

ANALISA VISUALISASI DATA AKADEMIK MENGUNAKAN *TABLEAU BIG DATA*

Dedy Hartama

Program Studi Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar.
Jln. Sudirman Blok A No 1,2 dan 3 Pematangsiantar 21111 - Indonesia
dedyhartama@yahoo.com

Abstract

This research explains the benefits of data analysis by visualizing Big data in performing optimization in the academic management environment. The data used is academic information system database that related with student status. In this study the authors use Tableau tools to perform data analysis based on the number of students worksheet, student status, student name table and generate student dashboard data. The results of the analysis obtained by using visualization in bentuk management graph is very fast and optimize data processing so mengetahui the development of academic database situation.

Keywords: *Visualization, Big data, Tableau analysis, worksheet, dashboard*

Abstrak

Penelitian ini menjelaskan manfaat analisis data secara visualisasi *Big data* dalam melakukan optimalisasi di lingkungan manajemen akademik. Data yang digunakan adalah database sistem informasi akademik yang berhubungan dengan status mahasiswa. Dalam penelitian ini penulis menggunakan tools *Tableau* untuk melakukan analisis data berdasarkan *worksheet* jumlah mahasiswa, status mahasiswa, tabel nama mahasiswa dan menghasilkan *dashboard* data mahasiswa. Hasil analisis yang diperoleh dengan menggunakan visualisasi dalam bentuk grafik manajemen sangat cepat dan mengoptimalkan pengolahan data sehingga mengetahui perkembangan keadaan database akademik.

Keywords: *Visualisasi, Big data, Tableau analisis, worksheet, dashboard*

1. PENDAHULUAN

Big data merupakan volume data yang besar volume dalam kisaran exabytes. Volume tersebut melebihi kapasitas penyimpanan sistem on-line saat ini dan proses sistem pengolahan data. Data, informasi, dan pengetahuan yang diciptakan dan dikumpulkan pada tingkat yang cepat mendekati kisaran exabyte pertahun. Tapi, penciptaan data dan agregasi mempercepat dan akan pendekatan kisaran zettabyte pertahun dalam beberapa tahun. Volume adalah hanya satu aspek dari data yang besar; atribut lain adalah berbagai, kecepatan, nilai dan kompleksitas. Penyimpanan dan data transportasi adalah masalah teknologi, yang tampaknya akan dipecahkan dalam waktu dekat, tapi mewakili tantangan jangka panjang yang memerlukan penelitian dan paradigma baru [1].

Defenisi *Big data* adalah sebagai jumlah data besar hanya di luar kemampuan teknologi untuk menyimpan, mengelola dan memproses secara efisien. Sesedikit 5 tahun yang lalu, kami yang hanya memikirkan puluhan hingga ratusan gigabyte penyimpanan untuk komputer pribadi kita. Hari ini, kami berpikir dalam puluhan hingga ratusan terabyte. Dengan demikian, data yang besar adalah target yang bergerak. Dengan kata lain, itu adalah jumlah data yang hanya di luar pemahaman kami segera, misalnya, kita harus bekerja keras untuk menyimpannya,

mengaksesnya, mengelola dan memproses [1]. Analisis *Big data* dapat dibedakan dari arsitektur tradisional data processing di sejumlah dimensi:

1. Kecepatan pengambilan yang sangat penting bagi para pengambil keputusan
2. Pengolahan kompleksitas karena ini memudahkan keputusan proses pembuatan
3. Transaksi data volume yang sangat besar
4. Data struktur data dapat terstruktur dan tidak terstruktur
5. Fleksibilitas terdiri dalam jumlah analisis yang dapat dilakukan pada pengolahan/analisis [2]

Analisis *Big data* inisiatif harus proyek bersama melibatkan TI dan bisnis. IT harus bertanggung jawab untuk menyebarkan alat analisis data tepat besar dan menerapkan praktek-praktek pengelolaan data suara. Kedua kelompok harus memahami bahwa keberhasilan akan diukur oleh nilai tambah oleh perbaikan bisnis yang disebabkan oleh inisiatif [3]. Untuk membuat setiap keputusan yang diinginkan ada adalah kebutuhan untuk membawa hasil penemuan pengetahuan untuk proses bisnis dan di jalan waktu yang sama dampak di berbagai *dashboard*, laporan dan analisis pengecualian yang sedang dipantau. Pengetahuan baru ditemukan melalui analisis mungkin juga memiliki bantalan pada business strategy, CRM strategi dan strategi finansial yang maju [3] Lihat gambar 1 berikut



Gambar 1. Manajemen *Big data* [3]

Dengan munculnya "*Big data*", visualisasi informasi yang muncul sebagai daerah yang tak seorang pun dapat mengabaikan. Lembaga yang menemukan bahwa pemegang saham utama ingin diberikan dengan metrik ketika diminta untuk membuat keputusan [4].

