

Proposal Thesis

**Pembangkitan *Valid Customer Information Files* (CIF)
Menggunakan Map-Reduce Fuzzy Matching pada E-CIF Bank Danamon**

Hendrik

G651160051



Dosen:

Dr Agus Buono MKom, MSi

Dr Yani Nurhadryani, SSi, MT

**Sekolah Pascasarjana
Institut Pertanian Bogor**

2017

Abstract

Managing information in an enterprise typically involves integrating data from across the enterprise and beyond, cleansing the data, matching the data to remove any duplicates, standardizing the data, enriching the data, making the data conform to legal and compliance requirements, and then storing the data in a centralized location with all the necessary security settings.

Bank Danamon Indonesia (BDI) as enterprise in banking has requirement to develop Information management in Customer Information Files (CIF). It's called as E-CIF system. E-CIF is developed to ensure the costumer information quality for the next business strategy and to support the business process runs effective and efficient.

1. PENDAHULUAN

Latar Belakang

Customer relationship management (CRM) tengah dikembangkan oleh Bank Danamon Indonesia. CRM dikembangkan agar BDI dapat menjaga hubungan dengan nasabah dengan memahami latar belakang nasabah dan perilakunya, meningkatkan layanan kepada nasabah, dan mengembangkan inovasi produk melalui implementasi layanan interaktif dan peningkatan teknologi pada basis data nasabah (7). CRM membutuhkan data nasabah yang valid dan esensial untuk memberikan hasil analisis yang akurat dan efektif sehingga memberikan layanan yang optimal untuk nasabah (BI, 2015).

BDI menyimpan informasi mengenai data nasabah pada *customer information files* (CIF). CIF merupakan sebuah fail, baik elektronik atau fisik, yang menyimpan semua informasi yang bersangkutan mengenai informasi pribadi dan rekening nasabah (8). CIF pada kasus kali ini merupakan fail elektronik yang tersimpan pada basis data. CIF pada BDI terdiri dari CIF nasabah individu dan CIF nasabah perusahaan. Sebagai contoh, CIF individu memuat informasi mengenai nama, alamat, tanggal lahir, tempat lahir dan ibu kandung. Sedangkan, CIF Perusahaan memuat nama, alamat, pemilik, dan NPWP perusahaan.

Dalam pengembangan CRM, BDI menghadapi tantangan berupa CIF yang terpisah di beberapa sumber data, yaitu Core Banking, Ascend, dan Adira. Selain itu, ukuran basis data yang besar (*Big Data*), yaitu mencapai 200 GB dan bertambah dengan cepat seiring dengan aktivitas perbankan. Kemudian, adanya duplikasi CIF di antara tiga sumber data dan CIF yang mengandung nilai *fuzzy*, menyebabkan BDI sulit untuk melakukan identifikasi CIF yang valid dan esensial. Duplikasi data akan mempengaruhi hasil ekstraksi informasi, mengacaukan hasil analisis, dan membuat peluang kesalahan pada pengambilan keputusan lebih tinggi (9).

Pencocokan data diterapkan untuk menjamin kualitas data, menghapus duplikasi dan inkonsistensi pada sumber data, dan mengurangi peluang kesalahan pada pengambilan keputusan (10). Penelitian mengenai pencocokan data yang memiliki nilai *fuzzy* dan duplikasi telah banyak dilakukan. Teknik pembersihan yang dikembangkan banyak mengadopsi atau peningkatan dari algoritma *fuzzy clustering*. Salah satu algoritma *fuzzy clustering* yang banyak digunakan adalah algoritma Fuzzy C-Means (FCM). Algoritma FCM dapat membagi sebuah himpunan terbatas dari n elemen ke dalam sebuah himpunan dari c kelompok *fuzzy* dengan memberikan beberapa kriteria.

Guo *et al* (2012) mengadopsi teknik *fuzzy c-means clustering* dikombinasikan dengan jarak Levenshtein untuk melakukan pembersihan duplikasi data. Teknik tersebut secara akurat dan cepat dapat mendeteksi dan menghilangkan duplikasi data. *Recall* dan *precision* deteksi duplikasi data lebih baik dibanding metode pembersihan data yang lain.

Di samping itu, untuk menangani Big Data, Prabha et al (2014) berhasil melakukan reduksi data menggunakan teknik peningkatan dari Fuzzy Clustering, yaitu *Incremental Weighted Fuzzy C-Means* (IWFCM). IWFCM memasukkan bobot sebagai parameter yang menggambarkan pengaruh masing-masing obyek di dalam *cluster*. IWFCM diterapkan pada *e-book dataset*. Dataset tersebut diolah menggunakan lingkungan Hadoop dengan *map reduce framework*. IWFCM berhasil menghasilkan *cluster* dengan minimum *run time* dan kualitas yang baik.

