

# Impianti Informatici

anno accademico 2015/2016

M. Arrigoni Neri & P. Borghese

# indice

- organizzazione e consigli
  - testi
  - obiettivi - modalità di esame
- contenuti del corso
- un approccio sistematico alla valutazione delle prestazioni

# Riferimenti

- Docente
  - Mario Arrigoni Neri
    - Ricevimento lunedì 10:00-12:00
    - [mario.arrigonineri@unibg.it](mailto:mario.arrigonineri@unibg.it)
- Collaboratori
  - Paolo Borghese
    - [paolo.borghese@promo.it](mailto:paolo.borghese@promo.it)

# Organizzazione

- materiale on-line <http://cs.unibg.it/arrigoni/impianti/>
  - slide delle lezioni
  - articoli di approfondimento
  - esercizi risolti e temi d'esame
- vostre responsabilità:
  - studio personale
    - leggere i capitoli preferibilmente prima delle lezioni
  - usare il web per approfondire gli argomenti trattati
  - partecipare attivamente
  - intervenire se qualche argomento non è chiaro (direttamente in aula, al “ricevimento” o via e-mail)
  - è possibile
    - svolgere un **progetto** o ricerca (per esempio: lettura di un articolo interessante soluzione di un problema proposto) e farne una breve presentazione in aula

# Orario (tentativo) – ora di inizio?

data	lez	es	argomenti lezione	argomenti esercitazione
29/02/16	2		introduzione al corso / Metriche	
07/03/16	2	2	dischi	metriche
14/03/16	2		cpu	
21/03/16	2	2	reti di comunicazione	es dischi / cpu
29/03/16		Sospensione per tesi		
04/04/16	2	2	Virtualizzazione e cloud	es dischi / cpu / reti
11/04/16	2		calcolo parallelo	
18/04/16	4		laboratorio bluemix	
25/04/16		Vacanze di Pasqua / festa della liberazione		
02/05/16	4		analisi operazionale	
09/05/16	2	2	reti di code	analisi operazionale
16/05/16	4		reti di code	
23/05/16	2	2	affidabilità	Analisi operazionale / reti code
30/05/16	2	2	raid	reti code
06/06/16	2	2	simulazione	affidabilità
08/06/16		2		preparazione al compito
	32	16		

# Libri

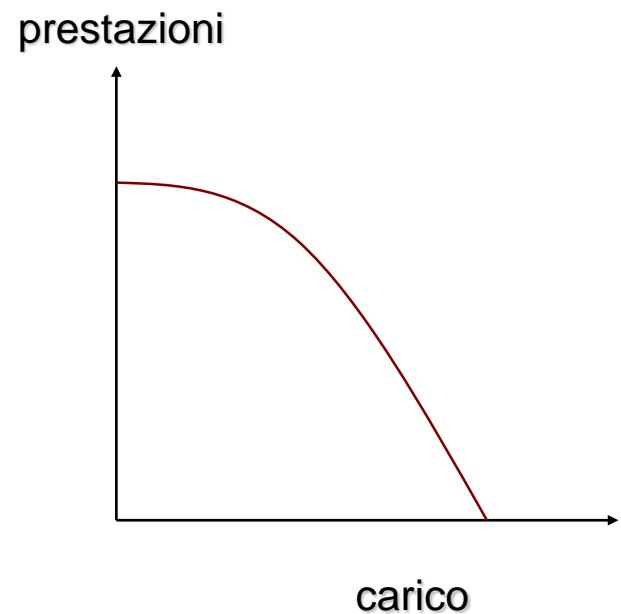
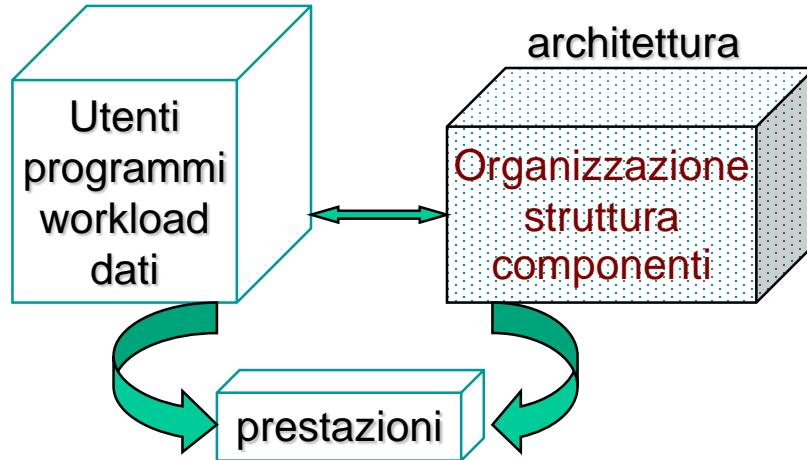
- **Testi di riferimento**
  - G. Iazeolla: *Impianti reti sistemi informatici*, Franco Angeli, 2004
  - E.D. Lazowska, J. Zahorjan, G.S. Graham, K. Sevcik: *Quantitative System Performance*, Prentice Hall, 1984
    - (capitoli da 1 a 8). Il libro può essere scaricato dall'indirizzo:  
<http://www.cs.washington.edu/homes/lazowska/qsp/>
  - K. Kant: *Introduction to Computer System Performance Evaluation*, McGraw-Hill, 1992
  - D.A. Patterson, J.L. Hennessy: *Computer Organization & Design: The Hardware/Software Interface*, Morgan Kaufmann, 2005
  - L. Kleinrock: *Queueing Systems*, Vol 1 & 2, John Wiley & Sons, 1975

