

# 1. Pendahuluan

## 1.1 Latar belakang

Penggunaan email semakin berkembang setiap tahunnya. Hampir semua orang memiliki email dengan banyak akun dari berbagai *provider*. Alamat email dibutuhkan untuk berbagai syarat dari penggunaan internet, salah satunya media sosial. Satu alamat email bisa menciptakan berbagai akun media sosial. Tentu data pengguna yang ditampung akan semakin berlipat ganda banyaknya, apalagi jika pengguna memiliki banyak alamat email. Kebanyakan pengguna sampai kesulitan untuk mencari email utama yang didaftarkan dalam suatu akun media sosial.

Salah satu *provider* email terbesar, google, sudah melengkapi pencarian email secara detil. Bayangkan satu alamat email bisa menampung ribuan email dari ratusan pengguna email dengan alamat email berbeda. Tentu database yang disajikan akan memakan banyak tabel dengan atribut yang sama, hanya isinya yang berbeda. Untuk satu alamat tujuan perlu dibentuk satu tabel baru yang menampung data email komunikasi antar pengguna. Lain halnya dengan menggunakan pemodelan dengan *graph* database [2].

*Graph* database adalah alat pemodelan dari data yang bersifat sederhana sampai yang kompleks. Pemodelan data kedalam *graph* bisa bermacam-macam. Dari kasus yang diambil (komunikasi data email), terlihat bahwa bentuk data adalah dinamis dan dalam skala yang besar bisa selalu terupdate. Sehingga diperlukan pendekatan sebuah *framework* relasional yang berbasis dekomposisi dan *Selektivitas-Aware* untuk pengolahan *sub-graph query* [Sakr, S. (2009)]. Menurut beliau, GraphREL adalah kerangka murni relasional untuk menyimpan dan query data grafik. GraphREL menurut [1] adalah satu-satunya metode *graph* database yang bisa menerapkan konsep B-Tree. Ini sangat menarik karena B-Tree biasa dipakai dalam pemodelan *relational database*.

Kasus dalam *relational database* tidak dapat menemukan pola untuk mengeksekusi *query* dalam *sub-graph*. Sehingga perlu adanya penerapan dengan pendekatan graphREL untuk memetakan pola data email kedalam bentuk *graph*, dan untuk mengetahui kecepatan suatu query dalam mengenali pola: satu akun mengirim email ke banyak akun, dan untuk mengenali pola satu akun yang hanya menjadi penerima (tidak pernah menjadi pengirim).

## 1.2 Perumusan masalah

Dengan mengacu latar belakang di atas, maka permasalahan yang dibahas dan diteliti adalah :

1. Bagaimana pemodelan data email dalam mencari *query*: pengguna email yang suka mengirim banyak email dan pengguna yang tidak pernah mendapat email (hanya sebagai pengirim)?
2. Bagaimana implementasi konsep B-Tree pada metode GraphREL?
3. Apakah pendekatan GraphREL lebih baik dari pendekatan *sub-graph query* lainnya?
4. Bagaimana hasil evaluasi pendekatan GraphREL terhadap komunikasi data email?

## 1.3 Tujuan

Tujuan yang ingin dicapai dalam tugas akhir ini, yaitu:

- Mengetahui cara memodelkan data email dalam mencari *query*: pengguna email yang suka mengirim banyak email dan pengguna yang tidak pernah mendapat email (hanya sebagai pengirim).
- Mengetahui hasil implementasi konsep B-Tree terhadap metode GraphREL.
- Membuktikan bahwa pendekatan GraphREL mempunyai keunggulan dalam proses *sub-graph query*.
- Mengetahui hasil evaluasi pendekatan GraphREL terhadap komunikasi data email.

## 1.4 Metodologi penyelesaian masalah

Metodologi yang digunakan dalam memecahkan masalah di atas adalah dengan menggunakan langkah-langkah berikut:

1. Studi literatur  
Pencarian literatur didapat dari perpustakaan Learning Center Telkom University, perpustakaan Lab DMC, jurnal-jurnal online gratis di internet maupun jurnal online berlangganan yang dimiliki Telkom University atau dari sumber-sumber lain.
2. Pengumpulan data  
Data email yang dibutuhkan berasal dari SNAP (Stanford Network Analysis Project)
3. Analisis dan perancangan sistem  
Melakukan analisis dan perancangan terhadap sistem yang dibangun, menganalisis metode yang akan digunakan untuk penyelesaian masalah, termasuk dalam menentukan bahasa pemrograman yang digunakan, arsitektur, fungsionalitas, dan antarmuka sistem
4. Implementasi dan pembangunan sistem

Secara umum prosedur yang dilakukan pada penelitian ini adalah:

- Melihat secara keseluruhan bagaimana dataset yang ada, apakah perlu dilakukan *feature selection* untuk menentukan *feature* apa saja yang akan terpakai dalam penelitian ini
- Melakukan analisis terhadap topik yang sedang dibicarakan dengan metode graphREL
- Melakukan *relational encoding graph data*.
- Menghitung frekuensi tiap *node*, *edge*, dan *edge connection*.
- Melakukan proses *index* dengan menggunakan metode *partitioned B-Tree*
- Melakukan analisis hasil *statistical summaries* dengan data sebenarnya berdasarkan dekomposisi dan *selectivity aware*.

## 5. Pengujian dan analisis

Pengujian dan analisis meliputi:

- a. Pengujian untuk menghitung jumlah frekuensi *node*, *edge*, dan *edge connection*
- b. Pengujian untuk menentukan apakah tabel bersifat *distinct*
- c. Pengujian untuk menentukan *pruning point* sebagai acuan *decomposition*
- d. Menganalisis *statistical summaries*
- e. Menganalisis penerapan *partitioned B-Tree indexing* dan *selectivity aware*
- f. Pengambilan kesimpulan dan penyusunan laporan Tugas Akhir

## 1.5 Sistematika Penulisan

Sistematika Penulisan pada Tugas Akhir ini terdiri atas beberapa bagian, yaitu:

### BAB 1. PENDAHULUAN

Bab ini berisi latar belakang, perumusan masalah, batasan masalah, tujuan pembahasan, hipotesa, metodologi penyelesaian masalah, dan sistematika penulisan.

### BAB 2. LANDASAN TEORI

Berisi penjelasan mengenai konsep-konsep yang digunakan pada Tugas Akhir ini.

### BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Berisi penjelasan mengenai kebutuhan dari sistem yang kemudian dilanjutkan ke tahap perancangan sistem dan implementasi.

#### **BAB 4 IMPLEMENTASI DAN ANALISIS**

Berisi penjelasan mengenai pengujian sistem yang dilakukan dan membandingkan dengan metode yang digunakan apakah sudah sesuai atau tidak.

#### **BAB 5 KESIMPULAN DAN SARAN**

Berisi kesimpulan yang diambil berkaitan dengan sistem yang dikembangkan serta saran-saran untuk pengembangan lebih lanjut.