

## IMPLEMENTASI EXTRACT, TRANSFORM, LOAD PROCESS PADA PERANCANGAN DATA WAREHOUSE TERKAIT KUALITAS PENDIDIKAN DI KABUPATEN SERANG

Raihan Abdul Ghani <sup>1</sup>, Rudi Kurniawan <sup>2</sup>

<sup>1</sup> Manajemen Informatika, STMIK IKMI Cirebon

<sup>2</sup> Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

Jalan Perjuangan No. 10B Karyamulya Kec. Kesambi Kota Cirebon, Indonesia

raihanabdulghani07@gmail.com

### ABSTRAK

Dalam era globalisasi, keberhasilan suatu organisasi sangat tergantung pada teknologi informasi yang mampu menyediakan laporan analisis data yang relevan. Dapodik merupakan aplikasi pendataan pendidikan yang dikelola oleh Kementerian Pendidikan, namun ketidakadaan sistem penyimpanan terpusat menghambat efektivitas analisis data. Ketidakefektifan akses dan analisis data sekolah yang tersebar dalam Dapodikdasmem membutuhkan solusi penyimpanan data yang terpusat. Hal ini mendorong kebutuhan akan perancangan *data warehouse* untuk memudahkan pengguna dalam melihat data dari berbagai lokasi. Penelitian bertujuan merancang *data pipeline* menggunakan *Extract Transform Load (ETL) process* yang menghasilkan *data warehouse*. Tujuannya adalah menciptakan sistem penyimpanan efektif yang mengintegrasikan data dari berbagai dimensi tabel, memungkinkan analisis data yang efisien dengan rentang waktu yang relevan. Perancangan data warehouse dilakukan dengan mengimplementasikan *ETL process* menggunakan *DataPrep by Trifacta* pada data sekolah yang kemudian disimpan dalam *BigQuery* sebagai sistem penyimpanan data. Pendekatan ini memberikan solusi penyimpanan data yang terpusat dan rapi untuk mengatasi permasalahan ketidakefektifan akses dan analisis data. Data warehouse yang dihasilkan memiliki struktur dimensi tabel yang terintegrasi, memberikan kemudahan analisis data dengan rentang waktu yang relevan. Proses ETL menggunakan *DataPrep by Trifacta* dan *BigQuery* berhasil mengatasi permasalahan data yang tersebar, menciptakan penyimpanan data yang lebih terpusat dan efisien.

**Kata kunci :** *Data warehouse, Data pipeline, ETL (Extract, Transform dan Load), Data integration, Transforming data, Database*

### 1. PENDAHULUAN

Data warehouse adalah konsep penyimpanan data terintegrasi yang dapat diandalkan untuk memenuhi semua kebutuhan analisis data. Maka dari itu penggunaan data warehouse sangat penting dalam aspek penelitian dan analisis. Penelitian ini bertujuan untuk merancang sebuah desain data pipeline untuk data warehouse.

Data dalam konteks ini adalah informasi yang dihasilkan oleh survey pemerintah tentang data sekolah. Data ini dapat digunakan dan siap untuk dianalisis. Data warehouse adalah sistem penyimpanan data yang dirancang untuk mendukung pengambilan keputusan dengan cara mengintegrasikan data dari berbagai sumber dan mentransformasikannya agar data bisa dianalisis secara efektif.

Penelitian terdahulu yang relevan dengan topik ini mencakup beberapa aspek penting. Salah satunya adalah penelitian oleh Vicky Novreza, Yuda Munarko, Lailatul Husniah pada tahun 2020 dengan judul "Data Warehouse Menggunakan Snowflake Schema Pada Virtual Shop"[1]. Penulis melakukan penelitian untuk merancang data warehouse untuk data transaksi sebuah online market yang menggunakan teknologi virtual reality. Mereka menggunakan snowflake schema untuk mengurangi redundansi data pada tabel dimensi. Mereka menggunakan MySQL sebagai database dan data warehousenya, Talend Open Data

Studio untuk data processing (ETL), dan Pentaho BI Server untuk menampilkan report OLAP Cube berbasis web.

