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 ${\rm A.A~2020/2021}$

Indice

L	Woı	rkload	Characterization	1
	1.1	Filtrag	ggio	1
		1.1.1	Colonne Identiche	1
		1.1.2	Outlier	2
	1.2	PCA		6
	1.3	Cluste	ring	7
		1.3.1	4 Componenti Principali	8
		1.3.2	5 Componenti Principali	8
		1.3.3	6 Componenti Principali	9
		1.3.4	Interpretazione	9
	1.4	Workl	oad Sintetico	9

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 è possibile rendersi conto della presenza di ben 4 colonne costanti:

- Active
- AnonPages
- AvbLatency
- Error

Tali colonne in quanto costanti 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:

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

1.1.2 Outlier

Gli outliers sono valori anomali all'interno dell'insieme di osservazioni, in altre parole sono valori che si discostano notevolmente dagli altri valori dell'insieme. Essendo valori anomali la loro frequenza di occorrenza è bassa rispetto agli altri valori e ciò li porta ad essere identificati all'esterno del range interquartile. In alcuni casi gli outlier possono essere eliminati ma ciò è possibile solo a monte di una analisi accurata. Tali valori infatti influiscono sulle analisi statistiche in modo considerevole e non sempre rappresentano situazioni trascurabili per l'analisi da svolgere . Osservando attraverso box plot e grafici di distribuzione l'andamento dei seguenti parametri :

- 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);

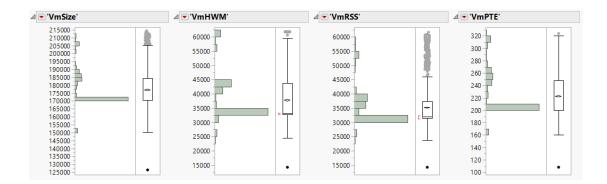


Figura 1.2: Grafici di distribuzione di VmSize, VmHWM, VmRSS, VmPTE

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. Tuttavia è 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.

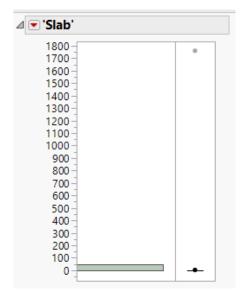


Figura 1.3: Grafico di distribuzione di Slab

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.

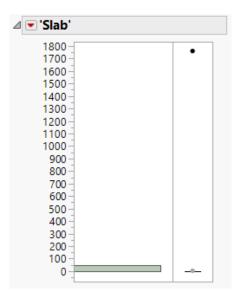


Figura 1.4: Grafico di distribuzione di Slab

Rispetto al precedente, tale outlier richiede un' analisi più approfondita visto che influenza significativamente l'andamento di parametri quali Mapped e Page Tables. 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.

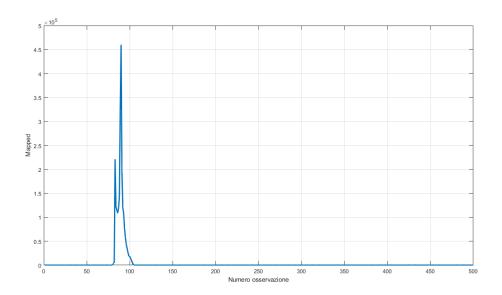


Figura 1.5: Grafico tra numero di osservazione e valore assunto dal parametro Mapped

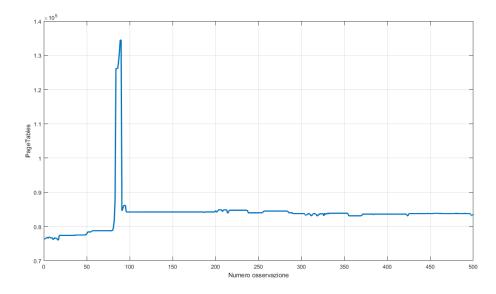


Figura 1.6: Grafico tra numero di osservazione e valore assunto dal parametro Page Tables

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 non trascurarlo.

In conclusione sono stati eliminati dal dataset solo 2 outlier che corrispondono a 38 osservazioni. 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 che non esprimevano varianza e alcune righe rappresentanti outlier trascurabili.

