

NoSQL Model Data Warehouse Metadata Survei Dinamis

Studi Kasus: Survei Rumah Tangga

Lutfi Rahmatuti Maghfiroh¹, Ibnu Santoso²

Politeknik Statistika STIS

¹lutfirm@stis.ac.id, ²ibnu@stis.ac.id

Abstract

A survey conducted between periods tends to experience changes in metadata. However, all data and metadata must be stored in an integrated manner for data retrieval efficiency. There are difficulty in query performed for data in between periods because of metadata changes. One query for one period data is certainly not efficient. On the other hand, one query for data for several periods can cause data consistency problems due to changes in metadata. In the latest study, we conduct further studies on several data warehouse metadata models that can accommodate dynamic metadata changes such as traditional data warehouse, Multi Version Data Warehouse (MVDW) and metadata management using non-relational models such as NoSQL. From the results of previous studies it was found that the non-relational model in the NoSQL database resulted in faster query time and more efficient storage than the relational model, but the model had not been able to handle different household survey models. This study aims to produce a data warehouse model that can manage changes in metadata in household surveys that vary dynamically. The results of the test show that the proposed model has been able to handle changes in metadata from household survey models that vary with time and place of storage that are not much different from the results of previous studies

Keywords: Metadata, Data warehouse, MVDW, NoSQL

Abstrak

Suatu survei yang dilakukan antar periode cenderung mengalami perubahan metadata. Meski demikian, seluruh data dan metadata tersebut haruslah disimpan secara terintegrasi untuk keperluan data retrieval yang efisien. Tantangan yang muncul dengan perubahan metadata ini adalah ketika dilakukan suatu query untuk data lebih dari satu periode. Satu query untuk data satu periode tentu tidak efisien. Di sisi lain, satu query untuk data beberapa periode dapat menyebabkan terjadinya masalah konsistensi data karena adanya perubahan metadata. Pada penelitian ini, kami melakukan studi lebih lanjut terhadap beberapa model metadata data warehouse yang dapat mengakomodasi perubahan metadata secara dinamis seperti data warehouse tradisional, Multi Version Data Warehouse (MVDW) dan Manajemen metadata menggunakan model non-relasional seperti NoSQL. Dari hasil penelitian sebelumnya didapatkan bahwa model non-relasional pada database NoSQL menghasilkan waktu query yang lebih cepat dan penyimpanan yang lebih efisien dibandingkan model relasional, namun model tersebut belum dapat menangani model survei rumah tangga yang berbeda-beda. Penelitian ini bertujuan untuk menghasilkan model data warehouse yang dapat mengelola perubahan metadata pada survei rumah tangga yang berbeda-beda secara dinamis. Dari hasil pengujian didapatkan hasil bahwa model yang diusulkan telah dapat menangani perubahan metadata dari model survei rumah tangga yang berbeda-beda dengan waktu dan tempat penyimpanan yang tidak jauh berbeda dari hasil penelitian sebelumnya.

Kata kunci: Metadata, Data warehouse, MVDW, NoSQL

1. Pendahuluan

Salah satu tugas dari Badan Pusat Statistik (BPS) adalah mendiseminasikan data hasil survei dan sensus yang telah dilaksanakan. Diseminasi data hasil survei selain dipublikasikan dalam bentuk naskah buku cetak juga disebarluaskan dalam bentuk raw data jika diperlukan oleh pengguna data. Pengguna data BPS yang membutuhkan raw data dapat mengajukan permohonan secara online lewat

aplikasi Silastik (<https://silastik.bps.go.id>) setelah sebelumnya menelusuri kuesioner, layout file, dan sampel data di <https://s.bps.go.id/layoutdata>.

Untuk mendapatkan raw data hasil survei/sensus yang utuh dan mudah dipahami, pengguna data harus mencocokkan terlebih dahulu antara kuesioner, layout file, dan sampel data untuk data periode tertentu. Setelah itu, petugas akan melakukan disclosure control untuk menghilangkan identitas individual pada raw data. Setelah itu, raw data yang telah melewati proses disclosure control dapat diserahkan kepada pengguna data.

SOP seperti ini tidak menjadi masalah jika pengguna data hanya membutuhkan raw data untuk satu periode saja. Masalah dapat muncul jika pengguna data membutuhkan data untuk multi periode atau data time series. Adanya perubahan data dan metadata untuk survei setiap periodenya menyebabkan pengguna data juga harus menelusuri kuesi

oner, layout file, dan sampel data untuk periode survei yang berbeda. Demikian juga petugas harus melakukan query secara terpisah untuk mendapatkan raw data survei periode berbeda karena database juga disimpan secara terpisah. Tabel 1 dan tabel 2 menunjukkan contoh kasus perbedaan data dan metadata untuk blok yang sama di survei yang sama di tahun yang berbeda.

Kasus yang pertama adalah adanya perbedaan penamaan variabel. Sebagai contoh, untuk pertanyaan umur pada periode yang satu adalah di rincian 5 disimpan dalam variabel B4AR5 sedangkan untuk periode yang lain adalah di rincian 407 disimpan dalam variabel R405. Hal ini akan menimbulkan masalah dan kesulitan bagi pengguna dalam mencari nama variabel yang sama karena dalam dua periode survei saja variabel yang digunakan sudah berbedanya. Belum lagi jika ada perbedaan variabel lagi di periode yang lain.

Kasus selanjutnya adalah adanya perbedaan pengkodean responden untuk jawaban yang sama dalam periode yang berbeda. Sebagai contoh, rincian 8 untuk periode 1 untuk jawaban “Ya” dan “Tidak” adalah kode 1 dan kode 2, sedangkan pada rincian 408 periode 2 kode untuk jawaban tersebut adalah kode 1 dan kode 5. Masalah yang hampir sama dengan kasus sebelumnya akan timbul lagi pada kasus ini, namun dengan item yang berbeda.

Selain itu, ada juga kasus tidak ada kontinuitas suatu variabel. Suatu variabel mungkin tidak ditanyakan lagi di periode survei yang lain. Sebagai contoh, rincian 10 dan 11 pada periode 1 tidak ditanyakan pada survei periode 2 dan terdapat perbedaan konsep definisi pada rincian 14-15 periode 1 dengan rincian 412-413 periode 2. Perubahan metadata ini berpotensi menimbulkan masalah inefisiensi dan juga potensi kesalahan data yang diambil pada proses data retrieval.

Namun bisa saja, kasus pertanyaan yang belum muncul di periode survei sebelumnya akan muncul di periode survei yang lain. Kasus lain yang berdampak hampir sama adalah bisa saja suatu pertanyaan muncul di suatu periode survei, kemudian tidak muncul di suatu periode survei yang lain, kemudian muncul kembali di suatu periode survei.

Untuk keperluan data retrieval yang efisien, data survei dan metadatanya untuk setiap periode seharusnya disimpan secara terintegrasi dalam suatu data warehouse. Data Warehouse (DW) adalah suatu tempat penyimpanan data yang terintegrasi dan berorientasi subyek. Penggunaan data warehouse untuk menyimpan dan menyusun data berdasarkan subyek bertujuan untuk mempermudah data retrieval dan juga analisis data. Namun, penyimpanan data pada data warehouse dengan perubahan metadata masih menjadi tantangan dalam proses data retrieval yang efisien.

Penelitian ini bertujuan untuk menganalisis model data warehouse yang cocok untuk metadata survei rumah tangga dinamis dan melakukan ujicoba terhadap model yang dihasilkan dari sisi validitas, kecepatan, dan kebutuhan tempat penyimpanan. Tujuan praktis dari penelitian ini adalah supaya nantinya pengguna data tidak perlu repot untuk mencocokkan kuesioner, layout file, dan sampel data untuk data survei yang berbeda periode. Demikian juga untuk petugas nantinya tidak

perlu untuk melakukan query secara terpisah untuk mendapatkan data survei yang berbeda periode.

