

# USULAN SOLUSI UNTUK MEMBANGUN *CUSTOMER MASTER FILE* PADA PT. ABC

Ardijan Handijono<sup>1</sup>, Zaldy Suhatman<sup>2</sup>

Sistem Informasi, Universitas Pamulang,  
Jl. Raya Puspitek No. 46 Buaran, Serpong, Tangerang  
Selatan, Banten, Indonesia, 15417

Email : dosen00853@unpam.ac.id<sup>1</sup>, zaldy@unpam.ac.id<sup>2</sup>

**ABSTRAK** - PT. ABC adalah perusahaan penjaminan kredit, yang menerima data pengajuan kredit dan data nasabah dari beberapa bank rekan bisnisnya. Setiap bank mengirimkan data nasabah dengan kelengkapan data dan struktur data yang berbeda-beda. Masalah timbul ketika data nasabah diketahui banyak yang duplikat, hal ini terjadi karena tidak setiap data nasabah yang diterima dari Bank memberikan nomor KTP sehingga setiap data nasabah yang baru diterima di anggap sebagai data nasabah baru. Pada kasus ini kita akan fokus pada masalah data nasabah. Dipasaran ada beberapa solusi untuk problem ini yaitu aplikasi Customer Data Management (CDI) namun aplikasi ini hanya bisa menunjang proses bisnis yang umum, tidak sesuai digunakan pada kasus ini. Proses menyatukan beberapa data nasabah yang duplikat menjadi satu biasanya disebut proses Data Deduplication. Masalah akan timbul jika masing-masing data nasabah tersebut telah mempunyai sejarah transaksi, maka kode identitas nasabah tidak bisa di ganti atau dihapus karena sudah terhubung ke tabel transaksi sebagai Foreign Key. Masalah ini dapat dirumuskan menjadi beberapa pertanyaan sebagai berikut: (1) Bagaimana cara mengidentifikasi beberapa data nasabah yang mirip adalah orang yang sama. (2) Bagaimana cara menjadikan satu Data duplikat tanpa menghilangkan transaksi history yang pernah dilakukan. (3) Bagaimana cara memberikan kode identitas yang unik setelah data nasabah disatukan. Untuk dapat menjawab pertanyaan-pertanyaan tersebut peneliti melakukan literatur review dari beberapa jurnal sebelumnya, melakukan analisa dan sintesa untuk bisa memberikan solusi, dan hasil akhirnya membuat kerangka desain sistem dalam bentuk proses flow dan desain database.

**Kata Kunci:** Data Deduplication, CDI, CIF Number

## 1. PENDAHULUAN

### Latar Belakang

Kualitas Data yang rendah memicu bertambahnya biaya operasional dan tidak dapat digunakan untuk menunjang pengambilan keputusan jika data tersebut diolah menjadi informasi. Problem data kotor akan semakin meningkat saat data diintegrasikan dari banyak sumber data. Duplikat record telah menjadi masalah besar pada data master, terutama data nasabah. Akan menjadi sangat penting jika keputusan-keputusan bisnis bersandar pada data. Suatu record diidentifikasi duplikat jika ditemukan beberapa record merepresentasikan objek yang sama.

Perusahaan ABC adalah perusahaan yang bergerak pada bisnis asuransi Kredit, yaitu memberikan jaminan kepada bank-bank atas Kredit yang disalurkan pada nasabah. Sehingga jika terjadi Kredit macet Bank tidak terdampak, karena Perusahaan ABC akan menanggung risikonya. Sebagian besar Kredit adalah program pemerintah yang menyalurkan kredit untuk UMKM (Usaha Mikro, Kecil dan Menengah) lewat-Bank-bank Pemerintah. Setelah terjadi kesepakatan

bisnis dengan suatu bank yang disebutnya Mitra, maka Mitra tersebut mulai mengajukan Penjaminan Kredit dengan mengirimkan data transaksi Kredit yang telah dikururkan ke UMKM beserta data Nasabahnya. Nasabah disini adalah pemilik usaha UMKM, namun bisa juga Instansi UMKM sendiri.

Dalam bisnis ini kinerja perusahaan ABC sangat tergantung pada Bank-Bank yang mengajukan penjaminan Kredit, sehingga ada kebijakan tidak tertulis bahwa organisasi akan menerima apapun bentuk dan kelengkapan data nasabah dari Bank-Bank. Sebagai contoh pada nasabah Personal beberapa Bank memberikan data nomor KTP, namun ada Bank yang tidak memberikan nomor KTP. Data pengajuan penjaminan Kredit dari Bank umumnya dikirim secara Batch lewat mekanisme Host-to-Host. Perusahaan ABC akan menerima data Host-to-Host tersebut ke tabel Penampungan. Setiap Bank akan menggunakan Tabel Penampungan tersendiri karena setiap Bank bisa mengirimkan data yang berbeda-beda kelengkapannya dan strukturnya, sehingga tabel Penampungan akan berbeda strukturnya untuk setiap Bank, bahkan satu Bank

