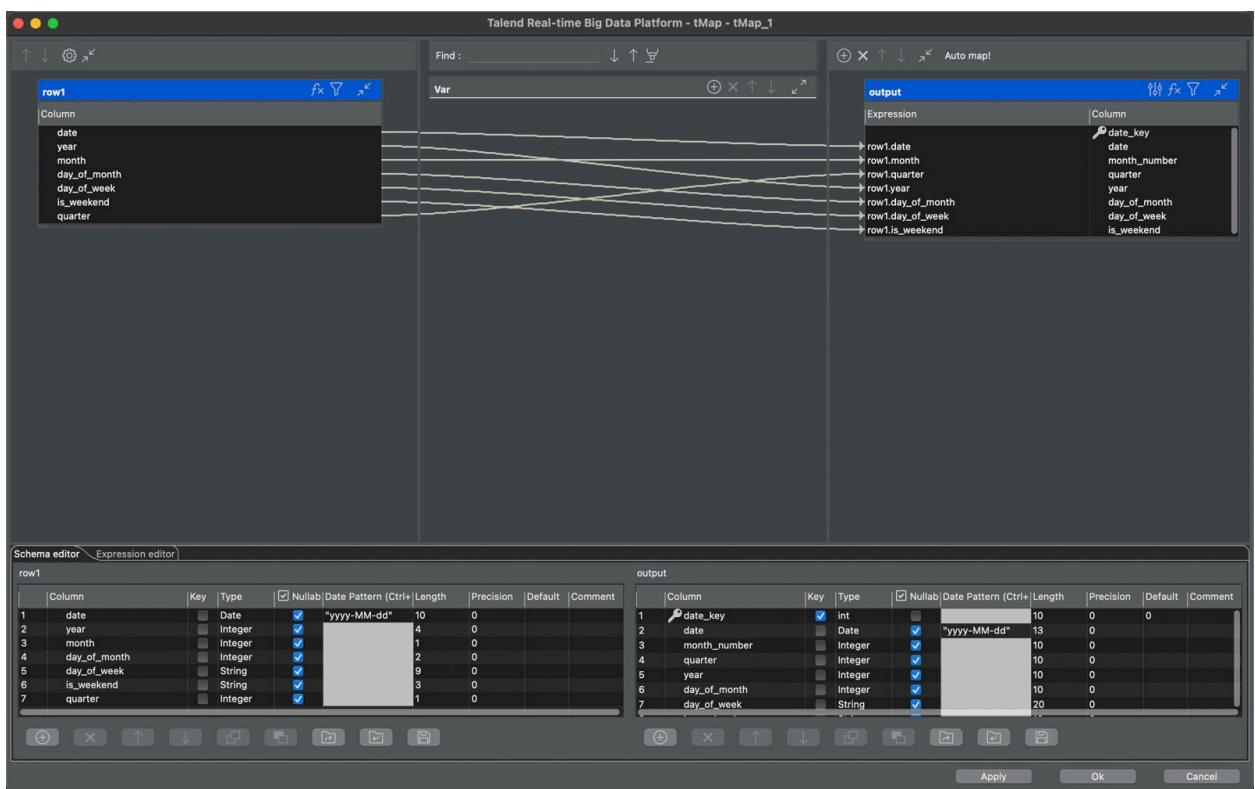
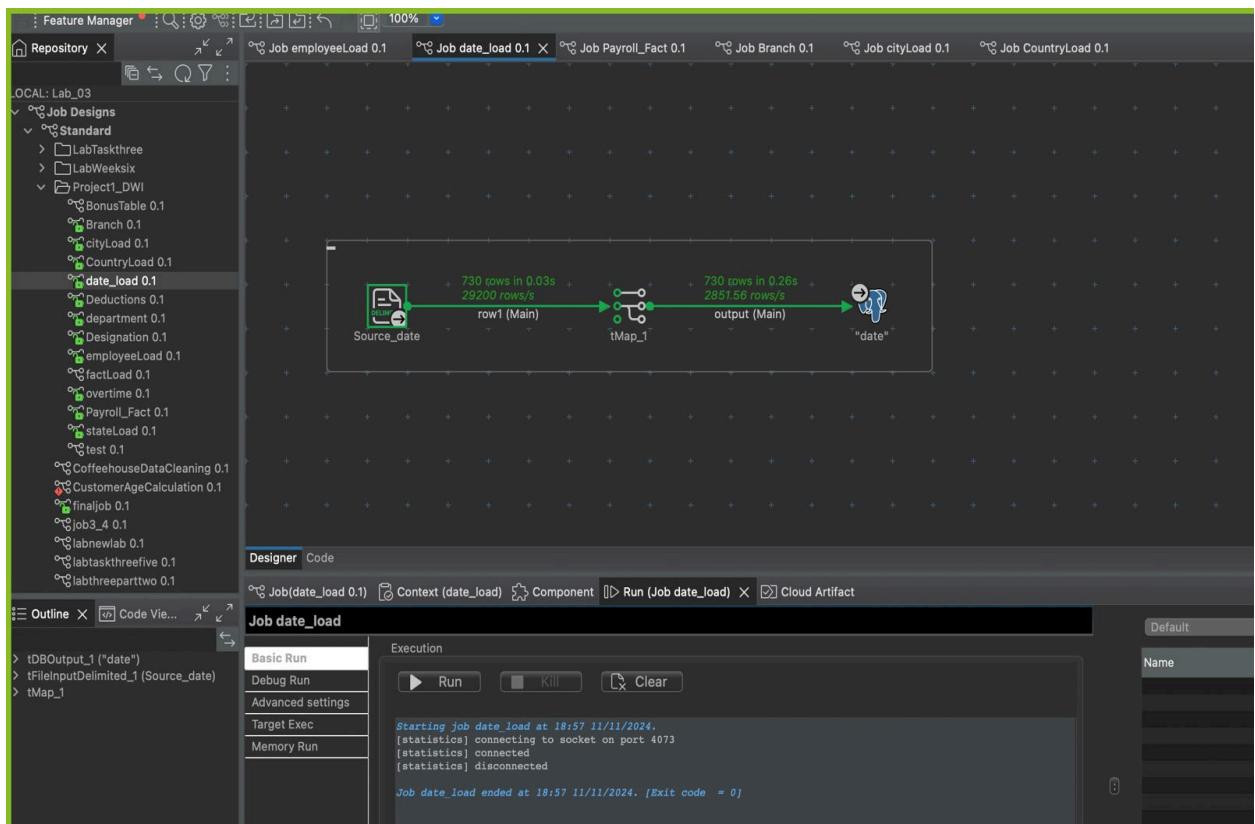


