

AOLANI DATA

www.aolanidata.com

DATA INTEGRATION

ETL DENGAN PENTAHO & EXCEL

STUDI KASUS
ABSENSI

Ni Wayan Sumartini Saraswati
Data Analyst
Practitioner

2023

DAFTAR ISI

DAFTAR ISI.....	2
DAFTAR GAMBAR.....	3
BAB I KASUS DATA ABSENSI.....	5
2.1 Penjelasan.....	5
2.2 Transformasi preprosesing data absensi.....	5
DAFTAR PUSTAKA	13

DAFTAR GAMBAR

Gambar 2.1 Detail Transformasi Preprosesing Data Absensi.....	5
Gambar 2.2 Microsoft Excel Input Pengaturan Tab Files	6
Gambar 2.3 Microsoft Excel Input Pengaturan Tab Sheet	6
Gambar 2.4 Microsoft Excel Input Pengaturan Tab Fields	6
Gambar 2.5 Hasil Preview Microsoft Excel Input sumber_data.xls.....	7
Gambar 2.6 Sort Rows Transformasi Awal	7
Gambar 2.7 Group By Jam Datang	8
Gambar 2.8 Group By Jam Pulang	9
Gambar 2.9 Preview Group By Jam Pulang	10
Gambar 2.10 Merge Join Jam Datang dan Jam Pulang	11
Gambar 2.11 Preview Merge Join.....	11
Gambar 2.12 Microsoft Excel Output jumlahjam.xls	12

KATA PENGANTAR

Buku ini dibuat untuk menjawab kegelisahan akan contoh penggunaan Pentaho dalam menyelesaikan permasalahan integrasi data pada dunia nyata. Dalam buku ini dijelaskan secara rinci dan jelas langkah demi langkah penyelesaian permasalahan absensi untuk mendapatkan bentuk data yang diharapkan. Penulis berharap dengan adanya buku ini mampu menambah pengetahuan pembaca yang ini memahami penggunaan Pentaho untuk pengolahan data sederhana.

Penulis

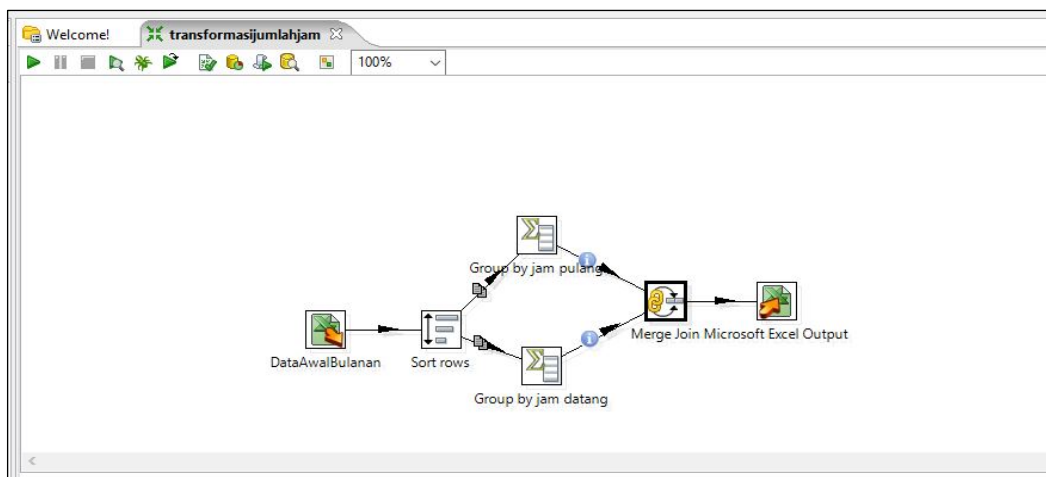
BAB I KASUS DATA ABSENSI

2.1 Penjelasan

Kasus data absensi dibuat berdasarkan permasalahan di mana data yang didapat dari output mesin sidik jari tidak mampu memenuhi kebutuhan akan laporan absensi. Output dari mesin sidik jari adalah berupa file Excel dan laporan absensi yang diinginkan juga dibuat dalam bentuk Excel. Dalam penyelesaian kasus ini pengenalan untuk penggunaan dasar pentaho seperti penggunaan transformation tidak dijabarkan lagi.

