

# **Fuentes de Datos y Aprovisionamiento**

## **Práctica 3 – IBM Talend**

**José Manuel Bustos Muñoz**

**Ejercicio IBM Talend:** Obtener los datos de ventas de clientes con valores negativos. Para ello se proporciona el software necesario (IBM Talend, PostgreSQL portable), y los datos para el ejercicio (ficheros de texto con ciertos datos, datos en Postgre u otros datos en Access).

Lo primero sería arrancar el software, tanto el PostgreSQL como el IBM Talend.

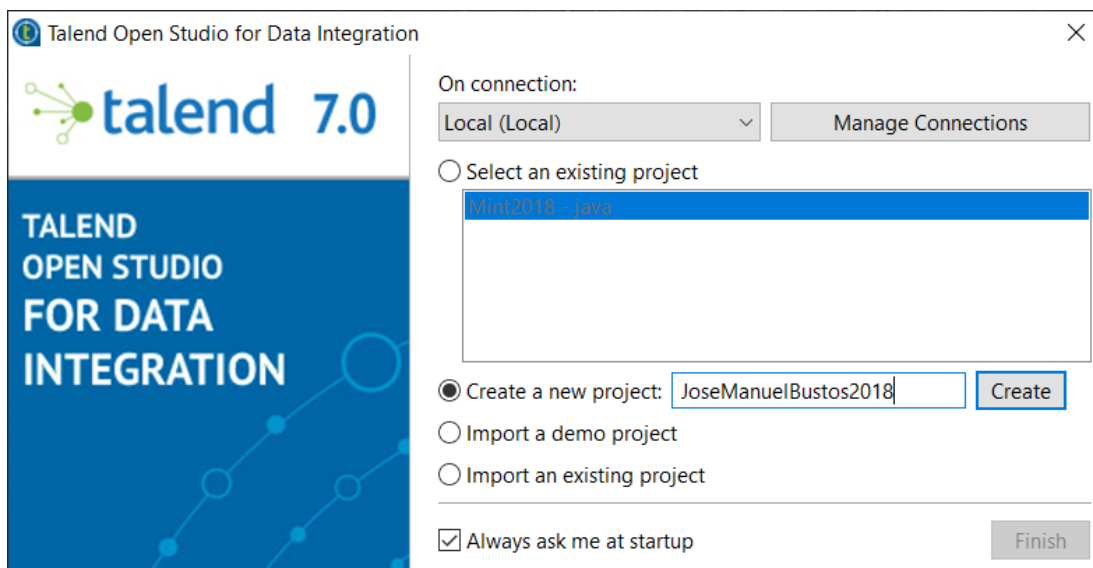
```
PostgreSQL Portable

Type \q to quit and shutdown server

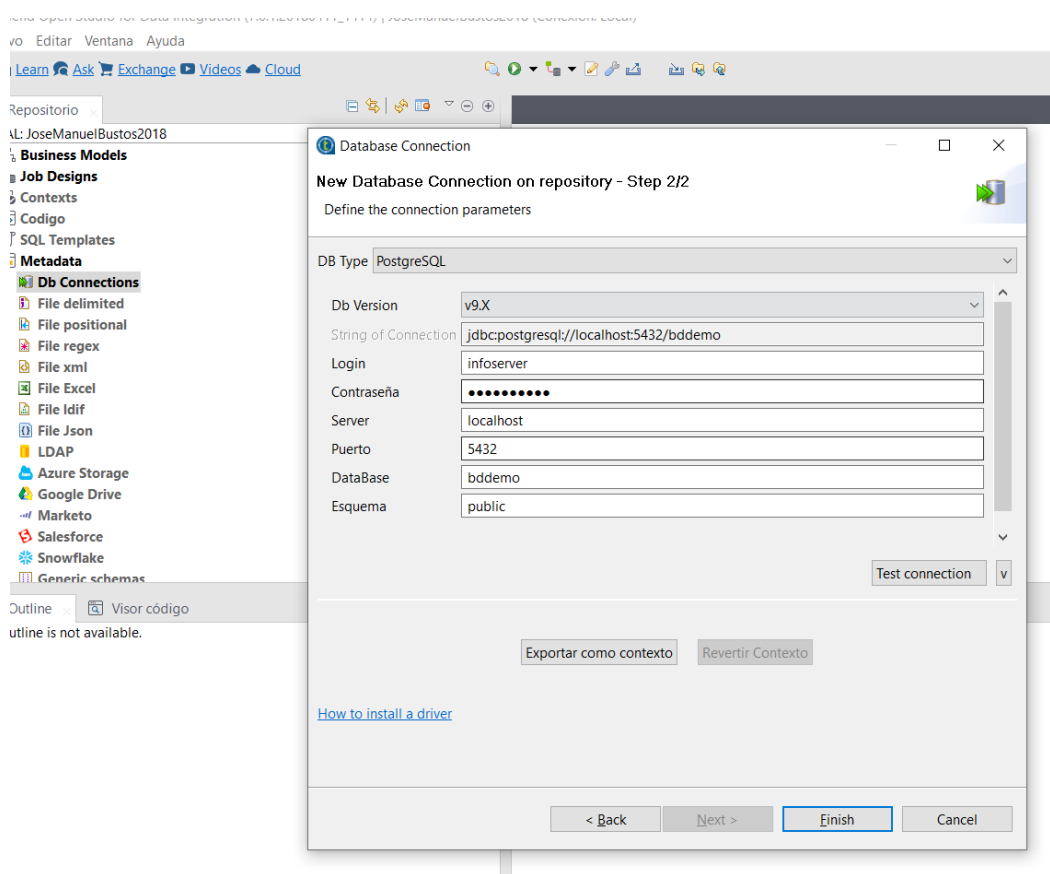
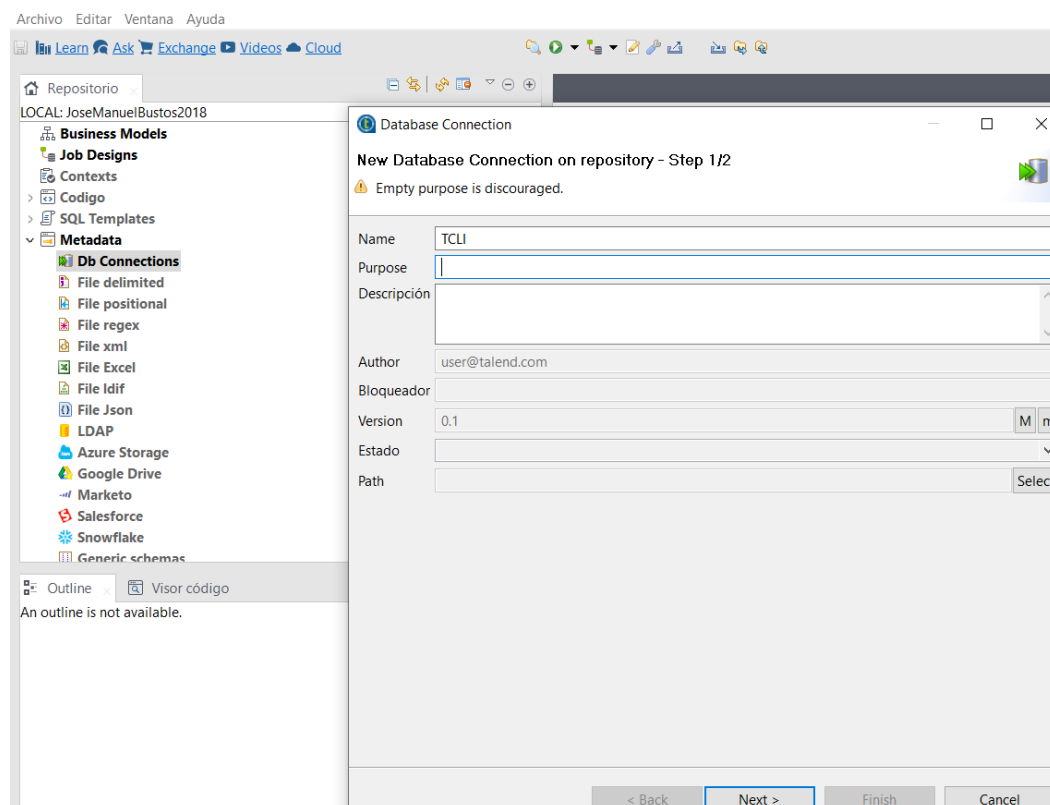
psql (9.6.1)
Digite «help» para obtener ayuda.

postgres=#
```