Data pada table-table tersebut akan menunjukkan data tentang sekolah di kabupaten serang provinsi banten, yang dapat digunakan untuk mengidentifikasi dan menganalisis kualitas dan kuantitas pendidikan data ini akan menjadi dasar untuk analisis lebih lanjut dalam penelitian ini.

Masalah yang dihadapi dalam penelitian ini adalah bahwa data sekolah perjenjang yaitu SD, SLB, SMP, SMA DAN SMK per kecamatan yang berjumlah 145 tabel tersebar di berbagai tabel dan memerlukan pengintegrasian dan juga transformasi terhadap data yang terintegrasi untuk analisis yang efektif. Oleh karena itu, diperlukan perancangan data warehouse yang efisien.

Tujuan dari jurnal ini adalah merancang dan mengimplementasikan data warehouse menggunakan ETL process untuk mengintegrasikan data sekolah dari semua jenjang di kabupaten serang. Hal ini bertujuan untuk mengurangi redundansi data, memfasilitasi akses dan analisis data, serta membantu kinerja penyimpanan yang lebih baik.

### 2. TINJAUAN PUSTAKA

#### 2.1. Landasan Teori

### 2.1.1. Data Warehouse

Data warehouse merupakan konsep penyimpanan data terintegrasi yang dirancang untuk mendukung kebutuhan analisis data. Menurut Inmon (2002) data Warehouse adalah sekumpulan data yang bersifat *integrated, subject-oriented, time variant* dan *nonvolatile* dalam mendukung pengambilan keputusan manajemen[2]. Subject-oriented berarti data diorganisir berdasarkan topik atau subjek tertentu, terintegrasi mencakup data dari berbagai sumber yang diubah menjadi format yang konsisten, non-volatile berarti data disimpan dalam jangka panjang tanpa modifikasi, dan time-variant menunjukkan bahwa data disimpan dengan informasi waktu. Pentingnya data warehouse terletak pada kemampuannya untuk menyediakan akses efisien dan cepat terhadap data yang tersebar, memfasilitasi analisis, dan mendukung pengambilan keputusan yang informasional.

### 2.1.2. Data Pipeline Architecture

Arsitektur data warehouse atau arsitektur data warehouse merupakan tahapan proses pengumpulan beberapa sumber yang terpisah kemudian disatukan kedalam satu tempat penyimpanan yang bertujuan untuk memudahkan proses load data[3]. Desain yang baik dari data pipeline sangat penting untuk memastikan kinerja optimal dan integritas data. Atau dengan kata lain, Data pipeline adalah rangkaian proses terhubung di mana keluaran dari satu atau lebih proses menjadi masukan untuk proses lainnya[4].

### 2.1.3. Etl Process (Extract, Transform, Load)

ETL merupakan suatu metodologi yang umum digunakan dalam desain data pipeline. Menurut Inmon (2005), ETL dapat membantu mengurangi redundansi data, meningkatkan aksesibilitas data, dan memungkinkan analisis data yang lebih efisien[5]. Tentu saja, ETL (Extract, Transform, Load) adalah proses penting dalam pengudangan data di mana sebagian besar pembersihan dan kurasi data dilakukan. ETL mencakup tiga langkah yaitu ekstraksi, transformasi, dan pengiriman data dari berbagai sumber ke satu penyimpanan sentral. Proses ini memungkinkan bisnis Anda untuk memiliki akses ke data yang lebih akurat dan terstruktur[6]. Pemilihan alat atau platform untuk ETL, seperti DataPrep by Trifacta, dapat memainkan peran kunci dalam efisiensi dan keberhasilan proses ini.