# argomento del corso: impianti informatici dal punto di vista delle **prestazioni** (in senso lato)

- utenti, gestori e progettisti dei sistemi informatici si pongono generalmente un obiettivo di ottimizzazione, per esempio di rendere massime le prestazioni per un dato costo o minimo il costo per determinate prestazioni
- le **prestazioni** sono perciò un criterio fondamentale nella progettazione, acquisizione ed uso di un sistema informatico
- è allora necessario che tutti coloro che hanno a che fare con i sistemi informatici dispongano di una sufficiente conoscenza della **terminologia**, delle **teorie** e delle **tecniche di valutazione delle prestazioni**
  - tali conoscenze sono di supporto all'intuizione
- l'attività di **modeling**, cioè la sperimentazione con un modello del sistema, costituisce una buona preparazione alle decisioni da prendere nel mondo reale

## impianti informatici dal punto di vista delle prestazioni (cont.)

- la valutazione delle prestazioni è un'attività spesso complessa perché in un sistema informatico interagiscono numerosi fattori, tutti interconnessi in modo non lineare e variabile nel tempo



# obiettivi del corso

- il corso si propone di rendervi in grado di:
  - porre le giuste domande
    - *un problema è in parte risolto se queste sono ben formulate*
  - conoscere le corrette metriche di valutazione di un sistema
  - comprendere la letteratura sull'analisi delle prestazioni e sulla modellazione dei sistemi informatici
  - usare con metodo e consapevolezza e al giusto livello di dettaglio tecnologie e strumenti di misura e previsione delle prestazioni
  - gestire un sistema informatico dal punto di vista della qualità del servizio (prestazioni, affidabilità, disponibilità, soddisfazione dell'utente in generale)

## note

- saranno affrontati gli aspetti più generali e astratti dei sistemi informatici che possono essere trattati con metodi matematici, si osservi che:
  - riconoscere l'identità in problemi apparentemente differenti è di grande valore pratico oltre che di profonda importanza concettuale;
  - se si è disposti a fare un piccolo sforzo per inquadrare i problemi in uno schema teorico, molti aspetti appariranno ovvi e ben connessi fra loro, invece che slegati e confusi;
  - ovviamente bisognerà introdurre alcune semplificazioni di cui si cercheranno di studiare le conseguenze e i limiti.
- il corso NON riguarda:
  - i dettagli degli strumenti di simulazione e di analisi (JMT);
  - l'implementazione di strumenti;
  - esperimenti e misure su sistemi reali

# contenuti del corso

- Il processo di valutazione delle prestazioni
- Metriche: indici di prestazione e misure
- Dischi: anatomia delle operazioni di I/O
  - Architetture storage: SAN - NAS
- CPU: come si misurano e confrontano le prestazioni di CPU (e sistemi)
- Affidabilità di un sistema: come si calcola
  - Esempio: dischi RAID - livelli - prestazioni e affidabilità
- Analisi operazionale: Leggi - applicazioni
- Reti di code: teoria elementare - caratteristiche principali e metodi di soluzione

# contenuti del corso (cont.)

- Simulazione: struttura e impiego di un simulatore
  - Esempio: Architetture Web - organizzazione e prestazioni
- Calcolo parallelo: caratteristiche
- Virtualizzazione e cloud computing
  - Eventuale intervento esterno
- Esame
  - prova scritta: teoria + soluzione di un problema reale
    - (materiale consultabile durante la prova)
  - presentazione in aula

# **un approccio sistematico alla valutazione delle prestazioni**

# il processo di valutazione delle prestazioni

- obiettivi specifici:
  - analisi e comprendere del comportamento di un sistema
  - individuazione dei colli di bottiglia (componenti che limitano la produttività)
  - ottimizzazione
  - confronto fra soluzioni, algoritmi, protocolli alternativi
  - capacity planning: calcolo della capacità residua
  - valutazione della scalabilità
    - relazione fra carico-prestazioni-capacità
  - calcolo delle prestazioni al variare delle condizioni e dei parametri (componenti, mix di carichi, priorità, ecc.)