Talend Real-time Big Data Platform - tMap - tMap\_1

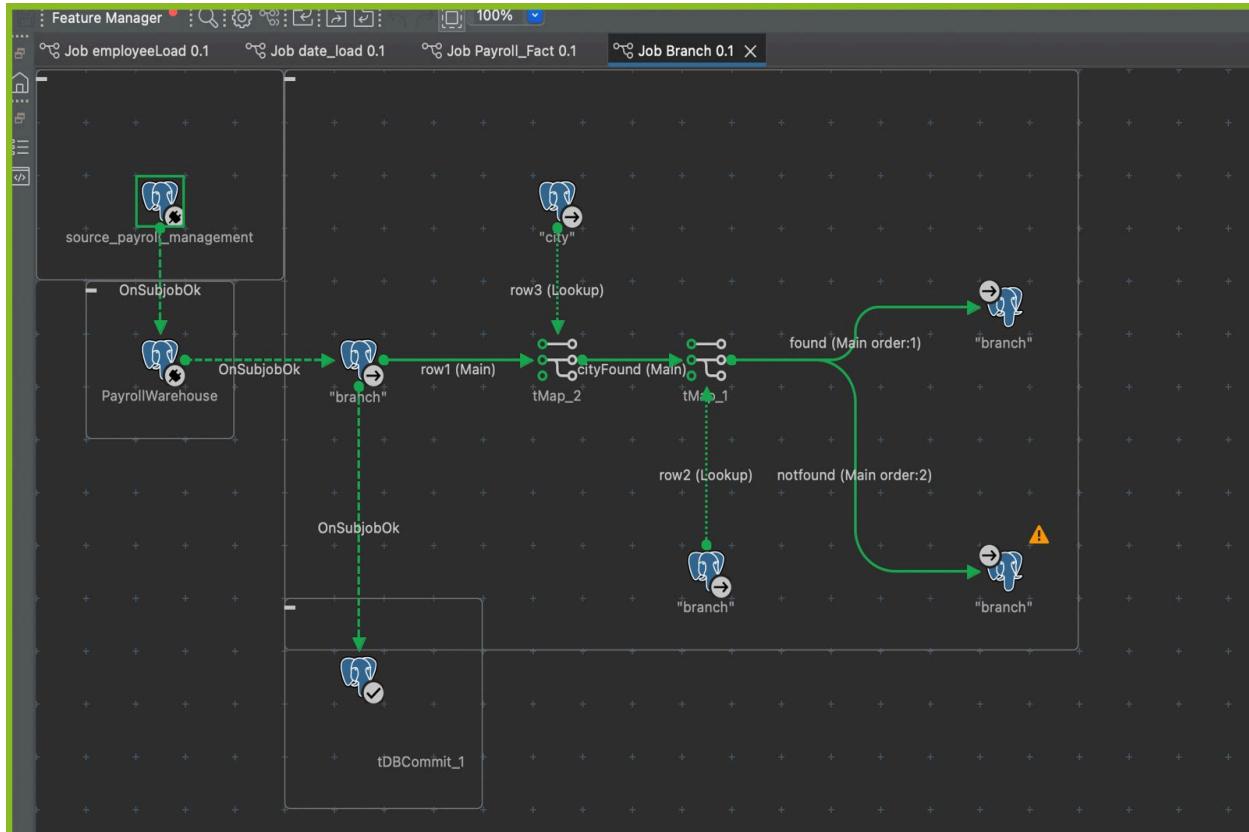
Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
overtime_id	int	<input checked="" type="checkbox"/>	✓		10	0	nextval('c')	
employee_id	Integer	<input checked="" type="checkbox"/>	✓		10	0		
overtime_hours	BigDecimal	<input checked="" type="checkbox"/>	✓		5	2		
overtime_rate	BigDecimal	<input checked="" type="checkbox"/>	✓		10	2		
overtime_pay	BigDecimal	<input checked="" type="checkbox"/>	✓		10	2		
overtime_date	Date	<input checked="" type="checkbox"/>	✓	"dd-MM-yyyy"	13	0		

Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
overtime_key	int	<input checked="" type="checkbox"/>	✓		10	0	nextval('c')	
overtime_id	Integer	<input checked="" type="checkbox"/>	✓		10	0		
overtime_date	Date	<input checked="" type="checkbox"/>	✓	"dd-MM-yyyy"	13	0		
overtime_hours	BigDecimal	<input checked="" type="checkbox"/>	✓		5	2		
overtime_rate	BigDecimal	<input checked="" type="checkbox"/>	✓		10	2		

## Insert/Update flow for Date Dimension



## Insert/Update flow for Branch Dimension



Talend Real-time Big Data Platform - tMap\_2

row1	Column	branch_id	branch_name	branch_city	branch_state
row3	Property	Value			
	Lookup Model	Load once			
	Match Model	Unique match			
	Join Model	Inner Join			
	Store temp data	false			
	Expr. key	row1.branch_city			
		city_key	city_name	state_key	

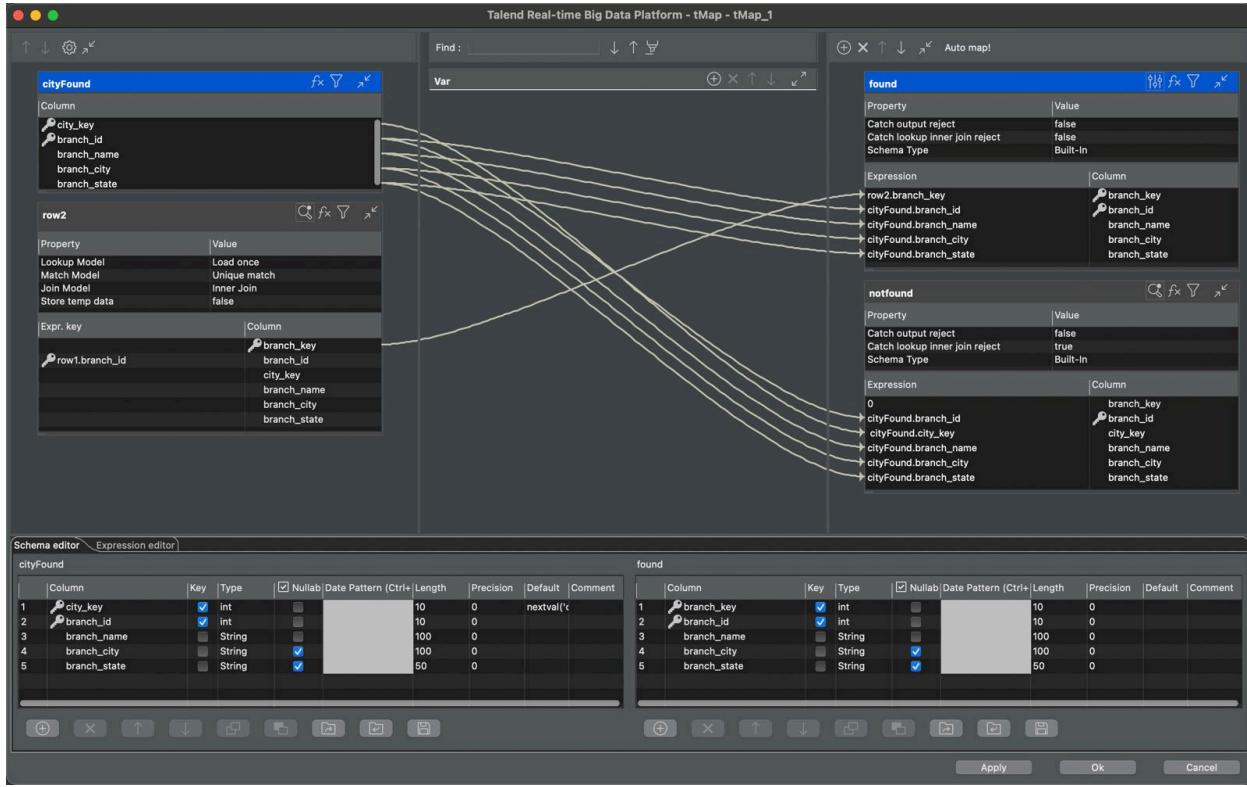
cityFound	Property	Value
	Catch output reject	false
	Catch lookup inner join reject	false
	Schema Type	Built-In
Expression	Column	
row3.city_key	city_key	
row1.branch_id	branch_id	
row1.branch_name	branch_name	
row1.branch_city	branch_city	
row1.branch_state	branch_state	

cityNotFound	Property	Value
	Catch output reject	false
	Catch lookup inner join reject	true
	Schema Type	Built-In
Expression	Column	
row1.branch_id	branch_id	
row1.branch_name	branch_name	
row1.branch_city	branch_city	
row1.branch_state	branch_state	