bisa mempunyai beberapa Tabel Penampungan karena produk yang berbeda, kelengkapan data nya berbeda sehingga struktur tabel Penampungannya juga berbeda. Dalam tabel Penampungan ini akan dilakukan beberapa validasi untuk memastikan data yang dikirimkan Bank sudah benar dan sesuai dengan kesepakatan. Pada sore hari setelah jam kerja selesai data dari semua tabel Penampungan ini akan di sinkronisasikan ke database Konsolidasi yang terpusat. Semua data transaksi di tabel Penampungan akan disimpan di beberapa tabel Transaksi pada Database Konsolidasi dan semua data nasabah dari beberapa tabel Penampungan akan disimpan di tabel nasabah pada Database Konsolidasi. ID nasabah baru yang berupa nomor urut (running number) akan diberikan sebagai pengikat dengan data transaksi Kreditnya, jika data nasabah tersebut masuk lagi pada waktu lain akan mendapatkan Nasabah ID yang berbeda lagi walaupun orangnya sama. Kondisi ini sudah berjalan puluhan tahun, tabel nasabah bersifat akumulatif, artinya data baru akan ditambahkan tanpa mengetahui apakah data nasabah tersebut sudah pernah tercatat atau belum. Dengan kondisi sistem seperti ini ukuran tabel nasabah akan terus tumbuh membesar dengan tingkat duplikasi data yang sangat tinggi. Saat ini tabel data nasabah lebih dari 60 juta record padahal diperkirakan ada lebih 60% adalah data duplikat. Kondisi ini berdampak pada performance database dan tidak efisiennya storage, lebih jauh menjadikan perhitungan Eksposur tidak tepat dan sistem Black-List yang longgar. Total pinjaman seorang nasabah disebut Eksposur, secara bisnis ada batas Eksposur pada seseorang yang ditentukan dengan metode scoring dari banyak variabel. Artinya saat nasabah mengajukan Kredit, pihak Bank harus menolak jika Eksposur nasabah melebihi limit yang telah ditentukan. Perusahaan ABC juga melakukan perhitungan eksposur ini untuk melihat apakah Kredit nasabah sudah melebihi limit atau belum. Permasalahan muncul karena Perusahaan tidak mempunyai single ID untuk nasabah karena nomor KTP tidak selalu diberikan Bank. Sehingga seseorang Kreditor yang sudah mengambil Kredit di beberapa Bank yang berbeda akan tercatat di tabel nasabah sebagai orang yang berbeda-beda, ini menjadikan perhitungan Eksposur tidak bisa akurat. Secara bisnis Perusahaan juga menerapkan sistem Black-List, yang datanya disimpan pada tabel Black-List. Seorang nasabah akan masuk Black-List jika mengajukan Klaim tapi Subrogasi belum dibayar lunas. Subrogasi adalah pengalihan hak tagih dari Bank ke PT. ABC. Dengan tidak adanya single ID nasabah maka nasabah yang sudah masuk Black-List masih bisa lolos jika Bank mengajukan penjaminan Kredit lagi walaupun untuk orang yang sama.

Dari kedua masalah tersebut organisasi membutuhkan *Master Data Management* (MDM) khususnya untuk Master data Nasabah atau CDI

(*Customer Data Integration*). Sistem tersebut harus mempunyai kemampuan: dapat membentuk Single ID untuk semua nasabah, Mampu melakukan identifikasi dengan cepat apakah data nasabah yang baru datang sudah tersimpan pada tabel Master Customer atau belum, mampu melakukan proses deduplikasi dengan mengidentifikasi data nasabah yang sebenarnya satu individu, mendukung *Matching Process* baik secara sistem atau secara manual. Masalah teknis yang mengemuka dalam proses pembentukan *Single Customer Master* adalah proses deduplikasi data, Manfaat lebih rinci deduplikasi data untuk organisasi meliputi (Viji & Revathy, 2019) :

- Mengurangi jumlah ruang data dalam sistem penyimpanan.
- Dalam sistem penyimpanan primer, maksimalkan kapasitas penyimpanan flash
- Konsumsi bandwidth lebih rendah.
- Periode penyimpanan data yang lebih tinggi.
- Pemulihan cepat.
- Mengurangi cadangan tape

Lebih jauh untuk bisnis PT ABC pembentukan *Single Customer Master* akan memberikan manfaat sebagai berikut:

- Meningkatkan kualitas pengambilan keputusan
- Memungkinkan organisasi untuk mengimplementasikan strategi bisnis CRM seperti Up-sell/Cross-Sell, Targeted Marketing, Customer Segmentation, dsb.
- Memungkinkan organisasi melakukan Customer Profitability Analysis
- Perhitungan Data Eksposur setiap nasabah lebih akurat
- Mendukung penerapan aturan Black-List lebih ketat

#### **Rumusan Masalah**

Rumusan masalah dari penelitian ini adalah sebagai berikut:

- 1) Bagaimana *Matching Process* untuk mengidentifikasi kemiripan nasabah?
- 2) Bagaimana proses konfirmasi untuk memastikan bahwa data nasabah suspek adalah orang yang sama?
- 3) Bagaimana desain Database untuk proses merging beberapa data nasabah yang telah diidentifikasi sebagai orang yang sama? Tanpa menghilangkan kunci identifier yang lama.
- 4) Bagaimana membuat desain sistem untuk proses Data Deduplikasi baik secara sistem atau secara manual.

#### **2. TINJAUAN PUSTAKA**

Data harus sangat akurat untuk menghasilkan Informasi berkualitas tinggi. Pengumpulan data semakin penting bagi banyak bisnis dan organisasi, dan penting untuk memiliki rencana untuk memantau kualitas data. Menurut Gartner, kualitas data yang buruk merugikan

organisasi rata-rata \$12,9 juta per tahun. (El-Hadi, 2023).