Tableau digunakan sebagai alat visualisasi untuk lebih baik mengelola jumlah data yang besar secara model grafik. Memanfaatkan visualisasi data ini efektif karena: Visualisasi grafik dari representasi data jauh lebih kuat daripada tampilan dalam bentuk angka. Sementara tabel memerlukan kita untuk membaca dan mempertimbangkan makna dan hubungan setiap nilai individu yang disajikan, visualisasi mengizinkan untuk memproses banyak nilai-nilai secara bersamaan. Lebih lanjut, manusia dapat lebih efisien dan efektif proses grafik dari tabel numerik teks. Dengan demikian, visualisasi memungkinkan analisis untuk mengenali tren, tempat pola, dan mengidentifikasi dengan cepat dan optimal [5].

2. METODOLOGI PENELITIAN

2.1. Permasalahan *Big data*

Big data menjadi tak terlihat seperti "tambang emas" nilai potensial yang dikandungnya. Dengan akumulasi dan pertumbuhan produksi, operasi, manajemen, pemantauan, penjualan, Layanan pelanggan, dan jenis lain dari data, serta peningkatan jumlah pengguna, menganalisis korelasi pola dan tren dari jumlah besar data memungkinkan untuk mencapai pengelolaan yang efisien, presisi pemasaran. Ini dapat menjadi kunci untuk membuka "tambang emas." tetapi, infrastruktur TI secara tradisional dan metode untuk pengelolaan data dan analisis tidak dapat beradaptasi dengan pertumbuhan cepat data besar [6]. Kita merangkum masalah besar data ke dalam tujuh kategori dalam tabel 1

Tabel 1. Permasalahan di *Big data* [6]

Klasifikasi masalah <i>Big data</i>	Keterangan
kecepatan	Masalah impor dan ekspor Masalah Analisis dan Statistik Masalah <i>Query</i> Masalah Real time
Jenis dan Struktur	Masalah multisource Masalah heterogenitas Masalah infrastruktur sistem
Volume dan fleksibilitas	Masalah skala linier Masalah penjadwalan secara Dinamis
Biaya	Biaya perubahannya antara mainframe dan server PC Biaya kontrol sistem yang asli
Nilai Data Mining	Analisis data mining Manfaat sebenarnya dari data mining. Terstruktur dan non terstruktur
Keamanan dan privasi	Keamanan Data Privasi
Konektivitas dan berbagi data	Data standar dan interface Protokol untuk berbagi Akses Kontrol

2.2. Teknologi *Big data*

Big data tidak hanya memberikan keuntungan tetapi juga sebuah tantangan. Pemrosesan data tradisional tidak mampu memenuhi permintaan pemrosesan data real-time yang sangat besar; kita membutuhkan generasi baru teknologi informasi untuk menangani permasalahan *Big data*. Tabel 2 mengklasifikasikan teknologi *Big data* menjadi lima kategori [6].

Tabel 2. Klasifikasi dari Teknologi *Big data* [6]

Klasifikasi Teknologi <i>Big data</i>	Teknologi <i>Big data</i> dan Alat
Dukungan infra struktur	Platform Komputasi awan Penyimpanan Cloud Teknologi Virtualisasi Teknologi Network Pengawasan sumber daya teknologi
Akuisisi data	Data Bus ETL Tools
Penyimpanan data	Sistem Distribusi File Database Relational Teknologi NoSQL Integrasi Relasi database dan Non Relasional database Database di memori
Perhitungan data	Data <i>Query</i> , Statistika, dan analisa data Data Mining dan Prediksi Grafik Analisis BI (Bisnis Inteligence)
Tampilan dan Interaksi	Grafik dan Laporan Tampilan Visualisasi

2.3. Evolusi Pemrosesan dari *Big data*

Tantangan utama yang ditimbulkan oleh masalah *Big data* adalah kemampuan untuk memproses aliran data yang ditandai dengan 3Vs variety, velocity, dan volume. Sumber *Big data* menghubungkan database terstruktur secara tradisional untuk memasukkan email, sensor, kamera video, media sosial, dan perangkat seluler (teks, video, dan audio) [7].

