



Impianti di elaborazione

Anno Accademico
2025/26

Rocco Lo Russo
Agostino D'Amora

Contents

Introduzione	5
I Performance	7
1 Valutazione delle Performance	9
1.1 System Evaluation	9
1.1.1 Passi per la valutazione di un sistema	9
1.1.2 Performance Analysis	10
1.1.3 Errori comuni nella System Evaluation	10

Introduzione

In questo documento verranno raccolti appunti presi a lezione del corso di *Impianti di elaborazione* tenuto nell'anno 2025-26 dai professori Cotroneo e Pietrantuono, con l'aggiunta di alcuni richiami e approfondimenti

Part I

Performance

Chapter 1

Valutazione delle Performance

La valutazione delle performance di un sistema (o system evaluation) è un argomento importante da dover trattare. Negli anni ci sono stati vari problemi ai sistemi che sono stati ideati, poichè o valutati in modo scorretto o progettati male. Lo scopo principale della system evaluation è quello di misurare le prestazioni di un determinato sistema, in modo che sia anche possibile confrontare i parametri con quelli di valutazione di altri sistemi (nella maniera più oggettiva possibile).

1.1 System Evaluation

Per la system evaluation, quindi, è importante impostare e delineare un metodo formale di valutazione. Ciò ci permette di ridurre gli errori legati a particolari operazioni e di poter definire una serie di "passi" da seguire per effettuare una corretta valutazione delle performance. In linea formale, la system evaluation si divide in due principali categorie:

- **Performance Analysis:** Tale valutazione presuppone che il sistema non possa avere alcun tipo di fallimento (failure-free). Il che va a valutare solo le performance legate al suo funzionamento. (Bisogna stare attenti quando si effettua Performance Analysis di non andare a valutare in alcun modo i casi di fallimento)
- **Dependability Analysis:** Tale valutazione ci permette di valutare per quanto tempo il sistema sia in grado di funzionare e quindi anche il caso di problematiche ed errori. (tra le analisi di dependability rientra anche la **reliability**, che definisce un parametro per identificare il tempo medio in cui il sistema fallisce [System Mean Time To Failure]

*Un esempio pratico per capire i concetti di **Performance Analysis** e **Dependability Analysis** è quello di un'auto di formula 1. Nel caso della Performance Analysis vado solo a valutare le specifiche performance (velocità massima, tenuta in curva, aerodinamica), senza tener conto in alcun modo di qualunque tipo di fallimento; mentre nel caso della Dependability Analysis si va a valutare quanto la macchina riesca a resistere in pista (durata delle gomme, tempo effettivo di funzionamento, casi di guasti imprevisti).*

1.1.1 Passi per la valutazione di un sistema

Come introdotto precedentemente, per la valutazione "corretta" (o di buona qualità) di un sistema, è ottimale definire una serie di passi da seguire, in modo da redere il criterio

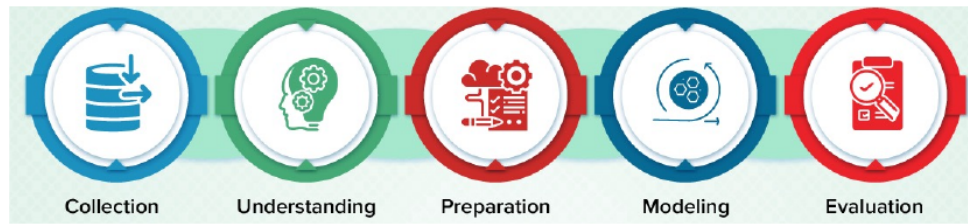


Figure 1.1: Passi da effettuare per la CM e la CP

di valutazione il quanto più formale possibile. I passi per valutare un sistema sono i seguenti:

1. Definire cosa bisogna valutare
2. Ottenere informazioni sul sistema
3. Effettuare le misurazioni
4. Analisi dei risultati
5. Trovare e valutare il corretto feedback da dare sulle considerazioni iniziali

1.1.2 Performance Analysis

Nell'analisi delle performance, quindi, si va a presupporre che il sistema di base funzioni e che quindi bisogna valutare solo il "come" tale sistema funziona (failure-free). L'analisi delle performance può essere suddivisa formalmente in due principali parti:

- **Capacity Management:** Fa in modo che le odierne risorse disponibili diano le migliori performance possibili (quindi si basa solo sulla valutazione del caso presente senza aver fatto alcuna supposizione sul carico futuro)
- **Capacity Planning:** Si assicura che ci sia un'allocazione delle risorse in base ai workload successivi (si va a fare delle previsioni sul possibile carico futuro)

*Nella spiegazione dei termini precedenti si è fatto riferimento ad una parola, ovvero **workload**. Il significato e la definizione vera e formale di workload sarà data nei capitoli successivi, per il momento possiamo definire workload come: richieste effettuate da utenti verso il sistema*

La Capacity Management e la Capacity Planning sono strutturate principalmente in 5 passi che portano alla corretta esecuzione della propria valutazione [1.1]

1.1.3 Errori comuni nella System Evaluation

Quando si fa systema evaluation, solitamente, si possono commettere degli errori, che possono portare poi ad avere uno scorretto parametro di confronto o giudizio del sistema. Gli errori che principalmente vengono fatti sono:

- **Nessun obiettivo:** senza obiettivi chiari non esiste un modello universale; gli obiettivi determinano tecniche, metriche e workload da usare, e non sono mai banali.