Kualitas data sering kali didefinisikan sebagai kesesuaian data untuk digunakan pada tujuan penggunaan tertentu yang diperlukan, yaitu harus bebas kesalahan, lengkap, benar, terkini, dan konsisten. (Azeroual & Abuosba, 2019)

Memahami cara memantau dan meningkatkan kualitas data adalah langkah pertama dalam menghasilkan data berkualitas baik sekaligus meningkatkan pendapatan dan produktivitas. Menurut Statistik Kualitas Data tahun 2022, ada beberapa temuan penting yang perlu dipertimbangkan oleh para pemimpin industri (El-Hadi, 2023) :

- Hanya 16% perusahaan yang menggolongkan data yang mereka gunakan sebagai “sangat baik.”
- 54% bisnis menyebut kualitas dan kelengkapan data sebagai tantangan pengelolaan data pemasaran terbesar mereka.
- Sekitar 47% data baru yang dikumpulkan oleh bisnis memiliki satu atau lebih kesalahan kritis.
- Masalah kualitas data merugikan perekonomian AS saja sekitar \$3,1 triliun per tahun.

Data memiliki banyak representasi – artinya data yang sama dapat direpresentasikan dengan cara yang berbeda. Ini adalah alasan utama mengapa ada record duplikat di database. Baik record gabungan dari beberapa sumber, atau dimasukkan ke dalam database tunggal namun seiring berjalannya waktu, keduanya berpotensi menyebabkan masalah duplikasi data yang kompleks. Idealnya, setiap record dalam database harus mewakili satu entitas unik. Namun karena sejumlah alasan maka satu entitas tersimpan pada beberapa record. Berikut adalah beberapa hal yang dapat menjadikan duplikat data:

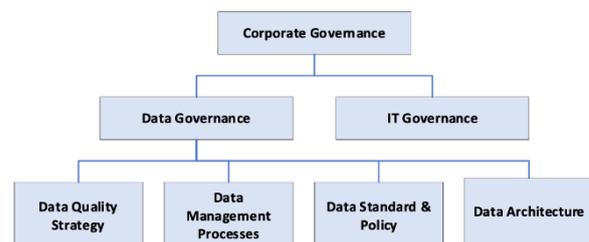
- **No unique identifiers** - Memiliki pengidentifikasi unik di database adalah cara terbaik untuk menghindari penyimpanan duplikat. Pengidentifikasi unik adalah bidang data yang selalu unik untuk suatu entitas (misalnya, Nomor Induk Kependudukan (NIK) untuk data pelanggan, Nomor NPWP untuk organisasi yang berbadan Hukum, Nomor Komponen Produsen (MPN) untuk data produk, dll.). Pada setiap entri data baru, dapat diperiksa apakah ada catatan dengan pengidentifikasi unik yang sama. Dan jika ya, sistem cukup memperbarui atau menggabungkannya, dan menghindari menyimpan catatan baru untuk entitas yang sama. Namun jika database tidak berisi pengidentifikasi unik tersebut, proses menghubungkan entitas baru yang masuk dengan entitas yang sudah ada menjadi tugas yang rumit.

- **No validation checks and integrity constraints** - Bahkan dengan adanya pengidentifikasi unik, Masih bisa terjadi duplikat di database. Hal ini terjadi ketika pengidentifikasi unik tidak sesuai dengan pola yang valid.
- **Error on Data entry** - Tingkat kesalahan entri data adalah sangat tinggi, jadi, meskipun terdapat pengidentifikasi unik, pemeriksaan validasi, dan batasan integritas, ada kemungkinan kesalahan operator dapat mengganggu dan memungkinkan duplikat dalam database.

Pada perusahaan di mana terdapat beberapa sumber data dan data didistribusikan ke beberapa sub-sistem, maka lingkungan data menjadi heterogen, model data, dan proses berbeda yang digunakan untuk mengelola data dalam perusahaan. Dalam keadaan ini, data pelanggan seringkali tidak lengkap, tidak akurat dan tidak konsisten di seluruh perusahaan. Permasalahan utama dalam pengelolaan data adalah kualitas data yang buruk, permasalahan lainnya melibatkan proses yang tidak memadai untuk memelihara data, kepemilikan data yang tidak jelas, dan tidak adanya pemeliharaan kualitas data yang berkelanjutan. (Sehat & Paves Flores, 2012)

#### Data Governance

Pengelolaan data Master seharusnya merupakan bagian dari *Data Governance*, dimana *Data Governance* merupakan perluasan dari *IT Governance* dengan mempertimbangkan aspek dan proses kualitas data yang didefinisikan secara khusus untuk memperjelas siklus hidup data. *Data Governance* memberikan kerangka kerja untuk keputusan pengelolaan terkait data dan bukan bagian dari *IT Data Governance*. Keduanya dipandang sama karena keduanya merupakan sub elemen dari *Data Governance Struktur Perusahaan*. (Traulsen, Tröbs, AG, Tucherpark, & Ganghoferstrasse, 2011)



