2D Data Visualization Tools Menggunakan Flask dan AngularJS

Jerryl Jeovano, *Departemen Teknik Informatika Institut Sains dan Teknologi Terpadu Surabaya*

***Abstrak* — Data merupakan bagian penting dari semua aplikasi. Namun seiring perkembangan aplikasi dan kecepatan stream data, volume data yang dihasilkan telah melebihi milyaran data. Semakin besar volume data, maka semakin sulit juga data tersebut dianalisa. Tujuan pembuatan software ini untuk memvisualisasikan data user dari file yang diupload kepada sistem. Memudahkan tim analis untuk memprediksi dan memajukan sebuah usaha. Visualisasi data berjalan fleksibel, sehingga user dapat menentukan sendiri data pada bagian axis-axisnya. Pembuatan website ini menggunakan framework Flask yang berbasis pada bahasa pemrograman Python, database cassandra merupakam database nosql yang mana digunakan untuk penyimpanan data website ini. Sedangkan visualisasi, website ini menggunakan library DevExpress dan Google Chart yang berbasis AngularJS. AngularJS sendiri merupakan perluasan dari bahasa pemrograman HTML. Metodologi yang akan digunakan pada pembuatan website ini adalah Scrum. Prose pembuatan website ini dilakukan dalam 4 sprint. Masing-masing sprint dalam penelitian ini berlangsung dalam jangka waktu 14 hari. Dari hasil penelitian ini dapat disimpulkan bahwa penggunaan metode scrum untuk membuat web untuk visualisasi data berjalan dengan efektif karena dalam proses pembuatan web ini lebih terstruktur sehingga time management menjadi lebih efisien. Library DevExpress lebih mudah digunakan daripada menggunakan library Google Chart. Dalam menampilkan chart, library DevExpress hanya dapat menampilkan sebuah chart dalam satu halaman.**

***Kata Kunci*—Data Visualization, Big Data, ETL, Python Flask.**

I. PENDAHULUAN

Data merupakan bagian penting dari semua aplikasi. Namun seiring perkembangan aplikasi dan kecepatan stream data, volume data yang dihasilkan telah melebihi milyaran data. Semakin besar volume data, maka semakin sulit juga data tersebut dianalisa. Penganalisaan data dilakukan agar dapat digunakan oleh tim analis untuk mendapatkan informasi-informasi yang dapat memajukan suatu bisnis.[1]

Visualisasi data memiliki peranan penting dalam menganalisa data. Visualisasi data yang di tampilkan harus divalidasi terlebih dahulu oleh user agar data yang tampil sesuai dengan yang diinginkan. Sering kali, visualisasi data yang digunakan tidak sesuai dengan kebutuhan user atau sistem sudah berkembang sehingga laporan-laporan yang disediakan tidak sesuai dengan kebutuhan user sekarang. Oleh karena itu, visualisasi data merupakan aktivitas yang seharusnya diatur oleh user dan berguna untuk kepentingan user itu sendiri.

Jerryl Jeovano, Departemen Teknik Informatika, Institut Sains dan Teknologi Terpadu Surabaya, Surabaya, Jawa Timur, Indonesia (e-mail: jerryl1@stts.edu)

Dengan adanya permasalahan tersebut maka munculah ide untuk membuat sebuah visualisasi data 2D dalam bentuk sebuah website.

Website ini diharapkan dapat membantu user dalam pembuatan laporan. Laporan ini nantinya dapat membantu tim analisis dalam hal mengambil sebuah keputusan yang mampu memajukan sebuah peluang usaha.

Pengolahan data pada website ini mengacu pada teori pengolahan data OLAP dan ETL yang akan dijelaskan pada bagian teori penunjang. Proses pengolahan data ini diperlukan sebelum memvisualisasikan data, agar data yang ditampilkan tidak ada yang redundant (sama). Dengan adanya pengolahan data, maka user akan mendapatkan hasil visual yang memiliki *value* terbaik.

II. TINJAUAN PUSTAKA

Di dalam bab ini akan dibahas beberapa teori terkait dengan penelitian ini. Detail pembahasan adalah sebagai berikut:

A. *Big Data* [1]

Data adalah jumlah, karakter, atau simbol tempat operasi dilakukan oleh komputer, yang dapat disimpan dan dikirim dalam bentuk sinyal listrik dan direkam pada media perekaman magnetik, optik, atau mekanis. Big Data juga data tetapi dengan ukuran besar. Big Data adalah istilah yang digunakan untuk mendeskripsikan kumpulan data yang berukuran sangat besar namun tumbuh secara eksponensial seiring waktu. Singkatnya, data tersebut sangat besar dan kompleks sehingga tidak ada alat manajemen data tradisional yang dapat menyimpan atau memprosesnya secara efisien.