# Terminologia

- **Sistema (impianto informatico):**
  - un complesso di componenti hardware e software interconnessi atto a eseguire un certo lavoro con date prestazioni
- **Modello:**
  - rappresentazione (matematica) di un concetto, fenomeno o sistema
- **Metriche:**
  - grandezze usate nella valutazione delle prestazioni di un sistema, criteri e misure impiegate
- **Workload (carico):**
  - richieste fatte dagli utenti al sistema (domande di risorse)
    - problema: come mappare le richieste applicative in domande di risorse

# Terminologia (cont.)

- **Parametri:**

grandezze caratteristiche del sistema e del carico da cui dipendono le prestazioni (da scegliere in base agli obiettivi)

- **Fattori:**

parametri che vengono variati in uno studio, in particolare quelli che rappresentano il comportamento degli utenti

- **ogni grandezza deve essere:**

- definita in base a una procedura di misurazione, cioè **in modo operativo**
- misurabile in modo semplice e non ambiguo
- studiata in base alle condizioni al contorno e ai suoi effetti sui parametri del sistema

# passi principali del processo di valutazione

- definizione degli obiettivi dello studio
- individuazione dei servizi e delle metriche appropriate
- identificazione dei parametri
- selezione dei metodi e strumenti di valutazione
- scelta e caratterizzazione dei carichi
- scelta dei periodi base delle misure (durata, carichi medi, di picco ecc.)
- misurazioni
- costruzione e soluzione di un modello
- progettazione degli esperimenti (passo fondamentale nella simulazione)
- analisi e presentazione dei risultati

# passi principali del processo (cont.)

- obiettivi dello studio
  - da definirsi in modo **preciso** perché da essi dipendono le decisioni successive sul livello di dettaglio, le scelte del tipo di modello e del metodo di soluzione, delle metriche e dei carichi da misurare
  - il sistema deve essere individuato unitamente ai suoi **confini** e interrelazioni col resto del mondo
  - gli obiettivi devono essere il più possibile:
    - specifici
    - accettabili
    - completi
    - misurabili
    - realizzabili

# passi principali del processo (cont.)

## ■ individuazione dei servizi

- identificazione dei servizi offerti dal sistema e dei possibili risultati, delle metriche e legami funzionali con l'architettura (struttura e organizzazione del sistema e delle applicazioni)

## ■ individuazione delle metriche appropriate

- quantità di servizio – *tassi di esecuzione del carico*
- frequenza di risultati corretti – affidabilità
- durata di attività – disponibilità
- efficienza nell'uso delle risorse – *utilizzo*
- bassa variabilità
  - si riduce il numero di misure / soluzioni
- non ridondanza
- completezza
- metriche di equità – *fairness*

# passi principali del processo (cont.)

- metriche più comuni
  - *tempo di risposta*
    - turnaround time
  - throughput
    - capacità nominale: con un carico ideale
    - capacità effettiva: con un tempo di risposta accettabile
  - *utilizzo (di un componente)*
    - tempo busy (del componente) / durata dell'osservazione
  - efficienza
    - capacità effettiva / capacità nominale
    - fattore di elongazione di un particolare carico (rapporto fra tempo di risposta misurato e quello minimo ottenuto in condizioni ottimali)

# passi principali del processo (cont.)

- identificazione dei parametri
  - parametri di sistema (caratteristiche del sistema che determinano le prestazioni) es. ampiezza di banda
  - parametri di carico (caratteristiche di uso e attributi del carico)
  - devono essere caratterizzati in base agli obiettivi e agli effetti sulle prestazioni
  - determinare il range della variazione
  - decidere una (o poche) variabili che variano mentre le altre sono fisse

# passi principali del processo (cont.)

- metodi e strumenti di valutazione
  - misure
  - modelli di simulazione
  - modelli analitici
- una combinazione di vari metodi
- la decisione deve essere guidata dagli obiettivi
- *bisogna tenere conto anche delle competenze a disposizione*

## misura diretta

- fornisce risultati realistici
- si possono testare i limiti del carico
- il sistema o un prototipo deve essere funzionante e disponibile per i test
- non è facile trovare correlazioni causa – effetto
- difficoltà:
  - definizione delle metriche appropriate
  - uso del carico adatto
  - strumenti statistici di analisi dei dati

# passi principali del processo (cont.)

- **modello**
  - rappresentazione idealizzata e semplificata della realtà, costruita attraverso un processo di astrazione
  - un buon modello deve:
    - contenere solo gli aspetti più importanti
    - essere abbastanza generale per affrontare un gamma sufficientemente ampia di fenomeni
    - essere risolubile in modo abbastanza semplice
  - lo stesso modello può rappresentare sistemi diversi
  - lo stesso sistema può essere rappresentato da differenti modelli