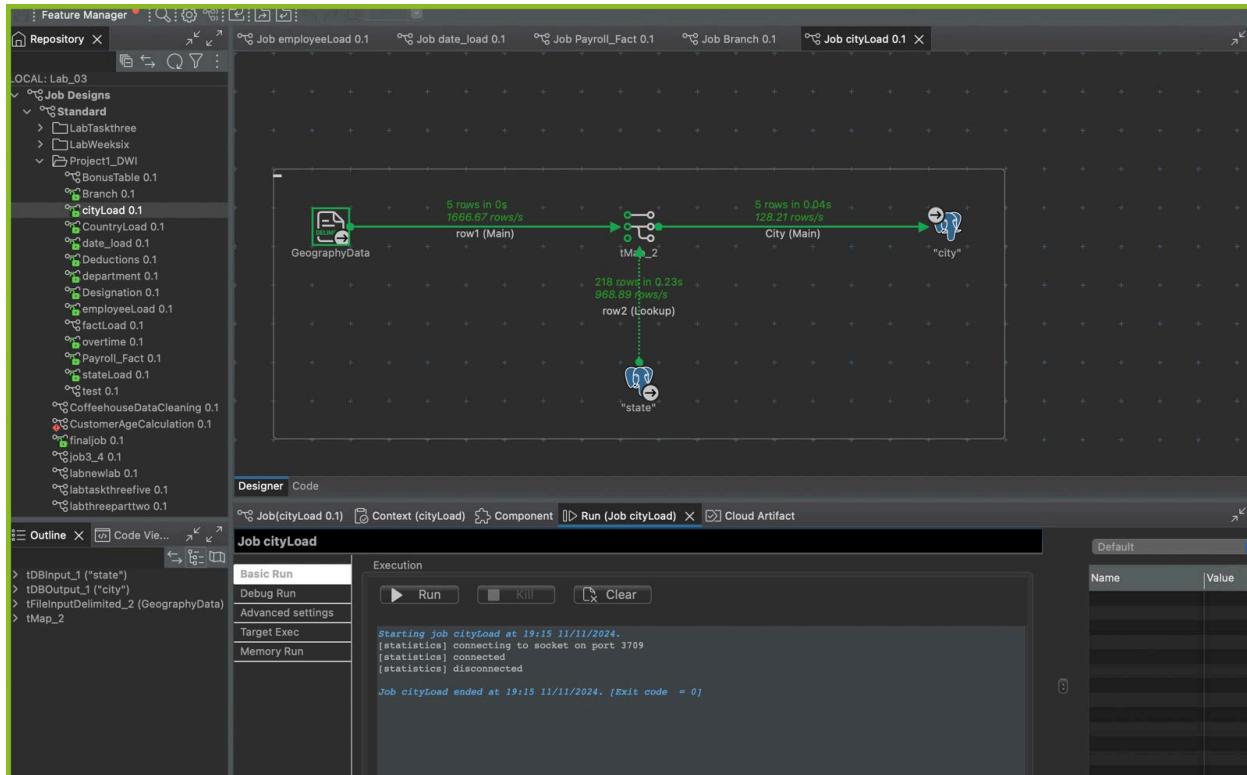
Schema editor Expression editor

row1	Column	Key	Type	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
1	branch_id	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>		10	0		
2	branch_name	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		100	0		
3	branch_city	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		100	0		
4	branch_state	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		50	0		

cityFound	Column	Key	Type	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
1	city_key	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>		10	0		nextval('c
2	branch_id	<input checked="" type="checkbox"/>	int	<input checked="" type="checkbox"/>		10	0		
3	branch_name	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		100	0		
4	branch_city	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		100	0		
5	branch_state	<input checked="" type="checkbox"/>	String	<input checked="" type="checkbox"/>		50	0		



## Insert/Update flow for State and City Dimension



Talend Real-time Big Data Platform - tMap - tMap\_2

**row1**

**row2**

**City**

**Schema editor**

Column	Type	Key	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
city	String				13	0		
state	String				13	0		
country	String				3	0		

Column	Type	Key	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
city_key	int				10	0		nextval('c...
city_name	String				50	0		
state_key	Integer				10	0		

