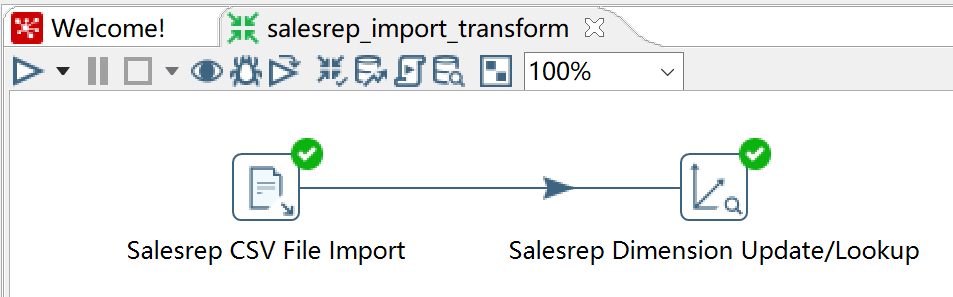
**CIS 9440 Homework #2 Summer 2020 Pentaho Data Integration – ETL**

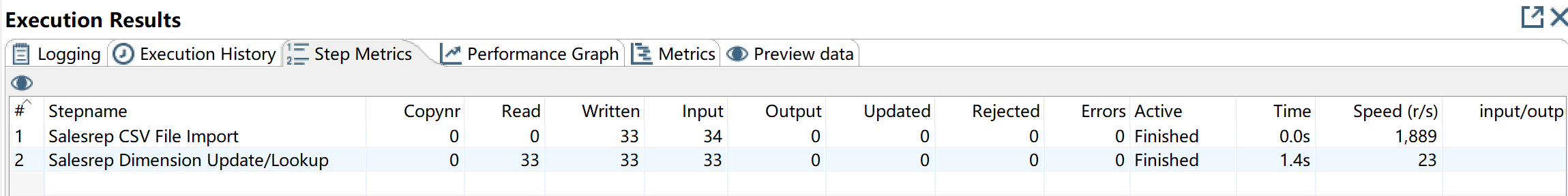
**Exercise 1 – Importing and Transforming Sales Representatives as Type 2 SCD**

**Answers:**

Import the salesrep csv file. Remember to uncheck lazy conversion. After getting fields, remember to adjust the data type, format, and length. Preview data to check errors.

To insert the dimension update (Type 2 SCD), connect to the Oracle Cloud, put in keys then get fields, execute SQL to create a table in the Oracle SQL Developer. Finish the transformation by running the process. Get a sample data by querying.