- **Obiettivi distorti:** porsi come obiettivo. Dimostrare che il nostro sistema è migliore di un altro porta a una valutazione di parte, in cui gli analisti fanno da giudici invece che da osservatori imparziali.
- **Approccio non sistematico:** condurre la valutazione senza un metodo strutturato (obiettivi → metriche → workload → esperimenti → analisi) porta a risultati incompleti o non riproducibili.
- **Analisi senza capire il problema:** raccogliere dati e produrre grafici senza aver compreso a fondo la natura del problema significa ottenere informazioni non utili alle decisioni.
- **Metriche di performance scorrette:** scegliere metriche che non riflettono gli aspetti importanti del sistema (es. guardare solo il throughput quando è cruciale la latenza) porta a conclusioni fuorvianti.
- **Workload non rappresentativo:** utilizzare un carico artificiale che non riproduce il comportamento reale degli utenti (picchi, mix di richieste, distribuzioni) rende i risultati poco affidabili.
- **Tecnica di valutazione sbagliata:** adottare un metodo inadeguato (analisi analitica, simulazione o misurazioni reali) rispetto agli obiettivi porta a risultati poco significativi o addirittura falsati.

Dati tali errori di valutazione si comprende il motivo a cui è legato il bisogno di definire un path formale per la performance evaluation. Pertanto si va a definire una serie di passi sistematici che ci spiega come poter realizzare la system evaluation senza andare in contro alle problematiche descritte in precedenza. I passi da seguire sono i seguenti:

1. Definire gli obiettivi e descrivere il sistema
2. Elencare i servizi e i risultati attesi
3. Selezionare le metriche
4. Elencare i parametri
5. Selezionare i fattori da studiare
6. Scegliere la tecnica di valutazione
7. Selezionare il workload
8. Progettare gli esperimenti
9. Analizzare e interpretare i dati
10. Presentare i risultati
11. Ripetere il processo

Guardando tali passi si comprende che bisogna effettuare alcune scelte fondamentali. Le scelte che principalmente bisogna effettuare sono legate a: Tecniche di valutazione, Metriche di performance e Performance richieste

Tecniche di valutazione

Per valutare le performance di un sistema si possono utilizzare diverse tecniche che racchiudono una metodologia differente di approccio rispetto al sistema, che ci permette di poter valutare le prestazioni prescindendo dal sistema stesso (o in parte). Le tecniche principali di valutazione sono:

- **Modellazione Analitica:** Si va a ricostruire il sistema mediante un modello matematico. Tale tecnica permette di avere una soluzione in forma chiusa (utilizza formule matematiche senza dover simulare o replicare un sistema). Tali sistemi, però, fanno delle assunzioni sul sistema, che permettono la semplificazione e la modellazione matematica
- **Simulazione:** Tale tecnica cerca di combinare la modellazione analitica del sistema con il mondo reale, cercando di emulare quanto più è possibile il caso reale tramite particolari software. È una buona soluzione, poichè richiede un costo intermedio per essere effettuata; se non fosse per il tempo che bisogna dedicargli per la ricostruzione del sistema
- **Misura:** Si vanno a valutare le performance con misurazioni sul sistema reale. Tale approccio è il più costoso, sia in termini di carico che di costo, ma è il più efficiente poichè si è a contatto con il caso reale effettivo

Per selezionare la tecnica adatta al nostro caso bisogna fare una valutazione completa del sistema cercando quella che è la tecnica più adatta in base al nostro criterio di valutazione richiesto. Talvolta può essere anche possibile utilizzare più tecniche insieme. (Un esempio potrebbero essere i Twin Systems, dove vado ad effettuare modifiche prima in simulazione, e se noto miglioramento delle performance applico le stesse scelte anche al sistema reale andando ad analizzare ulteriormente le performance).

Selezione della metrica

Altro parametro importante da scegliere è la metrica da utilizzare. È importante capire il corretto modo di scegliere una metrica dato che è il concetto su cui si basa il confronto tra vari sistemi. Le metriche possono basarsi su diversi criteri, i principali possono essere classificati come:

- **Metriche per le performance:** Si vanno a valutare dei parametri di valutazione discreti (non probabilistici), dipendenti da: Tempo, Processing Rate, Consumo di risorse ecc.
- **Metriche per la Dependability:** Si va a valutare un sistema in base alla sua "efficienza", vista nel senso di probabilità di diversi eventi (guasti, eccezioni ecc.), esempi di tali metriche sono: Availability, Performability, Reliability ecc.
- **Metriche per i costi:** Si basano i criteri sui costi impiegati per l'implementazione di particolari sistemi

Uno dei criteri di selezione della metrica è quello presentato nell'immagine [1.2], che ci permette di capire, in base a come reagisci il sistema, quale tipologia e quale classe di parametri andare a considerare. Fare attenzione all'immagine, essa suddivide le possibili

