Bab 7

Analisa Kinerja Multi Parameter dengan Teknik Karakterisasi Beban kerja (Workload)

Dr. Yeffry Handoko Putra , M.T

KARAKTERISASI BEBAN KERJA

Merupakan proses memodelkan beban kerja yang dapat diulang dalam kondisi riil

Contoh Komponen Beban kerja [Raj Jain]

- Perilaku Aplikasi: e-mail, text editing pengembangan program, updating. Periaku aplikasi dapat dikarakterisasi
- Situs: lokasi organisasi, lokasi web
- Sesi pengguna: sesi lengkap pengguna mulai dari log in sampai log out

TEKNIK KARAKTERISASI BEBAN KERJA

1. Perata-rataan (*Averaging*)
2. Spesifikasi Penyebaran (*Specifying Dispersion*)
3. Histogram parameter tunggal (*Single-parameter Histogram*)
4. Histrogram parameter ganda (*Multiparameter Histogram*)
5. Analisis Komponen Dasar (*Principal-Component Analysis*)
6. Model Markov (*Markov models*)
7. Pengkelompokan (*Clustering*)

Analisis Multi Parameter, Multi Kelas

Perata-rataan (*Averaging*)

Arithmetic mean:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Spesifikasi Penyebaran (*Specifying Dispersion*)

Variasi dinyatakan dengan varian (Var) dan ditulis sebagai kuadrat standar deviasi s^2

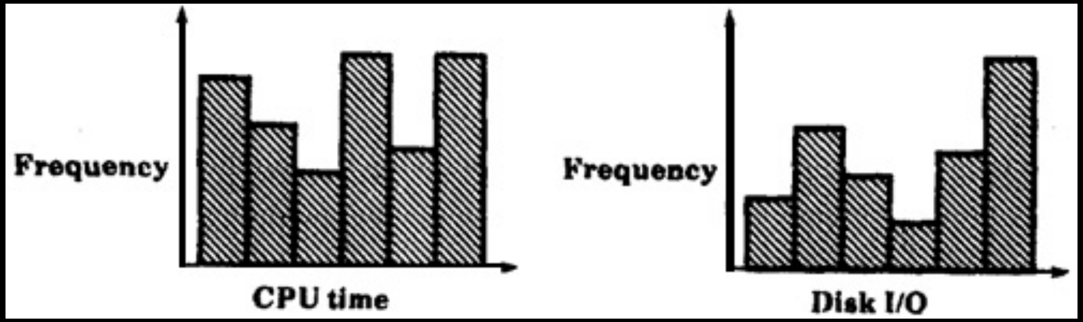
$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Rasio standar deviasi terhadap mean disebut COV (Coefficient of Variance)

$$COV = \frac{s}{\bar{x}}$$

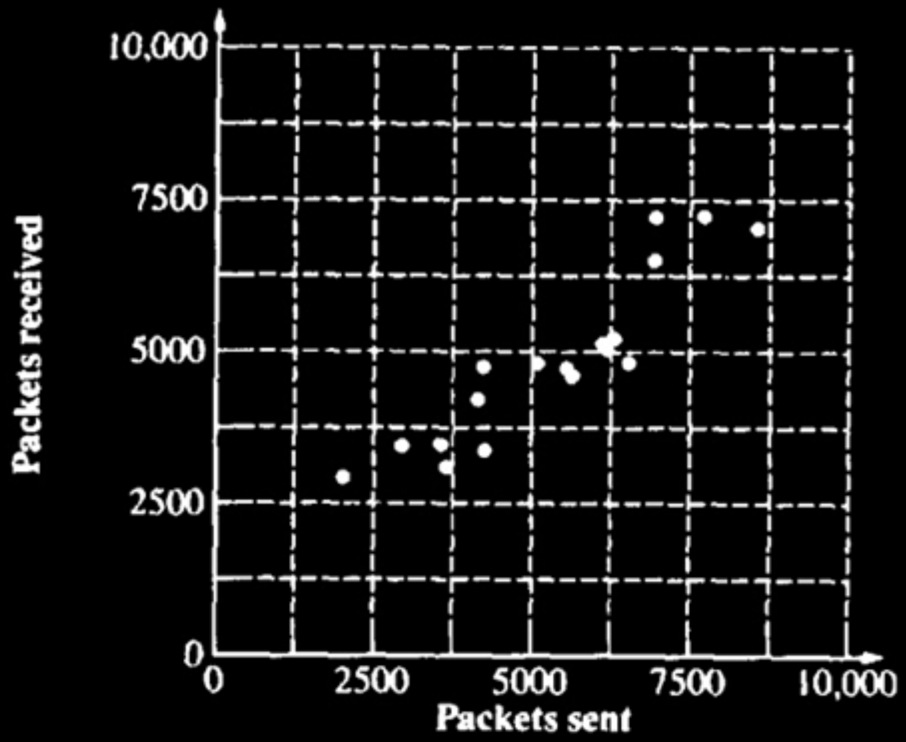
Jika COV bernilai nol berarti varian-nya nol dan parameter beban kerja bersifat konstan

Histogram parameter tunggal
(*Single-parameter Histogram*)



- Histogram menyatakan frekuensi dari beberapa nilai parameter. Untuk parameter kontinu rentang (cell/bucket) perlu dibagi lebih kecil lagi

Histrogram parameter ganda
(*Multiparameter Histogram*)



Matriks (histrogram) n-dimensi digunakan untuk menyatakan distribusi n-parameter beban kerja

Analisis Komponen Dasar
(*Principal-Component Analysis*)

|Digunakan untuk mengkarakterisasi / mengklasifikasi beban kerja menggunakan pembobotan (a_j)

$$y_i = \sum_{j=1}^{n} a_{ij} x_j$$

Digunakan untuk mengklasifikasi komponen beban kerja (x_j) menjadi kelas-kelas tertentu

