

Fondamenti di Informatica e Basi di Dati a.a. 2019/2020

DOCENTE: DOTT.SSA VALERIA FIONDA

PARZIALMENTE BASATE SUL MATERIALE DEL PROF. **MARCO DI FELICE**

Organizzazione del corso e modalità d'esame

Calendario Didattico

- I corsi iniziano il 30.09.2019 e terminano il 21.12.2019.
- La sessione invernale di esami va dal 07.01.2020 al 29.02.2020.

Orario delle lezioni

- Mercoledì 14.00 – 17.00, aula E, cubo 20B, IV piano.
- Venerdì 11.00 – 14.00, aula E, cubo 20B, IV piano / Laboratorio di Informatica, cubo 16 C, VI piano.
- Lezioni
 - Lezione 1: 02/10/2019
 - Lezione 2: 04/10/2019
 - Lezione 3: 09/10/2019
 - Lezione 4: 11/10/2019
 - Lezione 5: 16/10/2019
 - Lezione 6: 18/10/2019
 - Lezione 7: 23/10/2019
 - Lezione 8: 23/10/2019
 - Lezione 9: 30/10/2019
 - Lezione 10: 06/11/2019
 - Lezione 11: 08/11/2019
 - Lezione 12: 13/11/2019
 - Lezione 13: 15/11/2019

Orario delle lezioni

- Mercoledì 14.00 – 17.00, aula E, cubo 20B, IV piano.
- Venerdì 11.00 – 14.00, aula E, cubo 20B, IV piano / Laboratorio di Informatica, cubo 16 C, VI piano.
- Lezioni
 - Lezione 14: 20/11/2019
 - Lezione 15: 22/11/2019
 - Lezione 16: 27/11/2019
 - Lezione 17: 29/11/2019
 - Lezione 18: 04/12/2019
 - Lezione 19: 06/12/2019
 - Lezione 20: 11/12/2019
 - Lezione 21: 13/12/2019

Informazioni generali

- Ricevimento: ogni lunedì dalle 15 alle 16
- Pagina del corso: www.mat.unical.it/fionda → teaching → Fondamenti di Informatica e Basi di Dati
- Email docente: fionda@mat.unical.it / valeria.fionda@unical.it

Programma del corso

Parte 1 Basi di Dati

Lezioni dal 02/10/2019 – 08/11/2019

- Sistemi informativi e basi di dati
- Modello entità relazione
- Modello relazionale
- SQL

Parte 2 Introduzione alla programmazione e Python

Lezioni dal 13/11/2019 – 13/12/2019

- Introduzione all'Informatica.
- Il linguaggio dei computer.
- Problem solving
- Algoritmi
- Python

Testi Consigliati

Parte 1 (Basi di Dati)

Titolo: Basi di Dati

Autori: Paolo Atzeni, Stefano Ceri, Piero Fraternali, Stefano Paraboschi e Riccardo Torlone

Editore: McGraw-Hill

Parte 2 (Introduzione alla programmazione e Python)

Titolo: Think Python – How to Think Like a Computer Scientist

Autori: A. B. Downey

<https://runestone.academy/runestone/books/published/thinkcspy/index.html>

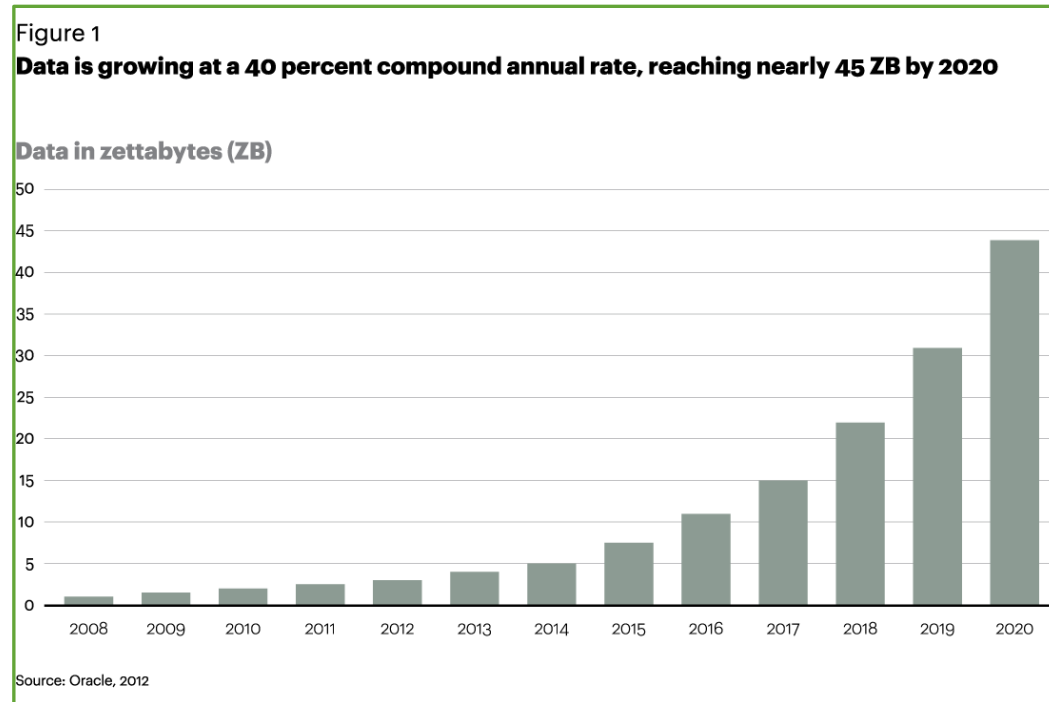
Modalità d'esame

- L'esame consta di:
 - Una prova scritta obbligatoria:
 1. Un esercizio di programmazione Python;
 2. Un esercizio di basi di dati (progettazione + interrogazione).
 - Una prova orale
 - Gli esiti della prova scritta saranno resi pubblici tramite il sistema di prenotazione degli esami esse3.

I dati

I dati nel mondo reale

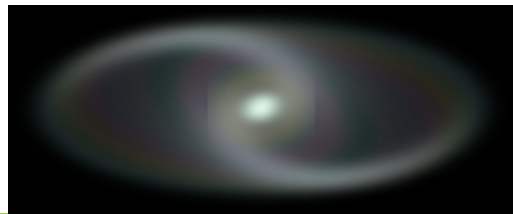
Source: <http://www.atkearney.it/>



WHAT'S A ZETTABYTE?

1 kilobyte	1,000,000,000,000,000,000
1 megabyte	1,000,000,000,000,000,000,000
1 gigabyte	1,000,000,000,000,000,000,000,000
1 terabyte	1,000,000,000,000,000,000,000,000,000
1 petabyte	1,000,000,000,000,000,000,000,000,000,000
1 exabyte	1,000,000,000,000,000,000,000,000,000,000,000
1 zettabyte	1,000,000,000,000,000,000,000,000,000,000,000,000

SOURCE: GIGABYTE



I dati nel mondo reale

✧ Chi produce questi dati?

1. Sistemi Informativi aziendali



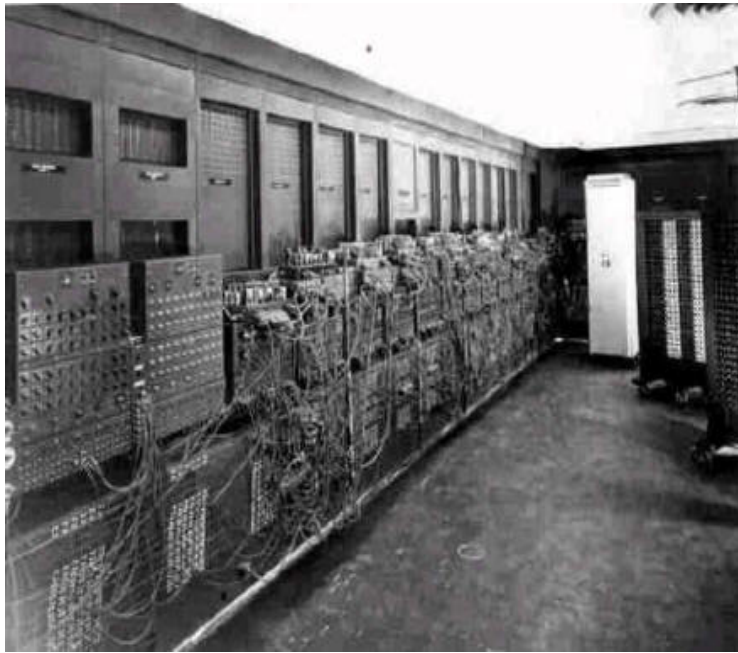
1. Social media



1. Sorgenti di big-data



L'Informatica di ieri



ENIAC (1954)
L'era dei supercalcolatori



IBM 5150 (1980)
L'era dei PC



IPHONE GSM (2007)
L'era degli smartphone

L'Informatica di Oggi

L'informatica **pervasiva** ...

