

Vincent Sirait

191402081

Instalasi dan Contoh Penggunaan Pentaho

1. Registrasi Pentaho

HITACHI
Inspire the Next

Hitachi Vantara Login

User Registration

Please complete all required fields

First Name*

Last Name*

Company Name*

Job Title*

Company Country or Region*

Company Zip/Postal Code*

Preferred Phone Number* Ext.

Time Zone*

Business Email*

Confirm Business Email*

Your email address will be your username when you login.
Please register with a company email address, not a
personal email address, to ensure proper access within the
portal. Community users, please use your community email
address to ensure proper access.

This site uses cookies from Hitachi and third parties for our own business purposes and to personalize your experience. By using this site, you agree to the use of cookies. For more information, visit [Hitachi Cookies Policy](#).

2. Login

HITACHI
Inspire the Next

Hitachi Vantara Login

User Registration

Please complete all required fields

First Name*

Last Name*

Company Name*

Job Title*

Company Country or Region*

Company Zip/Postal Code*

Preferred Phone Number* Ext.

Time Zone*

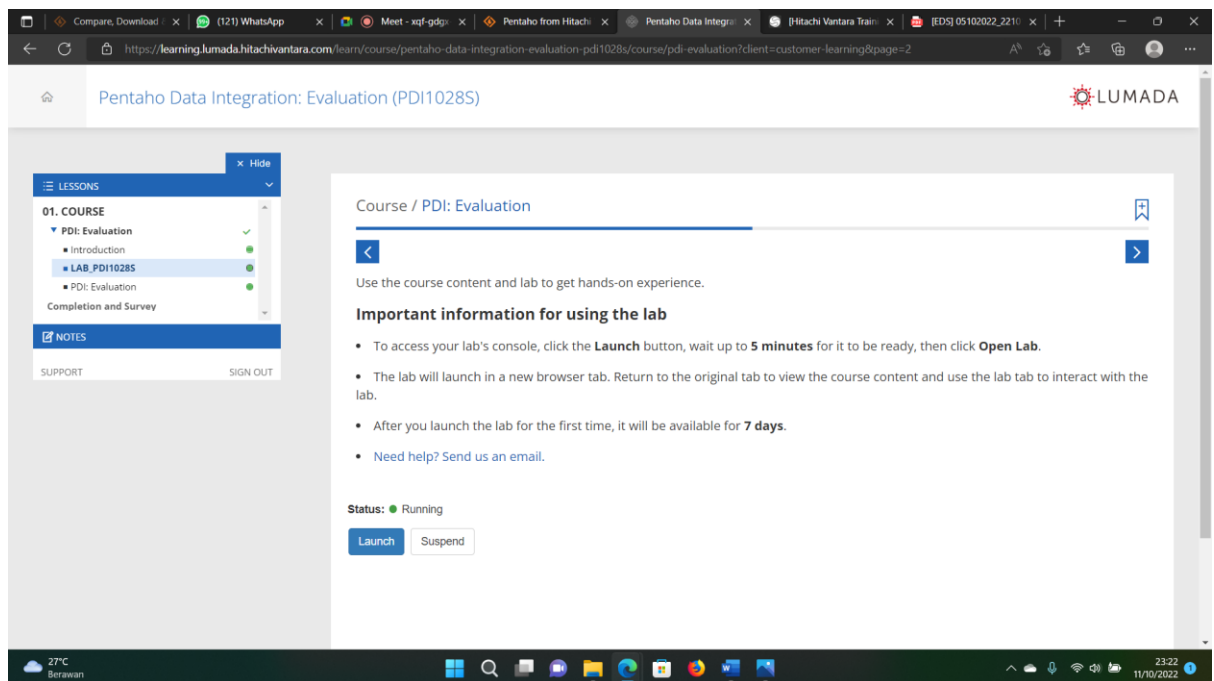
Business Email*

Confirm Business Email*

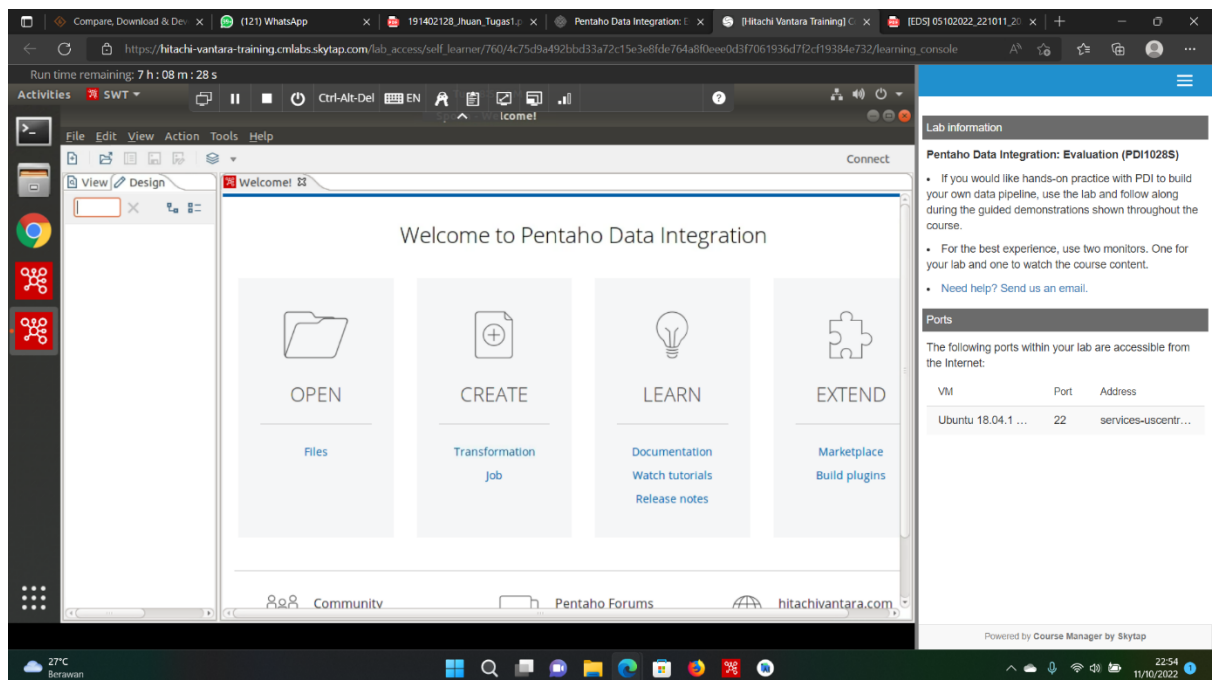
Your email address will be your username when you login.
Please register with a company email address, not a
personal email address, to ensure proper access within the
portal. Community users, please use your community email
address to ensure proper access.

This site uses cookies from Hitachi and third parties for our own business purposes and to personalize your experience. By using this site, you agree to the use of cookies. For more information, visit [Hitachi Cookies Policy](#).