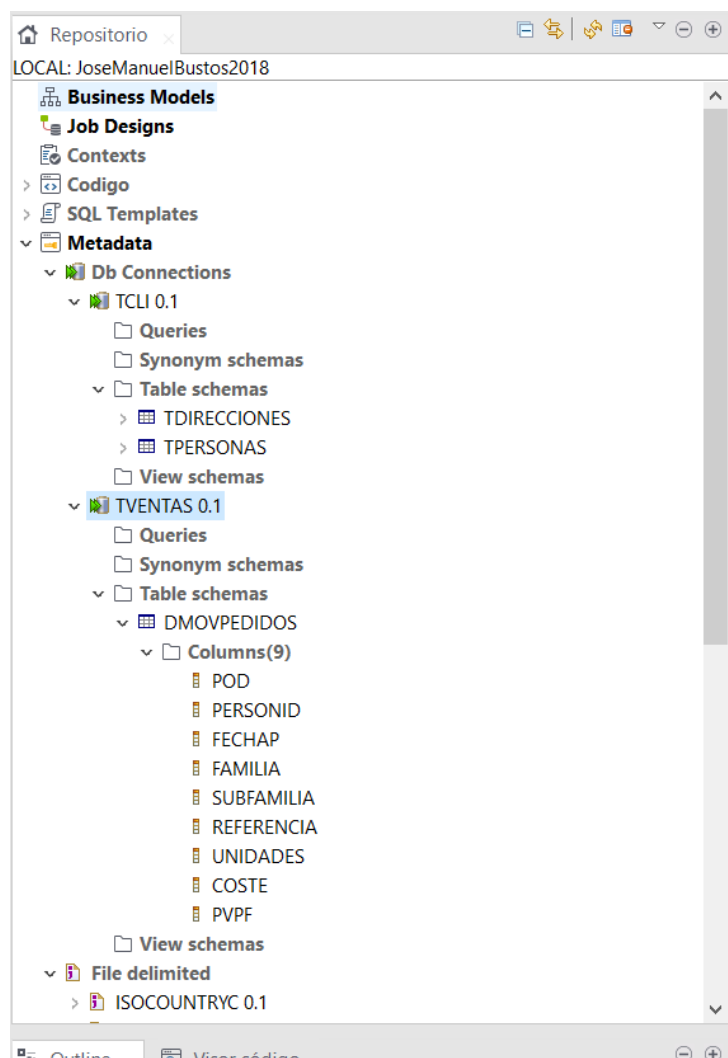
Al arrancar IBM Talend creamos el proyecto para la práctica.