Analisis Komponen Dasar adalah metoda untuk mencari pembobot (a_j) sehingga perbedaan antara masing-masing kelas menjadi jelas (maksimum)

Syarat yang harus dipenuhi:

Masing-masing kelas y_i harus tidak berkorelasi, yang dipenuhi dengan persamaan orthogonal:

$$\langle y_i | y_j \rangle = \sum_k a_{ik} x_{kj} = 0$$

Contoh:

Jumlah paket data yang dikirim dan diterima dinyatakan dengan x_s dan x_p. Paket ini dikirim dan diterima pada beberapa LAN. Dari gambar plot terdapat korelasi antara kedua variabel ini

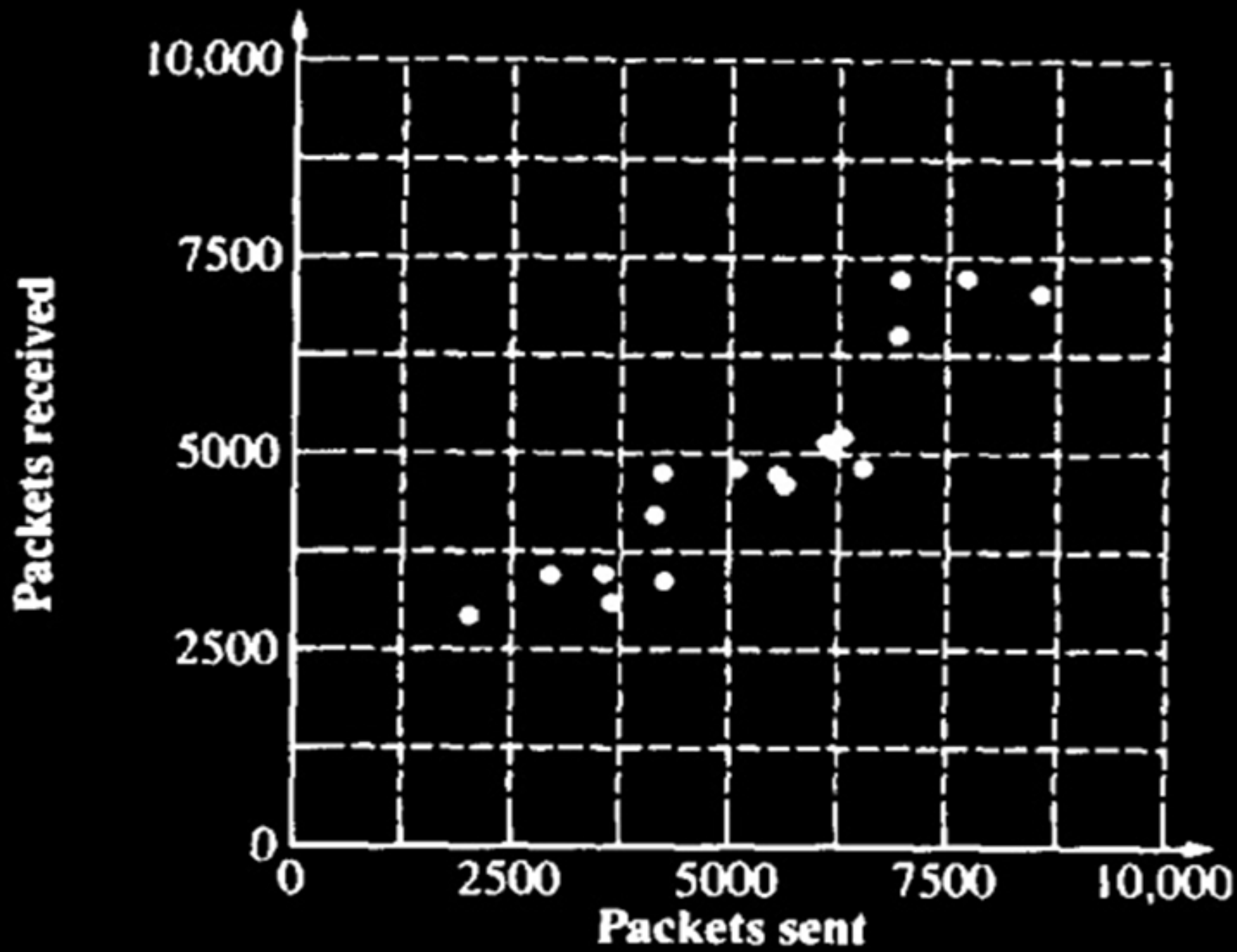

10,000
7500
5000
2500
0
Packets received
0
2500
5000
7500
10,000
Packets sent

TABLE 6.4 A Data for Principal-Component Analysis Example 6.1

Observation No.	Variables		Normalized Variables		Principal Factors	
	x_z	x_r	x'_z	x'_r	y_1	y_2
1	7718	7258	1.359	1.717	2.175	–0.253
2	6958	7232	0.922	1.698	1.853	–0.549
3	8551	7062	1.837	1.575	2.413	–0.186
4	6924	6526	0.903	1.186	1.477	–0.200
5	6298	5251	0.543	0.262	0.570	0.199
6	6120	5158	0.441	0.195	0.450	0.174
7	6184	5051	0.478	0.117	0.421	0.255
8	6527	4850	0.675	–0.029	0.457	0.497
9	5081	4825	–0.156	–0.047	–0.143	–0.077
10	4216	4762	–0.652	–0.092	–0.527	–0.396
11	5532	4750	0.103	–0.101	0.002	0.145
12	5638	4620	0.164	–0.195	–0.022	0.254
13	4147	4229	–0.692	–0.479	–0.828	–0.151
14	3562	3497	–1.028	–1.009	–1.441	–0.013
15	2955	3480	–1.377	–1.022	–1.696	–0.251
16	4261	3392	–0.627	–1.085	–1.211	0.324
17	3644	3120	–0.981	–1.283	–1.601	0.213
18	2020	2946	–1.914	–1.409	–2.349	–0.357
ςx	96,336	88,009	0.000	0.000	0.000	0.000
ςx^2	567,119,488	462,661,024	17.000	17.000	32.565	1.435
Mean	5352.0	4899.4	0.000	0.000	0.000	0.000
Standard Deviation	1741.0	1379.5	1.000	1.000	1.384	0.290