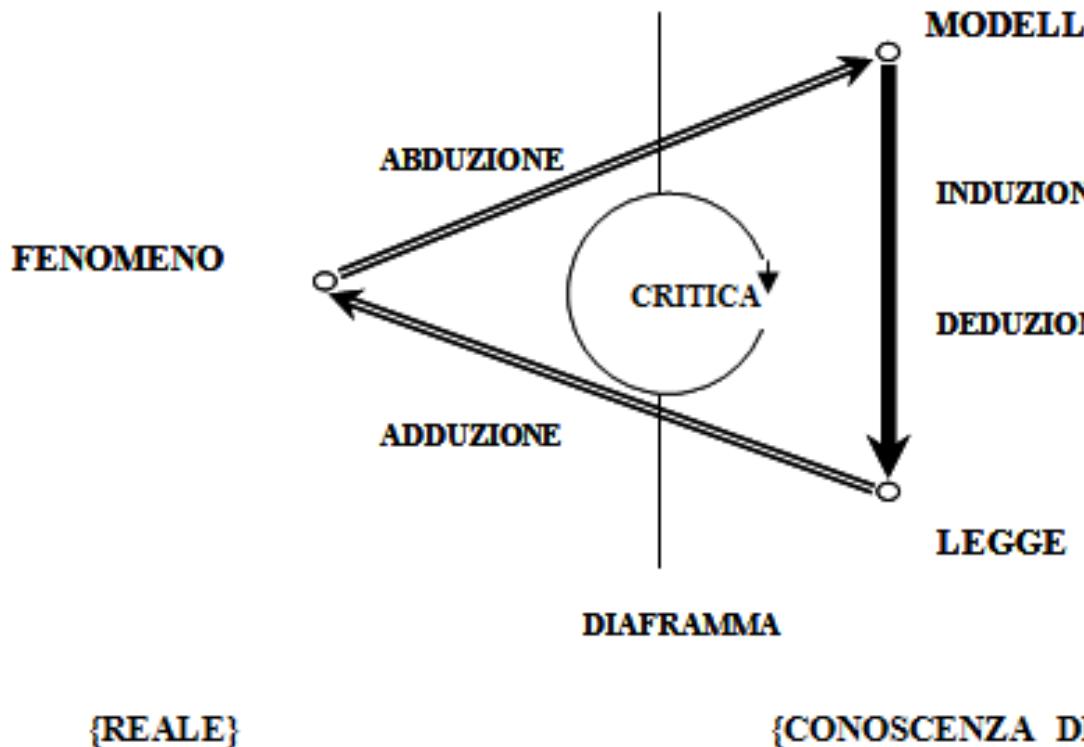
# passi principali del processo (cont.)

- **modello (cont.)**
  - un modello non deve essere complicato per produrre risultati significativi
  - ma è valido se cattura gli aspetti fondamentali
  - un modello complesso è giustificato solo dagli obiettivi e dagli input (misure) di cui si dispone
  - bisogna sempre aver presente che il modello matematico è altro che la realtà
  - la costruzione di un modello richiede anche competenze implicite (di lavoro) che non sono oggetto del corso es. “importanza” del tipo di carico
    - perciò verranno trattati modelli senza rendere esplicito nei dettagli il procedimento che ha portato alla loro costruzione

