

# Università degli Studi di Napoli "Federico II"

Scuola Politecnica e delle Scienze di Base  
Corso di Laurea Magistrale in Ingegneria Informatica

Giuseppe Francesco Di Cecio - M63001211  
Nicola D'Ambra - M63001223  
Emma Melluso - M63001176

Elaborato di *Impianti di Elaborazione*



Università degli Studi di Napoli *Federico II*

Napoli

A.A 2020/2021

# Indice

<b>1</b>	<b>Workload Characterization</b>	<b>1</b>
1.1	Filtraggio . . . . .	1
1.1.1	Colonne Identiche . . . . .	1
1.1.2	Outlier . . . . .	2
1.2	PCA . . . . .	5
1.3	Clustering . . . . .	6
1.3.1	4 Componenti Principali . . . . .	6
1.3.2	5 Componenti Principali . . . . .	7
1.3.3	6 Componenti Principali . . . . .	7
1.3.4	Interpretazione . . . . .	8
1.4	Workload Sintetico . . . . .	8

# Capitolo 1

## Workload Characterization

Il dataset di partenza è composto da **3000 righe** e **24 colonne**, ciascuna delle quali rappresenta uno dei parametri del sistema oggetto di studio. Si tratta di parametri caratterizzanti l'esecuzione di vari Threads su un sistema operativo.

In particolar modo le colonne con prefisso *Vm* rappresentano informazioni sulla memoria virtuale occupata e utilizzata dai Threads, mentre le altre colonne rappresentano informazioni di carattere generale, come memoria libera, numero di threads, pagine inattive ecc.

### 1.1 Filtraggio

#### 1.1.1 Colonne Identiche

Innanzitutto osservando il workload e effettuando un grafico delle distribuzioni ci rende conto della presenza di ben 4 colonne costanti:

- **Active**
- **AnonPages**
- **AvbLatency**
- **Error**

Essendo tali non spiegano varianza, dunque possono essere tranquillamente trascurate ai fini dell'analisi.

Osservando le distribuzioni dei parametri **WriteBack** e **MemFree** si sono notate alcune caratteristiche comuni. Per avere una maggiore chiarezza si è preferito calcolare la matrice delle correlazioni su questi due parametri.



	'MemFree'	'Writeback'
'MemFree'	1,0000	1,0000
'Writeback'	1,0000	1,0000

Figura 1.1: *Matrice di correlazione tra MemFree e WriteBack*

Osservando la matrice appare evidente che le due colonne sono esattamente identiche, fornendo quindi la stessa informazione. Per questo motivo si è deciso di trascurare una delle due, in particolare quella di WriteBack.

Le 24 colonne iniziali sono state ridotte a 19 colonne, riducendo il dataset di un numero di osservazioni pari a:

$$n_{dati} = 3.000 \times (24 - 19) = 15.000 \quad (1.1)$$

### 1.1.2 Outlier

Gli outliers sono valori che si discostano notevolmente dalle altre osservazioni. Non possono essere subito eliminati poiché influiscono sulle analisi statistiche in modo considerevole. Si possono però analizzare per capire se è possibile rimuoverli o meno. Osservando l'andamento dei seguenti parametri, attraverso box plot e grafici di distribuzione:

- **VmSize**, quanta memoria virtuale utilizza l'intero processo.
- **VmHWM**, di quanta RAM il processo necessita al massimo.
- **VmRSS**, quanta RAM il processo sta correntemente usando.
- **VmPTE**, quanta memoria Kernel è occupata dalle entries della tabella delle pagine.

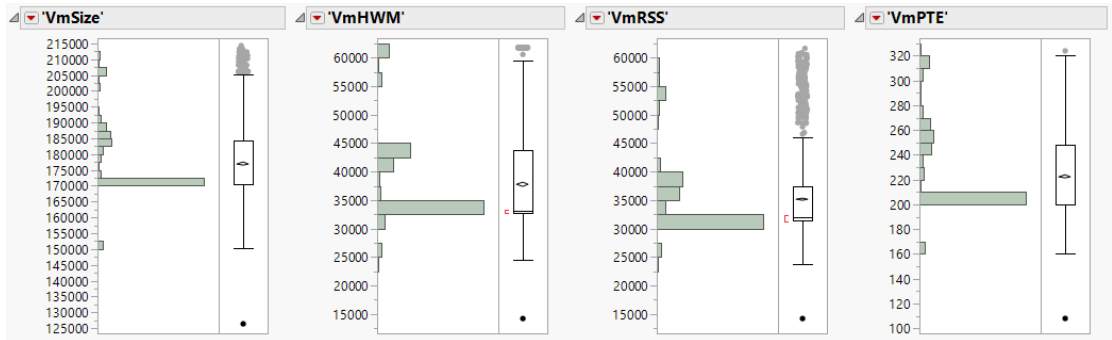
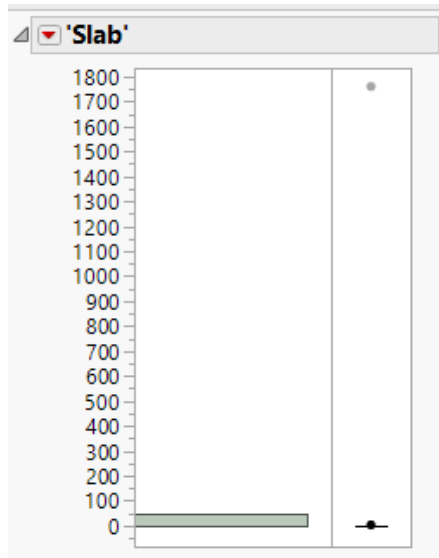


Figura 1.2: *Grafici di distribuzione di VmSize, VmHWM, VmRSS, VmPTE*

Si è notato che essi presentano un outlier isolato (in basso ad ogni grafico) in comune associato alla prima riga del dataset. Analizzando gli altri parametri (*MemFree*, *Dirty*, *PageTables*, *Buffer*, ...) è stato possibile evidenziare che anche per la maggior parte di essi lo è, ma non è un punto isolato.