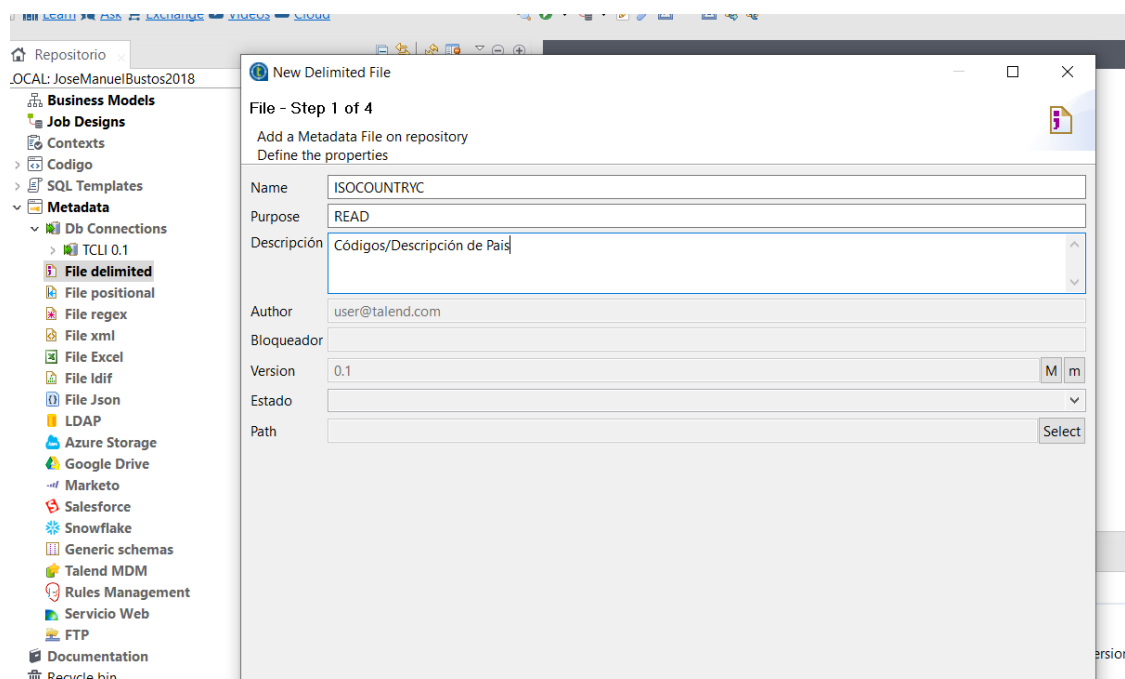
Lo primero sería crear las conexiones con BBDD, tanto con PostgreSQL de donde traemos las tablas de Direcciones y de Personas, como de Access de donde se obtienen los datos de las ventas. Al crear las conexiones y generar las tablas se infieren los esquemas correspondientes.



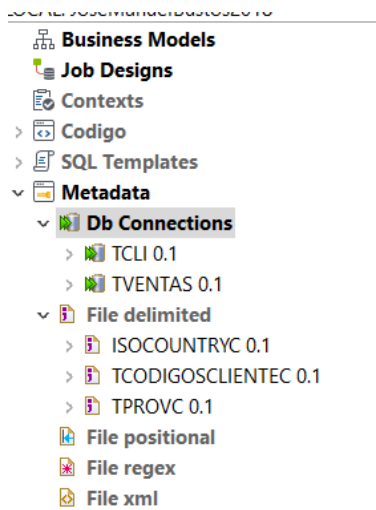
Una vez se crean las conexiones, tablas y esquemas, queda la siguiente jerarquía con las conexiones para los datos de clientes y para los datos de ventas:



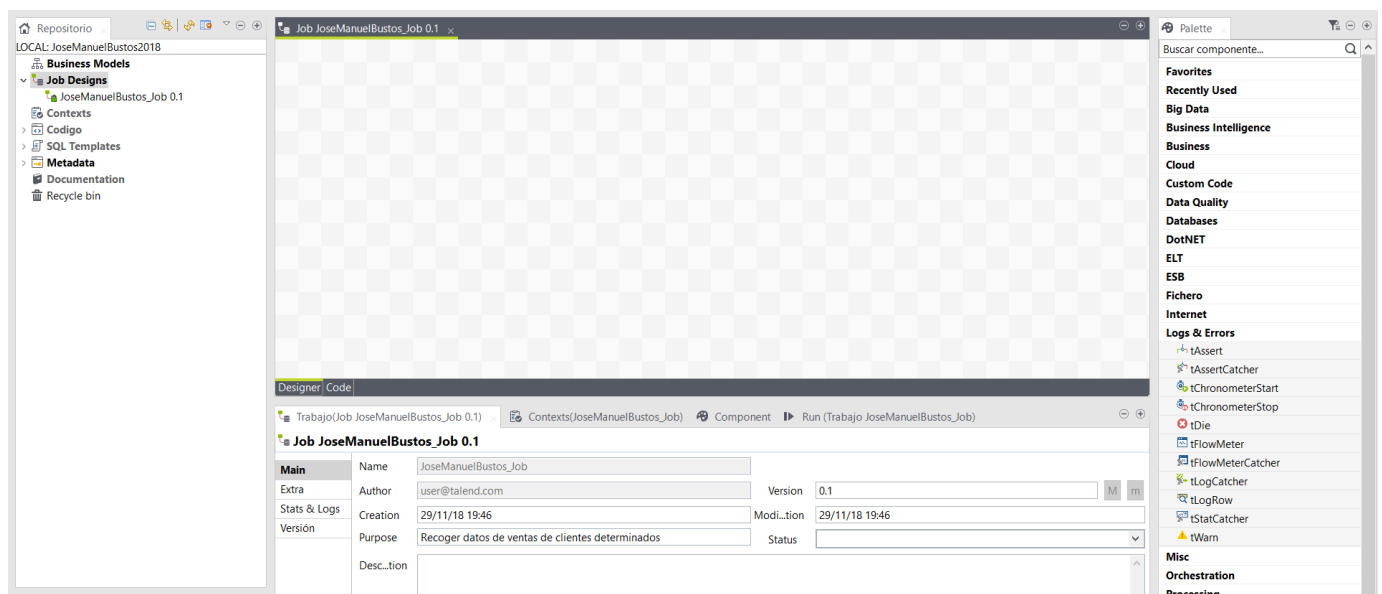
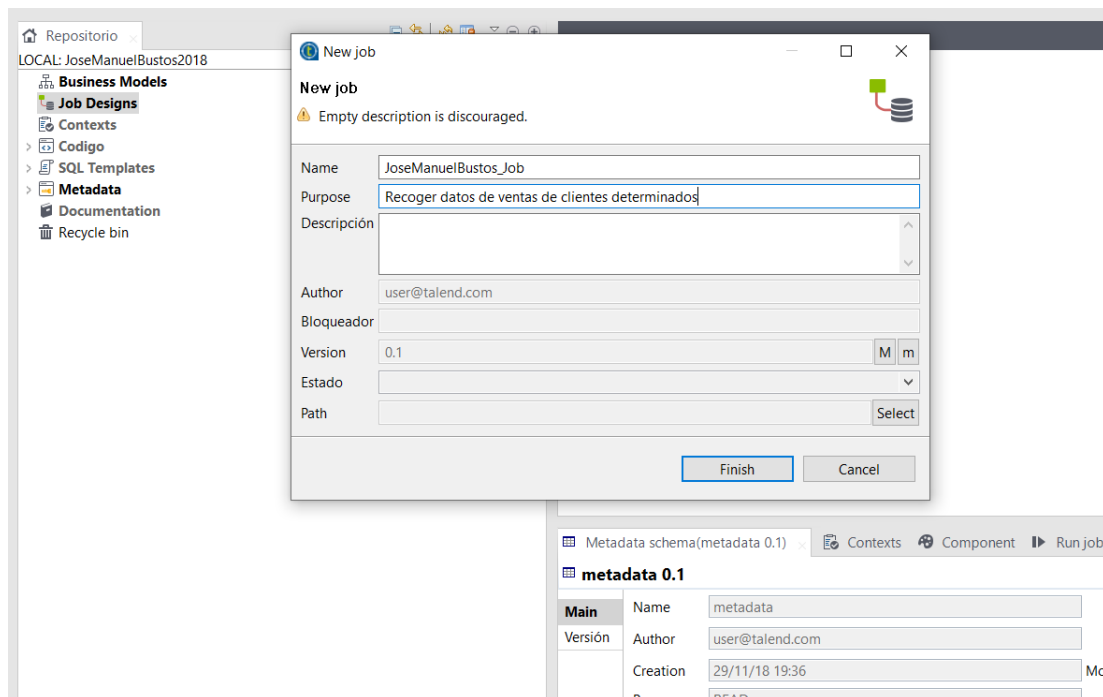
También se crean una serie de secuencias delimitados, para los datos de países, los códigos de los clientes y las provincias.