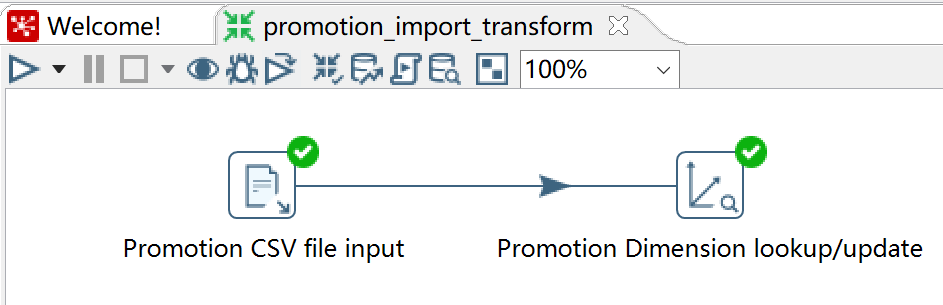
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| FIRST\_NAME | LAST\_NAME | EMAIL | PHONE\_NUMBER | JOB\_ID | COMMISSION\_PCT | SALESREP\_DIM\_ID | VERSION | DATE\_FROM | DATE\_TO | SALESREP\_ID | HIRE\_DATE | SALARY | MANAGER\_ID | DEPARTMENT\_ID |
|  |  |  |  |  |  | 0 | 1 |  |  |  |  |  |  |  |
| Alberto | Errazuriz | AERRAZUR | 011.44.1344.429278 | SA\_MAN | 0.3 | 1 | 1 | 01-JAN-00 | 31-DEC-99 | 147 | 10-MAR-05 | 12000 | 100 | 80 |
| Gerald | Cambrault | GCAMBRAU | 011.44.1344.619268 | SA\_MAN | 0.3 | 2 | 1 | 01-JAN-00 | 31-DEC-99 | 148 | 15-OCT-07 | 11000 | 100 | 80 |
| Eleni | Zlotkey | EZLOTKEY | 011.44.1344.429018 | SA\_MAN | 0.2 | 3 | 1 | 01-JAN-00 | 31-DEC-99 | 149 | 29-JAN-08 | 10500 | 100 | 80 |
| Peter | Tucker | PTUCKER | 011.44.1344.129268 | SA\_REP | 0.3 | 4 | 1 | 01-JAN-00 | 31-DEC-99 | 150 | 30-JAN-05 | 10000 | 145 | 80 |
| David | Bernstein | DBERNSTE | 011.44.1344.345268 | SA\_REP | 0.25 | 5 | 1 | 01-JAN-00 | 31-DEC-99 | 151 | 24-MAR-05 | 9500 | 145 | 80 |
| Peter | Hall | PHALL | 011.44.1344.478968 | SA\_REP | 0.25 | 6 | 1 | 01-JAN-00 | 31-DEC-99 | 152 | 20-AUG-05 | 9000 | 145 | 80 |
| Christopher | Olsen | COLSEN | 011.44.1344.498718 | SA\_REP | 0.2 | 7 | 1 | 01-JAN-00 | 31-DEC-99 | 153 | 30-MAR-06 | 8000 | 145 | 80 |
| Nanette | Cambrault | NCAMBRAU | 011.44.1344.987668 | SA\_REP | 0.2 | 8 | 1 | 01-JAN-00 | 31-DEC-99 | 154 | 09-DEC-06 | 7500 | 145 | 80 |
| Oliver | Tuvault | OTUVAULT | 011.44.1344.486508 | SA\_REP | 0.15 | 9 | 1 | 01-JAN-00 | 31-DEC-99 | 155 | 23-NOV-07 | 7000 | 145 | 80 |
| Janette | King | JKING | 011.44.1345.429268 | SA\_REP | 0.35 | 10 | 1 | 01-JAN-00 | 31-DEC-99 | 156 | 30-JAN-04 | 10000 | 146 | 80 |
| Patrick | Sully | PSULLY | 011.44.1345.929268 | SA\_REP | 0.35 | 11 | 1 | 01-JAN-00 | 31-DEC-99 | 157 | 04-MAR-04 | 9500 | 146 | 80 |
| Allan | McEwen | AMCEWEN | 011.44.1345.829268 | SA\_REP | 0.35 | 12 | 1 | 01-JAN-00 | 31-DEC-99 | 158 | 01-AUG-04 | 9000 | 146 | 80 |
| Lindsey | Smith | LSMITH | 011.44.1345.729268 | SA\_REP | 0.3 | 13 | 1 | 01-JAN-00 | 31-DEC-99 | 159 | 10-MAR-05 | 8000 | 146 | 80 |
| Louise | Doran | LDORAN | 011.44.1345.629268 | SA\_REP | 0.3 | 14 | 1 | 01-JAN-00 | 31-DEC-99 | 160 | 15-DEC-05 | 7500 | 146 | 80 |

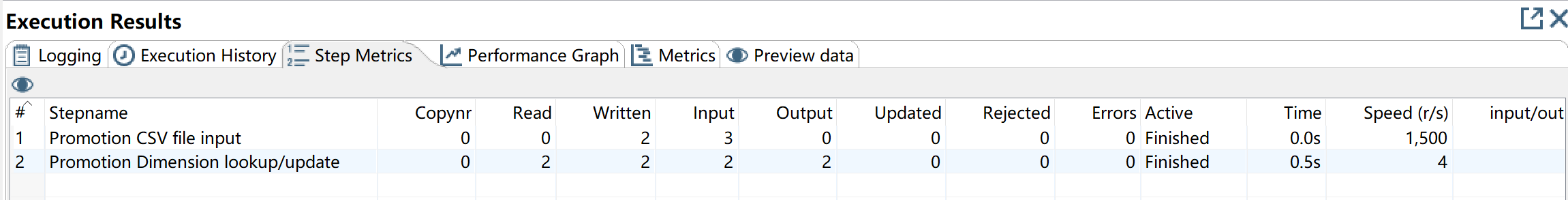
**Exercise 2 – Importing and Transforming Promotions as Type 1 SCD**

**Answers:**

Import the promotion csv file. Remember to uncheck lazy conversion. After getting fields, remember to adjust the data type, format, and length. Preview data to check errors.

