

PERLOMBONGAN PROSES

STQD6414 PERLOMBONGAN DATA



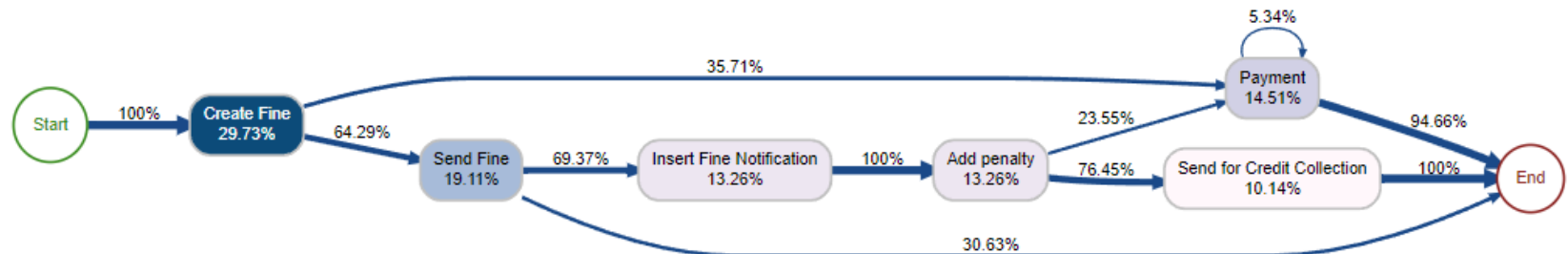
Prof. Madya Dr. Nurulkamal Masseran

Jabatan Sains Matematik

Universiti Kebangsaan Malaysia

PENGENALAN:

- Perlombongan proses merupakan kaedah perlombongan data yang digunakan untuk menganalisis data operasi suatu proses atau sistem.
- Teknik ini menganalisis proses operasi berdasarkan data log-peristiwa (*event logs data*).
- **Tujuan:** mengenalpasti kelemahan proses operasi, menilai tahap kepatuhan proses operasi dan menambah baik proses operasi.
- Data log-peristiwa untuk menunjukkan perkara yang dijalankan oleh pekerja, mesin atau aturan prosedur kerja dalam suatu organisasi pada masa tertentu.
- Perlombongan proses dijalankan dengan menukar data log-peristiwa kepada data cerapan untuk dianalisis.
- Hasil analisis akan digunakan sebagai maklumat untuk tindakan susulan.



PENGENALAN:

- Input untuk perlombongan proses ialah data log-peristiwa.
- Setiap peristiwa dalam log perlu mengandungi maklumat:
 - i) **Pengecaman unik:** ID kes.
 - ii) **Aktiviti:** Maklumat peristiwa yang sedang berlaku.
 - iii) **Rekod masa** (*timestamp*):
- Selain itu, atribut tambahan bagi peristiwa juga boleh dimasukkan dalam data seperti jenis sumber, kos, dan lain-lain.
- Data log-peristiwa boleh diekstrak daripada sistem maklumat yang menyokong proses operasi.