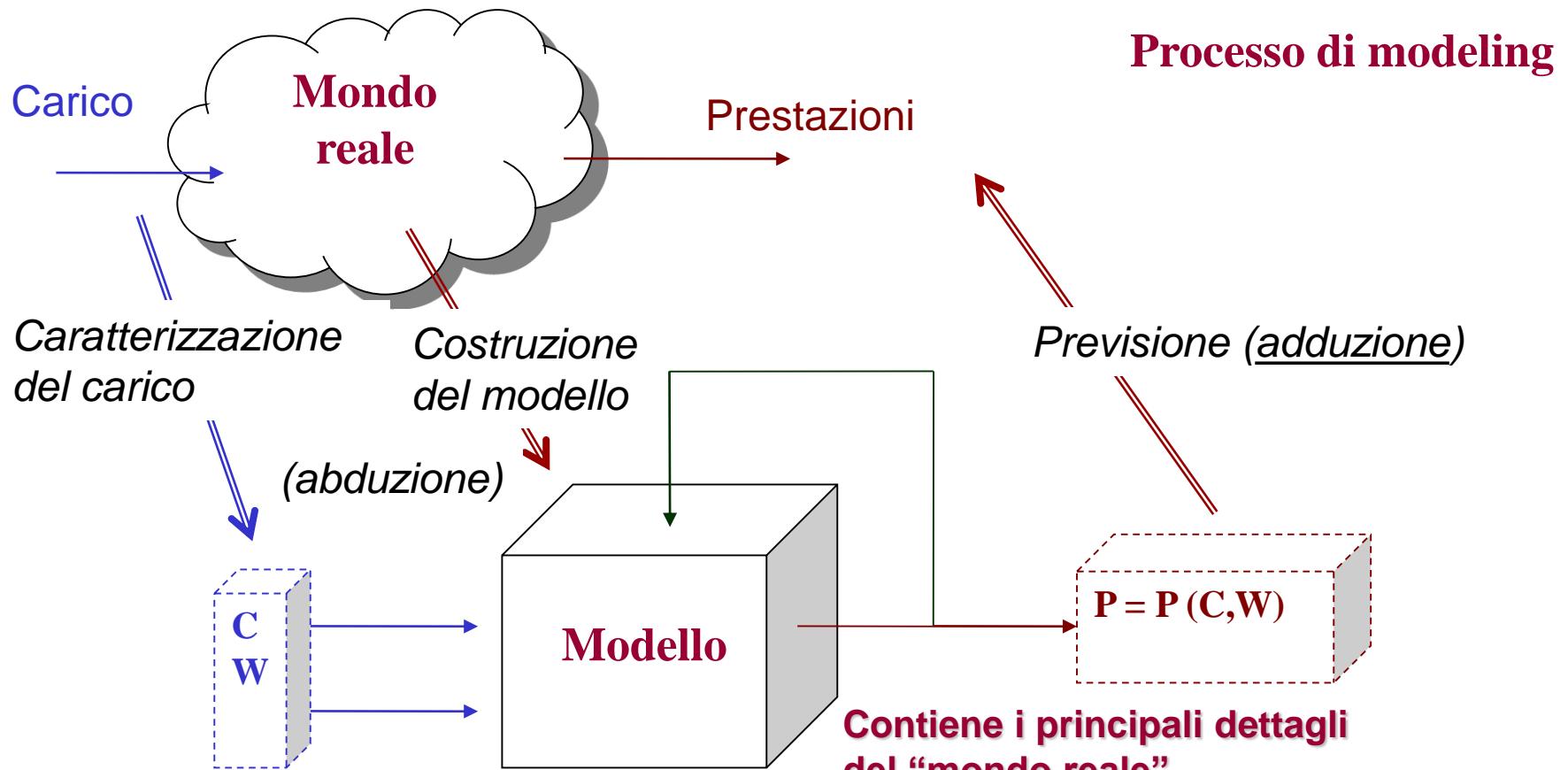
# alcune tipologie di modelli

- modelli funzionali: rappresentano i componenti e le relazioni
- modelli statistici: si basano su relazioni quantitative ricavate da misurazioni - non contengono la logica dei fenomeni
  - esempio: metodi di regressione
- modelli stocastici: basati su reti di code
  - le **code** rappresentano i componenti del sistema che vengono “visitati” dalle transazioni.
  - code attive (dispositivi - CPU, I/O,...)
  - code passive (memoria, canali,...)
  - generalmente in un dato istante una *transazione* si trova in una coda attiva ma può **contemporaneamente occupare risorse passive** (es. occupa la CPU e risiede in memoria)

**PARADIGMA EMPIRICO INDUTTIVO E DEDUTTIVO  
(GALILEO GALILEI)**



AFFERENZA = ABDUZIONE + ADDUZIONE  
INFERENZA = INDUZIONE + DEDUZIONE  
INTELLIGENZA = AFFERENZA + INFERENZA  
INTELLIGENZA = CREATIVITA' + FABBRICATIVITA'



C: Configurazione

W: Carico

P: Prestazioni

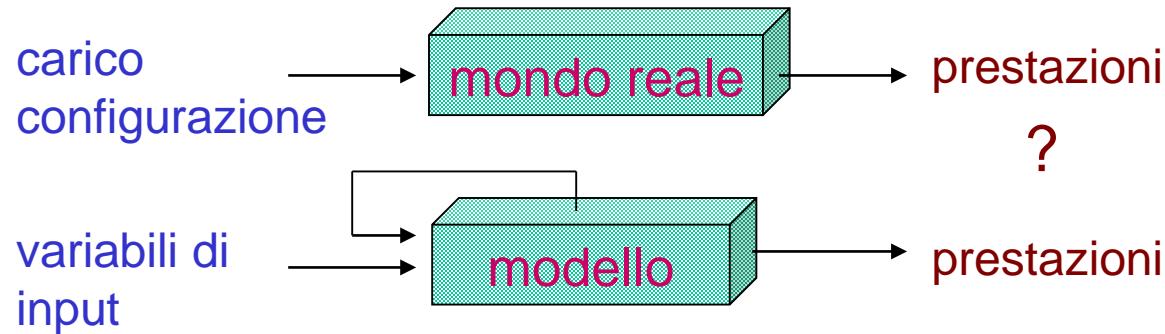
**Contiene i principali dettagli  
del “mondo reale”**