To insert the dimension punch through (Type 1 SCD), connect to the Oracle Cloud, put in keys then get fields, execute SQL to create a table in the Oracle SQL Developer. Finish the transformation by running the process. Get a sample data by querying.





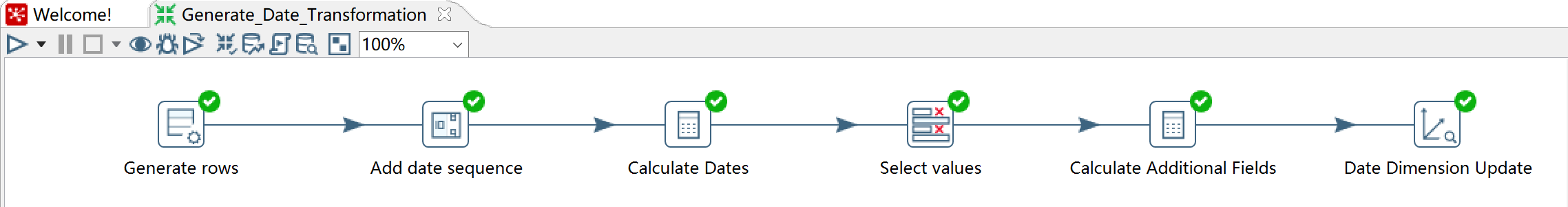
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PROMOTION\_DIM\_ID | VERSION | DATE\_FROM | DATE\_TO | PROMO\_ID | PROMO\_NAME |
| 0 | 1 |  |  |  |  |
| 1 | 1 | 01-JAN-00 | 31-DEC-99 | 1 | everyday low price |
| 2 | 1 | 01-JAN-00 | 31-DEC-99 | 2 | blowout sale |

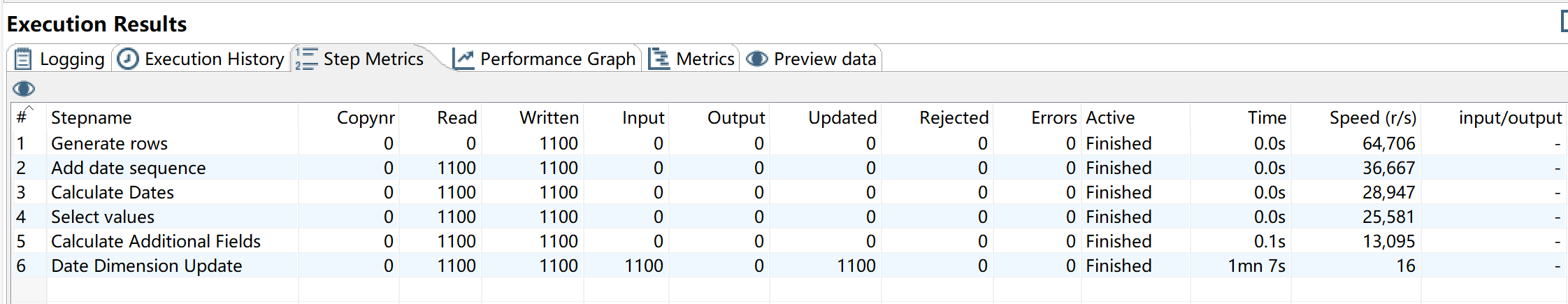
**Exercise 3 – Embellishing the Date Dimension with Additional Date information**

**Answers:**

Generate repetitive date rows. Add data sequence. Increment by 1. Calculate sales date by adding startdate and increment\_date. Select sales\_date to calculate other date fields.

