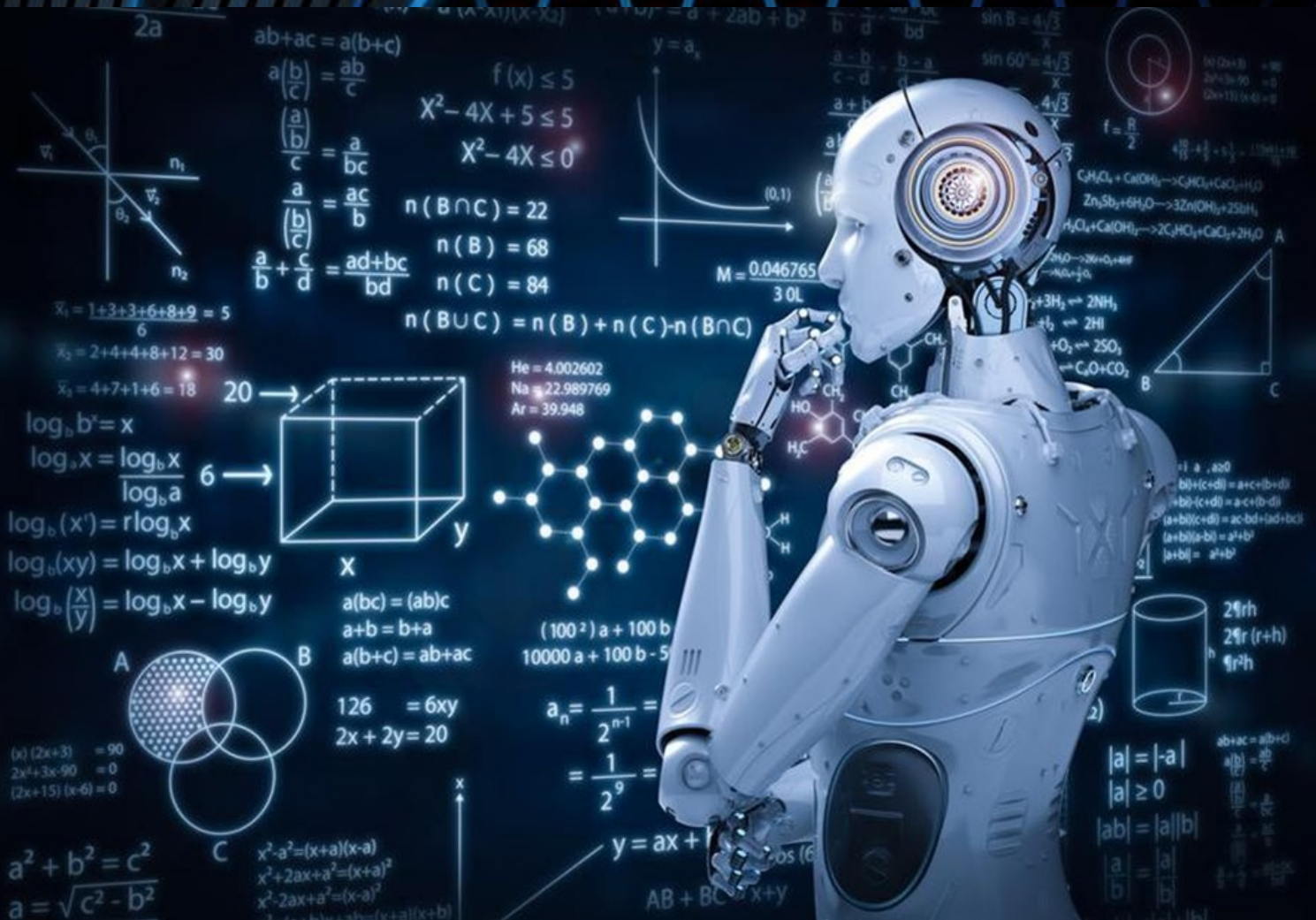


B. PERJASA

BULETIN PERSATUAN JURUANALISA SISTEM SEKTOR AWAM

PERCUMA UNTUK AHLI

BIL 1/2021



Topik Menarik

- Proses Analisis Data
- Jenis-jenis Analitik Data
- Perisian Berkaitan Analitik Data
- *Enhancing Application Analytics with Pandas*
- Keperluan Membina Tahap Ketersediaan Yang Tinggi (HA) Ke Atas Server Virtual Siri 2
- *Cyberattack Attribution: An Overview To Threat Intelligent*
- Meningkatkan Tahap Efisiensi Pusat Data Melalui Pengurusan Aliran Udara (*Airflow*)
- Indeks Pangkalan Data dan *Execution Plan*



ISI KANDUNGAN

Topik	Muka Surat
Perutusan Presiden	ii
Dari Meja Editor	iii
Sidang Redaksi	iii
Ahli Jawatankuasa PERJASA	iii
Aktiviti PERJASA	1 - 9
Program Kerjasama PERJASA	10
Proses Analisis Data	11
Jenis-jenis Analitik Data	13
Perisian Berkaitan Analitik Data	14
Memfaatkan Analitik Data Dalam Pengurusan Institusi Pengajian Tinggi	16
<i>Enhancing Application Analytics with Pandas</i>	20
Keperluan Membina Tahap Ketersediaan Yang Tinggi (HA) Ke Atas <i>Server Virtual</i> Siri 2	23
Amalan Terbaik Bagi Panduan Pengaturcaraan Selamat Dalam Persekitaran JSP	28
<i>Cyberattack Attribution: An Overview To Threat Intelligent</i>	31
Meningkatkan Tahap Efisiensi Pusat Data Melalui Pengurusan Aliran Udara (<i>Airflow</i>)	33
Indeks Pangkalan Data dan <i>Execution Plan</i>	35

PERUTUSAN PRESIDEN

Saya bersyukur ke hadrat Illahi di atas penerbitan Buletin PERJASA yang pertama untuk tahun 2021 ini.

Sesungguhnya kami selaku Jawatankuasa PERJASA 2021-2023 akan meneruskan dan menambah baik legasi Persatuan kita ini agar mencapai objektifnya. Saya percaya usaha dan kerja keras kurang bermakna andai dokumentasi dan komunikasi tidak dilaksanakan.

Oleh itu, Buletin PERJASA inilah antara *medium* komunikasi Persatuan dan peluang untuk ahli berkongsi ilmu dan pengalaman. Saya mengucapkan tahniah kepada barisan *editorial* dan juga ahli PERJASA yang sudi menyumbangkan artikel untuk Buletin PERJASA ini.

Saya berharap ia akan menjadi wadah perkongsian ilmu di kalangan warga teknologi maklumat.

Hanissull Jalis Binti Mohd Yusof
Presiden PERJASA 2021-2023



DARI MEJA KETUA EDITOR



Alhamdulillah syukur ke hadrat Illahi, akhirnya Buletin PERJASA edisi ini dapat diterbitkan.

Edisi kali ini menampilkan perkongsian yang hebat dan menarik oleh ahli-ahli PERJASA dan juga warga Skim Perkhidmatan Teknologi Maklumat Sektor Awam.

Edisi kali ini menampilkan sudut Laporan Khas Analisis Data yang memfokuskan artikel-artikel yang berkaitan dengan analisis data raya. Antara topik-topik perkongsian adalah yang berkaitan dengan Pengurusan Aliran Udara Pusat Data, *Threat Intelligence*, Pegaturcaraan Selamat dan juga laporan aktiviti PERJASA.

Artikel sambungan dari buletin yang lalu seperti Keperluan Membina Tahap Ketersediaan Yang Tinggi (HA) Ke Atas Server Virtual Siri 2 serta Indeks Pangkalan Data dan *Execution Plan* menunjukkan kesinambungan penerbitan Buletin PERJASA.

Semoga semua perkongsian ini menjadi manfaat kepada semua pembaca.

Kami di sidang editorial sangat menghargai sumbangan artikel oleh semua pihak. Kami juga berharap di keluaran akan datang, lebih banyak artikel yang lebih berkualiti dapat diterbitkan.

Ts. Mohd Naim Mohd Ibrahim
Ketua Editor Buletin PERJASA

SIDANG REDAKSI

Ketua Editor

Ts. Mohd Naim Mohd Ibrahim

Editor

Ts. Dr. Mohd Farizul Mat Ghani
Encik Mohd Firdaus Mohd Khairi

JAWATANKUASA PERJASA

Senarai Ahli Jawatankuasa Sesi 2021-2023

Presiden: Puan Hanissull Jalis Md Yusof (MAMPU)

Naib Presiden : Ts. Adi Azlan Mohd Ali (MAMPU)

Setiausaha : Ts. Wan Amishah Wan Mahmud (MAMPU)

Bendahari : Puan Rini Yudesia (JAKIM)

Penolong Setiausaha : Puan Felicia Chua Swee Suan (MAMPU)

AJK :

Ts. Mohd Naim Mohd Ibrahim (KDN)

Ts. Muzamir Mokhtar (JANM)

Puan Dorahasra Ramli (INTAN)

Encik Mohd Fairuz Mohd Yusof (ICU JPM)

Ts. Lindamaswana Masbahon (SPRM)

Ts. Mazwin Muhamad Yusof (BHEUU)

Ts. Masrulehsan Mamat (JPN Pahang)

Encik Zulhilmi Che Azizan (SUK Selangor)

Puan Roshaimieza Mat Adam (MAMPU)

Encik Farid Iqbal Ibrahimdam (MAMPU)

Ts. Hariadi Hinta (JPA)

Pemeriksa kira-kira:

Ts. Dr. Mohd Farizul Mat Ghani (KPM)

Puan Maizatul Sheela Mohamad (Jabatan Pendaftaran Negara)

AKTIVITI PERJASA MESYUARAT AGONG TAHUNAN KALI KE - 22

Mesyuarat Agong Tahunan PERJASA Kali Ke - 22 telah diadakan pada 7 April 2021. Seramai 225 orang ahli telah hadir secara maya. Seramai 10 orang telah dipilih untuk menerajui Jawatankuasa PERJASA yang baharu dan dua (2) orang pemeriksa kira-kira. Barisan Jawatankuasa yang baharu diterajui oleh Puan Hanissull Jalis sebagai Presiden dan YBr. Ts. Adi Azlan sebagai Naib Presiden. Mesyuarat Agong kali ini diserikan dengan persembahan kugiran yang diketuai oleh YBr. Tuan Haji Azih bin Yusof, Timbalan Ketua Pengarah (ICT) MAMPU.



AKTIVITI PERJASA

KUNJUNGAN HORMAT KEPADA KETUA PENGARAH MAMPU DAN TIMBALAN KETUA PENGARAH MAMPU (ICT)



PERJASA amat berbesar hati atas kesudian YBhg. Datuk Seri Dr. Yusof Ismail, Ketua Pengarah MAMPU merangkap Ketua Perkhidmatan ICT Sektor Awam menerima kunjungan hormat PERJASA pada 27 April 2021 di MAMPU Putrajaya.



Jawatankuasa PERJASA Sesi 2021/2023 yang baru dilantik telah mengadakan kunjungan hormat kepada YBrs. Tuan Azih bin Yusof, Timbalan Ketua Pengarah (ICT), MAMPU merangkap *Government Chief Information Officer (GCIO)* Sektor Awam Malaysia pada 23 April 2021.

Persetujuan Tuan Haji Azih sebagai Penasihat PERJASA memberi makna yang besar kepada PERJASA dan tahniah di atas pelantikan ini.

AKTIVITI PERJASA

SUMBANGAN KEBAJIKAN

Dalam memenuhi objektif penubuhan PERJASA iaitu menjaga kebajikan ahli-ahlinya, PERJASA turut cakna dalam membantu masyarakat melalui sumbangan ikhlas ahli-ahlinya. Beberapa rumah kebajikan dan sekolah telah dipilih untuk menerima sumbangan. PERJASA diwakili oleh YBrs. Ts. Lindamaswana selaku Ketua Biro Kebajikan.



6 Mei 2021 - Sumbangan kepada Rumah Amal Limpahan Kasih, Puchong, Selangor

6 Mei 2021 - Sumbangan kepada Rumah Kebajikan Noor Manzil, Puchong, Selangor



9 Mei 2021 - Sumbangan kepada SK Bandar Batu Pahat



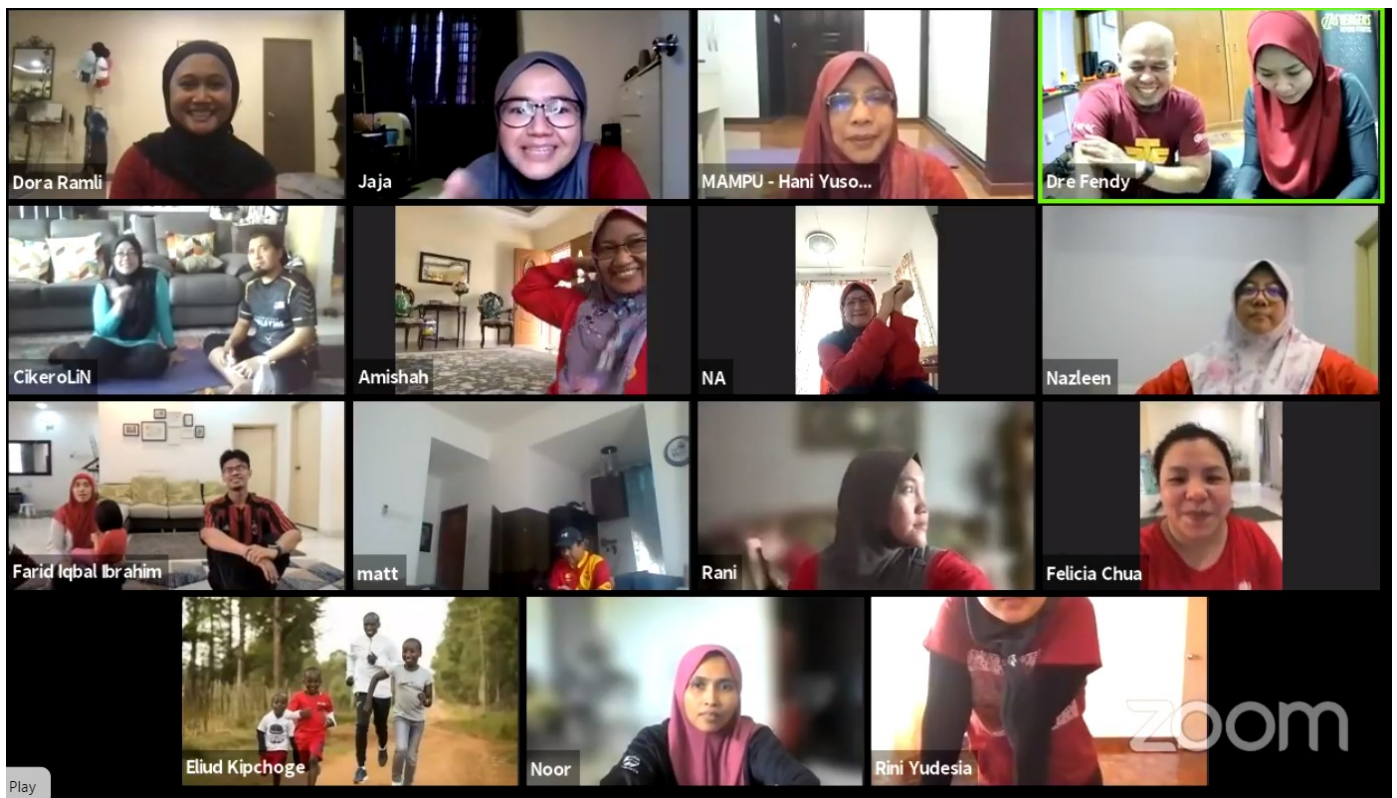
Pada 1 Ogos 2021, En. Abdullah Zaid bin Mohd. Amran Pegawai Teknologi Maklumat Gred F41 yang bertugas di Bahagian Audit ICT, Kementerian Pendidikan Malaysia telah Kembali ke rahmatullah. Sumbangan telah diberikan melalui PERJASA telah disalurkan kepada waris arwah.