1. Hitung mean dan standar deviasi:

$$\overline{x}_s = \frac{1}{n}\sum_{i=1}^{n} x_{si} = \frac{96{,}336}{18} = 5352.0$$

$$\overline{x}_r = \frac{1}{n}\sum_{i=1}^{n} x_{ri} = \frac{88{,}009}{18} = 4889.4$$

$$s_{x_s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{si} - \overline{x}_s)^2$$

$$= \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x_{si}^2\right) - n\overline{x}_s^2\right]$$

$$= \frac{567{,}119{,}488 - 18 \times 5352^2}{17} = 1741.0$$

Juga untuk

$$s^2_{xr} = \frac{462{,}661{,}024 - 18 \times 4889.4^2}{17} = 1379.5$$

2. *Normalisasi variabel terhadap rata-rata nol (zero mean) dan standar deviasi*. Nilai ternormalisasi x'_s dan x'_r, diberikan oleh :

$$x'_s = \frac{x_s - \bar{x}_s}{s_{x_s}} = \frac{x_s - 5352}{1741}$$

$$x'_r = \frac{x_r - \bar{x}_r}{s_{x_r}} = \frac{x_r - 4889}{1380}$$

3. Hitung korelasi antar variabel:

$$R_{x_s x_r} = \frac{(1/n)\sum_{i=1}^{n}(x_{si} - \bar{x}_s)(x_{ri} - \bar{x}_r)}{s_{x_s} s_{x_r}} = 0.916$$

4. Siapkan Matrik Korelasi

$$C = \begin{bmatrix} 1.000 & 0.916 \\ 0.916 & 1.000 \end{bmatrix}$$

5. Hitung Nilai Eigen dari matriks korelasi:

$$|\lambda I - C| = \begin{vmatrix} \lambda - 1 & -0.916 \\ -0.916 & \lambda - 1 \end{vmatrix} = 0 \quad \text{atau}$$

$$(\lambda - 1)^2 - 0.916^2 = 0$$

Diperoleh nilai Eigen: 1.916 dan 0.084

6. Hitung Vektor Eigen dari matriks korelasi:

Vektor Eigen untuk nilai Eigen λ_1=1.916 adalah

$$Cq_1 = \lambda_1 q_1 \quad \text{atau}$$

$$\begin{bmatrix} 1.000 & 0.916 \\ 0.916 & 1.000 \end{bmatrix} \begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix} = 1.916 \begin{bmatrix} q_{11} \\ q_{21} \end{bmatrix}$$

didapati $q_{11} = q_{21}$

sehingga

$$q_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Sama juga untuk mencari vektor Eigen dari λ_2=0.084

$$q_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Diperoleh factor dasar (principal factor) dengan mengkalikan vektor Eigen dengan vektor ternormalisasi:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{x_s - 5352}{1741} \\ \frac{x_r - 4889}{1380} \end{bmatrix}$$

Hasilnya adalah:

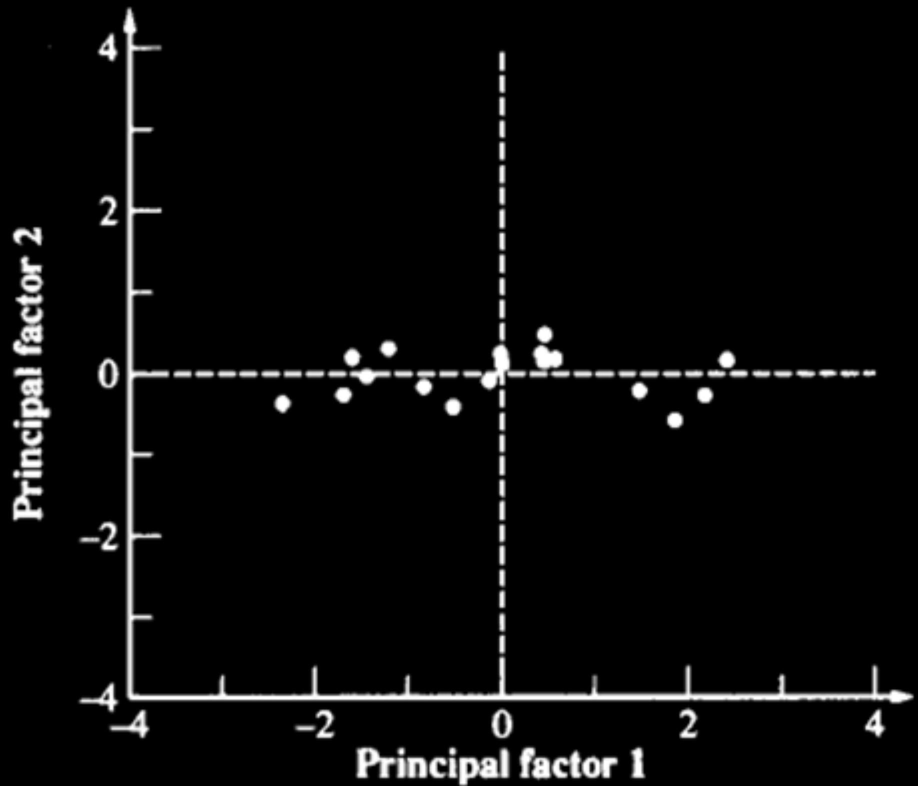
Observation No.	Variables		Normalized Variables		Principal Factors	
	x_s	x_r	x'_s	x'_r	y_1	y_2
1	7718	7258	1.359	1.717	2.175	−0.253
2	6958	7232	0.922	1.698	1.853	−0.549
3	8551	7062	1.837	1.575	2.413	0.186
4	6924	6526	0.903	1.186	1.477	−0.200
5	6298	5251	0.543	0.262	0.570	0.199
6	6120	5158	0.441	0.195	0.450	0.174
7	6184	5051	0.478	0.117	0.421	0.255
8	6527	4850	0.675	−0.029	0.457	0.497
9	5081	4825	−0.156	−0.047	−0.143	−0.077
10	4216	4762	−0.652	−0.092	−0.527	−0.396
11	5532	4750	0.103	−0.101	0.002	0.145
12	5638	4620	0.164	−0.195	−0.022	0.254
13	4147	4229	−0.692	−0.479	−0.828	−0.151
14	3562	3497	−1.028	−1.009	−1.441	−0.013
15	2955	3480	−1.377	−1.022	−1.696	−0.251
16	4261	3392	−0.627	−1.085	−1.211	0.324
17	3644	3120	−0.981	−1.283	−1.601	0.213
18	2020	2946	−1.914	−1.409	−2.349	−0.357
$\sum x$	96,336	88,009	0.000	0.000	0.000	0.000
$\sum x^2$	567,119,488	462,661,024	17.000	17.000	32.565	1.435
Mean	5352.0	4889.4	0.000	0.000	0.000	0.000
Standard Deviation	1741.0	1379.5	1.000	1.000	1.384	0.290