Tabel 1. Layout kuesioner dan metadata untuk survei x di periode 1

IV.A. KETERANGAN ANGGOTA RUMAH TANGGA																
No. urut	Nama anggota rumah tangga (Tulis siapa saja yang biasanya tinggal dan makan di rta ini baik dewasa, anak-anak maupun bayi)	Hubungan dengan kepala rumah tangga	Jenis Kelamin 1. Laki-laki 2. Perempuan	Umur (Tahun)	Status kawinan	Apakah menjadi korban kejahatan dalam setahun terakhir?	Jika Kol. (7) 1 sd 6, Apakah dilaporkan ke Polisi? 1. Ya 2. Tidak	Berapa kali bepergian ¹⁾ selama 3 bulan kalender? Jika tidak bepergian isikan "00"	Jika bepergian (Kol. (9)≠ 00) Tujuan utama bepergian yang terakhir	Provinsi tujuan utama pada bepergian yang terakhir	Apakah mempunyai akte kelahiran dari kantor catatan sipil? Boleh saya melihatnya?	Anggota ruta berumur 0 - 17 tahun Jika Kol. (12) berkode 3 atau 4 (tidak punya/TT), Apa alasan utamanya [Jawaban jangan dibacakan!]	Anggota ruta berumur 0 - 6 tahun Apakah pernah mengikuti pendidikan pra sekolah? Boleh 1. Ya, sedang 2. Ya, sudah 3. Tidak	Anggota ruta berumur 3-6 tahun dan Kolom 14 berkode 1 atau 2 Apakah mengikuti pendidikan pra sekolah dalam 3 bulan digunakan untuk sekolah?	Anggota ruta berumur 3-6 tahun dan Kolom 14 berkode 1, 2 atau 3, 4 Apakah mengikuti pendidikan pra sekolah dalam 3 bulan digunakan untuk sekolah? 1. Ya 2. Tidak	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)

Tabel 2. Layout kuesioner dan metadata untuk survei x di periode 2

BLOK IV. KETERANGAN DEMOGRAFI, AKTE KELAHIRAN, DAN PENDIDIKAN ANAK USIA DINI (PAUD)												
NAMA ANGGOTA RUMAH TANGGA												
No Urut	(Tuliskan terlebih dahulu semua orang yang biasanya tinggal dan makan di rumah tangga ini mulai dari kepala rumah tangga, pasangannya, anak yang belum menikah, anak yang sudah menikah, menantu, cucu, orang tua/mertua, famili lain, pembantu, dan lainnya)	APAKAH HUBUNGAN DENGAN KEPALA RUMAH TANGGA? (Kode)	APAKAH STATUS PERKAWINAN (nama)? 1. Belum kawin 2. Kawin 3. Cerai hidup 4. Cerai mati	APAKAH (nama) LAKI-LAKI ATAU PEREMPUAN? 1. Laki-laki 2. Perempuan	BERAPAKAH UMUR (nama)? (tahun) Umur harus diisi, jika ≥ 97 tahun, tulis 97	Jika status perkawinannya kawin (cek 404 berkode 2) APAKAH SUAMI/ISTRI (nama) BIASANYA TINGGAL DI RUMAH TANGGA INI? 1. Ya 5. Tidak	Jika pernah kawin (cek 404 berkode 2 s.d. 4) PADA UMUR BERAPA (nama) MELANGSUNGKAN PERKAWINAN PERTAMA?	Cek 407, untuk ART berumur 0-17 tahun APAKAH (nama) memiliki akte (410 berkode 3), APA ALASAN UTAMANYA SEKOLAH? (Kode)	Cek 407, untuk ART berumur 0-10 tahun APAKAH (nama) PERNAH MENGIKUTI PENDIDIKAN PRA SEKOLAH? (Kode)			
401	402	403	404	405	407	408	409	410	411	412	413	

Cukup satu query saja untuk mendapatkan data yang diperlukan meskipun berbeda periode.

Sebagaimana dijelaskan pada penelitian sebelumnya pada penelitian Maghfiroh & Baskara (2017), untuk menangani perubahan metadata dalam data warehouse, Kimbal & Ross (2013) memperkenalkan konsep slowly changing dimensions (SCD) untuk melacak perubahan. Namun, teknik ini masih memiliki keterbatasan yang hanya dapat menangani perubahan pada atribut struktur anggota pada model multidimensional. Teknik ini tidak dapat menangani perubahan pada struktur atribut non-anggota seperti menambahkan atribut, menghapus atribut. Body dkk (2002) memperkenalkan model multidimensional temporal untuk menangani perubahan struktur Data Warehouse. Namun, pendekatan ini terbatas hanya menggunakan satu tabel fakta untuk menyimpan semua versi data. Hal ini membatasi perubahan skema dan hanya mendukung perubahan struktur pada tabel dimensi.

Bebel dkk (2006) menangani perubahan data warehouse menggunakan multiversion data warehouse (MVDW) dengan menyediakan format model untuk membuat MVDW. Penelitian ini dilanjutkan oleh Wrembel dan Morzy (2006), yang menambahkan penanganan definisi dan melakukan query di MVDW. Penelitian lainnya oleh Wrembel dan Morzy (2006) mengusulkan query pada MVDW dan menggunakan MVDW untuk menangani metadata pada penelitian Wrembel dan Babel (2007). Model yang diusulkan diimplementasikan menggunakan model relasional pada Oracle PL/SQL.

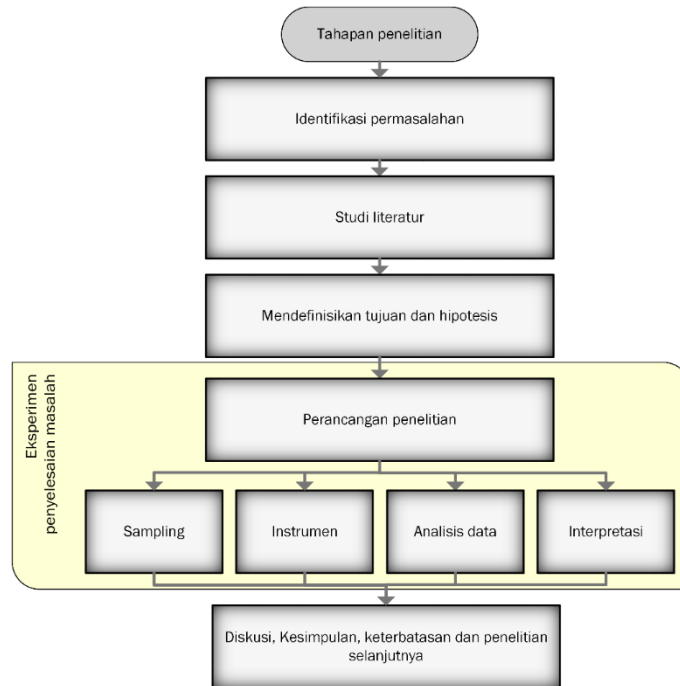
Kelemahan dari model relasional adalah adanya asumsi bahwa data telah terstruktur dan disimpan dalam bentuk baris dan kolom tabel. Tabel memiliki jumlah kolom yang sama dan tipe data yang sama pula (Parker dkk, 2013). Model ini belum dapat menangani data survei dengan berbagai struktur dengan perubahan variabel. NoSQL (*Not Only SQL*) dapat digunakan untuk mengatasi masalah ini karena model NoSQL fleksibel. Menurut Fitri (2013), Database NoSQL juga menyediakan mekanisme yang lebih mudah dan sederhana untuk menyimpan dan mengambil data dibandingkan database relasional. Beberapa penelitian lanjutan yang dilakukan oleh Chevalier dkk mengusulkan implementasi Data warehouse dan model multidimensional menggunakan NoSQL. Namun, penelitian tersebut belum dapat menangani perubahan metadata. Selain itu tidak terdapat cara menangani perubahan konsep dan definisi variabel yang ada dalam survei.