Su 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:

Numero	Autovalore	Percentuale	20	40	60	80	Percentuale cumulativa
1	9,8395	51,787			K.		51,787
2	3,6873	19,407					71,194
3	2,4576	12,935				1	84,129
4	0,7782	4,096				\	88,225
5	0,7075	3,724				\	91,949
6	0,6258	3,293					95,242
7	0,3409	1,794					97,036
8	0,2658	1,399					98,435
9	0,1612	0,848					99,283
10	0,0904	0,476					99,759
11	0,0178	0,094					99,853
12	0,0114	0,060					99,913
13	0,0078	0,041					99,954
14	0,0058	0,030					99,985
15	0,0014	0,007					99,992
16	0,0009	0,005					99,997
17	0,0006	0,003					100,000
18	0,0000	0,000					100,000
19	0,0000	0,000					100,000

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.

In questo caso la scelta è ricaduta sul prendere 5 componenti principali poiché rappresentano il 92% della devianza totale. Esso rappresenta un valore abbastanza elevato, ma è stato scelto per mantenersi in una regione di tolleranza durante la clusterizzazione.

Anche se il workload sintentico verrà costruito considerando 5 componenti principali, in seguito sono riportati i risultati di PCA e clusterizzazione anche nel caso in cui fosse stato scelto un numero diverso di componenti principali, ovvero:

• Prendere 4 PC

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

• Prendere 5 PC

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

• Prendere 6 PC

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

Per ognuno di questi 3 insiemi di Principal Components è stata effettuata la procedura di clustering .

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 numero di righe del dataset pari al numero di cluster scelti durante l'analisi.

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 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:

												∠ Riepilog	o cluster				
						⊿ Riepilog	o cluster						dei cluste				
						4 Medie	dei cluste					Cluster		Principale 1	Principale 2		
△ Riepilog	o cluster						Conteggio		Principale 2	Principale 3	Principale 4	1	82	-10,920 -0.740	8,451 12,018	-2,554 17,492	-0,241 4,744
⊿ Medie	dei cluster					ciustoi	82	-10.920		-2.554	-0,241	3	6	-0,740 5.681	12,018	17,492 37.033	4,744
Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	2	7	0.178		20,284	4,779	3	1	4.824	13,876	52,299	-20.679
1	82	-10.920		-2.554	-0.241	3	1	4.824		52.299	-20.679	5	1013		-0.981	0.393	-20,679
	7	0,178			4,779	4	1013	-1,560		0,393	0,650	6	786		-0,961	0,393	-0,551
3	1	4.824			-20,679	5	786	-1,282		0,100	-0,551	7	227	1,202	0.234	-0.467	-1,486
J	1013	-1.560		0.393	0.650	6		1,973		-0.467	-1,486	,	501	2,229	0,234	-0,467	-0,316
5		0.474			-0.514	7	618	2,158		-0.181	-0.110	9	117	1,852	0.009622	-0,00249	0,769
6		6.427			0.711	,		6,427	2,564	-0,181	0.711	10	264	6.427	2,564	-0,00249	0,769
												⊿ Riepilogo					
						⊿ Riepilog	o cluster					⊿ Medie d	ei cluster				
							dei cluste					Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4
Riepilogo	cluster											1	36	-10,872	8,544	-2,775	-0,955
						Cluster				Principale 3		2	46	-10,958	8,379	-2,381	0,318
△ Medie d						1	82	-10,920		-2,554	-0,241	3	6	-0,740	12,018	17,492	4,744
Cluster	Conteggio	Principale 1	Principale 2	Principale 3	Principale 4	2	6	-0,740		17,492	4,744	4	1	5,681	13,876	37,033	4,986
1	82	-10,920	8,451	-2,554	-0,241	3	1	5,681	13,876	37,033	4,986	5	1	4,824	19,063	52,299	-20,679
2	6	-0,740	12,018	17,492	4,744	4	1	4,824 -1.305		52,299 7.953	-20,679 1,779	6	4	-1,305	3,426	7,953	1,779
3	1	5,681	13,876	37,033	4,986	3	4 70		-1,163	0.607	1,687	/	70	-1,601	-1,163	0,607	1,687
4	1	4,824	19,063	52,299	-20,679	0						8	939	-1,558	-0,987	0,345	0,567
5	74	-1,585	-0,915	1,004	1,692	,	939	-1,558		0,345	0,567	9	786	-1,282	-0,852	0,100	-0,551
6	939	-1,558	-0,987	0,345	0,567	8	786	-1,282		0,100	-0,551	10	39	-0,604	-0,615	-0,150	-1,695
7	786	-1,282	-0,852	0,100	-0,551	9				-0,150	-1,695	11	188	2,508	0,410	-0,532	-1,442
8	227	1,973	0,234	-0,467	-1,486	10		2,508		-0,532	-1,442	12	501	2,229	0,269	-0,222	-0,316
9	501	2,229	0,269	-0,222	-0,316	11		2,229		-0,222	-0,316	13	117	1,852	0,009622	-0,00249	0,769
10	117	1,852	0,009622	-0,00249	0,769	12		1,852	0,009622	-0,00249	0,769	14	66	5,882	2,178	-0,773	1,603
11	77	6,308	2,598	-0,539	1,549	13		6,308		-0,539	1,549	15	- 11	8,866	5,116	0,865	1,229
12	187	6,476	2,550	-1,081	0.366	14	187	6,476	2,550	-1,081	0,366	16	187	6,476	2,550	-1,081	0,366