AKTIVITI PERJASA

VIRTUAL WORKOUT BERSAMA PERJASA 2021

Walaupun pandemik wabak Covid-19 sedang melanda dunia dan seluruh Malaysia sedang menjalani Perintah Kawalan Pergerakan, namun kesihatan tidak boleh diabaikan. Semua PTM haruslah sihat dan cergas untuk melaksanakan tugas secara Bekerja Dari Rumah. PERJASA telah mengambil inisiatif untuk menganjurkan program Virtual Workout Bersama PERJASA 2021 yang dibimbing oleh *Coach Linda & Coach Dref* pada 5 Jun 2021. Program ini telah diikuti oleh hampir 30 orang peserta secara maya.



AKTIVITI PERJASA WEBINAR

Peningkatan kemahiran dan profesionalisme Ahli PERJASA khususnya dan Pegawai Teknologi Maklumat amnya adalah menjadi matlamat utama PERJASA. Demi mencapai matlamat itu, pelbagai program webinar telah dianjurkan oleh PERJASA. Antaranya adalah :

1st Government Rendezvous



1st Government Rendezvous oleh MAMPU adalah sebuah acara anjuran Unit Pemodenan Tadbiran dan Perancangan Pengurusan Malaysia (MAMPU), Jabatan Perdana Menteri. Program ini melibatkan penyedia perkhidmatan awam seperti AWS, Microsoft, Google Cloud, TM One dan Awantec. *Malaysia Digital Economy Malaysia Corporation* (MDEC) telah menjadi rakan rasmi dan PERJASA sebagai penyedia platform. Pelbagai topik yang berkaitan dengan perkomputeran awan telah dibentangkan dan dibentang dan dibincang sepanjang 2 hari seminar yang berlangsung pada 24 dan 25 Mei 2021.

Pengenalan Kepada Organisasi Moden, Cloud Computing, Strategi Cloud First

Pengenalan kepada organisasi moden, *cloud computing*, strategi *Cloud First*

Siri Webinar MyDigital

Selasa, 8 Jun 2021
10:00 pagi - 1:00 ptg
Daftar - <https://bit.ly/akademi-alpha-connect>

Dibawa kepada anda oleh:

AKADEMI ALPHA CONNECT

Cloud cloud

Akademi Alpha Connect telah bekerjasama dengan PERJASA untuk menganjurkan webinar Pengenalan kepada Organisasi moden, *cloud computing* dan strategi *cloud first* pada 8 Jun 2021 yang lalu.

AKTIVITI PERJASA WEBINAR

LESSONS LEARNT DARI PELAKSANAAN APLIKASI TERPILIH DI PUBLIC CLOUD

Penceramah : Ts Wan Amishah (MAMPU)
Platform : Airmeeet MeetUp
Tarikh : 12 Jun 2021 (Sabtu)
Masa : 11.30 pg – 12 tghr
Live Viewers : 15 (Airmeeet) & 66 (Youtube)

Tiga aplikasi Kerajaan yang berskala besar telah menggunakan perkhidmatan *Public Cloud* sepenuhnya. Apakah cabaran dan persiapan setiap aplikasi berkenaan? Apakah Faktor memilih perkhidmatan *cloud* berbanding *self-hosting*? Bagaimana mereka bermula? Ketahui jawapan melalui program ini.



BUILDING THE ePAYMENT SYSTEM OF TOMORROW

Penceramah : En Faizal & En Luqman (PayNet)
Platform : Airmeeet MeetUp
Tarikh : 14 Jun 2021 (Isnin)
Masa : 10.00 pg – 11.30 pg
Live Viewers : 111 (Airmeeet) & 82 (Youtube)



The session will emphasize on empowering Malaysia's digital economy and accelerate adoption of epayment by introducing a national shared payment infrastructure which facilitates payments, collections and related messaging services.

EVOLVING CYBER SECURITY LANDSCAPE

Penceramah : Ts Adnan Shukor (Celcom)
Platform : Airmeeet MeetUp
Tarikh : 19 Jun 2021 (Sabtu)
Masa : 9.00 mlm – 10.30 mlm
Live Viewers : 36 (Airmeeet), 29 (Youtube) & 10 (FB)

Cyber security is a fast-moving sector, as both cybercriminals and security providers vie to outsmart each other. New threats – and innovative ways to combat them – emerge all the time. In this presentation, we will explore the current trends in cyber security and some highlight of incident reported in Malaysia.



AKTIVITI PERJASA WEBINAR

Payment Gateway Ala Shopee: Men- contohi PBTPay

Penceramah : En Muhamad Shukri &
En Mohamad Zubir (JKT,KPKT)
Platform : Airmeeet MeetUp
Tarikh : 22 Jun 2021 (Selasa)
Masa : 10.00 pg – 1 ptg
Live Viewers : 197 (Airmeeet),10 (Youtube) &
10 (FB)



PBTPay adalah sebuah portal pembayaran sepusat rasmi bagi Pihak Berkuasa Tempatan (PBT) yang dibangunkan oleh Jabatan Kerajaan Tempatan (JKT) di bawah Kementerian Perumahan dan Kerajaan Tempatan (KPKT). Ianya mematuhi Pekeliling Perbendaharaan PK2.2 dan telah mendapat kelulusan dari Jabatan Akauntan Negara. Lanjut di <https://pbtpay.kpkt.gov.my/>

DARI "DIGITAL FORENSICATOR" KEPADA "CYBER FORENSICATOR" - APAKAH BEZANYA?



Penceramah : Muhammad Muslim (Celcom)
Platform : Airmeeet MeetUp
Tarikh : 3 Jul 2021 (Sabtu)
Masa : 9.00 mlm – 11.00 mlm
Live Viewers : 42 (Airmeeet),40 (Youtube) &
23 (FB)

Terma Forensik Digital dan Forensik Siber makin kerap diperkatakan. Malah ramai yang teruja apabila watak - watak seperti Abby Schiuto (NCIS), Horatio Caine (CSI) diketengahkan dengan pelbagai kes berkaitan jenayah termasuk jenayah siber & digital dapat diselesaikan apabila pelbagai bukti dianalisa. Perkongsian di webinar kali ini akan mengupas secara santai evolusi kerjaya penganalisa forensik digital/ siber, berbincang sisi perbezaan antara kedua terma, dan apa yang diperlukan kepada mereka yang berminat menceburi.

PELAN STRATEGIK PENDIGITALAN SEKTOR AWAM 2021-2025

Penceramah : Puan Suhaiza binti Sudirman (MAMPU)
Platform : Airmeeet MeetUp
Tarikh : 9 Jul 2021 (Jumaat)
Masa : 10.00 pg – 12 tghr
Live Viewers : 171 (Airmeeet), 154 (Youtube) & 21 (FB)



Sesi Pelan Startegik Pendigitalan Sektor Awam 201-2025.
Apakah yang ingin diberi penekanan?

AKTIVITI PERJASA WEBINAR

URBAN FARMING

Penceramah : Fatin Farhana (Kebun Komuniti Rajawali)
Platform : Airmeeet MeetUp
Tarikh : 10 Jul 2021 (Sabtu)
Masa : 10.30 pg – 12 tghr
Live Viewers : 22 (Airmeeet), 35 (Youtube) & 11 (FB)



Pendedahan tentang cara pemilihan benih dan semaian, penentuan baja kompos dan seterusnya penerangan tentang konsep tanaman. Urban Farming jadi keperluan rakyat Malaysia terutamanya di waktu pandemik Covid-19.



INSIDE-OUT MYDIGITAL BLUEPRINT

Penceramah : En Fabian Bigar (CEO SCMO)
Platform : Airmeeet MeetUp
Tarikh : 11 Jul 2021 (Ahad)
Masa : 10.00 pg – 1.00 ptg
Live Viewers : 89 (Airmeeet)

SCMO is the agency established under EPU & tasked with delivering the Malaysia Digital Economy Blueprint (MyDigital) launched by the Prime Minister on February 19, 2021 from rhetoric to reality. It is the most important economic plan introduced by the government. It is much more an economic plan than a technology one. A plan to shape up, sharpen the competitiveness of the country, its companies and its people!

BERKAYUH PULANG KE MALAYSIA DARI UK

Penceramah : Mohd Azmeer
Platform : Airmeeet MeetUp
Tarikh : 27 Jul 2021 (Selasa)
Masa : 9.00 pg – 12.00 tghr
Live Viewers : 22 (Airmeeet), 6 (Youtube) & 11 (FB)



Tiga anak Malaysia yang mengembara dengan berbasikal selama 403 hari dengan jarak kira-kira 16,000 km, merentasi 13 negara menjadi bualan hangat pada tahun 2019. Tiga anak muda itu menzahirkan rasa bertuah dan syukur kerana dapat kembali ke tanah air setelah memulakan pengembaraan dari United Kingdom (UK) ke Malaysia selama lebih setahun.

AKTIVITI PERJASA WEBINAR

AKTA PELINDUNGAN DATA PERIBADI 2010 BERSAMA JPDP, KKMM

Penceramah : Sharil Mazmi (JPDP)
Platform : Google Meet
Tarikh : 4 Ogos 2021 (Rabu)
Masa : 2.00 ptg – 4.00 ptg
Live Viewers : 120 (GMeet) & 23 (Youtube)



Penjelasan tentang Akta Perlindungan Data Peribadi 2010 oleh wakil ICT di Jabatan Perlindungan Data Peribadi, KKMM

INISIATIF PLATFORM E-PEMBAYARAN SEKTOR AWAM MELALUI PELAKSANAAN SISTEM TERIMAAN ELEKTRONIK KERAJAAN PERSEKUTUAN (IPAYMENT)



Penceramah : En Azhar Fairuzz (JANM)
Platform : Airmeet MeetUp
Tarikh : 23 Sept 2021 (Khamis)
Masa : 9.30 pg – 11.00 pg
Live Viewers : Airmeet, Youtube) &
FB

Taklimat ini menerangkan inisiatif iPayment sebagai usaha kerajaan untuk merealisasikan hasrat government go-cashless dalam sistem penyampaian kepada orang awam. Sebagai agensi yang bertanggungjawab dalam menguruskan perakaunan dan kewangan kerajaan, Jabatan Akauntan Negara Malaysia (JANM) telah merancang inisiatif ini bermula pada September 2019 berikutan tren global orang awam yang menggunakan perkhidmatan e-dagang dan perbankan elektronik semakin meningkat. Adalah diharapkan melalui pelaksanaan iPayment ini, urusan kutipan hasil dan bukan hasil sektor awam dapat diuruskan dengan lebih cekap sekaligus membawa sistem penyampaian kerajaan seiring dengan teknologi dalam komuniti tanpa tunai.

AKTIVITI PERJASA

PROGRAM KERJASAMA

PERJASA

Libat Urus Kerjaya Skim F

Sesi libat urus bersama BSA MAMPU bagi pembangunan kerjaya, kepakaran dan kompetensi Skim F pada 21 Mei 2021.

Libat Urus CTSU

Sesi libat urus bersama *Computer Technician Scheme Union* (CTSU) bagi mendengar dan bekerjasama dalam pelbagai aktiviti akan datang pada 11 Jun 2021.

Bengkel Job Description

Terlibat bersama pihak MAMPU bagi Pegawai Klasifikasi Perkhidmatan Sistem Maklumat (F) pada 14 dan 15 Jun 2021

Libat Urus KPPTM

Sesi libat urus bersama Kesatuan Penolong Pegawai Teknologi Maklumat (KPPTM) bagi mendengar dan bekerjasama dalam pelbagai aktiviti akan datang pada 22 Jun 2021.

Virtual Conference Customer

Platform Provider kepada Program Awantec Cloud Day pada 20 Ogos 2021

ANALISIS DATA

Analisis data didefinisikan sebagai proses pembersihan, transformasi, dan pemodelan data untuk menemukan informasi yang berguna untuk membuat keputusan perniagaan. Tujuan Analisis Data adalah untuk mengekstrak maklumat yang berguna dari data dan mengambil keputusan berdasarkan analisis data.

PROSES ANALISIS DATA

Ts. Mohd Fitri Edros, Mohamed Roshimi Md Shahor, Ifahana Ishak
Pusat Pengurusan dan Pembangunan Digital, Universiti Malaysia Perlis

Proses Analisis Data adalah proses mengumpulkan maklumat dengan menggunakan aplikasi atau alat yang tepat yang membolehkan anda meneroka data dan mencari corak di dalamnya. Berdasarkan maklumat dan data, anda boleh membuat keputusan, atau anda dapat membuat kesimpulan yang muktamad. Analisis Data terdiri daripada beberapa fasa berikut :



Pengumpulan Data - Selepas mengumpul keperluan, anda akan mendapat idea yang jelas mengenai perkara yang perlu diukur dan apa yang harus menjadi penumpuan. Kini adalah masa untuk mengumpulkan data berdasarkan keperluan tersebut. Setelah mengumpulkan data anda, haruslah diingat bahawa data tersebut perlulah diproses atau disusun untuk di analisis. Semasa anda mengumpulkan data dari pelbagai sumber, anda mesti menyimpan log tarikh pengumpulan dan sumber data.