Pengolahan frameworks *Big data* juga harus berurusan dengan dataflows heterogen (misalnya, statis, streaming, dan transaksi), pengolahan data heterogen semantik (pengolahan batch di Hadoop, pengolahan aliran data secara realtime di mesin dan pengolahan transaksi di MySQL dan Cassandra). Tipe data heterogen (misalnya data terstruktur dari Twitter, data terstruktur dari tradisional SQL database dan data gambar dari kamera video) diatur dengan memvariasikan data volume, kecepatan data, dan jenis permintaan [7].

Untuk menjamin kinerja, penjadwalan kebutuhan platform untuk dapat memprediksi tuntutan dan perilaku kerangka yang mendasari sehingga mereka dapat cerdas mendistribusikan dan memprioritaskan beban kerja. Lebih jauh, hal ini tidak jelas bagaimana prioritas tersebut dapat dilestarikan di beberapa kerangka karena dataflows diproses di seluruh platform yang didistribusikan [7]

3. HASIL DAN PEMBAHASAN

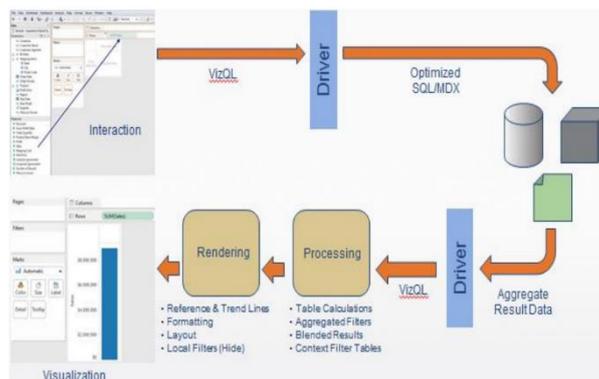
Menurut [8], *Big data* merupakan istilah yang berlaku untuk informasi yang tidak dapat diproses atau dianalisis menggunakan alat tradisional.

Menurut [9], *Big data* adalah data yang melebihi proses kapasitas dari kovensi sistem database yang ada. Data terlalu besar dan terlalu cepat atau tidak sesuai dengan struktur arsitektur database yang ada. Untuk mendapatkan nilai dari data, maka harus memilih jalan alternatif untuk memprosesnya.

Penggunaan *Big data* dalam melakukan survey penjualan Real estate di negara china. penelitiannya menjelaskan keuntungan dan kekuatannya dalam melakukan penjualan real estate dengan menggunakan aplikasi teknologi *Big data* [10]. Menjelaskan [11] peranan *Big data* dalam menganalisis dan mengembangkan data e-government untuk melakukan perubahan sehingga lebih efektif dan efisien dalam memberikan pelayanan. Penggunaan *Big data* untuk menciptakan kota pintar, yang memberikan berbagai service ilmu pengetahuan dalam menyimpan data yang begitu besar sehingga dapat memberikan informasi dari berbagai service dalam sebuah kota [12]. Menjelaskan [13] analisa *Big data* untuk kunjungan tourism yang berasal dari negara sweden. Dalam penelitian ini dibahas manfaat analisa daerah wisata dapat dengan cepat dan jelas kelebihan dan keuntungan dari masyarakat yang akan berkunjung ke daerah tersebut. Menjelaskan [14] teknologi *Big data* dalam pendekatan pencegahan penyebaran penyakit HIV dengan menggunakan teknologi digital seperti handphone.