2.2 Transformasi preprosesing data absensi

Dalam preprocessing data absensi kita akan membuat sebuah file transformation baru yang bisa kita beri nama transformasi jumlah jam. Adapun objek transformasi yang terlibat dalam preprocessing ini adalah Microsoft Excel input, sort rows, dua group by, merge join dan Microsoft Excel output seperti ditunjukkan oleh gambar 2.1

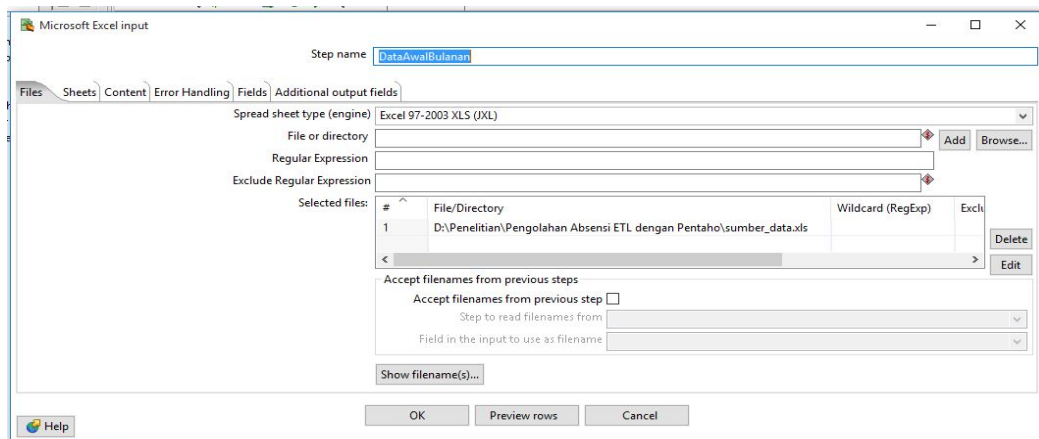


Gambar 2.1 Detail Transformasi Preprosesing Data Absensi

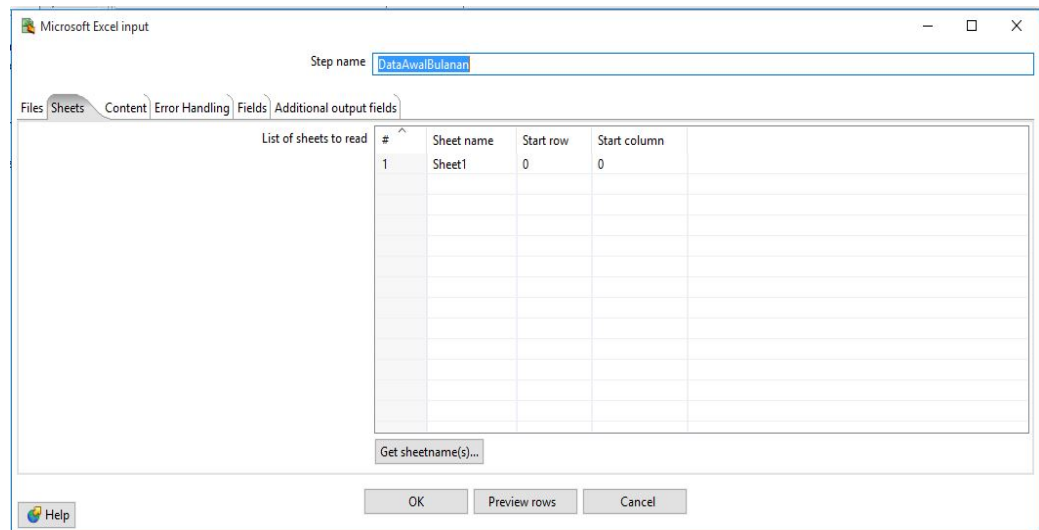
Microsoft Excel input digunakan untuk mengekstrak data awal keluaran mesin fingerprint. Langkah-langkah untuk mengkonfigurasi object Excel input adalah sebagai berikut

1. Klik ganda pada step
2. Seperti ditunjukkan oleh gambar 2.2 kemudian beri nama pada step (step name) sebagai data awal bulanan. Pada isian file directory klik browse untuk menemukan file yang dimaksud kemudian akhiri dengan menekan tombol add sehingga pada isian selected file akan terisi direktori file tersebut. Nama file yang di maksud dalam pengaturan ini adalah sumber_data.xls.
3. Pada tab sheets klik tombol “get sheets name”, pilih sheet 1 seperti ditunjukkan oleh gambar 2.3.