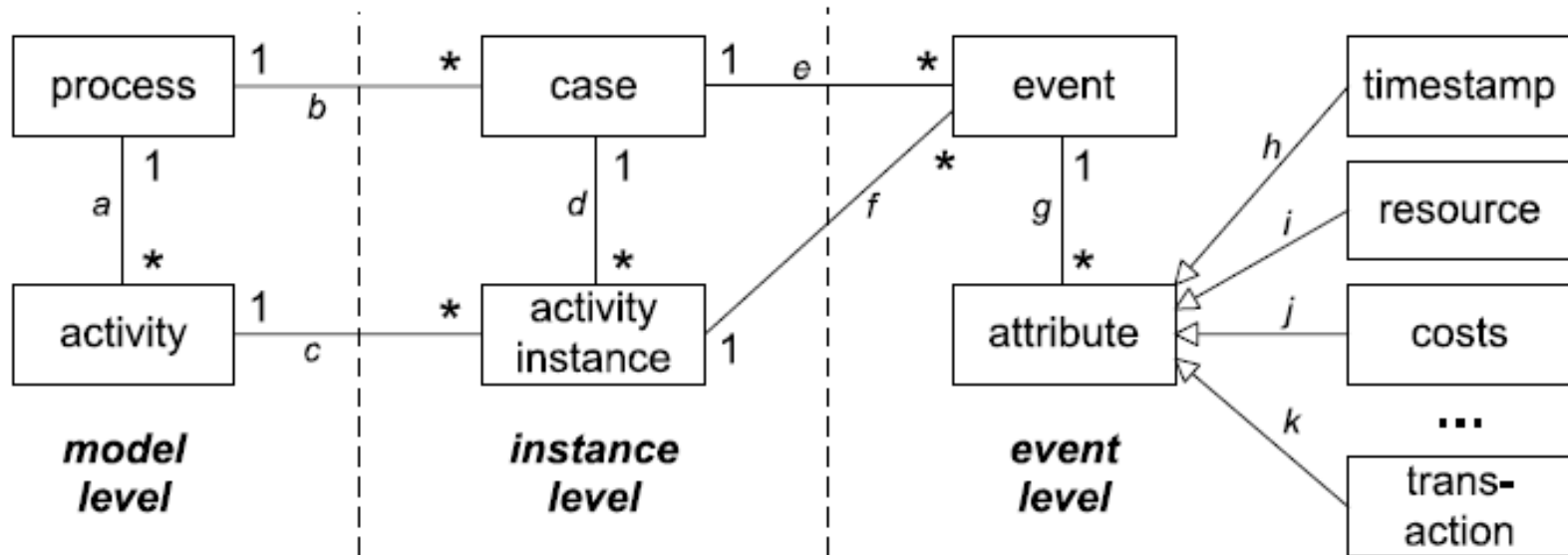
PENGENALAN:

- Berdasarkan data log-peristiwa, perlombongan proses boleh digunakan untuk menjawab pelbagai soalan berkaitan proses operasi.
- Tiga objektif utama dalam perlombongan proses:
 - i) Penemuan proses.
 - ii) Semakan pematuhan.
 - iii) Peningkatan proses.
- Perlombongan proses juga dikenali sebagai Perlombongan Aliran Kerja dan Penemuan Proses Perniagaan Automatik (*Workflow Mining and Automated Business Process Discovery, ABPD*)



LOG-PERISTIWA:

- Log-peristiwa merujuk kepada satu set peristiwa yang direkodkan dalam konteks proses operasi.
- Contoh:** proses operasi di sebuah Jabatan Kecemasan Hospital.
- Gambaran umum model proses operasi bagi data log-peristiwa:



MODEL DATA PERISTIWA:

i) **Kes:** Setiap peristiwa ditakrif sebagai kes.

- **Contoh:** Di jabatan kecemasan, contoh kes adalah lawatan pesakit.

ii) **Aktiviti:** Merujuk kepada langkah-langkah dalam proses.

- **Contoh:** terima pesanan, hantar pembayaran, jalankan MRI SCAN, dan lain-lain.

iii) **Kejadian aktiviti (*Activity instances*):** Pelaksanaan aktiviti khusus untuk kes tertentu.

Contoh: pembedahan dilakukan untuk kes kecemasan kritikal.

iv) **Peristiwa:** Kes yang dikaitkan dengan aktiviti yang berlaku dan dicirikan dengan rekod masa.

- **Contoh:** Pesakit X mula menjalani imbasan MRI pada masa t .



MODEL DATA PERISTIWA:

v) **Sumber:** Pekerja atau mesin yang digunakan untuk pelaksanaan suatu aktiviti.

- **Contoh:** ahli radiologi yang bertanggungjawab ke atas imbasan MRI.

vi) **Status kitaran (*Lifecycle status*):** Petunjuk status aktiviti yang berlaku bagi suatu peristiwa.

- Petunjuk yang biasa digunakan ialah masa mula dan tamat
- Petunjuk lain adalah penggantungan, disambung semula, dll.

vii) **Jejak (*Trace*):** Jujukan bagi aktiviti.

- Kejadian aktiviti bagi kes yang memberikan maklumat jejak tertentu apabila dितertibkan dalam suatu masa tertentu.