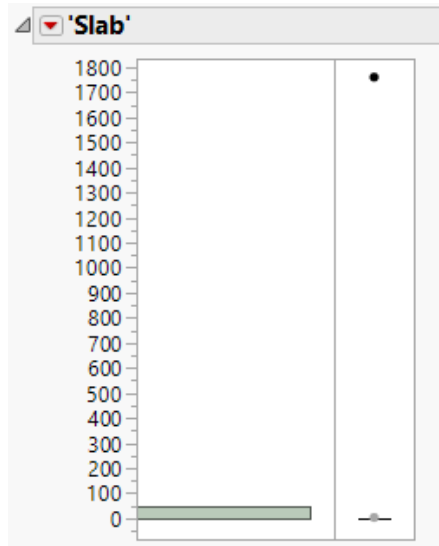
L'ipotesi fatta è che con molta probabilità le prime righe del dataset (da 0 a 100 circa), rappresentano la fase di avvio del processo e l'outlier oggetto di studio è la prima istanza di questa fase. Dato che l'obiettivo della caratterizzazione del workload è quello di analizzare le prestazioni a regime del sistema oggetto di studio (in questo caso), si è deciso di trascurare quel singolo outlier. E' bene notare che in ogni caso le informazioni riguardo questa fase di avvio non saranno del tutto perse dato che è stato rimosso un singolo punto e non tutti i punti che la rappresentano.

Un secondo outlier che può essere agevolmente rimosso è la riga 512 in cui il parametro **Slab** assume valore 4.

Figura 1.3: *Grafico di distribuzione di Slab*

Oltre ad avvicinarsi molto al valore medio assunto da Slab (zero), esso risulta essere un outlier solo per il parametro stesso dato che per gli altri è un valore compreso tra i quartili. Una sua rimozione quindi non influenza gli indici di caratterizzazione sintetica dei parametri del workload complessivo.

Un terzo outlier è il valore 1760 del parametro *Slab*, associato alla riga 90 del workload.

Figura 1.4: *Grafico di distribuzione di Slab*

Rispetto al precedente, tale outlier richiede un' analisi più approfondita visto che influenza significativamente l'andamento di parametri quali *Mapped* e *PageTables*. Per descrivere meglio la dipendenza tra questi parametri si può effettuare un grafico tra il numero dell'osservazione e il valore assunto da *Mapped*, analogo discorso con *PageTables*.

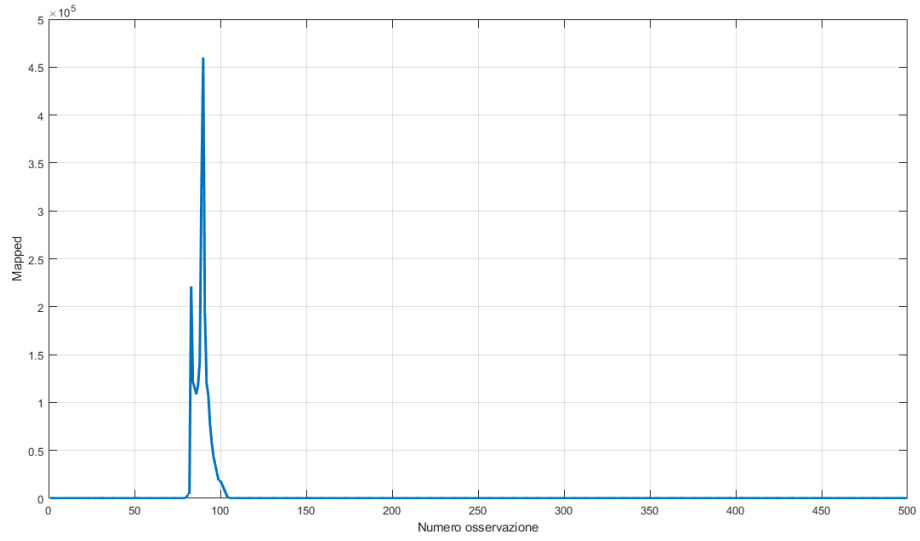


Figura 1.5: *Grafico tra numero di osservazione e valore assunto dal parametro Mapped*

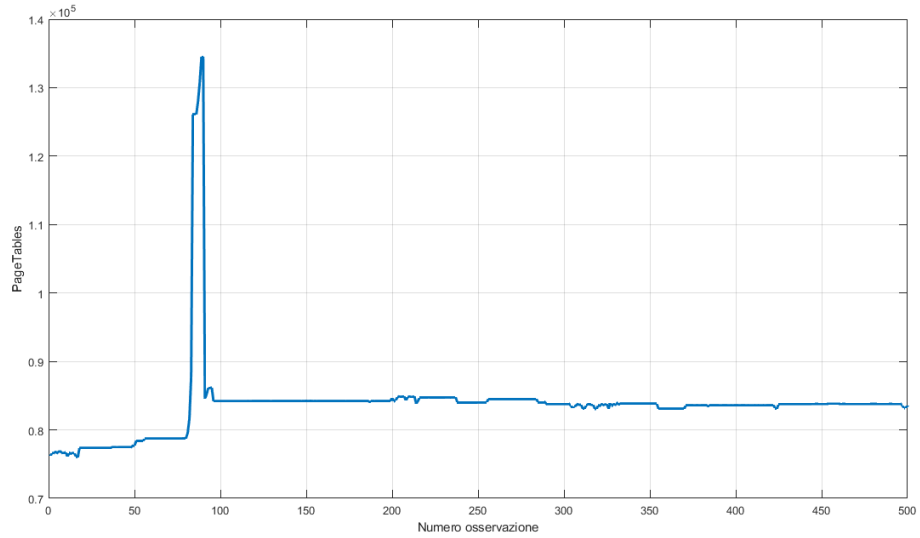


Figura 1.6: *Grafico tra numero di osservazione e valore assunto dal parametro PageTables*