Mengenalpasti Keperluan Data - Pertama sekali, anda perlulah berfikir mengapa anda mahu melakukan data analisis ini. Anda perlulah mengetahui tujuan atau sasaran dalam melakukan analisis. Anda harus tentukan jenis data analisis yang hendak digunakan. Pada fasa ini juga, anda perlu tentukan apa yang perlu dianalisis dan bagaimana mengukurnya, anda juga harus memahami mengapa anda mengkaji dan mengetahui apa jenis ukuran yang digunakan dalam analisis.



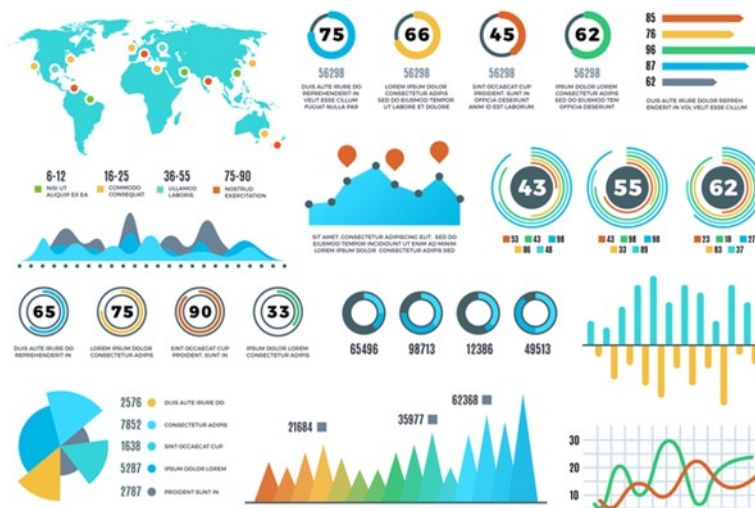
Pembersihan Data - Walau apapun data yang dikumpulkan pada masa ini, ia mungkin ada yang tidak berguna atau tidak relevan dengan tujuan analisis anda maka ia perlu dibersihkan. Data yang dikumpulkan mungkin mengandungi rekod bertindih, ruang kosong atau kesalahan. Data hendaklah dibersihkan dan bebas dari ralat. Fasa ini hendaklah dilakukan sebelum Analisis kerana melalui pembersihan data akan menjadikan hasil data lebih baik seperti yang diharapkan.

Tafsiran Data



Selepas anda menganalisis data, akhirnya tibalah masanya untuk mendapatkan keputusan. seterusnya. Anda boleh memilih cara menyatakan atau menyampaikan analisis data sama ada melalui penulisan mudah atau mungkin menggunakan jadual atau carta. Kemudian gunakan

Visualisasi Data



Visualisasi Data seperti dalam bentuk carta dan graf merupakan visual yang biasa kita gunakan dalam kehidupan harian. Dalam erti kata lain, data yang dipaparkan dalam bentuk grafik amatlah mudah difahami dan diproses oleh pemikiran manusia.

JENIS-JENIS ANALITIK DATA

Ts. Mohd Fitri Edros, Mohamed Roshimi Md Shahor, Ifahana Ishak
Pusat Pengurusan dan Pembangunan Digital, Universiti Malaysia Perlis



Deskriptif

Ini adalah yang paling asas dalam semua bentuk Data Analitik. Ia memberikan gambaran kepada penganalisa tentang metrik dan ukuran utama dalam perniagaan. Sebagai contoh, penyata bulanan untung dan rugi. Begitu juga, seorang penganalisis boleh memiliki data dengan skalar yang lebih besar tentang pelanggan. Memahami informasi demografik berkenaan pelanggan (sebagai contoh: 30% daripada pelanggan kita adalah bekerja sendiri) dapat dikategorikan sebagai Deskriptif Analitik. Menggunakan perisian visualisasi yang berkesan meningkatkan mesej penyampaian yang jelas tentang Deskriptif Analitik.

Diagnostik

Ini adalah langkah seterusnya yang lebih kompleks dalam analitik deskriptif. Pada penilaian data deskriptif, alatan analitik diagnostik akan memberi peluang kepada seorang penganalisis untuk menguraikan dan mengasingkan punca masalah. *Business Intelligent* dengan rekaan terbaik yang menggabungkan pembacaan data siri masa (iaitu data melalui beberapa titik berturut-turut dalam masa yang sama) dan ciri-ciri penapisan beserta kebolehan gerudi ke bawah (*drill down*) membolehkan analisis sedemikian dihasilkan.

Prediktif (Ramalan)

Prediktif Analitik adalah berkenaan jangkaan atau anggaran. Meramalkan sama ada kemungkinan sesuatu kejadian berlaku di masa depan, meramalkan jumlah yang dapat diukur atau menganggarkan titik waktu di mana sesuatu mungkin berlaku - semuanya dilakukan melalui model ramalan. Model ramalan biasanya menggunakan pelbagai data pembolehubah untuk membuat ramalan. Kebolehan komponen data berkait rapat dengan apa yang mungkin diramalkan (sebagai contoh, semakin tua seseorang, semakin mudah untuk mereka mengalami serangan jantung - kita boleh mengatakan bahawa usia mempunyai hubungan linear dengan risiko serangan jantung). Data-data ini kemudian disatukan menjadi skor atau ramalan. Dalam dunia yang tidak menentu, keupayaan membuat ramalan dapat membolehkan seseorang membuat keputusan yang lebih baik. Model ramalan adalah antara model yang paling penting digunakan dalam bidang tertentu.

Preskriptif

Jenis Data Analitik seterusnya yang lebih rumit adalah Preskriptif Analitik. Preskriptif Analitik menggunakan pemahaman tentang apa yang telah terjadi, mengapa ia berlaku dan apa yang mungkin berlaku, untuk membantu pengguna menentukan tindakan terbaik yang harus diambil. Preskriptif Analitik biasanya bukan hanya dengan satu tindakan individu, tetapi sebenarnya melibatkan tindakan bersama beberapa individu lain. Contoh yang baik adalah aplikasi lalu lintas yang membantu anda memilih jalan pulang yang terbaik dengan mengambil kira jarak setiap laluan, kelajuan di mana seseorang boleh melakukan perjalanan di setiap jalan dan, yang paling penting, kekangan lalu lintas semasa. Antara contoh lain ialah menghasilkan jadual waktu peperiksaan yang efektif sehingga tidak ada pelajar yang mengalami pertembungan jadual.

PERISIAN BERKAITAN ANALITIK DATA

Ts. Mohd Fitri Edros, Mohamed Roshimi Md Shahor, Ifahana Ishak
 Pusat Pengurusan dan Pembangunan Digital, Universiti Malaysia Perlis

Dengan meningkatnya permintaan Data Analitik dalam pasaran, banyak peralatan telah muncul dengan pelbagai fungsi dan kegunaan. Sama ada dari sumber terbuka atau mesra pengguna, peralatan terkemuka dalam pasaran data analitik terdiri daripada :

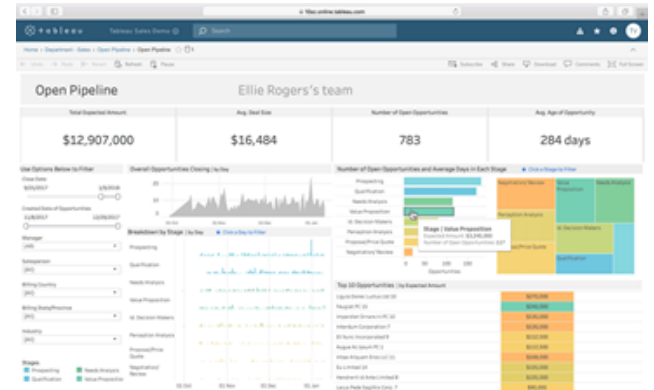
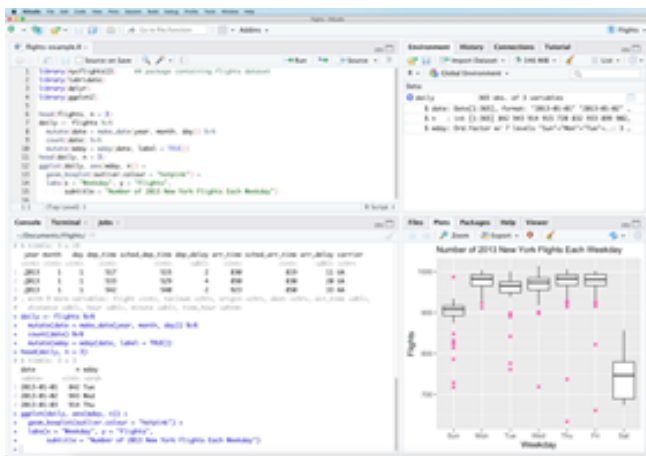


Tableau Public - Ini merupakan perisian percuma yang menyambungkan pada mana-mana sumber data seperti Excel, Gudang data korporat, etc. Ia kemudian mencipta visual, peta, dashboard etc dengan kemaskini pada laman sesawang dalam masa nyata.

R programming - Perisian ini merupakan perisian analitik yang paling terkemuka untuk statistik dan pemodelan data. R beroperasi pada kebanyakan platform seperti UNIX, Windows dan Mac OS. Ia juga menyediakan alat untuk memasang secara automatik semua paket mengikut keperluan pengguna.

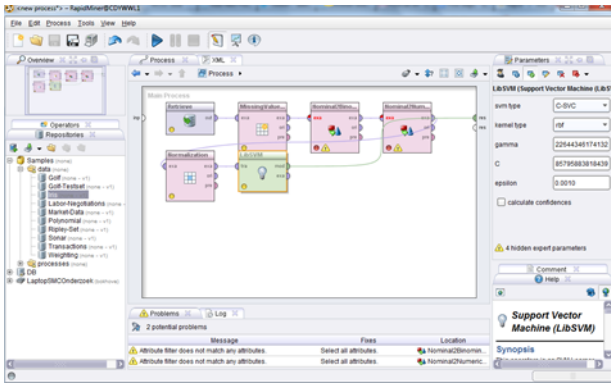
Date	Services	Products	Total	7-Day Avg	Goal
2016-08-19	1345	2789	2789	2,789	2000
2016-08-20	3076	2104	2104	2,104	2000
2016-08-21	1209	321	321	1,609	2000
2016-08-22	4578	507	507	1,520	2000
2016-08-23	3306	0	0	1,514	2000
2016-08-24	3799	0	0	845	2000
2016-08-25	987	416	416	784	2000
2016-08-26	8712	1306	1331	587	2000
2016-08-27	4509	2424	0	200	2000
2016-08-28	6712	0	2330	279	2000
2016-08-29	0	0	2814	985	2000
2016-08-30	0	0	514	996	2000
2016-08-31	0	0	443	1,060	2000
2016-09-01	0	0	1539	1,226	2000
2016-09-02	0	0	3286	1,502	2000

Microsoft Excel - Perisian ini merupakan salah satu daripada alat yang banyak digunakan dalam data analitik. Kebanyakan digunakan untuk data dalaman pelanggan, alat ini menganalisa tugas secara keseluruhan dengan merangkumi jadual pangsi.

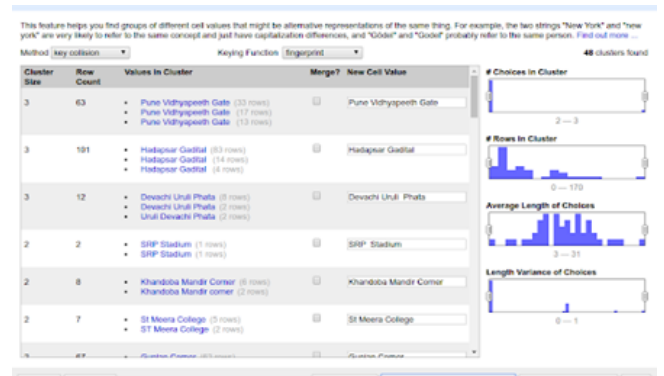
```

31 self.file = None
32 self.fingerprints = set()
33 self.logspaces = True
34 self.debug = debug
35 self.logger = logging.getLogger(__name__)
36
37 if path:
38     self.file = os.path.join(path, 'requests.log')
39     self.file.write('')
40     self.fingerprints.update(requests.log)
41
42 @classmethod
43 def from_settings(cls, settings):
44     debug = settings.getbool('requests.log')
45     return cls(job_dir(settings), debug)
46
47 def request_seen(self, request):
48     fp = self.request_fingerprint(request)
49     if fp in self.fingerprints:
50         return True
51     self.fingerprints.add(fp)
52     if self.file:
53         self.file.write(fp + os.linesep)
54
55 def request_fingerprint(self, request):
56     return request_fingerprint(request)
    
```

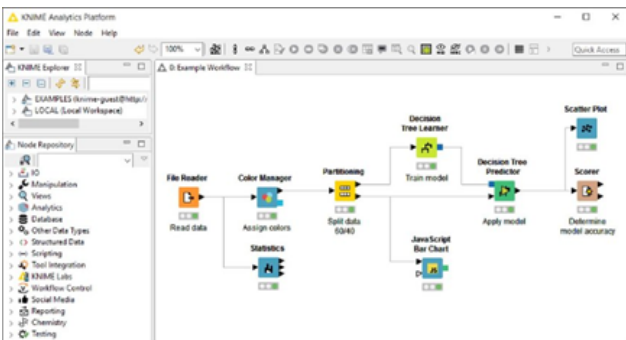
Python - Python merupakan bahasa pengaturcaraan sumber terbuka dan berorientasikan objek yang mudah dibaca, ditulis dan dikekalkan. Ia menyediakan pelbagai visual dan bahasa mesin seperti Scikit-learn, TensorFlow, Matplotlib, Pandas, Keras, etc. Ia juga boleh dipasang pada pelbagai platform seperti SQL server, pangkalan data MangoDB atau JSON.