Talend Real-time Big Data Platform - tMap - tMap\_1

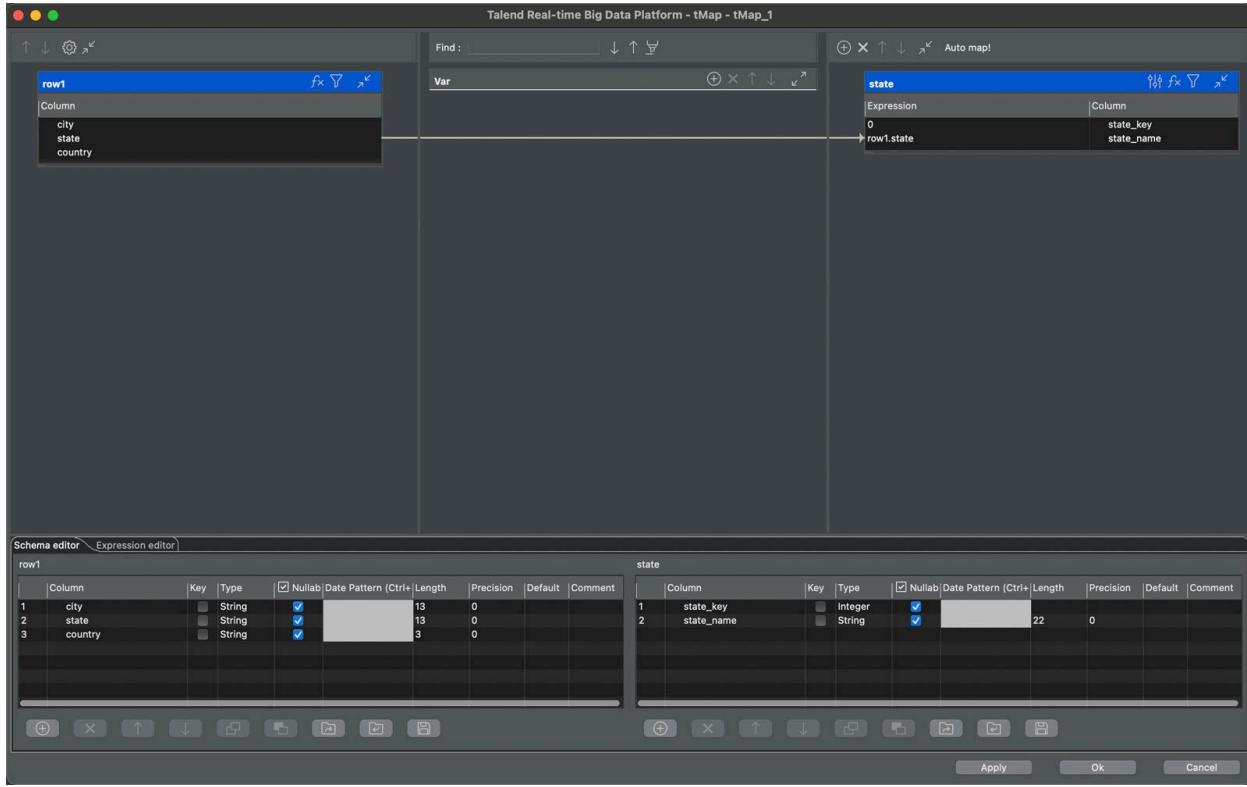
**row1**

**state**

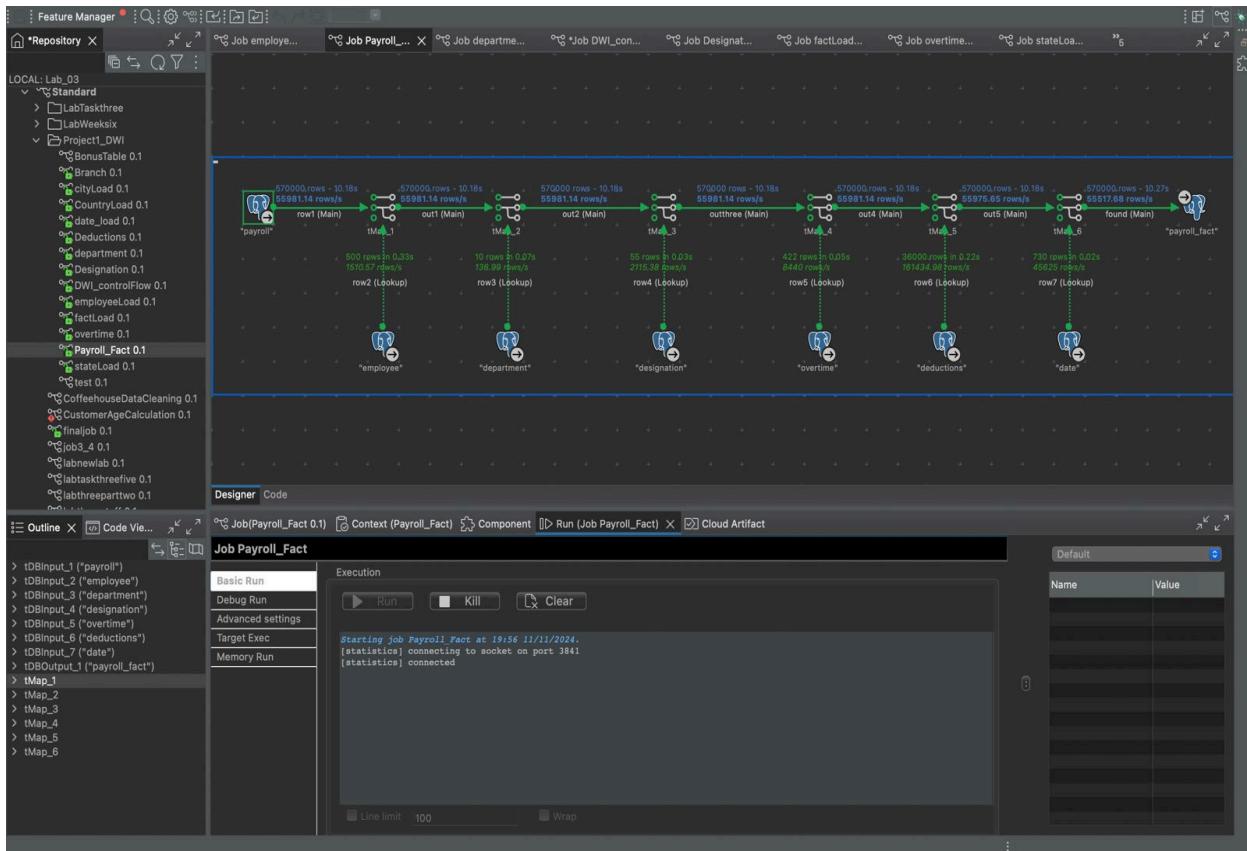
**Schema editor**

Column	Type	Key	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
_state_id	String				4	0		
_state_name	String				22	0		
_county_fips	String				7	0		
_county_name	String				22	0		
_lat	String				9	0		
_lng	String				11	0		
_population	String				10	0		
_density	String				7	0		

Column	Type	Key	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
state_id	Integer				10	0		
state_name	String				22	0		



## Insert/Update flow for Payroll (Fact Table)



Talend Real-time Big Data Platform - tMap - tMap\_1

**Row 1 (Input):**

- Columns: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount.
- Properties:
  - Lookup Model: Load once
  - Match Model: Unique match
  - Join Model: Inner Join
  - Store temp data: false
- Expr. key: row1.employee\_id
- Column: employee\_key (with dropdown for first\_name and last\_name)

**Row 2 (Input):**

- Property: Value
- Lookup Model: Load once
- Match Model: Unique match
- Join Model: Inner Join
- Store temp data: false
- Expr. key: row1.employee\_id
- Column: employee\_key (with dropdown for first\_name and last\_name)

**Out 1 (Output):**

- Expression: row1.payroll\_id, row1.employee\_id, row1.department\_id, row1.designation\_id, row1.deduction\_id, row1.overtime\_id, row1.pay\_end\_date, row1.base\_pay, row1.overtime\_hours, row1.overtime\_rate, row1.deduction\_amount.
- Column: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount.

**Out One (Output):**

- Expression: row1.employee\_id, row1.payroll\_id, row1.department\_id, row1.designation\_id, row1.deduction\_id, row1.overtime\_id, row1.overtime\_hours, row1.overtime\_rate, row1.deduction\_amount.