Si nota che in corrispondenza (in realtà nell'osservazione appena precedente) dell'outlier del parametro *Slab* i due parametri sopra indicati hanno un picco, durante la fase di avvio del sistema.

Lo *Slab* si riferisce ad un particolare meccanismo di allocazione/deallocazione della memoria nel Kernel. Dato che influenza in particolar modo altri parametri si è preferito di non trascurarlo.

In conclusione sono stati eliminati dal dataset solo 2 outlier che corrispondono a 38 campioni. Quindi il dataset è stato ridotto in totale di 15.038 elementi, provocando una diminuzione dei dati iniziali di poco più del 20%.

## 1.2 PCA

A seguito del filtraggio il dataset risulta ridotto grazie alla rimozione di alcune colonne e righe che rappresentano outlier.

Sul questo dataset si possono quindi iniziare a fare le prime considerazioni.

Utilizzando la tecnica della *Principal Component Analysis* il dataset può essere estremamente ridotto, sfruttando solo le *Componenti Principali* che mantengono più varianza. Il risultato della PCA è quindi:

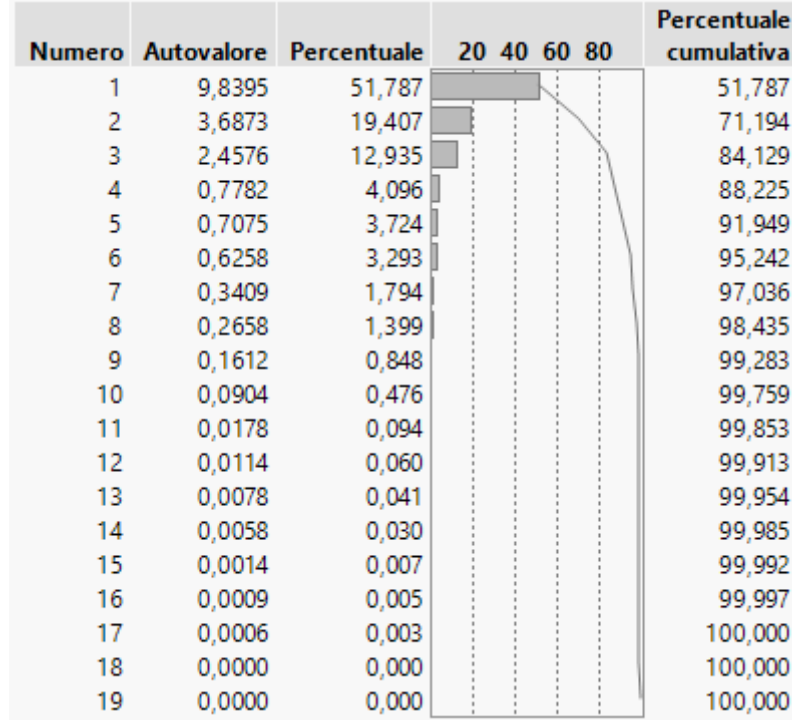


Figura 1.7: PCA applicata al dataset filtrato

La scelta del numero di componenti principali ricade in particolar modo sulla devianza che quelle componenti mantengono rispetto al dataset reale. Inoltre essa dipende anche dal tipo di osservazioni ed esperimento che è stato effettuato.

La scelta migliore ricade nel prendere 5 componenti principali poiché rappresentano il 92% della devianza totale. Esso rappresenta un valore anche abbastanza elevato, ma è stato scelto per mantenersi in una regione di tolleranza durante la clusterizzazione.

Per completezza si possono riportare le analisi considerando tre set di componenti principali nel seguente modo:

- Prendere 4 PC

$$DEV_{PCA-MANTENUTA} \approx 88\%$$

- Prendere 5 PC

$$DEV_{PCA-MANTENUTA} \approx 92\%$$

- Prendere 6 PC

$$DEV_{PCA-MANTENUTA} \approx 95\%$$

Sulla base di queste tre alternative, viene effettuata la procedura di clustering per ognuna di esse.

### 1.3 Clustering

Il clustering è una tecnica che consiste nel raggruppare osservazioni "simili" tra loro. La similitudine tra un elemento e un cluster, o tra un cluster e un altro cluster, può essere calcolata secondo varie tecniche. In questa analisi si è preferito utilizzare il **metodo di Ward**, il quale pesa la distanza tra due cluster in relazione al numero di elementi che li compongono.

Dati due cluster  $P$  e  $Q$  (un elemento non appartenente ad un cluster, può essere visto come un cluster di dimensione 1), sia  $|P|$  la cardinalità di  $P$ , analogo con  $|Q|$ , e sia  $\bar{x}_p$  il centroide di  $P$ , analogo con  $\bar{x}_q$ , la distanza tra  $P$  e  $Q$  viene calcolata come:

$$d(P, Q) = 2 \frac{|P| |Q|}{|P| + |Q|} \|\bar{x}_p - \bar{x}_q\|^2$$

Per ogni raggruppamento viene poi scelto una singola osservazione che la rappresenta, riducendo quindi il dataset pari al numero di cluster scelti durante l'analisi.