Harus nol

varian

Varian y1: 32.565/(32.565 + 1.435)=95.7%
Varian y2=4.3 %

Varian paling banyak di faktor dasar ke satu, terlihat:



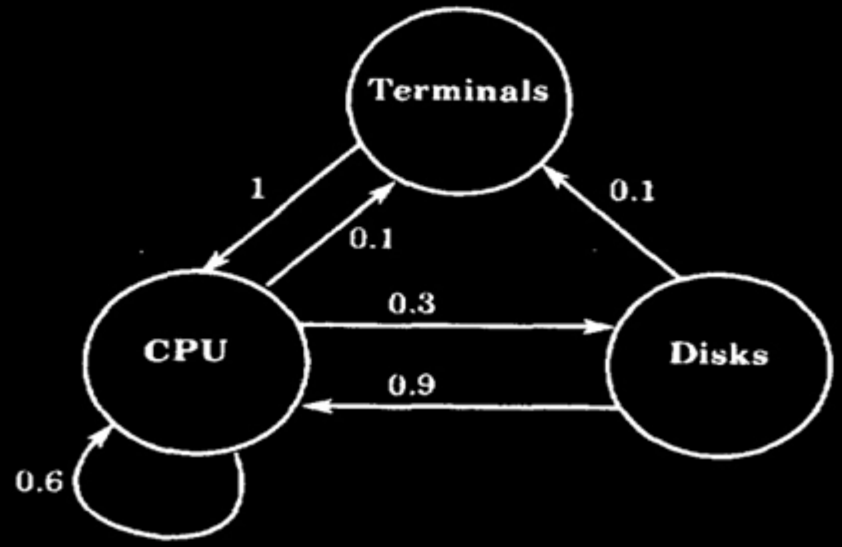
Model Markov (*Markov models*)

- Dalam Analisis kinerja terkadang kinerja sekarang ditentukan oleh kinerja sebelumya.
- Sehingga yang diperlukan bukan hanya jumlah permintaan layanan tetapi urutan kejadian lebih penting dan jika jumlah permintaan layanan berikutnya ditentukan hanya oleh layanan sebelumnya maka digunakan model Markov.
- Digunakan pada Analisis model Antrian
- P(transisi ke status berikutnya)=f(status sekarang)≠ f(semua status sebelumnya)

Contoh Model Markov

TABLE 6.5 Transition Probability Matrix

From/To	CPU	Disk	Terminal
CPU	0.6	0.3	0.1
Disk	0.9	0	0.1
Terminal	1	0	0



Contoh: Pada pengiriman paket data dibagi menjadi paket besar dan paket kecil, paket berukuran kecil mendominasi traffic sebanyak 80%. Rata-rata 4 paket kecil (k) dilanjuti oleh satu paket besar(b). Atau urutannya ditulis:

kkkkbkkkkbkkkk

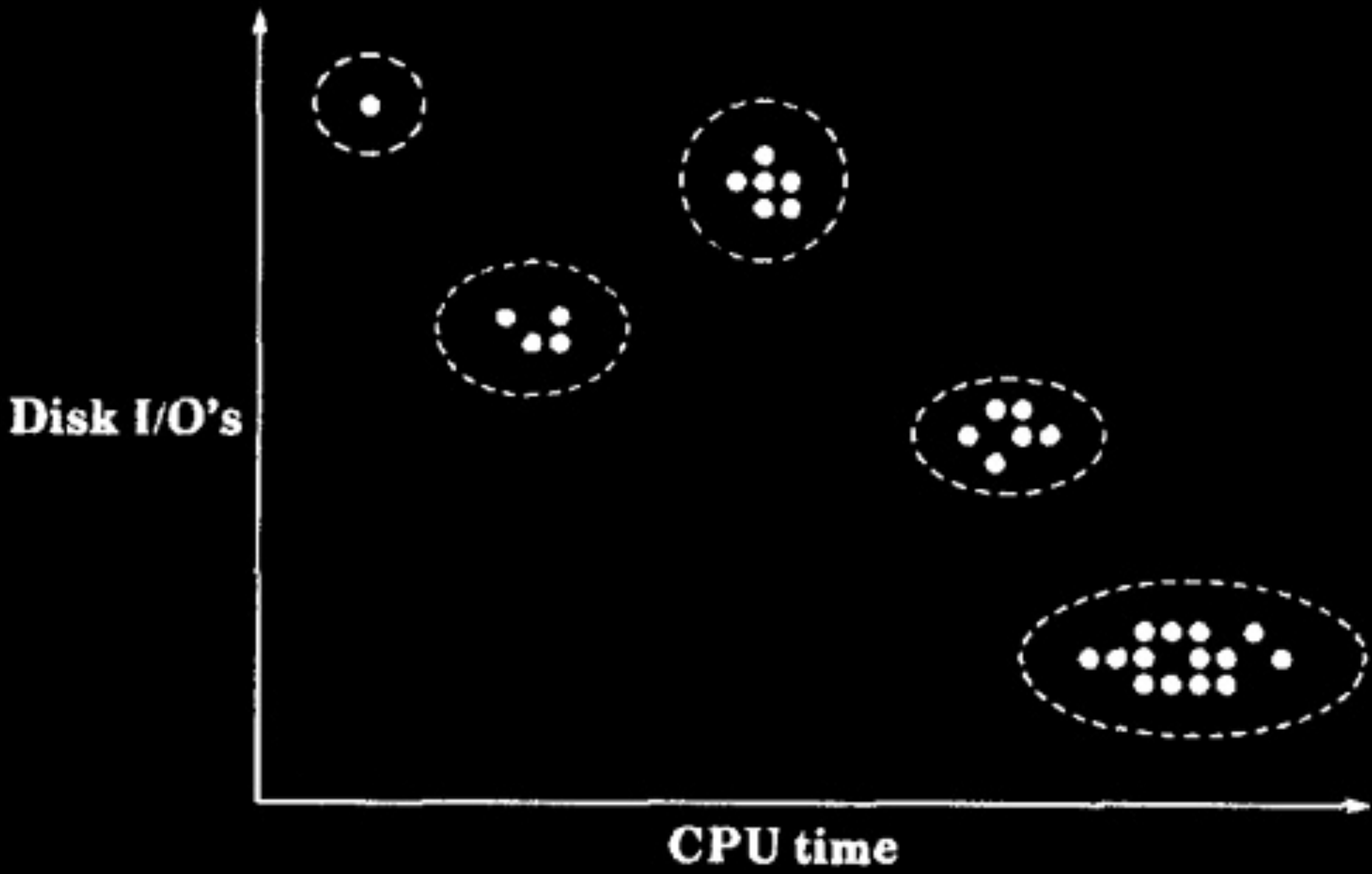
Sehingga matrik probabilitas transisinya:

	Paket berikutnya	
Paket Skrg	Kecil	Besar
Kecil	0.75	0.25
Besar	1	0

CLUSTERING

Klasifikasi Kinerja menjadi beberapa kelas kecil yang disebut klaster.

CPU dan I/O Disk membutuhkan 30 pekerjaan. Pekerjaan dapat dinyatakan hanya dengan 5 klaster



Untuk mengkarakterisasi data beban kerja menjadi klaster , dilakukan langkah-langkah berikut:

1. Ambil sampel dari beban kerja
2. Pilih parameter beban kerja
3. Transormasikan parameter bila perlu
4. Hilangkan pencilan (outlier/outstanding)
5. Skalakan semua pengamatan
6. Pilih jarak pengukuran
7. Tampilkan klaster
8. Interprestasikan klaster yang dihasilkan
9. Ubah parameter atau ubah jumlah klaster, ulangi tahap 3 sampai 7
10. Pilih komponen representative dari setiap klaster

1. Sampling
   Salah satu metoda sampling adalah pemilihan acak
2. Pemilihan Parameter
   Dua kunci pemilihan parameter adalah berdasarkan dampaknya dan variannya
3. Transformasi
   Jika distribusi dari parameternya condong maka perlu transformasi misalkan dengan transformasi logaritmik
4. Pencilan
   Data dengan nilai sangat ekstrim disebut pencilan / outlier. Hanya pencilan yang tidak mengambil porsi dari sumber daya sistem dapat dihilangkan

5. Penskalaan Data

a. Normalisasi ke rata-rata nol (zero mean) dan satuan varian: $x'_{ik} = \frac{x_{ik} - \bar{x}_k}{s_k}$

b. Pembobotan: $x'_{ik} = w_k x_{ik}$

c. Normalisasi Rentang, rentang diubah dari [$x_{\min,k}$, $x_{\max,k}$] jadi [0, 1]: $x'_{ik} = \frac{x_{ik} - x_{min,k}}{x_{max,k} - x_{min,k}}$

d. Normalisasi Percentile, Data diskalakan sehingga 95% nilainya berada di antara 0 dan 1:

$$x'_{ik} = \frac{x_{ik} - x_{2.5k}}{x_{97.5k} - x_{2.5k}}$$

6. Jarak Metriks, kedekatan data terhadap suatu klaster dinyatakan dengan pengukuran jarak. Tiga metoda yang sering digunakan:

   a. Jarak Euclidian:

$$d = \left\{ \sum_{k=1}^{n} \left(x_{ik} - x_{jk} \right)^2 \right\}^{0.5}$$

   b. Jarak Euclidian terboboti :

$$d = \left\{ \sum_{k=1}^{n} a_k \left(x_{ik} - x_{jk} \right)^2 \right\}^{0.5}$$

a_k = pembobot

c. Jarak Chi-Square:

$$d = \sum_{k=1}^{n} \left\{ \frac{\left(x_{ik} - x_{jk}\right)^2}{x_{ik}} \right\}$$