FITBIT



SMARTWATCH



SMART LIGHTING
(Philips Hue)



TESLA AutoPilot

Perchè i dati sono importanti?

R1. DATI come risorsa aziendale, *alla stessa maniera del capitale, degli impianti di produzione, delle persone, e dei beni prodotti dall'azienda.*

Necessità ad ogni livello
di **strumenti software** per
custodire, manipolare, analizzare i dati!

ANTHONY






Personale

Dati di **dettaglio**: ordini, dati fatturazioni, bilancio, contabilità, etc

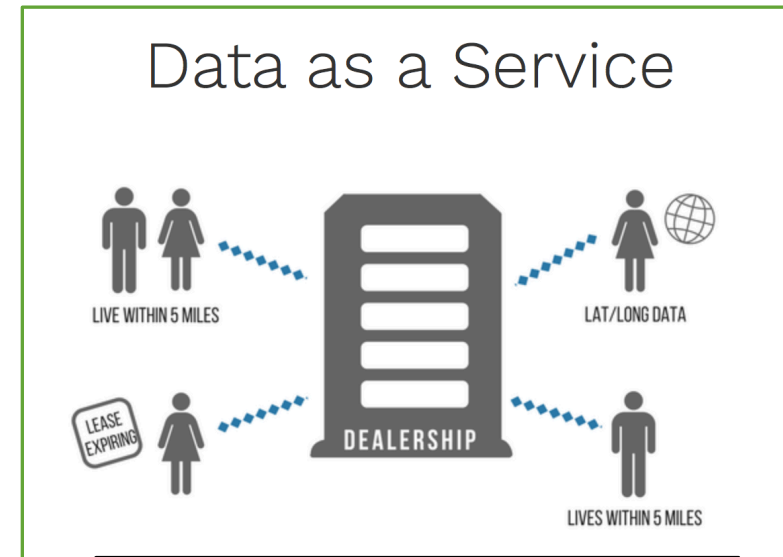
Perchè i dati sono importanti?

R2. DATI come bene prodotto dall'azienda e fonte di profitto!

Product Compare (0)

 100 Twitter Tweets We will deliver 100+ quality Twitter tweets, ideal for increasing your social si..	\$4.50	Add to Cart <input type="checkbox"/> Add to Wish List <input type="checkbox"/> Add to Compare
 200 Twitter Tweets We will deliver 200+ quality Twitter tweets ideal for increasing your social sig..	\$6.44	Add to Cart <input type="checkbox"/> Add to Wish List <input type="checkbox"/> Add to Compare
 300 Twitter Tweets We will deliver 300+ quality Twitter tweets, ideal for increasing your social si..	\$10.79	Add to Cart <input type="checkbox"/> Add to Wish List <input type="checkbox"/> Add to Compare
 400 Twitter Tweets We will deliver 400+ quality Twitter tweets, ideal for increasing your social si..	\$11.94	Add to Cart <input type="checkbox"/> Add to Wish List <input type="checkbox"/> Add to Compare
 600 Twitter Tweets We will deliver 600+ quality Twitter tweets, ideal for increasing your social si..	\$14.44	Add to Cart <input type="checkbox"/> Add to Wish List <input type="checkbox"/> Add to Compare

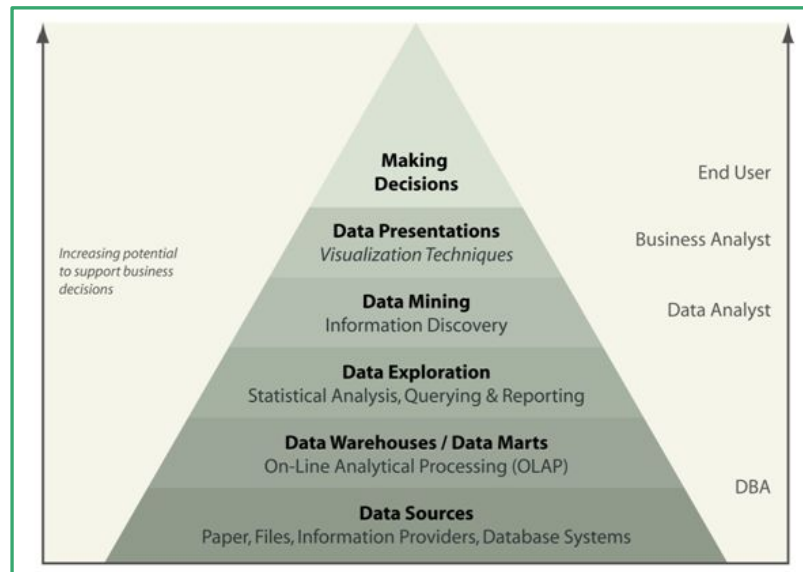
<http://retweets.pro/buy-twitter-tweets>



<http://www.bigdataexchange.com>

Perchè i dati sono importanti?

R3 → DATI = informazione = conoscenza = supporto alle decisioni!



Source: <http://www.conbusinessintelligence.com/>

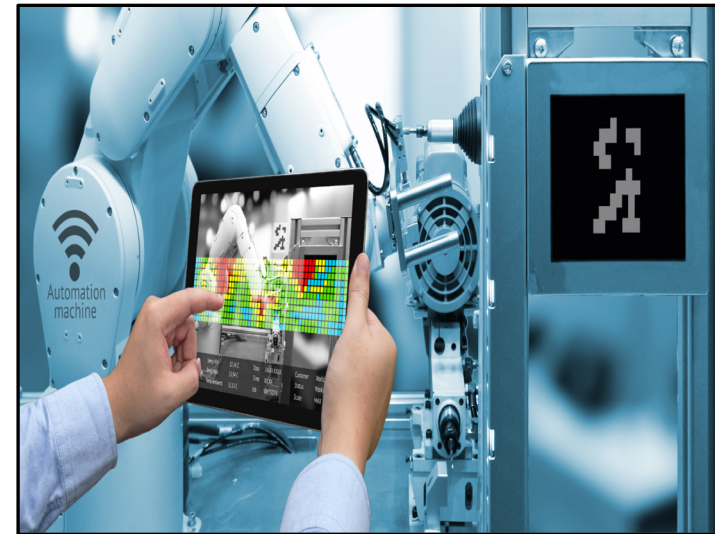
❑ **BUSINESS INTELLIGENCE (BI)** →
(def.) Insieme di **processi aziendali, metodologie tool** per raccogliere i dati di un'azienda, ed estrarre **informazioni di supporto alle decisioni strategiche**.

Perchè i dati sono importanti?

R3 → DATI = informazione = conoscenza = supporto alle decisioni!



RECOMMENDATION SYSTEMS



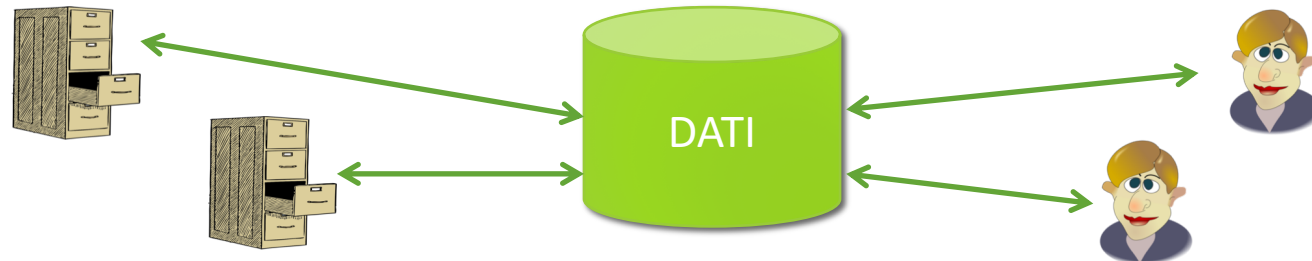
PREDICTIVE MAINTANANCE
(Industry 4.0)

Come gestire i dati?

- ✧ L'approccio alla **gestione dei dati basato su file** funziona bene nel caso di **sistemi di piccole dimensioni, prevalentemente single-user, o con scarsa necessità di condivisione dei dati.**

PROBLEMI:

- 1) Scalabilità?** → gestione di grandi moli di dati?
- 2) Concorrenza?** → accesso da parte di utenti/applicazioni?