Enterprise Customer Information Files (E-CIF) dikembangkan oleh BDI untuk menangani permasalahan duplikasi data tersebut. E-CIF merupakan sistem yang mengkonsolidasikan dan menggabungkan data nasabah di berbagai sumber data. E-CIF bertujuan untuk menciptakan satu data nasabah yang valid dan mencakup keseluruhan informasi nasabah dan relasi yang terkait dengan nasabah tersebut. Istilah satu data nasabah valid disebut *single customer information files* (CIF). Saat ini, BDI belum memiliki *single CIF* yang valid.

E-CIF sendiri merupakan pengembangan dari aplikasi Master Customer Information Files (M-CIF) yang saat ini sudah tersedia di Bank Danamon Indonesia. M-CIF dalam menentukan *single CIF* menggunakan metode *matching* atau pencocokan. Metode pencocokan yang sudah digunakan masih belum sesuai dengan kebutuhan BDI. CIF yang dihasilkan tidak sesuai dengan kebutuhan BDI, yaitu masih banyaknya duplikasi CIF.

BDI membutuhkan mekanisme pembangkitan *single CIF* baru pada E-CIF. Mekanisme baru ini dapat melakukan integrasi data, pembersihan data, pengelompokan data, pencocokan data, verifikasi data dan koreksi data. Mekanisme pembangkitan tersebut nantinya juga dapat menangani data yang memiliki nilai *fuzzy*. Data yang berkualitas yang diperlukan adalah data *single CIF* yang tidak memiliki duplikasi serta isi dari setiap data merupakan isian yang *valid*.

Mekanisme pembangkitan data berkualitas telah dikembangkan oleh Microsoft melalui *Enterprise Information Management (EIM)* pada Microsoft SQL Server 2012. EIM merupakan sebuah platform untuk melakukan pembangkitan data berkualitas. EIM memiliki tiga komponen *services* utama, yaitu *Master Data Service (MDS)*, *Integration Service (IS)* dan *Data Quality Services (DQS)*. EIM merupakan *platform* yang telah banyak digunakan oleh perusahaan skala *enterprise* untuk pengelolaan datanya.

EIM memberikan beberapa opsi operasi untuk melakukan pembangkitan data yang berkualitas. Operasi untuk melakukan manajemen data berbasis pengetahuan, pembersihan data, dan pencocokan data dapat dilakukan pada *data quality service (DQS)*. Selain itu, pada komponen *integration service (IS)*, dapat dilakukan pembersihan data menggunakan *fuzzy lookup*, dan *lookup fuzzy transformation* dan operasi tersebut dapat dilakukan secara *batch* tanpa pengawasan. Sedangkan, MDS memiliki kemampuan untuk melakukan pembuatan model dan basis pengetahuan.

Algoritme pengelompokan, pencocokan dan pembersihan data pada EIM berdasarkan pada prinsip Fuzzy. Sebagai contoh, pengelompokan data menggunakan Fuzzy grouping. Algoritme Fuzzy grouping telah digunakan Microsoft untuk operasi *geocode query* pada Bing Maps (2011). Disamping itu, operasi pencocokan menggunakan Fuzzy Matching. Sedangkan pembersihan data menggunakan Fuzzy Lookup. Fuzzy Lookup dikembangkan dan digunakan oleh Microsoft pada tahun 2012, metode ini digunakan untuk manajemen proyek master data internal untuk mencocokkan pelanggan baru dan pelanggan yang sudah ada.

Mekanisme EIM pada Microsoft SQL Server 2012 ini akan diterapkan dan dimodifikasi pada lingkungan *open source* Apache Spark dan Apache Hadoop untuk pembangkitan CIF data nasabah BDI. Saat ini Bank kompetitor sudah mengimplementasikan pembangkitan *single* CIF yang baik. Dengan demikian, BDI perlu mengembangkan E-CIF agar dapat bersaing dengan kompetitor yang sudah memiliki *single* CIF.

Manfaat Penelitian

ECI-F dapat mengidentifikasi dan mengombinasikan data maupun relasi nasabah sehingga menyediakan *single CIF* yang valid untuk aplikasi pada CRM di Bank Danamon atau pihak lain yang terkait (contoh: Manulife). Dengan demikian, CRM dapat memberikan analisis yang akurat sehingga membantu pertumbuhan bisnis Bank Danamon Indonesia secara signifikan.