3. Masuk ke dalam learning path dan ikuti setiap arahan yang ad

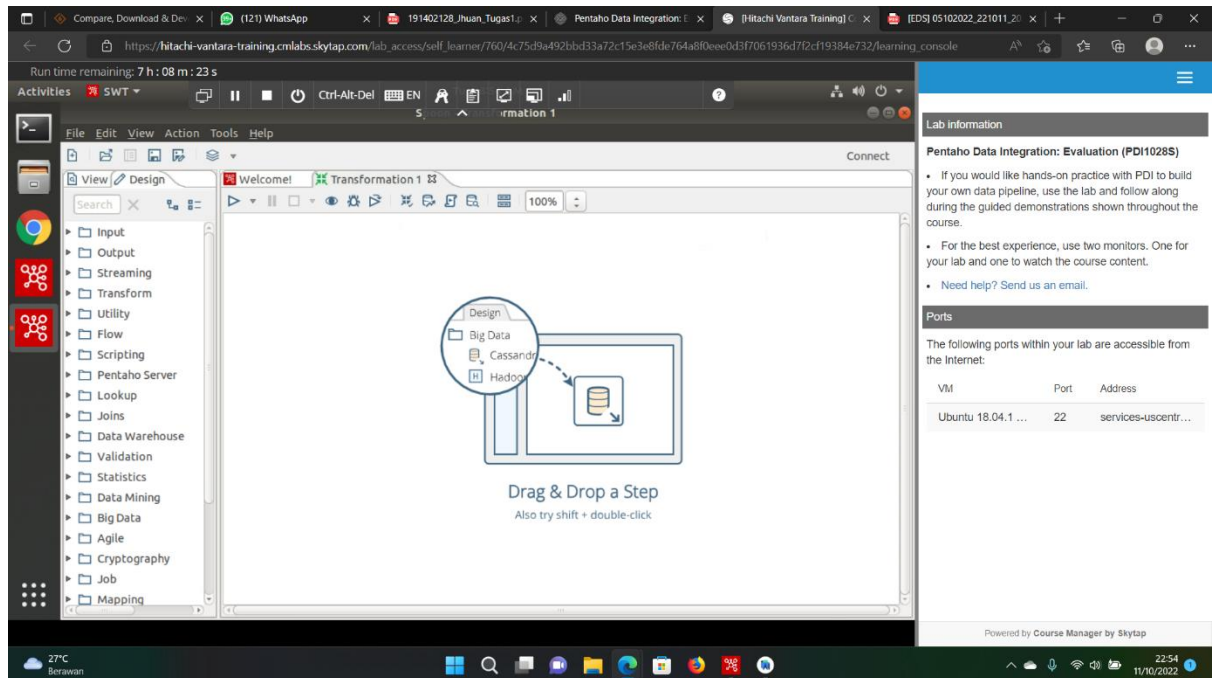


4. Laman tampilan pentaho

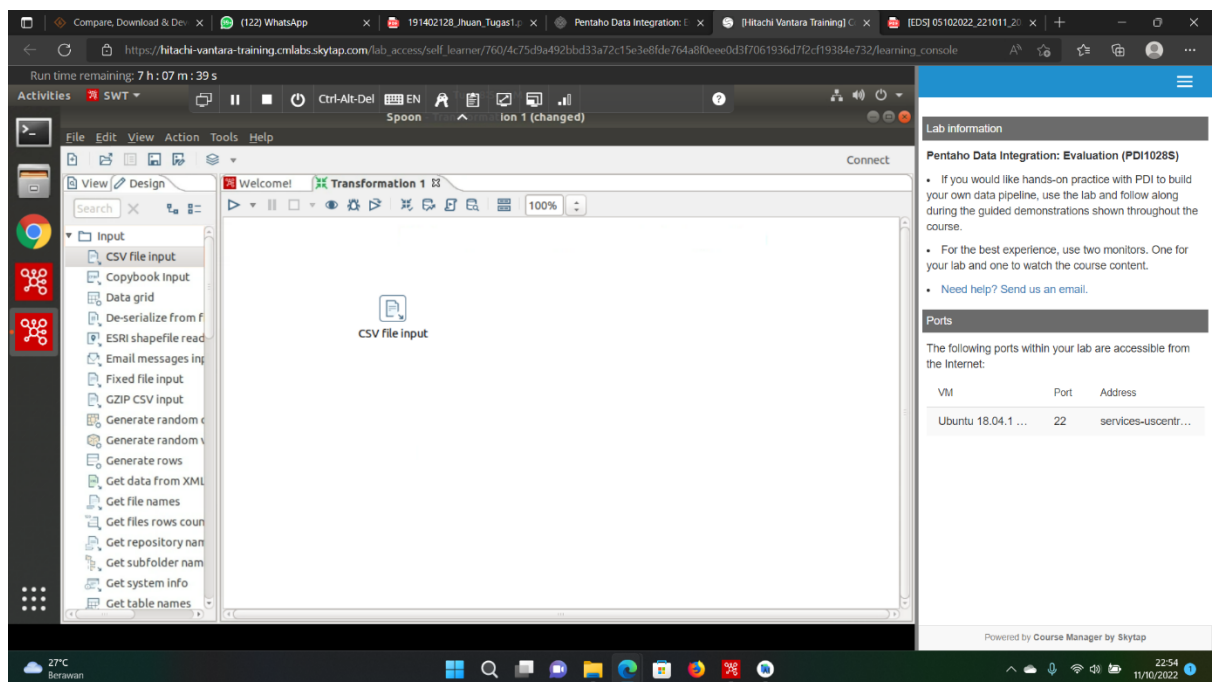


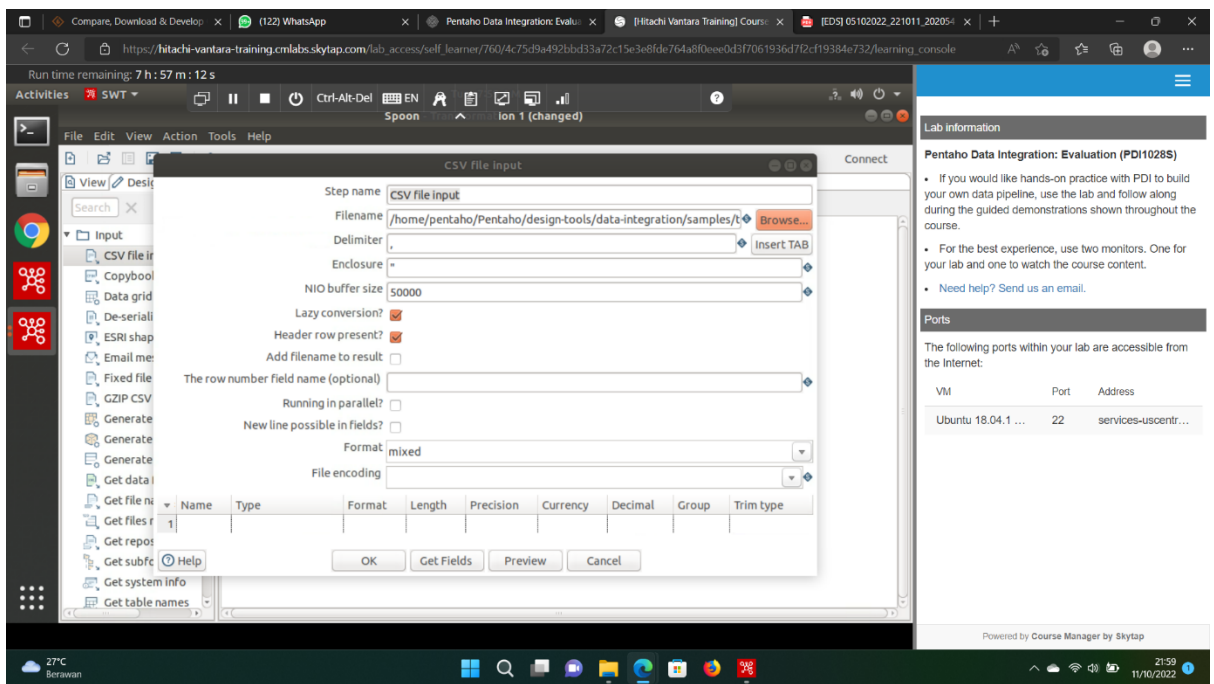
Pentaho Data Integration Evaluation. Pentaho Data Integration (PDI) adalah solusi dalam pengolahan serta perapian data sehingga data tersebut jauh lebih enak dipandang dan lebih mudah dipahami