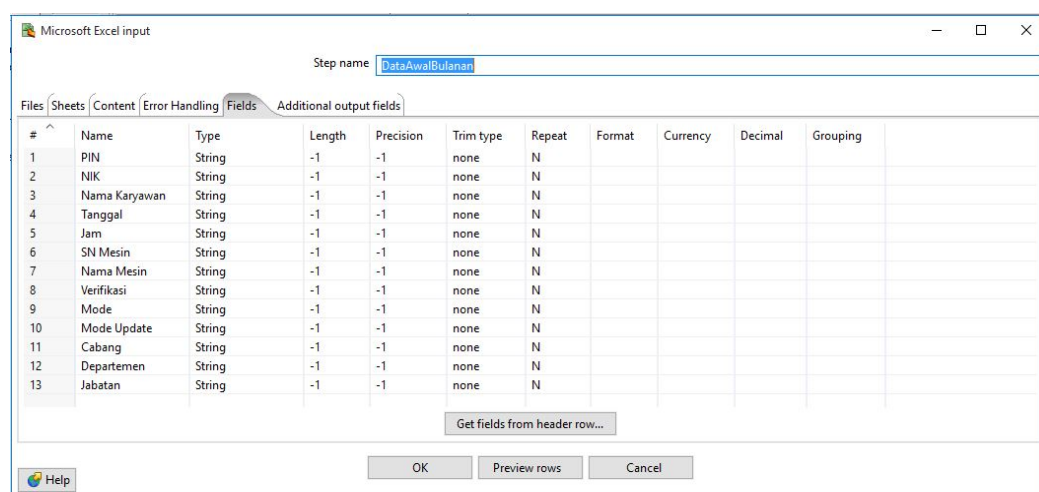
4. Pada tab field, klik “get field from header row” seperti ditunjukkan oleh gambar 2.4, kemudian akhiri dengan menekan ”preview rows”. Hasil akan ditunjukkan oleh gambar 2.5.



Gambar 2.2 Microsoft Excel Input Pengaturan Tab Files



Gambar 2.3 Microsoft Excel Input Pengaturan Tab Sheet



Gambar 2.4 Microsoft Excel Input Pengaturan Tab Fields

Examine preview data

Rows of step: DataAwalBulanan (1000 rows)

#	PIN	NIK	Nama Karyawan	Tanggal	Jam	SN Mesin	Nama Mesin	Verifikasi	Mode	Mode Update	Cabang	Departemen	Jabatan
1	6	1007151	I Nym Anom Fajaraditya	01/03/2018	00:11	0331142900355	MATRIK1	FP/PW/RF/FACE	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
2	99	1512251	Ketut Suartana	01/03/2018	05:37	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
3	42	1008157	Agus Harianto	01/03/2018	05:59	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
4	40	1310198	I Made Suartana	01/03/2018	06:01	6571150900355	PREMIER3	FP	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
5	100	1511250	Komang Catra Girawan	01/03/2018	06:01	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
6	37	1106170	I Ketut Merta	01/03/2018	06:01	6571150900355	PREMIER3	FP	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
7	120	1702274	Anak Agung Putu Susana	01/03/2018	06:03	6571150900355	PREMIER3	FP	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
8	52	1403206	I Dewa Nyoman Mudita	01/03/2018	06:07	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
9	54	1312203	Muhammad Rofi	01/03/2018	06:08	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
10	34	9007127	AA Ketut Putra	01/03/2018	06:09	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
11	41	9010134	Adrianus T. Balibo	01/03/2018	06:34	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
12	119	1701273	Dewa Ayu Giovany angga I	01/03/2018	07:11	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
13	126	1707280	Ni Luh Line Purwanti	01/03/2018	07:34	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
14	29	1210191	Ni Putu Ema Budiantari	01/03/2018	07:41	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
15	141	1710291	I Nyoman Tri Anindita Putra	01/03/2018	07:44	0331142900324	MATRIK2	FP/PW/RF/FACE	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
16	84	1507237	Kadek Ayu Aningsih	01/03/2018	07:57	0331142900324	MATRIK2	FP/PW/RF/FACE	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
17	99	1512251	Ketut Suartana	01/03/2018	07:57	6571150900355	PREMIER3	FP	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
18	42	1008157	Agus Harianto	01/03/2018	08:01	6571150900355	PREMIER3	FP	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
19	130	1707284	Yusta Suryono Afri Teo	01/03/2018	08:04	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
20	79	1506232	Ni Made Ery Indriawati	01/03/2018	08:04	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
21	86	1508239	I Kadek Agus Swartawan	01/03/2018	08:16	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
22	45	1207185	Ni Kadek Nita Noviani P	01/03/2018	08:34	0331142900324	MATRIK2	FP/PW/RF/FACE	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
23	103	1602253	Rury Ayuning Lati	01/03/2018	08:37	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
24	33	1007152	Dewa Gede Surya Damanik	01/03/2018	08:41	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
25	142	1712293	Gusti Ayu Shinta Dwi Astari	01/03/2018	08:50	0331142900324	MATRIK2	FP/PW/RF/FACE	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
26	145	1609267	I Wayan Agus Surya Darna	01/03/2018	09:01	0331142900324	MATRIK2	FP/PW/RF/FACE	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
27	31	1204182	Ni Kadek Ayu Nirvana	01/03/2018	09:02	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
28	144	1609264	Luh Putu Ariestari Pradnyadesvi	01/03/2018	09:02	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
29	74	1506233	Ida Ayu Gede Anindiatari	01/03/2018	09:09	6571150900355	PREMIER3	FP	Scan Masuk	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan
30	41	9010134	Adrianus T. Balibo	01/03/2018	09:09	6571150900355	PREMIER3	FP	Scan Keluar	Scan Masuk	Kantor Pusat	Fingerspot	Karyawan