MODEL DATA PERISTIWA:

- Dua jenis data yang digunakan untuk perlombongan proses ialah:

i) Data log-peristiwa:

- Data log-peristiwa dibina daripada bingkai data dengan baris mewakili satu peristiwa.
- Hanya mempunyai 1 rekod masa.

handling	patient	employee	handling_id	registration_type	time
Registration	16	r1	16	start	2017-01-13 12:09:49
Registration	16	r1	16	complete	2017-01-13 17:24:36
Triage and Assessment	16	r2	516	start	2017-01-14 06:37:26
Triage and Assessment	16	r2	516	complete	2017-01-14 13:54:22
Blood test	16	r3	1009	start	2017-01-15 02:41:17
Blood test	16	r3	1009	complete	2017-01-15 06:35:38

MODEL DATA PERISTIWA:

ii) Data log-aktiviti:

- Data log-aktiviti dibina daripada bingkai data dengan setiap baris mewakili satu aktiviti.
- Ia boleh mempunyai rekod masa berbilang yang disimpan dalam lajur yang berbeza.

handling	patient	employee	handling_id	complete	start
Registration	16	r1	16	2017-01-13 17:24:36	2017-01-13 12:09:49
Triage and Assessment	16	r2	516	2017-01-14 13:54:22	2017-01-14 06:37:26
Blood test	16	r3	1009	2017-01-15 06:35:38	2017-01-15 02:41:17



TEKNIK PENILAIAN DAQAPO:

- DaQAPO merujuk kepada Penilaian Kualiti Data untuk Proses Berorientasi (*Data Quality Assessment for Process-Oriented*).
- DaQAPO merupakan suatu set kerangka fungsi penilaian untuk mendapatkan maklumat perihalan berkaitan data log-peristiwa.
- Antara teknik penilaian berasaskan DaQAPO untuk mengenalpasti:
 - i) Kes ID yang tiada rekod/hilang.
 - ii) Keberlakuan aktiviti mengikut syarat.
 - iii) Nilai pencil tempoh masa aktiviti.
 - iv) Tempoh tak aktif.
 - v) Kes-kes tak lengkap.
 - vi) Nama aktiviti-aktiviti yang salah rekod.



PENILAIAN DATA-DAQAPO:

- Antara teknik penilaian berdasarkan DaQAPO untuk mengenalpasti:

viii. Nilai-nilai lenyap.

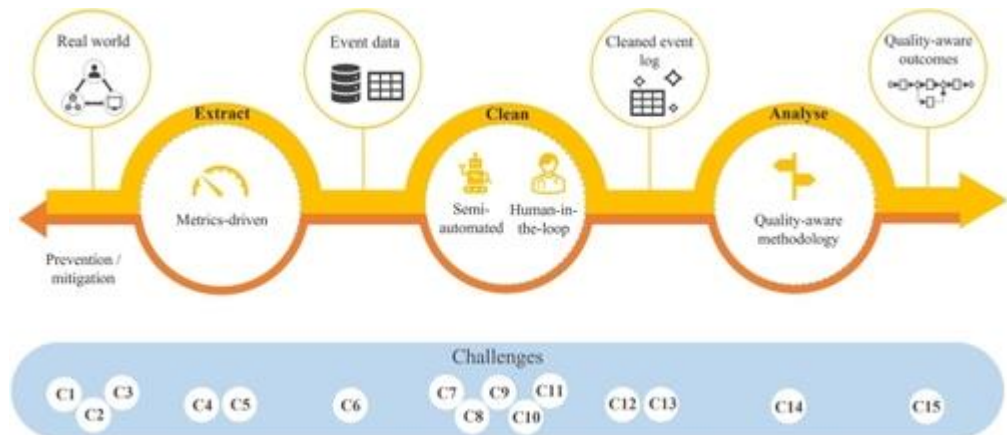
ix. Pendaftaran berganda.

x. Pertindanan.

xi. Anomali masa.

xii. Nilai-nilai unik.

xiii. Dan lain-lain.



SUBSET DATA PERISTIWA:

- Subset data log peristiwa boleh dibahagikan kepada dua jenis yang utama:
 - a) Tapisan peristiwa (*Event filters*).
 - b) Tapisan kes (*Case filters*).
- Tapisan peristiwa mengsubset sebahagian kes berdasarkan kriteria yang digunakan pada peristiwa (*Contoh*: apakah sumber yang digunakan dalam kes tertentu).
- Antara teknik bagi tapisan peristiwa:
 - i) Label aktiviti.
 - ii) Kekerapan aktiviti.
 - iii) Label sumber.
 - iv) Kekerapan sumber.
 - v) Pangkasan kes.