Obiettivi della prima parte del Corso

FOCUS sui DATI

- **modelli, linguaggi, e strumenti software**

al fine di: **progettare, implementare ed analizzare collezioni (Basi di Dati (DB)** che siano:

- *Persistenti*
- Potenzialmente *grandi*
- Potenzialmente *condivise* (tra utenti/applicazioni)

Contenuti della prima parte del corso

Forniremo un'introduzione al mondo delle basi di dati (DB) da **due punti di vista:**

- ❑ **Utente** → come interagire con un DB
(aggiungere/modificare informazioni, recuperare informazioni, etc).
- ❑ **Progettista** → come progettare un DB.

Contenuti della prima parte del corso

Forniremo un'introduzione al mondo delle basi di dati (DB) da **due punti di vista:**

- Utente** → come interagire con un DB
(aggiungere/modificare informazioni, recuperare informazioni, etc).
- Progettista** → come progettare un DB.

Contenuti della prima parte del corso

- Esistono **appositi strumenti software**, sviluppati fin dagli anni '60, per la gestione delle basi di dati.

Database Management System (DBMS) → sistema software che è in grado di gestire BD *grandi, condivise e persistenti, in maniera efficiente e sicura.*

- Da qui in avanti, per noi una **base di dati è una collezione di dati gestita da un DBMS ...**

Dati vs Informazione

Il concetto di *informazione*

- Che cos'è l'informazione?
- Che cos'è la conoscenza?

Il concetto di *informazione*

- Che cos'è l'informazione?
 - “questa pianta è buona da mangiare”
 - “il fuoco si accende in questo modo”
- L'informazione si caratterizza come una scelta tra diverse alternative
 - Viene fornita informazione quando si effettua una scelta
 - Maggiore è il numero di possibili scelte, maggiore sarà l'informazione fornita

Il concetto di *informazione*

- Definizione Treccani
 - “Notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere.”
- Che cos'è la conoscenza?

Il concetto di *informazione*

- Che cos'è la conoscenza?
 - La conoscenza è fatta di informazioni che vengono apprese, conservate e trasmesse

Rappresentazione delle informazioni

- Al fine di poter conservare e trasmettere l'informazione è necessario **rappresentarla** in qualche modo
 - Associare l'informazione ad un disegno (i graffiti preistorici)
 - Associare l'informazione a simboli

Rappresentazione delle informazioni

- In informatica le informazioni vengono rappresentate tramite dati:
 - Un dato necessita di essere interpretato

NOTA: Senza interpretazione, il dato non e' molto utile!

• DATO:



• Dato+contesto=informazione

Le unità di misura dell'informatica

(bit e byte)

1 bit (cifra che può assumere solo due valori, 0/1)

8 bit = 1 Byte = 1 carattere

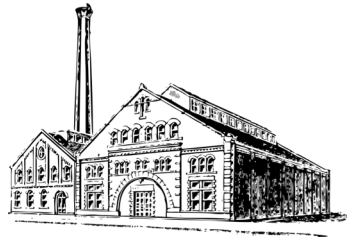
- L'aggregazione degli 8 bit necessari per definire un carattere alfanumerico viene definita byte e rappresenta l'unità pratica principale in informatica, in quanto permette di esprimere un singolo carattere alfanumerico.
- I multipli del byte, espressi con 2 (il numero delle cifre nel sistema di numerazione binaria) elevato alle potenze di 10, sono:
 - **2^{10} byte = 1024 byte = 1 Kilobyte = 1 KB**
 - **2^{20} byte = 1024 Kbyte = 1 Megabyte = 1 MB**
 - **2^{30} byte = 1024 Mbyte = 1 Gigabyte = 1 GB**
 - **2^{40} byte = 1024 Gbyte = 1 Terabyte = 1 TB**

Sistemi informativi

Sistemi Informativi

Un **Sistema Informativo (SI)** è una componente di un'organizzazione il cui scopo è quello di **gestire le informazioni** utili ad i fini dell'organizzazione stessa.

Organizzazione → Azienda, Ufficio, Ente, Università, etc

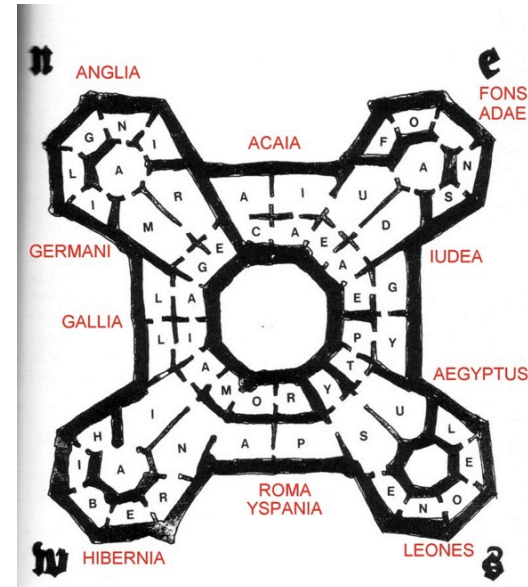


Sistemi Informativi

L'esistenza di un **Sistema Informativo** è indipendente dalla sua automatizzazione.



Biblioteca reale di Alessandria d'Egitto
IV-I secolo a.C
400000 rotoli presenti.



Sistemi Informativi

L'esistenza di un **Sistema Informativo** è indipendente dalla sua automatizzazione.

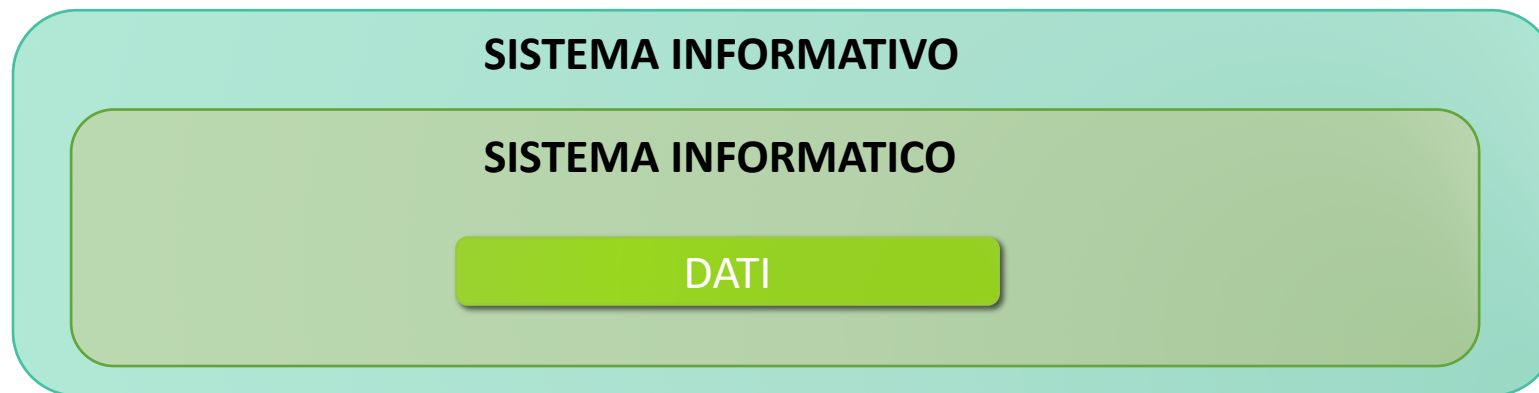
Censimenti e Registro Anagrafe.



Nell'Antica Roma, i **censimenti** venivano effettuati dalla fine del IV secolo a.c. Gli **elenchi dei censiti**, distinti secondo il possesso o meno dei diritti civili e politici, la classe patrimoniale e l'età, venivano utilizzati come liste elettorali e per determinare la ruoli per l'esenzione dei tributi e le liste di leva.

Sistemi Informativi

- La porzione **automatizzata** di un sistema informativo prende il nome di **Sistema Informatico**.
- All'interno di un sistema informatico, le informazioni sono rappresentate da **dati** ...



Approcci di gestione dei dati

Gran parte dei sistemi informatici hanno necessità di gestire **dati in maniera persistente**.



Persistente → Dati memorizzati su memoria secondaria

APPROCCI di GESTIONE

- Approccio **convenzionale** (basato su files)
- Approccio **strutturato** (basato su software di gestione dei dati)

Approcci di gestione dei dati

Gran parte dei sistemi informatici hanno necessità di gestire **dati in maniera persistente**.