Gambar 01-Ruang Lingkup Data Governance

#### Management Data Master (MDM)

Data Pelanggan, Data Suplier, Data Produk, adalah beberapa contoh dari Data Master, untuk menghindari duplikasi pada Data Master, solusi yang umum adalah menerapkan MDM adalah salah satu disiplin ilmu manajemen data yang paling banyak diadopsi dalam beberapa tahun terakhir. Hal ini karena definisi entitas bisnis yang

berdasarkan konsensus dan penerapannya secara konsisten di seluruh perusahaan merupakan faktor penentu keberhasilan aktivitas bisnis lintas fungsi yang penting, seperti Business Intelligence (BI), pandangan pelanggan yang lengkap, keunggulan operasional, optimalisasi rantai pasokan, peraturan, dan regulasi. pelaporan, kepatuhan, merger dan akuisisi, dan memperlakukan data sebagai aset perusahaan. Karena alasan bisnis yang menarik ini, banyak organisasi telah menerapkan solusi MDM generasi pertama atau kedua. (Jasim, Fahad, & Applications, 2018)

MDM dapat didefinisikan sebagai serangkaian prosedur dan teknologi untuk mengumpulkan, dan memelihara data master berkualitas tinggi. MDM tidak hanya melibatkan teknologi Komputer saja, namun juga memerlukan organisasi untuk menerapkan kebijakan dan prosedur untuk mengendalikan bagaimana data master dibuat dan digunakan. Bahkan folkus MDM adalah manajemen dan bukan teknologi. Penggunaan teknologi baru tanpa melibatkan pengelolaan data tidak akan menghasilkan sistem informasi yang terpadu. Walaupun MDM bukan proyek teknologi, akan tetapi inisiatif ini tidak dapat dilakukan tanpa dukungan teknologi Komputer.

Salah satu tujuan utama sistem MDM adalah mempublikasikan kumpulan data master yang terintegrasi, akurat, dan konsisten untuk digunakan oleh pemangku kepentingan. Dengan tata kelola yang tepat, data master dapat dianggap sebagai kumpulan data terpadu yang dapat diandalkan oleh semua aplikasi untuk mendapatkan informasi yang konsisten dan berkualitas tinggi. MDM yang sukses bergantung pada hal-hal berikut:

- Metode untuk identifikasi data master
- Definisi data terpadu di seluruh unit bisnis
- Teknologi integrasi berkualitas tinggi
- Kerangka tata kelola untuk mengelola integrasi data perusahaan yang berkelanjutan ke dalam lingkungan data master

Fitur utama Data /master pada umumnya adalah Stabilitas, Kompleksitas, dapat digunakan berulang-ulang, mempunyai nilai tinggi bagi organisasi, kemandirian, dan perilaku. Fitur Utama tersebut adalah: (Aholo, 2022a)

- **Stability** - Data master cukup stabil karena tidak sering berubah dan volumenya agak konstan dibandingkan data transaksional.
- **Complexity** - Biasanya, data master ada dan digunakan di lebih dari satu area bisnis dalam suatu organisasi. Data master yang sama dapat disimpan di beberapa sistem sekaligus, misalnya di sistem penjualan dan sistem penagihan. Data master mungkin berbeda dalam organisasi antar unit bisnis yang berbeda.
- **Reusability** - Biasanya, data master ada dan digunakan di lebih dari satu area bisnis dalam suatu organisasi. Data master biasanya digunakan kembali .

- **High value for organization** - Objek data master dapat dilihat sebagai hal yang paling penting bagi suatu organisasi. Data master menggambarkan entitas inti perusahaan.
- **Independence** - Data master berbeda dari tipe data lainnya berdasarkan independensinya. Data master bisa ada dengan sendirinya, tetapi misalnya, data transaksional tidak bisa ada karena bergantung pada data master.
- **Behavior** - Data transaksi bergantung pada data master. Data master memungkinkan pelaporan bisnis dengan menyediakan sumber data dimensi

Data master adalah data paling penting bagi organisasi karena menggambarkan objek bisnis utama yang digunakan dalam berbagai proses bisnis, dan transaksi bisnis bergantung pada data master. Data pelanggan adalah salah satu objek data master yang khas. Data master biasanya cukup kompleks karena digunakan di banyak area bisnis dalam organisasi, dan biasanya disimpan di beberapa sistem informasi pada waktu yang bersamaan. Karena digunakan dan disimpan di tempat berbeda dalam organisasi, data mungkin berbeda antar unit bisnis, dan kualitas data master mungkin dipengaruhi oleh sistem yang berbeda dan peningkatan jumlah pengguna yang menangani data. MDM mencoba mengatasi tantangan data master ini. Manajemen data master adalah aktivitas manajemen seluruh organisasi yang memungkinkan organisasi mencapai satu versi kebenaran dalam objek data master. Organisasi, proses, dan sistem yang berbeda telah menciptakan informasi yang terpisah.

Tujuan MDM adalah menyatukan informasi yang berbeda dan terpisah serta memastikan informasi tersebut mudah diakses dan tersedia di seluruh organisasi bila diperlukan. MDM terdiri dari serangkaian praktik manajemen data terbaik. Perusahaan ini mencoba mengatasi masalah data master dengan mengelola proses bisnis dan kualitas data serta menstandarisasi operasi dan menciptakan integrasi antara sistem informasi yang terpisah.

Tata kelola data master adalah salah satu praktik manajemen penting di MDM. Tata kelola data master terdiri dari aturan, kebijakan, tanggung jawab, dan kepemilikan data master. Tujuan tata kelola data adalah untuk memastikan bahwa data telah dikelola dengan benar.