### 2.1.4. Data Integration

Data integration merujuk pada proses menggabungkan data yang memiliki banyak label, misalnya, fusi data, asimilasi, kombinasi, atau integrasi[7]. Pengintegrasian data diperlukan untuk memastikan bahwa data yang disimpan dalam data warehouse dapat diandalkan dan relevan. Inmon (2005) mencatat bahwa integrasi data melibatkan pemodelan data, pemilihan metode integrasi, dan implementasi teknologi untuk mengatasi perbedaan format dan struktur data[8].

### 2.1.5. Transforming Data

Transformasi data adalah proses mengubah struktur atau nilai data agar sesuai dengan kebutuhan analisis dan penyimpanan dalam data warehouse. Beberapa contoh transformasi data yang mungkin dilakukan adalah menambahkan konstanta, mengalikan, mengkuadratkan atau memangkatkan, mengonversi ke skala logaritmik, membalik dan memantulkan, mengambil akar kuadrat dari nilai, dan bahkan menerapkan transformasi trigonometri seperti transformasi gelombang sinus[9].

### 2.1.6. Database

Pemilihan database untuk penyimpanan data sangat penting dalam desain data warehouse. Basis data adalah daftar garis yang sangat akurat, yang selain posisi garis dan intensitas yang akurat, menyediakan parameter advance line-shape termasuk pencampuran garis[10]. Berbagai jenis database, seperti BigQuery, menawarkan skema penyimpanan yang berbeda dan tingkat kinerja yang beragam. Database yang efisien dapat meningkatkan performa pengambilan data dan analisis.

### 2.1.7. Hasil Penelitian Terdahulu

"Data Warehouse Menggunakan Snowflake Schema Pada Virtual Shop"[1] Penelitian ini dilakukan oleh Vicky Novreza, Yuda Munarko, dan Lailatul Husniah pada tahun 2020. Mereka merancang data warehouse untuk data transaksi sebuah online market yang menggunakan teknologi virtual reality. Dalam penelitiannya, mereka menggunakan snowflake schema untuk mengurangi redundansi data pada tabel dimensi. MySQL digunakan sebagai database dan data warehouse, Talend Open Data Studio digunakan untuk proses ETL, dan Pentaho BI Server digunakan untuk menampilkan report OLAP Cube berbasis web. Hasil penelitian ini memberikan wawasan tentang pemilihan skema dan alat yang tepat dalam merancang data warehouse.

## 3. METODE PENELITIAN

### 3.1. Sumber Data

Data yang digunakan dalam studi ini adalah data sekunder yang diambil dari DAPODIKDASMEN. Data tersebut mencakup data sekolah semua jenjang per kecamatan di Kabupaten Serang dari sumber dataset url: Data Sekolah Prov. Banten - Dapodikdasmen (kemdikbud.go.id)

### 3.2. Teknik Pengumpulan Data

Metode yang digunakan adalah studi literature untuk menemukan data yang akurat dan memenuhi kriteria analisis kebutuhan data.

### 3.3. Tahapan Perancangan

Tahapan perancangan dalam studi ini mencakup serangkaian langkah untuk merancang dan mengimplementasikan data warehouse yang sesuai dengan ETL Process. Berikut adalah tahapan perancangannya:

**3.3.1. Data Selection**

Pada proses ini dilakukan pemilihan data, pengecekan data dan identifikasi kebutuhan data untuk memahami jenis data yang diperlukan, struktur data yang diinginkan, serta kebutuhan analisis yang akan dilakukan terhadap data sekolah di Kabupaten Serang.

**3.3.2. Pemilihan Data Pipeline Architecture**

Berdasarkan kebutuhan data dan ketersediaan sumber daya, pada proses ini pemilihan arsitektur data pipeline akan dilakukan. Data Pipeline Architecture yang tepat dapat membantu memastikan bahwa data dapat diproses dengan cepat, akurat, dan efisien.