SUBSET DATA PERISTIWA:

- Antara teknik bagi tapisan kes:
 - i) Masa pemprosesan.
 - ii) Panjang jejak.
 - iii) Keberlakuan suatu aktiviti.
 - iv) Titik akhir.
 - v) Duluan.
 - vi) Kekerapan jejak.
- Setiap teknik tapisan mempunyai hujah songsang, yang membolehkan penapisan terbalik dapat dijalankan.



PENGAGREGATAN DATA PERISTIWA:

- Proses pengagregatan dijalankan untuk menggabungkan sebahagian aktiviti dengan aktiviti yang lain.

- Antara teknik pengagregatan:

i) Pengagregatan Is-a:

- Dibuat dengan mengeluarkan perbezaan antara jenis aktiviti yang serupa atau menukarnya kepada nama seragam.

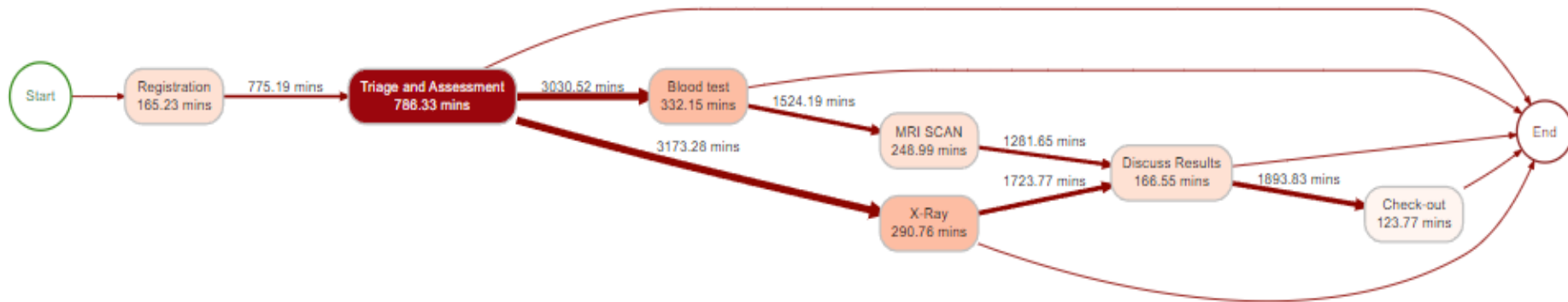
ii) Pengagregatan Part-of:

- Dibuat dengan menggabungkan aktiviti yang tergolong bersama sebagai sub-proses di bawah nama aktiviti peringkat lebih tinggi.



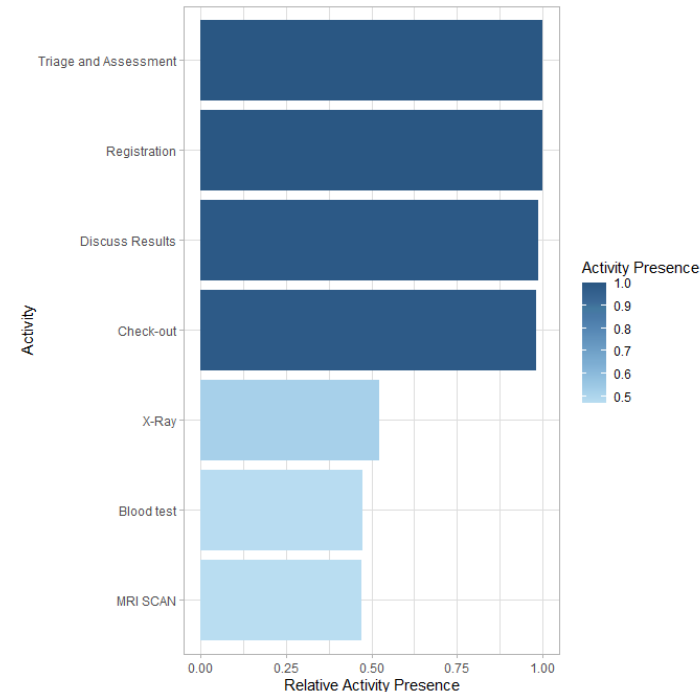
PEMETAAN KEKERAPAN PROSES:

- Antara kaedah pemetaan proses adalah berdasarkan:
 - Kekerapan mutlak.
 - Kekerapan relatif.
 - Profil prestasi.