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:

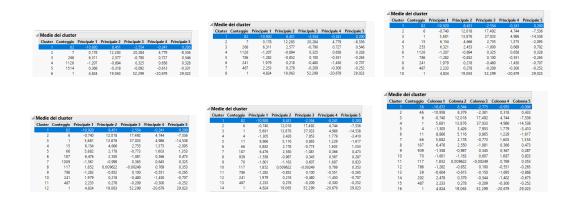


Figura 1.9: Numero di cluster e dimensione per diversi valori

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:

																△ Riepilog	o cluster						
								⊿ Riepiloc	a aluatas							⊿ Medie	dei cluster	r					
																Cluster	Conteggio			Principale3			
Riepilog	go cluster								dei cluste							1	82	-10,920	8,451	-2,554	-0,241		
Modio	dei cluste							Cluster	Conteggio						Principale6	2	7	0,178	12,283	20,284	4,779		-0,74
			Principale2	Principale3		Principale5	Principale6	1	82	-10,920	8,451	-2,554	-0,241	0,200	0,291	3	152	6,518	2,801	-0,692			0,63
Citister		-10.920	8.4514	-2,554	-0.241		0,291	2		0,178	12,283	20,284	4,779	-8,536	-0,748	4	497	2,190	0,258	-0,203	-0,310		0,54
	1 82	0.178	12.2830	20.284	4.775		-0.748	3	152	6,518	2,801	-0,692	0,505	0,121	0,639	3		2,354 6,040	0,326 2,283	-0,523 -0,919	-1,427 1,019		-0,03 -2.36
-	863	2,993	0.7229	-0.369	-0.444		0.419	4	711	2,239	0,279	-0,299	-0,646	-0,390	0,373		116 639	-1.718	-0.946	0,329			-2,30
3	1 116	6,040	2.2835	-0,309	1,019		-2,367	5	116 1126		2,283 -0.894	-0,919 0.325	1,019	1,102	-2,367 0.252		487	-0.536	-0,946	0,329			0.72
	1929	-1.241	-0.8744	0.228	0.141		-0.057	6	803	-1,207				-0.279		° a	803	-1,290	-0.847	0,092	-0.584		-0.48
	1929	4,824	19.0628	52,299	-20.679		3.044	,		-1,290 4,824	-0,847 19.063	0,092 52,299	-0,584 -20,679	29.023	-0,489 3.044	10	803	4,824	19.063	52.299			3.04
		4,004	15,0000	34,433	-20,012	,	3,044	0		4,024	19,003	32,299	-20,079	29,025	3,044	10		4,004	15,005	22,230	-20,012	20,023	3,04
aniloge	o cluster								dei cluster								Conteggio	Principale1			Principale4	Principale5	
Medie o										Principale1	Principale2 F	rincipale3 Pi	rincipale4 I	Principale5	Principale6	1	82	-10,920	8,451	-2,554	-0,241	0,200	0,291
Cluster	dei cluster								Conteggio						Principale6	1 2		-10,920 -0,740	8,451 12,018	-2,554 17,492	-0,241 4,744	0,200 -7,536	0,291
							rincipale6			Principale 1 -10,920 -0.740	Principale2 F 8,451 12.018	rincipale3 Pr -2,554 17,492	-0,241 4,744	0,200 -7.536	Principale6 0,291 -0.298	1 2 3		-10,920 -0,740 5,681	8,451 12,018 13,876	-2,554 17,492 37,033	-0,241 4,744 4,986	0,200 -7,536 -14,538	0,291 -0,296 -3,445
1		-10,920	8,451	-2,554	-0,241	0,200	0,291		Conteggio	-10,920	8,451	-2,554	-0,241	0,200	0,291	1 2 3 4	82 6 1 15	-10,920 -0,740 5,681 6,154	8,451 12,018 13,876 4,666	-2,554 17,492 37,033 2,755	-0,241 4,744 4,986 1,375	0,200 -7,536 -14,538 -2,095	0,291 -0,296 -3,445 -0,730
1 2		-10,920 -0,740	8,451 12,018	-2,554 17,492	-0,241 4,744	0,200 -7,536	0,291 -0,298		Conteggio	-10,920 -0,740	8,451 12,018	-2,554 17,492	-0,241 4,744	0,200 -7,536	0,291 -0,298	1 2 3 4 5	82 6 1	-10,920 -0,740 5,681 6,154 6,558	8,451 12,018 13,876	-2,554 17,492 37,033 2,755 -1,070	-0,241 4,744 4,986	0,200 -7,536 -14,538 -2,095 0,364	0,291 -0,296 -3,445 -0,730 0,789