Tujuan Penelitian

Tujuan dari penelitian adalah membuat sebuah mekanisme penyatuan, pembersihan, pengelompokan, dan pencocokan berbagai data nasabah menggunakan *Modified Enterprise Information Management (EIM)* untuk menghasilkan *single* CIF.

Ruang Lingkup Penelitian

E-CIF dikembangkan dengan menggunakan data nasabah Bank Danamon Indonesia dan digunakan untuk Bank Danamon Indonesia. E-CIF dibangun pada lingkungan teknologi Microsoft .NET dan SQL Server Enterprise 2014. Sumber data nasabah berasal dari tiga sumber data, yaitu Core Banking, Ascend, dan Adira. Data yang digunakan adalah data nasabah individual.

Tinjaun Pustaka

Enterprise Information Management (EIM)

Enterprise Information Management (EIM) merupakan sebuah manajemen pembersihan, pencocokan dan penyatuan berbagai sumber data untuk menghasilkan suatu data yang berkualitas tinggi. Data berkualitas tinggi yang dimaksudkan adalah data yang bernilai benar, paling baru, dan tidak memiliki duplikasi. EIM terdapat pada Microsoft SQL Server 2012. EIM memiliki tiga komponen utama yaitu, *Integration Service (IS)*, *Master Data Service (MDS)* dan *Data Quality Service (DQS)*.

Master Data Services (MDS) adalah sebuah platform manajemen data master yang memungkinkan pembuatan dan pemeliharaan, standar tunggal untuk sebuah data bisnis. MDS dapat memelihara dan mengelola data master dalam kondisi "known good" melalui sebuah aturan dan UI berbasis Excel. Menurut definisi, data master adalah non-transaksi referensi data yang berubah perlahan. MDS dapat menyinkronisasi dan menyelesaikan perbedaan antara berbagai sumber data master.

Data Quality Services (DQS) merupakan platform untuk membersihkan, memperkaya, dan pencocokan data. DQS bekerja untuk meningkatkan kualitas data dengan cara melacak aturan dan pengetahuan lainnya untuk memperbaiki permasalahan data "kotor", dan untuk menerapkan pengetahuan dalam skenario interaktif dan otomatis. DQS adalah sistem berbasis pengetahuan yang melakukan proses pembersihan secara interaktif. Pada DQS dapat dibangun basis pengetahuan dengan menambahkan pengetahuan yang telah dibuat, melakukan penemuan pengetahuan, dan menambahkan aturan yang sesuai, yang memanfaatkan algoritma pada DQS dan data referensi yang dibangun ke dalam setiap basis pengetahuan. DQS juga dapat membangun pengetahuan yang diperoleh dalam operasi pembersihan yang sedang berlangsung ke dalam basis pengetahuan, untuk mencapai dan mempertahankan standar kualitas data yang tinggi.

Integration Services (IS) adalah platform untuk membangun integrasi data pada tingkat perusahaan dan solusi transformasi data. IS dikembangkan untuk memecahkan masalah bisnis yang kompleks dengan menyalin atau mengunduh fail, mengirim pesan e-mail dalam menanggapi peristiwa, memperbarui *data ware-house*, pembersihan dan *data mining*, dan mengelola SQL Server objek dan data. Paket dapat bekerja sendiri atau dengan paket lain untuk mengatasi kebutuhan bisnis yang kompleks. IS dapat mengekstrak dan mengubah data dari berbagai sumber seperti file XML data, file datar, dan sumber data relasional, dan kemudian memuat data ke dalam satu atau lebih tujuan.

Big Data

“Big Data” adalah sebuah istilah untuk merepresentasikan sebuah koleksi data yang berukuran besar, berstruktur kompleks dan mengalir dengan kecepatan tinggi yang mana tidak

bisa dikelola menggunakan teknologi *traditional data-warehouse*. Big Data adalah dunia data yang berada di luar tradisional *data warehouse* dan *enterprise*. Data tersebut dihasilkan oleh aktivitas perangkat, berita pada blog dan sosial media, data sensor, dan transaksi perdagangan. Big Data berbentuk tidak terstruktur, tidak terfilter dan berbentuk *non-relational*.