4.1. Visualisasi Data

Analisis dapat menerapkan pengolahan data secara realtime dengan cara memvisualisasikan hasil dari pengolahan data [15]. Dalam melakukan visualisasi data untuk mengoptimalkan permasalahan di ilmu komputer khususnya dibidang akademik penulis menggunakan *Tableau*. Arsitektur *Tableau* dalam melakukan visualisasi data dapat dilihat pada gambar 2 dibawah ini :



Gambar 2. Arsitektur Pengolahan data *Tableau* [16]

Pada tahun 2003 *Tableau* berasal dari Stanford University dengan VizQL™, sebuah teknologi yang benar-benar melakukan perubahan yang bekerja dengan data dengan memungkinkan sederhana drag and drop fungsi untuk membuat visualisasi yang canggih. Inovasi yang mendasar adalah bahasa dipatenkan *query* yang diterjemahkan tindakan ke *query* database kemudian menyampaikan tanggapan grafis. Terobosan berikutnya adalah kemampuan untuk melakukan analisis pada jutaan baris data dalam detik dengan Mesin Data *Tableau* [16]. *Tableau* adalah sistem komersial, berdasarkan Polaris [17] untuk memvisualisasikan isi dari database. Antarmuka *Tableau* mencakup daftar bidang database tersedia dan ruang

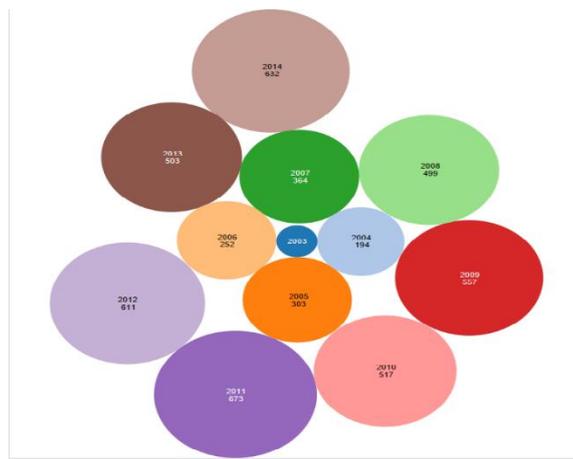
kerja di mana pengguna dapat memilih bidang dan menyeret mereka ke rak-rak yang sesuai untuk pengkodean visual seperti posisi, warna, bentuk dan ukuran. Tablo didasarkan pada spesifikasi bahasa disebut VizQL. VizQL pernyataan yang dihasilkan dari isi rak antarmuka dan mereka menentukan data yang harus divisualisasikan (sebagai perintah *query* database) dan bagaimana visualisasi akan muncul (sebagai pernyataan spesifikasi visual). Pengolahan data *Tableau* mendukung berbagai visualisasi, termasuk *bar charts*, *time series*, *scatter plots*, and *heat maps*, serta analisis operasi seperti penyaringan, penyortiran, dan pencarian [17].

Mirip dengan Microsoft Excel, *Tableau* mendukung beberapa lembar kerja. Masing-masing *worksheet Tableau* digambarkan oleh pernyataan VizQL [18]. Dalam melakukan visualisasi data tahapan yang dilakukan :

1. Melakukan koneksi dengan database SQL server yang berjalan secara real time.
2. Menyiapkan tiga workshet yang terdiri dari *worksheet* :
 - a. Jumlah mahasiswa
 - b. Status mahasiswa
 - c. Nama mahasiswa
3. Mendesign *dashboard* dari *worksheet* yang telah dibuat. Dalam visualisasikan data di *dashboard* menggunakan action yang menghubungkan dari semua *worksheet*.

A. *Worksheet* Jumlah mahasiswa

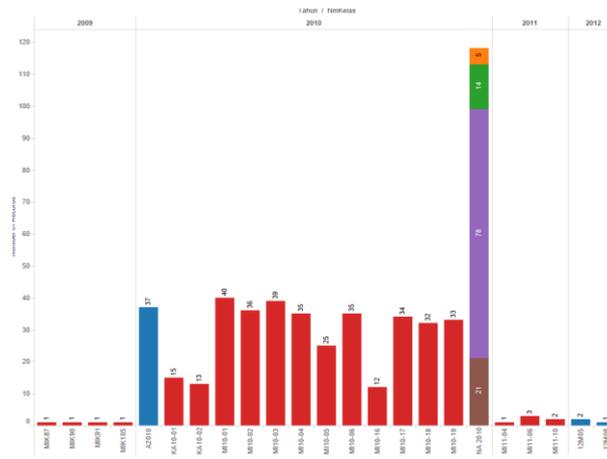
Dalam membuat *worksheet* mahasiswa, jumlah mahasiswa di visualisasikan berdasarkan tahun masuk yang di filter berdasarkan tahun masuk dan status mahasiswa. Jumlah mahasiswa kita menggunakan chart circle, seperti gambar 3 dibawah ini :