**scopi del modello:**

**analisi e comprensione  
del comportamento**

**stima delle prestazioni  
delle applicazioni**

# processo di modeling

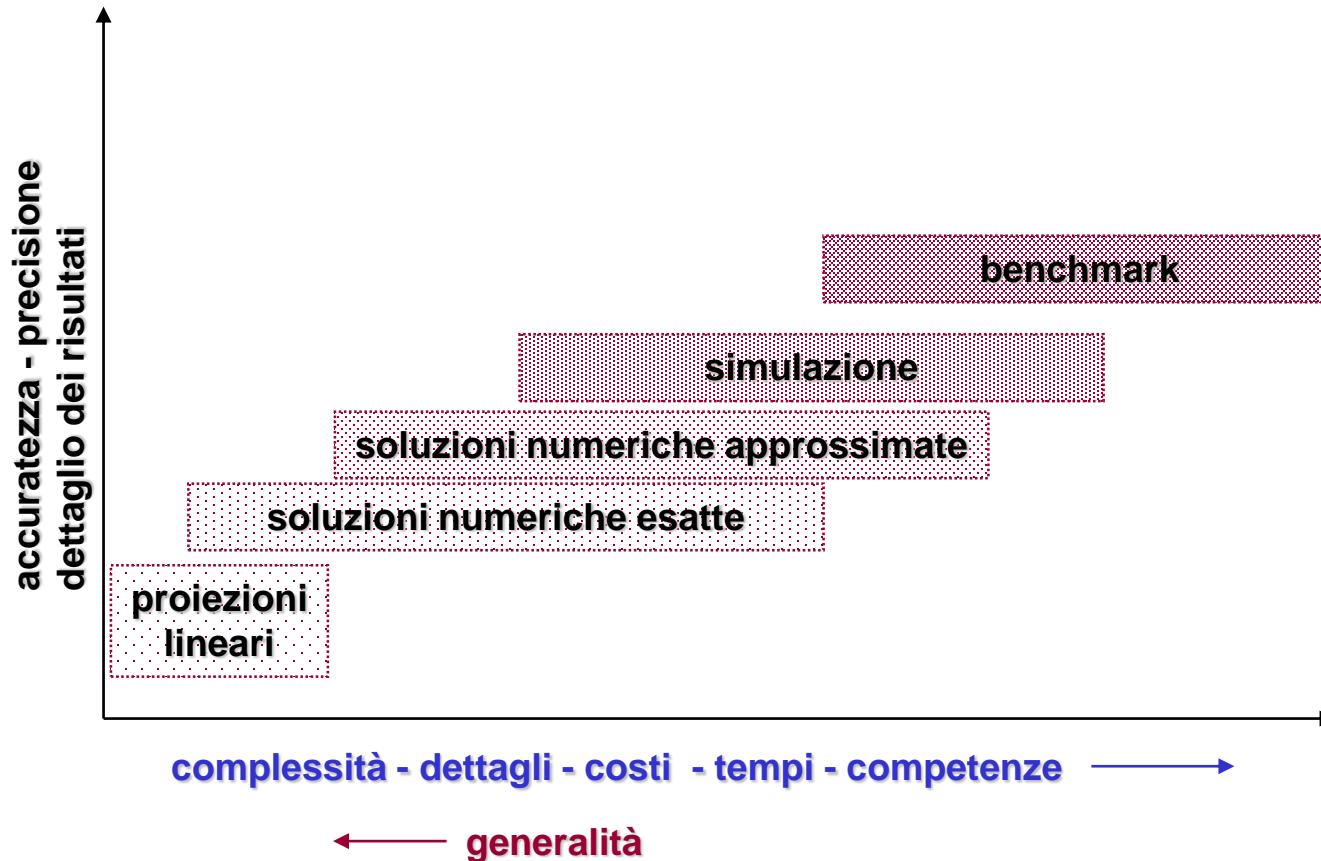


- *parametrizzazione*: caratterizzazione del carico
- *verifica* della consistenza interna del modello
- *calibrazione*: stima e tuning di parametri (es.suddivisione dei consumi di risorse per overhead)
- *convalida*: confronto con misurazioni reali, studio della sensibilità della risposta alle variazioni dei parametri
- *progettazione* ed esecuzione degli esperimenti
- *analisi* dei risultati

# errori di un modello

- errori di **rappresentazione**, dovuti alle approssimazioni usate nello stabilire la corrispondenza fra la realtà e la sua rappresentazione astratta (modello). A questo tipo di errori appartengono quelli indotti dal livello di dettaglio sia dei comportamenti logici sia delle misure. Lo studio delle prestazioni attraverso metodi **analitici** è generalmente poco costoso ma le semplificazioni (es. arrivi casuali) introdotte per rendere risolubile il modello possono essere, in qualche caso, drastiche;
- **errori della soluzione**, dipendenti dal metodo usato (solo l'analisi esatta non è soggetta a tale tipo di errore);
- errori dovuti ad **approssimazioni** e stime imperfette **dell'ambiente futuro**; tale tipo di errore è normalmente quello maggiore
  - questa è una delle ragioni per cui spesso non è consigliabile usare modelli molto raffinati e dettagliati a scopo di dimensionamento.