B. *OLAP*

OLAP (*Online Analytical Processing*) [2] adalah metode pendekatan untuk menyajikan jawaban dari permintaan proses analisis yang bersifat dimensional secara cepat, yaitu desain dari aplikasi dan teknologi yang dapat mengoleksi, menyimpan, memanipulasi suatu data multidimensi untuk tujuan analis. OLAP adalah bagian dari kategori yang lebih global dari pemikiran bisnis, yang juga merangkum hubungan antara pelaporan dan penggalian data.

C. *Extract, Transform, Load (ETL)*

Extract Transform and Load (ETL) adalah sekumpulan proses yang harus dilalui dalam pembentukan data warehouse. Dengan tujuan mengumpulkan, menyaring, mengolah dan menggabungkan data-data yang relevan dari berbagai sumber untuk disimpan ke dalam data warehouse.

D. *Python Flask*

Python Flask [3] adalah *framework* atau API Python untuk aplikasi web. Flask juga merupakan aplikasi web WSGI yang

ringan. Yang berarti Flask memberi bantuan berupa *tools, library* dan *technology* yang memungkinkan untuk membangun aplikasi web. Flask merupakan bagian dari kategori kerangka kerja mikro. Kerangka kerja micro biasanya kerangka kerja dengan sedikit atau tidak ada ketergantungan ke *library* eksternal. Kerangka kerja flask lebih cepat dan mudah dengan kemampuan meningkatkan aplikasi yang kompleks daripada kerangka kerja Django dan juga lebih mudah dipelajari.

E. AngularJS

AngularJS adalah kerangka kerja struktural untuk aplikasi web dinamis dan open-source. Ini memungkinkan untuk menggunakan HTML sebagai bahasa template dan memungkinkan untuk memperluas sintaksis HTML untuk mengekspresikan komponen aplikasi dengan jelas dan ringkas. Pengikatan data dan injeksi ketergantungannya menghilangkan banyak kode yang saat ini harus ditulis. Dan itu semua terjadi dalam browser, menjadikannya mitra yang ideal dengan teknologi server apa pun.

F. Apache Cassandra

Apache Cassandra adalah sistem basis data terdistribusi open-source yang sangat kuat yang bekerja sangat baik untuk menangani volume besar catatan yang tersebar di beberapa server komoditas. Dapat dengan mudah ditingkatkan untuk memenuhi peningkatan permintaan yang tiba-tiba dengan menggunakan multi-node Cassandra cluster dan memenuhi persyaratan ketersediaan tinggi, tanpa satu titik kegagalan. Ini adalah salah satu database NoSQL paling efisien yang tersedia saat ini. DataStax menawarkan distribusi paket gratis dari Apache Cassandra. Ini juga mencakup berbagai alat lain seperti Pemasang Windows, DevCenter, dan dokumentasi profesional DataStax.

G. Trello

Trello adalah alat kolaborasi untuk mengatur dan kolaborasi proyek berbasis *cloud* untuk membantu pekerjaan tim serta memberi gambaran visual tentang apa yang sedang dikerjakan, siapa yang mengerjakannya, dan seberapa jauh mereka telah mencapai itu. Hal ini memberi pengguna visibilitas penuh dan kontrol atas proyek mereka, memungkinkan mereka untuk mengirimkan proyek tepat waktu dan sesuai anggaran.

H. Data Visualization

Visualisasi data adalah penyajian data dalam format gambar atau grafik. Ini memungkinkan para pembuat keputusan untuk melihat analitik yang disajikan secara visual, sehingga dapat memahami konsep-konsep yang sulit atau mengidentifikasi pola-pola baru. Dengan visualisasi interaktif, maka dengan mengambil konsep ini selangkah lebih maju dengan menggunakan teknologi untuk menelusuri bagan dan grafik untuk lebih detail. Tujuan utamanya adalah untuk menyaring dataset besar ke dalam grafik visual untuk memudahkan pemahaman hubungan yang kompleks dalam data. [4]

I. Library werkzeug

Library werkzeug.utils (secure_filename) adalah library yang digunakan untuk mengamankan nama file. Sehingga ketika file disimpan di sistem, pengaksesan file lebih mudah. Library werkzeug.security (generate_password_hash, check_password_hash) adalah library yang digunakan untuk mengencrypt password-password pengguna website. Sedangkan check_password_hash digunakan untuk melakukan pengecekkan terhadap password yang terdaftar dengan password yang diinput oleh pengguna website ketika akan masuk ke dalam website.

J. Library flask_mail

Library flask_mail adalah library yang disediakan oleh Matt Wright yang akan memudahkan untuk menghubungkan aplikasi dengan provider email. Library ini merupakan API yang digunakan untuk mengatur SMTP pada aplikasi. Meskipun disebut sebagai API tetapi tidak sepenuhnya sebagai API pada umumnya, masih diperlukan melakukan beberapa perubahan opsi agar mailer dapat berjalan dengan baik. Library ini juga support untuk mengirimkan email dengan format HTML.