Gambar 3. *worksheet* jumlah mahasiswa

B. *Worksheet* status mahasiswa

Dalam membuat *worksheet* status mahasiswa, status mahasiswa di visualisasikan berdasarkan nama kelas dengan jumlah mahasiswa yang berada dikelas tersebut. Filter dan warna berdasarkan tahun masuk dan status mahasiswa. Nama kelas kita menggunakan Stacked Bars, seperti gambar 4 dibawah ini :



Gambar 4. *worksheet* status mahasiswa

C. Worksheet nama mahasiswa

Dalam membuat *worksheet* nama mahasiswa, status nama mahasiswa di visualisasikan berdasarkan nim, nama, nama kelas, sesi dan status. Filter berdasarkan tahun masuk dan status mahasiswa. Data mahasiswa menggunakan table sheet, seperti gambar 5 dibawah ini :

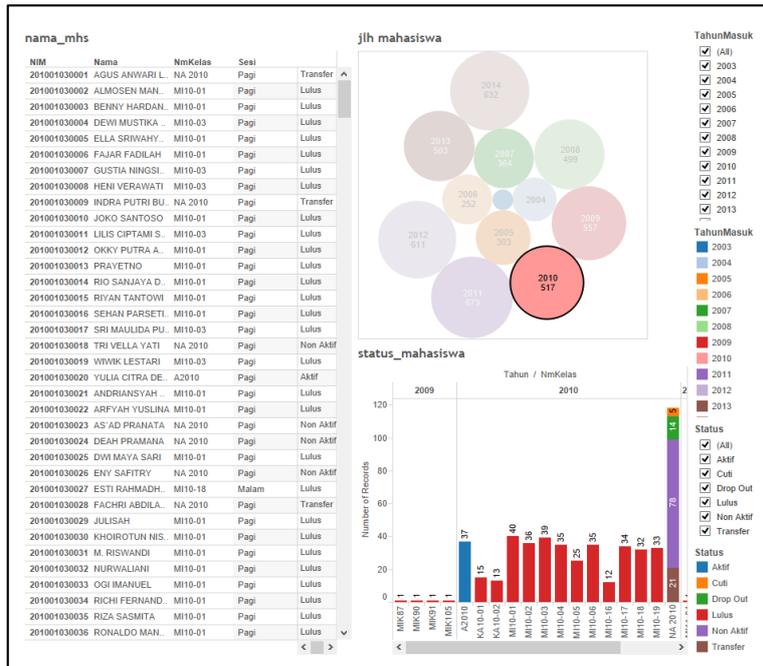
NIM	Nama	NmKelas	Sesi	Status
201001030001	AGUS ANWARI L. TOBING	NA 2010	Pagi	Transfer
201001030002	ALMOSEN MANOGI GRAZIO..	MI10-01	Pagi	Lulus
201001030003	BENNY HARDANI HUTAPEA	MI10-01	Pagi	Lulus
201001030004	DEWI MUSTIKA SARI	MI10-03	Pagi	Lulus
201001030005	ELLA SRIWAHYUNI	MI10-01	Pagi	Lulus
201001030006	FAJAR FADILAH	MI10-01	Pagi	Lulus
201001030007	GUSTIA NINGSIH BATUBARA	MI10-03	Pagi	Lulus
201001030008	HENI VERAWATI	MI10-03	Pagi	Lulus
201001030009	INDRA PUTRI BUDIASIH RA..	NA 2010	Pagi	Transfer
201001030010	JOKO SANTOSO	MI10-01	Pagi	Lulus
201001030011	LILIS CIPTAMI SUYANTI SIA..	MI10-03	Pagi	Lulus
201001030012	OKKY PUTRA ARIYANA	MI10-01	Pagi	Lulus
201001030013	PRAYETNO	MI10-01	Pagi	Lulus
201001030014	RIO SANJAYA DAMANIK	MI10-01	Pagi	Lulus
201001030015	RIYAN TANTOWI	MI10-01	Pagi	Lulus
201001030016	SEHAN PARSETIA	MI10-01	Pagi	Lulus
201001030017	SRI MAULIDA PURWANING..	MI10-03	Pagi	Lulus
201001030018	TRI VELLA YATI	NA 2010	Pagi	Non Aktif
201001030019	WIWIK LESTARI	MI10-03	Pagi	Lulus
201001030020	YULIA CITRA DEWI	A2010	Pagi	Aktif
201001030021	ANDRIANSYAH RUKMAN	MI10-01	Pagi	Lulus
201001030022	ARFYAH YUSLINA	MI10-01	Pagi	Lulus
201001030023	AS'AD PRANATA	NA 2010	Pagi	Non Aktif
201001030024	DEAH PRAMANA	NA 2010	Pagi	Non Aktif