Gambar 2.5 Hasil Preview Microsoft Excel Input sumber_data.xls

- Setelah kita berhasil membuat Microsoft Excel input langkah berikutnya kita menambahkan step sort rows. Hal ini berguna untuk mengurutkan jam kedatangan dan jam kepulangan dari input yang tersedia. Dari Microsoft Excel input kita tambahkan hop yang mengarah kepada step sort rows. Pada setting seperti ditunjukkan gambar 2.6, kita memilih yang pertama adalah nama karyawan ascending kemudian diikuti dengan tanggal dan yang terakhir adalah jam.

Sort rows

Step name:

Sort directory:

TMP-file prefix:

Sort size (rows in memory):

Free memory threshold (in %):

Compress TMP Files? ☐

Only pass unique rows? (verifies keys only) ☐

Fields:

#	Fieldname	Ascending	Case sensitive compare?	Presorted?
1	Nama Karyawan	Y	N	N
2	Tanggal	Y	N	N
3	Jam	Y	N	N

Gambar 2.6 Sort Rows Transformasi Awal

6. Untuk membuat jam kedatangan dan jam kepulangan dalam satu baris record karyawan per tanggal maka kita menggunakan dua buah step group by. Dari step sort rows dihubungkan dengan hop dengan pilihan copy all record untuk kedua grup by seperti ditunjukkan oleh gambar 2.1. Group by yang pertama diberi nama jam pulang sedangkan group by yang kedua diberi nama jam datang. Untuk mengkonfigurasi group by lakukan klik ganda pada objek, sehingga muncul tampilan seperti ditunjukkan oleh gambar 2.7. Pada isian “the field that make up the group” kita mengisikan nama karyawan dan tanggal yang berarti kita mengelompokkan data berdasarkan nama karyawan dan tanggal. Pada isian aggregates kita mengisikan nama sebagai “jamdatang”, subject “jam” dan type adalah “first value”. Karena kita telah mengurutkan data per jam absen maka first value akan mengembalikan jam yang lebih kecil.

Gambar 2.7 Group By Jam Datang

- [illegible]

Gambar 2.8 Group By Jam Pulang

Examine preview data			
Rows of step: Group by (1000 rows)			
#	Nama Karyawan	Tanggal	JamPulang
1	AA Gede Bgs Ariana	24/03/2018	15:59
2	AA Gede Bgs Ariana	31/03/2018	12:22
3	AA Ketut Putra	01/03/2018	14:04
4	AA Ketut Putra	02/03/2018	22:05
5	AA Ketut Putra	03/03/2018	22:02
6	AA Ketut Putra	06/03/2018	06:02
7	AA Ketut Putra	07/03/2018	14:02
8	AA Ketut Putra	09/03/2018	22:02
9	AA Ketut Putra	10/03/2018	22:02
10	AA Ketut Putra	11/03/2018	21:43
11	AA Ketut Putra	12/03/2018	21:47
12	AA Ketut Putra	13/03/2018	06:02
13	AA Ketut Putra	14/03/2018	22:07
14	AA Ketut Putra	15/03/2018	14:02
15	AA Ketut Putra	20/03/2018	06:01
16	AA Ketut Putra	21/03/2018	22:00
17	AA Ketut Putra	22/03/2018	14:02
18	AA Ketut Putra	23/03/2018	22:06
19	AA Ketut Putra	24/03/2018	14:13
20	AA Ketut Putra	25/03/2018	21:35
21	AA Ketut Putra	26/03/2018	06:07
22	AA Ketut Putra	28/03/2018	14:02
23	AA Ketut Putra	29/03/2018	14:06
24	AA Ketut Putra	30/03/2018	22:03
25	AA Ketut Putra	31/03/2018	22:00
26	Adigoci Mnaka	05/03/2018	15:22
27	Adigoci Mnaka	06/03/2018	15:46
28	Adigoci Mnaka	07/03/2018	15:25
29	Adigoci Mnaka	08/03/2018	15:15
30	Adigoci Mnaka	09/03/2018	15:38