Persistente → Dati memorizzati su memoria secondaria

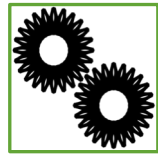
APPROCCI di GESTIONE

- **Approccio convenzionale (basato su file)**
- **Approccio strutturato (basato su software di gestione dei dati)**

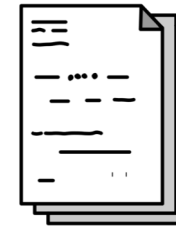
Approccio basato su files

➤ Approccio **Convenzionale** (basato su **files**)

APPLICAZIONE



FILES



Operazioni di **Letture/Scrittura**
su file mediante supporto del
Sistema Operativo

- Nessuna chiara distinzione tra **dati ed applicazioni**.
- L'applicazione contiene al suo interno la **logica di gestione e memorizzazione** dei dati stessi (es. formato dei dati).
- Il **Sistema Operativo** offre le primitive di base per l'accesso ai files ed i meccanismi di sicurezza del **file-system**.

Approccio basato su files

➤ Approccio **Convenzionale** (basato su **files**)

PROBLEMA1: Gestione di **grandi quantità di dati**?

Qualche esempio “estremo”:



AMAZON

59 Milioni di clienti iscritti
Oltre 42 Terabyte di dati



AT&T

323 Terabyte di dati
1.9 trillioni di record relative a chiamate

Ovvi problemi di **scalabilità** ed **efficienza** ...

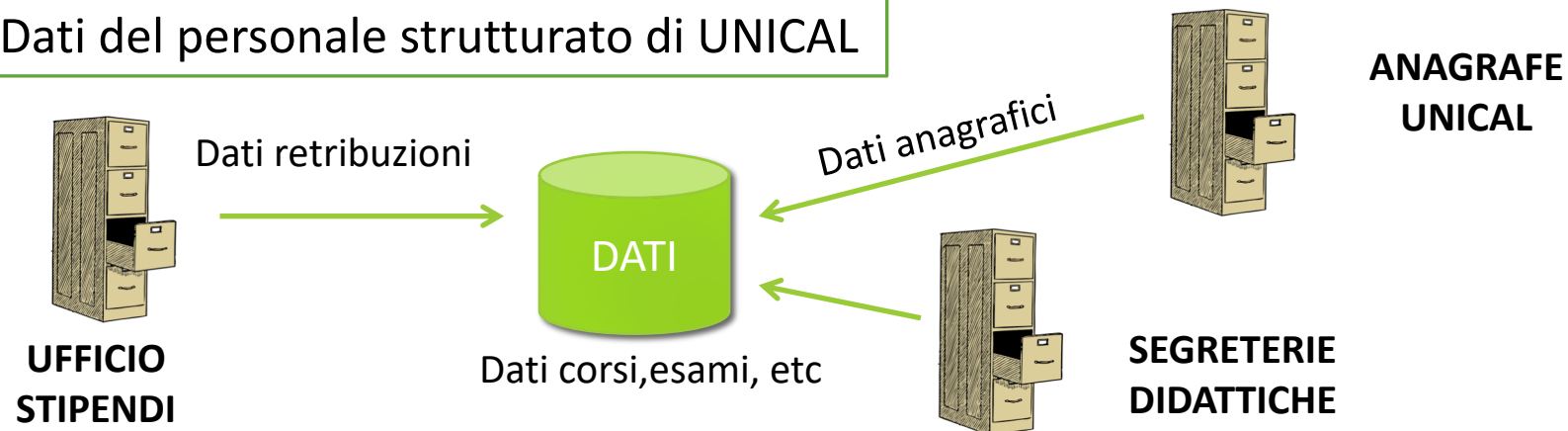
Approccio basato su files

➤ Approccio **Convenzionale** (basato su **files**)

PROBLEMA2: **Condivisione** ed **accesso** concorrente?

In molti scenari pratici, i dati devono essere a disposizione di una moltitudine di utenti/applicazioni per accessi concorrenti.

Es. Dati del personale strutturato di UNICAL



Approccio basato su files

➤ Approccio **Convenzionale** (basato su **files**)

PROBLEMA2: **Condivisione** ed **accesso** concorrente?

LIMITAZIONI

- Accesso a file condivisi avviene attraverso le politiche di accesso del file-system → Lock a livello di file, **bassa granularità di concorrenza**, prestazioni limitate!
- Applicazioni diverse devono conoscere l'esatta collocazione e formato dei dati → **Aggiornamento del formato dei dati?**
- In alternativa: replica dei dati presso i vari sistemi/utenti che ne fanno utilizzo → **Consistenza delle repliche?**

Approcci di gestione dei dati

Gran parte dei sistemi informatici hanno necessità di gestire **dati in maniera persistente**.



Persistente → Dati memorizzati su memoria secondaria

APPROCCI di GESTIONE

- Approccio **convenzionale** (basato su files)
- Approccio **strutturato** (basato su software di gestione dei dati)

Caratteristiche dei DBMS

Un **DBMS** è un sistema software che è in grado di gestire collezioni di dati *grandi, condivise e persistenti, in maniera efficiente e sicura.*

(ALCUNE) FUNZIONALITA':

- Creazione di una base di dati e memorizzazione su memoria secondaria
- Accesso in lettura/scrittura ad i dati
- Condivisione di dati tra diversi utenti/applicazioni
- Protezione dei dati da accessi non autorizzati
- Reliability dei dati in caso di guasti (hardware/software)
- ...

Caratteristiche dei DBMS

Un **DBMS** è un sistema software che è in grado di gestire collezioni di dati *grandi, condivise e persistenti, in maniera efficiente e sicura.*

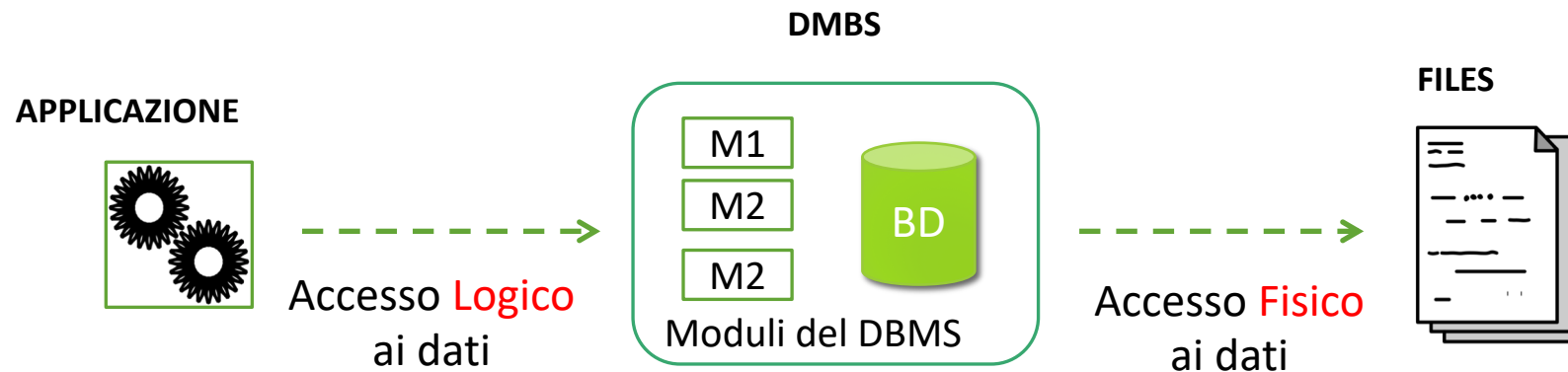
(ALCUNE) FUNZIONALITA':

- Creazione di una base di dati e memorizzazione su memoria

Da qui in avanti nel corso, una base di dati è una collezione di dati gestita da un DBMS!