To get the dates updated, connect to the Oracle Cloud, put in keys then get fields, execute SQL to create a table in the Oracle SQL Developer. Finish the transformation by running the process. Get a sample data by querying.





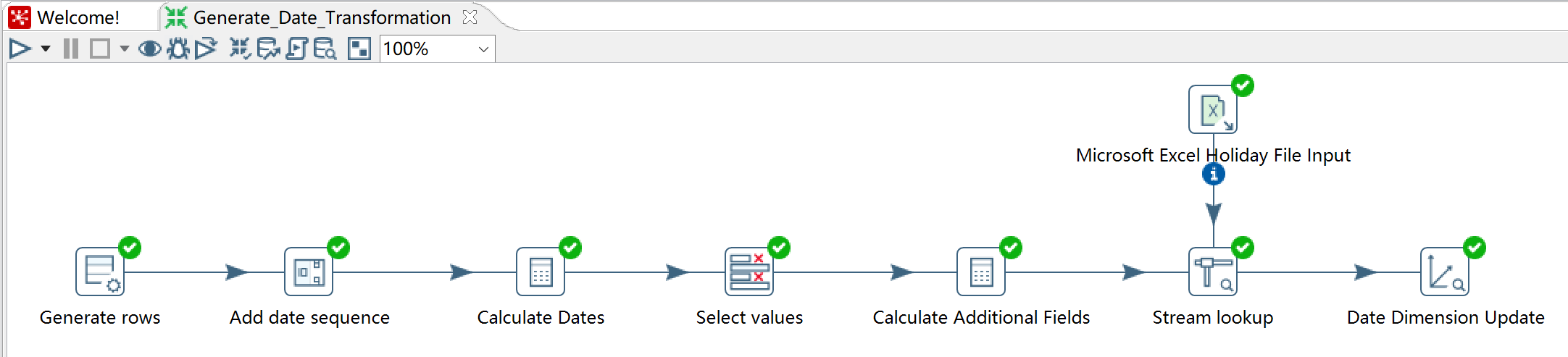
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SALES\_MONTH\_NAME | SALES\_DAY\_OF\_WEEK\_NAME | DATE\_DIM\_ID | VERSION | DATE\_FROM | DATE\_TO | SALES\_DATE | SALES\_DAY\_OF\_YEAR | SALES\_MONTH | SALES\_YEAR | SALES\_QUARTER | SALES\_DAY\_OF\_WEEK | SALES\_DAY\_OF\_MONTH |
|  |  | 0 | 1 |  |  |  |  |  |  |  |  |  |
| January | Monday | 1 | 1 | 01-JAN-00 | 31-DEC-99 | 02-JAN-06 | 2 | 1 | 2006 | 1 | 2 | 2 |
| January | Tuesday | 2 | 1 | 01-JAN-00 | 31-DEC-99 | 03-JAN-06 | 3 | 1 | 2006 | 1 | 3 | 3 |
| January | Wednesday | 3 | 1 | 01-JAN-00 | 31-DEC-99 | 04-JAN-06 | 4 | 1 | 2006 | 1 | 4 | 4 |
| January | Thursday | 4 | 1 | 01-JAN-00 | 31-DEC-99 | 05-JAN-06 | 5 | 1 | 2006 | 1 | 5 | 5 |
| January | Friday | 5 | 1 | 01-JAN-00 | 31-DEC-99 | 06-JAN-06 | 6 | 1 | 2006 | 1 | 6 | 6 |
| January | Saturday | 6 | 1 | 01-JAN-00 | 31-DEC-99 | 07-JAN-06 | 7 | 1 | 2006 | 1 | 7 | 7 |
| January | Sunday | 7 | 1 | 01-JAN-00 | 31-DEC-99 | 08-JAN-06 | 8 | 1 | 2006 | 1 | 1 | 8 |
| January | Monday | 8 | 1 | 01-JAN-00 | 31-DEC-99 | 09-JAN-06 | 9 | 1 | 2006 | 1 | 2 | 9 |
| January | Tuesday | 9 | 1 | 01-JAN-00 | 31-DEC-99 | 10-JAN-06 | 10 | 1 | 2006 | 1 | 3 | 10 |
| January | Wednesday | 10 | 1 | 01-JAN-00 | 31-DEC-99 | 11-JAN-06 | 11 | 1 | 2006 | 1 | 4 | 11 |
| January | Thursday | 11 | 1 | 01-JAN-00 | 31-DEC-99 | 12-JAN-06 | 12 | 1 | 2006 | 1 | 5 | 12 |
| January | Friday | 12 | 1 | 01-JAN-00 | 31-DEC-99 | 13-JAN-06 | 13 | 1 | 2006 | 1 | 6 | 13 |
| January | Saturday | 13 | 1 | 01-JAN-00 | 31-DEC-99 | 14-JAN-06 | 14 | 1 | 2006 | 1 | 7 | 14 |
| January | Sunday | 14 | 1 | 01-JAN-00 | 31-DEC-99 | 15-JAN-06 | 15 | 1 | 2006 | 1 | 1 | 15 |
| January | Monday | 15 | 1 | 01-JAN-00 | 31-DEC-99 | 16-JAN-06 | 16 | 1 | 2006 | 1 | 2 | 16 |
| January | Tuesday | 16 | 1 | 01-JAN-00 | 31-DEC-99 | 17-JAN-06 | 17 | 1 | 2006 | 1 | 3 | 17 |
| January | Wednesday | 17 | 1 | 01-JAN-00 | 31-DEC-99 | 18-JAN-06 | 18 | 1 | 2006 | 1 | 4 | 18 |
| January | Thursday | 18 | 1 | 01-JAN-00 | 31-DEC-99 | 19-JAN-06 | 19 | 1 | 2006 | 1 | 5 | 19 |
| January | Friday | 19 | 1 | 01-JAN-00 | 31-DEC-99 | 20-JAN-06 | 20 | 1 | 2006 | 1 | 6 | 20 |
| January | Saturday | 20 | 1 | 01-JAN-00 | 31-DEC-99 | 21-JAN-06 | 21 | 1 | 2006 | 1 | 7 | 21 |