Maghfiroh & Baskara (2017) sudah membangun model DW dengan NoSQL untuk dapat menangani perubahan data dan metadata survei. Namun ternyata model yang diusulkan masih belum dinamis atau belum bisa menangani perubahan data dan metadata yang terjadi pada survei dengan model yang berbeda, baru hanya satu model survei saja.

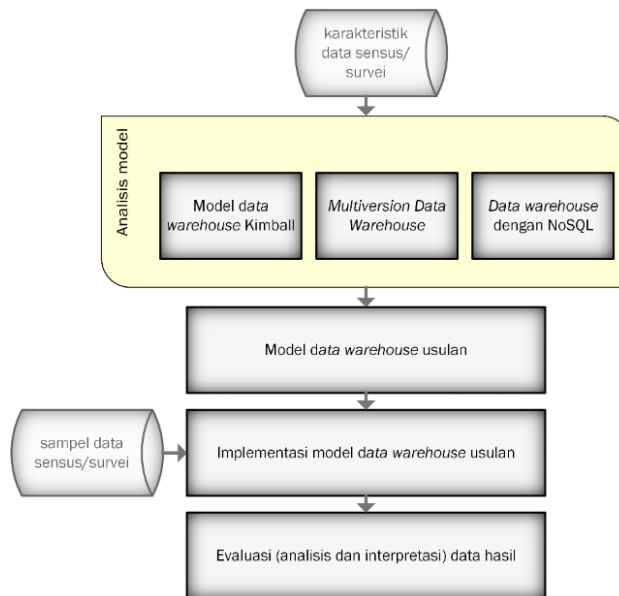
2. Metode Penelitian

Tahapan penelitian ini dimulai dari identifikasi permasalahan, kemudian melakukan studi literatur, mendefinisikan tujuan dan hipotesis. Desain penelitian yang digunakan adalah experimental design yang dalam hal ini terbagi dalam pemilihan kasus data, instrumen yang digunakan, analisis data dan interpretasi sebagaimana ditunjukkan pada gambar 1.

Dalam *experimental design* ini, tahapan yang dilakukan adalah mempelajari karakteristik data sensus/survei, kemudian melakukan analisis model untuk tiga model utama Data warehouse yang ada yaitu Model Data warehouse Kimball, *Multiversion Data Warehouse*, dan Data warehouse dengan NoSQL. Dari ketiga model tersebut kami telaah mana yang kiranya paling sesuai sebagai dasar usulan model data warehouse. Kemudian model data warehouse usulan diimplementasikan pada sampel data dan metadata survei kemudian dilakukan evaluasi (analisis dan interpretasi) hasilnya. Hasil evaluasi terhadap model usulan akan dibandingkan dengan eksperimen yang sama pada model sebelumnya. Dengan demikian kami dapat mengevaluasi model yang kami usulkan. Proses dalam tahap experimental design ini ditunjukkan pada gambar 2.



Gambar 1. Tahapan Penelitian



Gambar 2. Tahapan Experimental Design

3. Hasil dan Pembahasan

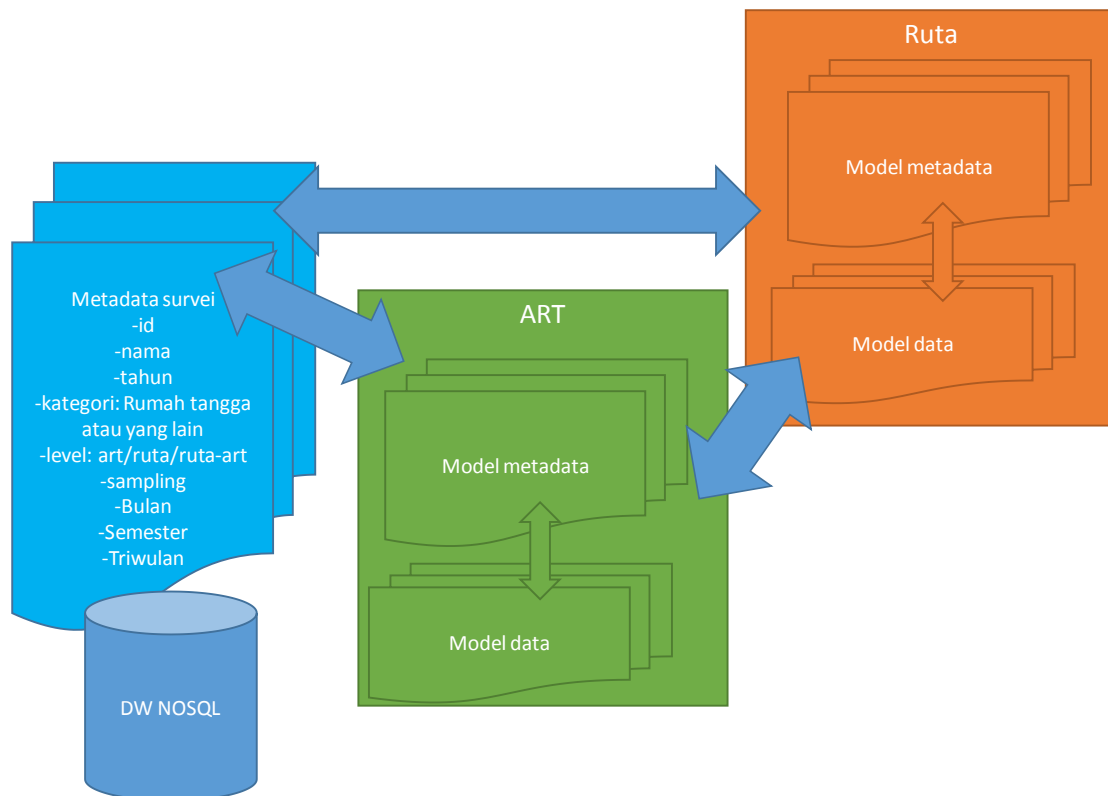
3.1 Model Data Warehouse

Untuk menangani perubahan metadata dan data yang ada, kami merancang model DW dengan NoSQL berbasis dokumen. Model yang kami rancang terdiri dari 3 Bagian. Bagian yang pertama menyimpan informasi tentang survei, yang tersimpan dalam Koleksi Metadata Survei.

Bagian kedua menyimpan informasi tentang rumah tangga, yang tersimpan dalam Koleksi Metadata Rumah Tangga dan Koleksi Data Rumah Tangga. Koleksi Metadata Rumah Tangga menyimpan metadata responden rumah tangga dan perubahan yang terjadi. Koleksi Data Rumah Tangga menyimpan data responden rumah tangga dan perubahan yang terjadi.

Bagian ketiga menyimpan informasi tentang anggota rumah tangga, yang tersimpan dalam Koleksi Metadata Anggota Rumah Tangga dan Koleksi Data Anggota Rumah Tangga. Koleksi Metadata Anggota Rumah Tangga menyimpan metadata responden anggota rumah tangga dan perubahan yang terjadi. Koleksi Data Anggota Rumah Tangga menyimpan data responden anggota rumah tangga dan perubahan yang terjadi.