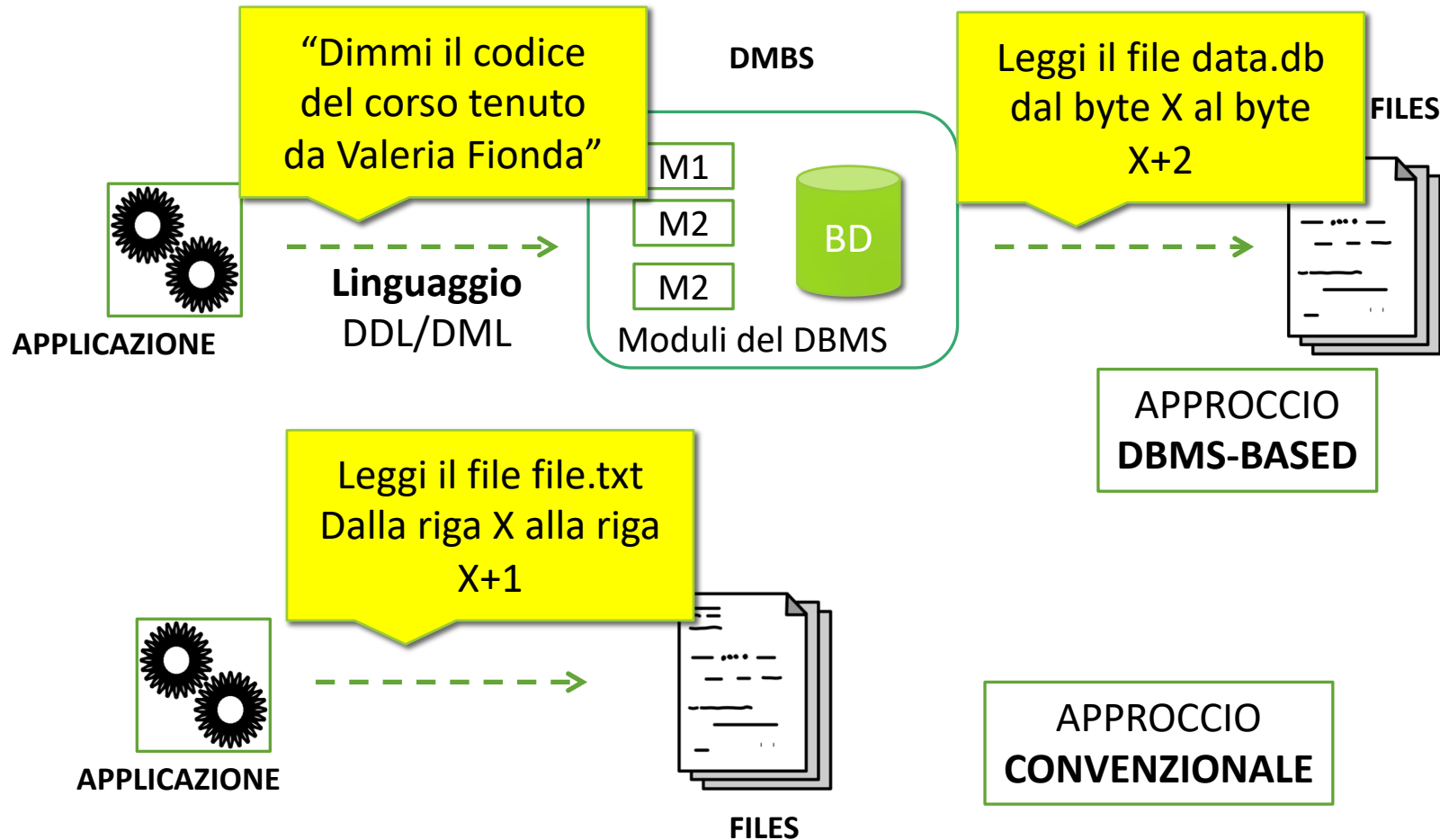
- Protezione dei dati da accessi non autorizzati
- Reliability dei dati in caso di guasti (hardware/software)
- ...

Separazione Dati/Applicazioni con DBMS



- Tramite i DBMS, è possibile implementare un **paradigma di separazione di dati ed applicazioni ...**
- Le applicazioni non necessitano di conoscere la struttura fisica dei dati (es. come e dove sono memorizzati su disco) ma **solo la struttura logica** (cosa rappresentano).

Separazione Dati/Applicazioni con DBMS



DBMS: Quali?

4th Dimension

Adabas D

Alpha Five

Apache Derby

Aster Data

Altibase

BlackRay

CA-Datcom

Clarion

Clustrix

CSQL

CUBRID

Daffodil database

DataEase

Database Management

Dataphor

Java DB

Empress Embedded

Database

EnterpriseDB

eXtremeDB

FileMaker Pro

Firebird

Greenplum

GroveSite

H2

Helix database

HSQLDB

IBM DB2

IBM Lotus Approach

Infobright

Informix

Ingres

InterBase

InterSystems Caché

GT.M

Linter

MariaDB

MaxDB

Microsoft Access

Microsoft Jet Database

Engine (part of Microsoft

Access)

Microsoft SQL Server

Lista (parziale) dei DBMS più utilizzati.

DBMS: Quali?

Microsoft SQL Server Express

Microsoft Visual FoxPro

Mimer SQL

MonetDB

mSQL

MySQL

Netezza

NonStop SQL

Openbase

OpenLink Virtuoso

OpenLink Virtuoso Server

OpenOffice.org Base

Oracle

Oracle Rdb for OpenVMS

Panorama

PostgreSQL

Progress Software

RDM Server

SAND CDBMS

Sav Zigzag

ScimoreDB

SmallSQL

SQLBase

SQLite

Sybase SQL AdvantageServer

Teradata

TimesTen

txtSQL

Unisys RDMS 2200

UniData

UniVerse

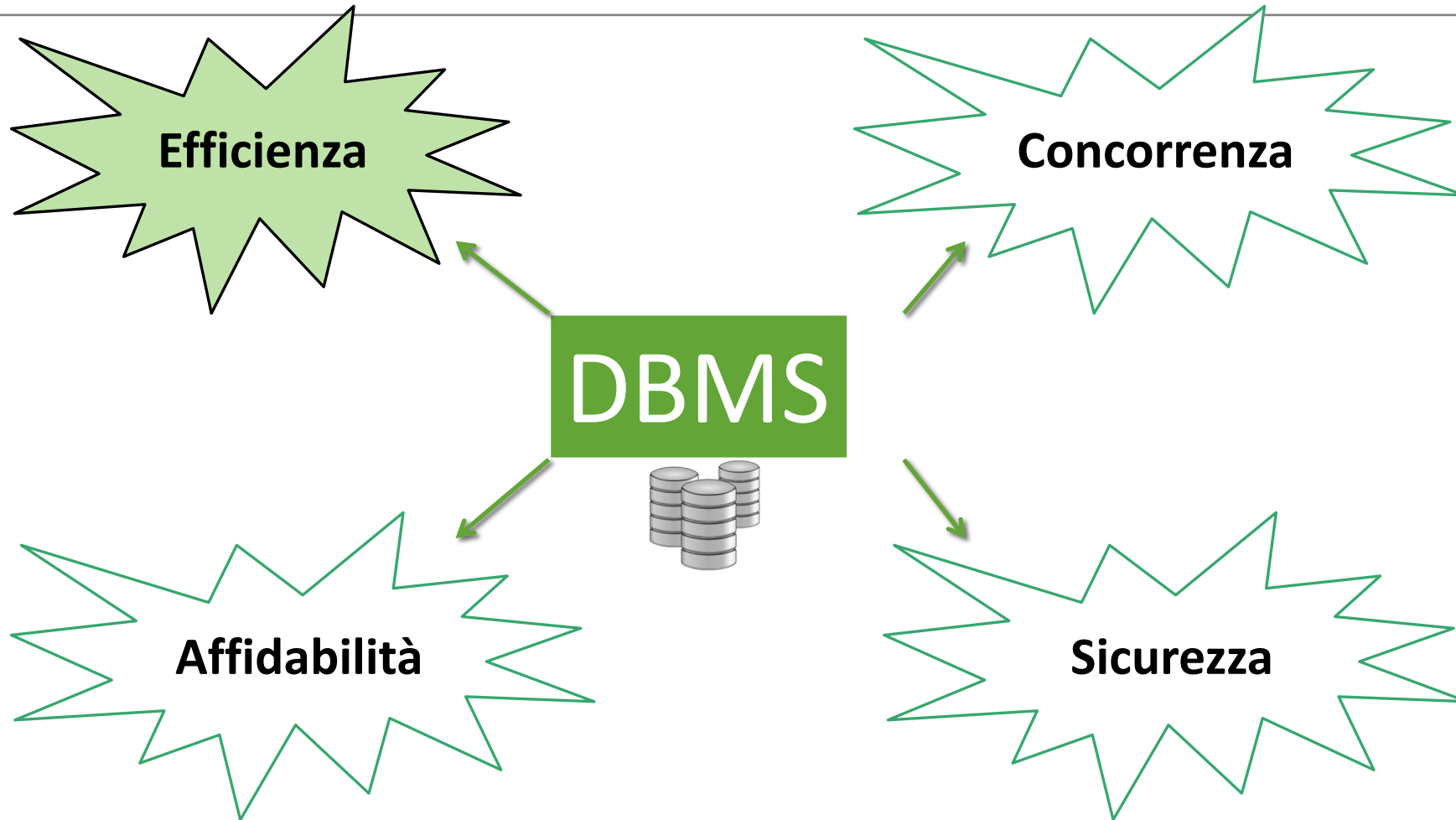
Vertica

VMDS

VISTADB

Lista (parziale) dei DBMS
più utilizzati.