A tal proposito il numero di cluster da scegliere può dipendere da vari fattori:

- Omogeneità dei cluster: i cluster devono raggruppare un numero di osservazioni quanto il più possibile omogeneo rispetto agli altri cluster. Avere un cluster con un numero di elementi di vari ordini di grandezza rispetto ad un altro cluster non sempre può portare a buoni risultati (in termini di devianza).
- Devianza mantenuta: a seguito della PCA parte della devianza nei dati viene persa. Dato che il clustering viene effettuato sulle *Componenti Principali* allora esso produce un'ulteriore perdita di devianza nel risultato finale.

#### Devianza Persa

Per effettuare il calcolo della devianza totale persa bisogna prima calcolare la devianza intra-cluster (la somma delle devianze per ogni cluster) e sulla base di questa si può calcolare la quantità richiesta.

Matematicamente, definita  $DEV_{PCA-PERSA}$  la devianza persa (in termini percentuali) a causa della PCA, viceversa  $DEV_{PCA-MANTENUTA}$  la devianza mantenuta dalla PCA, e  $DEV_{INTRA}$  la devianza intra-cluster, la devianza totale persa percentuale vale:

$$DEV_{PCA-LOST} + DEV_{INTRA} \times DEV_{PCA-MANTENUTA}$$

Essa può essere calcolata in MATLAB passando ad uno script i cluster, le componenti principali e il dataset iniziale.

Per dare valore ai fattori sopra citati il clustering viene effettuato scegliendo un numero di cluster che varia da 6 a 16 (per ogni gruppo di componenti principali) su cui poi viene calcolata la devianza persa.

#### 1.3.1 4 Componenti Principali

Utilizzando le prime quattro PC si ha:



Figura 1.8: Numero di cluster e dimensione per diversi valori

La devianza persa durante la clusterizzazione varia in relazione al numero di cluster scelti. Si possono racchiudere le informazioni in un'unica tabella:

6 Cluster	8 Cluster	10 Cluster	12 Cluster	14 Cluster	16 Cluster
26%	17%	16%	15%	14.5%	14%

### 1.3.2 5 Componenti Principali

Utilizzando le prime quattro PC si ha:

<b>▲ Medie dei cluster</b> <table> <tr> <th>Cluster</th><th>Conteggio</th><th>Principale 1</th><th>Principale 2</th><th>Principale 3</th><th>Principale 4</th><th>Principale 5</th></tr> <tr><td>1</td><td>82</td><td>-10.920</td><td>8.451</td><td>-2.554</td><td>-0.241</td><td>0.200</td></tr> <tr><td>2</td><td>7</td><td>0.178</td><td>12.283</td><td>20.284</td><td>4.779</td><td>-8.536</td></tr> <tr><td>3</td><td>1</td><td>4.824</td><td>19.063</td><td>52.299</td><td>-20.679</td><td>29.023</td></tr> <tr><td>4</td><td>1013</td><td>-1.560</td><td>-0.981</td><td>0.393</td><td>0.650</td><td>0.328</td></tr> <tr><td>5</td><td>786</td><td>-1.282</td><td>-0.852</td><td>0.100</td><td>-0.551</td><td>-0.265</td></tr> <tr><td>6</td><td>227</td><td>1.973</td><td>0.234</td><td>-0.467</td><td>-1.486</td><td>-0.707</td></tr> <tr><td>7</td><td>618</td><td>2.158</td><td>0.220</td><td>-0.181</td><td>-0.110</td><td>-0.252</td></tr> <tr><td>8</td><td>264</td><td>6.427</td><td>2.564</td><td>-0.923</td><td>0.711</td><td>29.023</td></tr> </table>							Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5	1	82	-10.920	8.451	-2.554	-0.241	0.200	2	7	0.178	12.283	20.284	4.779	-8.536	3	1	4.824	19.063	52.299	-20.679	29.023	4	1013	-1.560	-0.981	0.393	0.650	0.328	5	786	-1.282	-0.852	0.100	-0.551	-0.265	6	227	1.973	0.234	-0.467	-1.486	-0.707	7	618	2.158	0.220	-0.181	-0.110	-0.252	8	264	6.427	2.564	-0.923	0.711	29.023																																																								
Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5																																																																																																																							
1	82	-10.920	8.451	-2.554	-0.241	0.200																																																																																																																							
2	7	0.178	12.283	20.284	4.779	-8.536																																																																																																																							
3	1	4.824	19.063	52.299	-20.679	29.023																																																																																																																							
4	1013	-1.560	-0.981	0.393	0.650	0.328																																																																																																																							
5	786	-1.282	-0.852	0.100	-0.551	-0.265																																																																																																																							
6	227	1.973	0.234	-0.467	-1.486	-0.707																																																																																																																							
7	618	2.158	0.220	-0.181	-0.110	-0.252																																																																																																																							
8	264	6.427	2.564	-0.923	0.711	29.023																																																																																																																							
<b>▲ Medie dei cluster</b> <table> <tr> <th>Cluster</th><th>Conteggio</th><th>Principale 1</th><th>Principale 2</th><th>Principale 3</th><th>Principale 4</th><th>Principale 5</th></tr> <tr><td>1</td><td>82</td><td>-10.920</td><td>8.451</td><td>-2.554</td><td>-0.241</td><td>0.200</td></tr> <tr><td>2</td><td>6</td><td>-0.740</td><td>12.018</td><td>17.492</td><td>4.744</td><td>-7.536</td></tr> <tr><td>3</td><td>1</td><td>5.681</td><td>13.876</td><td>37.033</td><td>4.986</td><td>-14.538</td></tr> <tr><td>4</td><td>1</td><td>4.824</td><td>19.063</td><td>52.299</td><td>-20.679</td><td>29.023</td></tr> <tr><td>5</td><td>4</td><td>-1.305</td><td>3.426</td><td>7.953</td><td>1.779</td><td>-3.410</td></tr> <tr><td>6</td><td>70</td><td>-1.601</td><td>-1.163</td><td>0.607</td><td>1.687</td><td>0.833</td></tr> <tr><td>7</td><td>939</td><td>-1.558</td><td>-0.987</td><td>0.345</td><td>0.567</td><td>0.287</td></tr> <tr><td>8</td><td>786</td><td>-1.282</td><td>-0.852</td><td>0.100</td><td>-0.551</td><td>-0.265</td></tr> <tr><td>9</td><td>39</td><td>-0.604</td><td>0.615</td><td>0.150</td><td>-1.695</td><td>-0.868</td></tr> <tr><td>10</td><td>188</td><td>2.508</td><td>0.410</td><td>-0.532</td><td>-1.442</td><td>0.769</td></tr> <tr><td>11</td><td>501</td><td>2.229</td><td>0.269</td><td>-0.222</td><td>-0.316</td><td>0.353</td></tr> <tr><td>12</td><td>117</td><td>1.852</td><td>0.009622</td><td>-0.00249</td><td>0.769</td><td>0.353</td></tr> <tr><td>13</td><td>77</td><td>6.308</td><td>2.508</td><td>-0.539</td><td>1.549</td><td>-0.252</td></tr> <tr><td>14</td><td>187</td><td>6.476</td><td>2.550</td><td>-1.081</td><td>0.366</td><td>29.023</td></tr> </table>							Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5	1	82	-10.920	8.451	-2.554	-0.241	0.200	2	6	-0.740	12.018	17.492	4.744	-7.536	3	1	5.681	13.876	37.033	4.986	-14.538	4	1	4.824	19.063	52.299	-20.679	29.023	5	4	-1.305	3.426	7.953	1.779	-3.410	6	70	-1.601	-1.163	0.607	1.687	0.833	7	939	-1.558	-0.987	0.345	0.567	0.287	8	786	-1.282	-0.852	0.100	-0.551	-0.265	9	39	-0.604	0.615	0.150	-1.695	-0.868	10	188	2.508	0.410	-0.532	-1.442	0.769	11	501	2.229	0.269	-0.222	-0.316	0.353	12	117	1.852	0.009622	-0.00249	0.769	0.353	13	77	6.308	2.508	-0.539	1.549	-0.252	14	187	6.476	2.550	-1.081	0.366	29.023														
Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5																																																																																																																							
1	82	-10.920	8.451	-2.554	-0.241	0.200																																																																																																																							
2	6	-0.740	12.018	17.492	4.744	-7.536																																																																																																																							
3	1	5.681	13.876	37.033	4.986	-14.538																																																																																																																							
4	1	4.824	19.063	52.299	-20.679	29.023																																																																																																																							
5	4	-1.305	3.426	7.953	1.779	-3.410																																																																																																																							
6	70	-1.601	-1.163	0.607	1.687	0.833																																																																																																																							
7	939	-1.558	-0.987	0.345	0.567	0.287																																																																																																																							
8	786	-1.282	-0.852	0.100	-0.551	-0.265																																																																																																																							
9	39	-0.604	0.615	0.150	-1.695	-0.868																																																																																																																							
10	188	2.508	0.410	-0.532	-1.442	0.769																																																																																																																							
11	501	2.229	0.269	-0.222	-0.316	0.353																																																																																																																							
12	117	1.852	0.009622	-0.00249	0.769	0.353																																																																																																																							
13	77	6.308	2.508	-0.539	1.549	-0.252																																																																																																																							
14	187	6.476	2.550	-1.081	0.366	29.023																																																																																																																							
<b>▲ Medie dei cluster</b> <table> <tr> <th>Cluster</th><th>Conteggio</th><th>Principale 1</th><th>Principale 2</th><th>Principale 3</th><th>Principale 4</th><th>Principale 5</th></tr> <tr><td>1</td><td>82</td><td>-10.920</td><td>8.451</td><td>-2.554</td><td>-0.241</td><td>0.200</td></tr> <tr><td>2</td><td>6</td><td>-0.740</td><td>12.018</td><td>17.492</td><td>4.744</td><td>-7.536</td></tr> <tr><td>3</td><td>1</td><td>5.681</td><td>13.876</td><td>37.033</td><td>4.986</td><td>-14.538</td></tr> <tr><td>4</td><td>1</td><td>4.824</td><td>19.063</td><td>52.299</td><td>-20.679</td><td>29.023</td></tr> <tr><td>5</td><td>4</td><td>-1.305</td><td>3.426</td><td>7.953</td><td>1.779</td><td>-3.410</td></tr> <tr><td>6</td><td>70</td><td>-1.601</td><td>-1.163</td><td>0.607</td><td>1.687</td><td>0.833</td></tr> <tr><td>7</td><td>939</td><td>-1.558</td><td>-0.987</td><td>0.345</td><td>0.567</td><td>0.287</td></tr> <tr><td>8</td><td>786</td><td>-1.282</td><td>-0.852</td><td>0.100</td><td>-0.551</td><td>-0.265</td></tr> <tr><td>9</td><td>39</td><td>-0.604</td><td>0.615</td><td>0.150</td><td>-1.695</td><td>-0.868</td></tr> <tr><td>10</td><td>188</td><td>2.508</td><td>0.410</td><td>-0.532</td><td>-1.442</td><td>0.769</td></tr> <tr><td>11</td><td>501</td><td>2.229</td><td>0.269</td><td>-0.222</td><td>-0.316</td><td>0.353</td></tr> <tr><td>12</td><td>117</td><td>1.852</td><td>0.009622</td><td>-0.00249</td><td>0.769</td><td>0.353</td></tr> <tr><td>13</td><td>77</td><td>6.308</td><td>2.508</td><td>-0.539</td><td>1.549</td><td>-0.252</td></tr> <tr><td>14</td><td>187</td><td>6.476</td><td>2.550</td><td>-1.081</td><td>0.366</td><td>29.023</td></tr> </table>							Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5	1	82	-10.920	8.451	-2.554	-0.241	0.200	2	6	-0.740	12.018	17.492	4.744	-7.536	3	1	5.681	13.876	37.033	4.986	-14.538	4	1	4.824	19.063	52.299	-20.679	29.023	5	4	-1.305	3.426	7.953	1.779	-3.410	6	70	-1.601	-1.163	0.607	1.687	0.833	7	939	-1.558	-0.987	0.345	0.567	0.287	8	786	-1.282	-0.852	0.100	-0.551	-0.265	9	39	-0.604	0.615	0.150	-1.695	-0.868	10	188	2.508	0.410	-0.532	-1.442	0.769	11	501	2.229	0.269	-0.222	-0.316	0.353	12	117	1.852	0.009622	-0.00249	0.769	0.353	13	77	6.308	2.508	-0.539	1.549	-0.252	14	187	6.476	2.550	-1.081	0.366	29.023														
Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5																																																																																																																							
1	82	-10.920	8.451	-2.554	-0.241	0.200																																																																																																																							
2	6	-0.740	12.018	17.492	4.744	-7.536																																																																																																																							
3	1	5.681	13.876	37.033	4.986	-14.538																																																																																																																							
4	1	4.824	19.063	52.299	-20.679	29.023																																																																																																																							
5	4	-1.305	3.426	7.953	1.779	-3.410																																																																																																																							
6	70	-1.601	-1.163	0.607	1.687	0.833																																																																																																																							
7	939	-1.558	-0.987	0.345	0.567	0.287																																																																																																																							
8	786	-1.282	-0.852	0.100	-0.551	-0.265																																																																																																																							
9	39	-0.604	0.615	0.150	-1.695	-0.868																																																																																																																							
10	188	2.508	0.410	-0.532	-1.442	0.769																																																																																																																							
11	501	2.229	0.269	-0.222	-0.316	0.353																																																																																																																							
12	117	1.852	0.009622	-0.00249	0.769	0.353																																																																																																																							
13	77	6.308	2.508	-0.539	1.549	-0.252																																																																																																																							
14	187	6.476	2.550	-1.081	0.366	29.023																																																																																																																							
<b>▲ Medie dei cluster</b> <table> <tr> <th>Cluster</th><th>Conteggio</th><th>Principale 1</th><th>Principale 2</th><th>Principale 3</th><th>Principale 4</th><th>Principale 5</th></tr> <tr><td>1</td><td>36</td><td>-10.872</td><td>8.544</td><td>-2.775</td><td>-0.955</td><td>-0.988</td></tr> <tr><td>2</td><td>46</td><td>-10.958</td><td>8.379</td><td>-2.381</td><td>0.318</td><td>0.433</td></tr> <tr><td>3</td><td>6</td><td>-0.740</td><td>12.018</td><td>17.492</td><td>4.744</td><td>-7.536</td></tr> <tr><td>4</td><td>1</td><td>5.681</td><td>13.876</td><td>37.033</td><td>4.986</td><td>-14.538</td></tr> <tr><td>5</td><td>1</td><td>4.824</td><td>19.063</td><td>52.299</td><td>-20.679</td><td>29.023</td></tr> <tr><td>6</td><td>4</td><td>-1.305</td><td>3.426</td><td>7.953</td><td>1.779</td><td>-3.410</td></tr> <tr><td>7</td><td>70</td><td>-1.601</td><td>-1.163</td><td>0.607</td><td>1.687</td><td>0.833</td></tr> <tr><td>8</td><td>939</td><td>-1.558</td><td>-0.987</td><td>0.345</td><td>0.567</td><td>0.287</td></tr> <tr><td>9</td><td>786</td><td>-1.282</td><td>-0.852</td><td>0.100</td><td>-0.551</td><td>-0.265</td></tr> <tr><td>10</td><td>39</td><td>-0.604</td><td>0.615</td><td>0.150</td><td>-1.695</td><td>-0.868</td></tr> <tr><td>11</td><td>188</td><td>2.508</td><td>0.410</td><td>-0.532</td><td>-1.442</td><td>0.769</td></tr> <tr><td>12</td><td>501</td><td>2.229</td><td>0.269</td><td>-0.222</td><td>-0.316</td><td>0.353</td></tr> <tr><td>13</td><td>117</td><td>1.852</td><td>0.009622</td><td>-0.00249</td><td>0.769</td><td>0.353</td></tr> <tr><td>14</td><td>202</td><td>2.478</td><td>0.379</td><td>-0.544</td><td>-1.402</td><td>-0.675</td></tr> <tr><td>15</td><td>487</td><td>2.233</td><td>0.278</td><td>-0.209</td><td>-0.300</td><td>-0.252</td></tr> <tr><td>16</td><td>1</td><td>4.824</td><td>19.063</td><td>52.299</td><td>-20.679</td><td>29.023</td></tr> </table>							Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5	1	36	-10.872	8.544	-2.775	-0.955	-0.988	2	46	-10.958	8.379	-2.381	0.318	0.433	3	6	-0.740	12.018	17.492	4.744	-7.536	4	1	5.681	13.876	37.033	4.986	-14.538	5	1	4.824	19.063	52.299	-20.679	29.023	6	4	-1.305	3.426	7.953	1.779	-3.410	7	70	-1.601	-1.163	0.607	1.687	0.833	8	939	-1.558	-0.987	0.345	0.567	0.287	9	786	-1.282	-0.852	0.100	-0.551	-0.265	10	39	-0.604	0.615	0.150	-1.695	-0.868	11	188	2.508	0.410	-0.532	-1.442	0.769	12	501	2.229	0.269	-0.222	-0.316	0.353	13	117	1.852	0.009622	-0.00249	0.769	0.353	14	202	2.478	0.379	-0.544	-1.402	-0.675	15	487	2.233	0.278	-0.209	-0.300	-0.252	16	1	4.824	19.063	52.299	-20.679	29.023
Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	Principale 5																																																																																																																							
1	36	-10.872	8.544	-2.775	-0.955	-0.988																																																																																																																							
2	46	-10.958	8.379	-2.381	0.318	0.433																																																																																																																							
3	6	-0.740	12.018	17.492	4.744	-7.536																																																																																																																							
4	1	5.681	13.876	37.033	4.986	-14.538																																																																																																																							
5	1	4.824	19.063	52.299	-20.679	29.023																																																																																																																							
6	4	-1.305	3.426	7.953	1.779	-3.410																																																																																																																							
7	70	-1.601	-1.163	0.607	1.687	0.833																																																																																																																							
8	939	-1.558	-0.987	0.345	0.567	0.287																																																																																																																							
9	786	-1.282	-0.852	0.100	-0.551	-0.265																																																																																																																							
10	39	-0.604	0.615	0.150	-1.695	-0.868																																																																																																																							
11	188	2.508	0.410	-0.532	-1.442	0.769																																																																																																																							
12	501	2.229	0.269	-0.222	-0.316	0.353																																																																																																																							
13	117	1.852	0.009622	-0.00249	0.769	0.353																																																																																																																							
14	202	2.478	0.379	-0.544	-1.402	-0.675																																																																																																																							
15	487	2.233	0.278	-0.209	-0.300	-0.252																																																																																																																							
16	1	4.824	19.063	52.299	-20.679	29.023																																																																																																																							