Jika survei hanya terdiri dari data rumah tangga (level: ruta), maka metadata dan data yang terisi hanya untuk Koleksi Metadata Survei, Koleksi Metadata Rumah Tangga dan Koleksi Data Rumah Tangga. Jika survei hanya terdiri dari data anggota rumah tangga (level: art), maka metadata dan data yang terisi hanya untuk Koleksi Metadata Survei, Koleksi Metadata Anggota Rumah Tangga dan Koleksi Data Anggota Rumah Tangga. Jika survei terdiri dari data rumah tangga dan anggota rumah tangga (level: ruta-art), maka semua koleksi metadata dan data harus terisi. Gambar 3 mengilustrasikan hubungan antara metadata survei, metadata rumah tangga, dan metadata anggota rumah tangga dalam database NoSQL.



Gambar 3. Metadata Survei, Metadata Ruta, dan Metadata ART dalam DB NoSQL

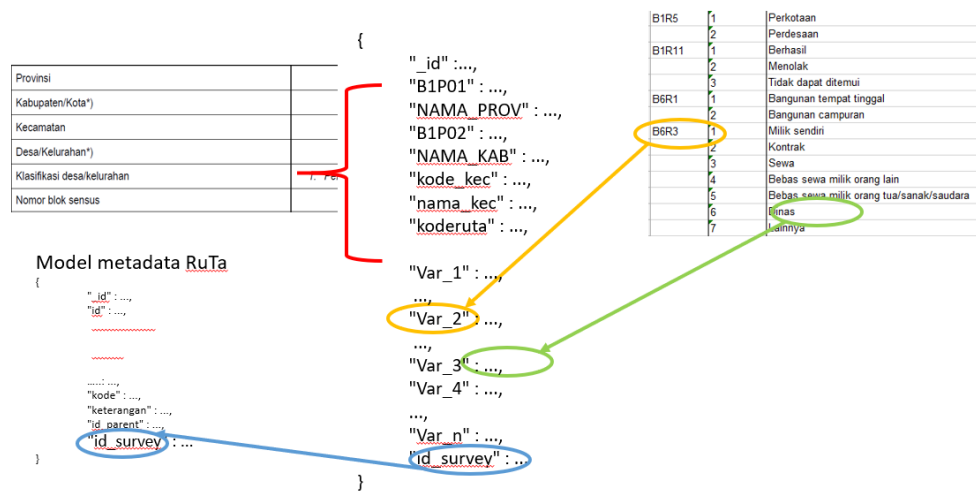
3.2. Metadata Survei

Metadata survei menyimpan informasi tentang:

- id
- nama
- tahun
- kategori: Rumah tangga atau yang lain
- level: art/ruta/ruta-art
- sampling
- Bulan
- Semester
- Triwulan

3.3. Model Data Rumah Tangga

Model Data Rumah Tangga menyimpan hasil respon survei untuk rumah tangga. Kode pertanyaan yang ada akan dijadikan kolom dan nilai jawaban akan menjadi isi dari kolom tersebut. Model Data Rumah Tangga dan Model Metadata Rumah Tangga akan terhubung dengan kolom id_survei. Gambar 4 menunjukkan model data Rumah Tangga yang digunakan.



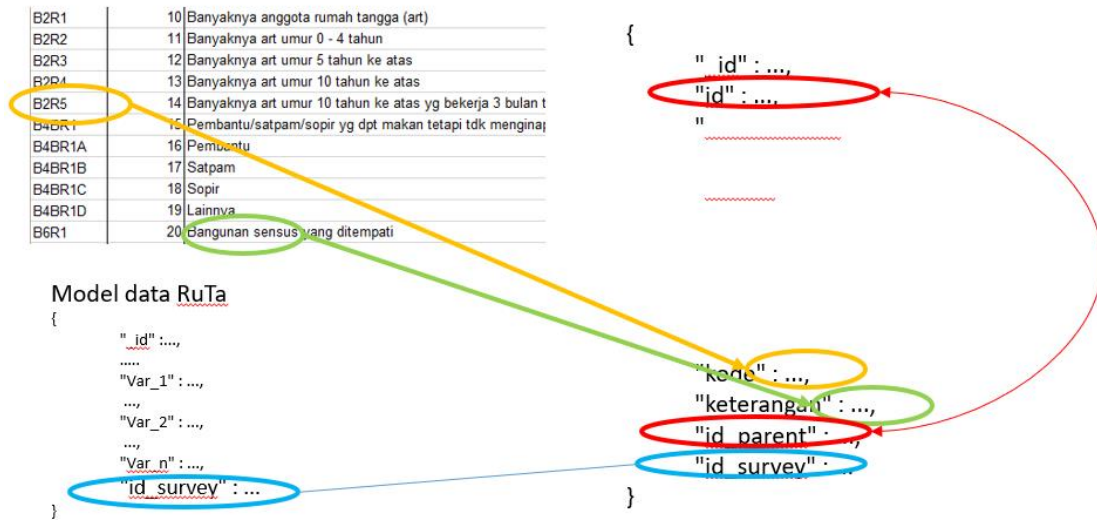
Gambar 4. Model Data Rumah Tangga

3.4. Model Metadata Rumah Tangga

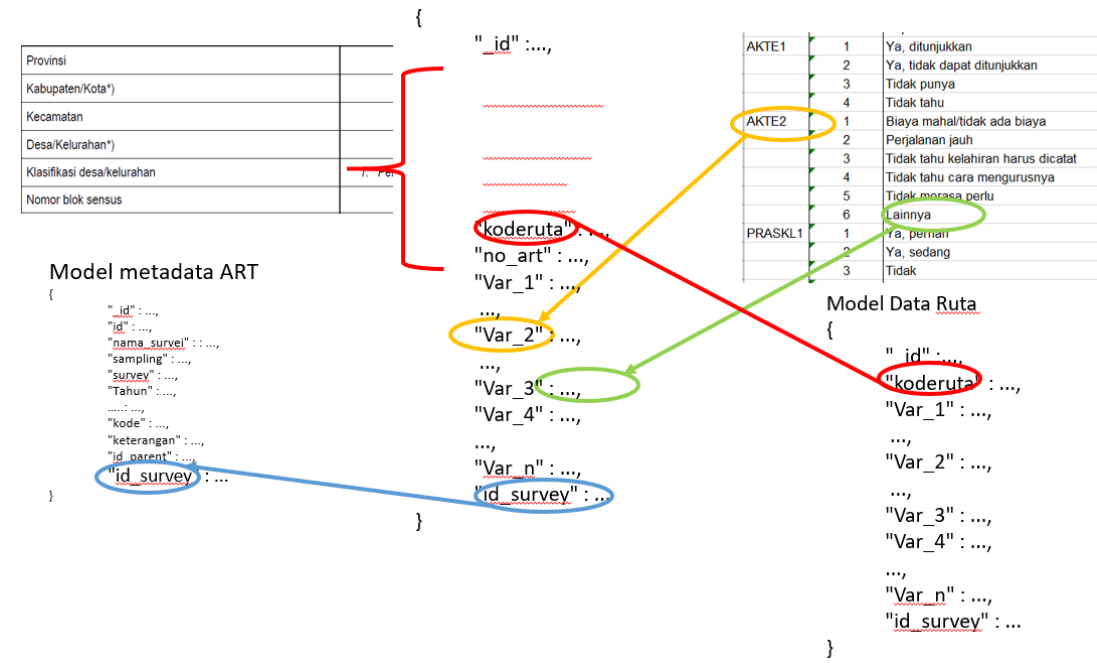
Model Metadata Rumah Tangga menyimpan informasi metadata yang ada dan perubahan metadata yang terjadi antar survei yang sama pada periode yang berbeda. Informasi yang disimpan di Model ini adalah kode variabel, keterangan dan kode induk dari kode tersebut di periode survei sebelumnya. Model Data Rumah Tangga dan Model Metadata Rumah Tangga akan terhubung dengan kolom id_survei. Gambar 5 menunjukkan model metadata rumah tangga.

