

SKRIPSI

PENGUKURAN APLIKASI USANG DI HTTPARCHIVE



Vinson Tandra

NPM: 2016730042

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022**

UNDERGRADUATE THESIS

OBSOLETE APP MEASUREMENT IN HTTPARCHIVE



Vinson Tandra

NPM: 2016730042

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022**

LEMBAR PENGESAHAN

PENGUKURAN APLIKASI USANG DI HTTPARCHIVE

Vinson Tandra

NPM: 2016730042

Bandung, 1 Juli 2022

Menyetujui,

Pembimbing

Digitally signed
by Pascal
Alfadian Nugroho

Pascal Alfadian, Nugroho, M.Comp.

Ketua Tim Penguji
Digitally signed
by Raymond
Chandra Putra

Raymond Chandra Putra, M.T.

Anggota Tim Penguji
Digitally signed
by Chandra
Wijaya

Chandra Wijaya, M.T.

Mengetahui,

Ketua Program Studi
Digitally signed
by Mariskha Tri
Adithia

Mariskha Tri Adithia, P.D.Eng

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

PENGUKURAN APLIKASI USANG DI HTTPARCHIVE

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 1 Juli 2022



Vinson Tandra
NPM: 2016730042

ABSTRAK

Di masa teknologi saat ini, banyak perusahaan yang menggunakan *website* sebagai tempat untuk mencari informasi. Terdapat banyak *website* yang menggunakan aplikasi yang sudah usang. HTTPArchive adalah sebuah proyek yang bersifat *open source* untuk melihat bagaimana *website* dibuat. Di dalam HTTPArchive terdapat data-data historis yang disediakan untuk menunjukkan bagaimana *website* terus berkembang dan proyek ini sering digunakan untuk penelitian.

Didalam HTTPArchive terdapat dataset yang berisi jutaan *web* setiap bulan dan dapat dianalisis menggunakan teknologi BigQuery. BigQuery adalah salah satu produk dari Google yang berbasis *cloud* dan dapat digunakan untuk menganalisis data. BigQuery menggunakan SQL (*Structured Query Language*) untuk mengakses data yang diberikan dengan melakukan *query*. Maka dari itu, data dari HTTPArchive dapat dianalisis untuk mendapatkan informasi versi dari setiap aplikasi yang masih didukung dan tidak serta mencari *website* yang sudah menggunakan aplikasi yang sudah didukung. Untuk mendapatkan data-data tersebut dapat melakukan *query* pada teknologi BigQuery.

Pada skripsi ini dibuat sebuah aplikasi yang menggunakan ReactJS untuk membuat tampilan dan NodeJS untuk pengambilan data. Langkah-langkah yang dilakukan adalah mengumpulkan daftar website, mencari aplikasi yang digunakan, mencari aplikasi yang digunakan website, mengelompokkan berdasar nama semua aplikasi yang dipakai, mencari data tentang versi aplikasi yang masih didukung, dan melakukan perbandingan antara versi aplikasi yang masih dipakai sekarang dengan aplikasi yang masih didukung (*version compare*). Setelah langkah tersebut dilakukan, didapatkan hasil yaitu pada aplikasi Nginx, untuk ruang lingkup di Indonesia versi 1.10 paling banyak digunakan url, sedangkan untuk ruang lingkup dunia versi 1.14 yang paling banyak digunakan. Adapun beberapa aplikasi yang ditambahkan pada aplikasi ini yaitu Python dan JQuery Migrate. Pada aplikasi Python jumlah versi yang paling banyak digunakan adalah versi 2.7 dan pada aplikasi JQuery Migrate versi aplikasi yang paling banyak digunakan adalah versi 1.4. Data yang dihasilkan menunjukkan bahwa terdapat 4.280 *website* yang menggunakan aplikasi yang semua aplikasinya masih didukung.

Kata-kata kunci: BigQuery, HTTPArchive, Structured Query Language, Node.js, React.js, NPM, Version Compare

ABSTRACT

In today's technological era, many companies use website as a place to find information. There are many websites that use outdated applications. HTTPArchive is a open source project to see how a website is created. HTTPArchive contains historical data provided to show how website continues to evolve and this project is often used for research.