Figura 1.9: Numero di cluster e dimensione per diversi valori

La devianza persa durante la clusterizzazione varia in relazione al numero di cluster scelti. Si possono racchiudere le informazioni in un'unica tabella:

6 Cluster	8 Cluster	10 Cluster	12 Cluster	14 Cluster	16 Cluster
25.5%	16%	15%	12%	11%	10%

### 1.3.3 6 Componenti Principali

Utilizzando le prime quattro PC si ha:

Riepilogo cluster						
Medie dei cluster						
Cluster	Conteggio	Principale1	Principale2	Principale3	Principale4	Principale5
1	82	-10.920	8.451	-2.554	-0.241	0.200
2	7	0.178	12.283	20.284	4.779	-8.536
3	863	2.980	0.7229	-0.369	-0.444	-0.2988
4	116	6.040	2.283	-0.919	1.019	1.102
5	1929	-1.241	-0.744	0.238	0.141	0.073
6	1	4.824	19.083	52.299	-20.679	29.023

Riepilogo cluster						
Medie dei cluster						
Cluster	Conteggio	Principale1	Principale2	Principale3	Principale4	Principale5
1	82	-10.920	8.451	-2.554	-0.241	0.200
2	7	0.178	12.283	20.284	4.779	-8.536
3	132	6.518	2.801	-0.882	0.505	0.121
4	711	2.239	0.279	-0.209	-0.466	-0.380
5	116	6.040	2.283	-0.919	1.019	1.102
6	1126	-1.207	-0.884	0.325	0.058	0.338
7	803	-1.290	-0.847	0.082	-0.584	-0.279
8	1	4.824	19.083	52.299	-20.679	29.023

Riepilogo cluster						
Medie dei cluster						
Cluster	Conteggio	Principale1	Principale2	Principale3	Principale4	Principale5
1	82	-10.920	8.451	-2.554	-0.241	0.200
2	7	0.178	12.283	20.284	4.779	-8.536
3	152	6.518	2.801	-0.882	0.505	0.121
4	116	6.040	2.283	-0.919	1.019	1.102
5	1929	-1.241	-0.744	0.238	0.141	0.073
6	1	4.824	19.083	52.299	-20.679	29.023

Riepilogo cluster						
Medie dei cluster						
Cluster	Conteggio	Principale1	Principale2	Principale3	Principale4	Principale5
1	82	-10.920	8.451	-2.554	-0.241	0.200
2	6	-0.740	12.018	17.482	4.744	-7.536
3	1	5.681	13.876	37.033	4.986	-14.538
4	15	6.154	4.666	2.755	1.375	-0.085
5	137	6.558	2.597	-1.070	0.409	0.364
6	497	2.190	0.258	-0.203	-0.310	-0.258
7	214	2.354	0.326	-0.533	-1.427	-0.695
8	116	6.040	2.283	-0.919	1.019	1.102
9	639	-1.718	-0.946	0.329	0.039	0.363
10	487	-0.536	-0.626	0.330	0.082	0.281
11	803	-1.290	-0.847	0.082	-0.584	-0.279
12	1	4.824	19.083	52.299	-20.679	29.023