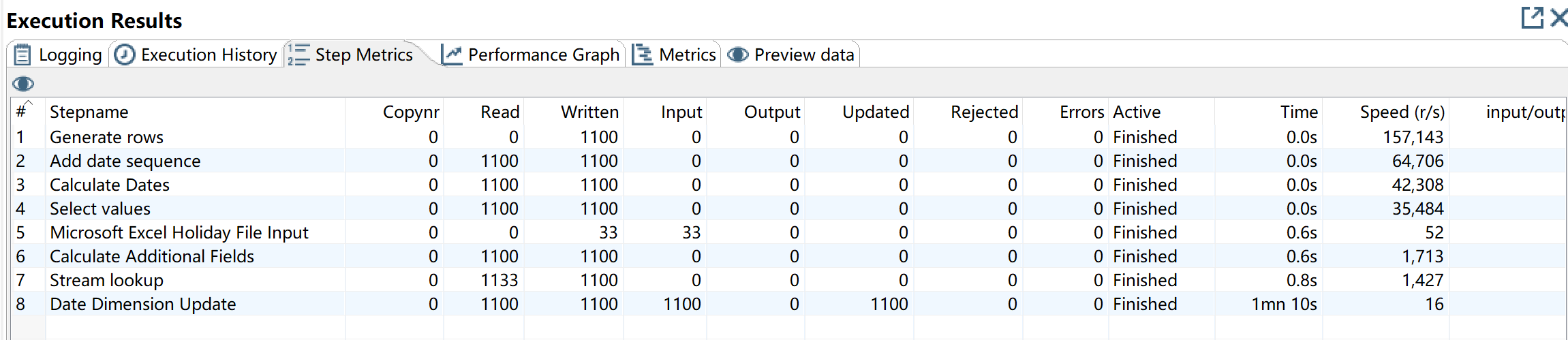
**Exercise 4 – Embellishing the Date Dimension with Holiday information**

**Answers:**

Input 3 new fields into the dimension by importing excel. Add a stream lookup to match the sales date with holiday date to show whether it is a holiday or not.