1 2 3	Conteggio 82 6	-10,920 -0,740 5,681	8,451 12,018 13,876	-2,554 17,492 37,033	-0,241 4,744 4,986	0,200 -7,536 -14,538	0,291 -0,298 -3,445		Conteggio 82 6 1	-10,920 -0,740 5,681	8,451 12,018 13,876	-2,554 17,492 37,033	-0,241 4,744 4,986	0,200 -7,536 -14,538	0,291 -0,298 -3,445	1 2 3 4 5 6	82 6 1 15 137	-10,920 -0,740 5,681 6,154	8,451 12,018 13,876 4,666 2,597	-2,554 17,492 37,033 2,755	-0,241 4,744 4,995 1,375 0,409	0,200 -7,536 -14,538 -2,095	0,291 -0,296 -3,445 -0,730 0,789
1 2 3 4	82 6 1	-10,920 -0,740 5,681 6,154	8,451 12,018 13,876 4,666	-2,554 17,492 37,033 2,755	-0,241 4,744 4,986 1,375	0,200 -7,536 -14,538 -2,095	0,291 -0,298 -3,445 -0,730		Conteggio 82 6 1 15	-10,920 -0,740 5,681 6,154 6,558 2,190	8,451 12,018 13,876 4,666 2,597 0,258	-2,554 17,492 37,033 2,755 -1,070 -0,203	-0,241 4,744 4,966 1,375	0,200 -7,536 -14,538 -2,095	0,291 -0,298 -3,445 -0,730	1 2 3 4 5 6 7	82 6 1 15 137 150	-10,920 -0,740 5,681 6,154 6,558 2,223	8,451 12,018 13,876 4,666 2,597 0,281	-2,554 17,492 37,033 2,755 -1,070 -0,270	-0,241 4,744 4,985 1,375 0,409 -0,493	0,200 -7,536 -14,538 -2,095 0,364 -0,393	0,291 -0,296 -3,445 -0,730 0,789 1,118 0,301
1 2 3 4 5	82 6 1 15 137	-10,920 -0,740 5,681 6,154 6,558	8,451 12,018 13,876 4,666 2,597	-2,554 17,492 37,033 2,755 -1,070	-0,241 4,744 4,986 1,375 0,409	0,200 -7,536 -14,538 -2,095 0,364	0,291 -0,298 -3,445 -0,730 0,789		82 6 1 15 137	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354	8,451 12,018 13,876 4,666 2,597 0,258 0,326	-2,554 17,492 37,033 2,755 -1,070 -0,208 -0,523	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034	1 2 3 4 5 6 7 8	82 6 1 15 137 150 347	-10,920 -0,740 5,681 6,154 6,558 2,223 2,176	8,451 12,018 13,876 4,666 2,597 0,281 0,249	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174	-0,241 4,744 4,995 1,375 0,409 -0,493 -0,232	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200	0,291 -0,298 -3,445 -0,730 0,789 1,118 0,301 -0,034
1 2 3 4 5	82 6 1 15 137 497	-10,920 -0,740 5,681 6,154 6,558 2,190	8,451 12,018 13,876 4,666 2,997 0,258	-2,554 17,492 37,033 2,755 -1,070 -0,203	-0,241 4,744 4,986 1,375 0,409 -0,310	0,200 -7,536 -14,538 -2,095 0,364 -0,258	0,291 -0,298 -3,445 -0,730 0,789 0,548		82 6 1 15 137 497 214 116	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367	1 2 3 4 5 6 7	82 6 1 15 137 150 347 214	-10,920 -0,740 5,681 6,154 6,558 2,223 2,176 2,354	8,451 12,018 13,876 4,666 2,597 0,281 0,249 0,326	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174 -0,523	-0,241 4,744 4,995 1,375 0,409 -0,493 -0,232 -1,427	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200 -0,695	-0,296 -3,445 -0,730 0,789 1,118
1 2 3 4 5 6 7	82 6 1 15 137 497 214	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354	8,451 12,018 13,876 4,666 2,597 0,258 0,326	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034	Cluster 1 2 3 4 5 6 7 8	82 6 1 15 137 497 214 116 639	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,946	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111	1 2 3 4 5 6 7 8 9 10	82 6 1 15 137 150 347 214 116	-10,920 -0,740 5,681 6,154 6,558 2,223 2,176 2,354 6,040 -1,718 -1,601	8,451 12,018 13,876 4,666 2,597 0,281 0,249 0,326 2,283 -0,946 -1,163	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174 -0,523 -0,919 0,329 0,607	-0,241 4,744 4,986 1,375 0,409 -0,493 -0,232 -1,427 1,019 0,639 1,687	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200 -0,695 1,102 0,363 0,833	0,291 -0,298 -3,445 -0,730 0,789 1,118 0,301 -0,034 -2,367 -0,111 0,703