# spettro dei metodi di soluzione



# spettro dei metodi di soluzione (cont.)

- **benchmark:**
  - indica sia il processo di misurazione di un sistema reale che riproduce e modella il sistema che ci interessa sia l'insieme di applicazioni e dati che sono usati per confrontare diverse soluzioni / ipotesi
  - per generare il carico applicativo viene utilizzato un sistema esterno che deve contenere la logica del comportamento degli utenti (almeno entro certi limiti)

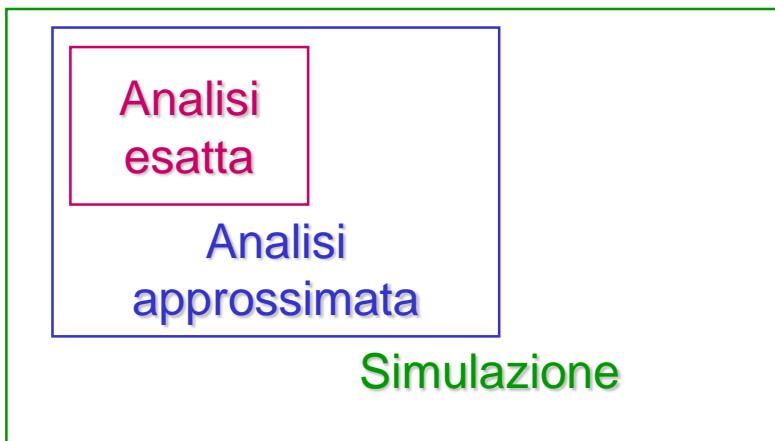
# metodi di soluzione di modelli basati su reti di code

- **soluzione analitica esatta: formula chiusa - algoritmo**
  - è praticabile solo se si accettano alcune ipotesi restrittive
  - l'analisi operazionale è una tecnica di soluzione delle reti
- **soluzione analitica approssimata**
  - la copertura delle tecniche approssimate è molto più ampia, ma la determinazione degli errori di soluzione è difficile
- **simulazione:** esperimento che consiste nella osservazione del comportamento del modello al trascorrere del tempo simulato
  - input al modello:
    - traccia ottenuta dal mondo reale (**trace driven**)
    - sintesi statistica ottenuta da misure e/o ipotesi (**data driven**)
  - output del modello:
    - indici di prestazione

# passi principali del processo (cont.)

## ■ soluzione

- nei modelli **analitici**:
  - derivazione matematica dei risultati (stime dei comportamenti del sistema)
- nelle misure e simulazioni:
  - analisi **statistica** dei dati raccolti
  - sommarizzazione e presentazione dei risultati ottenuti



## modello analitico

- può dare garanzie più forti sul comportamento atteso
  - (cioè mostra agevolmente l'effetto delle variazioni)
- non richiede la costruzione di un prototipo
- può fornire indicazioni su relazioni causa – effetto
- le previsioni dipendono decisamente dalla adeguatezza del modello
- difficoltà:
  - aspetti matematici
  - scelta del modello corretto
  - l'apprendimento richiede molto impegno (tempo) perché il livello di astrazione è elevato