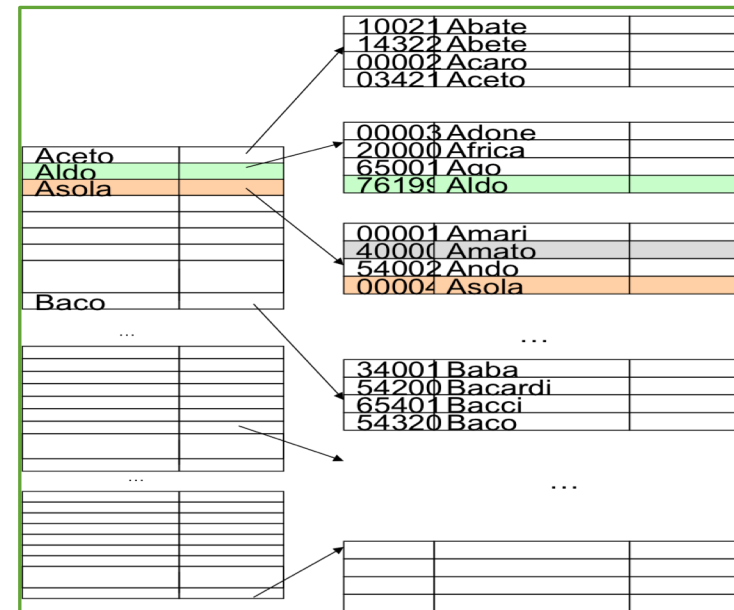
Componenti di un DBMS



Caratteristiche di un DBMS: **Efficienza**

Efficienza di un DBMS nella gestione dei dati ...

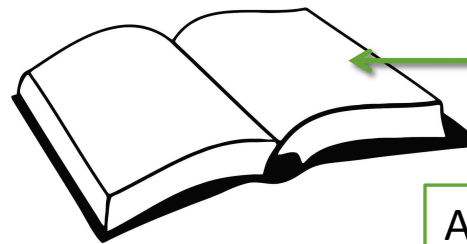
- DBMS forniscono adeguate **strutture dati** per organizzare i dati all'interno dei file, e per supportare le operazioni di ricerca/aggiornamento.
- In genere, parliamo di **strutture dati ad albero** o **tabelle hash**.



Caratteristiche di un DBMS: **Efficienza**

Indice → struttura che contiene **informazioni sulla posizione di memorizzazione delle tuple** sulla base del valore del campo **chiave**.

Q. A che serve un indice?

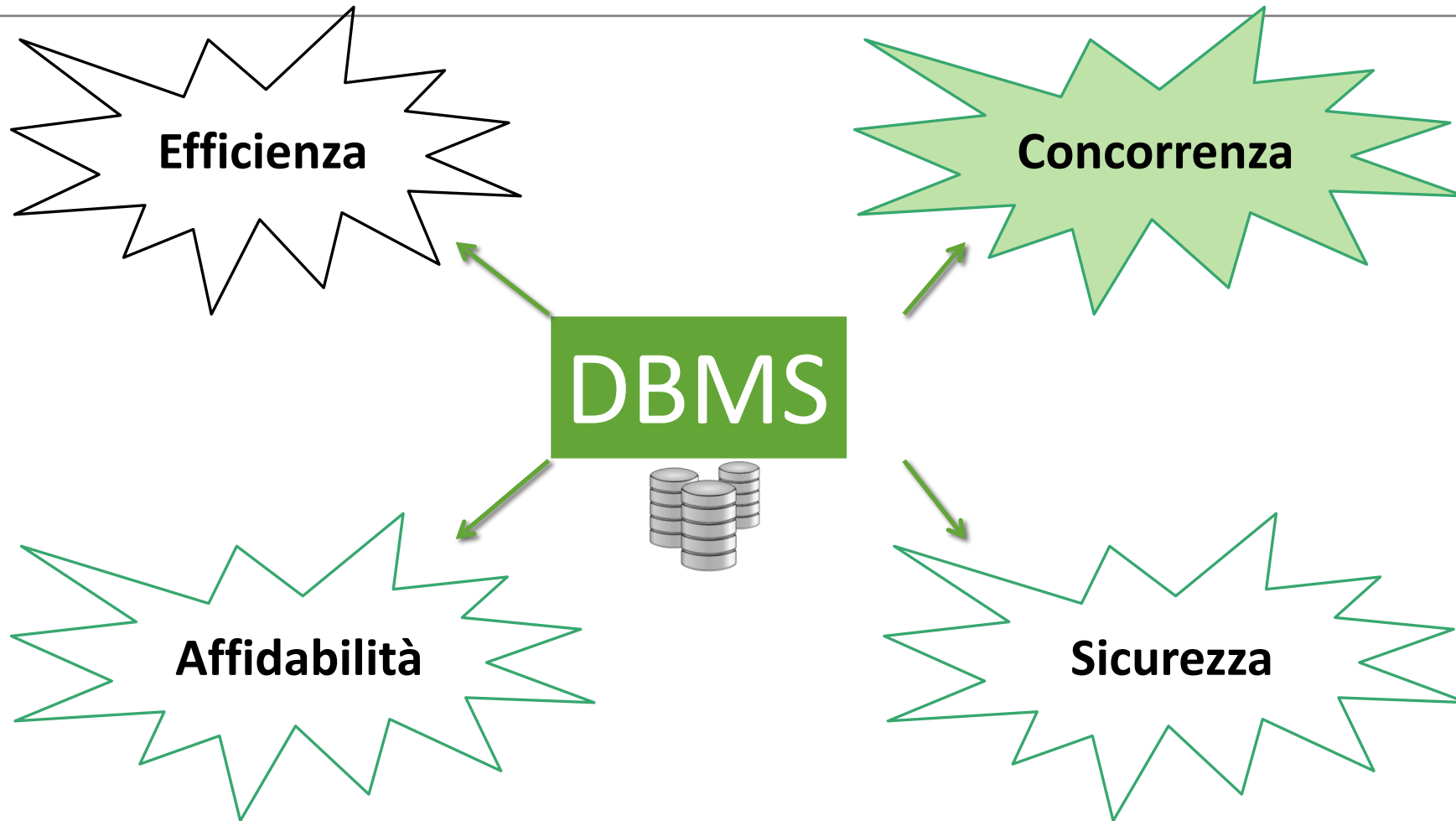


ACCESSO DIRETTO

Indice

Introduzione	1
Capitolo 1	20
Capitolo 2	40
Capitolo 3	60
Conclusioni	65

Componenti di un DBMS



Caratteristiche di un DBMS: **Gestione della Concorrenza**

- In molti sistemi è fondamentale **gestire operazioni concorrenti** di accesso ai dati ...



Processing di più di 7.7 milioni di pagamenti ogni giorno

La maggior parte dei DBMS forniscono un **livello di granularità di locking** più fine di quello convenzionale (a livello di tabella, pagina, o singola entry).

Caratteristiche di un DBMS: **Gestione della Concorrenza**

- Al tempo stesso, un DBMS deve garantire il fatto che **accessi da parte di applicazioni diverse non interferiscano tra loro**, lasciando il sistema in uno **stato inconsistente** ...

Es. Sistema informativo dei conti bancari

- 2 richieste da gestire al tempo t:
 - Prelievo di 100 euro dal conto X
 - Prelievo di 80 euro dal conto X
- Saldo del conto X at tempo t: 120 euro

Caratteristiche di un DBMS: Gestione della Concorrenza

ESEMPIO di ESECUZIONE (non corretta!!)

OP1

Leggi X
Calcola X-100
Scrivi X

OP2

Leggi X
Calcola X-80
Scrivi X

Schedule:

OP1: Leggi X
OP2: Leggi X
OP1: Calcola X-100
OP2: Calcola X-80
OP1: Scrivi X
OP2: Scrivi X

Valore X:

120
120
120
120
20
40 (????)