Riepilogo cluster						
Medie dei cluster						
Cluster	Conteggio	Principale1	Principale2	Principale3	Principale4	Principale5
1	82	-10.920	8.451	-2.554	-0.241	0.200
2	6	-0.740	12.018	17.482	4.744	-7.536
3	1	5.681	13.876	37.033	4.986	-14.538
4	15	6.154	4.666	2.755	1.375	-0.085
5	137	6.558	2.597	-1.070	0.409	0.364
6	497	2.190	0.258	-0.203	-0.310	-0.258
7	214	2.354	0.326	-0.533	-1.427	-0.695
8	116	6.040	2.283	-0.919	1.019	1.102
9	639	-1.718	-0.946	0.329	0.039	0.363
10	487	-0.536	-0.626	0.330	0.082	0.281
11	803	-1.290	-0.847	0.082	-0.584	-0.279
12	1	4.824	19.083	52.299	-20.679	29.023

Riepilogo cluster						
Medie dei cluster						
Cluster	Conteggio	Principale1	Principale2	Principale3	Principale4	Principale5
1	82	-10.920	8.451	-2.554	-0.241	0.200
2	6	-0.740	12.018	17.482	4.744	-7.536
3	1	5.681	13.876	37.033	4.986	-14.538
4	15	6.154	4.666	2.755	1.375	-0.085
5	137	6.558	2.597	-1.070	0.409	0.364
6	497	2.190	0.258	-0.203	-0.310	-0.258
7	214	2.354	0.326	-0.533	-1.427	-0.695
8	116	6.040	2.283	-0.919	1.019	1.102
9	639	-1.718	-0.946	0.329	0.039	0.363
10	487	-0.536	-0.626	0.330	0.082	0.281
11	803	-1.290	-0.847	0.082	-0.584	-0.279
12	1	4.824	19.083	52.299	-20.679	29.023

Figura 1.10: Numero di cluster e dimensione per diversi valori

6 Cluster	8 Cluster	10 Cluster	12 Cluster	14 Cluster	16 Cluster
22%	15%	13%	12%	11%	9%

### 1.3.4 Interpretazione

All'inizio dell'analisi sono state effettuate delle ipotesi che hanno trovato riscontro nella procedura di clustering:

1. La fase iniziale del sistema (descritto dalle prime righe) trova riscontro con il cluster numero uno qualunque siano le componenti principali e qualunque sia il numero di cluster scelto. Questo quindi prova l'ipotesi definita inizialmente
2. L'ultimo cluster contiene sempre un elemento singolo. Questo accade a causa del fatto che è stato identificato un picco nella fase iniziale delle misure. Esso inoltre è stato definito grazie all'outlier nel parametro Slab che non è stato eliminato, di conseguenza il picco viene racchiuso in un unico cluster in tutte le situazioni.

In conclusione si può costruire una tabella che racchiude le informazioni riguardanti le PCA e la clusterizzazione in termini di percentuale di devianza persa.

	6 Cluster	8 Cluster	10 Cluster	12 Cluster	14 Cluster	16 Cluster
4 PC	26%	17%	16%	15%	14.5%	14%
5 PC	25.5%	16%	15%	12%	11%	10%
6 PC	22%	15%	13%	12%	11%	9%

## 1.4 Workload Sintetico

Come già deciso nel paragrafo precedente, sono state scelte 5 PC. A tal proposito per non perdere troppa varianza, e non sfociare in un numero di cluster molto elevato, si è scelto di considerare 10 cluster. In tal caso la perdita di devianza con PCA e Cluster è di circa del 15%.

In conclusione dopo aver calcolato i centroidi con uno script MATLAB, il workload sintetico risulta essere:

VmPeak'	VmSize'	VmHWM'	VmRSS'	VmPTE'	Threads'	MemFree'	Buffers'	Cached'	Inactive'	Dirty'	Mapped'	Slab'	PageTables'	Commit ted_AS'	NumOf Alloc...	proc- fd'	avgThr ough...	avgEl aps...	Cluster
152272	150216	25228	25208	160	58	5604588	33440	334672	141440	304136	148	0	77472	23772	27852	7192	294280	1530	1
153484	151436	26612	26468	160	20	5044360	83684	732680	365572	576940	115360	0	126116	28324	91172	7984	347336	2040	2
170344	170340	31144	31140	200	52	4655664	89480	1090572	541220	773332	324400	1760	134500	29956	106788	8028	365264	2040	10
170344	170340	31144	31140	200	43	4471580	94884	1259568	597900	891332	459632	0	134504	29956	116216	8024	365692	1530	3
172392	170340	32696	31516	200	35	4608100	158180	1117364	424808	934596	176	0	83860	23740	110212	7208	318956	1020	7
172392	170340	32696	31340	200	35	4605508	160628	1117388	427232	934468	152	0	83684	23740	110320	7208	318508	1530	6
192256	184832	43804	37584	252	87	4557264	198720	1118136	488652	918124	56	0	89880	23848	113836	7260	329604	510	8
192384	185460	43804	37396	256	97	4555172	201052	1118212	494300	914696	36	0	89736	23848	113556	7264	330928	1020	9
215352	210232	61784	59084	316	116	4527992	204160	1118424	526896	908536	148	0	112848	23996	113028	7772	356604	510	4
215352	206136	61784	53204	312	9	4539096	204400	1118460	520240	908140	148	0	105548	23848	110300	7320	348476	510	5

Figura 1.11: *Workload sintetico*