1 2 3 4 5 6 7	Conteggio 82 6 1 15 137 497 214 116	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367	Cluster 1 2 3 4 5 6 7 8 9 10	Conteggio 82 6 1 15 137 497 214 116 639 188	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718 0,557	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,946 -0,431	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329 0,226	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639 1,108	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363 0,529	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111 0,954	1 2 3 4 5 6 6 7 8 9 10 11 12	82 6 1 15 137 150 347 214 116 639 70 118	-10,920 -0,740 5,681 6,154 6,538 2,223 2,176 2,354 6,040 -1,718 -1,601 1,836	8,451 12,018 13,876 4,666 2,597 0,281 0,249 0,326 2,283 -0,946 -1,163 0,002459	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174 -0,523 -0,919 0,607 1,331e-5	-0,241 4,744 4,986 1,375 0,409 -0,493 -0,232 -1,427 1,019 0,639 1,687 0,764	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200 -0,695 1,102 0,363 0,363 0,349	0,291 -0,298 -3,445 -0,730 0,789 1,118 0,301 -0,034 -2,367 -0,111 0,703 1,103
1 2 3 4 5 6 7 8	82 6 1 15 137 497 214 116 639	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,946	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111	Cluster 1 2 3 3 4 4 5 6 7 7 8 8 9 10 11	Conteggio 82 6 1 15 137 497 214 116 639 188 299	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718 0,557 -1,223	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,446 -0,431 -1,074	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329 0,226 0,380	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639 1,108 0,415	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363 0,529 0,126	0,291 -0,296 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111 0,954 0,585	1 2 3 4 5 6 7 7 8 9 10 11 12 13	82 6 1 15 137 150 347 214 116 639 70 118 299	-10,920 -0,740 5,681 6,154 6,558 2,223 2,176 2,354 6,040 -1,718 -1,601 1,836 -1,223	8,451 12,018 13,876 4,666 2,597 0,281 0,249 0,326 2,283 -0,946 -1,163 0,002459 -1,074	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174 -0,523 -0,919 0,329 0,607 1,331e-5 0,380	-0,241 4,744 4,986 1,375 0,409 -0,493 -0,232 -1,427 1,019 0,639 1,687 0,764 0,415	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200 -0,695 1,102 0,363 0,833 0,349 0,126	0,291 -0,298 -3,445 -0,730 0,789 1,118 0,301 -0,034 -2,367 -0,111 0,703 1,103 0,585