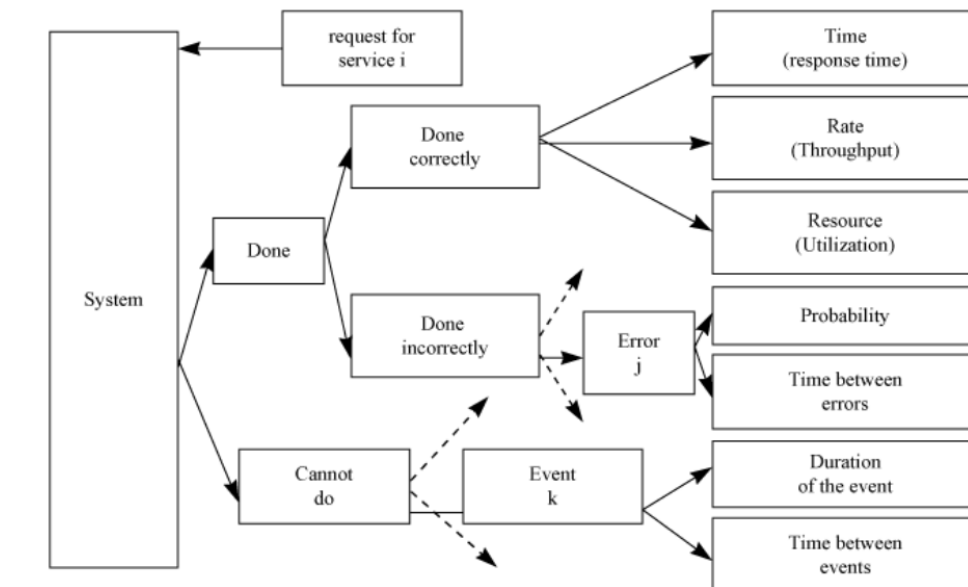


Figure 1.2: Criterio di selezione della metrica

metriche in tre macroaree, la prima è quella riguardanti le metriche deterministiche, ovvero quelle metriche che vengono valutate su casi effettivi e risultati non di natura probabilistica; a differenza delle altre 2 categorie che, avendo a che fare con errori ecc., ricadono nei casi di dover andare a valutare il sistema mediante dei valori di natura probabilistica.

Le metriche presentate, quindi, sono le seguenti:

- **Response Time:** Tale parametro ci permette di capire ogni quanto di tempo un sistema produce un risultato valido. Per la valutazione di tale parametro, però, possono essere effettuate varie tipologie di valutazione:
 - **Response time:** Misurazione basata sul tempo tra l'inizio e la fine della richiesta
 - **Reaction time:** Fine della richiesta dall'inizio del suo processamento
 - **Turnaround time:** Inizio della richiesta fino alla fine della risposta

Generalmente la Response Time cresce all'aumentare del carico sul sistema, per tale caso è stato definito quello che verrà chiamato **Strech Factor** (permette di capire quando non saranno più usabili un certo numero di risorse)

- **Processing Rate:** Il processing rate non rappresenta altro che il throughput associato al mio sistema, e che quindi calcola la quantità di lavoro svolto da un singolo componente per unità di tempo.

Talvolta, utilizzare sia il throughput che il response time può risultare ridondante, pertanto si decide di utilizzare un unico parametro, dipendente da entrambi, chiamato **potenza**, che si può calcolare come $\frac{Throughput}{ResponseTime}$. Pertanto è giusto andare a ridefinire anche i diversi punti di evoluzione della potenza, che racchiudono la nostra attenzione

- **Capacità nominale:** Rappresenta il throughput massimo raggiungibile in condizioni di carico di lavoro ideali. Tuttavia, a questo livello di throughput, il tempo di risposta è generalmente troppo elevato.
- **Capacità utilizzabile:** È il throughput massimo che si può ottenere, quindi è il punto massimo dopo il quale il sistema potrebbe andare in crash (quindi ad esempio ha tempi di risposta molto lunghi ma riesce a gestire molto più carico)
- **Capacità di "ginocchio":** È il punto operativo ottimale, considerato il miglior equilibrio tra un alto throughput e un basso tempo di risposta. Questo punto corrisponde al valore massimo della metrica Power.

Oltre la potenza un ulteriore parametro utile è la **fairness**, che ci permette di capire se un particolare sistema distribuisce bene il carico o meno. Ciò lo veniamo a scoprire mediante il calcolo del preciso valore di fairness che è normalizzato, e quindi compreso tra 0 ed 1.

A livello formale, si definisce:

- x_i , frazione del throughput associato ad x_i
- n , numero di utenti nel sistema

A questo punto comprendo se sto usando il throughput in maniera fair, calcolando la **fairness**, che mi dice che se ogni utente ha a disposizione una porzione eguale di throughput, allora sarà 1, mentre nel caso opposto sarà vicino allo 0. Per calcolare la fairness si utilizza la seguente formula:

$$fairness = \frac{(\sum_{i=1}^n x_i)^2}{n \sum_{i=1}^n x_i^2}$$