1. Buka pentaho
2. Buat transformasi baru

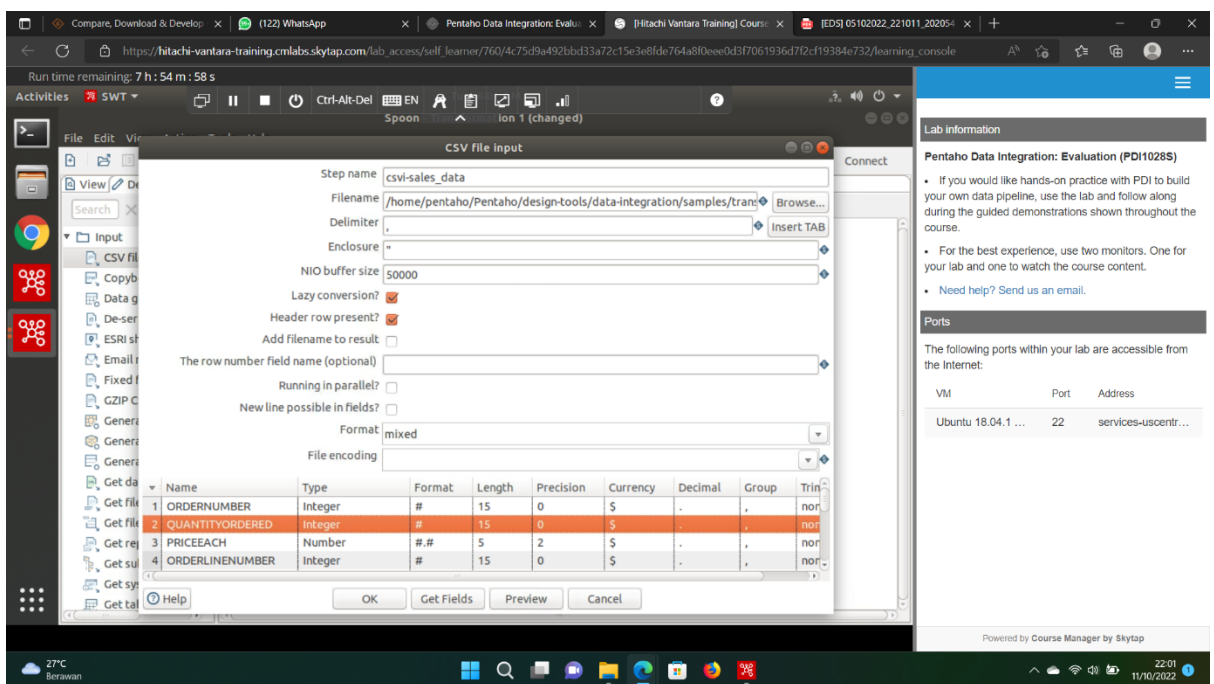


3. Input file CSV yang akan diolah (Drag and Drop ataupun double click, setelahnya double click)





4. Buka file yang akan diolah



5. Ambil sample data, dan buatget field sebesar 0 data dari sample yang ada

The screenshot shows the 'CSV input' step configuration in Pentaho Data Integration. The step name is 'csv-sales_data'. The filename is '/home/pentaho/Pentaho/design-tools/data-integration/samples/transactions.csv'. The delimiter is set to 'Insert TAB'. The NIO buffer size is 50000. The 'Header row present?' checkbox is checked. The 'Format' is set to 'mixed'. The 'File encoding' is set to 'UTF-8'. The 'Get Fields' button is highlighted.