3.5. Model Data Anggota Rumah Tangga

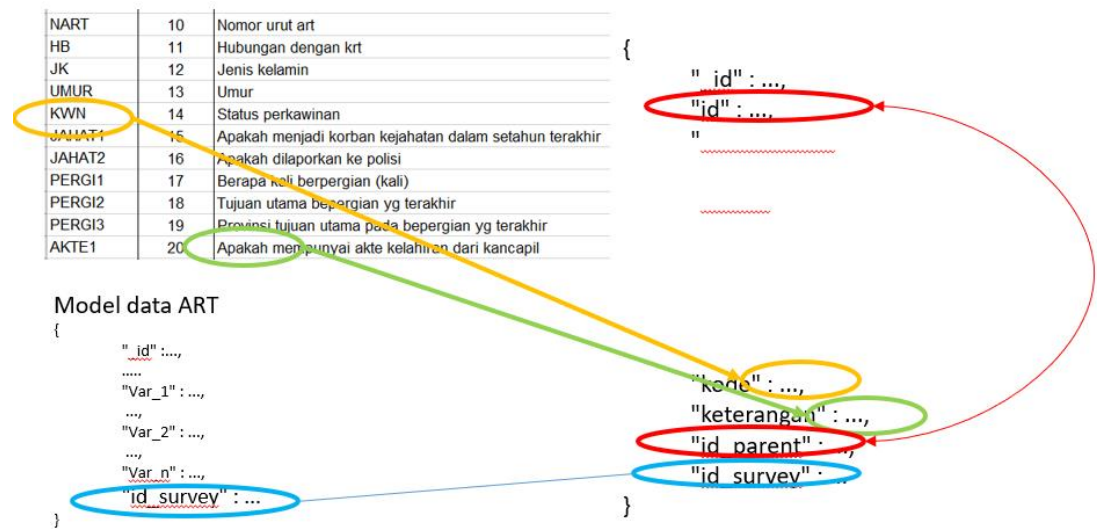
Model Data Anggota Rumah Tangga menyimpan hasil respon survei untuk anggota rumah tangga. Kode pertanyaan yang ada akan dijadikan kolom dan nilai jawaban akan menjadi isi dari kolom tersebut. Model Data Anggota Rumah Tangga dan Model Metadata Anggota Rumah Tangga akan terhubung dengan kolom id_survei. Di model ini juga terdapat kode ruta yang menghubungkan Model Data Anggota Rumah Tangga dan Model Data Rumah Tangga jika ada sebagaimana ditunjukkan pada gambar 6.



Gambar 5. Model Metadata Rumah Tangga



Gambar 6. Model Data Anggota Rumah Tangga



Gambar 7. Model Metadata Anggota Rumah Tangga

3.6. Model Metadata ART

Model Metadata Anggota Rumah Tangga menyimpan informasi metadata yang ada dan perubahan metadata yang terjadi antar survei yang sama pada periode yang berbeda. Informasi yang disimpan di Model ini adalah kode variabel, keterangan dan kode induk dari kode tersebut di periode survei sebelumnya. Model Data Anggota Rumah Tangga dan Model Metadata Anggota Rumah Tangga akan terhubung dengan kolom `id_survei` sebagaimana ditunjukkan pada gambar 7.

3.7. Algoritma

Untuk mendapatkan data yang ada dengan perubahan metadata yang ada, kami merancang algoritma. Input yang digunakan adalah query yang mengambil data anggota rumah tangga (`art`), rumah tangga (`ruta`) atau keduanya.

Jika query hanya mengambil data `ruta` maka daftar nama atribut akan di ekstrak dari query. Dari daftar nama atribut yang ada, kemudian kami cek koleksi metadata `ruta` untuk melihat perubahan metadata yang ada. Dari metadata, kami mendapatkan daftar nama atribut yang baru yang dapat kita gunakan di query. Selanjutnya dari query dengan nama atribut yang baru tersebut kami ambil data ke koleksi data `ruta` untuk mendapatkan data yang diperlukan.

Jika query hanya mengambil data `art` maka daftar nama atribut akan di ekstrak dari query. Dari daftar nama atribut yang ada, kemudian kami cek koleksi metadata `art` untuk melihat perubahan metadata yang ada. Dari metadata, kita mendapatkan daftar nama atribut yang baru yang dapat kita gunakan di query. Selanjutnya dari query dengan nama atribut yang baru tersebut kita ambil data ke koleksi data `art` untuk mendapatkan data yang diperlukan.

Jika query mengambil data `ruta` dan `art`, setelah hasil query didapatkan, kami juga mengambil kode `ruta` dari koleksi data `ruta` untuk digunakan dalam query yang dijalankan dalam koleksi data `art`. Dengan demikian query akan mendapatkan data `ruta` dan `art`. Alur yang digunakan ditunjukkan pada gambar 8. Algoritma usulan untuk mendapatkan data dengan model yang diusulkan dapat dilihat di gambar 10. Algoritma ini lebih kompleks dibandingkan dengan algoritma pada model sebelumnya. Algoritma dari model sebelumnya dapat dilihat pada gambar 9.

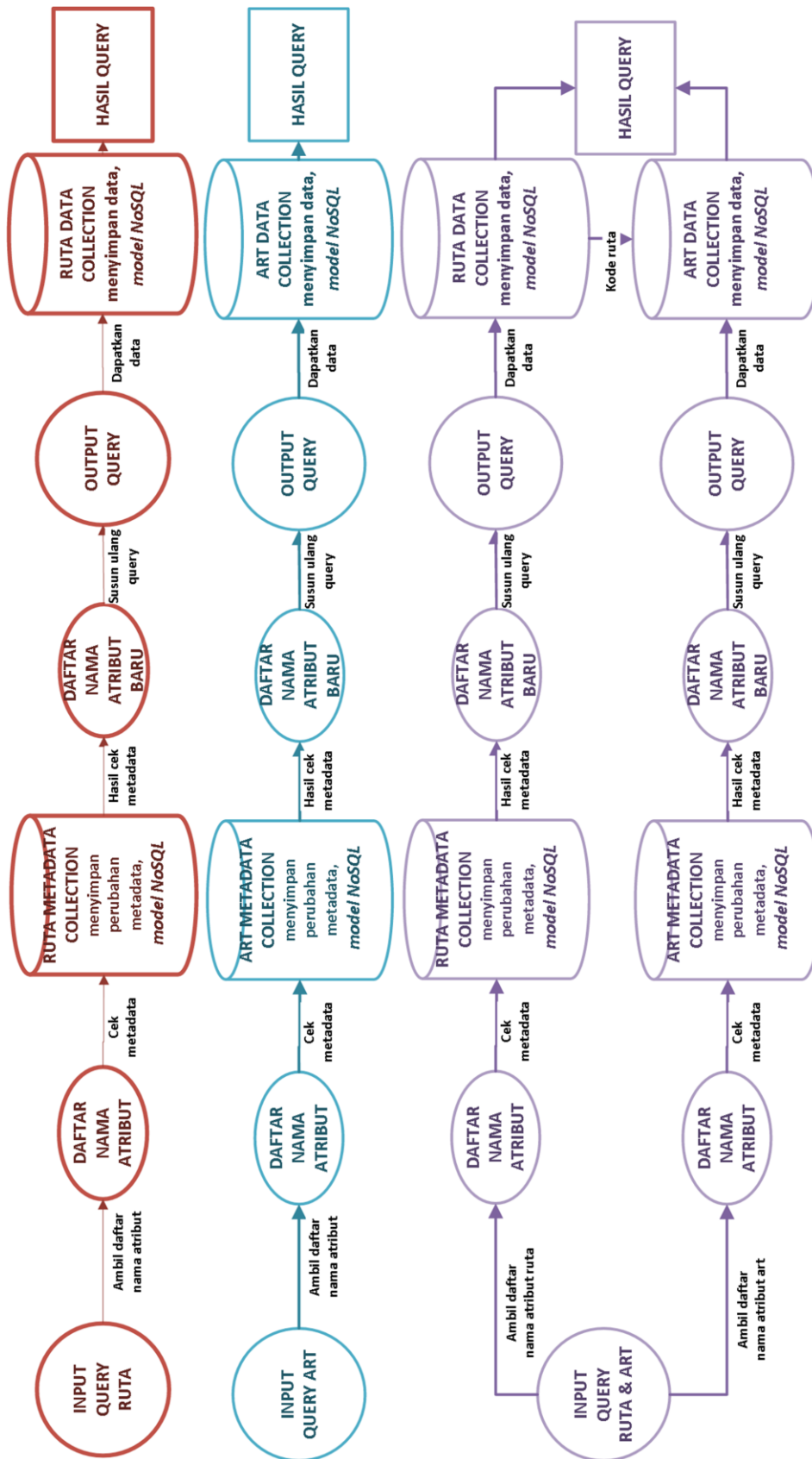
Untuk mengimplementasikan model yang sebelumnya dan yang kami buat, kami menggunakan MongoDB versi 3.2. Untuk mengimplementasikan algoritma, kami menggunakan Bahasa pemrograman java dengan Netbeans 8.2. skema pengujian yang dilakukan dapat dilihat pada gambar 11.