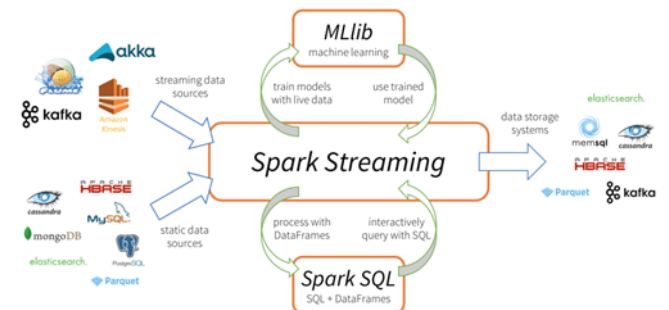
RapidMiner - Satu platform yang kuat dan bersepadu yang dapat disatukan dengan mana-mana sumber data seperti Access, Excel, Microsoft SQL, Tera data, Oracle, Sybase etc. Alat ini kebanyakan digunakan untuk analisis ramalan (analitik predictive) seperti analisis teks, perlombongan data (data mining) dan pembelajaran mesin



OpenRefine - Juga dikenali sebagai GoogleRefine, perisian pembersihan data ini akan membantu anda membersihkan data untuk dianalisis. Ia digunakan untuk membersihkan data yang tidak kemas, menghuraikan dan transformasi data dari laman web.



KNIME - Konstanz Information Miner (KNIME) ialah platform data analitik sumber terbuka di mana menawarkan kepada anda analisis data dan data model.



Apache Spark - Satu daripada enjin pemprosesan data berskala besar. Perisian ini melakukan aplikasi dalam kluster Hadoop 100 kali lebih pantas dalam memori dan 10 kali lebih pantas dalam cakera. Perisian ini juga popular dalam saluran data dan pembangunan model pembelajaran mesin.

MEMANFAATKAN DATA ANALITIK DALAM PENGURUSAN INSTITUSI PENGAJIAN TINGGI

Ts. Mohd Fitri Edros, Mohamed Roshimi Md Shahor, Ifahana Ishak
Pusat Pengurusan dan Pembangunan Digital, Universiti Malaysia Perlis

PENDAHULUAN

Pada era digital, data umpama sumber minyak yang baharu. Samada ia dihasilkan oleh organisasi berskala besar atau dijana oleh individu, ia perlu diproses daripada pelbagai aspek untuk menukarkan data tersebut kepada sesuatu yang boleh memberikan manfaat kepada semua sektor amnya dan sektor pengajian tinggi khususnya. Persoalannya ialah bagaimanakah kaedah untuk menganalisa data. Seterusnya menggunakan hasil analisa tersebut untuk menambahbaik operasi institusi pengajian tinggi. Disinilah munculnya istilah Data Analitik. Artikel ini akan cuba mengupas tentang konsep Data Analitik, jenis-jenis Data Analitik, proses Data Analitik, perisian berkaitan Data Analitik dan kegunaan Data Analitik dalam persekitaran institusi pengajian tinggi.

KONSEP DATA ANALITIK

Data Analitik merujuk kepada teknik dan proses kualitatif dan kuantitatif yang digunakan untuk meningkatkan produktiviti dan prestasi organisasi. Data diekstrak dan dikategorikan untuk mengenalpasti dan menganalisis pola dan corak tingkah laku. Tekniknya berbeza mengikut keperluan organisasi. Data Analitik juga dikenali sebagai analisis data.



Antara peranan Data Analitik ialah seperti berikut:

Mengumpul Pandangan Tersirat (*Hidden Insight*)

Pandangan tersirat daripada data dikumpul dan dibuat analisis mengikut keperluan organisasi.

Menjana Pelaporan

Laporan dihasilkan daripada data dan dihantar kepada individu dan pasukan tertentu untuk tindakan selanjutnya bagi mendapatkan impak yang tinggi dalam pengurusan organisasi.

Melaksanakan Analisis Pasaran

Analisa Pasaran boleh dilaksanakan untuk mengenalpasti kekuatan dan kelemahan pesaing.

Menambahbaik Keperluan Organisasi

Analisis data dapat menambahbaik keperluan dan pengalaman Organisasi kepada Pelanggan.



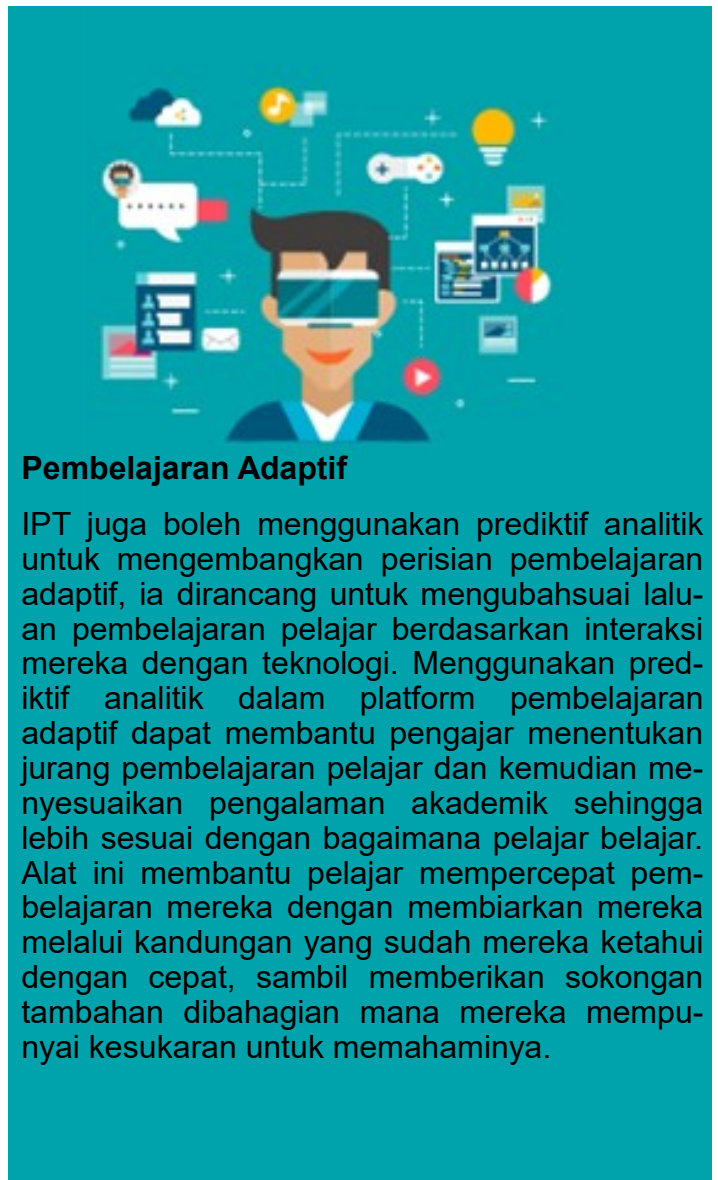
Data Analitik Dalam Persekitaran Institusi Pengajian Tinggi

Institusi pengajian tinggi boleh menggunakan data analitik sebagai cara untuk bertindak balas terhadap pelbagai perubahan organisasi dan operasi yang berlaku dalam sektor pendidikan. Berikut adalah kegunaan data analitik dalam membantu pengurusan Institusi Pengajian Tinggi:



Mengenalpasti Pelajar Yang Memerlukan Khidmat Nasihat

Terdapat IPT yang mempunyai bilangan Penasihat Akademik yang terbatas. Ini mengakibatkan pelajar yang sepatutnya menerima nasihat akademik tidak disenaraikan sebagai mentee kepada Penasihat Akademik. Satu kajian yang dikendalikan oleh *National Academic Advising Association* (NACADA), yang merupakan sebuah persatuan berpangkalan di Amerika Syarikat mendapati bahawa kadar antara Pelajar dan Penasihat Akademik di negara itu ialah 296 nisbah 1. Walaubagaimanapun, perkara ini dapat ditangani dengan penggunaan data analitik jenis prediktif contohnya pembangunan sistem analitik untuk mengenalpasti pelajar yang betul-betul memerlukan bantuan akademik dengan ciri-ciri atau parameter tertentu.



Pembelajaran Adaptif

IPT juga boleh menggunakan prediktif analitik untuk mengembangkan perisian pembelajaran adaptif, ia dirancang untuk mengubahsuai laluan pembelajaran pelajar berdasarkan interaksi mereka dengan teknologi. Menggunakan prediktif analitik dalam platform pembelajaran adaptif dapat membantu pengajar menentukan jurang pembelajaran pelajar dan kemudian menyesuaikan pengalaman akademik sehingga lebih sesuai dengan bagaimana pelajar belajar. Alat ini membantu pelajar mempercepat pembelajaran mereka dengan membiarkan mereka melalui kandungan yang sudah mereka ketahui dengan cepat, sambil memberikan sokongan tambahan dibahagian mana mereka mempunyai kesukaran untuk memahaminya.



Perancangan Pendaftaran Pelajar

IPT juga boleh menggunakan prediktif analitik untuk membantu merancang pengurusan pendaftaran dengan optimum. Ia dapat membantu pihak fakulti untuk meramal bilangan kelas yang diperlukan. Ia juga dapat digunakan untuk menasaskan promosi hanya kepada bakal pelajar yang diramalkan mempunyai kecenderungan yang tinggi untuk mendaftar ke IPT. Selain itu, data analitik dapat menjangkakan keperluan kewangan dan meramal samada pelajar layak atau tidak layak untuk menerima bantuan



Mensasarkan Pelajar Yang Layak Menerima Biasiswa

Menggunakan prediktif analitik untuk mengenalpasti pelajar yang layak menerima biasiswa berdasarkan data kelayakan akademik, data kehadiran, data aktiviti ko-kurikulum dan data pendapatan keluarga. Hasilnya, biasiswa akan diberikan kepada pelajar yang benar-benar layak dan memerlukan bantuan tersebut.



Penambahbaikan Urusan Pengambilan Pelajar

Dengan mewujudkan satu set permodelan yang mengambilkira demografi, sejarah akademik dan faktor lain untuk populasi pelajar yang berbeza (pelajar baru, pelajar berhenti dan sebagainya) terhadap prospek yang berpotensi, institusi pendidikan tinggi dapat mensasarkan pelajar yang mempunyai kebarangkalian tertinggi untuk menerima tawaran dan hadir semasa hari pendaftaran. Ini



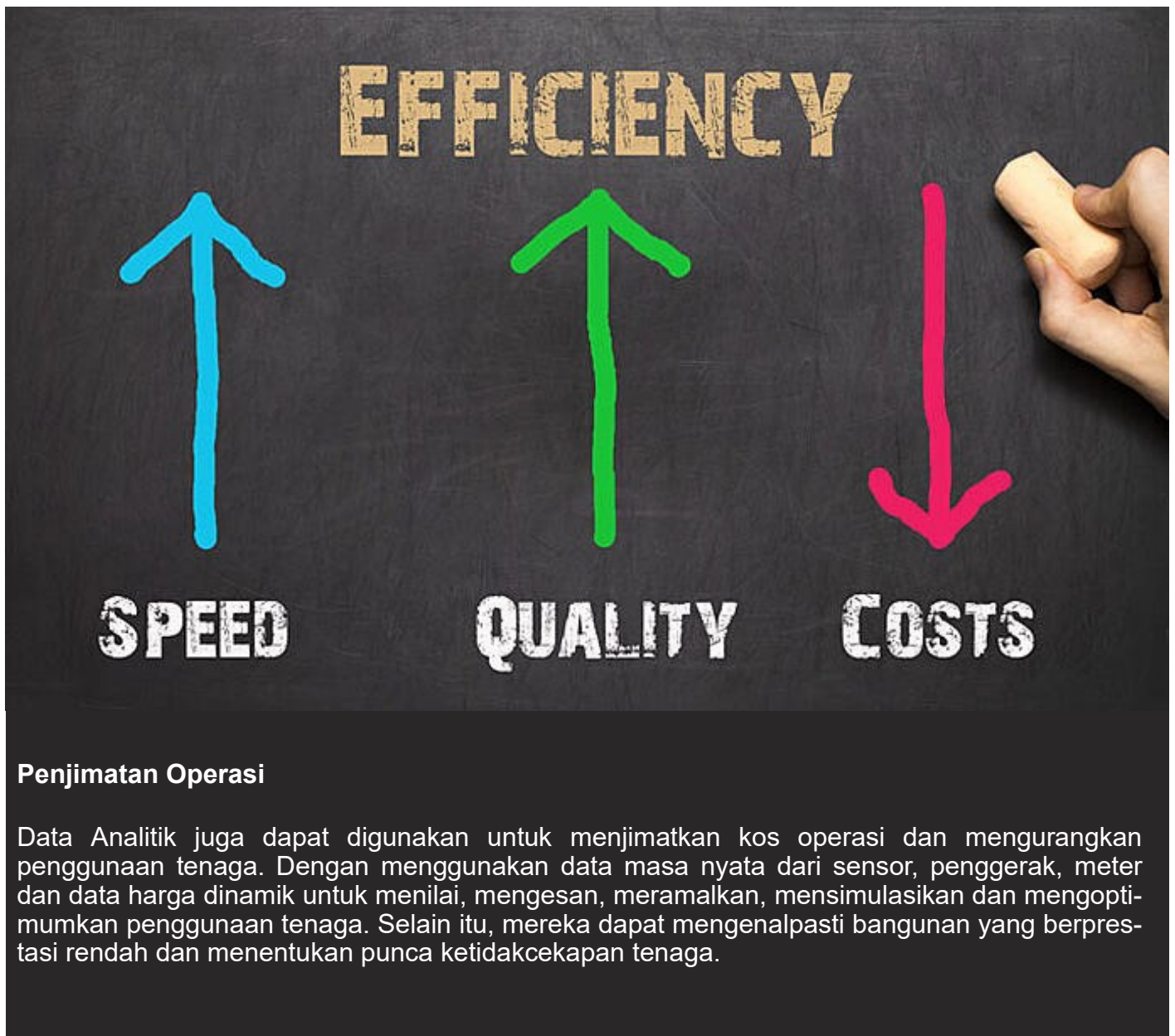
Mengenalpasti Pelajar Berisiko

Institusi boleh mengenalpasti pelajar yang berisiko untuk berhenti sebelum semester baru bermula dengan merujuk sejarah prestasi akademik masa lalu pelajar, rekod kehadiran ke kelas atau perpustakaan dan rekod prestasi aktiviti ko-kurikulum pelajar. Setelah dikenalpasti, penasihat akademik boleh mengambil langkah untuk memasukkannya pelajar tersebut ke kelas yang akan mempersiapkan pelajar dengan kelas tambahan atau motivasi.