#### **Customer Data Master (CDM)**

Pelanggan merupakan salah satu objek master data. Data master pelanggan misalnya dapat mencakup nama perusahaan, nomor pajak, alamat, kontak, akun, dan hierarki perusahaan. Karena pelanggan adalah salah satu objek bisnis yang paling penting, pelanggan dapat ditangani dalam berbagai operasi berbeda dalam suatu organisasi, seperti dalam operasi pemasaran, penjualan, dan penagihan. Semua operasi yang

berbeda ini mungkin memiliki kebutuhan data pelanggan yang berbeda, dan menyeimbangkan kebutuhan ini adalah hal yang penting. Agar tidak memiliki data master pelanggan dalam beberapa sistem berbeda dengan detail berbeda, semua operasi bisnis harus menyetujui semua entitas dan cara menyimpan data sehingga akurat dan tersedia untuk semua operasi. Salah satu tantangan utama saat mengelola data pelanggan adalah mengumpulkan data yang relevan dan akurat. Selain itu, keakuratan data perlu dipertahankan di semua sistem yang mencakup data tersebut, meskipun ada perubahan. Permasalahan utama lainnya adalah definisi proses pengumpulan dan pemeliharaan data yang tidak jelas dan tidak jelas, serta kurangnya kejelasan dalam kepemilikan data. Selain itu, pemeliharaan data merupakan tantangan karena jumlah data yang kompleks dan terus meningkat. Tantangan terbesar untuk mengintegrasikan data dari berbagai sumber ke dalam satu data induk adalah tantangan organisasional, bukan teknologi. Dengan CDM yang tepat, perusahaan dapat mencapai "single version of the truth" tentang pelanggan mereka dengan mengumpulkan dan memelihara data pelanggan yang jelas, akurat dan terpusat.

### **Deduplication**

Metode deduplikasi digunakan dalam proyek yang direalisasikan melalui penelitian dan industri. Deduplikasi data merupakan tantangan tersendiri kompleksitas komputasi dan kompleksitas penemuannya metode yang paling memadai untuk membandingkan record dan komputasi kesamaan record- record ini. Nilai kesamaan dua record merupakan nilai majemuk yang perhitungannya didasarkan pada kesamaan nilai atribut individu.

Deduplikasi data juga dikenal sebagai resolusi entitas. Dalam penelitian literatur, jalur deduplikasi data dasar telah diusulkan. Ini telah menjadi saluran referensi dalam beberapa proyek deduplikasi data. (Andrzejewski, Bębel, Boiński, Sienkiewicz, & Wrembel, 2023). Pipeline mencakup empat tugas dasar, yaitu:

- 1) Pemblokiran (alias pengindeksan) - ini membentuk record ke dalam kelompok-kelompok, sedemikian rupa sehingga setiap kelompok menyertakan record yang mungkin berpotensi duplikat.
- 2) Pemrosesan blok (alias pemfilteran) - tujuannya adalah untuk menghilangkan record yang tidak perlu dibandingkan.
- 3) Pencocokan entitas (alias komputasi kesamaan) - ini com-nilai kesamaan putes antar record dibandingkan secara berpasangan, yaitu, nilai setiap atribut dalam satu record dibandingkan dengan nilai atribut terkait di rekaman kedua.
- 4) Pengelompokan entitas – bertujuan untuk menciptakan kelompok-kelompok record yang

serupa, dari pasangan record yang mewakili kemungkinan besar duplikat.

Menurut (Boiński et al., 2022) proses Data Deduplikasi juga bisa dilakukan dengan menggunakan Pipeline :

- 1) Memilih pengelompokan atribut,
- 2) Memilih atribut yang digunakan untuk membandingkan pasangan record,
- 3) Memilih metode untuk membandingkan record,
- 4) Memilih ukuran kesamaan untuk membandingkan nilai pasangan atribut,
- 5) Menentukan bobot atribut untuk menghitung kesamaan record dan memilih ambang kesamaan, Membangun pasangan rekaman yang memiliki nilai kesamaan tinggi,
- 6) Membangun klik dari record serupa,
- 7) Selanjutnya menggabungkan klik dari record serupa.

### **3. METODE**

Metode penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu. Maksud dari cara ilmiah adalah bahwa kegiatan penelitian bersandar pada ciri-ciri keilmuan, yakni rasional, sistematis dan empiris. Dalam penelitian ini penulis menggunakan metode penelitian *Literature Review* (Tinjauan literatur).

*Literature Review* adalah fitur penting dari penelitian akademis, karena pengembangan ilmu pengetahuan harus dibangun di atas penelitian yang sudah ada sebelumnya. Dengan meninjau literatur yang relevan, kami memahami luas dan dalamnya bidang ilmu yang ada dan mengidentifikasi celah untuk dijelajahi. Dengan meringkas, menganalisis, dan mensintesis sekelompok literatur terkait, kita dapat menguji hipotesis tertentu untuk mengembangkan teori baru (Xiao, Watson, & research, 2019)

Dalam melakukan penelitian ini digunakan metodologi tinjauan literatur dari studi pustaka dengan tahapan dan detail pelaksanaan tersebut adalah sebagai berikut:

#### **3.1. Menentukan Tujuan Penelitian**

Tujuan utama penelitian ini adalah untuk membuat rancangan sistem dan arsitektur untuk membangun mekanisme untuk mengelola Customer Data Master yang dikumpulkan dari berbagai tabel Customer yang tersebar di banyak tabel, desain juga akan dilengkapi dengan proses *Deduplication*, *Data Enrichment*, *Data Cleansing* dan *Data House Keeping*

#### **3.2. Menentukan Pertanyaan Penelitian dan Metode Penelitian**

Mengidentifikasi masalah yang ada atau peluang bisnis yang akan didapat dari hasil penelitian .