**3.3.3. Implementasi Data Pipeline**

Pada tahap ini dilakukan implementasi data pipeline architecture yang telah dibuat:

- a) Load Dataset kedalam Data Lake  
 Pada tahap ini dilakukan load dataset kedalam untuk dimuat kedalam data lake untuk dilakukan proses lebih lanjut.
- b) Desain dan data Load Data Staging  
 Pada tahap ini dilakukan desain data staging, Dimana data dalam data lake dimuat pada tabel yang sudah dibuat .
- c) Implementasi ETL Process  
 Pada tahap ini dilakukan ETL process, Dimana data dalam data staging di extract, transform, lalu akan di load kedalam data warehouse.
- d) Uji Validasi Data  
 Setelah implementasi, uji kinerja data warehouse akan dilakukan untuk memastikan bahwa proses ETL berjalan dengan baik dan data yang dimuat ke dalam data warehouse valid. Pengujian ini melibatkan analisis akurasi dan kualitas data.
- e) Optimalisasi Data Pipeline  
 Langkah terakhir adalah melakukan optimalisasi data pipeline. Proses ini mencakup batch processing untuk menjaga integritas data secara terjadwal.

Dengan mengikuti tahapan perancangan ini, data warehouse yang diimplementasikan dapat memenuhi kebutuhan analisis data sekolah di Kabupaten Serang dan memberikan kontribusi positif dalam pengelolaan dan peningkatan kualitas pendidikan.

**4. HASIL DAN PEMBAHASAN**

**4.1. Hasil Tujuan 1**

**4.1.1. Hasil Pengumpulan Data (Data Collecting)**

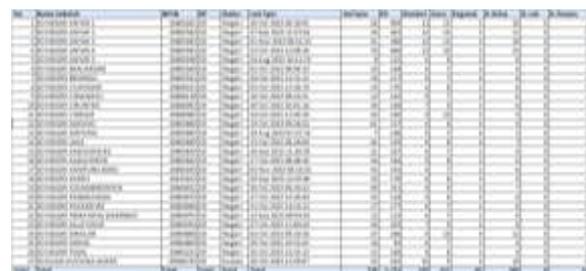
Data yang digunakan pada penelitian ini menggunakan data asli sekolah semester 2023/2024 Ganjil di Kabupaten Serang. Atribut yang ada pada

data ini terdiri dari No, Nama Sekolah, NPSN, BP, Status, Last Sync, Jml Sync, PD, Rombel, Guru, Pegawai,, R. Kelas, R. Lab dan R. Perpus. Dengan keseluruhan berjumlah 145 tabel. Tabel di bawah ini adalah atribut dari data sekolah Semester 2023/2024 Ganjil yang dikumpulkan oleh KEMENDIKBUD:

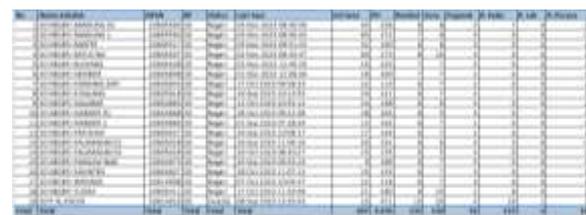
Tabel 1. atribut table sd\_kecamatan\_anyar

Nama kolom	Deskripsi
No	Nomor baris
NPSN	Nomor Pokok Sekolah Nasional
BP	Jenjang pendidikan
Status	Status kepemilikan sekolah
Last Sync	Sinkronisasi terakhir
Jml Sync	Jumlah keseluruhan sinkronisasi
PD	Jumlah peserta didik
Rombel	Jumlah rombongan belajar
Guru	Jumlah tenaga pengajar
Pegawai	Jumlah pegawai
R. Kelas	Jumlah ruang kelas
R. Lab	Jumlah ruang laboratorium
R.Perpus	Jumlah ruang perpustakaan

Berikut pada Gambar 1 dan Gambar 2 diperlihatkan contoh data per kecamatan dan per jenjang yang ada:



Gambar 1. SD Kecamatan Anyar



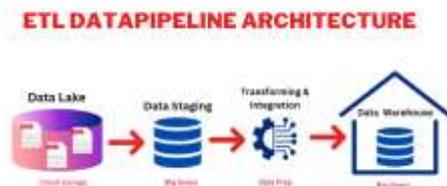
Gambar 2. SD Kecamatan Bandung

Dengan kumpulan dataset diatas bisa disimpulkan bahwa dataset yang dikumpulkan sudah dalam bentuk *data mart* yaitu *dataset* yang sudah diproses dan diorganisir sedemikian rupa sehingga dapat dengan mudah diakses atau bagian terkecil dari *data warehouse*. Dan dengan tujuan penelitian ini yang berfokus dalam kualitas pendidikan dalam lingkup fasilitas dan tenaga pengajar, maka dalam hal ini kebutuhan analisis hanya mencakup data sekolah, jumlah ruang tertentu dalam lingkup sekolah dan jumlah pengajar dan staff.

4.2. Hasil Tujuan 2

4.2.1. Hasil Merancang Data Pipeline

Berdasarkan kebutuhan data dan ketersediaan sumber daya, pemilihan arsitektur data pipeline dilakukan. Dalam hal ini, Extract Transform Load (ETL) process akan menjadi metodologi utama untuk mengintegrasikan data dari berbagai sumber dan juga *data transforming* pada data yang terintegrasi. Arsitektur data pipeline bisa dilihat pada gambar dibawah ini:



Gambar 3. ETL Data Pipeline Architecture

4.3. Hasil Tujuan 3

4.3.1. Implementasi ETL process dalam data pipeline pada Google Cloud Platform

Berikut dijelaskan mengenai Langkah kerja pengimplementasian ETL process pada Google Cloud Platform, guna untuk melihat cara dan hasil dari kerja ETL process:

a. Load data ke Data Lake(Google Cloud Storage)

Langkah pertama yaitu Load Dataset kedalam Data Lake, masuk ke bucket Google Cloud Storage lalu drag and drop folder kedalam bucket.



Gambar 4. Dataset dalam Data Lake

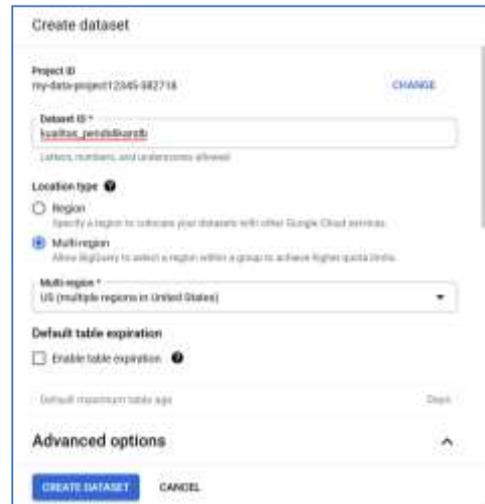
b. Desain dan data Load ke Data Staging

Membuat data staging pada Bigquery dilakukan seperti berikut:



Gambar 5. Pembuatan data staging di BigQuery

Pada antarmuka BigQuery, klik simbol titik tiga disamping nama project lalu klik *create dataset* seperti pada Gambar 5.



Gambar 6. Pembuatan Data Staging di BigQuery

Pada antarmuka *create dataset* seperti pada Gambar 6 pilih *Multi-region* lalu klik *create dataset*.



Gambar 7. Pembuatan Tabel pada Data Staging di BigQuery

Pada tahap pembuatan table pada Data Staging. Dalam *query editor* ketikkan query untuk membuat table dan load data dari data lake ke table tertentu seperti pada Gambar 7 diatas.