3.8. Ujicoba

Untuk melakukan validasi terhadap model yang kami rancang, maka kami melakukan eksperimen pengujian terhadap model yang kami rancang dan model sebelumnya. Kami melakukan eksperimen dengan lingkungan perangkat lunak sistem operasi Windows 10 64 bit dan perangkat keras prosesor Inter ® Core ™ i7-8550U CPU @1.88GHz 1.99GHz dengan RAM 8GB dan penyimpanan 128TB SSD+1TB HDD.

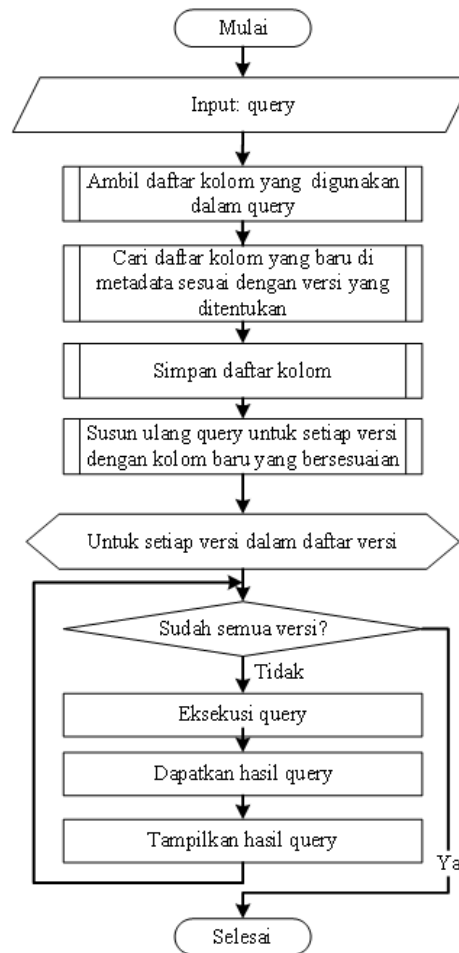
Kami menjalankan 11 kasus query yang harus mendapatkan hasil yang sama antara model yang sebelumnya dengan model yang kami buat. Kami menggunakan data Survei Angkatan Kerja Nasional tahun 2010, 2012 dan 2015. Kami melakukan iterasi sebanyak 1000 kali terhadap algoritma yang dijalankan untuk setiap query.

- Kasus 1: Menampilkan kode hubungan dengan KRT dan jenis kelamin
- Kasus 2: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan jenis kelamin
- Kasus 3: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan umur untuk penduduk laki-laki



Gambar 8. Alur mendapatkan data dengan adanya perubahan data dan metadata

- Kasus 4: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan umur untuk penduduk laki-laki diurutkan berdasarkan kode kabupaten/kota



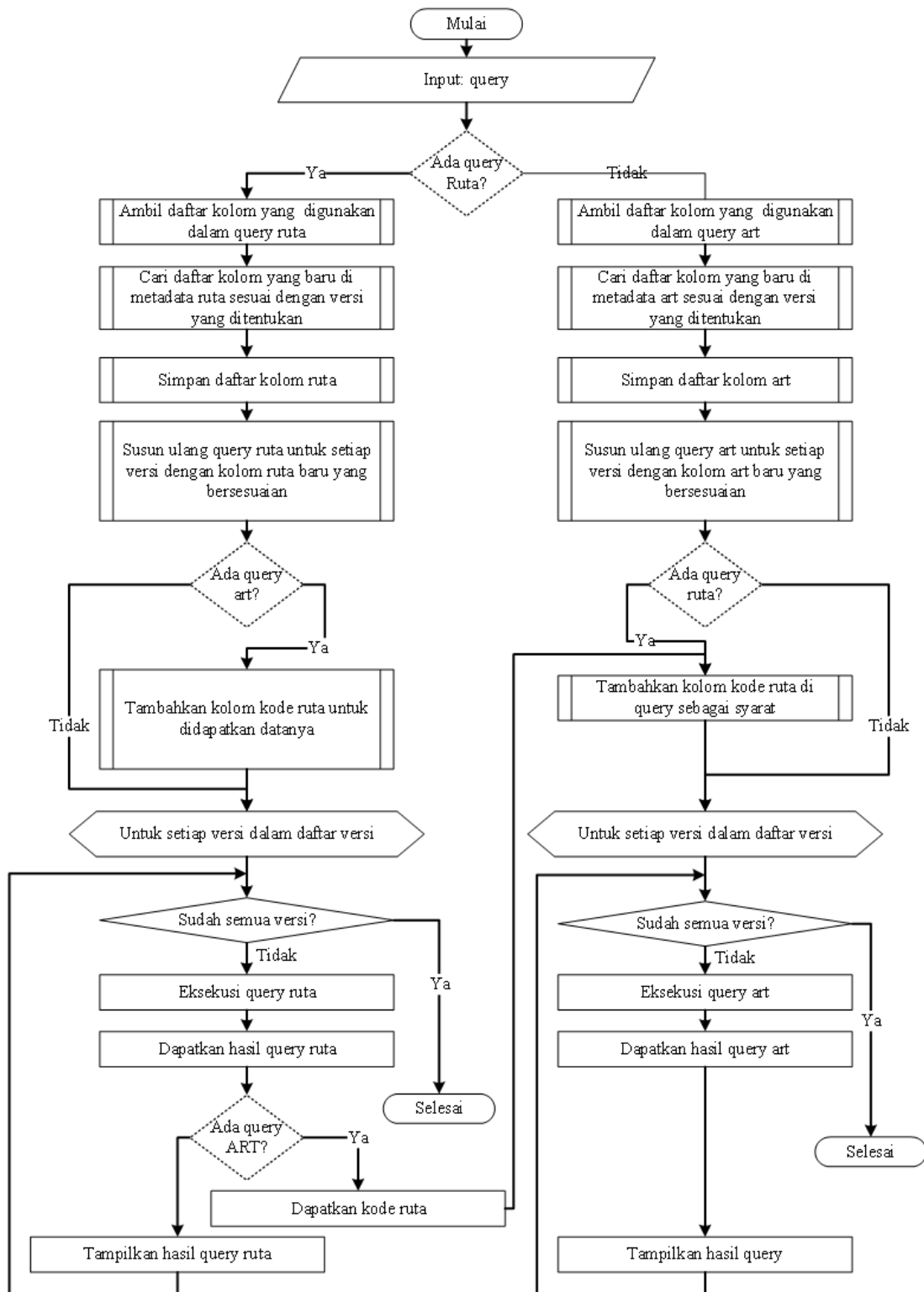
Gambar 9. Algoritma mendapatkan data dari Maghfiroh & Baskara (2017)