Gambar 5. *Worksheet* nama mahasiswa

4.2. Analisis Visualisasi

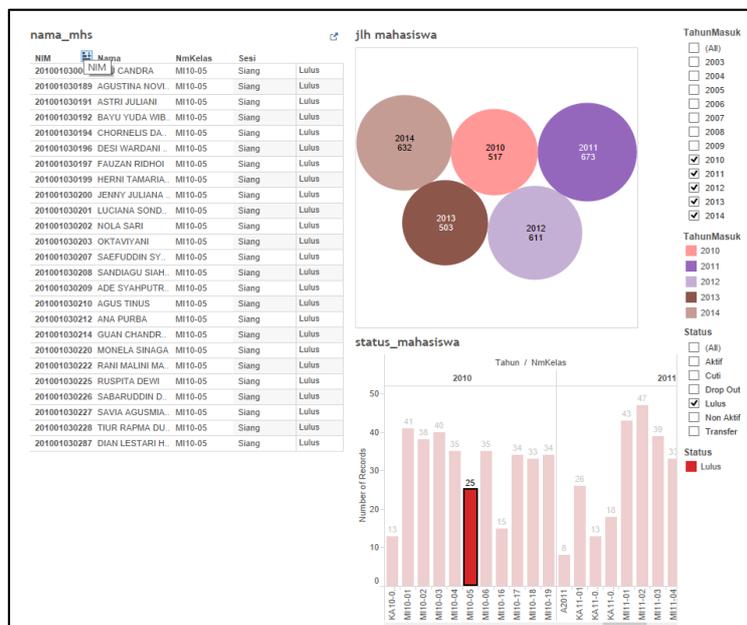
Untuk mendapatkan hasil dari gabungan bebarapa *worksheet*, langkah selanjutnya agar lembar kerja bekerja secara optimal terhadap data yang digunakan langkah selanjutnya memindahkan semua *worksheet* kedalam *dashboard*. Sebuah *dashboard* adalah kumpulan dari beberapa lembar kerja dan informasi pendukung yang ditampilkan dalam satu tempat sehingga dapat membandingkan dan memantau berbagai data secara bersamaan. Misalnya, kita memiliki satu set pandangan bahwa analisis data diperiksa setiap hari. Daripada membalik setiap lembar, kita dapat membuat sebuah *dashboard* yang menampilkan semua pandangan sekaligus. Setiap tampilan yang di tambahkan ke *dashboard* terhubung ke lembar kerja yang sesuai. Bahwa berarti bila mengubah *worksheet*, *dashboard*

diperbarui dan ketika mengubah tampilan di *dashboard*, *worksheet* diperbarui [16], seperti pada gambar 6 di bawah ini.



Gambar 6. Dashboard data mahasiswa

Dengan adanya *dashboard* dan dapat dibaca oleh *Tableau* reader maka manajemen dapat dengan cepat mengetahui jumlah mahasiswa perangkatan, status mahasiswa berdasarkan aktif, cuti, lulus, non aktif, transfer dan drop out. Manajemen juga dapat dengan cepat nim, nama sesi dan status dari data akademik dari nim dan nama mahasiswa sebagai kunci primary key seperti pada gambar 7 dibawah ini



Gambar 7. Dashboard data mahasiswa lulus

Dashboard pada gambar 7 diatas dapat menampilkan dengan cepat data mahasiswa lima tahun terakhir yang sudah lulus dengan menampilkan nama mahasiswa di table sheet nama mahasiswa. Cara kerja menampilkan diatas adalah :