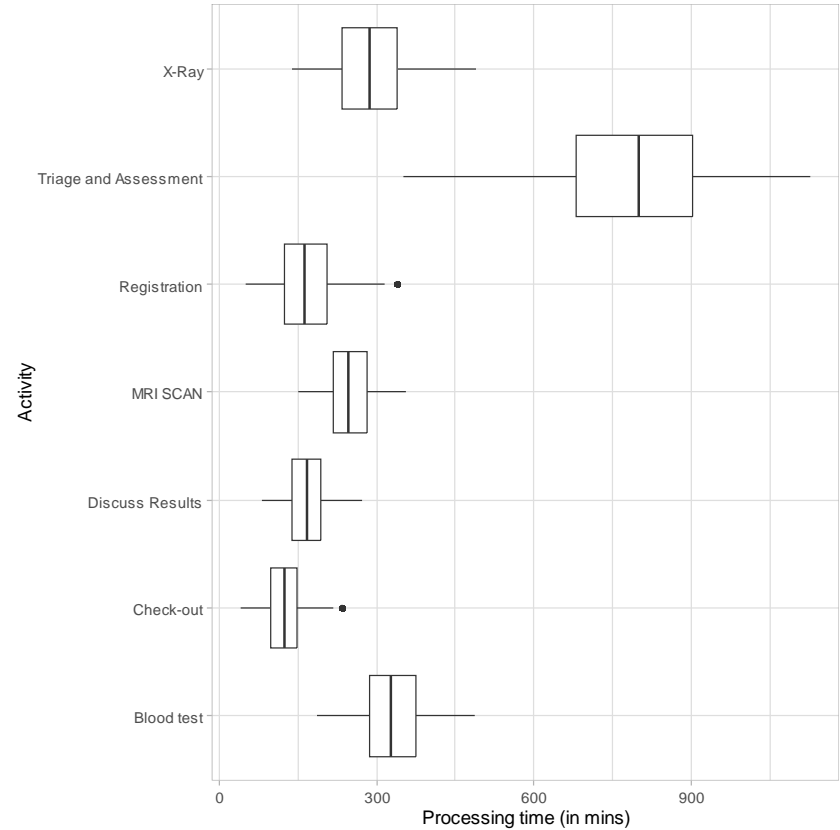
ANALISIS KAWALAN-ALIRAN:

- Analisis kawalan-aliran (*control flow analysis*) terhadap data log-peristiwa meliputi teknik:
 - i) **Perkadaran aktiviti:** peratus kes aktiviti tertentu.
 - ii) **Liputan dan panjang jejak:** menunjukkan hubungan antara bilangan urutan aktiviti yang berbeza (iaitu jejak) dan bilangan kes yang diliputinya.
 - iii) **Pematuhan berasaskan-aturan:** semak pematuhan aturan dalam log-peristiwa.
 - iv) **Penapisan berasaskan-aturan:** tapis kes yang mematuhi satu atau lebih peraturan.



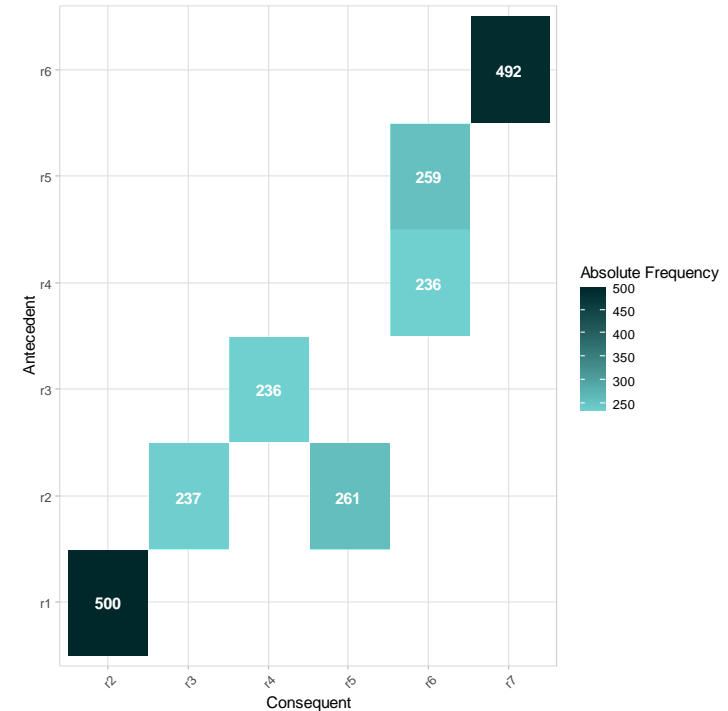
ANALISIS PRESTASI:

- Analisis prestasi (*performance analysis*) terhadap data log peristiwa diukur menerusi tiga ukuran tersebut :
 - Masa pemprosesan**: jumlah tempoh semua aktiviti.
 - Masa keseluruhan kes (*throughput*)**: masa antara aktiviti pertama kes hingga aktiviti terakhir.
 - Masa terbiar (*idle*)**: : masa apabila tiada aktiviti aktif.



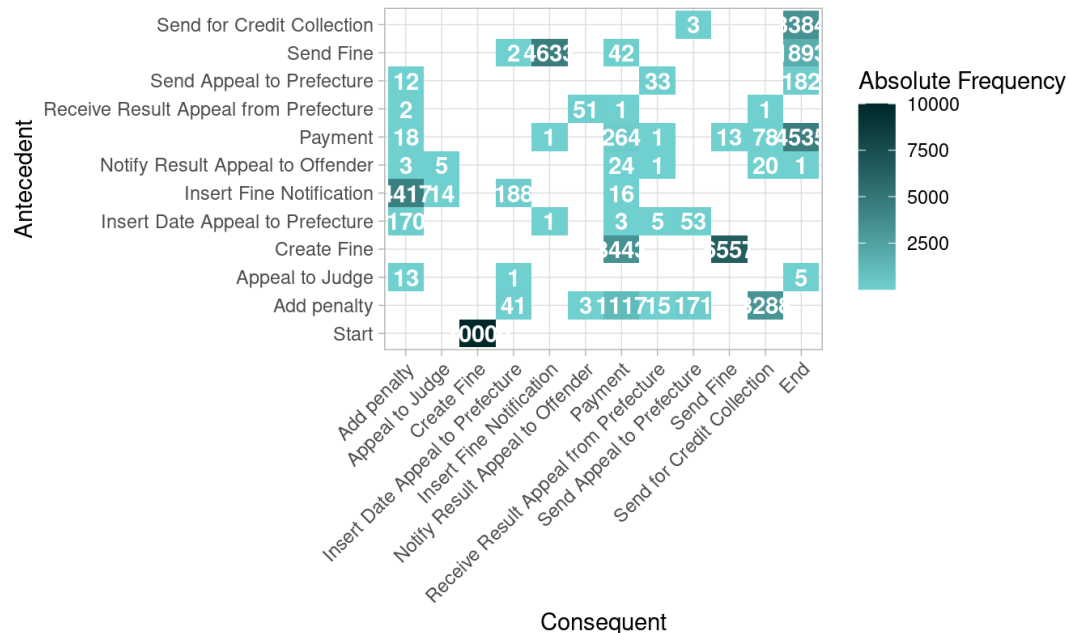
ANALISIS ORGANISASI:

- Analisis organisasi (*Organizational analysis*) terhadap data log peristiwa dijalankan untuk menilai:
 - i) **Kekerapan sumber:** bilangan kes yang diurus oleh setiap sumber.
 - ii) **Penglibatan sumber:** perkadaran relatif sumber yang terlibat dalam semua kes.
 - iii) **Pengkhususan sumber:** kenalpasti sama ada sumber dikhususkan dalam aktiviti tertentu atau tidak.
 - iv) **Rangkaian penyerahan kerja:** menunjukkan pemetaan aktiviti peralihan sumber.
 - v) **Matriks duluan sumber:** paparan yang lebih ringkas berkenaan pemetaan sumber.