- Column: employee\_id, payroll\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate.

**Schema Editor:**

Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
3 department_id	Integer		✓		50	0		
4 designation_id	Integer		✓		13	0		
5 deduction_id	Integer		✓		13	0		
6 overtime_id	Integer		✓		10	2		
7 pay_end_date	Date		✓	"dd-MM-yyyy"	10	2		

**Schema Editor (Out 1):**

Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
3 department_id	Integer		✓		50	0		
4 designation_id	Integer		✓		13	0		
5 deduction_id	Integer		✓		13	0		
6 overtime_id	Integer		✓		10	2		
7 pay_end_date	Date		✓	"dd-MM-yyyy"	10	2		

**Buttons:** Apply, Ok, Cancel

Talend Real-time Big Data Platform - tMap - tMap\_2

**Out 1 (Input):**

- Columns: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount.
- Properties:
  - Employee key: payroll\_id

**Row 3 (Input):**

- Expr. key: out1.department\_id
- Column: department\_key (with dropdown for department\_id, branch\_key, department\_name, no\_of\_employees)

**Out 2 (Output):**

- Expression: out1.payroll\_id, out1.employee\_id, out1.department\_id, out1.designation\_id, out1.deduction\_id, out1.overtime\_id, out1.pay\_end\_date, out1.base\_pay, out1.overtime\_hours, out1.overtime\_rate, out1.deduction\_amount.
- Column: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount.

**Out Three (Output):**

- Expression: out1.employee\_key, out1.department\_id, out1.designation\_id, out1.deduction\_id, out1.overtime\_id, out1.overtime\_hours, out1.overtime\_rate, out1.deduction\_amount.
- Column: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount.