K. Library Pandas

Library pandas [5] adalah paket Python yang menyediakan struktur data yang cepat, fleksibel, dan ekspresif yang dirancang untuk membuat bekerja dengan data "relasional" atau "berlabel" mudah dan intuitif. Ini bertujuan untuk menjadi blok bangunan tingkat tinggi yang mendasar untuk melakukan analisis data dunia nyata yang praktis dengan Python. Selain itu, ia memiliki tujuan yang lebih luas untuk menjadi alat analisis / manipulasi data open source yang paling kuat dan fleksibel yang tersedia dalam bahasa apa pun.

L. Library Datetime

Library datetime digunakan untuk memanipulasi tanggal dan waktu. Fokus implementasi adalah pada ekstraksi atribut yang efisien untuk pemformatan dan manipulasi output. Objek *naive* tidak mengandung informasi yang cukup untuk secara jelas menemukan dirinya relatif terhadap objek tanggal / waktu lainnya.

M. Library gviz_api

Library gviz_api adalah library bantuan untuk pembentukan chart dari library Google Chart. Library ini membantu terhadap pemrosesan data dari json kepada data table. Selain membantu mempermudah proses data, library ini juga berfungsi untuk mengirimkan data tersebut ke dalam halaman HTML dan JavaScript. Meskipun terdapat library ini, tetap membutuhkan pemrosesan data pada Python yang tergolong cukup rumit. Dikarenakan banyak requirement yang harus dipenuhi agar grafik dapat terbentuk.

N. The Truthful Art Data, Charts, and Maps for Communication [6]

Aspek dari pemikiran kontekstual menyangkut dalam mengidentifikasi keadaan proyek yang memiliki beberapa faktor yang mempengaruhi dan mencirikan kondisi proyek yang dilakukan. Hal ini akan menentukan batas-batas ambisi kreatif dalam sebuah proyek visualisasi. Dalam melakukan visualisasi data, terdapat tiga buah role yang bertugas dalam menentukan visualisasi data yang diinginkan yaitu stakeholder, audience, dan visualiser. Berikut ini merupakan penjelasan untuk setiap rolenya:

1) Stakeholder
   Dalam sebuah situasi mengembangkan visualisasi, sangat penting sekali untuk membangun pemahaman tentang siapa yang terlibat. Hal ini merupakan sebuah

pertanyaan kunci untuk dijawab. Pelanggan akan selalu diinvestasikan secara khusus dalam suatu hal yang akan dikembangkan. Pelanggan biasanya akan menentukan apakah pekerjaan itu berada di jalur yang benar, menurut pandangan mereka. Mereka juga akan menentukan secara kritis kapan pekerjaan pada akhirnya memiliki kualitas yang cukup untuk dipertimbangkan selesai. Jika tidak ada pelanggan, maka hal ini merupakan upaya tunggal dan akan ada lebih banyak otonomi yang dapat membebaskanyang akan menjadi tanggung jawab lebih besar untuk mengarahkan semua masalah sendiri.

2) Audience
Terdapat beberapa karakteristik target yang diharapkan dari audience yang perlu dipertimbangkan dengan cermat. Hal ini berguna agar proyek yang dibuat tidak akan menjadi irasional dan inkonsisten. Terkadang jika kita memiliki pengetahuan langsung tentang audience, hal tersebut dapat dengan mudah menandai kebutuhan mereka dan membentuk pilihan yang sesuai dengan keinginan audience. Misalnya, jika memiliki grup pemirsa tetap yang kita kenal akan memahami konteks teknis dari data yang disajikan. Sehingga kita tidak perlu menyertakan penjelasan terperinci yang diperlukan untuk audience yang kurang berpengetahuan.

3) Visualiser
Desain visualisasi data benar-benar merupakan upaya multidisiplin. Variasi inilah yang memicu kekayaan subjek dan menjadikannya tantangan yang sangat menarik. Untuk menguasainya diperlukan serangkaian keterampilan, pengetahuan, dan sikap berbeda yang mendominasi berbagai tahap proses ini. Jika kolaborasi dimungkinkan, ada keuntungan yang jelas dalam menggabungkan berbagai kemampuan menjadi tantangan bersama. Visualisasi yang berfungsi paling baik akan menawarkan perpaduan keterampilan yang seimbang. Keberhasilan akan sulit dicapai jika sebuah tim terdiri dari dominasi yang condong pada keragaman kemampuan individu.
Saat memulai suatu proyek, mungkin tidak semua keadaan yang berpotensi mempengaruhi pekerjaan Anda akan dapat ditentukan. Banyak hal berubah, hal itu sebabnya kita perlu siap untuk mengakomodasi dampak faktor-faktor baru di setiap titik dalam proses. Tentu saja, semakin banyak hal yang dapat Anda definisikan, semakin banyak hal diperbaiki dan ini mengurangi ketidakpastian. Idealnya, hal ini digunakan untuk menghilangkan sebanyak mungkin hal yang tidak diketahui. Kendala tidak selalu merupakan hal yang buruk, namun sering dapat membantu untuk berinovasi. Berikut ini merupaka kendala-kendala yang terdapat dalam sebuah proyek visualisasi data:

1) Timescale
Kendala utama biasanya adalah berapa banyak waktu yang dibutuhkan untuk mengembangkan solus. Sebagian besar proyek memiliki tenggat waktu, apakah ini dipaksakan, disepakati bersama atau ditetapkan sendiri. Bahkan jika tidak perlu mematuhi tenggat waktu, masih dapat berguna untuk menentukan tanggal target untuk membantu mempertajam kemajuan Anda. Di ujung timeline, ada tanggal mulai. Mungkin harus menunggu untuk kondisi tertentu sebelum memulai pekerjaan. Memperkirakan durasi proyek hingga tingkat yang dapat dipercaya adalah hal yang sulit untuk dinilai. Biasanya kita tidak tahu berapa lama proyek akan berlangsung sampai selesai, yang seringkali terlambat untuk berguna. Bahkan dengan pengalaman bekerja di beragam proyek, proyek yang tampaknya serupa dapat berakhir dengan durasi tugas yang sangat berbeda.

2) Pressure
Pressure bergantung pada konteks proyek, faktor biaya tertentu mungkin ada. Beberapa proyek mungkin memiliki anggaran yang dialokasikan dan oleh karena itu biaya kegiatan staf terkait perlu dikelola secara bijaksana. Biaya apa yang akan dikeluarkan untuk membayar suatu barang yang dibutuhkan dalam membuat proyek tersebut. Tekanan lebih lanjut dapat muncul dari politik di sekitar kita sehingga kita dapat menerima panduan dari orang lain bahwa pesan-pesan tertentu perlu diremehkan atau diperkuat. Ini bisa menjadi masalah sulit untuk ditangani karena kita ingin menghormati persyaratan yang diterima, tetapi kita juga tidak ingin merusak integritas apa yang kita wakili. Mungkin ada kepekaan budaya untuk dipertimbangkan dalam membuat karya untuk audience dari berbagai daerah. Masalah seputar penggunaan citra, konotasi warna, atau simbologi bentuk-bentuk tertentu mungkin perlu ditangani dengan hati-hati. Mungkin juga ada pertimbangan lingkungan, terutama yang menyangkut hasil pekerjaan kita yang perlu diperhatikan.

3) Design
Pembatasan sekitar pilihan desain tertentu merupakan hal yang umum, sering diinformasikan oleh pedoman gaya yang harus dipatuhi melalui penggunaan warna, jenis huruf dan font tertentu. Mungkin seperti perlu memasukkan logo, yang dapat mengambil ruang berharga dan tidak seimbang komposisinya, tetapi aturan adalah aturan dan oleh karena itu kita perlu tahu tentang hal-hal ini di awal. Pembatasan layout atau size mungkin juga ada untuk menentukan ruang di mana Anda harus bekerja. Misalnya, ketika memproduksi grafik untuk jurnal atau untuk keluaran digital yang perlu bekerja pada tablet atau smartphone, mungkin memiliki ruang yang cukup kecil untuk digunakan. Sebaliknya, output Anda mungkin harus sangat besar, yang dapat memperkenalkan berbagai tantangan dengan keterbacaan dan kualitas resolusi.

4) Technology
Terdapat banyak sekali alat, aplikasi, dan perpustakaan pemrograman dalam visualisasi data, yang menawarkan lanskap beragam kemampuan. Teknologi yang dimiliki aksesnya akan memengaruhi seberapa ambisius pekerjaan secara digital dan seberapa efisien yang dapat membuatnya.
Bagian berikutnya adalah penggunaan data yang akan divisualisasikan dengan melakukan empat langkah berbeda yang secara komprehensif memperkenalkan dengan potensi data mentah. Langkah pertama adalah data acquisition yang

melihat asal-usul dan metode berbeda untuk mengakses data, seperti melalui unduhan sistem, diperoleh dengan menggunakan API, atau melibatkan metode pengumpulan secara manual. Langkah berikutnya adalah data examination yang memetakan berbagai karakteristik yang menentukan jenis, ukuran, dan kondisi data. Untuk membedakan berbagai jenis data, mnemonik 'TNOIR diusulkan yaitu textual, nominal, ordinal, interval, dan ratio.

Langkah ketiga adalah data transformation yang dibangun di atas data examination dimana hal ini berguna untuk mengidentifikasi cara-cara memodifikasi dan meningkatkan data untuk mempersiapkannya untuk digunakan. Data transformation meliputi cleaning yang menghapus data yang bermasalah, creating yang membuat sebuah kalkulas dan nilai konversi yang baru, dan cosolidating yang berguna untuk memperkenalkan data lebih lanjut untuk memperluas atau menambahkan apa yang sudah dimiliki. Langkah terakhir adalah data exploration yang membahas nilai menggunakan teknik visualisasi untuk melengkapi pendekatan statistik sebagai cara untuk menemukan lebih banyak tentang kualitas dan wawasan yang terdapat pada data yang akan divisualisasikan.