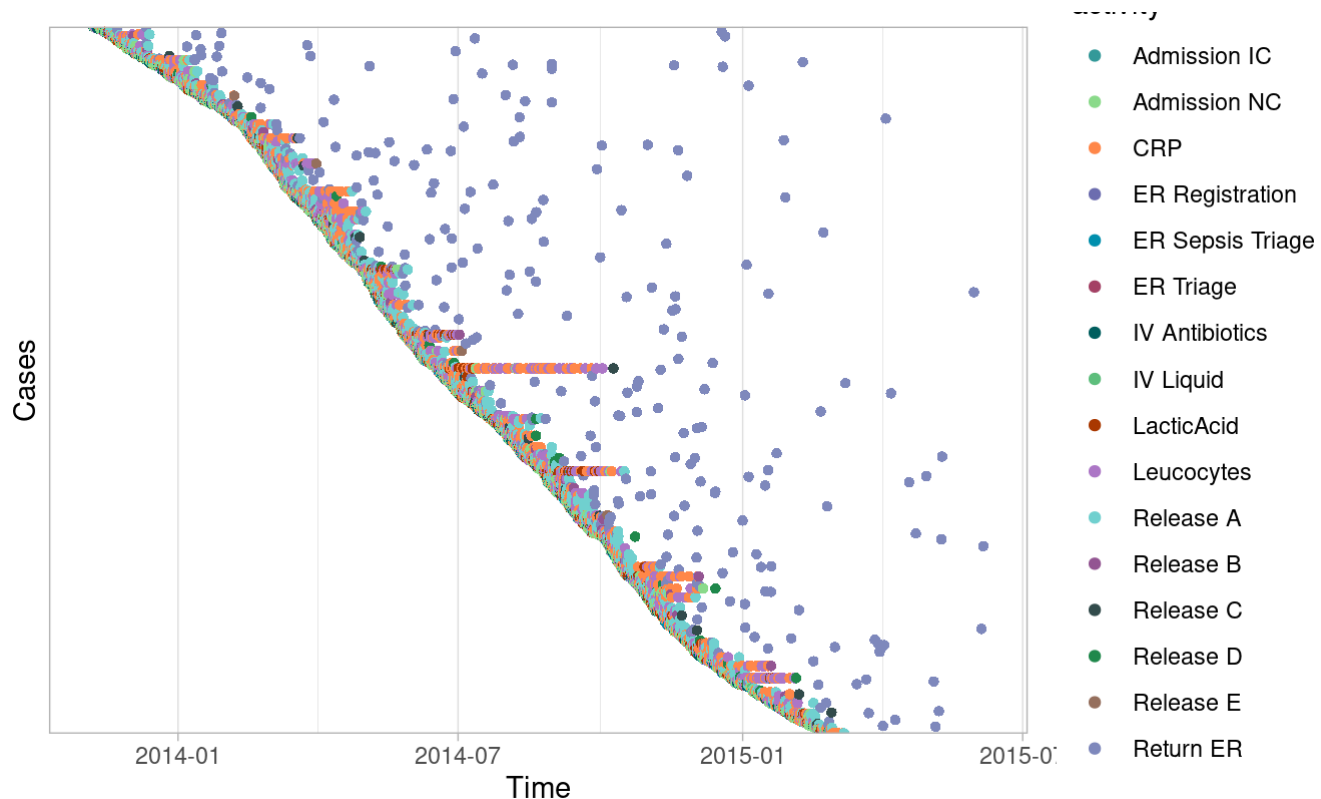
PLOT MATRIKS PROSES:

- Matriks proses ialah matriks dua dimensi yang menunjukkan aliran antara aktiviti berdasarkan:
 - i) Kekerapan aliran mutlak.
 - ii) Kekerapan kes relatif aliran.
 - iii) Kekerapan relatif aliran untuk setiap anteseden.
 - iv) Kekerapan relatif aliran untuk setiap konsekuen.