Mengenalpasti Penderma dan Pelabur

Dalam penjanaan kewangan, institusi pendidikan tinggi dapat menggunakan analisis data untuk memeriksa pangkalan data alumni mereka dan mengenalpasti individu tertentu dengan kecenderungan terbesar untuk memberikan sumbangan dan dengan potensi terbesar untuk meningkatkan sumbangan mereka ketika kerjaya mereka sedang berkembang. Demikian juga, prediktif analitik dapat diterapkan untuk mengesan alumni yang mungkin tidak berminat untuk menyumbang. Analisis preskriptif dapat menawarkan cara untuk memastikan mereka menyumbang secara aktif dan berterusan. Hasilnya, lebih banyak penderma memberikan lebih banyak wang dalam jangka masa panjang.



Penjimatan Operasi

Data Analitik juga dapat digunakan untuk menjimatkan kos operasi dan mengurangkan penggunaan tenaga. Dengan menggunakan data masa nyata dari sensor, penggerak, meter dan data harga dinamik untuk menilai, mengesan, meramalkan, mensimulasikan dan mengoptimumkan penggunaan tenaga. Selain itu, mereka dapat mengenalpasti bangunan yang berprestasi rendah dan menentukan punca ketidakcekapan tenaga.

PENUTUP

Kesimpulannya, data analitik merupakan suatu konsep yang sangat penting dan memberi manfaat yang besar kepada IPT. Ia dapat digunakan dalam pelbagai aspek dan permasalahan dalam membantu pengurusan IPT. Terdapat pelbagai perisian yang boleh digunakan untuk mengaplikasikan dan merealisasikan konsep data analitik. Data Analitik ini, jika digunakan dengan kaedah atau proses yang betul, ia mampu untuk membantu pihak pengurusan IPT dalam membuat keputusan berpacuan data. Dengan kebolehan untuk membuat keputusan yang lebih tepat, seterusnya akan melonjakkan prestasi IPT secara keseluruhan ke tahap yang lebih tinggi

Enhancing Application Analytics with Pandas

Samsuddin Wira Samsuddin
Bahagian Pengurusan Maklumat,
Kementerian Pertahanan Malaysia

Pandas (<https://pandas.pydata.org/about/index.html>) is one of the greatest data analytic tools that has been widely used all over the world. So it comes as no surprise that this project is currently sponsored by Mark Zuckerberg himself through the Chan Zuckerberg Foundation.

One of the features that makes it great and powerful is High Performance. The backend of this library is written in C which makes all instructions go directly to the machine. Hence, data slicing, merging, dropping, searching, indexing, subsetting, aggregating, reshaping, etc can be executed instantaneously. As for this article, we will highlight this characteristic based on our experience and how pandas has impacted our application accomplishment (iMAS to be exact).

Mindef Auditing System (or also known as iMAS) is a system that is specifically created for auditing EKSA achievement within the Ministry of Defence. Previously, the auditing was executed manually where all auditors will get the audit questions, print it, go to the designated auditee sites and bring back the results to EKSA secretariat for final assessment. All these processes took some time. Therefore, iMAS was introduced in 2018 to expedite the assessments and results. It turns out the time required to complete an EKSA session has been reduced from 1-2 months to just 1-2 weeks. Remarkable!

Why Python/Django? This question always comes up when we present iMAS. The main reason why we develop it in Python/Django is because we would like to develop a system that is fast and secure.

Worldwide, May 2021 compared to a year ago:

Rank	Change	Language	Share	Trend
1		Python	29.9 %	-1.2 %
2		Java	17.72 %	-0.0 %
3		JavaScript	8.31 %	+0.4 %
4		C#	6.9 %	-0.1 %
5	↑	C/C++	6.62 %	+0.9 %
6	↓	PHP	6.15 %	+0.1 %
7		R	3.93 %	+0.0 %
8		Objective-C	2.52 %	+0.1 %
9		Swift	1.96 %	-0.2 %
10	↑	TypeScript	1.89 %	+0.0 %
11	↓	Matlab	1.71 %	-0.2 %
12		Kotlin	1.62 %	+0.1 %

Figure 1 - *Popularity of Programming Language (PYPL) index, as of May 2021*

As a start, we decided to develop a system (in Python/Django) that is small in size (less functions and users) and operates once or twice a year. In addition, the best way to learn a new language is to develop it. Therefore, iMAS was built. Worth to mention is that Python has gained its maturity and popularity among IT developers globally. Based on PYPL Popularity of Programming Language index, Python is the most searched language based on google trends, followed by JAVA, PHP, R and Ruby (Figure 1).

Although iMAS has successfully reduced time consumption for MINDEF EKSA auditing activities, a new problem raised: Reporting. After an auditing session is completed, the secretariat will generate a report that will depict the scores for all EKSA auditees. This report will be the reference for the ministry EKSA auditing outcome. However, the report took about 11.9 seconds to be generated. Despite the fact that the user (the secretariat) does not complain about it, it is still unacceptable as we want a quicker system.

After a deep analysis, we found that the main reason why the report took so long is because the non-count scores were determined by the auditors who assess the auditee. Once they are on site then they will know which components, subcomponents of questions that are accountable and not accountable. Different sites have different criteria. So, the calculations must be flexible and the logic will be different for each site. The more flexible the system is, the more complicated the system will be and the more time is required to complete the calculations (as more looping and logics may be desired). Last but not least, but most importantly, is the incorrect instrument. Data analytic requires its own setup and features so that it meets its purpose. Only these tools can handle enormous data with velocity. This is where the search for a better tool begins.

Earlier this year (2021), our ministry was introduced to big data analytics and one of the tools presented is *pandas*. As explained earlier, *pandas* is rich with analytics features that make reporting much easier and promptly. In fact, less looping and logics needed to execute complicated formulas. All are within the *pandas* itself and it can be installed using PIP (*Pip Install Packages for python*) standard installation. Thus, we have chosen *pandas* for our iMAS performance upgrade test. Besides, *pandas* is one of the python libraries which makes it easy as iMAS also developed in python. The way to enable the *pandas* module in iMAS is shown in Figure 2.

```
#Importing models and pandas
from Urusetia.models import SubRefSkala,Auditee
from Penilaian.models import Penilaian
from Pentadbir.models import Sistem
from django.db.models import Max
import datetime
import pandas as pd
import numpy as np
import math, json
```

Figure 2 - pandas and numpy modules are imported in iMAS

After importing pandas, we need to convert all the assessment results to *pandas* data frame. Then, aggregating it (with group by) before executing all the calculations. The best part is that no looping was required (based on iMAS functionality) compared to the later algorithm.

Yes, again, no looping is required. Can you believe it? We think this is why pandas are so powerful and excel for data analytics. The introduction of *JupyterLab* tool also makes our job easy. We managed to visualize the outcome for each step or commands used (Refer Figure 3). This will reduce errors while coding. It is an interactive

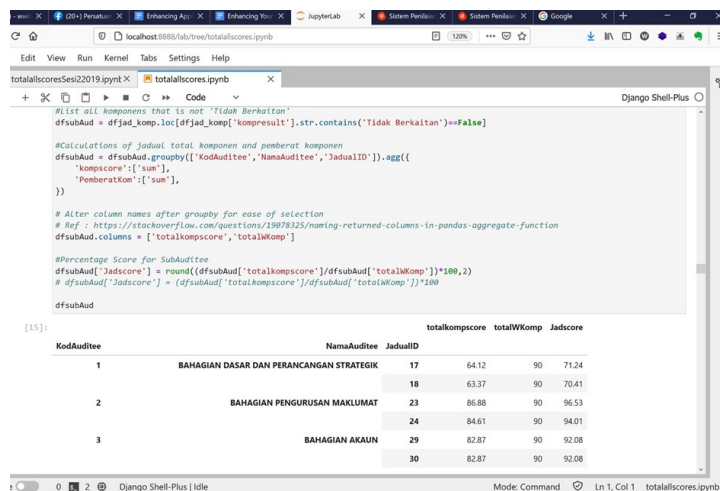


Figure 3 - JupyterLab was used to calculate results for each auditee and the outcome can be visualized through it. As we can see, 'dfsubAud' is a pandas data frame that was created to obtain Sub Auditee scores.

The results after pandas consumption in iMAS are quite shocking. Previously, 11.9 seconds were required to generate all auditees' reports. Now, it requires only 0.94 seconds. 10 times faster than before. Awesome! We believe this outcome should be lesser if it does not require conversion to html (Bootstrap). The evidence can be depicted in Figure 4 (Without pandas) and Figure 5 (with pandas).

“The results after pandas consumption in iMAS are quite shocking”

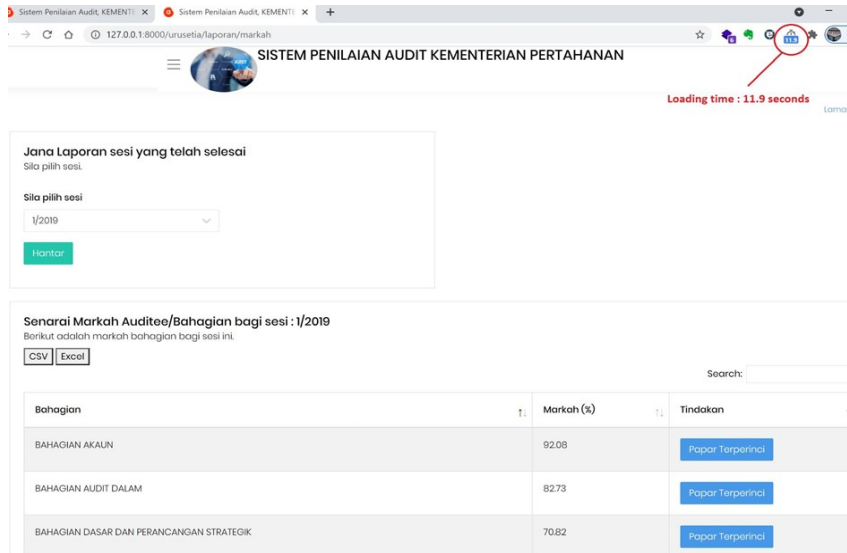


Figure 4 - The loading time for Auditee report without pandas : 11.9 seconds

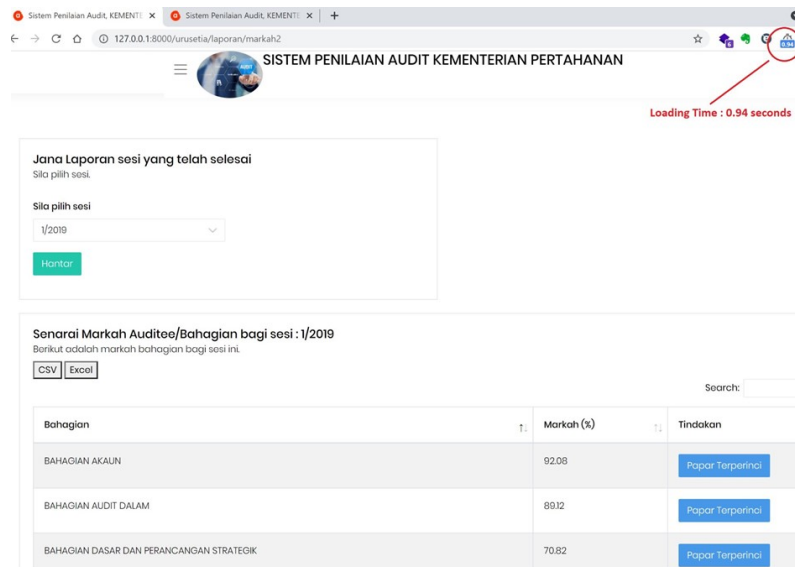


Figure 5 - The loading time for Auditee report with pandas : 0.94 seconds

CONCLUSION

As a conclusion, this finding will be a benchmark and reference for our future application development particularly when it comes to data analytic. Basically, we plan to use stored procedures to augment our application analysis performance. But it comes with one set back. The dependency will be too much on the database itself. Therefore, less python will be utilized. With *pandas*, we are free to use any database as *pandas* does not depend on them too much. It stands by itself. Additionally, *pandas* can replace SQL statements with lesser statements and line of codes. It is worth a try!

KEPERLUAN MEMBINA TAHAP KETERSEDIAAN YANG TINGGI (HA) KE ATAS SERVER VIRTUAL SIRI 2

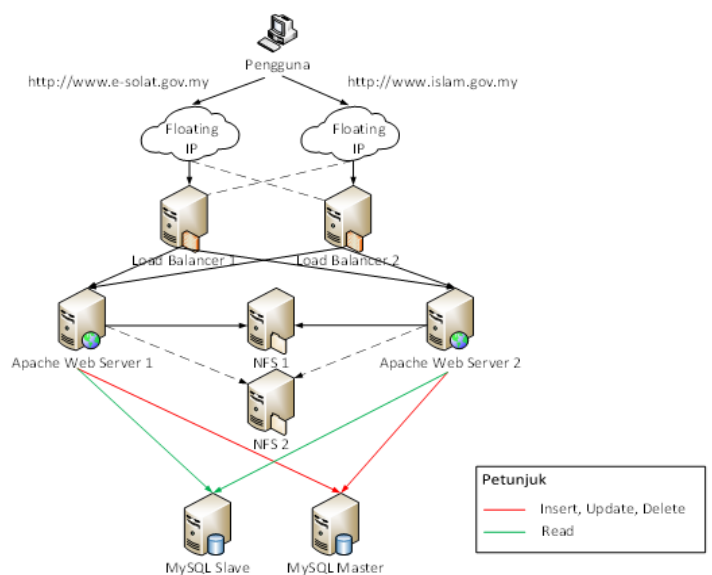
Mohd Rizal Kadis
Jabatan Kemajuan Islam Malaysia

Dalam artikel sebelum ini telah diterangkan asas pembinaan *Active-Passive High Availability (HA)* ke atas *server virtual* serta beberapa kelemahan yang telah dikenal pasti disebabkan oleh senibina tersebut. Pada siri kali ini akan diterangkan bagaimana untuk mengatasi sebahagian daripada kelemahan tersebut untuk menjadikan fungsi HA lebih efisien dan mesra pengguna. Artikel ini akan menerangkan penambahbaikan yang boleh dilaksanakan dengan menggunakan kaedah *Active-Active HA*. Buat masa ini, artikel ini hanya akan membincangkan apa yang boleh dilaksanakan pada sistem aplikasi berbanding dengan sistem pangkalan data. Ini kerana, sistem pangkalan data memerlukan teknik khusus terutamanya apabila melibatkan operasi Sisip, Kemaskini, Padam dan Pilih (*Insert, Update, Delete and Select*) atau lebih dikenali sebagai DML (*Data Manipulation Language*)

Seperti yang dibincangkan dalam artikel sebelum ini, masalah utama yang terjadi kepada sistem aplikasi adalah seperti berikut:-

- *Server Failover* tidak akan digunakan sehinggalah *primary server* rosak, ini akan menyebabkan pembaziran sumber.
- Terdapat dua set kod sumber masing-masing dalam *Apache Web Server 1* dan *2*, ini menyebabkan Pentadbir Sistem perlu memuat naik kod sumber kepada dua lokasi berlainan.
- Oleh kerana fail sidang (*session*) disimpan di server *primary* sahaja, pengguna akan terkeluar daripada sistem secara tiba-tiba dan perlu login semula.