Update the fields in Date Dimension Update and run the process. Get a sample data by querying.



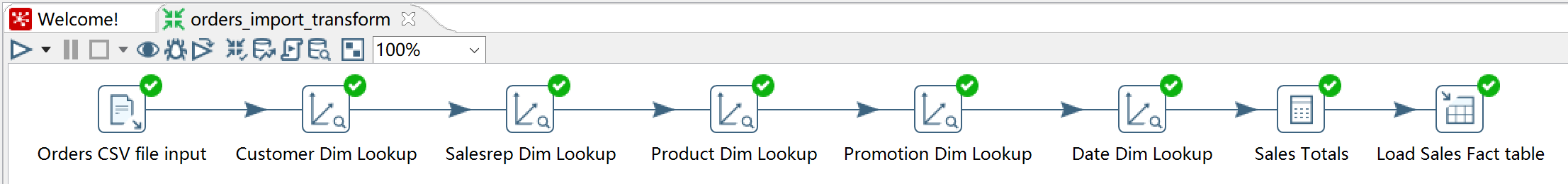


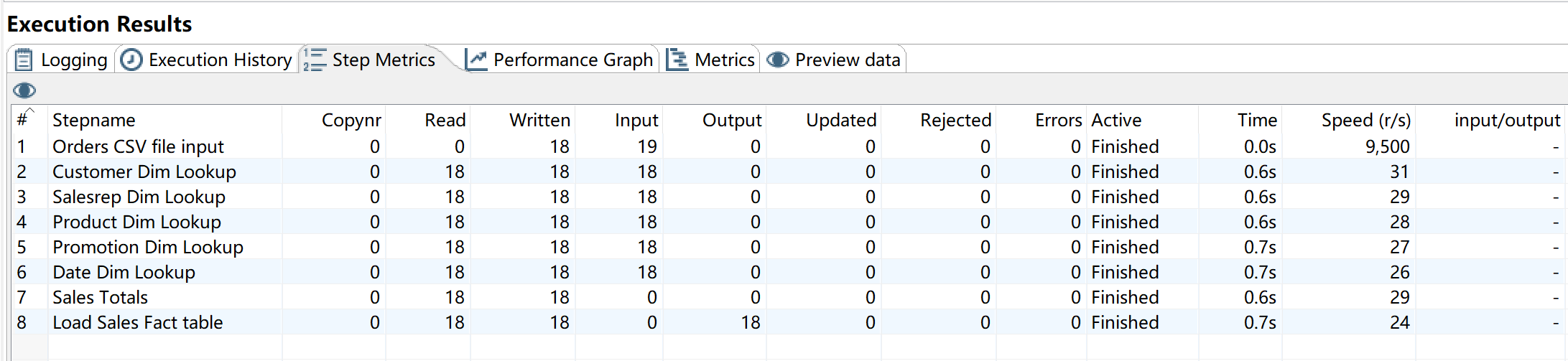
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SALES\_MONTH\_NAME | SALES\_DAY\_OF\_WEEK\_NAME | HOLIDAY\_DESCRIPTION | IS\_A\_HOLIDAY | DATE\_DIM\_ID | VERSION | DATE\_FROM | DATE\_TO | SALES\_DATE | SALES\_DAY\_OF\_YEAR | SALES\_MONTH | SALES\_YEAR | SALES\_QUARTER | SALES\_DAY\_OF\_WEEK | SALES\_DAY\_OF\_MONTH |
|  |  |  |  | 0 | 1 |  |  |  |  |  |  |  |  |  |
| January | Monday | New Year's Day | 1 | 1 | 1 | 01-JAN-00 | 31-DEC-99 | 02-JAN-06 | 2 | 1 | 2006 | 1 | 2 | 2 |
| January | Tuesday |  | 0 | 2 | 1 | 01-JAN-00 | 31-DEC-99 | 03-JAN-06 | 3 | 1 | 2006 | 1 | 3 | 3 |
| January | Wednesday |  | 0 | 3 | 1 | 01-JAN-00 | 31-DEC-99 | 04-JAN-06 | 4 | 1 | 2006 | 1 | 4 | 4 |
| January | Thursday |  | 0 | 4 | 1 | 01-JAN-00 | 31-DEC-99 | 05-JAN-06 | 5 | 1 | 2006 | 1 | 5 | 5 |
| January | Friday |  | 0 | 5 | 1 | 01-JAN-00 | 31-DEC-99 | 06-JAN-06 | 6 | 1 | 2006 | 1 | 6 | 6 |
| January | Saturday |  | 0 | 6 | 1 | 01-JAN-00 | 31-DEC-99 | 07-JAN-06 | 7 | 1 | 2006 | 1 | 7 | 7 |
| January | Sunday |  | 0 | 7 | 1 | 01-JAN-00 | 31-DEC-99 | 08-JAN-06 | 8 | 1 | 2006 | 1 | 1 | 8 |
| January | Monday |  | 0 | 9 | 1 | 01-JAN-00 | 31-DEC-99 | 10-JAN-06 | 10 | 1 | 2006 | 1 | 3 | 10 |