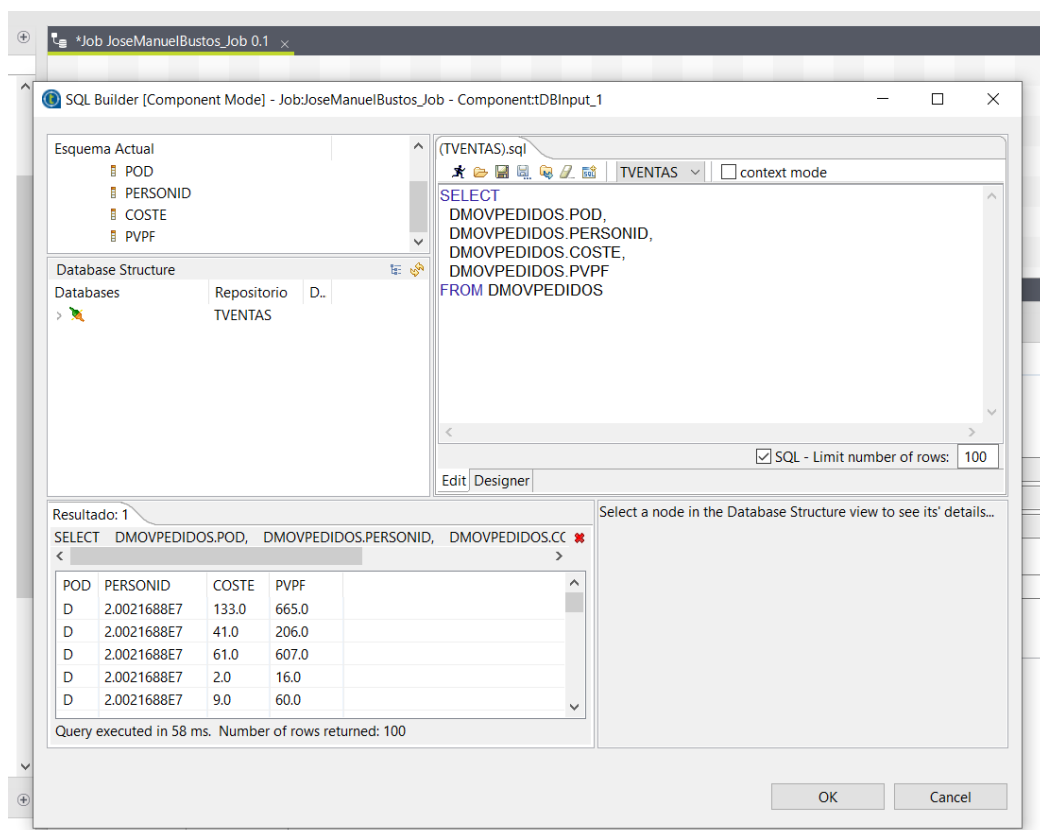
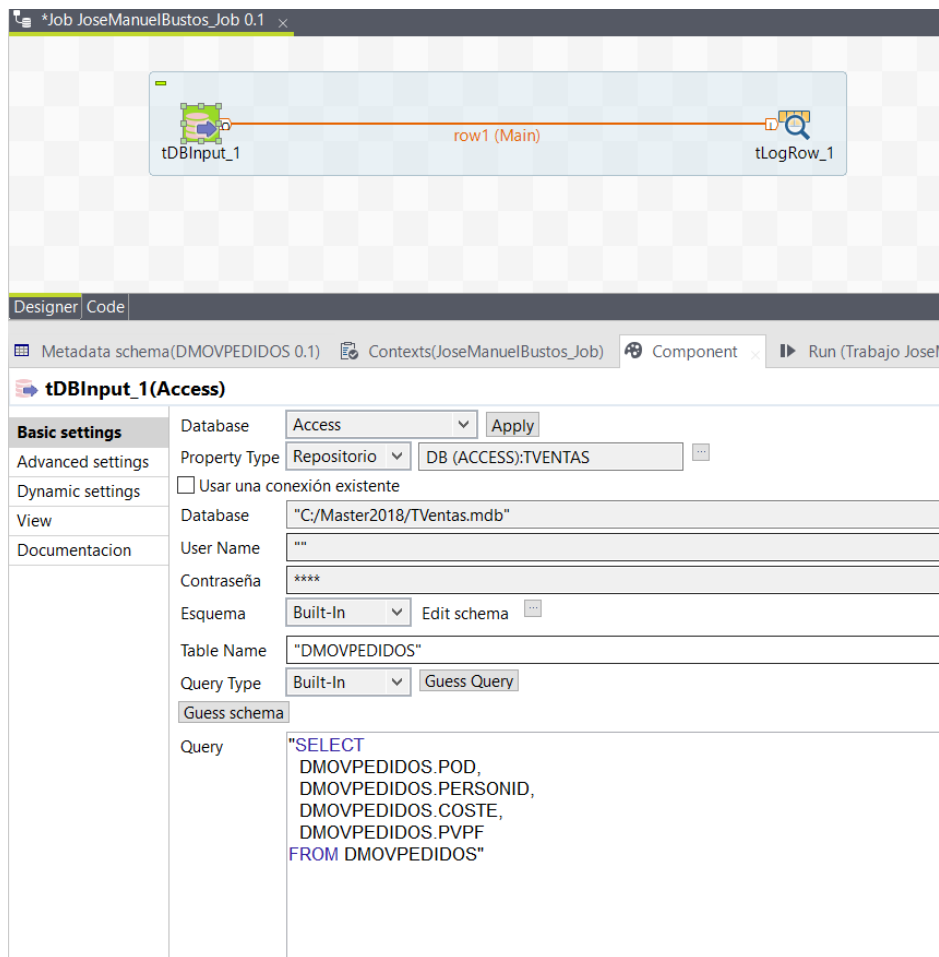
Finalmente, una vez se crean los delimitadores secuenciales, además de las conexiones con BBDD queda así:



Ahora generamos el job (proceso) con el que a partir de los datos obtenidos y el uso de los distintos componentes de la herramienta intentaremos acabar generando el documento Excel con los datos requeridos.



Lo primero que hacemos es poner un elemento de lectura de datos Access, para obtener los datos de ventas, y conectarlo con el componente de log, para escribir en un log de ejecución y ver lo que ocurre en el proceso. En las propiedades del componente Access se configura la query en la que se obtienen los datos del Access.



Al correr el job, se obtienen los datos con la select y se pueden ver en la pantalla de resultado.

The screenshot shows the Talend Studio interface during the execution of a job. The top bar displays the job name 'Job JoseManuelBustos\_Job 0.1' and the execution progress: '100031 rows in 6,48s' and '15427,36 rows/s row1 (Main)'. The main window is divided into several sections. On the left, there is a 'Designer' tab and a 'Code' tab. Below these, the job name 'Trabajo JoseManuelBustos\_Job' is displayed. The 'Execution' tab is active, showing a list of data rows. The rows are displayed in a table format with columns for 'P', '23923275', '43.0', and '289.0'. The status bar at the bottom indicates 'Job JoseManuelBustos\_Job ended at 20:03 29/11/2018. [exit code=0]'. On the right side, there is a 'Default' tab with a table structure showing 'Nombre' and 'Valor' columns.

| Nombre   | Valor                     |
|----------|---------------------------|
| P        | 23923275                  |
| 23923275 | 43.0                      |
| 43.0     | 289.0                     |
| 289.0    | P                         |
| P        | 23923275                  |
| 23923275 | 0.0                       |
| 0.0      | 3.0                       |
| 3.0      | P                         |
| P        | 23923275                  |
| 23923275 | 224.0                     |
| 224.0    | 1120.0                    |
| 1120.0   | P                         |
| P        | 23923275                  |
| 23923275 | 11.0                      |
| 11.0     | 108.0                     |
| 108.0    | P                         |
| P        | 23931833                  |
| 23931833 | 6.0                       |
| 6.0      | 40.0                      |
| 40.0     | P                         |
| P        | 23931833                  |
| 23931833 | 211.0                     |
| 211.0    | 1056.0                    |
| 1056.0   | P                         |
| P        | 23931833                  |
| 23931833 | 34.0                      |
| 34.0     | 170.0                     |
| 170.0    | P                         |
| P        | 23931833                  |
| 23931833 | 5.0                       |
| 5.0      | 24.0                      |
| 24.0     | P                         |
| P        | 23931833                  |
| 23931833 | 12.0                      |
| 12.0     | 77.0                      |
| 77.0     | P                         |
| P        | 23931833                  |
| 23931833 | 1.0                       |
| 1.0      | 10.0                      |
| 10.0     | P                         |
| P        | 23931833                  |
| 23931833 | 9.0                       |
| 9.0      | 46.0                      |
| 46.0     | P                         |
| P        | 23931833                  |
| 23931833 | 4.0                       |
| 4.0      | 22.0                      |
| 22.0     | P                         |
| P        | 23931833                  |
| 23931833 | 36.0                      |
| 36.0     | 182.0                     |
| 182.0    | P                         |
| P        | 23931833                  |
| 23931833 | 5.0                       |
| 5.0      | 47.0                      |
| 47.0     | P                         |
| P        | 23933724                  |
| 23933724 | 11.0                      |
| 11.0     | 113.0                     |
| 113.0    | P                         |
| P        | 23933724                  |
| 23933724 | 19.0                      |
| 19.0     | 94.0                      |
| 94.0     | P                         |
| P        | 23933724                  |
| 23933724 | 32.0                      |
| 32.0     | 216.0                     |
| 216.0    | P                         |
| P        | 23933724                  |
| 23933724 | 22.0                      |
| 22.0     | 224.0                     |
| 224.0    | P                         |
| P        | 23933724                  |
| 23933724 | 13.0                      |
| 13.0     | 63.0                      |
| 63.0     | [statistics] disconnected |

Job JoseManuelBustos\_Job ended at 20:03 29/11/2018. [exit code=0]

Añadiendo los objetos tmap se pueden mapear y hacer transformaciones entre datos de entrada y de salida.