Kadangkala pembaziran sumber berlaku apabila sumber yang disediakan melebihi daripada keperluan sistem aplikasi. Sebenarnya, perkongsian server maya (*virtual*) untuk beberapa sistem aplikasi adalah lebih efisien berbanding daripada memperuntukkan server maya untuk sistem aplikasi tertentu sahaja. Ini juga dapat mengurangkan tenaga kerja penyelenggaraan server kerana jumlah server yang terlibat tidak terlalu banyak. Selain daripada itu, mungkin terdapat sistem aplikasi yang hanya mengalami capaian yang tinggi secara bermusim seperti musim peperiksaan, perayaan dan sebagainya. Bagi sistem seperti ini, kita boleh memberi keutamaan secara sementara untuk sistem tersebut kepada beberapa server maya sehingga musim tersebut tamat.



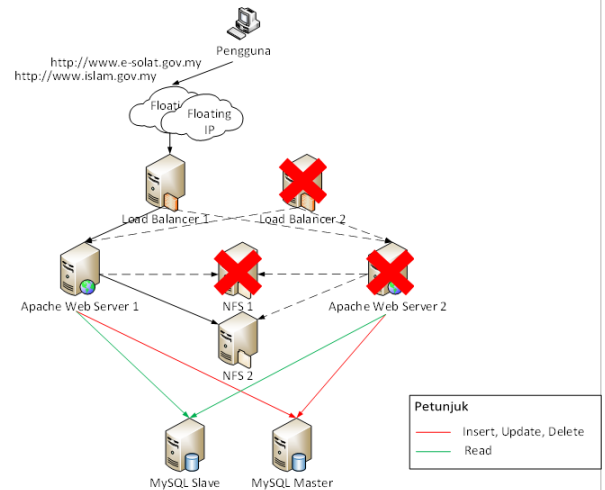
Gambarajah 1: Topologi Server berteraskan Active-Active HA bagi sistem aplikasi dengan menggunakan HAProxy dan NFS

Gambarajah 1 adalah senibina bagi membentuk *Active-Active HA* kepada sistem aplikasi. Fungsi bagi setiap server adalah seperti berikut:-

- *Load Balancer* – perisian yang digunakan adalah HAProxy dan Keepalived.
- *Apache Web Server* – Mengandungi perisian sumber terbuka *server* HTTP dan enjin PHP. Boleh juga menggunakan mana-mana server web lain seperti Nginx.
- *MySQL Master* – Pangkalan data MySQL yang bertindak sebagai tuan (*master*).
- *MySQL Slave* – Pangkalan data MySQL yang bertindak sebagai *slave*.

Perhatikan bahawa terdapat pertambahan dua server pengimbang beban (*load balancer*) dan NFS (Network File Share) server bagi meningkatkan tahap HA dan menyeimbangkan aliran trafik. Berbanding kepada kaedah sebelum ini, kod sumber kini hanya perlu dimuat naik kepada server NFS 1 sahaja dan seterusnya akan di disegerakkan "*sync*" ke server NFS 2. Selain daripada itu, fail sidang (*session*) juga disimpan di dalam server NFS. Terdapat juga kaedah lain untuk menempatkan fail sidang (*session*) secara berpusat namun ianya tidak akan dibincangkan dalam artikel ini.

Dalam hal ini, server NFS 1 akan bertindak sebagai *primary server* manakala server NFS 2 akan bertindak sebagai *failover server*. Server juga berkemampuan untuk menempatkan sistem aplikasi yang berlainan dengan aliran trafik yang berbeza



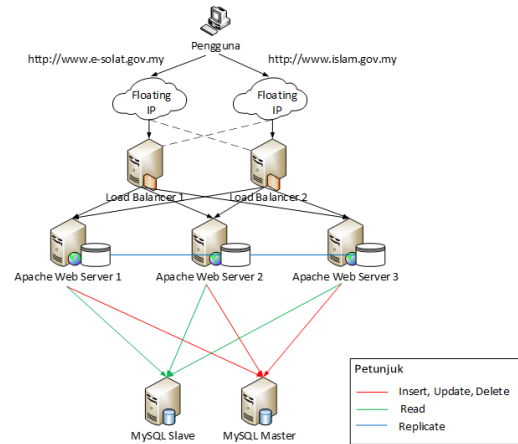
Gambarajah 2: Senario kerosakan kepada Load Balancer 2, Apache Web Server 2 dan NFS 1

Gambarajah 2 adalah senario tahap toleransi terhadap capaian sistem sekiranya berlaku kegagalan server. Walaupun berlaku kerosakan kepada load balancer, server NFS dan *Apache Web Server*, system masih boleh berfungsi dengan campur tangan Pentadbir Server disebabkan NFS perlu dilekapkan "*mount*" semula secara manual pada server web. Ini membolehkan pengguna menggunakan sistem secara normal dalam tempoh menunggu yang tidak terlalu lama. Disebabkan fail *session* disimpan secara berpusat iaitu di dalam server NFS, pengguna tidak akan terkeluar daripada sistem walaupun berlaku kerosakan kepada Apache Web Server.

Walau bagaimanapun, senibina yang ditunjukkan masih mempunyai kelemahan iaitu proses *failover* tidak akan berlaku secara automatik sekiranya berlaku kegagalan kepada server NFS. Ia perlu dilekapkan "*mount*" semula oleh Pentadbir Server secara manual. Ini bermaksud, pengguna perlu menunggu sehingga Pentadbir Server selesai melaksanakan proses lekap "*mount*" tersebut. Bayangkan sekiranya sistem dikaunter tergendala atau kerosakan berlaku diwaktu malam atau semasa Pentadbir Server tidak dapat dihubungi.

Sudah tentu kakitangan dibarisan hadapan akan menghadapi situasi sukar untuk berhadapan dengan pelanggan.

Selain daripada itu, capaian kepada kod sumber daripada *Apache Web Server* adalah bergantung sepenuhnya kepada tahap kelajuan rangkaian yang mungkin akan menjejaskan sedikit tahap kelajuan pengkompil (*compiler*) kod sumber. Namun begitu, masalah ini dapat diatasi sekiranya kelajuan rangkaian adalah tinggi dan sumber server juga banyak.



Gambarajah 4: Senibina Server Cluster menggunakan GlusterFS.

Di antara alternatif yang boleh diambil untuk membina Active-Active HA adalah dengan menggunakan kaedah pengelompokan server (*server clustering*) seperti penggunaan *filesystem* GFS2 atau OCFS2, atau menggunakan kaedah *ethernet-based storage server* seperti GlusterFS atau Ceph.

GFS2

GFS2 adalah singkatan kepada *Global File System 2* yang dimiliki oleh Red Hat dan digunakan sebagai *filesystem* kepada infrastruktur pengelompokan (*cluster*) bagi sistem pengoperasian linux. GFS2 bertujuan untuk mewujudkan Active-Active HA yang teguh dan stabil serta mempunyai prestasi kelajuan yang lebih baik daripada NFS. Ia menggunakan konsep perkongsian storan tunggal di antara setiap server yang biasanya dihubungkan secara terus kepada SAN Storage sama ada melalui RAW Mapping atau perkongsian cakera (*disc*) yang sama di dalam simpanan data (*datastore*) (sebenarnya agak sukar juga untuk menjelaskan kedua-dua terma ini di dalam artikel ini). Dengan menggunakan SAN Storage, ianya lebih menikmati penggunaan protokol *fiber channel* yang lebih teguh dan kurang kependaman (*latency*) berbanding protokol *ethernet*. Bagi memudahkan urusan penyelenggaraan dikemudian hari, kaedah mencipta perkongsian cakera di dalam simpanan data (*datastore*) adalah menjadi pilihan utama.

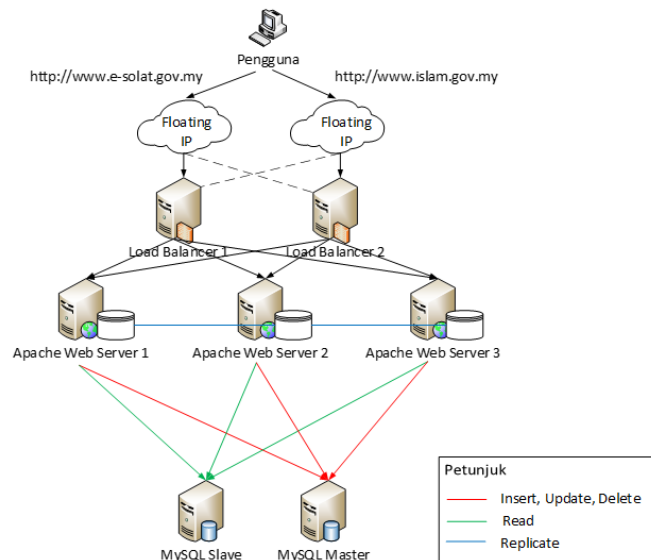
Gambarajah 3 menunjukkan senibina bagi *filesystem* GFS2 dimana beberapa Apache Web Server berkongsi storan yang sama dan membentuk pengelompokan (*cluster*). Perlu diketahui bahawa jumlah minimum server yang diperlukan untuk membentuk pengelompokan (*cluster*) ini adalah 3 unit.

Di antara kelebihan kaedah ini adalah ianya teguh dan stabil serta dijamin oleh pemiliknya iaitu Red Hat. Ia juga adalah lebih selamat kerana tidak mudah diakses oleh penceroboh melalui rangkaian pejabat. Kesemua server *Apache Web Server* tersebut adalah active-active dimana sekiranya berlaku kegagalan kepada salah satu server, sistem masih boleh diakses walaupun berlaku sedikit waktu gendala

Namun begitu, ia juga mempunyai kekurangan seperti berikut:-

- Keseluruhan konfigurasi adalah kompleks dan memerlukan Pentadbir Server yang benar-benar mahir untuk mengendalikannya
- Disebabkan ianya menggunakan storan tunggal, risiko untuk berlakunya single point of failure juga adalah tinggi.
- Risiko untuk berlakunya split brain iaitu setiap server tidak bertindak mengikut aturan yang telah ditetapkan juga tinggi sekiranya berlaku gangguan rangkaian. Misalnya, terdapat dua atau lebih server yang cuba membuat operasi write ke atas blok fail yang sama akan menyebabkan keseluruhan filesystem corrupt. Namun begitu, Red Hat telah mengatasi masalah ini dengan mewujudkan konsep fencing.

- Kos untuk mewujudkan pengelompokan (cluster) ini adalah tinggi.
- Penyelenggaraan juga adalah sukar.
- Disebabkan storan menggunakan multi-write atau RAW Mapping, perisian san-dar (backup) atau fungsi vMotion (VMware) mungkin tidak dapat berfungsi.



GlusterFS

GlusterFS adalah singkatan kepada GNU Cluster Filesystem yang menggunakan konsep *replication volumes* kepada setiap server untuk menjamin ketersediaan dan kebolehpercayaan data. Berbanding dengan GFS2, ia menggunakan rangkaian ethernet dan menggunakan konsep storan teragih berbanding GFS2 yang menggunakan konsep storan tunggal.

Gambarajah 4 menunjukkan senibina bagi GlusterFS dimana beberapa Apache Web Server bertindak sebagai server storage dan membentuk pengelompokan (*cluster*). Sama seperti GFS2, GlusterFS juga memerlukan sekurang-kurangnya 3 unit server.

Kaedah ini mempunyai kelebihan dimana ianya lebih murah dan mudah untuk diselenggara berbanding GFS2. Proses failover juga berlaku secara automatik tanpa memerlukan penglibatan Pentadbir Server.

Disamping itu, beban rangkaian juga diagihkan secara sama rata. Ia juga dapat mengurangkan risiko berlakunya *single point of failure* kerana setiap fail telah di salin ke setiap server.

Namun begitu, ia juga mempunyai kekurangan seperti berikut:-

- GlusterFS memerlukan keupayaan prestasi rangkaian yang tinggi kerana ianya menggunakan protokol ethernet.
- Tahap keselamatan perlu sentiasa dipertingkatkan kerana ianya terdedah kepada capaian tidak sah melalui rangkaian pejabat.
- Prestasi I/O adalah lebih perlahan daripada GFS2. Namun begitu, ianya tidak menunjukkan perbezaan yang ketara di dalam persekitaran rangkaian yang pantas.

Ia juga memerlukan Pentadbir Server yang benar-benar mahir untuk mengendalikannya

Kesimpulan

Pemilihan senibina HA yang ingin dibangunkan adalah bergantung kepada penilaian risiko di agensi masing-masing. Ia perlulah selari dengan tahap kepentingan sistem aplikasi tersebut kepada pelanggan, tahap keupayaan sumber manusia serta kedudukan kewangan.

Di artikel seterusnya kita akan membincangkan tentang HA bagi pangkalan data.

Tech Empowerment Series (TES) (Advanced Users)

Join us as we explore topics to help ease getting to know Google Cloud and it's many features within the platform.