O. Data Visualization A Handbook for Data Driven Design [6]

Selama beberapa dekade, banyak yang beranggapan bahwa visualisasi bekerja; bahwa mempelajari cara melakukannya dengan benar dapat membuat siapa pun menjadi lebih baik tidak hanya dalam mengomunikasikan fakta yang relevan tentang dunia, tetapi juga dalam mengeksplorasinya dan bahwa itu adalah kerajinan yang tidak didasarkan pada kecenderungan artistik, tetapi pada merangkul prinsip-prinsip dan heuristik tertentu yang berasal dari pengalaman dan penyelidikan ilmiah. Beberapa orang percaya ide-ide tersebut benar. Salah satu contohnya adalah The Hockey Stick Chart [7], yang dirancang oleh profesor Michael E. Mann, Raymond S. Bradley, dan Malcolm K. Hughes.

The hockey stick chart adalah salah satu visualisasi paling ikonik dan persuasif. Visualisasi data ini menjadi yang sangat baik dikarenakan memiliki kualitas tertentu. Tanpa adanya kualitas-kualitas ini dapat mengurangi nilai visualisasi sebuah data seperti kurang memahami arti dari sebuah data yang divisualisasikan. Berikut ini merupakan kualitas yang hatus dimiliki dalam sebuah data yang akan divisualisasikan:

1) Truthful
   Hal ini merupakan bagian yang paling penting dalam melakukan visualisasi data. Seluruh data yang dimiliki merupakan data yang asli secara menyeluruh. Menjadi jujur melibatkan dua strategi yang berbeda tetapi terhubung erat. Strategi yang pertama adalah menghindari penipuan diri. Kita sebagai manusia berevolusi untuk melihat pola noise yang tidak berarti. Kami langsung mengambil kesimpulan. Kami melihat hubungan sebab-akibat di mana hanya ada korelasi. Grafik yang jujur dibuat oleh orang-orang yang melakukan yang terbaik untuk mengatasi kelemahan intelektual dan bias kognitif dan ideologis dengan menerapkan pemikiran yang kritis. Strategi yang kedua adalah jujur dengan audiens. Tunjukkan pada mereka pemahaman terbaik tentang apa kenyataannya.
2) Functional
   Hal berikutnya yang penting dalam sebuah data yang akan divisualisasikan adalah fungsionalitas dari sebuah data. Fungsional dari sebuah data dapat membantu dalam penggambaran data yang akurat, dan dibuat dengan cara yang memungkinkan orang melakukan operasi yang bermakna berdasarkan data seperti melihat perubahan waktu. Tujuan dari visualisasi sebuah data dalam bentuk grafik adalah memandu dalam mengambil keputusan untuk membentuk sebuah informasi.
3) Beautiful
   Dalam melakukan visualisasi data harus menarik, dan menyenangkan secara estetika bagi audiens yang dituju. Beberapa hal bisa menjadi indah hanya karena menjadi efisien. Salah satunya yaitu dengan memberikan banyak hal menggunakan jumlah elemen minimum. Banyak perancang visualisasi dan ahli teori menekankan pentingnya kesederhanaan.
4) Insightful
   Tujuan visualisasi adalah wawasan, bukan gambar [8]. Hal ini berarti bahwa visualisasi yang baik membersihkan jalan untuk membuat penemuan-penemuan berharga yang tidak akan dapat diakses jika informasi disajikan dengan cara yang berbeda. Visualisasi yang menawarkan pesan yang jelas dan sepele tidak berharga. Secara implisit, desainer meminta untuk berinvestasi dalam upaya membacanya dengan imbalan sedikit pengembalian.
5) Enlightening
   Grafik yang mencerahkan adalah konsekuensi dari memperhatikan empat kualitas sebelumnya. Grafik yang jujur, fungsional, indah, dan berwawasan luas juga berpotensi menjadi pencerahan. Dengan memperhatikan empat kualitas sebelumnya, audience juga dapat memahami data yang divisualisasikan.

P. Visualize This: The FlowingData Guide to Design, Visualization, and Statistics [9]

Sebelum memulai bekerja untuk bagian visualisasi, hal yang pertama kali dibutuhkan adalah melakukan pengumpulan data yang akan divisualisasikan. Data merupakan hasl yang sangat penting untuk membuat sebuah visualisasi menjadi menarik. Jika tidak memiliki data yang menarik, maka visualisasi ini hanya akan menjadi sebuah graph yang terlupakan dan sebuah gambar yang tidak berguna.