Gambar 2.9 Preview Group By Jam Pulang

8. Langkah berikutnya adalah menggabungkan hasil kedua group by tersebut menggunakan step Merge join. Untuk first step kita isikan group by jam datang second step kita isikan group by jam pulang. Tipe joint adalah full outer. Keys kita isikan nama karyawan dan tanggal seperti ditunjukkan oleh gambar 2.10. Hasil dari proses ini dapat dilihat dalam gambar 2.11. Jam datang dan jam pulang yang semula berbeda record akan menjadi satu record yang berdampingan.

Gambar 2.10 Merge Join Jam Datang dan Jam Pulang

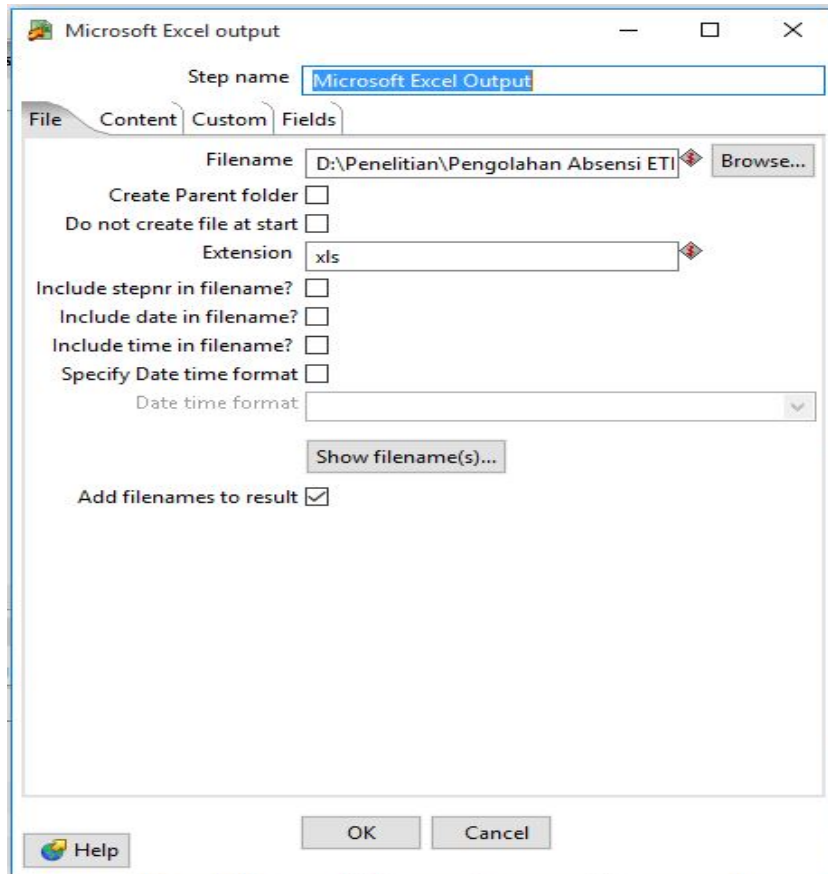
Examine preview data

Rows of step: Merge Join (1000 rows)