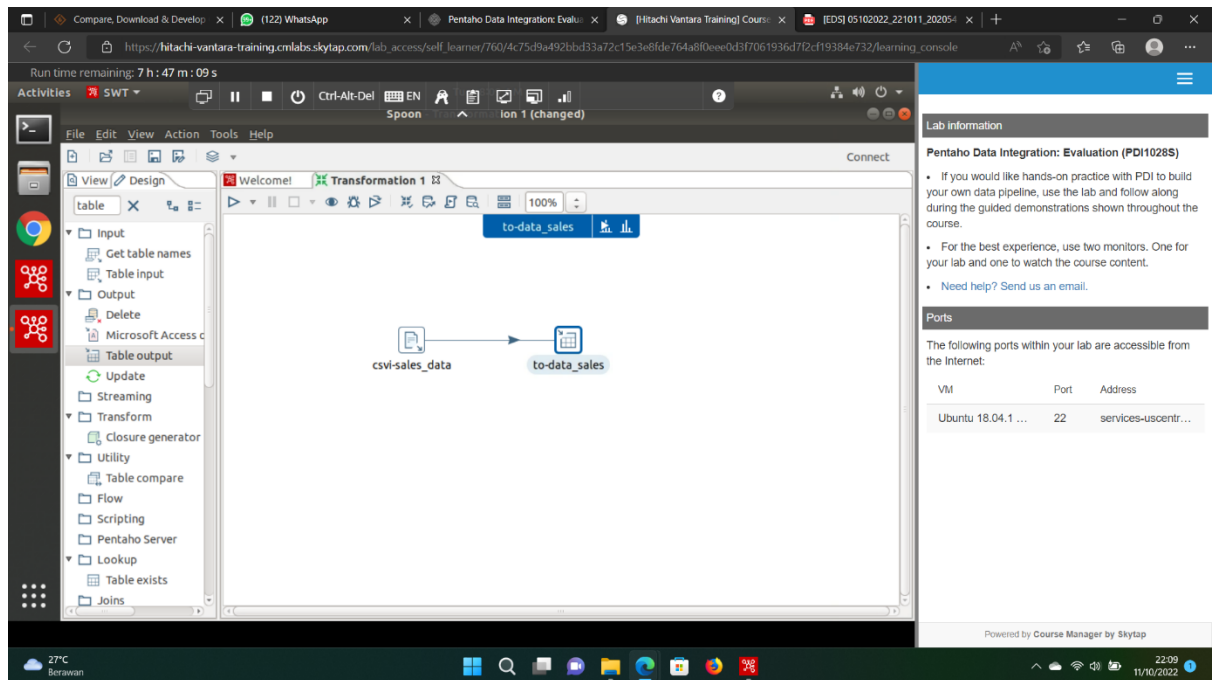
Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trin
1. ORDERNUMBER	Integer	#	15	0	\$	-	,	non
2. QUANTITYORDERED	Integer	#	15	0	\$	-	,	non
3. PRICEEACH	Number	##	5	2	\$	-	,	non
4. ORDERLINENUMBER	Integer	#	15	0	\$	-	,	non