| January | Tuesday |  | 0 | 10 | 1 | 01-JAN-00 | 31-DEC-99 | 11-JAN-06 | 11 | 1 | 2006 | 1 | 4 | 11 |
| January | Wednesday |  | 0 | 11 | 1 | 01-JAN-00 | 31-DEC-99 | 12-JAN-06 | 12 | 1 | 2006 | 1 | 5 | 12 |
| January | Thursday |  | 0 | 12 | 1 | 01-JAN-00 | 31-DEC-99 | 13-JAN-06 | 13 | 1 | 2006 | 1 | 6 | 13 |
| January | Friday |  | 0 | 13 | 1 | 01-JAN-00 | 31-DEC-99 | 14-JAN-06 | 14 | 1 | 2006 | 1 | 7 | 14 |
| January | Saturday |  | 0 | 14 | 1 | 01-JAN-00 | 31-DEC-99 | 15-JAN-06 | 15 | 1 | 2006 | 1 | 1 | 15 |
| January | Sunday | Martin Luther King, Jr. Day | 1 | 15 | 1 | 01-JAN-00 | 31-DEC-99 | 16-JAN-06 | 16 | 1 | 2006 | 1 | 2 | 16 |
| January | Monday |  | 0 | 16 | 1 | 01-JAN-00 | 31-DEC-99 | 17-JAN-06 | 17 | 1 | 2006 | 1 | 3 | 17 |
| January | Tuesday |  | 0 | 17 | 1 | 01-JAN-00 | 31-DEC-99 | 18-JAN-06 | 18 | 1 | 2006 | 1 | 4 | 18 |
| January | Wednesday |  | 0 | 18 | 1 | 01-JAN-00 | 31-DEC-99 | 19-JAN-06 | 19 | 1 | 2006 | 1 | 5 | 19 |
| January | Thursday |  | 0 | 19 | 1 | 01-JAN-00 | 31-DEC-99 | 20-JAN-06 | 20 | 1 | 2006 | 1 | 6 | 20 |
| January | Friday |  | 0 | 20 | 1 | 01-JAN-00 | 31-DEC-99 | 21-JAN-06 | 21 | 1 | 2006 | 1 | 7 | 21 |

**Exercise 5 – Importing new Sales Data**

**Answers:**

Input the new order data into the system. After the dimensional tables match the keys with the columns in the csv file. The new sales totals has been calculated and the results has been loaded into the sales fact table.