#	Nama Karyawan	Tanggal	JamDatang	Nama Karyawan_1	Tanggal_1	JamPulang
1	AA Gede Bgs Ariana	24/03/2018	12:02	AA Gede Bgs Ariana	24/03/2018	15:59
2	AA Gede Bgs Ariana	31/03/2018	12:22	AA Gede Bgs Ariana	31/03/2018	12:22
3	AA Ketut Putra	01/03/2018	06:09	AA Ketut Putra	01/03/2018	14:04
4	AA Ketut Putra	02/03/2018	14:05	AA Ketut Putra	02/03/2018	22:05
5	AA Ketut Putra	03/03/2018	14:02	AA Ketut Putra	03/03/2018	22:02
6	AA Ketut Putra	06/03/2018	06:02	AA Ketut Putra	06/03/2018	06:02
7	AA Ketut Putra	07/03/2018	06:01	AA Ketut Putra	07/03/2018	14:02
8	AA Ketut Putra	09/03/2018	13:44	AA Ketut Putra	09/03/2018	22:02
9	AA Ketut Putra	10/03/2018	13:46	AA Ketut Putra	10/03/2018	22:02
10	AA Ketut Putra	11/03/2018	21:43	AA Ketut Putra	11/03/2018	21:43
11	AA Ketut Putra	12/03/2018	06:07	AA Ketut Putra	12/03/2018	21:47
12	AA Ketut Putra	13/03/2018	06:02	AA Ketut Putra	13/03/2018	06:02
13	AA Ketut Putra	14/03/2018	13:57	AA Ketut Putra	14/03/2018	22:07
14	AA Ketut Putra	15/03/2018	05:50	AA Ketut Putra	15/03/2018	14:02
15	AA Ketut Putra	20/03/2018	06:01	AA Ketut Putra	20/03/2018	06:01
16	AA Ketut Putra	21/03/2018	13:50	AA Ketut Putra	21/03/2018	22:00
17	AA Ketut Putra	22/03/2018	05:56	AA Ketut Putra	22/03/2018	14:02
18	AA Ketut Putra	23/03/2018	14:03	AA Ketut Putra	23/03/2018	22:06
19	AA Ketut Putra	24/03/2018	05:58	AA Ketut Putra	24/03/2018	14:13
20	AA Ketut Putra	25/03/2018	21:35	AA Ketut Putra	25/03/2018	21:35
21	AA Ketut Putra	26/03/2018	06:07	AA Ketut Putra	26/03/2018	06:07
22	AA Ketut Putra	28/03/2018	06:12	AA Ketut Putra	28/03/2018	14:02
23	AA Ketut Putra	29/03/2018	05:55	AA Ketut Putra	29/03/2018	14:06
24	AA Ketut Putra	30/03/2018	22:03	AA Ketut Putra	30/03/2018	22:03
25	AA Ketut Putra	31/03/2018	06:01	AA Ketut Putra	31/03/2018	22:00
26	Adigoci Mnaka	05/03/2018	06:14	Adigoci Mnaka	05/03/2018	15:22
27	Adigoci Mnaka	06/03/2018	06:16	Adigoci Mnaka	06/03/2018	15:46
28	Adigoci Mnaka	07/03/2018	06:11	Adigoci Mnaka	07/03/2018	15:25
29	Adigoci Mnaka	08/03/2018	06:27	Adigoci Mnaka	08/03/2018	15:15

Gambar 2.11 Preview Merge Join

9. Untuk membuat hasil transformasi tersimpan ke dalam file Excel kita membutuhkan Microsoft Excel output. Setting Microsoft Excel output ditunjukkan oleh gambar 22 berikut ini. Kita tinggal menyambungkan hop dari merge join ke dalam Microsoft Excel output kemudian setting nama file output dengan klik tombol browse dan beri nama jumlahjam.xls.



Gambar 2.12 Microsoft Excel Output jumlahjam.xls

DAFTAR PUSTAKA

- Anonym, <http://blog.fingerspot.com/8-cara-kerja-sistem-absensi-sidik-jari/> [Diunduh pada tanggal 19/04/2019].
- Anonym, <http://e-journal.uajy.ac.id/1692/2/1TF05139.pdf> [Diunduh pada tanggal 19/04/2019].
- Anonym, https://id.wikipedia.org/wiki/Microsoft_Excel [Diunduh pada tanggal 19/04/2019].
- Mulyana JRP, 2014. *Pentaho: Solusi Open Source untuk Membangun Data Warehouse*. Bandung : Penerbit Andi.
- Nurtriana Hidayati. *Pentaho Sebagai Solusi Masalah Pengolahan Database*. Fakultas Teknologi Informasi dan Komunikasi Universitas Semarang.
- Preeti Dhanda & Neetu Sharma. *Extract Transform Load Data with ETL Tools*. International Journal of Advanced Research in Computer Science Volume 7, No. 3, May-June 2016
- Vishal Gour dkk. *Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse*. International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 786-789.