7. Teknik Klaster

Dasar dari *clustering* adalah pemilahan yang membuat anggota dari suatu klas semirip mungkin dan jauh berbeda dengan klas yang lain. Meminimalkan varian intragroup dalam:

Total varian=varian intragroup+varian intergroup

Metoda Minimum Pohon Rentang

(*Minimum Spanning Tree Method*)

1. Mulai dengan k = n klaster
2. Cari titik pusat klaster ke-i, i=1, 2, …, k

   Nilai titik pusat sama dengan rata-rata semua titik di dalam klaster
3. Hitung matriks jarak intercluster
4. Cari jarak terkecil selain nol dari matriks jarak, d_{lm}

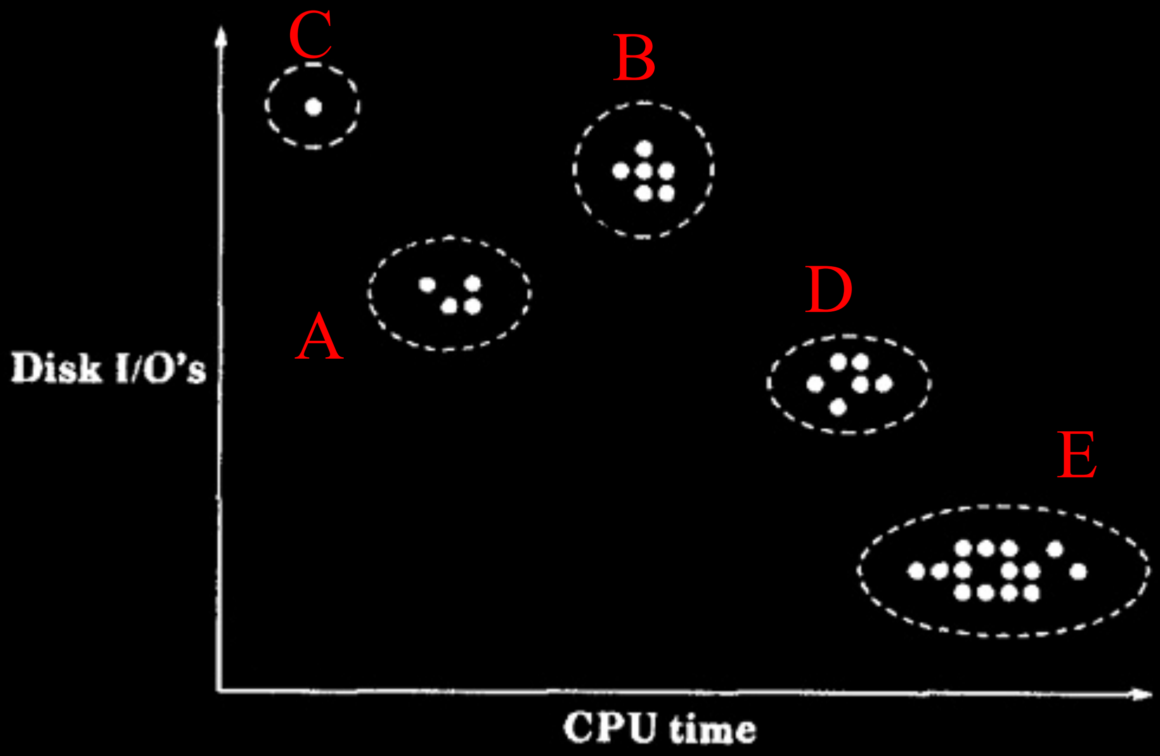
5. Ulangi langkah 2-4 sehingga semua menjadi bagian suatu klaster

Contoh: Suatu beban kerja terdiri dari 5 komponen, 2 parameter. CPU time dan jumlah I/O Disk diukur untuk 5 program. Nilai parameter yang diskalakan :

Program	CPU Time	Disk I/O
A	2	4
B	3	5
C	1	6
D	4	3
E	5	2

Langkah 1: Anggap ada 5 klaster dengan klaster ke-i hanya beranggotakan program ke-i