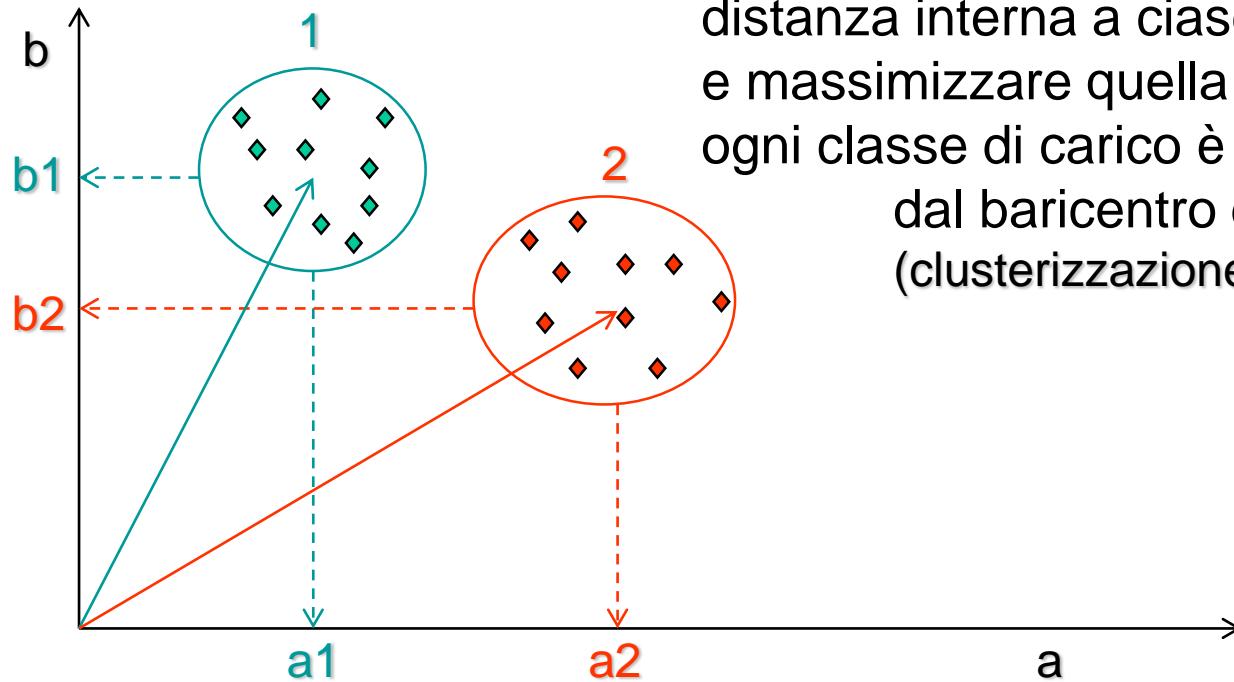
# modello di simulazione

- meno dispendioso della costruzione di un prototipo
- si possono testare più scenari di carico
- il modello è solo un oggetto simile al sistema reale costruito generalmente a partire da un linguaggio opportuno
- non può essere usato per *garantire* le prestazioni attese
- difficoltà:
  - uso del modello
  - sua correttezza
  - rappresentazione dei risultati
  - apprendimento degli strumenti

## passi principali del processo (cont.)

### ■ selezione e caratterizzazione dei carichi

- ogni unità di carico (job, transazione) è rappresentata da un vettore in uno *spazio di parametri*
- il carico viene suddiviso in classi in base a opportune considerazioni statistiche in modo da minimizzare la distanza interna a ciascun gruppo e massimizzare quella tra gruppi  
ogni classe di carico è rappresentata dal baricentro del gruppo (clusterizzazione)



## passi principali del processo (cont.)

- progettazione degli esperimenti (design)
  - si propone di ottenere la **massima informazione** con il minimo sforzo
  - la preparazione degli esperimenti di misura richiede molto lavoro
  - bisogna tentare di ridurne il numero
  - occorre ripetere l'esperimento con certe modalità (se si tratta di misura o simulazione) un numero sufficiente di volte perché abbia validità statistica
  - non dimenticare gli obiettivi

# passi principali del processo (cont.)

- **presentazione dei risultati**
  - con i metodi analitici:
    - descrizione degli algoritmi, validità e limiti
    - mostrare gli effetti della modifica dei parametri
    - esporre i dati di convalida
  - con misure e simulazioni:
    - spiegazione chiara gli obiettivi degli esperimenti
    - lista delle ipotesi e assunzioni
    - presentazione grafica
  - in tutti i casi bisogna:
    - discutere le implicazioni per gli utenti
    - presentare le conclusioni (cosa si è appreso, sorprese, nuove direzioni di indagine)
    - discutere le limitazioni e il lavoro futuro

# processo di valutazione delle prestazioni (cont.)

- conclusioni:
- il processo è un **ciclo** composto da varie fasi, generalmente ripercorso più volte con un livello di **dettaglio** via via crescente:
  - all'inizio possiamo limitarci ad individuare le soluzioni possibili;
  - poi eliminare quelle meno adatte;
  - determinare la capacità necessaria per eseguire un certo carico di lavoro rispettando livelli di servizio prefissati;
  - studiare i colli di bottiglia e i comportamenti asintotici;
  - in un momento successivo, quando avremo a disposizione dati più precisi, saremo in grado di stimare i valori medi dei tempi di risposta applicativi;
  - infine ottenere altri indici caratteristici delle distribuzioni delle grandezze (varianze, percentili).

## processo di valutazione delle prestazioni (cont.)

- procedendo per approssimazioni successive si producono risultati validi in ogni fase dello studio che può avere inizio quando i dettagli della realtà non sono ancora ben definiti e terminare quando si siano raggiunti risultati con un livello di dettaglio sufficiente e non più migliorabile a un costo ragionevole.
- siete invitati a studiare e usare lo strumento di modeling JMT (Java Modelig Tool) “scaricabile” da: <http://jmt.sourceforge.net/>