Sebuah data yang menarik tidak akan datang secara langsung dan didapatkan secara cuma-cuma. Untuk memperolehnya memerlukan sebuah usaha yang keras dan biaya yang besar. Sebuah data yang menarik didapatkan dengan melakukan pengumpulan data yang didapatkan dari search engine, universitas, olahraga, dan sebagainya.

Setelah mendapatkan data yang diinginkan, selanjutnya adalah melakukan formatting pada data yang akan divisualisasikan. Bagian kedua ini merupakan melakukan format pada sebuah data agar dapat dibaca oleh komputer sehingga dapat dipahami dan diproses oleh komputer. Tiga buah format yang cocok agar komputer dapat memproses adalah dengan menggunakan delimiter text, JavaScript Object Notation (JSON), dan Extensible Markup Language (XML). Pada bagian ini format yang akan digunakan adalah JSON untuk melakukan proses agar data dapat divisualisasikan

Q. Perbandingan Sistem yang Dibuat dengan Tableau dan PowerBI

Perbedaan yang mendasar dari PowerBI [10] dan Tableau [11] adalah interface dari setiap software, target pasar, penggunaan software, dan jumlah data yang diproses. Tabel I menunjukkan perbedaan untuk PowerBI dan Tableau.

Tabel 1
Tabel Perbandingan PowerBI dan Tableau

Keterangan	PowerBI	Tableau
Volume Data	Volume data yang dapat diproses terbatas max 250 MB.	Volume data yang dapat diproses sangat besar.
Interface	Interface sederhana dan mudah untuk dipahami	Interface kurang baik dan sulit untuk dipahami
Target Pasar	Tidak direkomendasi untuk user yang memprioritaskan analisis data	Direkomendasi untuk user yang memprioritaskan analisis data
Penggunaan software	Mudah dipahami	Sulit dipahami

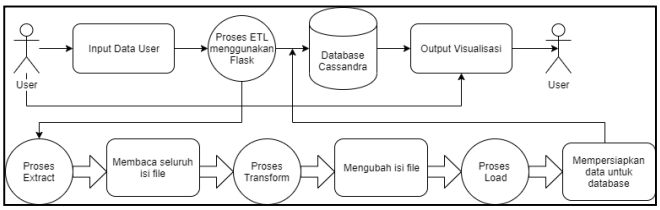
Perbedaan yang pertama antara PowerBI dengan Tableau adalah volume data yang dapat diakses. PowerBI hanya dapat memproses data maksimal sebesar 250 MB. Sedangkan untuk Tableau tidak memiliki batasan volume data yang dapat diproses. Perbedaan berikutnya yaitu design interface yang dimiliki setiap software. PowerBI memiliki interface yang sederhana sehingga lebih mudah dipahami oleh pengguna. Sedangkan untuk software Tableau, interface lebih rumit sehingga pengguna yang pemula akan sedikit lebih bingung ketika menggunakan software Tableau. Perbedaan yang ketiga adalah target pasar yang dituju. Software PowerBI lebih mengutamakan untuk kelas menengah kebawah dikarenakan jumlah data yang dapat diproses terbatas. Sedangkan untuk software Tableau lebih mengutamakan untuk melakukan pemrosesan untuk big data dikarenakan tidak memiliki batas data yang dapat diproses

III. IMPLEMENTASI

A. Arsitektur sistem

Untuk menggunakan fitur-fitur yang disediakan pada website, customer memerlukan ID yang terdaftar. Untuk mendaftar, customer bisa melakukan registrasi sendiri dengan memasukan data-data yang diminta oleh website. Data-data yang diminta antara lain adalah alamat email, nama lengkap, dan lain sebagainya. Setelah terdaftar barulah customer dapat melakukan login kedalam sistem. Setelah login, user dapat melakukan pembuatan project baru dengan melakukan pemberian nama pada project tersebut dan melakukan upload file(s) yang akan digunakan.

Setelah melakukan pembuatan project, customer akan diarahkan ke halaman baru yang akan berisikan list file yang telah diupload dan 10 baris teratas data dari file tersebut. Pada halaman yang sama juga user dapat melakukan penguploadan file baru dan beberapa fitur terhadap file yang ada. Setelah proses ETL maka data yang telah diproses tadi akan diupload ke dalam database. Kemudian customer dapat melakukan pemilihan grafik yang diinginkan beserta data yang akan digunakan pada file tersebut. Jika data telah teruplod ke database, maka customer tidak perlu mengulangi proses ETL ketika akan membuat grafik lagi, melainkan customer langsung memilih grafik dan data yang akan digunakan. Arsitektur sistem ini dapat dilihat pada gambar 1.