**Schema Editor:**

Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
3 department_id	Integer		✓		50	0		
4 designation_id	Integer		✓		13	0		
5 deduction_id	Integer		✓		13	0		
6 overtime_id	Integer		✓		10	2		
7 pay_end_date	Date		✓	"dd-MM-yyyy"	10	2		
8 base_pay	BigDecimal		✓		10	2		
9 overtime_hours	BigDecimal		✓		10	2		

**Schema Editor (Out 1):**

Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
3 department_id	Integer		✓		50	0		
4 designation_id	Integer		✓		13	0		
5 deduction_id	Integer		✓		13	0		
6 overtime_id	Integer		✓		10	2		
7 pay_end_date	Date		✓	"dd-MM-yyyy"	10	2		

**Schema Editor (Out Two):**

Column	Type	Key	Nullab	Date Pattern (Ctrl+L)	Length	Precision	Default	Comment
3 department_id	Integer		✓		50	0		
4 designation_id	Integer		✓		13	0		
5 deduction_id	Integer		✓		13	0		
6 overtime_id	Integer		✓		10	2		
7 pay_end_date	Date		✓	"dd-MM-yyyy"	10	2		

**Buttons:** Apply, Ok, Cancel

Talend Real-time Big Data Platform - tMap - tMap\_3

**out2**

- Column: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount, employee\_key, department\_key

**row4**

- Expr. key: out2.designation\_id
- Column: designation\_key, designation\_id, designation\_name, job\_level, base\_pay

**outhree**

- Expression: out2.payroll\_id, out2.employee\_id, out2.department\_id, out2.designation\_id, out2.deduction\_id, out2.overtime\_id, out2.pay\_end\_date, out2.base\_pay, out2.overtime\_hours, out2.overtime\_rate, out2.deduction\_amount, out2.employee\_key, out2.department\_key

**outhree**

- Property: Catch output reject (false), Catch lookup inner join reject (true), Schema Type (Built-In)
- Expression: out2.department\_id

**Schema editor**   **Expression editor**

**out2**

Column	Key	Type	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
1 payroll_id	✓ int	int	✓		10	0	nextval('t	
2 employee_id	✓ integer	integer	✓		10	0		
3 department_id	✓ integer	integer	✓		50	0		
4 designation_id	✓ integer	integer	✓		13	0		
5 deduction_id	✓ integer	integer	✓		13	0		
6 overtime_id	✓ integer	integer	✓		10	2		
7 pay_end_date	✓ Date	date	✓ "dd-MM-yyyy"		10	2		
8 base_pay	✓ BigDecimal	decimal	✓		10	2		
9 overtime_hours	✓ BigDecimal	decimal	✓		10	2		
10 overtime_rate	✓ BigDecimal	decimal	✓		10	2		
11 deduction_amount	✓ BigDecimal	decimal	✓		10	2		
12 employee_key	✓ int	int	✓		10	0	nextval('t	
13 department_key	✓ int	int	✓		10	0		

**outhree**

Column	Key	Type	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
1 payroll_id	✓ int	int	✓		10	0	nextval('t	
2 employee_id	✓ integer	integer	✓		10	0		
3 department_id	✓ integer	integer	✓		50	0		
4 designation_id	✓ integer	integer	✓		13	0		
5 deduction_id	✓ integer	integer	✓		13	0		
6 overtime_id	✓ integer	integer	✓		10	2		
7 pay_end_date	✓ Date	date	✓ "dd-MM-yyyy"		10	2		
8 base_pay	✓ BigDecimal	decimal	✓		10	2		
9 overtime_hours	✓ BigDecimal	decimal	✓		10	2		
10 overtime_rate	✓ BigDecimal	decimal	✓		10	2		
11 deduction_amount	✓ BigDecimal	decimal	✓		10	2		
12 employee_key	✓ int	int	✓		10	0		
13 department_key	✓ int	int	✓		10	0	nextval('t	
14 designation_key	✓ int	int	✓		10	0	nextval('t	

Apply   Ok   Cancel

Talend Real-time Big Data Platform - tMap - tMap\_4

**outhree**

- Column: payroll\_id, employee\_id, department\_id, designation\_id, deduction\_id, overtime\_id, pay\_end\_date, base\_pay, overtime\_hours, overtime\_rate, deduction\_amount, employee\_key, department\_key, designation\_key

**row5**

- Expr. key: cuthree.overtime\_id
- Column: overtime\_key, overtime\_id, overtime\_date, overtime\_hours, overtime\_rate

**out4**

- Expression: outhree.payroll\_id, outhree.employee\_id, outhree.department\_id, outhree.designation\_id, outhree.deduction\_id, outhree.overtime\_id, outhree.pay\_end\_date, outhree.base\_pay, outhree.overtime\_hours, outhree.overtime\_rate, outhree.deduction\_amount, outhree.employee\_key, outhree.department\_key, outhree.designation\_key

**out4**

- Expression: outhree.overtime\_id, outhree.employee\_id, outhree.department\_id, outhree.designation\_id, outhree.deduction\_id, outhree.overtime\_id, outhree.pay\_end\_date, outhree.base\_pay, outhree.overtime\_hours, outhree.overtime\_rate