- Kasus 5: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan pendidikan terakhir
- Kasus 6: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan pendidikan terakhir untuk penduduk laki-laki
- Kasus 7: Menampilkan kode kabupaten/kota, nama kabupaten/kota, hubungan dengan KRT dan pendidikan terakhir untuk penduduk laki-laki diurutkan berdasarkan kode kabupaten/kota
- Kasus 8: Menampilkan jumlah penduduk dikelompokkan berdasarkan kabupaten/kota
- Kasus 9: Menampilkan jumlah penduduk dikelompokkan berdasarkan kabupaten/kota dan jenis kelamin
- Kasus 10: Menampilkan jumlah penduduk dikelompokkan berdasarkan kabupaten/kota dan jenis kelaminnya untuk penduduk yang pendidikan terakhirnya S1
- Kasus 11: Menampilkan jumlah penduduk dikelompokkan jenis kelamin dan pendidikannya

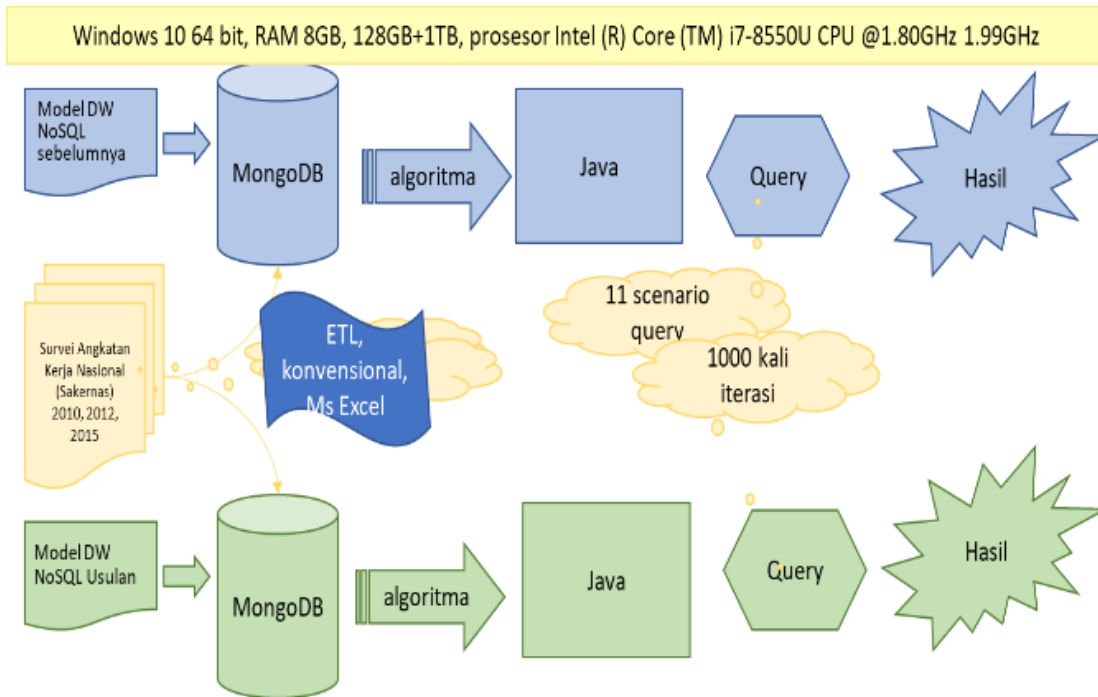
Hasil dari eksperimen tersebut kami bandingkan hasilnya untuk mengevaluasi model yang kami buat terhadap model yang sebelumnya.

3.9. Hasil Ujicoba

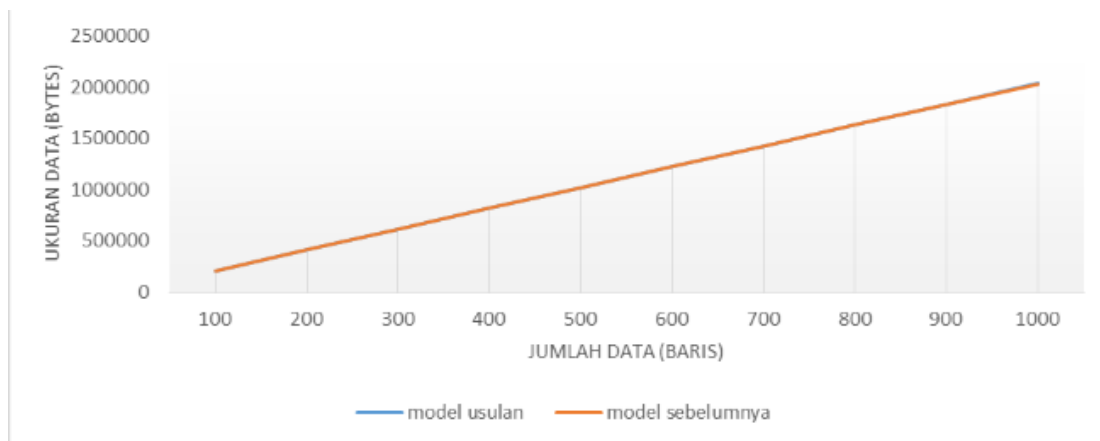
Dari hasil eksperimen, kami melakukan evaluasi terhadap ukuran data yang tersimpan dan waktu untuk menjalankan query dari iterasi yang dilakukan dengan algoritma yang diusulkan.