Menurut Gartner, “Big Data adalah aset data yang memiliki volume tinggi, kecepatan tinggi, dan keragaman tinggi yang membutuhkan metode pemrosesan baru untuk memungkinkan pembuatan keputusan, penemuan pengetahuan, dan optimasi proses. Selain itu, menurut Steve Stood, Big data adalah ketika aplikasi normal dari teknologi yang sedang digunakan tidak mampu membuat pengguna mendapatkan kebutuhan pengetahuan dari data yang dengan cepat dan biaya yang murah.

Apache Hadoop

Apache Hadoop adalah kerangka perangkat lunak *open-source* yang digunakan untuk penyimpanan terdistribusi dan pemrosesan *dataset* data besar dengan menggunakan model pemrograman MapReduce. Ini terdiri dari kumpulan komputer yang dibangun dari perangkat keras. Semua modul di Hadoop dirancang dengan asumsi mendasar bahwa kegagalan perangkat keras adalah kejadian biasa dan harus ditangani secara otomatis oleh kerangka kerja.

Inti dari Apache Hadoop terdiri dari bagian penyimpanan, yang dikenal sebagai Hadoop Distributed File System (HDFS), dan bagian pemrosesan yang merupakan model pemrograman MapReduce. Hadoop membagi file menjadi blok besar dan mendistribusikannya ke node dalam sebuah cluster. Kemudian transfer kode paket ke dalam node untuk memproses data secara paralel. Pendekatan ini memanfaatkan wilayah data, [3] dimana node memanipulasi data yang mereka akses. Hal ini memungkinkan dataset diproses lebih cepat dan lebih efisien daripada arsitektur superkomputer yang lebih konvensional yang bergantung pada sistem file paralel dimana komputasi dan data didistribusikan melalui jaringan berkecepatan tinggi. [4]

Dasar kerangka Apache Hadoop terdiri dari modul berikut:

1. Hadoop Common - berisi perpustakaan dan utilitas yang dibutuhkan oleh modul Hadoop lainnya,
2. Hadoop Distributed File System (HDFS) - sistem file terdistribusi yang menyimpan data pada mesin komoditas, memberikan bandwidth agregat yang sangat tinggi ke seluruh cluster,
3. Hadoop YARN - platform yang bertanggung jawab untuk mengelola sumber daya komputasi dalam kelompok dan menggunakannya untuk menjadwalkan aplikasi pengguna, dan
4. Hadoop MapReduce - sebuah implementasi dari model pemrograman MapReduce untuk pemrosesan data berskala besar.

Fuzzy Matching

Fuzzy matching adalah sebuah mekanisme pencocokan yang dibantu oleh komputer, untuk mencocokkan kata, frasa, kalimat atau sebagian bagian teks dari kalimat pada sebuah basis data. Algoritma pencocokan fuzzy digunakan untuk membandingkan dua *string* dengan mengukur jumlah karakter yang harus dimodifikasi (menambahkan, menghapus atau mengubah) *source string* agar terlihat seperti *target string*. Sebagai contoh, jika *source string* adalah 'tour' dan *target string* adalah 'tow', algoritma akan mengembalikan *score* 2, karena melakukan dua penggantian karakter.

Fuzzy c-Means Clustering

Fuzzy C-means (FCM) clustering memungkinkan sebuah grup data masuk ke dalam dua atau lebih dari dua grup. Metode ini diusulkan oleh Dunn pada tahun 1973, dan dikembangkan oleh Bezdek pada tahun 1981. Metode ini merupakan pengelompokan didasarkan pada pembagian dan bertujuan agar obyek yang mirip masuk ke dalam kelompok yang sama menjadi maksimum, dan obyek yang berbeda masuk ke kelompok sama menjadi minimum. Hal ini bergantung pada fungsi obyektif yang meminimalkan, dapat dilihat pada gambar di bawah ini:

$$(1) \quad J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|_2$$

di mana m adalah setiap bilangan *real* lebih besar dari 1, m diberikan nilai 2.00. u_{ij} adalah tingkat j dari anggota grup x_i . x adalah pengukuran dimensi i ; C adalah ij dimensi klaster, * pusat adalah untuk mengekspresikan kesamaan antara setiap kriteria, setiap pusat pengukuran. Fungsi tujuan yang partisi kabur dioptimalkan melalui daftar fungsi berulang seperti persamaan (1), memperbarui anggota u_{ij} menggunakan persamaan (2) dan pusat grup C_j diperbaharui menggunakan persamaan (3):

$$(2) \quad u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

$$(3) \quad C = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m}$$

Iterasi berhenti ketika

$$(4) \quad \max_{ij} = \left\{ \left| u_{ij}^{(k-1)} - u_{ij}^k \right| \right\} \varepsilon$$

Fuzzy rule based classification systems (FRCBs) for big data

FRBCSs (Ishibuchi et al. 2004) merupakan perluasan ke sistem berbasis aturan klasik, karena mereka berurusan dengan "IF-THEN" aturan, tapi anteseden dan konsekuen terdiri dari pernyataan logika *fuzzy*, bukan yang klasik. FRBCSs terdiri dari basis pengetahuan (KB), termasuk baik informasi dari *fuzzy* set [(yang terkandung dalam data base (DB)) dan aturan [(dalam peraturan dasar (RB)], dan sistem inferensi.