**Schema editor**   **Expression editor**

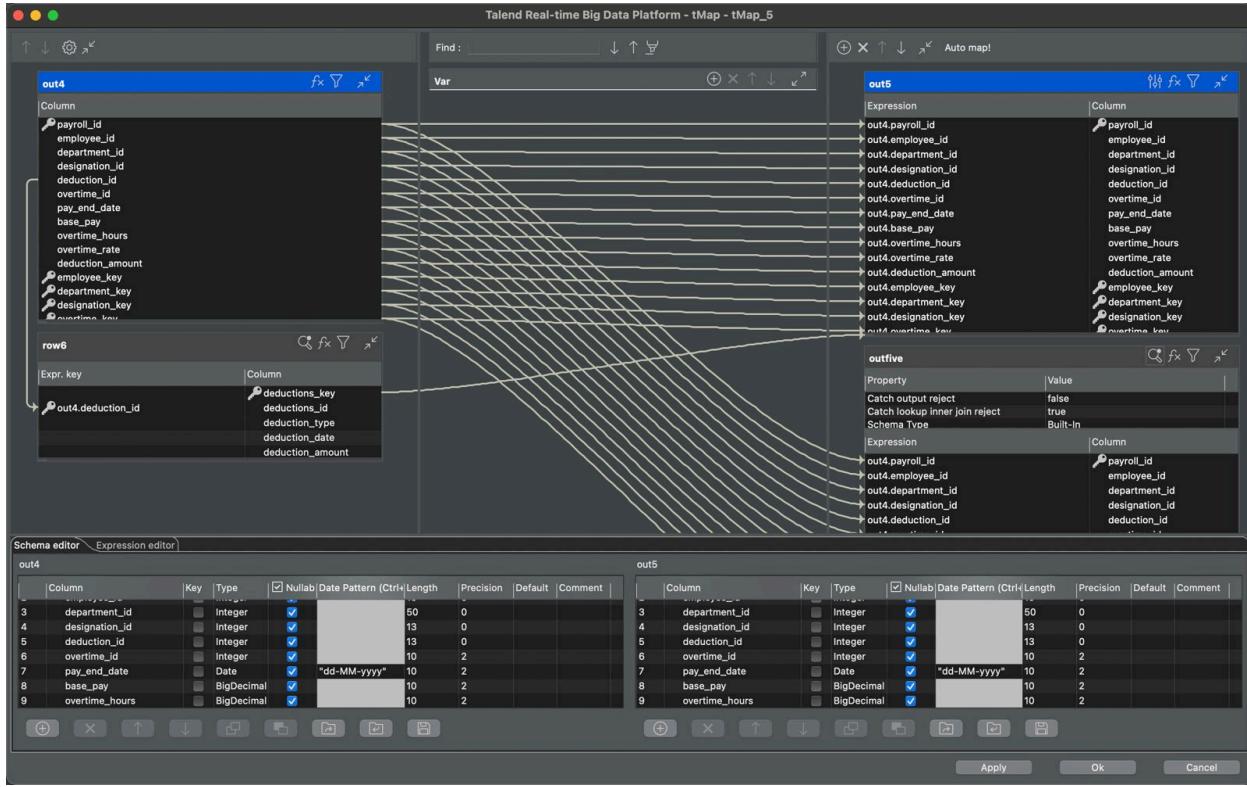
**outhree**

Column	Key	Type	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
3 department_id	✓ integer	integer	✓		50	0		
4 designation_id	✓ integer	integer	✓		13	0		
5 deduction_id	✓ integer	integer	✓		13	0		
6 overtime_id	✓ integer	integer	✓		10	2		
7 pay_end_date	✓ Date	date	✓ "dd-MM-yyyy"		10	2		
8 base_pay	✓ BigDecimal	decimal	✓		10	2		
9 overtime_hours	✓ BigDecimal	decimal	✓		10	2		

**out4**

Column	Key	Type	Nullab	Date Pattern (Ctrl+)	Length	Precision	Default	Comment
3 department_id	✓ integer	integer	✓		50	0		
4 designation_id	✓ integer	integer	✓		13	0		
5 deduction_id	✓ integer	integer	✓		13	0		
6 overtime_id	✓ integer	integer	✓		10	2		
7 pay_end_date	✓ Date	date	✓ "dd-MM-yyyy"		10	2		
8 base_pay	✓ BigDecimal	decimal	✓		10	2		
9 overtime_hours	✓ BigDecimal	decimal	✓		10	2		

Apply   Ok   Cancel



## Control Flow



## 2. Data Sources Used

Used 3 sources of data

- CSV file for Date Dimension
- CSV file for State and City Dimensions
- Postgres Database with Statically generated data

## 3. Additional Features: Calculated Measures

Below are the Calculated measures

- Average Base Pay in Fact Table: Calculated by averaging the Base pay from the Payroll table in the source system
- Average Deductions in Fact Table: Calculated by averaging the Deductions from the Payroll table in the source system
- Overtime Hours:

out5.overtime\_hours.multiply(out5.overtime\_rate)

```
out5.base_pay.add(out5.overtime_hours.multiply(out5.overtime_rate)).subtract(out5.deduction_amount)
```

## 4. Transformations

For the transformation component in this ETL process, we implemented the following:

Sequential to Primary Key Transformation: Transformed sequential values from SEQ to a SERIAL PRIMARY KEY for all dimension tables. This approach establishes unique identifiers for each record, ensuring data integrity and supporting optimized indexing.

Date Format Standardization: Standardized all date fields to the YYYY-MM-DD format to maintain consistency across tables, enhancing readability and simplifying date-related queries.

Data Type Conversion: Converted and standardized data types during the ETL process to ensure compatibility with the target schema, which helped prevent data inconsistency and streamlined query processing.