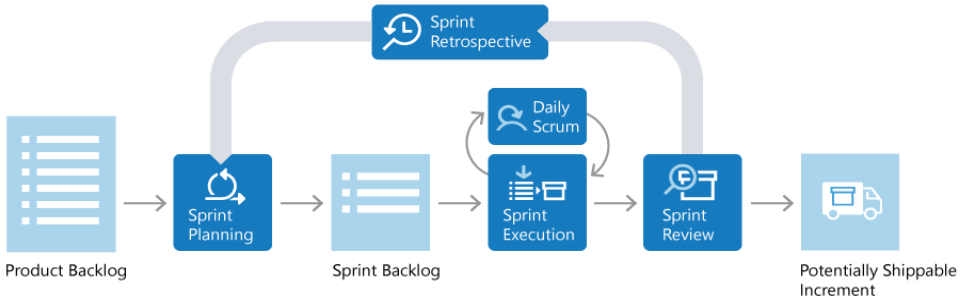
Gambar 1. Arsitektur Sistem

Proses ETL akan berjalan dalam beberapa langkah. Langkah pertama adalah ekstraksi file, sistem akan membaca file yang telah diupload dan dipilih. Selanjutnya sistem akan melakukan langkah kedua yaitu transform. Pada langkah kedua ini, data akan diproses pembersihkan dan penyesuaian data. Setelah kedua langkah tadi, maka data bisa dibilang siap untuk dimasukkan kedalam database. Langkah terakhir adalah proses penguploadan kedalam database.

B. ETL Proses

Extract Transform and Load (ETL) adalah sekumpulan proses yang harus dilalui dalam pembentukan data warehouse. Dengan tujuan mengumpulkan, menyaring, mengolah dan menggabungkan data-data yang relevan dari berbagai sumber untuk disimpan ke dalam data warehouse.

1) Extract
   Pada langkah ini, data diekstraksi dari sistem sumber ke area pementasan. Transformasi jika ada dilakukan di staging area sehingga kinerja sistem sumber tidak terdegradasi. Juga, jika data yang rusak disalin langsung dari sumber ke dalam database gudang data, rollback akan menjadi tantangan. Staging area memberi peluang untuk memvalidasi data yang diekstraksi sebelum pindah ke gudang Data.
2) Transform
   Pada langkah ini, akan menerapkan serangkaian fungsi pada data yang diekstraksi. Data yang tidak memerlukan transformasi apa pun disebut sebagai pemindahan langsung atau melalui data. Pada langkah transformasi, proses dapat melakukan operasi khusus pada data. Misalnya, jika pengguna menginginkan jumlah total penjualan yang tidak ada dalam basis data. Atau jika nama depan dan nama belakang dalam sebuah tabel ada di kolom yang berbeda. Dimungkinkan untuk menggabungkannya sebelum memuat.
3) Load
   Memuat data ke dalam basis data house data target adalah langkah terakhir dari proses ETL. Dalam data warehouse yang khas, volume data yang besar perlu dimuat dalam periode yang relatif singkat. Karenanya, proses pemuatan harus dioptimalkan untuk kinerja. Sehingga dalam hal kegagalan pemuatan, mekanisme pemulihan harus dikonfigurasikan untuk memulai kembali dari titik kegagalan tanpa kehilangan integritas



Gambar 2. Scrum Lifecycle

data. Admin database perlu memantau, melanjutkan, membatalkan beban sesuai kinerja server yang berlaku.

C. *Fitur*

Fitur – fitur yang disediakan oleh website adalah user dapat masuk atau mendaftarkan diri dengan menggunakan emailnya sendiri. Selain itu, user juga dapat masuk dengan akun Google miliknya untuk kemudahan dan kepraktisan pengguna. Website ini juga sudah mendukung untuk big data, sehingga user tidak perlu kebingungan untuk masalah jumlah data yang besar.

Website juga memfasilitasi penguploadan file dalam jumlah yang banyak dalam sekali upload. Setelah proses upload, user akan diminta untuk memilih proses mengesktrak, merubah, dan memuat (ETL) data agar data siap untuk di visualisasikan. Grafik yang dapat dibuat oleh sistem dalam bentuk 2D dengan opsi – opsi chart tertentu sesuai keinginan user itu sendiri.

D. *Desain dan Garis Besar Scrum*

Scrum [12] adalah kerangka kerja proses yang telah digunakan untuk mengelola pengembangan produk kompleks sejak awal tahun 1990-an. Scrum bukanlah sebuah proses, ataupun teknik. Akan tetapi Scrum adalah sebuah kerangka kerja dimana dapat menggunakan bermacam proses dan teknik di dalamnya. Scrum mengeksps ketidak efektifan dari manajemen produk dan teknik kerja yang ada, sehingga dapat secara terus menerus meningkatkan kinerja produk, tim, dan lingkungan kerja. Lifecycle scrum dapat dilihat pada gambar 2.