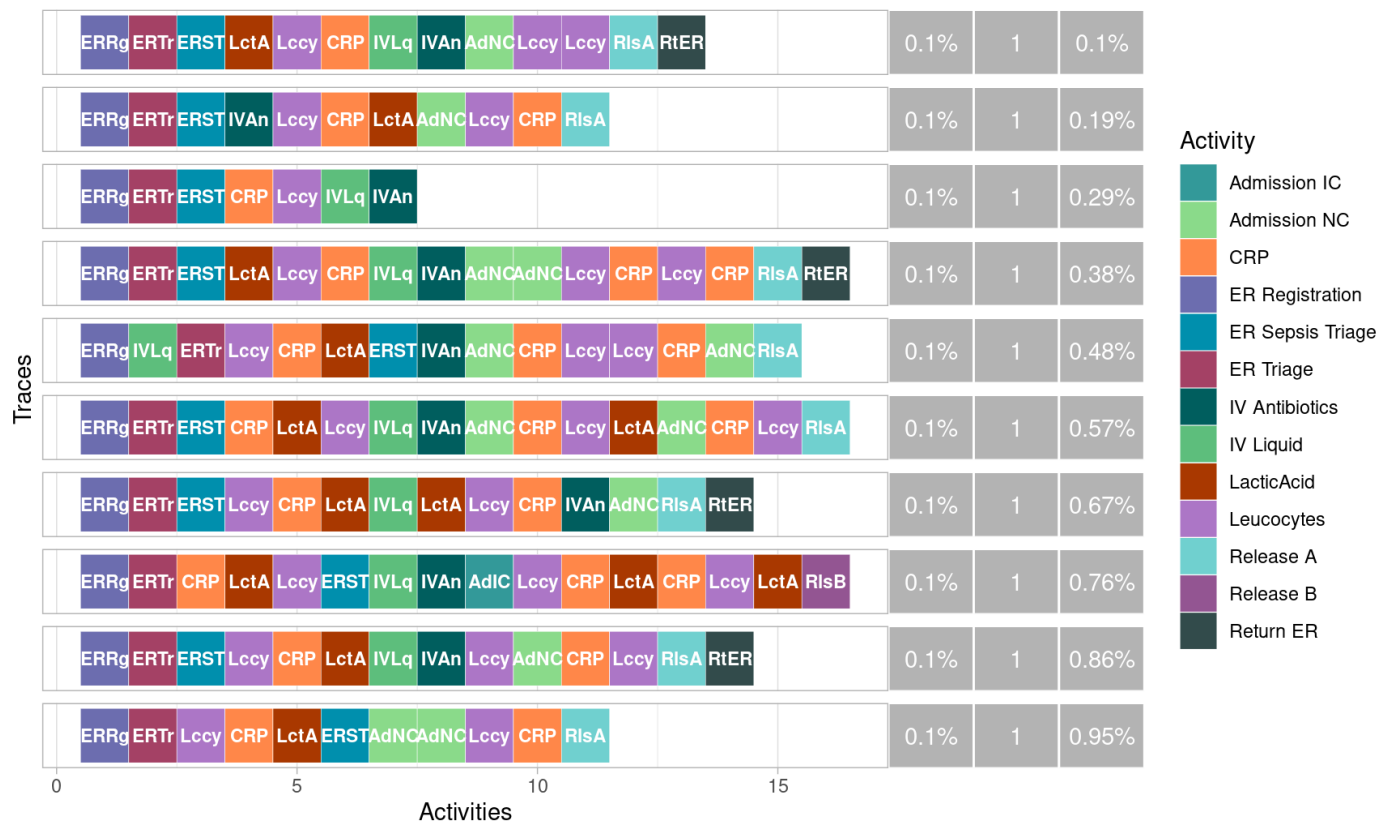
PLOT CARTA TITIK:

- Plot carta bertitik ialah graf di mana setiap kejadian aktiviti dipaparkan dengan titik.
- Paksi-x merujuk kepada aspek masa, manakala paksi-y merujuk kepada kes.



PLOT JEJAK PERISTIWA:

- Plot jejak peristiwa memaparkan urutan aktiviti yang berbeza dalam data log-peristiwa.
- Ia digunakan untuk meneroka jejak aktiviti yang kerap dan juga jejak aktiviti yang jarang berlaku.



RUJUKAN:

- Okoye, K. (2020). *Applications and Developments in Semantic Process Mining*. Engineering Science Reference.
- Reinkemeyer, L. (2020). *Process Mining in Action: Principles, Use Cases and Outlook*. Springer.
- Ronny S., Mans, van der Aalst, W.M.P, & Rob J.B. Vanwersch. (2015). *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*, Springer.
- van der Aalst, W. (2016) *Process mining: data science in action. Second Edition*, Springer.
- van der Aalst, W.M.P., & Carmona, J. (2022). *Process Mining Handbook*, Springer.



PEPERIKSAAN AKHIR:

TARIKH: 22 FEB 2025

MASA: 9.00 AM - 12.00 PM

**TEMPAT: MAKMAL BETA,
JABATAN SAINS MATEMATIK**

SELAMAT MAJU JAYA!!!