HTTPArchive contains a dataset containing millions of web every month and can be analyzed using BigQuery technology. BigQuery is a product from Google which is based on cloud and can be used to analyze data. BigQuery uses SQL (Structured Query Language) to access the given data by doing query. Therefore, data from HTTPArchive can be analyzed to get version information of each application that is still supported and not and look for website that is already using a supported application. To get these data, you can do query on BigQuery technology.

In this thesis, an application is made that uses ReactJS to create views and NodeJS to retrieve data. The steps taken are to collect a list of websites, search for applications that are used, search for applications that are used by the website, group by name of all applications used, look for data on application versions that are still supported, and make comparisons between the versions of applications that are still in use today and the application. which is still supported version compare. After these steps were taken, the results were obtained, namely the Nginx application, for the scope of Indonesia version 1.10 the most widely used url, while for the world scope version 1.14 was the most widely used. There are several applications added to this application, namely Python and JQuery Migrate. In the Python application the most widely used version is version 2.7 and in the JQuery Migrate application the most widely used version of the application is version 1.4. The resulting data shows that there are 4,280 websites that use applications where all applications are still supported

Keywords: BigQuery, HTTPArchive, Structured Query Language, Node.js, React.js, NPM, Version Compare

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan bimbingannya yang diberikan selama ini kepada penulis sehingga dapat menyelesaikan penulisan skripsi dengan judul Pengukuran Aplikasi Usang di HTTPArchive. Skripsi ini ditulis untuk memenuhi syarat kelulusan program sarjana (S1).

Selama penulisan skripsi ini banyak hambatan yang harus dihadapi namun semua hambatan tersebut dapat dilewati dengan adanya bimbingan dan bantuan dari berbagai pihak baik secara moral maupun spiritual. Untuk itu saya ingin mengucapkan terima kasih kepada :

- Mama, Papa dan Kakak yang selalu memberikan doa dan dukungan kepada penulis selama penyusunan skripsi ini.
- Bapak Pascal Alfadian Nugroho, sebagai dosen pembimbing yang selalu sabar dan selalu memberikan waktu dalam memberikan bimbingan dan masukan selama pengerjaan skripsi ini.
- Chrissandi, Zaki, Jonathan, Dian, Lara, dan Dini sebagai rekan seperjuangan selama kuliah yang telah membantu penulis selama kuliah.

Bandung, Juli 2022

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi	2
1.6 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 BigQuery	5
2.1.1 <i>Cloud Storage System</i>	5
2.1.2 <i>SQL (Structured Query Language)</i>	5
2.2 HTTPArchive	9
2.3 Pengukuran Aplikasi Usang Pada Beberapa <i>Website</i> Populer Di Indonesia	10
2.3.1 <i>Research Method</i>	10
2.3.2 Hasil Keseluruhan	11
2.4 Node.js	12
2.4.1 <i>Node Package Manager</i>	13
2.4.2 NPM CLI	13
2.5 ReactJS	13
2.5.1 JSX	13
2.5.2 Melakukan <i>Render</i> Elemen	14
2.5.3 <i>Components and Props</i>	15
2.5.4 Penanganan Event	15
2.5.5 <i>Render</i> Bersyarat	16
2.5.6 <i>List</i> dan <i>Keys</i>	17
2.5.7 Form	18
2.5.8 Hooks	19
2.5.9 NPX	21
2.6 ChartJS	21
2.7 Express.js	21
3 PERCOBAAN AWAL	23
3.1 Eksplorasi Teknologi	23
3.2 <i>Dataset</i> yang Digunakan pada HTTPArchive	31
3.3 Langkah-Langkah Query Yang Dilakukan	31
3.3.1 Mengumpulkan Daftar Website	31