6. Lalu kita coba lihat review dari 1000 sample yang ada pada data

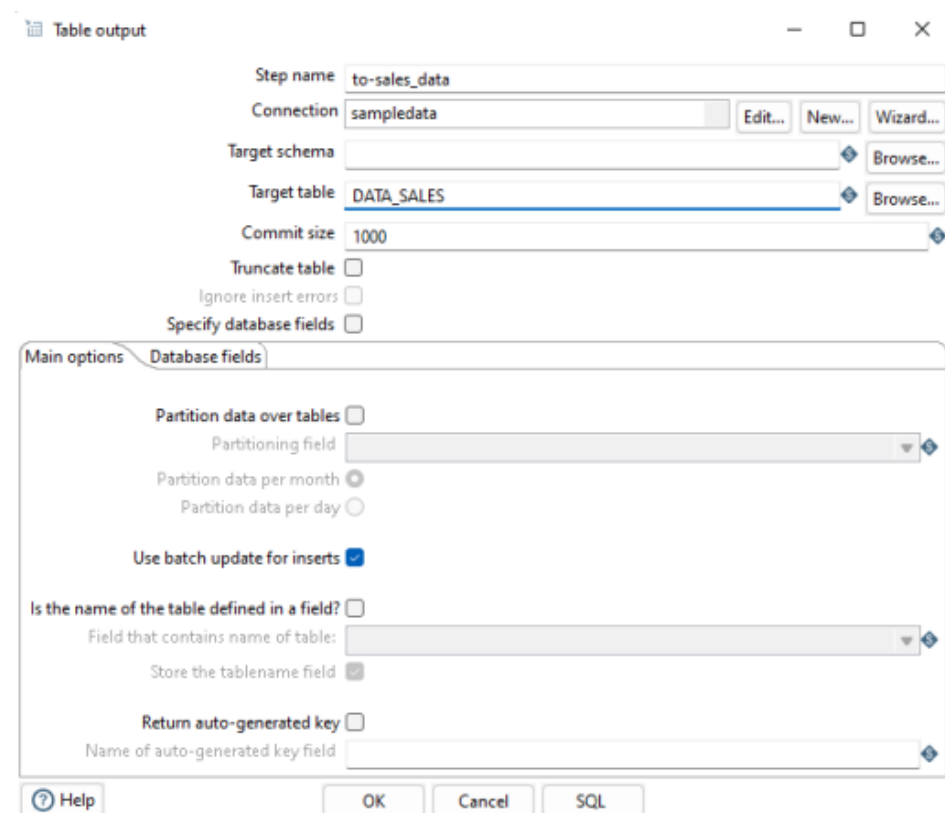
The screenshot shows the 'Rows of step: csv-sales_data (1000 rows)' view in Pentaho Data Integration. The data is displayed in a table with columns: ORDERNUMBER, QUANTITYORDERED, PRICEEACH, ORDERLINENUMBER, SALES, ORDERDATE, STATUS, QTR_ID, and MONTH. The first 22 rows are visible.

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	STATUS	QTR_ID	MONTH
1	10107	30	95.7	2	2871	2/24/2003 0:00	Shipped	1	
2	10121	34	81.3	5	2765.9	5/7/2003 0:00	Shipped	2	
3	10134	41	94.7	2	3884.3	7/1/2003 0:00	Shipped	3	
4	10145	45	83.3	6	3746.7	8/25/2003 0:00	Shipped	3	
5	10159	49	100	14	5205.3	10/10/2003 0:00	Shipped	4	
6	10168	36	96.7	1	3479.8	10/28/2003 0:00	Shipped	4	
7	10180	29	86.1	9	2497.8	11/11/2003 0:00	Shipped	4	
8	10188	48	100	1	5512.3	11/18/2003 0:00	Shipped	4	
9	10201	22	98.6	2	2168.5	12/1/2003 0:00	Shipped	4	
10	10211	41	100	14	4708.4	1/15/2004 0:00	Shipped	1	
11	10223	37	100	1	3965.7	2/20/2004 0:00	Shipped	1	
12	10237	23	100	7	2333.1	4/5/2004 0:00	Shipped	2	
13	10251	28	100	2	3188.6	5/18/2004 0:00	Shipped	2	
14	10263	34	100	2	3676.8	6/28/2004 0:00	Shipped	2	
15	10275	45	92.8	1	4177.4	7/23/2004 0:00	Shipped	3	
16	10285	36	100	6	4099.7	8/27/2004 0:00	Shipped	3	
17	10299	23	100	9	2597.4	9/30/2004 0:00	Shipped	3	
18	10309	41	100	5	4394.4	10/15/2004 0:00	Shipped	4	
19	10318	46	94.7	1	4358	11/2/2004 0:00	Shipped	4	
20	10329	42	100	1	4396.1	11/15/2004 0:00	Shipped	4	
21	10341	41	100	9	7737.9	11/24/2004 0:00	Shipped	4	
22	10361	20	72.5	13	1451	12/17/2004 0:00	Shipped	4	