Dalam karya ini, DB akan diperoleh ad hoc dengan cara label kabur linguistik (fungsi keanggotaan segitiga) homogen sepanjang rentang masing-masing variabel. RB akan diekstrak menggunakan prosedur pembelajaran dari contoh masukan, membangun aturan formulir berikut:

$$(4) \quad \text{Rule } R_j : \text{If } x_1 \text{ is } A_j^1 \text{ and } \dots \text{ and } x_n \text{ is } A_j^n \\ \text{lalu Class} = C_j \text{ with } RW_j$$

di mana R_j adalah label dari aturan- j , $x = (x_1, \dots, x_n)$ adalah vektor pola n -dimensi, A_i adalah himpunan *fuzzy* yg, C_j adalah label kelas, dan RW_j adalah berat aturan (Ishibuchi dan Nakashima, 2001). Secara khusus, kita akan memanfaatkan metode heuristik yang dikenal sebagai Kepastian Factor Penalized (Ishibuchi dan Yamamoto 2005):

$$(5) \quad RW_j = PCF_j = \frac{\sum_{x_p \in C_j} \mu_{A_j}(x_p) - \sum_{x_p \notin C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^m \mu_{A_j}(x_p)}$$

dimana $\mu_{A_j}(x_p)$ adalah derajat keanggotaan x_p , yaitu misalnya p -th dari pelatihan diatur dengan anteseden dari aturan dan C_j adalah kelas ditentukan oleh aturan j .

Akhirnya, mengenai sistem inferensi kita akan menerapkan metode *fuzzy* penalaran dari aturan wining tunggal (Cordón et al. 1999). Metodologi ini memprediksi sebagai kelas *output* yang dari aturan yang memiliki nilai yang cocok tertinggi secara keseluruhan dengan contoh.

Alat dan Data

Pada penelitian ini menggunakan alat berupa perangkat lunak yang terdiri atas *integrated developent environment* (IDE), *database management system* (DBMS), dan penyimpanan Big Data. IDE yang digunakan adalah Apache Spark, penyimpanan dan manajemen data pada Hadoop HDFS. Apache Spark merupakan sebuah *open source tool* yang kaya akan *machine learning libraries* dan dikenal memiliki performance yang baik untuk pemrosesan data pada lingkungan Hadoop. Sedangkan Hive, unit manajemen *relational* pada lingkungan Hadoop.



Data

Data yang digunakan untuk penelitian ini adalah data nasabah dari tiga basis data di Bank Danamon Indonesia. Tiga basis data tersebut adalah Ascend, NCBS, dan Adira. Sedangkan nasabah yang digunakan adalah nasabah individual/personal.

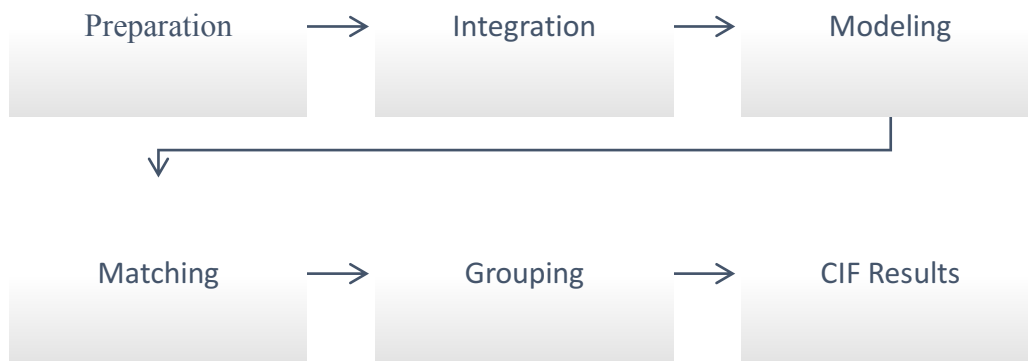
Sampel struktur data nasabah personal yang ada di basis data NCBS dapat dilihat pada Gambar 1. Pada Gambar 1 juga dapat dilihat jenis isian dari setiap kolom yang penting.