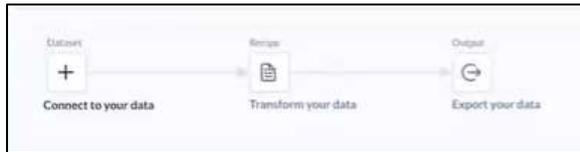


Gambar 8. Pembuatan Tabel pada Data Staging di BigQuery

Dikarenakan BigQuery cenderung merubah nama kolom berdasarkan nama kolom yang ada pada data asli ketika di load. Maka perlu merubah lagi nama kolom pada table seperti pada Gambar 8.

c. Implementasi ETL Process

Pada tahap ini dilakukan ETL process, Dimana hal pertama yang dilakukan adalah membuat flow yang terdiri dari input, receipt dan output:



Gambar 9. Flow pada Dataprep

Pada gambar 9 klik tanda tambah di *Connect your data*.



Gambar 10. Import datasets pada Dataprep

Lalu klik *import datasets* seperti pada gambar 10 Setelah selesai meng-*import* semua *datasets*(tabel) lalu lakukan union semua *datasets* dengan cara klik kanan pada receipt lalu klik *Append Union* lalu akan muncul seperti gambar 11 klik *Add data* untuk menambahkan *datasets* untuk di *Union*-kan lalu klik *Add to recipe* mengintegrasikan semua table kedalam *recipe*.



Gambar 11. Union Data pada Dataprep

Pada tahap *Transforming* data, klik *recipe* lalu klik *edit recipe* pada antarmuka *Details* di sebelah kanan layar. Dan selanjutnya akan muncul *recipe* data yang sudah terintegrasi seperti pada Gambar 12.



Gambar 12. Data sebelum *transform*

Setelah tahap diatas selesai, dilakukan proses transform sebagai berikut:



Gambar 13. proses *data transform* pada *recipe*

Pada tahap *transforming*, dilakukan beberapa tahap *data cleaning* pada *recipe* agar data yang di-*load* kedalam data warehouse sesuai dengan kebutuhan, berikut beberapa penjelasan seperti pada gambar 13:

1. *Union with 144 tables*: pada *dataset* *sd\_kecamatan\_cikeusal - 4* di-*union*-kan dengan 144 *dataset* lainnya.
2. *Delete jml\_sync*: hapus kolom *jml\_sync*.
3. *Set no to IFMISMATCHED(\$col, ['Integer'] '')*:menentukan apakah nilai dalam kolom tertentu cocok dengan tipe data yang diinginkan atau tidak, dalam hal ini adalah kolom-kolom yang berisi nilai berbentuk angka.
4. *Delete rows with missing values in no*: menghapus baris yang memiliki nilai kosong pada kolom *no*.
5. *Set last\_sync to IFMISMATCHED(\$col, ['Datetime.dd-mm-yy hh:mm:ss"dd"shortMonth"yyyy"HH:MM:SS'])*:digunakan untuk memeriksa apakah nilai dalam kolom tertentu cocok dengan format tanggal dan waktu yang diinginkan atau tidak.
6. *Lock bp type to String*: kunci tipe data pada kolom *bp* kedalam *String*.
7. *Replaces matches of "." From pd with ""*: ganti semua nilai dalam kolom *pd* yang berisi "." Dengan "".
8. *Delete rows with missing values in last\_sync*: menghapus baris yang memiliki nilai kosong pada kolom *last\_sync*.
9. *Set pd to IF(pd == '0', NULL(), \$col)*:digunakan untuk mengubah nilai dalam kolom tertentu menjadi NULL jika nilai tersebut sama dengan '0'.
10. *Delete rows with missing values in pd*: menghapus baris yang memiliki nilai kosong pada kolom *pd*.

Jika proses data transform selesai, maka data pada dataprep terlihat seperti pada gambar 14 dibawah:



Gambar 14. Data setelah transform

Setelah keseluruhan proses transform selesai, maka dilakukan output ke dalam data warehouse di BigQuery. Sebelumnya lakukan pembuatan dataset untuk data warehouse seperti pada gambar 5 dan 6 dengan nama dataset kualitas\_pendidikandwh.



Gambar 15. Publishing action dalam output setting

Selanjutnya buat output untuk data warehouse, klik output pada flow lalu buat create new table pada publishing action lalu klik Truncate the table every run, Dimana data dalam tabel akan dihapus dan diganti dengan data yang baru dalam tabel yang sama setiap kali ETL process berjalan lalu klik update seperti pada gambar 15.