The screenshot shows the Talend Studio interface during the configuration of a job. The 'PedidosIN' table is mapped to the 'SaleOPT' table. The mapping is defined by the expression 'PedidosIN.POD.contains("P"))?1:-1' and the variable 'PosNeg'. The 'Auto map!' button is visible on the right side of the interface.

| Expresión                         | Tipo | Variable |
|-----------------------------------|------|----------|
| PedidosIN.POD.contains("P"))?1:-1 | int  | PosNeg   |

También haremos uso de objetos fileInputDelimited para obtener los registros que cumplan la estructura de datos y enviar al log las entradas erróneas.

The screenshot shows the SAP Business Builder Designer interface. At the top, a diagram shows a connection from 'CodigosClienteC' to 'tLogRow\_2' labeled 'row3 (Main)'. Below the diagram, the 'Designer' tab is active, showing the configuration for 'CodigosClienteC(tFileInputDelimited\_1)'. The 'Basic settings' tab is selected, displaying the following properties:

- Property Type: Repositorio (dropdown)
- Esquema: Repositorio (dropdown)
- Nombre de Archivo/Flujo: "C:/Master2018/CodigosClientesC.txt"
- Separador de Fila: "\n"
- Separador de Campo: "|" (dropdown)
- Encabezado: 1 (dropdown)
- Pie de Página: 0 (dropdown)
- Limit: (empty field)
- Options: ☒ Skip empty rows, ☐ Descomprimir como fichero zip, ☐ Die on error

Y para escribir o generar archivos Excel con la información que configuremos, hacemos uso de los componentes FileOutputExcel.

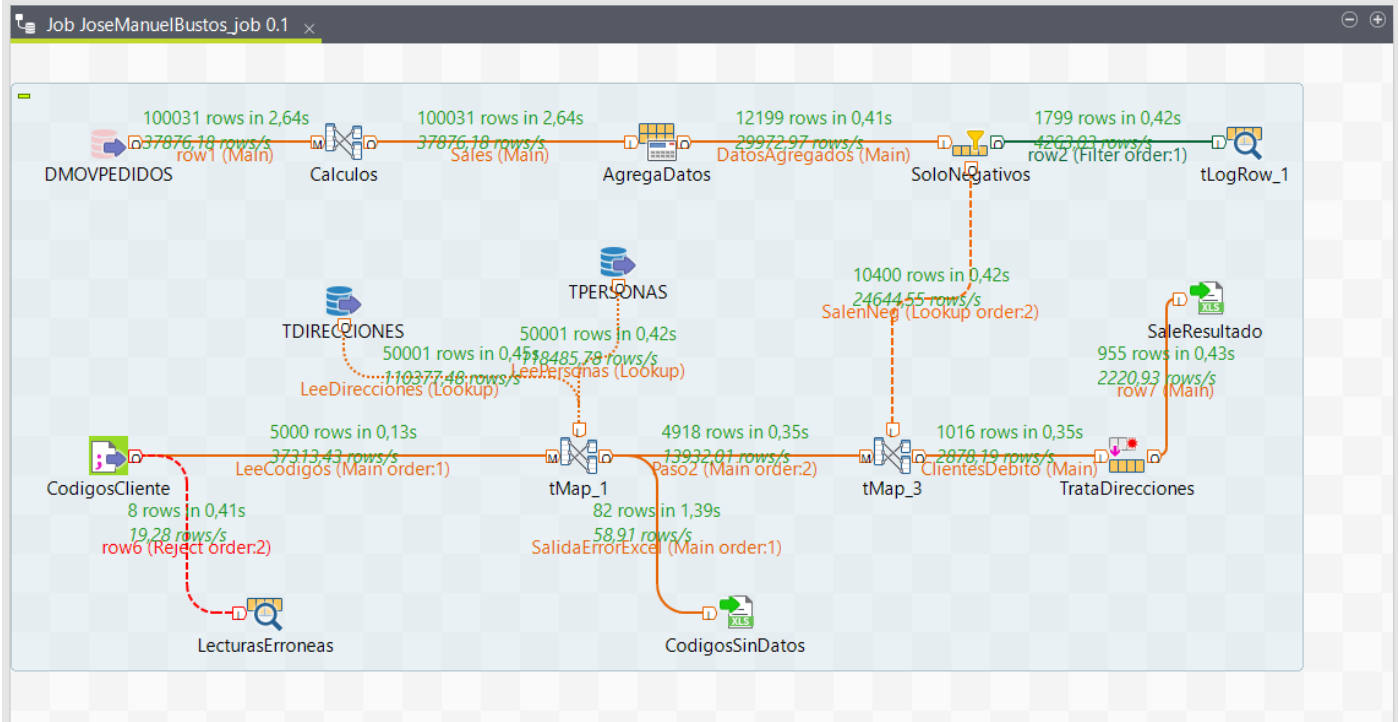
The screenshot shows the SAP Business Builder Designer interface. At the top, a diagram shows a connection from 'TPERSONAS' to 'TDIRECCIONES' to 'SalenNeg (Lookup order:2)' to 'tFileOutputExcel\_2'. Below the diagram, the 'Designer' tab is active, showing the configuration for 'tFileOutputExcel\_2'. The 'Basic settings' tab is selected, displaying the following properties:

- Property Type: Built-In (dropdown)
- Advanced settings: ☒ Write excel2007 file format(xlsx), ☐ Use Output Stream
- Nombre de Archivo: "C:/Master2018/Resultado.xlsx"
- Nombre Hoja: "ClientesNegativos"
- Options: ☒ Include header, ☐ Añadir al fichero existente, ☐ Is absolute Y pos.
- Font: None (dropdown)
- Define all columns auto size: ☐ Auto size
- Define column auto size: (table with columns and checkboxes)

| Column           | Auto size                |
|------------------|--------------------------|
| personid         | <input type="checkbox"/> |
| docidentid       | <input type="checkbox"/> |
| birthdate        | <input type="checkbox"/> |
| deathdate        | <input type="checkbox"/> |
| sex              | <input type="checkbox"/> |
| addressnum       | <input type="checkbox"/> |
| NombreyApellidos | <input type="checkbox"/> |
| address          | <input type="checkbox"/> |
| city             | <input type="checkbox"/> |
| pc               | <input type="checkbox"/> |
| province         | <input type="checkbox"/> |
| country          | <input type="checkbox"/> |
| COSTE            | <input type="checkbox"/> |



El proceso completo quedaría de la siguiente forma, añadiendo los componentes para leer de las tablas de PostgreSQL, la escritura en fichero Excel de los resultados del job y de posibles errores, y de algún otro componente como el “tExtractRegexFields” para el tratamiento de expresiones regulares, que en este caso se utiliza para estandarizar y validar las direcciones.



Vistazo a la Excel de salida:

| Resultado.xlsx - Excel  |          |            |            |            |     |                                 |            |                                 |                           |       |          |         |       |       |  |
|---|----------|------------|------------|------------|-----|---------------------------------|------------|---------------------------------|---------------------------|-------|----------|---------|-------|-------|--|
| Inicio  |          |            |            |            |     |                                 |            |                                 |                           |       |          |         |       |       |  |
| Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista ¿Qué desea hacer? |          |            |            |            |     |                                 |            |                                 |                           |       |          |         |       |       |  |
| Pegar Cortar Copiar Copiar formato Fuente Alineación Número Estilos Celdas Edición      |          |            |            |            |     |                                 |            |                                 |                           |       |          |         |       |       |  |
| A1  | personid |            |            |            |     |                                 |            |                                 |                           |       |          |         |       |       |  |
|   | A        | B          | C          | D          | E   | F                               | G          | H                               | I                         | J     | K        | L       | M     | N     |  |
| 1   | personid | docidentid | birthdate  | deathdate  | sex | NombreyApellidos                | addressnum | address                         | city                      | pc    | province | country | COSTE | PVPF  |  |
| 2   | 20656818 | 00079287Y  | 09-11-1961 | 01-01-1900 | F   | AMELIAAVILAMEDINA               | 0          | AVDA DEL TRIUNFO 12 D           | DOS HERMANAS              | 41700 | 41       | ES      | 427   | 3483  |  |
| 3   | 20610240 | 00234985V  | 08-10-1965 | 01-01-1900 | F   | ANA ESPERANZAGAVIRAPEREZ        | 0          | CALLE DUNAS 2 B                 | ARBATE                    | 11160 | 11       | ES      | 452   | 4528  |  |
| 4   | 20422094 | 00343578G  | 10-05-1950 | 01-01-1900 | 1   | LORENZOGOMEZGOMEZ               | 0          | CALLE PEPITA JIMENEZ 7 M        | MARCHENA                  | 41620 | 41       | ES      | 592   | 4332  |  |
| 5   | 20535115 | 00369511Q  | 19-07-1945 | 01-01-1900 | 2   | MERCEDESFLORESVELASCO           | 0          | CALLE LEPANTO 65 L              | LA LINEA DE LA CONCEPCION | 11300 | 11       | ES      | 128   | 1090  |  |
| 6   | 20508669 | 00375280N  | 02-03-1948 | 22-01-2004 | 1   | JOSE AMADEOFIGUEROACUENCA       | 0          | AV VILLARRASA 31 R              | ROCIANA DEL CONDADO       | 21720 | 21       | ES      | 262   | 1480  |  |
| 7   | 20671242 | 01013444H  | 01-02-1906 | 01-01-1900 | 1   | MANUELCASTROLOPEZ               | 0          | CL ALCALA DE CHIVERT 3 A        | ALCALA DE GUADAIRA        | 41500 | 41       | ES      | 578   | 3349  |  |
| 8   | 20573658 | 02516292T  | 27-01-1959 | 01-01-1900 | 1   | JUAN IGNACIOGARCIAMATEU         | 0          | CL CANARIAS 4 1º DCHA S         | SAN FERNANDO              | 11100 | 11       | ES      | 387   | 2081  |  |
| 9   | 20539343 | 03030274R  | 16-01-1940 | 01-01-1900 | 1   | IGNACIOMOLINABENITO             | 0          | CALLE FUENSANTA 6 C             | CORCOYA                   | 41599 | 41       | ES      | 272   | 1734  |  |
| 10  | 20604856 | 03386536Q  | 07-06-1940 | 01-01-1900 | 1   | PEDROHERREROLLAMAS              | 0          | CL C NTRA SRA DE LAS NIEVES 9 B | ORMUJOS                   | 41930 | 41       | ES      | 656   | 4203  |  |
| 11  | 20629178 | 03452151N  | 30-07-1969 | 04-10-2001 | 2   | INESROBLEDOGONZALEZ             | 0          | PLZ ENCINA DEL REY 1 PTA 13 S   | SEVILLA                   | 41019 | 41       | ES      | 126   | 1085  |  |
| 12  | 20382575 | 04545311M  | 10-07-1956 | 24-07-2007 | 1   | RODOLFORODRIGUEZSEGARRA         | 0          | CL ACUEDUCTO 8 3 A S            | SEVILLA                   | 41008 | 41       | ES      | 569   | 5315  |  |
| 13  | 20596117 | 05242462A  | 24-05-1960 | 24-10-2001 | 1   | PABLOFERNANDEZCASTRO            | 0          | CALLE PASTORA 7 1 DCHA C        | CADIZ                     | 11002 | 11       | ES      | 360   | 2855  |  |
| 14  | 20588271 | 06523231W  | 19-04-1965 | 24-10-2001 | 1   | FRANCISCO JAVIERHERNANDEZSUAREZ | 0          | CL CARLOS I ESPAÑA 11 D         | DOS HERMANAS              | 41700 | 41       | ES      | 158   | 1185  |  |
| 15  | 20672842 | 06788530B  | 14-02-1921 | 01-01-1900 | 2   | PRUDENCIOJORGUÑUEZ              | 0          | UR LAS PARCELAS DEL GOLF 46 C   | CHICLANA DE LA FRONTERA   | 11130 | 11       | ES      | 62    | 421   |  |
| 16  | 20373235 | 06929119R  | 11-12-1945 | 01-01-1900 | 2   | MARIA HILARIAREDONDOGALINDO     | 0          | CL J M PASCUAL Y PASCUAL 3 1C C | CADIZ                     | 11008 | 11       | ES      | 1962  | 11358 |  |
| 17  | 20581710 | 07468818M  | 01-07-1967 | 01-01-1900 | 2   | MARIA DEL CARMENPEREZFLORES     | 0          | CL SANTA ISABEL 47 V            | VILLACARRILLO             | 23300 | 23       | ES      | 637   | 4568  |  |
| 18  | 20663290 | 07723380A  | 22-09-1941 | 01-01-1900 | 1   | JOSE MANUELGONZALEZMARTIN       | 0          | CL ARQUITECTO JOSE GALNARES C/  | SEVILLA                   | 41013 | 41       | ES      | 141   | 1088  |  |
| 19  | 20449353 | 07751224V  | 07-07-1945 | 12-06-2001 | 1   | RAFAELSCUERSEGUERA              | 0          | CAMNO CAMINO DE LA ATUNARA 15   | LA LINEA DE LA CONCEPCION | 11300 | 11       | ES      | 218   | 1371  |  |
| 20  | 20519283 | 08329606H  | 08-12-1928 | 01-01-1900 | 2   | JUANADELGADODURAN               | 0          | CALLE VIOLETA 1 2 D A           | ALCALA DE GUADAIRA        | 41500 | 41       | ES      | 173   | 1553  |  |
| 21  | 20397806 | 08442019F  | 09-01-1929 | 01-01-1900 | 2   | RUFINOFELIXMORILLON             | 0          | PASEO VELADA 5 1ºQ L            | LA LINEA DE LA CONCEPCION | 11300 | 11       | ES      | 432   | 3043  |  |
| 22  | 20458035 | 08917072H  | 18-01-1969 | 14-02-2002 | 1   | FRANCISCO JOSECOLLADOCASTELAO   | 0          | CALLE MAESTRA JOSEFA NAVARRO 2  | CORIA DEL RIO             | 41100 | 41       | ES      | 137   | 1004  |  |
| 23  | 20351048 | 08918559X  | 21-01-1971 | 20-07-2001 | 2   | ELENAROMEROGARCIA               | 0          | CL DALI 10 L                    | LA ALGABA                 | 41980 | 41       | ES      | 206   | 2066  |  |
| 24  | 20609931 | 08922357J  | 22-10-1974 | 01-01-1900 | 1   | ANGEL MANUELJIMENEZGUERRERO     | 0          | CL JOSE LEGASSA 13 M            | MANZANARES                | 13200 | 13       | ES      | 246   | 1345  |  |
| 25  | 20346949 | 10450559A  | 12-12-1936 | 17-08-2007 | 1   | MACARIO DIAZALONSO              | 0          | AV TOMAS GARCIA FIGUERAS 11 1 B | JEREZ DE LA FRONTERA      | 11407 | 11       | ES      | 713   | 4570  |  |
| 26  | 20566479 | 12863675M  | 16-05-1931 | 24-10-2001 | 1   | JUANGONZALEZGUTIERREZ           | 0          | CL CALLEJAS 6 A                 | ARCOS DE LA FRONTERA      | 11630 | 11       | ES      | 497   | 4794  |  |

También se genera la Excel con los errores encontrados:

ErroresLK01.xlsx

ArchivoInicioInsertarDiseño de páginaFórmulasDatosRevisarVista¿Qué desea hacer?

Cortar

Copiar

Copiar formato

Portapapeles

Calibri

11

A A

N K S

A

Fuente

Ajustar texto

Combinar y centrar

Alineación

General

% 000

0,00

0,0

Número

Formato condicional

A1

X ✓ fx

CODCLI

|    | A        | B        | C          | D | E | F | G | H | I | J | K | L |
|----|----------|----------|------------|---|---|---|---|---|---|---|---|---|
| 1  | CODCLI   | personid | personid_1 |   |   |   |   |   |   |   |   |   |
| 2  | 20430109 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 3  | 20518633 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 4  | 20532848 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 5  | 20390089 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 6  | 20638195 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 7  | 20437225 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 8  | 20416701 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 9  | 20428443 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 10 | 20662291 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 11 | 20662128 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 12 | 20351526 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 13 | 20498161 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 14 | 20525775 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 15 | 20513512 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 16 | 20368489 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 17 | 20508586 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 18 | 20409419 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 19 | 20510169 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 20 | 20325575 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 21 | 20388300 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 22 | 20440348 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 23 | 20644598 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 24 | 20553475 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 25 | 20394045 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 26 | 20437477 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 27 | 20455519 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 28 | 20348071 | 0        | 0          |   |   |   |   |   |   |   |   |   |
| 29 | 20463734 | 0        | 0          |   |   |   |   |   |   |   |   |   |

PrimerError