No	Personal	Tipe
1	ID	Fuzzy
2	Nama	Fuzzy
3	Tanggal Lahir	Exact
4	Alamat	Fuzzy
5	Ibu Kandung	Fuzzy

Tabel 1. Struktur prioritas data nasabah personal

Metode

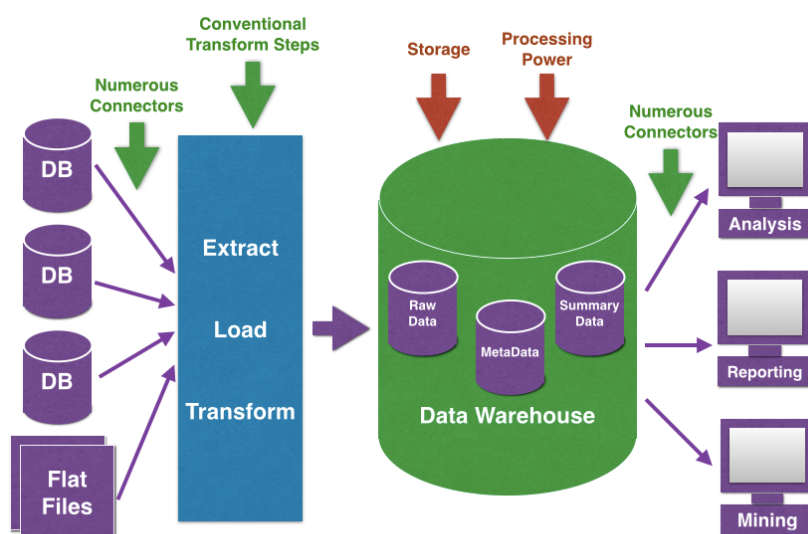
Metode untuk menghasilkan *single* CIF meliputi tahapan penyatuan, pemodelan, pembersihan, pencocokan, dan pengelompokan data disebut *modified enterprise information management (Mod-EIM)*. Skema Mod-EIM dapat dilihat pada Gambar 1.



Gambar 1. Skenario Mod-EIM

Preparation and Integration

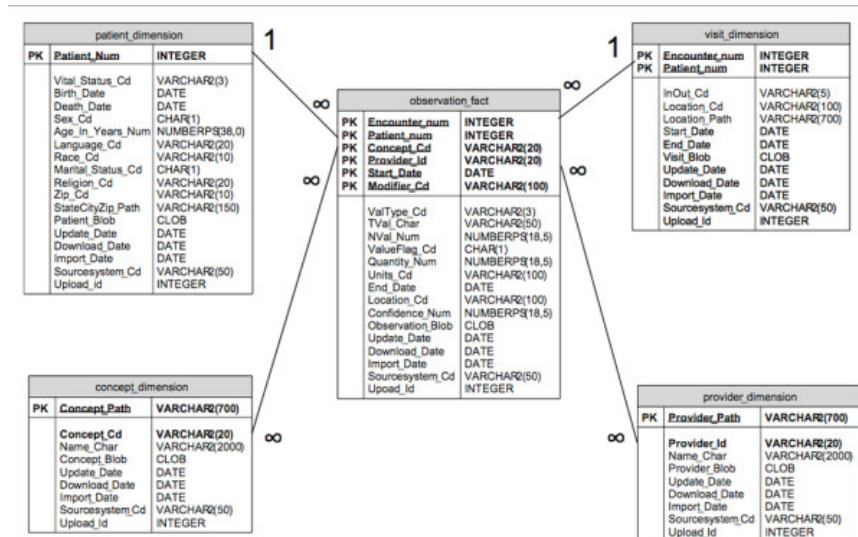
Skenario Mod-EIM dimulai dari persiapan dan integrasi tiga basis data nasabah di BDI, yaitu Ascend, Adira dan NCBS. Tahapan integrasi menggunakan *framework* Hadoop. Hadoop memindahkan data berukuran besar dari berbagai sumber ke tempat sentral (*data warehouse*). Hadoop memiliki keunggulan, yaitu menyimpan data berukuran besar dengan biaya yang relatif murah pada HDFS, dan memungkinkan pengolahan secara paralel dan terdistribusi. Skenario integrasi dapat dilihat pada Gambar 2.