1 2 3 4 5 6 7 8 9	Conteggio 82 6 1 15 137 497 214 116 639 487	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718 -0,536	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,946 -0,826	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329 0,320	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639 0,682	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363 0,281	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111 0,727	Cluster 1 2 3 4 4 5 6 6 7 7 8 9 10 11 12	Conteggio 82 6 1 15 137 497 214 116 639 188 299 400	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718 0,557 -1,223 -1,064	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,946 0,431 -1,074 -0,887	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329 0,226 0,380 0,131	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639 1,108 0,415 -0,645	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363 0,529 0,126 -0,371	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111 0,585 -0,149	1 2 3 4 5 6 7 8 9 10 11 12 13	82 6 1 15 137 150 347 214 116 639 70 118 299 400	-10,920 -0,740 5,681 6,154 6,558 2,223 2,176 2,354 6,040 -1,718 -1,601 1,836 -1,223 -1,064	8,451 12,018 13,676 4,666 2,997 0,281 0,249 0,326 2,283 -0,946 -1,163 0,002459 -1,074 -0,887	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174 -0,523 -0,919 0,329 0,607 1,331e-5 0,380 0,131	-0,241 4,744 4,906 1,375 0,409 -0,493 -0,232 -1,427 1,019 0,639 1,687 0,764 0,415 -0,645	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200 -0,695 1,102 0,363 0,833 0,349 0,126 -0,371	0,291 -0,296 -3,445 -0,730 0,789 1,118 0,301 -0,034 -2,367 -0,111 0,703 1,103 0,585 -0,149
1 2 3 4 5 6 7 8 9 10	82 6 1 15 137 497 214 116 639	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,946	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363	0,291 -0,298 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111	Cluster 1 2 3 3 4 4 5 6 7 7 8 8 9 10 11	Conteggio 82 6 1 15 137 497 214 116 639 188 299	-10,920 -0,740 5,681 6,154 6,558 2,190 2,354 6,040 -1,718 0,557 -1,223	8,451 12,018 13,876 4,666 2,597 0,258 0,326 2,283 -0,446 -0,431 -1,074	-2,554 17,492 37,033 2,755 -1,070 -0,203 -0,523 -0,919 0,329 0,226 0,380	-0,241 4,744 4,986 1,375 0,409 -0,310 -1,427 1,019 0,639 1,108 0,415	0,200 -7,536 -14,538 -2,095 0,364 -0,258 -0,695 1,102 0,363 0,529 0,126	0,291 -0,296 -3,445 -0,730 0,789 0,548 -0,034 -2,367 -0,111 0,954 0,585	1 2 3 4 5 6 7 7 8 9 10 11 12 13	82 6 1 15 137 150 347 214 116 639 70 118 299	-10,920 -0,740 5,681 6,154 6,558 2,223 2,176 2,354 6,040 -1,718 -1,601 1,836 -1,223	8,451 12,018 13,876 4,666 2,597 0,281 0,249 0,326 2,283 -0,946 -1,163 0,002459 -1,074	-2,554 17,492 37,033 2,755 -1,070 -0,270 -0,174 -0,523 -0,919 0,329 0,607 1,331e-5 0,380	-0,241 4,744 4,986 1,375 0,409 -0,493 -0,232 -1,427 1,019 0,639 1,687 0,764 0,415	0,200 -7,536 -14,538 -2,095 0,364 -0,393 -0,200 -0,695 1,102 0,363 0,833 0,349 0,126	0,291 -0,296 -3,445 -0,730 0,785 1,118 0,301 -0,034 -2,367 -0,111 0,703 1,103 0,585

Figura 1.10: Numero di cluster e dimensione per diversi valori

	6 Cluster	8 Cluster	10 Cluster	12 Cluster	14 Cluster	16 Cluster	
ĺ	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 anticipato nel paragrafo precedente, sono state scelte 5 PC e per non perdere troppa varianza ma al tempo stesso 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:

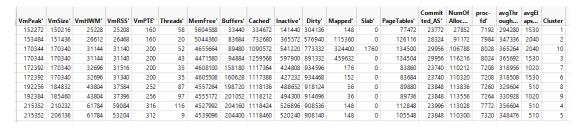


Figura 1.11: Workload sintetico