7. Cari table output, lalu double-click table output. Table output akan langsung terintegrasi dengan data



8. Buka Table Output



9. Tambahkan Target Table

Table output

Step name: to-sales_data

Connection: sampledata

Target schema:

Target table: DATA_SALES

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☐

Main options Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field?: ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Help OK Cancel SQL

10. Lalu klik SQ dan execute tablenya

Simple SQL editor

SQL statements, separated by semicolon ';' :

```
CREATE TABLE DATA_SALES
(
  ORDERNUMBER BIGINT
  QUANTITYORDERED BIGINT
  PRICEEACH DOUBLE
  ORDERLINENUMBER BIGINT
  SALES DOUBLE
  ORDERDATE VARCHAR(15)
  STATUS VARCHAR(10)
  QTR_ID BIGINT
  MONTH_ID BIGINT
  YEAR_ID BIGINT
  PRODUCTLINE VARCHAR(12)
  MSRP BIGINT
  CUSTOMERNAME VARCHAR(30)
  PHONE VARCHAR(16)
  ADDRESSLINE1 VARCHAR(40)
  ADDRESSLINE2 VARCHAR(9)
  CITY VARCHAR(14)
  STATE VARCHAR(10)
  POSTALCODE VARCHAR(8)
  COUNTRY VARCHAR(13)
  TERRITORY VARCHAR(5)
  CONTACTLASTNAME VARCHAR(10)
  CONTACTFIRSTNAME VARCHAR(9)
)
```

Line 1 column 0

Execute Clear cache Close

Table output

Step name: to-sales_data

Connection: sampledata

Target schema:

Target table: DATA_SALES

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☐

Main options Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field?: ☐

Field that contains name of table:

Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

Help OK Cancel SQL

11. Lalu jalankan PDI nya maka akan muncul Hasil PDI yang berjalan dengan baik seperti berikut

The screenshot shows the Pentaho Data Integration (PDI) interface. The main workspace displays a data flow diagram with two steps: 'csv-sales_data' (input) and 'to-data_sales' (output). The 'Execution Results' pane at the bottom shows the following log:

```
2022/10/11 08:10:52 - to-data_sales.0 - Connected to database [sampledata] (commit=1000)
2022/10/11 08:10:52 - csv-sales_data.0 - Header row skipped in file /home/pentaho/Pentaho/design-tools/data-integration/sampledata.csv
2022/10/11 08:10:52 - csv-sales_data.0 - Finished processing (I=2824, O=0, R=0, W=2823, U=0, E=0)
2022/10/11 08:10:52 - to-data_sales.0 - Finished processing (I=0, O=2823, R=2823, W=2823, U=0, E=0)
2022/10/11 08:10:52 - Spoon - The transformation has finished!!
```

The right sidebar contains 'Lab information' and 'Ports' sections. The 'Ports' section lists the following ports accessible from the Internet:

VM	Port	Address
Ubuntu 18.04.1 ...	22	services-uscentr...

The screenshot shows the Pentaho Data Integration (PDI) interface. The main workspace displays a data flow diagram with two steps: 'csv-sales_data' (input) and 'to-data_sales' (output). The 'Execution Results' pane at the bottom shows the following table:

Stepname	Copypnr	Read	Written	Input	Output	Updated	Rejected	Errors	Active
1 csv-sales_data	0	0	2823	2824	0	0	0	0	Finished
2 to-data_sales	0	2823	2823	0	2823	0	0	0	Finished

The right sidebar contains 'Lab information' and 'Ports' sections. The 'Ports' section lists the following ports accessible from the Internet:

VM	Port	Address
Ubuntu 18.04.1 ...	22	services-uscentr...