Gambar 16. Pembuatan Table pada Data Staging di BigQuery

Setelah semua langkah-langkah diatas selesai, maka flow akan terlihat seperti pada gambar 16 lalu klik Run.

d. Uji Kinerja Data Pipeline

Untuk melihat apakah data pipeline bekerja dengan benar untuk merancang data warehouse, dilakukan pengecekan pada tabel kualitas\_pendidikandwh pada

dataset kualitas\_pendidikandwh di BigQuery. Ketik query pada query editor di bigQuery seperti pada gambar gambar 17 dibawah.



Gambar 17. query untuk mengecek data pada data warehouse

e. Optimalisasi Data Pipeline

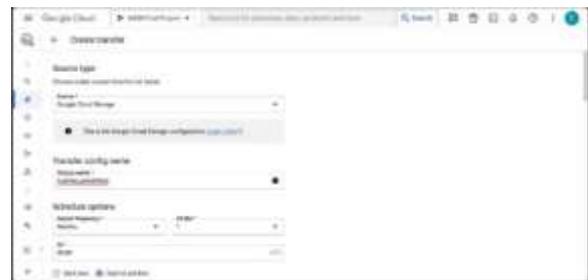
Pada Langkah terakhir ini dilakukan optimalisasi pada data pipeline dengan menggunakan metode batch processing, yaitu mesin atau dalam hal ini Google Cloud akan menyelesaikan tugas data yang berulang-ulang secara berkala. Ada dua tahapan dalam batch scheduling dalam penelitian ini yaitu transfer scheduling dari data lake ke data staging dan scheduling untuk ETL process, Berikut tahapannya:

a. Transfer scheduling dari data lake ke data staging.



Gambar 18. console BigQuery

masuk ke antarmuka di sebelah kiri di console BigQuery seperti pada gambar 4.16



Gambar 18. Data Transfer Scheduling from Data Lake to Data Staging

lalu atur Source pada Google Cloud Storage, ganti display name dengan kualitas\_pendidikan dan terakhir atur waktu agar data transferring mulai setiap satu bulan sekali seperti pada gambar 18.



Gambar 20. Data Transfer Scheduling from Data Lake to Data Staging

Atur Destination Setting pada data staging yang ingin kita masukkan.



Gambar 21. Data Transfer Scheduling from Data Lake to Data Staging

Dan terakhir klik save seperti pada Gambar 21

- b. *Scheduling* untuk *ETL process*, masuk ke flow lalu klik *schedule* dibagian kanan atas lalu akan muncul pop-up seperti gambar 22 lalu atur *ETL process* agar berjalan setiap satu bulan sekali.



Gambar 22. ETL Process Scheduling

## 5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk merancang dan mengimplementasikan data warehouse menggunakan ETL Process dalam konteks analisis kualitas pendidikan di Kabupaten Serang. Berdasarkan hasil pembahasan, ditemukan kesimpulan utama mengenai pentingnya data warehouse sebagai penyimpanan data terintegrasi yang mendukung pengambilan keputusan, dengan implementasi konsep yang mengikuti prinsip-prinsip integrasi, subject-oriented, time variant, dan nonvolatile. Pemilihan desain data pipeline, khususnya ETL Process, dianggap krusial untuk memastikan kinerja optimal dan integritas data. Penggunaan ETL Process dipilih berdasarkan umumnya ditemui dan sesuai dengan sumber daya yang dimiliki. Pengaruh

pemilihan teknologi, seperti alat atau platform ETL seperti DataPrep by Trifacta, terbukti signifikan terhadap efisiensi dan keberhasilan proses ETL, meskipun terdapat kendala terkait ketergantungan pada layanan Google Cloud Platform.