Gambar 2. Skenario Integrasi

Modeling

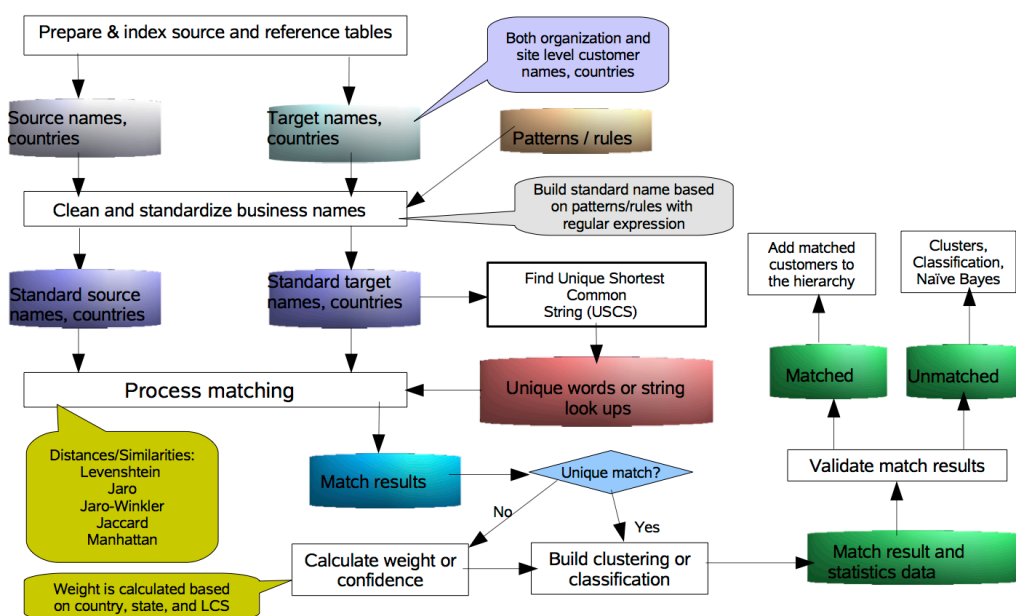
Modeling adalah tahapan pembentukan *schema* data nasabah. *Schema* dibuat untuk mendefinisikan setiap kolom yang digunakan pada data nasabah untuk dijadikan CIF. Skema modeling data nasabah pada Hive dapat dilihat pada Gambar 3.



Gambar 3. Skema modeling

Matching

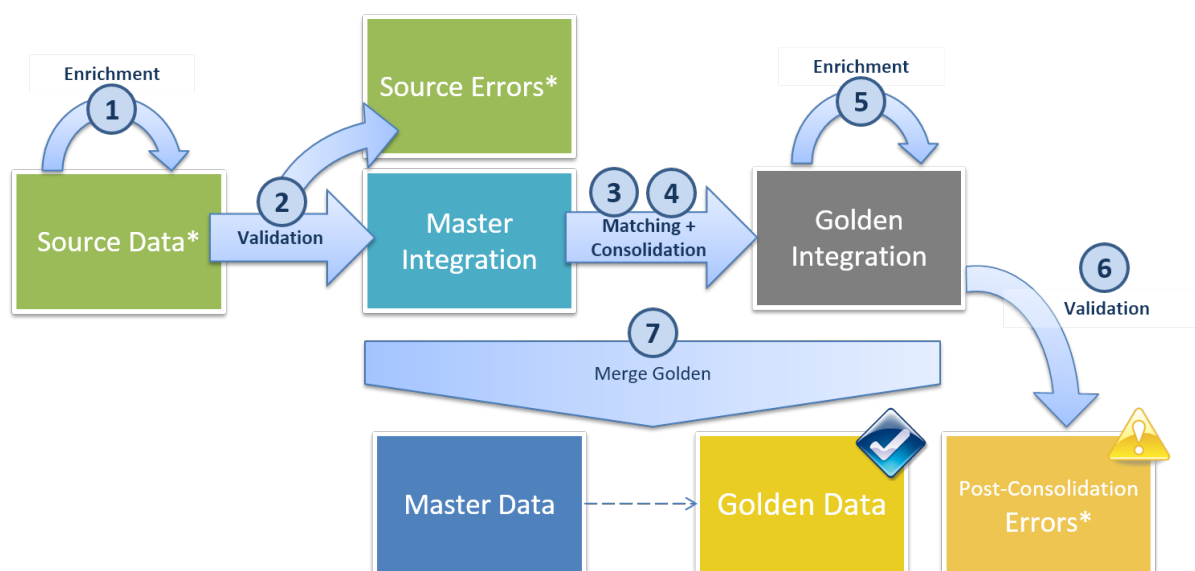
Tahapan ke empat adalah *matching* atau pencocokan data. Skenario pada Gambar 4 menggambarkan alur pencocokan data nasabah yang akan diterapkan. Skenario tersebut memberikan luaran berupa data nasabah yang memiliki *score* atau nilai *matching*.