Per prevenire tali situazioni, i DBMS implementano **algoritmi di controllo della concorrenza** che **operazioni sui dati (transazioni)** eseguite in concorrenza producano lo stesso risultato di un'esecuzione seriale.

Caratteristiche di un DBMS: Gestione della Concorrenza

Lock Manager → componente del DBMS responsabile di gestire i lock alle risorse del DB, e di rispondere alle richieste delle transazioni.

OP1

Lock(x)

Leggi X

Calcola X-100

Scrivi X

Unlock(x)

OP2

Lock(x)

Leggi X

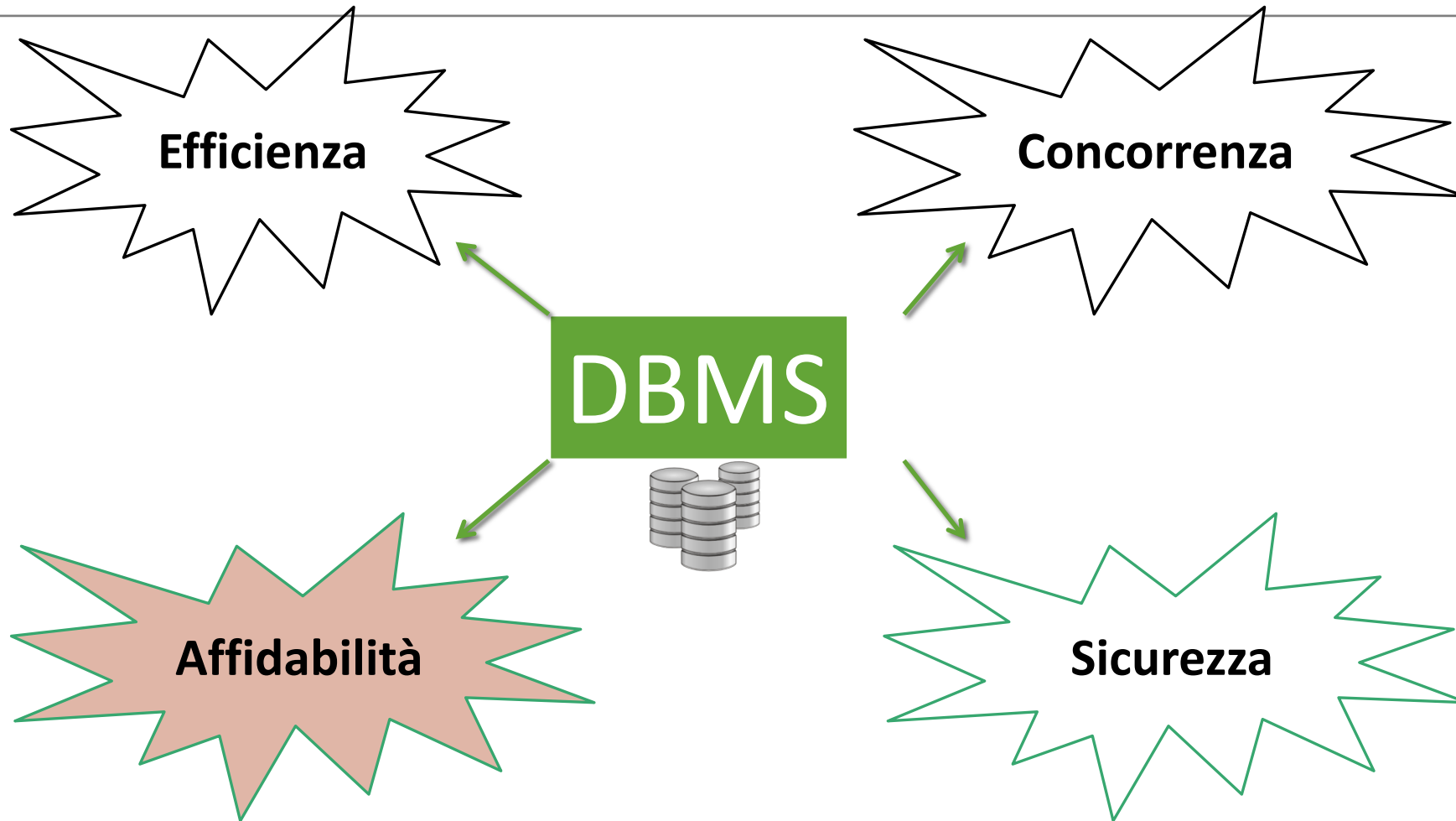
Calcola X-80

Scrivi X

Unlock(x)

➤ Utilizzo di **lock** in lettura/scrittura per accesso a risorse condivise (dati).

Componenti di un DBMS



Caratteristiche di un DBMS: Meccanismi di Roll-back

- Alcune operazioni sui dati sono **particolarmente delicate**, e devono essere gestite in maniera opportuna, secondo la regola del **tutto o niente**.

Es. Trasferimento di denaro (100\$) dal conto X al conto Y.

Op1: $X = X - 100$

Op2: $Y = Y + 100$

CRASH!



Caratteristiche di un DBMS: Meccanismi di **Roll-back**

- Alcune operazioni sui dati sono **particolarmente delicate**, e devono essere gestite in maniera opportuna, secondo la regola del **tutto o niente**.
- Per questo, i DBMS devono fornire appositi strumenti per **annullare operazioni non completate** e fare **roll-back** dello stato del sistema ...



Caratteristiche di un DBMS: **Persistenza ed Affidabilità**

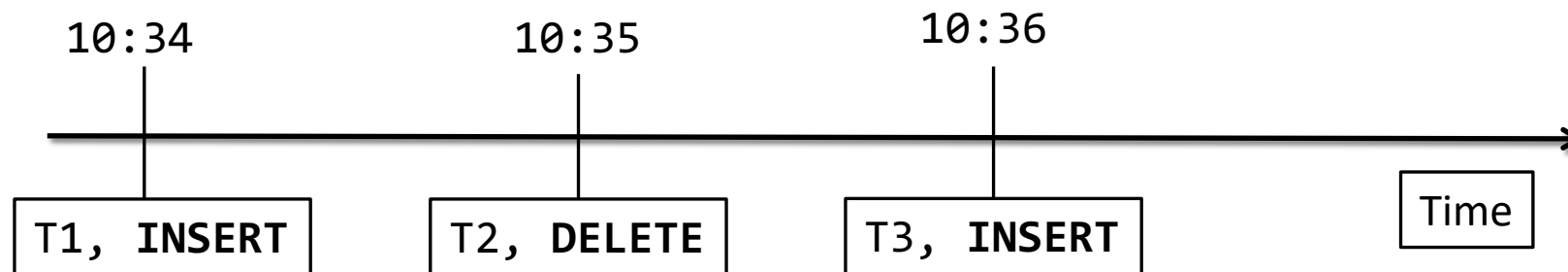
In molti casi i DBMS mettono a disposizione appositi **strumenti ed algoritmi per garantire la persistenza dei dati** anche in presenza di **malfunzionamenti hardware/software**.

Il controllore di affidabilità utilizza dei **log**, nel quale sono indicate tutte le **operazioni svolte dal DBMS**.

Caratteristiche di un DBMS: **Persistenza ed Affidabilità**

Il controllore di affidabilità utilizza un **log**, nel quale sono indicate tutte le **operazioni svolte dal DBMS**.

LOG, struttura logica



Tramite il log, e' possibile fare **do/undo** delle operazioni...

Caratteristiche di un DBMS: **Persistenza ed Affidabilità**

Il controllore di affidabilità utilizza un **log**, nel quale sono indicate tutte le **operazioni svolte dal DBMS**.

LOG, struttura logica

Q. Dove/come memorizzare il file di log?



Tramite il log, e' possibile fare **do/undo** delle operazioni...