9am - 11am

Why you should join us?

- Industry specialists from Google conducting sessions to benefit you**
 With our partners over at Google, we have invited experienced industry practitioners to grace the stage and share their knowledge and insights for your benefit.
- Relevant topics to the changing technology landscape**
 Topics in technology can be a tricky thing, lucky for you, the topics during our August - October sessions are specifically curated to suit the time and advancement of tech.
- It is free of charge**
 Need we say more?

Speakers



2 September 2021
Complex Analytics made easy with Looker and Big Query
Min Chen
 Customer Engineer, Public Sector, Google Cloud



12 October 2021
Recaptcha for Website Security: Help protect your website from fraudulent activity, spam, and abuse using Recaptcha
Lee Wai Theng
 Customer Solutions Consultant, Government, Google Cloud



7 September 2021
Managing multi-cloud workloads at scale with Anthos
Jonathan Chin
 Customer Engineer, Application Modernization, Public Sector, Google Cloud



26 October 2021
Meet Google's AI Learning Platform
Cedric Yang
 Education Lead, Public Sector, Google Cloud



21 September 2021
Introduction to API Management with Apigee
Lee Wai Theng
 Customer Solutions Consultant, Government, Google Cloud



Participants will receive a **RM 10 Grab voucher** during Q&A session.
*Terms and conditions apply



E-certificate of attendance from Awantec.



Secure your seat now

<https://forms.gle/aq7t8dRRt2s2TED89>

Like, Share & Follow Us!



@awantecmy

AMALAN TERBAIK BAGI PANDUAN PENGATURCARAAN SELAMAT DALAM PERSEKITARAN JSP

Ts. Najdawati Mohd Fadzil
Pusat Pengurusan dan Pembangunan Digital,
Universiti Malaysia Perlis

Pengenalan

Setiap hari, penjenayah siber akan melancarkan serangan yang berbeza untuk mencari kelemahan dalam sistem maklumat tanpa kita sedari. Amalan panduan pengaturcaraan yang selamat adalah salah satu kaedah terbaik untuk mencegah serangan dan kesalahan sebanyak yang mungkin. Sehubungan itu, pembangun sistem haruslah mengekalkan dan mengemaskini garis panduan pengaturcaraan yang selamat semasa membangunkan sistem sekerap yang mungkin. Terdapat pelbagai alatan yang tersedia setakat ini yang dapat mengautomasikan proses ini.

Apakah Panduan Pengaturcaraan Selamat?


Garis panduan pengaturcaraan selamat digunakan untuk menulis pengaturcaraan yang mempunyai ketahanan tinggi terhadap serangan oleh penjenayah siber. Pengaturcaraan tanpa jaminan keselamatan yang tinggi akan menyebabkan capaian tidak sah oleh penjenayah siber yang digunakan untuk tujuan mencerooboh operasi pelayan atau komputer pengguna tanpa kebenaran. Ini akan mengakibatkan kebocoran maklumat sulit serta kerosakan pada sistem perkhidmatan pengguna. Pengaturcaraan yang selamat akan membantu melindungi data pengguna dari sebarang kecurian atau kerosakan dan sangat penting bagi semua perisian sama ada untuk kegunaan peribadi ataupun komersial.

Panduan Pengaturcaraan Selamat

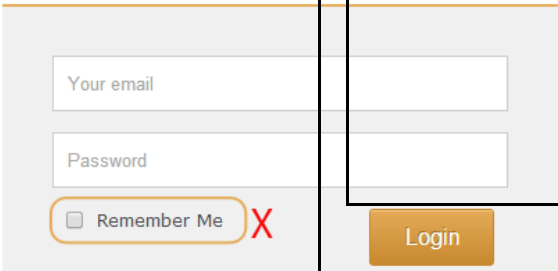
Terdapat beberapa panduan pengaturcaraan selamat beserta contoh yang sangat berguna dan boleh dipraktikkan khususnya di persekitaran JSP. Berikut adalah

- Pengesahan Input / *Input Validation*
- Pengesahan dan Pengurusan Kata Laluan / *Authentication and Password Management*
- Pengurusan Sesi / *Session Management*
- Kawalan Capaian / *Access Control*
- Konfigurasi Sistem / *System Configuration*
- Keselamatan Pangkalan Data / *Database Security*

Pengesahan Input / *Input Validation*

Item	Contoh
Kodkan data input ke jenis input yang berkaitan sebelum pengesahan dilaksanakan.	Contoh: Jenis data bagi kata laluan <pre><input maxLength=20 name=upass type=password size="20"></pre>
Semua kegagalan pengesahan input harus mengakibatkan penolakan input.	
Sahkan untuk jenis data yang diharapkan.	<ul style="list-style-type: none"> • Sekiranya input teks kosong atau tidak • Sekiranya input teks adalah semua nombor • Sekiranya input teks adalah semua huruf
Sahkan julat data, jika sekiranya boleh.	<ul style="list-style-type: none"> • Semua tarikh yang lalu tidak boleh dipilih semula. • Tidak akan dapat memulakannya selepas tarikh akhir ataupun akhir sebelum tarikh mula.

Pengesahan dan Pengurusan Kata Laluan / *Authentication and Password Management*

Item	Contoh
Pengesahan diperlukan untuk semua halaman dan sumber, kecuali bagi halaman yang dikhaskan untuk paparan umum sahaja.	<pre><%@ include file="inc/chkvalid.jsp"%></pre>
Tindakbalas semasa kegagalan pengesahan tidak seharusnya menunjukkan bahagian data pengesahan mana yang salah.	Sebagai contoh, bukannya "Nama pengguna tidak sah" atau "Kata laluan tidak sah" , gunakan "Nama pengguna dan / atau kata laluan tidak sah" untuk kedua-duanya.
Pengisian kata laluan harus dikaburkan di skrin pengguna.	Contohnya, pada borang berasaskan web, gunakan jenis input "kata laluan"
Menyahaktifkan fungsi "Ingat Saya / Remember Me" bagi medan kata laluan.	<p>IDC login</p> 

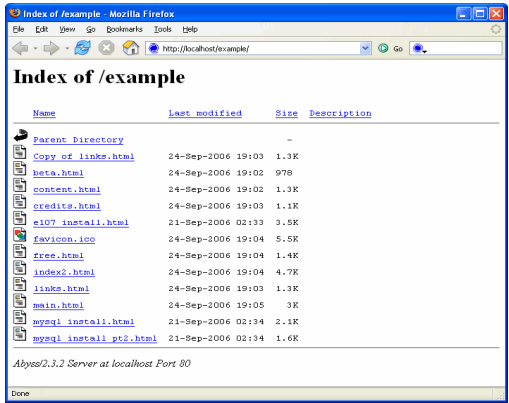
Pengurusan Sesi / *Session Management*

Item	Contoh
Fungsi log keluar seharusnya menghentikan sesi atau sambungan yang berkaitan secara sepenuhnya.	<pre><%session.invalidate();%></pre>
Penetapan masa tamat bagi setiap aktiviti secara sesingkat mungkin, dan ia seharusnya berdasarkan keperluan fungsi.	<p>Secara umumnya, untuk Aplikasi bernilai tinggi (<i>High Value Apps</i>): 5 minit, untuk Aplikasi bernilai sederhana (<i>Medium Value Apps</i>): 10 minit dan untuk Aplikasi bernilai rendah (<i>Low Value Apps</i>): 15 minit (OWASP)</p> <p>Contohnya;</p> <pre>String idmy = request.getParameter("myid"); String idgrp = request.getParameter("grpid"); session.setMaxInactiveInterval(300); //300 secs = 5 mins session.setAttribute("themyid", idmy); session.setAttribute("thegrpid", idgrp);</pre>

Kawalan Capaian / Access Control

Item	Contoh
Hadkan capaian ke fungsi / <i>functions</i> yang dilindungi hanya untuk pengguna yang dibenarkan.	Senarai capaian untuk aplikasi / sistem. Contoh: Senarai Pengguna dan Peringkat Capaian
Hadkan capaian ke perkhidmatan / <i>services</i> hanya untuk pengguna yang dibenarkan.	Senarai capaian untuk aplikasi pelayan. Contoh SSH, FTP
Hadkan capaian ke aplikasi data / <i>data application</i> hanya untuk pengguna yang dibenarkan.	Senarai capaian untuk pelanggan pangkalan data. Contoh Schema

Konfigurasi Sistem / System Configuration

Item	Contoh
Pastikan pelayan / <i>servers</i> , kerangka kerja / <i>frameworks</i> dan komponen sistem berfungsi pada versi terbaru yang telah diluluskan.	
Pastikan pelayan / <i>servers</i> , kerangka kerja / <i>frameworks</i> dan komponen sistem mempunyai semua <i>patch</i> yang dikeluarkan untuk versi yang digunakan.	
Pastikan capaian ke senarai direktori ditutup.	
Keluarkan semua fungsi dan fail yang tidak diperlukan.	e.g., old version of source code
Keluarkan kod ujian atau apa-apa fungsi yang tidak digunakan pada persekitaran sebenar sistem yang sebelum ini mungkin digunakan pada persekitaran pengujian sistem.	Contoh, *.war
Laksanakan sistem kawalan perubahan perisian untuk mengurus dan merekod perubahan pada kod dalam persekitaran sebenar sistem.	Contoh., SVN Tools

Kesimpulan

Salah satu kesalahan besar yang sering dilakukan ialah memandang remeh aspek keselamatan semasa proses pembangunan sistem. Keselamatan adalah salah satu aspek penting yang perlu dititikberatkan dan harus menjadi keutamaan dari awal kitaran hidup pembangunan sistem. Sehubungan itu, bahagian teknologi maklumat bagi semua agensi kerajaan di Malaysia haruslah mengambil inisiatif untuk meningkatkan daya tahan sistem maklumat masing-masing dengan menggunakan panduan pengaturcaraan yang selamat semasa pembangunan sistem.

CYBERATTACK ATTRIBUTION: AN OVERVIEW TO THREAT INTELLIGENCE

Mohamed Kheirulnaim Mohamed Danial
National Cyber Security Agency (NACSA), Majlis Keselamatan Negara

When the recent Palestine-Israel conflict escalating quickly throughout the whole week, we have witnessed the battle is fought on two different front, kinetic and cyber. While previously Israel Defense Forces (IDF) used to retaliate Hamas cyber offensive to kinetic attack (Kovacs, 2019), the current cyber attack campaign now is spearheaded by multiple group of hacktivists from Muslim-majority countries including Malaysia which targeting both on technical vulnerability and public perception to maximise the campaign effectiveness.

Up until now, around 360 websites has been defaced and gigabytes of personal data belonging to Israeli citizen has been leaked and distributed on multiple channel. This indicate that local Threat Actor (TA) has improved significantly in term of capabilities and coordination compared to the infamous #OpMalaysia 10 years ago.

Cyberattack attribution is the process of tracking and identifying on the perpetrator of a cyberattack or other hacking exploit. Establishing attribution for cyberattack is difficult but not impossible as there's no automated solution or simple technical process for determining the responsibility of the operation. It demands a painstaking effort, in many cases requires weeks or months of analysing all the information at hand. The lack of data collected usually become a common pitfall that will directly affect the correlated information as the accuracy and confidence of the attribution will vary depending on available data. But still, the most common reason to do attribution--and also the one that drive the motivation throughout the process--is our own curious nature that need to know who is the perpetrators. (The Attribution Process, 2020).

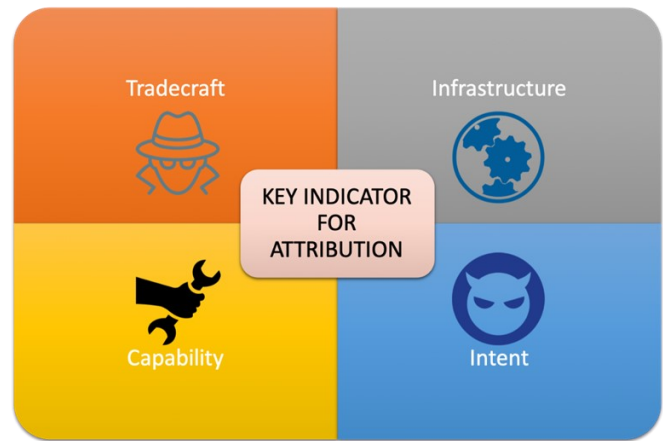
CLEARED FOR RELEASE: We thwarted an attempted Hamas cyber offensive against Israeli targets. Following our successful cyber defensive operation, we targeted a building where the Hamas cyber operatives work.

HamasCyberHQ.exe has been removed.



Analysts typically rely on three sources of data: technical intelligence (TECHINT) that relies on artefact analysis, open-source intelligence (OSINT), which is publicly available information on the internet, and proprietary data available only to the organisations involved with the incident. National-level Information Sharing and Analysis Centre (ISAC) also serve as another source of intelligence because they have more information and additional resources. An ISAC is an industry-specific organisation that gathers and shares information on cyber threats to critical infrastructure. ISACs also facilitate the sharing of data between public and private sector groups. In Malaysia, this task falls to the National Cyber Coordination & Command Centre (NC4), National Cyber Security Agency (NACSA) under the purview of National Security Council (NSC) that are established under the NSC Directive No. 24 to become the national ISAC and responsible with the coordination and sharing cyber threat information to the Critical National Information Infrastructure (CNII) and public during peace and crisis time.

Every kind of cyberattack will leave a digital trail. The analysts will make use of this information, along with the knowledge of previous events and including the Tactics, Techniques, and Procedures (TTP) of known threat actors, to attempt to trace these attacks back to the sources. To do this, the key indicator is needed to guide analysts in seeking timely, accurate attribution. The primary indicators for attribution are **Tradecraft**, **Infrastructure**, **Capability**, and **Intent**. Indicators from external sources, such as open-source reports from the private cybersecurity firms also help to assist in attribution process.



- **Tradecraft:** Frequent behaviour of the threat actor used to conduct cyberattack. This is the most important indicator because habits are more difficult to change than technical tools.
- **Infrastructure:** The physical and/or virtual communication structures used to deliver a cyber capability or maintain persistence link on victim environment. Actors can buy, lease, share, and compromise servers and networks to build their own network of infrastructure scattered around the world.
- **Capability:** The capability describes TTP of the actors used in the event. The TTP can reveal attack patterns for each threat actor including all means to affect the victim from the most manual methods (e.g., manual password guessing) to the most sophisticated automated techniques.
- **Intent:** Actor's commitment to carry out certain actions based on Socio-Political context. Cyberattacks often are launched against victim before or during the conflict as some form of harassment, suppression & protest.
- **Indicators from External Sources:** Reports from the private industry, the media, and academia can be used to provide more visibility on threat landscape or share hypotheses about the perpetrators.