|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DOLLARS\_SOLD | AMOUNT\_SOLD | PROMOTION\_DIM\_ID | ORDER\_ID | CUSTOMER\_DIM\_ID | SALESREP\_DIM\_ID | PRODUCT\_DIM\_ID | DATE\_DIM\_ID |
| 6765 | 165 | 0 | 2421 | 166 | 0 | 185 | 435 |
| 4464 | 9 | 0 | 2422 | 1 | 7 | 10 | 714 |
| 828 | 18 | 0 | 2422 | 1 | 7 | 179 | 714 |
| 205 | 5 | 0 | 2422 | 1 | 7 | 185 | 714 |
| 357.5 | 5 | 0 | 2422 | 1 | 7 | 240 | 714 |
| 2106 | 39 | 0 | 2422 | 1 | 7 | 164 | 714 |
| 425 | 25 | 0 | 2422 | 1 | 7 | 154 | 714 |
| 1247 | 29 | 0 | 2422 | 1 | 7 | 76 | 714 |
| 1050 | 35 | 0 | 2422 | 1 | 7 | 162 | 714 |
| 506 | 11 | 0 | 2422 | 1 | 7 | 175 | 714 |
| 312 | 8 | 1 | 2423 | 2 | 14 | 166 | 689 |
| 96 | 3 | 0 | 2423 | 2 | 14 | 157 | 689 |
| 2788.5 | 13 | 0 | 2423 | 2 | 14 | 107 | 689 |
| 2145 | 33 | 0 | 2423 | 2 | 14 | 81 | 689 |
| 416 | 16 | 0 | 2423 | 2 | 14 | 129 | 689 |
| 1638 | 21 | 0 | 2423 | 2 | 14 | 133 | 689 |
| 2972.2000000000003 | 14 | 0 | 2423 | 2 | 14 | 108 | 689 |
| 7623 | 11 | 0 | 2424 | 320 | 7 | 78 | 689 |
| 1332 | 12 | 0 | 2424 | 320 | 7 | 64 | 689 |
| 4869 | 9 | 0 | 2424 | 320 | 7 | 28 | 689 |

**How long it took you to complete the assignment (in units of hours working). What was the most difficult part of the assignment?**

I think I also spent like 10 hours for the assignment like HW #1. Although this assignment seems easier than the first one, it took me plenty of time to debug and fix the errors. I think the most difficult part was that sometimes I thought I was right but there were always errors and it took me hours to find out why. Sometimes I deleted or restarted Pentaho it actually helped, but most of time was hard to figure out where were the problems and how to solve it. It was useful to learn how to ETL source data (OLTP) into dimensional data (OLAT) though.

CREATE VIEW v2

AS

SELECT sf.order\_id, sf.dollars\_sold, sf.amount\_sold,

cd.customer\_id, cd.cust\_first\_name, cd.cust\_last\_name, cd.street\_address, cd.postal\_code, cd.city, cd.state\_province, cd.country\_id, cd.country\_name, cd.region\_id, cd.nls\_language, cd.nls\_territory, cd.credit\_limit, cd.cust\_email, cd.primary\_phone\_number, cd.phone\_number\_2, cd.account\_mgr\_id, cd.location\_gtype, cd.location\_srid, cd.location\_x, cd.location\_y,

dd.sales\_date, dd.sales\_day\_of\_year, dd.sales\_month, dd.sales\_year, dd.sales\_quarter, dd.sales\_month\_name, dd.sales\_day\_of\_week\_name, dd.sales\_day\_of\_week, dd.sales\_day\_of\_month, dd.holiday\_description, dd.is\_a\_holiday,

pd.product\_id, pd.product\_name, pd.language\_id, pd.min\_price, pd.list\_price, pd.product\_status, pd.supplier\_id, pd.warranty\_period, pd.weight\_class, pd.product\_description, pd.category\_id, pd.catalog\_url, pd.sub\_category\_name, pd.sub\_category\_description, pd.parent\_category\_id, pd.category\_name,

prd.promo\_id, prd.promo\_name,

sd.salesrep\_id, sd.first\_name, sd.last\_name, sd.email, sd.phone\_number, sd.hire\_date, sd.job\_id, sd.salary, sd.commission\_pct, sd.manager\_id, sd.department\_id

FROM admin.sales\_fact sf

INNER JOIN admin.customer\_dim cd

USING (customer\_dim\_id)

INNER JOIN admin.date\_dim dd

USING (date\_dim\_id)

INNER JOIN admin.product\_dim pd

USING (product\_dim\_id)

INNER JOIN admin.promotion\_dim prd

USING (promotion\_dim\_id)

INNER JOIN admin.salesrep\_dim sd

USING (salesrep\_dim\_id);