Gambar 10. Algoritma usulan untuk mendapatkan data



Gambar 11. Skenario Pengujian

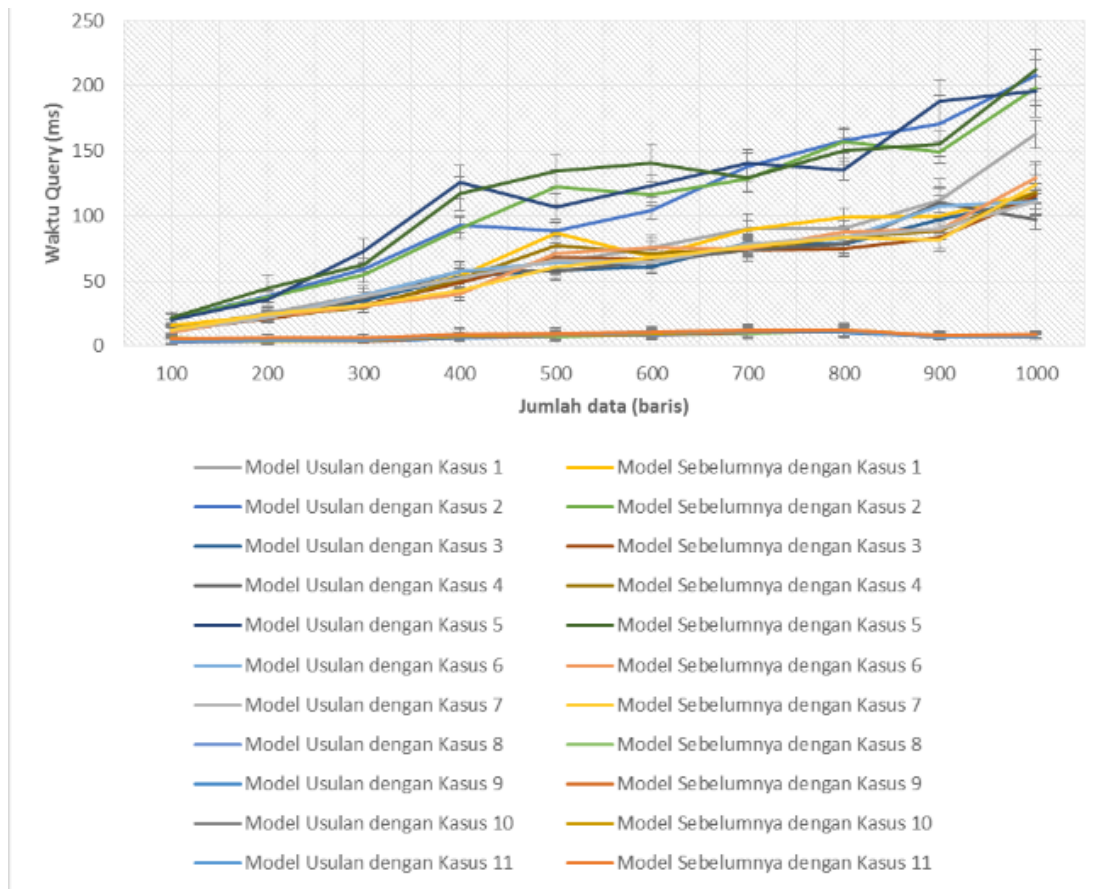


Gambar 12. Ukuran Data

Berdasarkan implementasi data yang ada dengan model yang sebelumnya dan model yang diusulkan. Ukuran data dari kedua model tersebut tidak jauh berbeda, bahkan hampir sama. Perbandingan ukuran data ditunjukkan pada gambar 12.

Dengan 11 kasus, baik dilihat secara keseluruhan maupun satu per satu kasus, waktu untuk query dengan model dan algoritma usulan cenderung sedikit lebih tinggi dibandingkan dengan model dan algoritma sebelumnya. Hal ini dapat dimungkinkan karena model dan algoritma yang diusulkan lebih kompleks dibandingkan dengan model dan algoritma usulan. Model dan algoritma yang lebih kompleks ini juga diakibatkan karena adanya kebutuhan untuk mengakomodasi model survei yang berbeda-beda, dibandingkan dengan model dan algoritma sebelumnya yang hanya mengakomodasi satu model survei saja. Hasil secara umum dapat dilihat pada gambar 13

Berdasarkan hasil ujicoba, meskipun ukuran data yang hampir sama dan waktu query yang cenderung sedikit lebih tinggi, namun model dan algoritma yang diusulkan dianggap lebih baik karena dapat menangani perubahan metadata dan data dari model survei yang berbeda-beda, dibandingkan dengan model dan algoritma sebelumnya yang hanya bisa menanganai satu model survei.



Gambar 13. Waktu Query

4. Kesimpulan

Untuk menangani perubahan metadata dan data yang ada, kami merancang model DW dengan NoSQL berbasis dokumen. Model yang kami rancang terdiri dari 3 Bagian, Metadata Survei, Data dan Metadata Rumah Tangga serta Data dan Metadata Anggota Rumah Tangga

Untuk mendapatkan data yang ada dengan perubahan metadata yang ada, kami merancang algoritma. Input yang digunakan adalah query yang mengambil data anggota rumah tangga (art), rumah tangga (ruta) atau keduanya.

Ukuran data dari kedua model tersebut tidak jauh berbeda bahkan hampir sama dan waktu untuk query dengan model dan algoritma usulan cenderung sedikit lebih tinggi dibandingkan dengan model dan algoritma sebelumnya.

Meskipun model yang diusulkan sudah dapat mengakomodasi kebutuhan, namun waktu yang dibutuhkan untuk menjalankan query masih sedikit lebih tinggi. Hal ini dikarenakan algoritma yang digunakan memang lebih kompleks dibandingkan dengan algoritma yang digunakan dalam penelitian sebelumnya. Oleh karena itu, untuk penelitian selanjutnya kami menyarankan untuk melakukan penelitian lebih lanjut untuk merancang algoritma yang lebih sederhana sehingga dapat menjalankan query dengan lebih cepat.

Selain itu, penelitian yang kami lakukan baru terbatas pada survei berbasis rumah tangga. Untuk penelitian selanjutnya, dapat dilakukan penelitian untuk perancangan model DW dinamis untuk model survei dengan basis yang lain atau bahkan model yang dapat digunakan semua model survei tanpa ada batasan.

Daftar Rujukan

- B. Beçbel, Z. Królikowski, and R. Wrembel. 2006. Formal approach to modelling a multiversion data warehouse. in *Bulletin of the Polish Academy of Sciences, Technical Sciences*, vol. 54, pp. 51–62.
- M. Body et al. 2002. A Multidimensional and Multiversion Structure for OLAP Applications, in *Proceedings of the 5th ACM International Workshop on Data Warehousing and OLAP*, ser. DOLAP '02. New York, NY, USA: ACM, pp. 1–6.
- L.R. Maghfiroh and I.G.B. Baskara 2017. Survei Data and Metadata Modelling using Document Oriented NoSQL. *The International Conference on Data and Information Science ICoDis:2017*.
- R. Kimball and M. Ross 2013 *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Indianapolis: Wiley.
- R. Wrembel and B. Beçbel. 2007. Metadata Management in a Multiversion Data Warehouse. in *Journal on Data Semantics VIII*, ser. Lecture Notes in Computer Science, S. Spaccapietra et al., Eds. Springer Berlin Heidelberg, no. 4380, pp. 118–157.
- R. Wrembel and T. Morzy. 2006. Managing and Querying Versions of Multiversion Data Warehouse. in *Advances in Database Technology EDBT 2006*, ser. *Lecture Notes in Computer Science*, Y. Ioannidis et al., Eds. Springer Berlin Heidelberg, no. 3896, pp. 1121–1124.
- M. Chevalier et al. 2016. Document-oriented Models for Data Warehouses - NoSQL Document-oriented for Data Warehouses. in *Proceedings of the 18th International Conference on Enterprise Information Systems*. Rome, Italy: SCITEPRESS - Science and and Technology Publications, pp. 142–149.
- . 2015. Implementing Multidimensional Data Warehouses into NoSQL. in *Proceedings of the 17th International Conference on Enterprise Information Systems*. Barcelona, Spain: SCITEPRESS - Science and and Technology Publications, pp. 172–183.
- . 2015. Implementation of Multidimensional Databases with DocumentOriented NoSQL. in *Big Data Analytics and Knowledge Discovery*, ser. *Lecture Notes in Computer Science*, S. Madria and T. Hara, Eds. Springer International Publishing, no. 9263, pp. 379–390.
- . 2015. Implementation of Multidimensional Databases in ColumnOriented NoSQL Systems. in *Advances in Databases and Information Systems*, ser. *Lecture Notes in Computer Science*, M. Tadeusz, P. Valduriez, and L. Bellatreche, Eds. Springer International Publishing, pp. 79–91.
- . 2015. How Can We Implement a Multidimensional Data Warehouse Using NoSQL?. in *Enterprise Information Systems*, ser. *Lecture Notes in Business Information Processing*, S. Hammoudi et al., Eds. Springer International Publishing, no. 241, pp. 108–130.
- M. O. Fitri. 2013. Trend Penggunaan Nosql Untuk Basis Data Non Relasional. *Fakultas Sains dan Teknologi UIN ALauddin*, vol. 7, No 1 2013: Januari.

Z. Parker, S. Poe, and S. V. Vrbsky. 2013. Comparing NoSQL MongoDB to an SQL DB. in *Proceedings of the 51st ACM Southeast Conference*, ser. ACMSE '13. New York, NY, USA: ACM, pp. 5:1–5:6.