In summary, analysts need to compare the new information to existing knowledge, weigh the evidence to determine a confidence level for their judgments, and consider alternative hypotheses and ambiguities to produce cyber attribution assessments. This process becomes more demanding and time consuming as the analysts move on from attributing a local hacktivist group to a Nation-state Advance Persistence Threat (APT).

REFERENCES

- Kovacs, E. (2019, 5 6). *Israel Bombs Hamas Cyber HQ in Response to Cyberattack*. Retrieved from SECURITYWEEK: <https://www.securityweek.com/israel-bombs-hamas-cyber-hq-response-cyberattack>
- The Attribution Process. (2020). In T. Steffens, *Attribution of Advanced Persistent Threats* (p. 24). Springer Vieweg.

MENINGKATKAN TAHAP EFISIENSI PUSAT DATA MELALUI PENGURUSAN ALIRAN UDARA (AIRFLOW)

Ts. Rizman bin Yusam
Jabatan Pendaftaran Negara Malaysia

Pusat Data merupakan fasiliti yang menempatkan segala peralatan ICT yang merangkumi server, storan, peralatan rangkaian dan pelbagai sistem yang menyokong urusniaga utama sesebuah agensi.

Antara cabaran dalam pengurusan pusat data adalah memastikan peralatan ICT mendapat aliran udara (airflow) sejuk yang mencukupi bagi menjamin prestasi sistem sentiasa berada pada tahap optimum. "Efisien" dalam konteks pengurusan aliran udara pusat data membawa maksud setiap unit penyejukan dapat membekalkan udara sejuk yang mencukupi yang diperlukan oleh ICT.

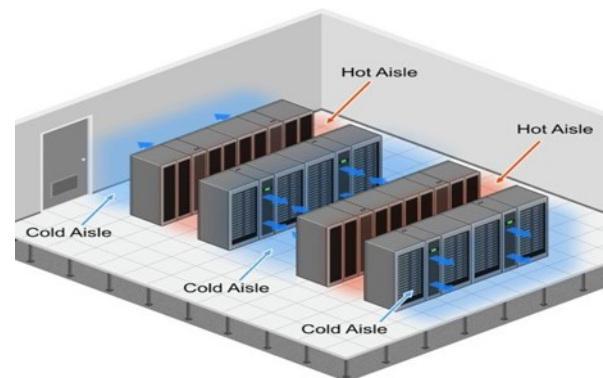
Masalah yang sering berlaku bagi pusat data yang menggunakan sistem penyejukan secara *down floor* atau *room cooling* yang tidak mempunyai pengurusan aliran udara yang baik adalah seperti berikut :

- Titik panas (*hotspot*) berlaku di merata tempat
- Aliran udara sejuk tidak sekata
- Server yang berada di kedudukan rak bahagian atas tidak mendapat udara sejuk (*top of rack*)
- Server kerap mengalami peningkatan suhu
- Bil utiliti yang tinggi

Bagaimana masalah seperti di atas dapat diatasi? Beberapa Langkah berikut boleh dilakukan oleh agensi bagi meningkatkan efisiensi aliran udara dalam pusat data :

Susun Atur Rak Server

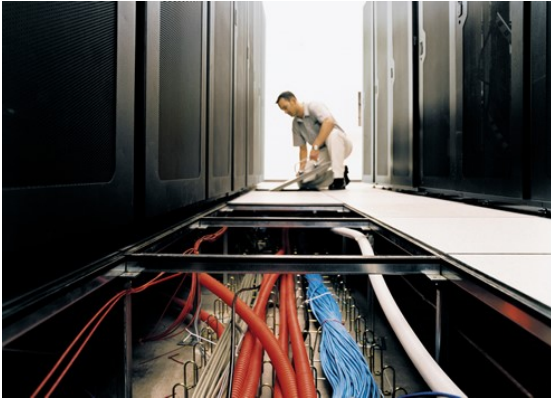
Kaedah penyusunan rak server memainkan peranan yang amat besar dalam memastikan aliran udara yang efisien. Susunan secara front-to-front / back-to-back mampu memperbaiki tahap efisiensi dengan mengurangkan percampuran udara sejuk dan udara panas dalam persekitaran pusat data.



Gambarajah 1 – Cold/Hot Aisle

Lokasi dan Bilangan Perforated Tiles

Perforated tiles perlu di pastikan hanya berada pada hadapan rak server kerana udara sejuk diperlukan oleh peralatan ICT. Peralatan bukan ICT tidak memerlukan udara sejuk dan ini secara langsung dapat mengurangkan bypass airflow yang tidak sepatutnya berlaku. Bilangan *perforated* yang berlebihan mampu mengurangkan tekanan statik (*static pressure*) dan secara langsung memberi kesan ke atas tekanan aliran udara sejuk yang naik ke permukaan melalui *perforated tiles*.



Halangan (*Obstruction*) Di Bawah *Raised Floor*

Halangan dibawah *raised floor* seperti kabel kuasa, kabel rangkaian dan laluan kabel boleh menjadi penyumbang kepada aliran udara sejuk yang tidak efisien. Pastikan semua kabel yang berada di bawah *raised floor* tersusun rapi dengan pengurusan kabel yang baik tanpa menghalang udara sejuk yang naik melalui *perforated tiles*.

Rak Kosong / Ruang Antara Server

Ruang (*gap*) antara server atau pada rak server yang tidak ditutup menggunakan *blanking panel* menjadi penyumbang utama *bypass airflow* dan aliran udara sejuk yang tidak sekata.

Before



After



Gambarajah 2 – *Blanking Panels* (Before/After)



Laluan Keluar Kabel

Semua laluan kabel yang keluar daripada bawah *raised floor* ditutup antaranya dengan memasang *brush grommet* bagi mengurangkan *bypass airflow*.

Penutup

Kesimpulannya, pengurusan aliran udara merupakan satu langkah awal yang utama perlu dititik beratkan oleh agensi yang mempunyai pusat data bagi memastikan peralatan ICT dapat berfungsi secara optimum tanpa ada masalah penyejukan yang memberi kesan ke atas jangka hayat peralatan ICT dalam jangka masa panjang.

Indeks Pangkalan Data dan *Execution Plan*

Ts. Mohd Naim Mohd Ibrahim
Kementerian Dalam Negeri

Sambungan kepada artikel yang lalu bertajuk Pengenalan Kepada Indeks Pangkalan Data, kali ini kita akan membincangkan tentang bagaimana untuk mengenal pasti adakah indeks yang dicipta digunakan oleh injin pangkalan data ataupun tidak.

Jika kita telah mengisytiharkan atau mencipta indeks untuk satu *table*, tidak semestinya injin pangkalan data akan menggunakan indeks tersebut. Ia mungkin memutuskan untuk tidak menggunakan indeks yang dicipta kerana tidak bersesuaian dengan kuiry yang dilarikan.

Sebenarnya kita boleh mengetahui sama ada kuiry yang kita jalankan itu menggunakan indeks ataupun tidak. Kaedahnya sedikit berbeza mengikut pangkalan data.

EXECUTION PLAN

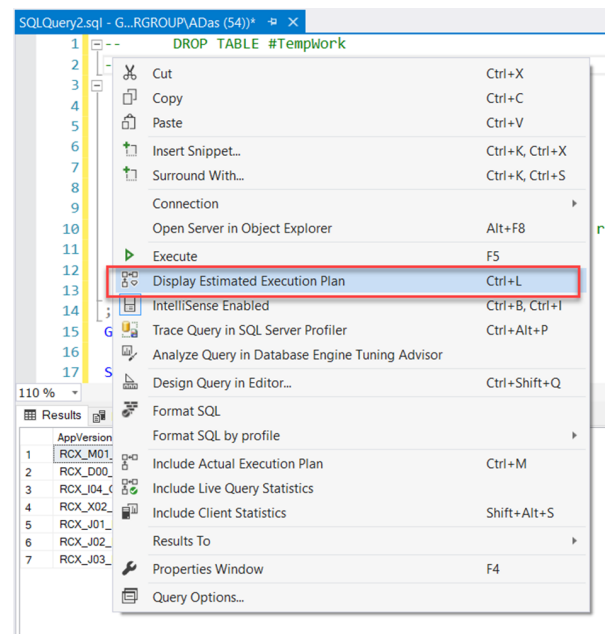
Untuk melarikan sesuatu *SQL statement*, injin pangkalan data mempunyai pelan bagaimana arahan itu akan dilaksanakan, di mana data akan diperolehi dan apakah kaedah untuk mendapatkan data tersebut. Semua ini dipanggil *execution plan*.

MySQL dan PostgreSQL

Kaedah untuk mengetahui *execution plan* untuk MySQL dan PostgreSQL adalah amat mudah. Kita hanya perlu menambah arahan **EXPLAIN** di hadapan *SQL statement* yang ingin kita larikan.

Microsoft SQL Server

Untuk mendapatkan *execution plan* untuk Microsoft SQL Server, kita boleh menekan “CTRL”+“L” ataupun klik kanan pada tettingkap kuiry dan pilih “**Display Estimated Execution Plan**” seperti rajah di bawah.



Gambarajah 1 – SQL Server Studio *Execution Plan*

Oracle

Bagi pangkalan data Oracle, *execution plan* boleh diperolehi dengan menambah **EXPLAIN PLAN FOR** di hadapan *SQL statement* kita.

BAGAIMANA MEMBACA DAN MEMAHAMI EXECUTION PLAN

Oleh kerana ruang yang terhad, pada kali ini penulis hanya akan menerangkan bagaimana untuk membaca dan memahami *execution plan* untuk pangkalan data MySQL.

Apabila arahan **EXPLAIN** di letakkan di hadapan *SQL statement*, output yang akan dipaparkan adalah *execution plan* dan bukan lagi data yang ingin dicapai melalui *SQL statement* asal. Contohnya untuk *SQL statement* :

```
SELECT * FROM table_a
STRAIGHT_JOIN table_b ON table_a.Item
= table_b.Item;
```

Dengan menambah **EXPLAIN** di hadapan *SQL statement* ini

```
EXPLAIN SELECT * FROM table_a
STRAIGHT_JOIN table_b ON table_a.Item
= table_b.Item;
```

Outputnya adalah seperti berikut :

id	select_type	table	type	possible_keys	key	key_len	ref	rows	Extra
1	SIMPLE	table_a	ALL					6	
2	SIMPLE	table_b	ALL					3	Using where; Using join buffer

id = Nilai turutan yang diberikan oleh pangkalan data MySQL.

select_type = jenis query

Type	Penerangan
SIMPLE	Plain, tanpa union atau subquery
PRIMARY	SELECT yang pertama
UNION	SELECT yang kedua atau seterusnya
DERIVED	SELECT dalam table yg ada FROM
DEPEND- ENT UNION	SELECT yang kedua atau seterusnya yg bergantung kpd OUTER query
SUBQUERY	SELECT yang pertama utk subquery
DEPEND- ENT SUBQUERY	SELECT yang pertama utk subquery bergantung kpd OUTER QUERY

table = nama table atau aliasnya

type = jenis akses ke data (ini adalah yang paling penting sekali dalam execute explain)

- System = sumber data dari table System
- const = digunakan bila table ada *primary key* (atau unique index). Sangat pantas
- eq_ref = digunakan apabila satu join menggunakan semua parts dlm unique dan index tanpa null, Cuma 1 row perlu dibaca dari table untuk setiap kombinasi row yg dipulangankan oleh join yg lain. Paling efisien dalam semua jenis join
- ref = macam eq_ref tetapi satu atau lebih row dibaca dari table berbanding hanya satu row.
- Fulltext = join yg dijalankan menggunakan fulltext
- ref_or_null = seperti ref tetapi dengan tambahan searching untuk null
- index_merge = pelbagai index untuk table yg sama digunakan untuk mendapatkan data
- Index_subquery = seperti unique_subquery cuma indexnya tidak unik
- range = digunakan bila suatu index wujud range nilai yg dikenhendaki. Ia termasuklah =, <>, >, <, IS NULL, <=>, BETWEEN, LIKE dan IN
- index = ianya ialah *full scan* untuk index. Lebih laju dari *full table scan*
- ALL = *Full table scan*. Ianya terjadi mungkin disebabkan table tiada index yg sesuai atau tiada kluasa WHERE atau query optimizer fikir ianya adalah kaedah terbaik. Ini yang patut dielakkan.

possible keys = senarai index yang boleh digunakan utk mendapatkan *rows* dari *table*.

key_len = jumlah bytes untuk index yg dipilih yg digunakan oleh MySQL sebenarnya.

rows = anggaran rows yg optimizer fikir ia perlu periksa untuk mendapatkan rows/data yg dikehendaki.

Extra = maklumat lain berkenaan table tersebut

Extra	Penerangan
distinct	MySQL tak perlu cari row lain jika ia sudah jumpa row yg match dgn query mahu
not exists	MySQL tak perlu cari row lain jika ia sudah jumpa row yg match JOIN criteria
range checked for each record	MySQL tak jumpa index yg sesuai, tetapi semak index utk setiap row. Lambat ttapi lbh laju dari full table scan
using filesort	MySQL perlu menjalankan tugas tambahan utk isih row
using index	MySQL boleh dapatkan row/data dari index. Berita baik!
using temporary	MySQL perlu cipta table sementara utk memegang data. Selalu terjadi jika ada GROUP BY atau ORDER
using where	MySQL menggunakan nilai dalam WHERE untuk hadkan row yg perlu dipulangkan
using where	MySQL menggunakan nilai dalam WHERE utk hadkan row yg perlu dipulangkan

Kesimpulan

Dengan menggunakan memaparkan *execution plan*, kita boleh mengetahui sama ada kuir yang dilarikan menggunakan indeks ataupun full table scanning yang memakan masa yang lama. Jika berkesempatan, kita akan membincangkan bagaimana caranya untuk menala indeks supaya injin pangkalan data mememanfaatkannya untuk mencepatkan carian.