Gambar 4. Skema *mathcing*

Grouping

Grouping adalah tahapan pengelompokan data nasabah berdasarkan nilai *matching* yang dihasilkan pada tahap sebelumnya. Data nasabah dipisahkan menggunakan *validation* dengan nilai batasan *threshold* yang ditentukan. Nilai *threshold* terdiri atas *low bound threshold* dan *upper bound threshold*. Skenario grouping dapat dilihat pada Gambar 5.



Gambar 6. Skenario *grouping*

Setiap data nasabah yang nilai pencocokannya diatas nilai *threshold*, maka data tersebut akan masuk ke dalam kelompok *golden data* atau otomatis menjadi *single* CIF. Sedangkan, data nasabah yang terletak diantara *upper bound threshold* dan *low bound threshold* akan menjadi *potential data*. Potensial data akan diproses menjadi *single* CIF pada aplikasi E-CIF berbasis *website* melalui penilaian manusia. Setiap *golden data* dan *potensial data* memiliki *key_out* untuk penomoran kelompoknya. Sampel hasil *grouping* dapat dilihat pada Gambar 7.

_key_in	_key_out	_score	FirstName	LastName	Address	City	Zip	FirstName_clean	LastName_clean
2	2	1	Kevin S.	Goff	111 Main St.	Harrisburg	17111	Kevin S.	Goff
3	2	0.8768038	Kevin	Goff	111 Main St.	Harrisburg	17111	Kevin S.	Goff
1	2	0.7196349	Kevin	Goff	111 Main Street	Harrisburg	17111	Kevin S.	Goff
4	4	1	Kevin	Goff	67 Albert Avenue	Phoenix	87644	Kevin	Goff
5	5	1	James	Goff	643 Wilson Lane	Detroit	93234	James	Goff
6	6	1	Jenny	Wilson	561 Hamilton St	New York	30091	Jenny	Wilson
7	6	0.9173867	Jennifer	Wilson	561 Hamilton St	New York	30091	Jenny	Wilson
8	8	1	Jennifer	Wilson-Smith	789 Allendale Lane	New York	30092	Jennifer	Wilson-Smith
9	9	1	Max	Crennel	3 Philadelphia Ave	Newark	13323	Max	Crennel
10	10	1	Cathy	Burks	19 Lancaster Ave	Los Angeles	91233	Cathy	Burks
11	11	1	Catherine	Burks	19 Lancaster Ave Apt 7	Los Angeles	91233	Catherine	Burks
12	12	1	Jon	DeYoung	2921 Oxford Circle	Miami	23030	Jon	DeYoung
13	12	0.9453965	John	DeYoung	2921 Oxford Circle	Miami	23030	Jon	DeYoung
14	12	0.8315697	Jonathan	DeYoung	2921 Oxford Circle	Miami	23030	Jon	DeYoung
15	15	1	John	DeYoung	15 Main Street	Seattle	11033	John	DeYoung

Gambar 7. Hasil pengelompokan

Daftar Pustaka

1. Ishibuchi H, Mihara S, Nojima Y (2013) Parallel distributed hybrid fuzzy gbml models with rule set
2. Migration and training data rotation. IEEE Trans Fuzzy Syst 21(2):355–368 Ishibuchi H, Nakashima T (2001) Effect of rule weights in fuzzy rule-based classification systems. IEEE
3. Trans Fuzzy Syst 9(4):506–515 Ishibuchi H, Nakashima T, Nii M (2004) Classification and modeling with linguistic information granules:
4. Advanced approaches to linguistic data mining. Springer, Berlin Ishibuchi H, Yamamoto T (2005) Rule weight specification in fuzzy rule-based classification systems. IEEE
5. Trans Fuzzy Syst 13:428–435
6. Stojmenovic, X. Lin. Power-aware localized routing in wireless networks, (2001) 12, No.11,1122-1133
7. D. Peppers and M. Rogers, The One to One Manager: Real-World Lessons in Customer Relationship Management (Doubleday, New York, 1999)
8. P.N. Spring, P.C. Verhoef, J.C. Hoekstra and P.S.H. Leeftang, The Commercial Use of Segmentation and Predictive Modeling Techniques for Database Marketing, Working Paper (University of Groningen, 2000)