Berdasarkan kesimpulan di atas, beberapa saran untuk pengembangan selanjutnya dapat diajukan. Pertama, disarankan untuk mengimplementasikan Extract, Load, and Transform (ELT) Process sebagai inovasi dalam desain data warehouse dan ETL process, mengingat perkembangan pesat dalam teknologi. Evaluasi berkala terhadap arsitektur data pipeline perlu dilakukan untuk memastikan adopsi teknologi terbaru, memastikan data warehouse selalu diperbarui dan dapat diakses serta dianalisis dengan efisien. Kedua, penelitian lebih lanjut dapat melibatkan perbandingan berbagai alat ETL selain DataPrep by Trifacta, seperti Talend Data Integration, Pentaho Data Integration, atau alat-alat ETL lainnya, untuk memberikan wawasan lebih mendalam tentang keunggulan dan kelemahan setiap alat. Ketiga, disarankan penelitian terkait penggunaan Non-Cloud Environment dalam Data Pipeline Architecture sebagai alternatif selain Cloud environment, terutama Google Cloud Platform. Dengan menerapkan saran-saran ini, diharapkan dapat memperkuat implementasi data warehouse dan memberikan dampak positif yang lebih besar dalam mendukung analisis kualitas pendidikan di Kabupaten Serang.

## DAFTAR PUSTAKA

- [1] V. Novreza, Y. Munarko, and L. Husniah, "Data Warehouse Menggunakan Snowflake Schema Pada Virtual Shop," *J. Repos.*, vol. 2, no. 1, p. 67, 2020, doi: 10.22219/repository.v2i1.180.
- [2] U. L. Q. Di, "RANCANGAN DATA WAREHOUSE UNTUK PENUNJANG SISTEM INFORMASI EKSEKUTIF PADA YAYASAN," vol. 4, no. 1, 2018.
- [3] I. P. A. E. Pratama and N. P. N. D. Widhiasih, "Perancangan Data Warehouse Untuk Prediksi Penjualan Pada Orba Express Menggunakan Pentaho," *JUSS (Jurnal Sains dan Sist. Informasi)*, vol. 3, no. 2, pp. 43–48, 2020, doi: 10.22437/juss.v3i2.8147.
- [4] A. R. Munappy, J. Bosch, and H. H. Olsson, "Data Pipeline Management in Practice: Challenges and Opportunities," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12562 LNCS, pp. 168–184, 2020, doi: 10.1007/978-3-030-64148-1\_11.
- [5] A. Kuntadi, Y. Feriandi, J. Ir, H. J. No, K. Coblong, and J. B. Indonesia, "Penggunaan SQL Server dalam Pengolahan Data Warehouse yang Praktis dan Berkelanjutan," pp. 457–468.
- [6] M. Souibgui, F. Atigui, S. Zammali, S. Cherfi, and S. Ben Yahia, "Data quality in ETL process: A preliminary study," *Procedia Comput. Sci.*, vol. 159, pp. 676–687, 2019, doi:

- 10.1016/j.procs.2019.09.223.
- [7] N. J. B. Isaac *et al.*, “Data Integration for Large-Scale Models of Species Distributions,” *Trends Ecol. Evol.*, vol. 35, no. 1, pp. 56–67, 2020, doi: 10.1016/j.tree.2019.08.006.
- [8] B. Prihantoro, “Evaluasi penerapan konsep integrasi data menggunakan dhis2 di kementerian kesehatan,” *J. Inf. Syst. Public Heal.*, vol. 5, no. 1, p. 43, 2021, doi: 10.22146/jisph.33959.
- [9] J. W. Osborne and D. Ph, “Notes on the use of data transformations . Data transformation and normality,” *Pract. Assessment, Res. Eval.*, vol. 8, no. 6, pp. 1–7, 2008.
- [10] I. E. Gordon *et al.*, “The HITRAN2020 molecular spectroscopic database,” *J. Quant. Spectrosc. Radiat. Transf.*, vol. 277, p. 107949, 2022, doi: 10.1016/j.jqsrt.2021.107949.