1. Berikan check list lima tahun terakhir dari tahun 2014 sampai dengan 2010 di *worksheet* jlh mahasiswa.
2. Pilih status lulus di *worksheet* status mahasiswa.
3. Pilih grafik bar yang menunjukkan jumlah mahasiswa di kelas tersebut yang telah lulus dan akan menampilkan nim, nama mahasiswa secara lengkap dan cepat seperti pada gambar dibawah ini

4. SIMPULAN

Kesimpulan yang dapat ditarik dari studi ini adalah :

1. Visualisasi dalam bentuk laporan grafik dapat mengoptimalkan permasalahan dalam manajemen pendidikan.
2. Permasalahan dalam ilmu komputer dengan teknologi *Big data* dapat dengan cepat menghasilkan analisis yang memberikan pengetahuan baru terhadap *Big data*.
3. Arsitektur *Tableau* dapat menyelesaikan dan mengoptimalkan permasalahan data yang besar dalam lingkungan manajemen bisnis.

DAFTAR PUSTAKA

- [1] S. Kaisler, F. Amour, J. a. Espinosa dan W. Money, "*Big data : Issues and Challenges Moving Forward*," *Hawaii International Conference on System Sciences*, vol. 46 th, pp. 995-999, 2013.
- [2] 4syth.com, "For *Big data* Analytics There's No Such Thing as Too Big The Compelling Economics and Technology of *Big data* Computing," *Emerging Big data thought White Paper*, March 2012.
- [3] ULARU, E. Geanina, PUICAN, F. Camelia, A. Anca dan V. Manole, "Perspectives on *Big data* and *Big data* Analytics," *Database Systems Journal*, vol. III, no. 4, pp. 3-13, 2014.
- [4] N. Sibley, *Using Tableau for Visual Analytics in Libraries*, Simmons Colleg, 2013.
- [5] S. A. Murphy, "Data Visualization and Rapid Analytics: Applying *Tableau* Desktop to Support Library Decision-Making," *Journal of Web Librarianship*, no. 7:4, pp. 465-476, 2013.
- [6] Zhao, D. W. Tian dan D. Yong, *Optimized Cloud Resource Management and Scheduling*, Elsevier Inc, 2015.
- [7] R. Ranjan, "Modeling and Simulation in Performance Optimization of *Big data* Processing Frameworks," dalam *IEEE Cloud Computing*, IEEE Computer Society, 2014, pp. 76-81.
- [8] C. Eaton, D. Dirk, D. Tom, L. George dan Z. Paul, *Understanding Big data*, Mc Graw Hill.
- [9] E. Dumbill, *Big data Now Current Perspective*, O'Reilly Media, 2012.
- [10] D. Du, A. li dan L. Z, "Survey on the Application Of *Big data* in Chinese real Estate Enterprise", "*International Conference on data Science ICDS*, pp. 24-33, 2014.
- [11] A. J. R. C dan N. Jhonson, "*Big data* and Transformational Government," *IEEE*, 2013.
- [12] F. D. C. a. Xhafa, "Intellegent Services For *Big data* Science," *Future Generation Computer system*, pp. 198-209, 2014.
- [13] F. M, H. W dan L. M, "*Big data* Analytics For Knowledge Generation in Tourism Destination A case from Sweden," *Journal of Destination Marketing and Management*, pp. 198-209, 2014.

- [14] Y. S. D, "A *Big data* approach to HIV epidemiology and prevention," *Preventive Medicine*, pp. 17-18, 2015.
- [15] S. Kandel, A. Paepcke, J. M. Hellerstein dan a. J. Heer, "Enterprise Data Analysis and Visualization: An Interview Study," *IEEE Transactions On Visualization And Computer Graphics*, vol. 18, no. 12, pp. 2917-2926, 2012.
- [16] C. T. Center, "*Tableau* Desktop Fundamental," dalam *Tableau Classroom Training*, Jakarta, Cybertrend Training Center , 2015.
- [17] S. C., T. D. dan H. P. Polaris, "A System for *Query*, Analysis,," *IEEE Trans. on Visualization and Comp. Graphics and Visualization of Multidimensional Relational Databases*, vol. 8, no. 1, pp. 52-65, 2002.
- [18] J. Heer, J. D. Mackinlay, C. Stolte dan a. M. Agrawala, "Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation," *IEEE Transactions On Visualization And Computer Graphics*, vol. 14, no. 6, pp. 1189-1996, 2008.