#### **3.3. Mencari Jurnal yang Relevan**

Pada tahap ini terbagi menjadi beberapa kegiatan pertama-tama menentukan di

perpustakaan digital mana akan digunakan sebagai sumber pencarian, menentukan kalimat kunci pencarian, melakukan pencarian awal, menyempurnakan kalimat kunci pencarian dan mengambil daftar awal studi utama dari perpustakaan digital yang sesuai dengan kalimat kunci pencarian. Untuk mendapatkan serangkaian studi seluas mungkin pencarian akan dilakukan pada basis data literatur paling populer di bidang ini. Berikut adalah perpustakaan digital yang digunakan pada penelitian ini: Google Scholar, e-Resource Perpustakaan nasional, ProQuest, dan Springer Nature. Pencarian jurnal akan difokuskan pada jurnal-jurnal yang berisi artikel penelitian paling relevan, dengan kata kunci "Data Quality", "Data Cleansing", "Data Deduplication", "Master Data Management" dan "Customer Data Integration".

### 3.4. Studi Pustaka

Mempelajari tentang Master Data Management, problem yang umum pada data master dan cara cara mengatasinya, mempelajari tentang berbagai cara dalam melakukan Data Cleansing, mempelajari mengenai mekanisme pengelolaan Master Data Management khususnya Customer Master Data. Mempelajari bagaimana cara membentuk Customer Id yang universal namun bersifat unik. Mengumpulkan berbagai teknik Data Deduplication dari berbagai jurnal. Pengutipan referensi yang relevan dengan pokok masalah dan yang relevan dengan solusi yang akan digunakan.

#### Perancangan Solusi

Akan dibuat desain database Customer Master, dilanjutkan desain sistem Data Deduplication, Process Data Flow mulai data datang dari mitra, pengecekan duplikasi data secara sistem, proses penambahan identitas global (CIF Id), mekanisme konfirmasi data deduplikasi secara manual.

### 3.5. Implementasi Solusi

Selain sistem ini dapat digunakan untuk PT. ABC, Solusi ini juga dapat diimplementasikan pada hypermart atau toko waralaba. Terutama pada organisasi yang data nasabahnya tersebar di berbagai aplikasi, atau organisasi yang datanya berasal dari beberapa organisasi lain seperti PT. ABC.

### 3.6. Hasil dan Pembahasan

Desain Sistem dan arsitektur yang terbentuk meliputi Pipeline Data Deduplication, Teknik Similarity Detection baik by sistem atau manual dan proses Data Cleansing berkelanjutan untuk mengkonfirmasi record duplicate yang diidentifikasi oleh sistem. Detailnya akan diuraikan pada tulisan ini.

## 4. HASIL DAN PEMBAHASAN

MDM mendukung organisasi yang mengintegrasikan dan berbagi data secara akurat, tepat waktu, dan konsisten dengan menerapkannya kebijakan, layanan, dan infrastruktur. Ini memandu organisasi dalam membuat data Master. Teknologi adalah bagian penting dari MDM. Namun, proses organisasi adalah bagian terpenting dari MDM, bukan hanya teknologi dan sistem informasi. Berikut adalah langkah-langkah untuk keberhasilan proses MDM. (Hikmawati, Santosa, & Hidayah, 2021)

- 1) Identifikasi kebutuhan organisasi.
- 2) Identifikasi data Master organisasi dan sistem yang digunakan.
- 3) Mendefinisikan tata kelola MDM.
- 4) Definisikan proses pemeliharaan.
- 5) Menentukan standar data.
- 6) Menentukan perbaikan kualitas data saat ini di masa depan.
- 7) Perencanaan arsitektur MDM (aplikasi, aliran data, keamanan data, dan masalah privasi data).
- 8) Pelatihan dan komunikasi bagi seluruh pemangku kepentingan.
- 9) Roadmap/strategi pengembangan MDM.
- 10) Menentukan karakteristik MDM (antarmuka pengguna, alur kerja, pengeditan data, rekonsiliasi, integrasi)

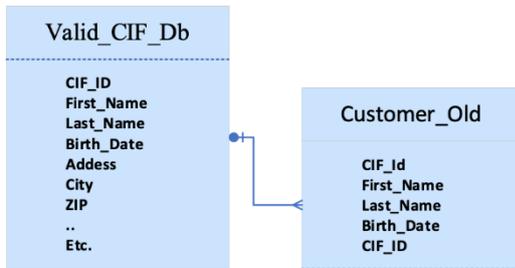
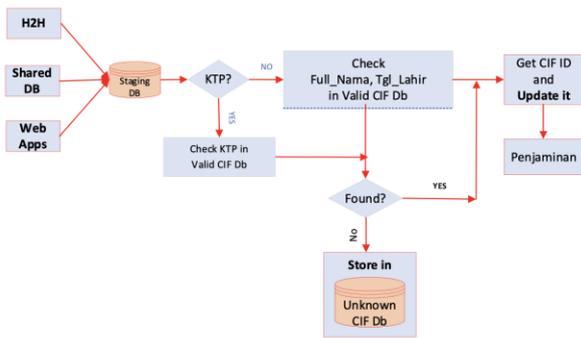
Ketika Data master dikelola secara efisien, siklus hidup data master dimulai sebelum pembuatan data dengan merencanakan, menentukan, dan mengaktifkan pengambilan data. Gambar 1 merangkum semua langkah siklus hidup data master. (Ahola, 2022b).