3.3.2	Mencari Aplikasi Yang Digunakan <i>Website</i>	32
3.3.3	Mengelompokkan Berdasarkan Nama Semua Aplikasi Yang Dipakai	33
3.3.4	Mencari Data Tentang Versi Aplikasi Yang Masih Didukung	35
3.3.5	Melakukan Perbandingan Antara Versi Aplikasi Yang Masih Dipakai Sekarang Dengan Versi Aplikasi Yang Masih Didukung	35
3.4	Hasil <i>Sample Data</i> Apache	36
4	PENGGALIAN DATA	39
4.1	Langkah-Langkah <i>Query</i> Yang Dilakukan Dengan Data Yang Lebih Besar	39
4.1.1	Mengumpulkan <i>List Website</i>	39
4.1.2	Mencari Aplikasi Yang Digunakan <i>Website</i>	40
4.1.3	Mengelompokkan Berdasarkan Nama Semua Aplikasi Yang Dipakai	41
4.1.4	Mencari Data Tentang Versi Aplikasi Yang Masih Didukung	43
4.1.5	Melakukan Perbandingan Antara Versi Aplikasi Yang Masih Dipakai Sekarang Dengan Versi Aplikasi Yang Masih Didukung	43
4.2	Hasil <i>Sample Data</i>	44
4.2.1	Apache dan Nginx	44
4.2.2	PHP dan Python	46
4.2.3	jQuery dan jQuery Migrate	47
5	PEMBANGUNAN PERANGKAT LUNAK	49
5.1	Analisis dan Perancangan Perangkat Lunak	49
5.1.1	Use Case	49
5.1.2	Perancangan Antarmuka	50
5.2	Implementasi Perangkat Lunak	53
5.2.1	BigQuery	54
5.2.2	React.js	58
5.2.3	Node.js	62
5.3	Pengujian Fungsional	64
5.3.1	Daftar Aplikasi	64
5.3.2	Daftar URL dengan Jumlah Hasil Perbandingan	66
5.3.3	Daftar Aplikasi yang Populer	68
6	KESIMPULAN DAN SARAN	71
6.1	Kesimpulan	71
6.2	Saran	71
	DAFTAR REFERENSI	73
	A DATA APLIKASI	75
	B KODE PROGRAM	101

DAFTAR GAMBAR

2.1	<i>Algoritma untuk membandingkan versi yang dipakai dengan versi yang masih didukung</i>	11
3.1	Halaman Awal Google Cloud Project	23
3.2	Memilih <i>Project</i>	24
3.3	Membuat <i>Project</i>	24
3.4	Membuka <i>Console</i>	25
3.5	Data Terlihat Pada Dashboard	25
3.6	Google Cloud Project Page	26
3.7	Create atau Open Project	26
3.8	Membuka BigQuery	27
3.9	Membuat Dataset Baru	27
3.10	Membuat Tabel Baru	28
3.11	Memilih Save Result As BigQuery Table	28
3.12	Export Table	29
3.13	Dashboard Table	29
3.14	Create Table	30
3.15	Pilih Upload	30
3.16	Pilih Lokasi dan Format File	31
3.17	Data Sample Jumlah Aplikasi Dengan Versi yang Dipakai	37
4.1	Aplikasi Nginx	44
4.2	Aplikasi Apache	45
4.3	Aplikasi Apache dari [1]	45
4.4	Aplikasi Nginx dari [1]	45
4.5	Aplikasi PHP	46
4.6	Aplikasi Python	46
4.7	Aplikasi PHP dari [1]	47
4.8	Aplikasi jQuery	47
4.9	Aplikasi jQuery Migrate	48
4.10	Aplikasi jQuery dari [1]	48
5.1	Use Case Diagram	49
5.2	Daftar Aplikasi	50
5.3	Menampilkan Chart Ketika Menekan Tombol View	51
5.4	URL dengan Hasil Keseluruhan Aplikasi	52
5.5	Aplikasi yang Populer	53
5.6	Menampilkan <i>Chart</i>	65
5.7	Melakukan Tombol Next	66
5.8	Melakukan Tombol Prev	66
5.9	Melakukan Tombol Next	67
5.10	Melakukan Tombol Prev	67
5.11	Melakukan Tombol Next	68
5.12	Melakukan Tombol Prev	69

BAB 1

PENDAHULUAN

Pada bab ini dijelaskan latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, dan sistematika pembahasan.