Kerangka kerja Scrum terdiri dari Scrum team dan peran-peran, acara-acara, artefak-artefak, dan aturan-aturan terkait. Setiap komponen di dalam kerjangka kerja ini memiliki tujuan tertentu dan sangat penting bagi keberhasilan penggunaan Scrum. Aturan Scrum mengikat peran-peran, acara-acara, artefak-artefak, serta menjaga hubungan dan interaksi antar komponen tersebut Ujicoba

Uji coba terhadap website dilakukan secara bertahap mengikuti jumlah sprint yang telah ditetapkan. Untuk website proses pembuatan website ini dibutuhkan 4 buah sprint. Yang masing – masing sprint memiliki waktu pengerjaan 14 hari. Dari 1 sprint dibagi menjadi bagian yang tediri dari Sprint Planning, Daily Scrum, Sprint Test, Sprint Deploy, dan Sprint Review.

Sprint 1 merupakan permulaan pengerjaan website ini. Pada sprint ini, tim memulai untuk pembelajaran, dan pemahaman terhadap bahasa permrograman, database, dan tools yang akan digunakan untuk membantu proses pengerjaan.

Sprint 2 merupakan permulaan untuk membangun website penelitian ini. Pada sprint ini, pekerjaan tim adalah memulai pembuatan prototype halaman – halaman HTML dan penggabungan modul yang telah dipelajari pada sprint 1.

Sprint 3 merupakan peralihan halaman – halaman HTML dari prototype menjadi menggunakan templates. Dan penambahan beberapa halaman baru.

Sprint 4 merupakan sprint terakhir dari proses pengerjaan website ini, pada sprint ini dapat disebut sebagai finishing dari website ini. Ringkasan uji coba yang dilakukan dari sprint 1 hingga sprint 4 dapat dilihat pada tabel 2.

Tabel 2
Ringkasan uji coba yang dilakukan dari sprint 1 hingga sprint 4

Skenario Pengujian	Yang Diharapkan	Kesimpulan
Navigasi halaman	Sistem dapat melakukan navigasi halaman	Berhasil
Pengubahan format penulisan tanggal	Sistem dapat mengubah berbagai format tanggal menjadi ‘YYYY-MM-DD’	3 format tanggal Berhasil 3 format tanggal Gagal
Pengiriman data dalam bentuk JSON	Sistem dapat mengirimkan data dalam bentuk JSON dari Python menuju JavaScript	Berhasil
Pembentukan Grafik	Sistem dapat membuat dan menampilkan chart tersebut	Berhasil

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil tersebut dapat disimpulkan beberapa poin yang akan disebutkan dibawah ini.

1) Penggunaan metode scrum untuk membuat web untuk visualisasi data berjalan dengan efektif karena dalam proses pembuatan web ini lebih terstruktur sehingga time management menjadi lebih efisien.
2) Library DevExpress lebih mudah digunakan daripada menggunakan library Google Chart.
3) Dalam menampilkan chart, library DevExpress hanya dapat menampilkan sebuah chart dalam satu halaman.

Berdasarkan hasil tersebut mendapatkan saran beberapa poin yang akan disebutkan dibawah ini.

1) Library visualisasi Google Chart lebih baik digantikan dengan library lainnya yang dapat melakukan proses zoom in dan zoom out lebih baik.
2) Library ETL yang dapat melakukan prosesnya secara

otomatis dan memiliki kemampuan mempelajari pola – pola user.

DAFTAR PUSTAKA

[1] J. D. Miller, *Big data visualization*. Packt Publishing Ltd, 2017.

[2] S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," *ACM Sigmod Rec.*, vol. 26, no. 1, pp. 65–74, 1997.

[3] M. Grinberg, *Flask web development: developing web applications with python*. " O'Reilly Media, Inc.," 2018.

[4] J. Bertin, "Semiology of graphics: diagrams, networks, maps.," 1984.

[5] W. McKinney and others, "pandas: a foundational Python library for data analysis and statistics," *Python high Perform. Sci. Comput.*, vol. 14, no. 9, pp. 1–9, 2011.

[6] A. Cairo, *The truthful art: Data, charts, and maps for communication*. New Riders, 2016.

[7] M. E. Mann, *The Hockey Stick and the Climate Wars*. 2012.

[8] S. K. Card, J. Mackinlay, and B. Shneiderman, "Reading in information visualization: using vision to Think (Chapter 1)," *Morgan Kaufmann*, vol. 6, no. 4, 1999.

[9] N. Yau, *Visualize this: the FlowingData guide to design, visualization, and statistics*. John Wiley \& Sons, 2011.

[10] A. Ferrari and M. Russo, *Introducing Microsoft Power BI*. Microsoft Press, 2016.

[11] S. Batt, T. Grealis, O. Harmon, and P. Tomolonis, "Learning Tableau: A data visualization tool," *J. Econ. Educ.*, vol. 51, no. 3–4, pp. 317–328, 2020.

[12] L. Gonçalves, "Scrum," *Control. \& Manag. Rev.*, vol. 62, no. 4, pp. 40–42, 2018.