Gambar-02 Data Master Life Cycle Development

Bank-bank sebagai partner bisnis PT ABC mengirimkan data Transaksi pengajuan Kredit beserta data Nasabahnya. Metode pengiriman data ini setiap Bank menggunakan cara yang berbeda beda demikian juga struktur datanya. Beberapa Bank menggunakan cara Host-to-Host, ada yang menggunakan Share Database, dan ada juga yang menggunakan aplikasi Web.

Data yang sudah diterima ini kemudian disimpan pada database Staging, yang berisi beberapa tabel Penampungan. Data dari masing-masing Bank disimpan pada tabel Penampungan yang berbeda-beda karena struktur data dari Bank tidak ada yang sama.



Gambar-03 Process Flow Diagram

Pada tabel Penampungan ini akan ada beberapa Validasi terkait keberadaan beberapa field-field yang mandatory, Validasi format tanggal. Validasi nomor KTP, nomor NPWP, dan sebagainya. Pada tahap ini semua data nasabah yang masuk akan mendapatkan nomor urut (running number Id) sebagai nasabah Id dan Transaksi Id. Jika data sudah lolos dari semua Validasi maka sistem akan memberikan respon yang menunjukkan bahwa data telah diterima sukses.

KTP adalah satu-satunya kode unik yang dimiliki setiap nasabah, namun sayangnya tidak

Gambar-04-Matching Proses

semua Bank memberikan data nomor KTP ini pada PT.ABC, sehingga harus dicari cara untuk membuat nomor Unik pada setiap nasabah. Oleh karenanya langkah pertama dari sistem ini adalah melakukan pengecekan apakah ada data KTP apa tidak. Pada sistem ini data Nasabah yang diproses disimpan pada dua tabel yaitu tabel Valid\_CIF\_Db yang menyimpan data nasabah yang sudah dibuatkan nomor CIF dan tabel Unknown\_CIF\_Db yang berisi data nasabah yang baru datang.

Jika ada data nomor KTP maka dilakukan pengecekan apakah nasabah dengan nomor KTP tersebut sudah tersimpan di tabel Valid\_CIF\_Db, jika ditemukan maka ambil nomor CIF nya dan tambahkan pada data pengajuan penjaminan. Sebaliknya jika tidak ditemukan simpan data nasabah ke tabel Valid\_CIF\_Db.

Jika tidak ada nomor KTP, lakukan pengecekan dengan Nama Lengkap dan Tanggal Lahir ke tabel Valid\_CIF\_Db, jika ditemukan maka ambil nomor CIF nya dan tambahkan pada data pengajuan penjaminan. Sebaliknya jika tidak ditemukan simpan data nasabah ke tabel

Valid\_CIF\_Db. Tugas berikutnya adalah verifikasi secara manual pada data-data nasabah yang tersimpan pada tabel Unknown\_CIF\_Db. Pada tahap ini terdiri dari beberapa sub-proses yaitu:

- *Matching Process* - Proses menentukan apakah data nasabah yang baru ada kesamaan dengan data yang sudah diberikan nomor CIF di tabel Valid\_CIF\_Db.
- *Merging Process* - Proses menggabungkan dua record data nasabah menjadi satu, karena sesungguhnya ke dua record merefer ke orang yang sama.

### Matching Process

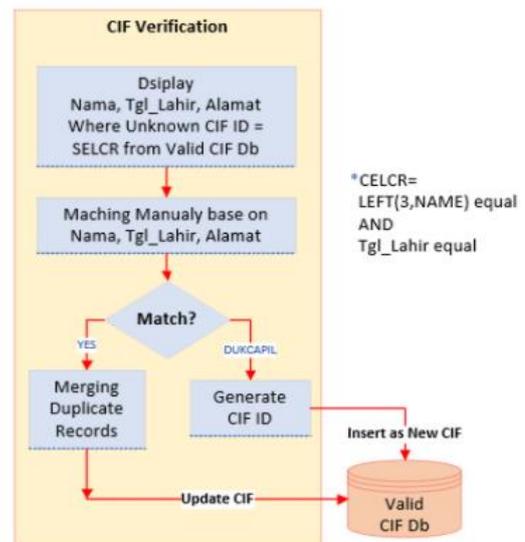
Proses verifikasi dimulai dengan menampilkan data semua record yang mirip dengan data nasabah baru, hal ini dilakukan dengan query ke tabel Valid\_CIF\_Db dengan kriteria 3 huruf pertama dari nama depan dan tanggal lahir yang tepat sama. Jika ada record yang sesuai dengan kriteria ini akan ditampilkan data Nama Lengkap, Tanggal lahir dan alamat.