1.1 Latar Belakang

Di masa teknologi saat ini, banyak perusahaan yang menggunakan *website* sebagai tempat untuk mencari informasi. Terdapat banyak *website* yang menggunakan aplikasi yang sudah usang. HTTPArchive memiliki dataset penggunaan aplikasi setiap *website* dari *desktop* dan *mobile* pada bulan Januari tahun 2016 sampai sekarang yang dapat dilihat menggunakan teknologi BigQuery. Dataset pada HTTPArchive didapatkan dari *Chrome User Experience Report* (CrUX). CrUX merupakan *public dataset* dari *user experience* data pada jutaan *website*. Pengujian pada *dataset* tersebut dilakukan dengan menggunakan *Chrome* pada *desktop* dan *mobile*.

HTTPArchive ¹ adalah sebuah *project* yang bersifat *open source* untuk melihat bagaimana *website* dibuat. Di dalam HTTPArchive terdapat data-data historis yang disediakan untuk menunjukkan bagaimana *website* terus berkembang dan *project* ini sering digunakan untuk penelitian. Didalam HTTPArchive terdapat *dataset* yang berisi jutaan *website* setiap bulan dan dapat dianalisis menggunakan teknologi BigQuery.

BigQuery [2] adalah salah satu produk dari Google untuk menyimpan kumpulan data yang berbasis *cloud* dan dapat digunakan untuk menganalisis data. BigQuery dapat menjalankan *query* dalam skala *terabyte* dalam hitungan detik dan *petabyte* dalam hitungan menit. BigQuery digunakan untuk menganalisis data yang besar dengan melakukan *query*.

Pada [1] dilakukan penelitian tentang seberapa besar penggunaan aplikasi usang pada *website* di Indonesia. Data diambil dari *website* Alexa, dari 1.500 situs teratas menurut peringkat Alexa untuk pengunjung situs di Indonesia dan mengidentifikasi jenisnya aplikasi yang dipakai beserta nomor versinya, lebih dari setengah atau 63% aplikasi yang digunakan berhasil dibandingkan dengan skrip yang telah dibuat dan hasilnya aplikasi tidak lagi didukung oleh pengelolanya.

Beberapa aplikasi sudah menyediakan fitur untuk memperbarui ke versi yang paling baru tanpa harus menginstal ulang. Dalam kebanyakan kasus, versi aplikasi yang semakin baru sudah memperbaiki banyak kerentanan yang sudah diketahui. Beberapa aplikasi usang tidak memiliki pemberitahuan untuk meng-*update* sehingga pengguna tidak mengetahui jika terdapat *update*. Aplikasi yang baik biasanya memberikan *update* otomatis dan memberikan pesan yang efektif jika terjadi *update*.

Pada skripsi ini, dibuat sebuah replikasi dari [1] tetapi dengan data yang lebih besar. Data dapat diambil dari HTTPArchive dengan melakukan *query* pada BigQuery. Pada penelitian ini dilakukan perhitungan pada jumlah aplikasi yang sudah diberi versi dan belum diberi versi. Terdapat beberapa aplikasi yang informasi versinya tidak dapat ditentukan. Versi aplikasi yang tidak dapat ditentukan disebut NON-CONCLUSIVE. Versi aplikasi yang NON-CONCLUSIVE biasanya berisi simbol. Selain itu terdapat beberapa versi aplikasi yang kosong. Versi aplikasi yang kosong disebut sebagai NOT-VERSIONED Versi aplikasi yang dipakai setiap *website* juga dibandingkan dengan versi

¹https://github.com/HTTPArchive/httparchive.org/blob/main/docs/gettingstarted_bigquery.md

aplikasi yang masih didukung berdasarkan *official website*-nya. Pada penelitian ini hanya berfokus pada *semantic version*. *Semantic version* merupakan tata cara penentuan urutan pada aplikasi. Pada *semantic version* terdapat tiga komponen pengurutannya, yaitu *major*, *minor*, dan *patch*. Kemudian hasil tersebut ditampilkan dalam bentuk *bar chart*.