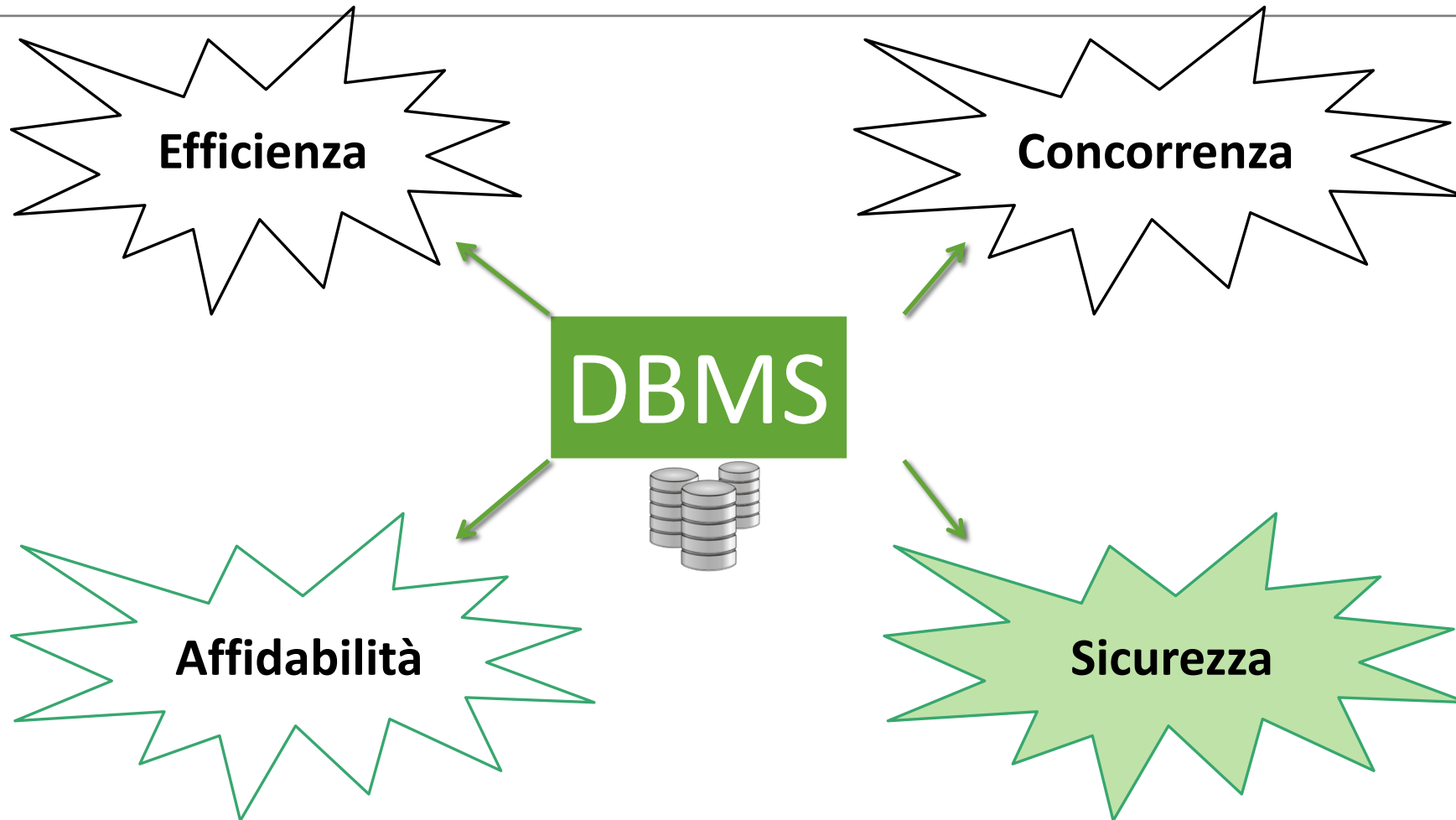
Caratteristiche di un DBMS: **Persistenza ed Affidabilità**

Transazione → insieme di operazioni (lettura/scrittura) eseguite su un DB dal DBMS.

Proprietà **ACID** di un sistema transazionale

- **Atomicita'** → La transazione deve essere eseguita con la regola del “**tutto o niente**”.
- **Consistenza** → La transazione deve lasciare il DB in uno stato **consistente**, vincoli di integrità sui dati non devono essere violati.
- **Isolamento** → L'esecuzione di una transazione deve essere **indipendente** dalle altre.
- **Persistenza** → L'effetto di una transazione conclusa con successo non deve essere perso.

Componenti di un DBMS



Caratteristiche di un DBMS: Multi-utenza e Sicurezza

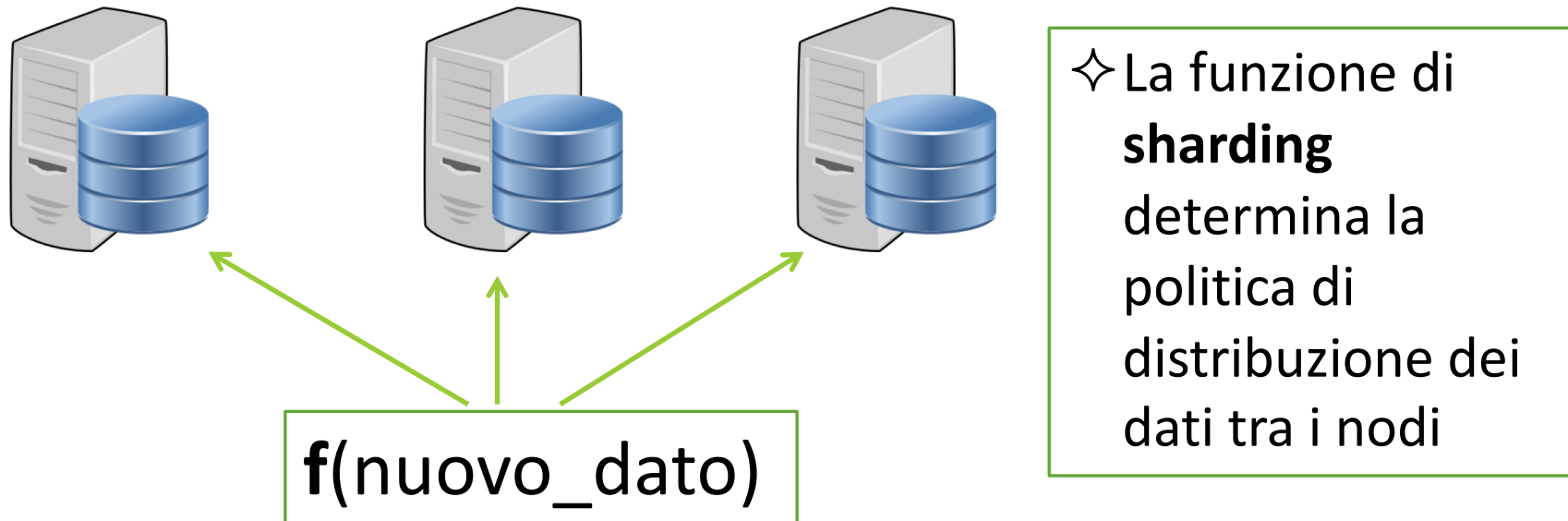
La maggior parte dei DBMS implementa **politiche di controllo degli accessi** ad i dati mediante **sistemi di permessi**:

- Quali **operazioni sono consentite** all'utente X?
- Quali **dati appartengono** all'utente X?

Utente	OPERAZIONE	DATO	PERMESSO
Utente X	Lettura	Stipendio di X	Consentito
Utente X	Lettura	Stipendio di Y	Consentito
Utente Y	Scrittura	Stipendio di Y	Negato

Caratteristiche di un DBMS: **Scalabilità** **(orizzontale)**

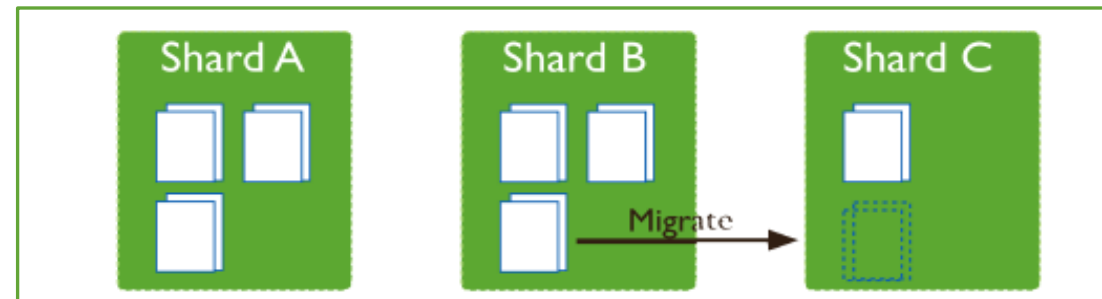
Possibilità di gestire grandi moli di dati aumentando il **numero di nodi** usati per lo storage (**database distribuito**).



Caratteristiche di un DBMS: **Scalabilità** **(orizzontale)**

Ulteriori funzionalità di un database distribuito.

1. Meccanismi di **Load-balancing**
2. Meccanismi di **Gestione delle repliche dati**



PROBLEMA. Come gestire la **consistenza** delle repliche dati in presenza di partizionamenti della rete e perdita di messaggi?