Gambar 05-Tabel Customer Master

Secara visual diamati dan dianalisa secara manual apakah ada record yang berindikasi adalah data dari orang yang sama dengan data nasabah baru. Jika ada maka akan dilakukan proses Merge dengan salah satu record yang ditampilkan, jika tidak ada satupun kemiripan maka dapat disimpulkan bahwa data nasabah baru adalah orang baru yang belum ada di database dan akan di buatkan nomor CIF ID yang baru.

### Generate New CIF ID

Proses membentuk nomor CIF mengikuti aturan bahwa nomor tersebut harus bersifat unik tidak ada yang sama dan berlaku di level korporat apapun bisnis prosesnya. Formula untuk



membentuk nomor CIF adalah sebagai berikut:  
**CIF\_ID = YYYYMMDDHHMM999999**

- *Time-stamp akan digunakan urutan Tahun, Bulan, Tanggal, Jam, Menit agar bisa urutan secara kronologis*
- *Index Number (999999) sebanyak 6 digit artinya bisa mengantisipasi jika ada duplikat sebanyak 1 juta varian per menit.*
- *Setiap Index number mencapai maksimal yaitu 999999 maka akan diset secara otomatis kembali menjadi 000001*

### Merging Process

Dua atau lebih record data nasabah yang secara manual di pastikan adalah orang yang sama maka akan dijadikan menjadi satu record dengan nomor CIF ID yang sama, namun dalam proses penggabungan ini tidak boleh ada informasi yang hilang karena kedua record dipastikan masing masing mempunyai sejarah transaksi sendiri-sendiri. Untuk mengatasi masalah tersebut akan digunakan dua tabel yang terkait sebagai berikut, yaitu tabel Valid\_CIF\_Db dan tabel Customer\_Old, dengan relasi sebagai berikut:

Tabel Valid\_CIF\_Db adalah tabel Customer yang sudah clean dan tidak ada duplikasi dan sudah ditambahkan nomor CIF ID, sedangkan tabel Customer\_Old adalah tabel yang menyimpan nomor CIF\_Id lama. Setiap record pada tabel Valid\_CIF\_Db dapat mempunyai relasi ke beberapa record tabel Customer\_Old sebanyak record duplikat yang telah di verifikasi secara manual. Jika data customer tidak ada yang duplikat maka tabel pada record Valid\_CIF\_Db tidak ada relasi ke tabel Customer\_Old. Dengan cara demikian maka semua nomor CIF\_Id lama yang diidentifikasi adalah orang yang sama tetap disimpan di sistem, sehingga masih bisa digunakan untuk pelaporan atau pelacakan transaksi masa lalu.

## 5. KESIMPULAN DAN SARAN

### Kesimpulan

Dari semua tahapan perancangan sistem ini maka penulis dapat menyimpulkan beberapa hal sebagai berikut:

- 1) Data kotor bisa terjadi disemua orgainisasi, termasuk didalamnya adalah data duplikat.
- 2) Untuk validasi dan koreksi data kotor akan membutuhkan biaya yang tidak sedikit, kerugian paling besar adalah organisasi tidak akan mendapatkan informasi yang tepat dan akurat jika Data diolah menjadi informasi.
- 3) Untuk mengatasi masalah data kotor secara menyeluruh dan berkesinambungan organisasi perlu mengimplemantasikan Good Data Governance.
- 4) Data cleansing adalah proses yang harus dilakukan secara berkelanjutan karena setiap

hari akan ada data data baru yang dihasilkan dari semua proses bisnis.

### Saran

Sistem ini masih bisa ditingkatkan kualitas dan performa dengan memanfaatkan **Machine Learning**, yaitu pada tahap *Matching process* yang saat ini masih dilakukan secara manual.

## DAFTAR PUSTAKA

- Ahola, S. (2022a). Developing A Customer Master Data Management Model.
- Andrzejewski, W., Bębel, B., Boiński, P., Sienkiewicz, M., & Wrembel, R. (2023). Text similarity measures in a data deduplication pipeline for customers records.
- Azeroual, O., & Abuosba, M. (2019). Improving the data quality in the research information systems. *arXiv preprint arXiv:1901.07388*.
- Boiński, P., Sienkiewicz, M., Bębel, B., Wrembel, R., Gałczowski, D., & Graniszewski, W. (2022). *On customer data deduplication: Lessons learned from a r&d project in the financial sector*. Paper presented at the Proc. of the Workshops of the EDBT/ICDT 2022 Joint Conference (CEUR Workshop Proceedings).
- El-Hadi, M. (2023). How Good Data Quality Achieves Better Results and Boosts Productivity. *31(31)*, 79-80.
- Hikmawati, S., Santosa, P. I., & Hidayah, I. J. I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. *5(3)*, 90-95.
- Jasim, H. A., Fahad, A. A. J. I. J. o. A. C. S., & Applications. (2018). New techniques to enhance data deduplication using content based-TTDD chunking algorithm. *9(5)*.
- Sehat, M., & Paves Flores, R. (2012). Customer Data Management. In.
- Traulsen, S., Tröbs, M., AG, U., Tucherpark, A., & Ganghoferstrasse, C. (2011). *Implementing data governance within a financial institution*. Paper presented at the GI-Jahrestagung.
- Viji, D., & Revathy, S. (2019). *Various data deduplication techniques of primary storage*. Paper presented at the 2019 International conference on communication and electronics systems (ICCES).
- Xiao, Y., Watson, M. J. J. o. p. e., & research. (2019). Guidance on conducting a systematic literature review. *39(1)*, 93-112.