1.2 Rumusan Masalah

Berikut ini adalah rumusan masalah dari penelitian ini:

1. Bagaimana cara mengambil nilai informasi untuk mendukung pencarian aplikasi yang usang?
2. Bagaimana mereplikasi jurnal [1] dengan menggunakan data yang lebih besar?
3. Berapa banyak *website* pada HTTPArchive yang menggunakan aplikasi yang semua aplikasinya masih didukung?

1.3 Tujuan

Berikut ini adalah tujuan dari penelitian ini:

1. Mengambil nilai informasi dengan cara melakukan query untuk mengumpulkan daftar *website*, mencari aplikasi yang digunakan, mencari aplikasi yang digunakan *website*, mengelompokkan berdasar nama semua aplikasi yang dipakai, mencari data tentang versi aplikasi yang masih didukung, dan melakukan perbandingan antara versi aplikasi yang masih dipakai sekarang dengan aplikasi yang masih didukung.
2. Mereplikasi jurnal [1] dengan menggunakan data yang lebih besar, dikarenakan jurnal [1] menggunakan data yang lebih sedikit, sehingga *chart* dari aplikasi yang ditampilkan juga lebih sedikit. Data pada jurnal [1] ruang lingkupnya hanya sebatas wilayah Indonesia saja, sedangkan pada skripsi ini dilakukan penelitian dengan ruang lingkup global atau dunia.
3. Mencari jumlah *website* pada HTTPArchive yang menggunakan aplikasi yang semua aplikasinya masih didukung.

1.4 Batasan Masalah

Berikut ini adalah batasan masalah dari penelitian ini:

1. Data HTTPArchive yang digunakan adalah data *technologies* pada bulan Agustus tahun 2020.
2. Versi aplikasi berisi simbol merupakan NON-CONCLUSIVE.
3. Pada penelitian ini hanya berfokus pada *semantic version* yang dimana digit *major*, *minor*, dan *patch* adalah angka.
4. *Chart* pada aplikasi ini tidak menunjukkan data yang *UNVERSIONED* dan *NON-CONCLUSIVE* karena datanya terlalu besar sehingga mengakibatkan data lain tidak terlihat.

1.5 Metodologi

Bagian-bagian pekerjaan skripsi ini adalah sebagai berikut:

1. Melakukan studi literatur mengenai HTTPArchive.
2. Melakukan studi literatur mengenai Query
3. Melakukan studi literatur mengenai BigQuery.
4. Melakukan studi literatur mengenai ReactJS.
5. Melakukan studi literatur mengenai NodeJS.
6. Melakukan studi literatur mengenai ChartJS.
7. Melakukan studi literatur mengenai *website* yang dikatakan usang.
8. Menganalisis beberapa *website* yang dikatakan usang.

9. Membandingkan versi aplikasi pada data HTTPArchive yang dipakai sekarang dengan versi aplikasi yang ada pada dokumentasi.
10. Membuat perangkat lunak untuk menampilkan data.
11. Menulis dokumen skripsi.

1.6 Sistematika Pembahasan

Laporan penelitian tersusun ke dalam enam bab secara sistematis sebagai berikut.

- Bab 1 Pendahuluan
Berisi latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi penelitian, dan sistematika pembahasan.
- Bab 2 Dasar Teori
Berisi teori BigQuery, teori HTTPArchive, teori Node.js, teori React.js, Chart.js, dan Express.js.
- Bab 3 Percobaan Awal
Berisi eksplorasi BigQuery, dataset yang digunakan di HTTPArchive, langkah *query* yang dilakukan dengan data terbatas untuk membandingkan versi aplikasi, dan hasil *sample* data Apache.
- Bab 4 Penggalan Data
Berisi langkah-langkah query yang dilakukan untuk mengumpulkan data yang besar, penjelasan tentang sample data antara aplikasi apache dan nginx, php dan python, dan jquery dan jquery migrate.
- Bab 5 Pembangunan Perangkat Lunak
Berisi perancangan perangkat lunak yang dibangun, masukan dan keluaran dari perangkat lunak.
- Bab 6 Kesimpulan dan Saran
Berisi kesimpulan dari awal hingga akhir penelitian dan saran untuk penelitian berikutnya.