Langkah 2: Titik Pusatnya adalah {2,4}, {3,5}, {1,6}, {4,3}, dan {5,2}



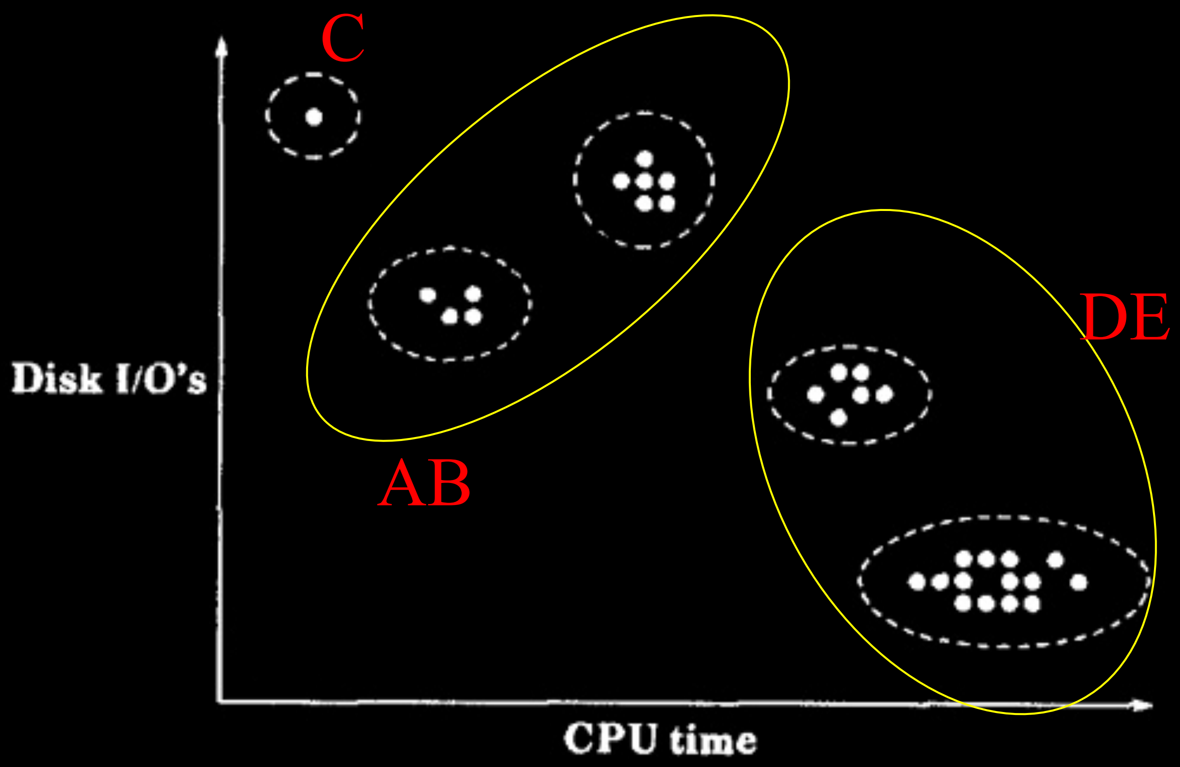
Langkah 3: Menggunakan pengukuran jarak Euclidian , matriks jaraknya:

	Program				
Program	A	B	C	D	E
A	0	$\sqrt{2}$	$\sqrt{5}$	$\sqrt{5}$	$\sqrt{13}$
B		0	$\sqrt{5}$	$\sqrt{5}$	$\sqrt{13}$
C			0	$\sqrt{18}$	$\sqrt{32}$
D				0	$\sqrt{2}$
E					0

Langkah 4: Jarak Interklaster minimum $\sqrt{2}$ antara A dan B, D dan E. Kedua pasangan titik ini kemudian digabung

Langkah 2: Titik pusat pasangan AB adalah {(2+3)/2, (4+5)/2} = {2.5, 4.5}. Juga pasangan DE adalah {4.5, 2.5}. Titik Pusat yang lain tetap sama

Langkah 3: Sehingga sekarang terdapat 3 klaster


C
AB
DE
Disk I/O's
CPU time

Dengan Matriks Jaraknya sekarang:

	Program		
Program	AB	C	DE
AB	0	$\sqrt{4.5}$	$\sqrt{8}$
C		0	$\sqrt{24.5}$
DE			0

Langkah 4: Jarak Interklaster minimum sekarang $\sqrt{4.5}$ antara AB dan C. Kedua klaster ini kemudian digabung

Langkah 2: Titik Pusat ABC adalah: {(2+3+1)/3, (4+5+6)/3} = {2,5}

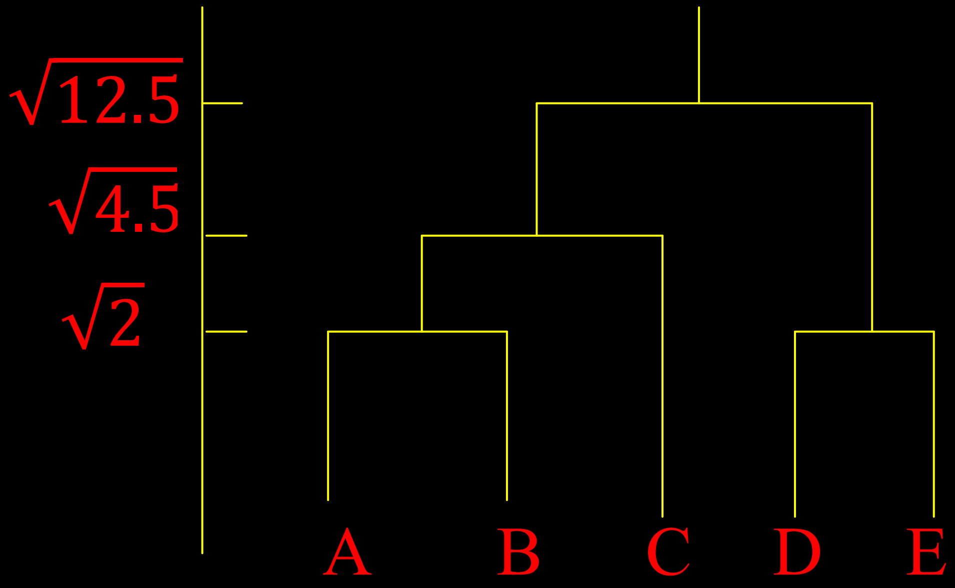
Langkah 3: Matriks jaraknya sekarang:

	Program	
Program	ABC	DE
ABC	0	$\sqrt{12.5}$
DE		0

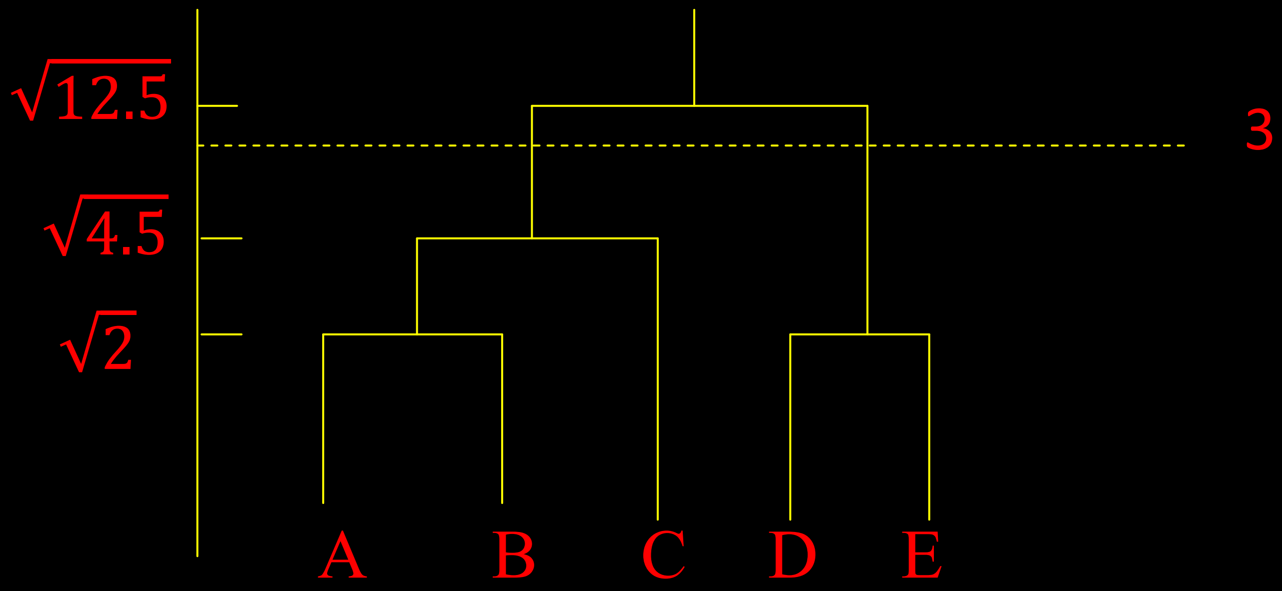
Langkah 4: Jarak Interklaster minimum: $\sqrt{12.5}$.
Penggabungan ABC dan DE jadi satu klaster ABCDE.

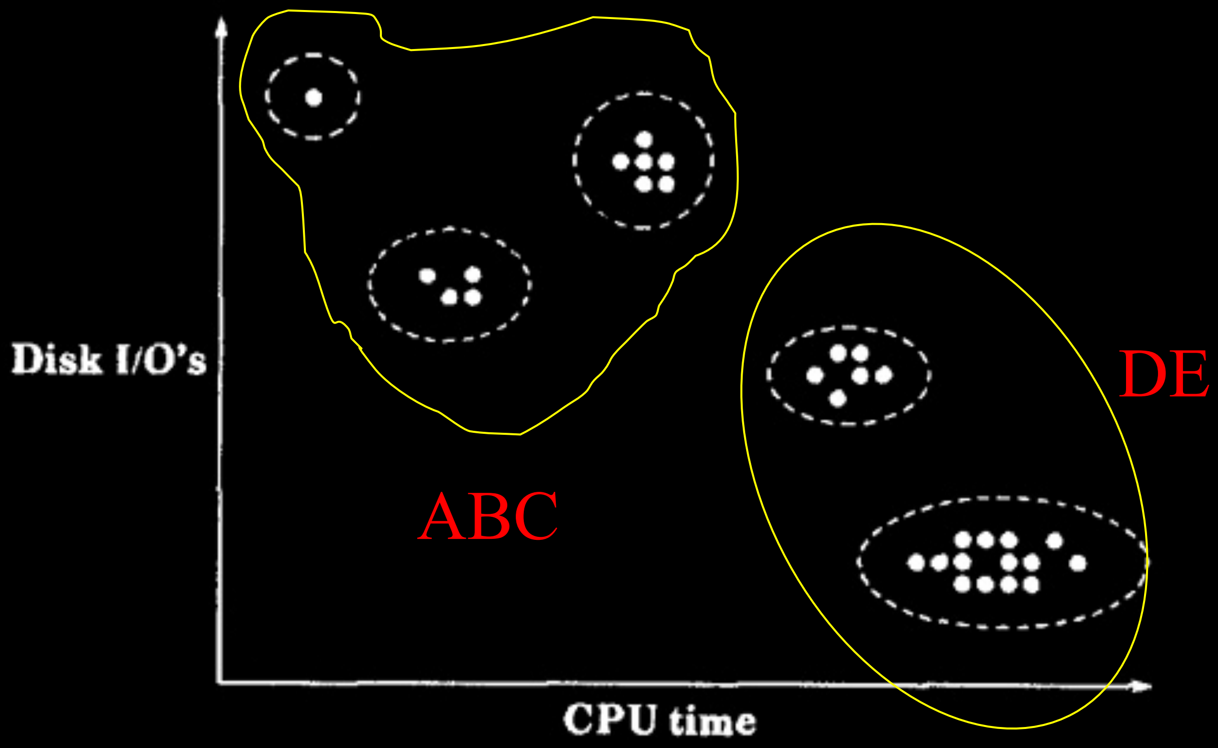
Jumlah klaster ditentukan dari jarak minimum interklaster atau jarak maksimum intraklaster

Dalam bentuk pohon rentang (spanning tree) digambarkan dalam bentuk dendogram :



Jika dipilih jarak maksimum intrakluster adalah 3 maka terdapat dua klaster yaitu ABC dan DE





Klastering berdasarkan permintaan sumber daya (*resource demand*) bukan jumlah populasi dalam klaster. Meski ada satu anggota tapi berdampak pada permintaan sumber daya, maka tidak dapat diabaikan