

**UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA**

Dipartimento di Ingegneria «Enzo Ferrari»

Corso di laurea in Ingegneria Informatica

# Formato SIARD

Un nuovo modo di preservare database

Candidato: Davide Bucciarelli  
Relatore: Prof. Sonia Bergamaschi  
Correlatore: PhD. Luca Gagliardelli

**Anno Accademico 2021/2022**



# Introduzione

---

Lo scopo di librerie, archivi ed altre simili istituzioni è quello di preservare documenti in modo tale che essi siano accessibili dalle future generazioni come sono disponibili a noi oggi. La conservazione di antichi libri in forma cartacea richiede tecniche estremamente costose per limitare il loro degrado.

Da questo punto di vista, le informazioni digitali hanno proprietà ideali: flussi di bit possono facilmente essere conservati senza che vi sia alcuna perdita di informazioni.

Quindi, qual è il problema?

# La preservazione di dati digitali

Fra cinquant'anni, esisterà ancora un dispositivo capace di leggere le informazioni conservate digitalmente su un supporto realizzato al giorno d'oggi, e presentarlo in un formato comprensibile ad un umano?







# Problemi specifici della conservazione di database

- Dati provenienti da fonti diverse
- Trasferimento dei record
- Velocità di accesso all'archivio
- Dimensione dei dati archiviati

# Due soluzioni classiche

## **Emulazione**

- Si conserva una copia, fisica o simulata, dell'hardware originale, sul quale poi potrà essere eseguito il software da conservare o poter accedere ai dati preservati.

## **Migrazione**

- Trasformare il materiale archiviato in un nuovo formato, ogni volta che il formato di partenza non è più supportato



# La normalizzazione

---

La normalizzazione è un approccio che mira alla conservazione di dati in modo tale che il materiale archiviato rimanga invariato per sempre.

Informazione preservata in una forma "normale" : uno standard sulla base del quale, in qualunque momento si possa programmare un software di conversione per poter de-normalizzare l'oggetto e riportarlo ad una forma fruibile dall'utente finale.

Dato da preservare



Processo di normalizzazione

Dato in forma normalizzata



Processo di de-normalizzazione

Dato rigenerato

# Normalizzazione applicata ai database

Cosa serve effettivamente preservare?

# Criterio di completa preservazione

Vogliamo che il risultato di qualsiasi query SELECT sia identico su DB di partenza e DB rigenerato

## **Conservazione di metadati**

Necessario conservare per ogni tabella:

- Nome della tabella
- Nome di ogni colonna
- Tipo di ogni colonna

Ogni altra informazione può essere molto utile ma non è indispensabile

## **Conservazione di dati primari**

Il risultato di query SELECT è da considerarsi identico sulla base del confronto tramite l'operatore '=' in SQL, quindi:

- Tipo di dato del DB rigenerato può differire dall'originale, ma deve essere confrontabile
- Problemi per gli oggetti non direttamente confrontabili con tale operatore, come i BLOBS



# Il formato SIARD

Sviluppato da AFS, sulla base dei principi di normalizzazione dei database, con i seguenti requisiti:

- Produrre un singolo file
- Non conservare ciò che è inaccessibile da query SQL SELECT
- Basato su standard Open source (XML e ZIP)
- Conservazione delle chiavi primarie ed esterne

# Sperimentazione sul formato SIARD

La sperimentazione si è basata in particolare su quattro DBMS: SQL Server, MySQL, PostgreSQL e Oracle. Sono state analizzate le differenze fra:

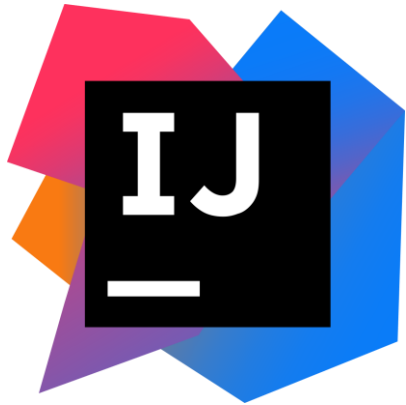
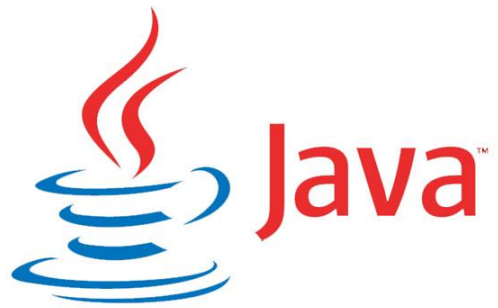
- Database originale
- Database in forma normalizzata
- Database recuperato e de-normalizzato su diversi DBMS

# Risultati della sperimentazione

Operazione	File SIARD	SQLServer	MySQL	ORACLE	PostgreSQL
Primary Keys	Mantenuto	Mantenuto	Mantenuto	Mantenuto	Mantenuto
Foreign Keys	Mantenuto	Mantenuto	Mantenuto	Mantenuto in parte	Mantenuto in parte
Vincoli NOT NULL	Mantenuto	Mantenuto	Mantenuto	Mantenuto	Mantenuto
Tipi di dato 'string'	Mantenuto	Mantenuto	Mantenuto in parte	Mantenuto in parte	Mantenuto
Tipi di dato 'numerici'	Mantenuto	Mantenuto	Mantenuto	Mantenuto	Mantenuto
Vincoli CHECK	NON mantenuto	NON mantenuto	NON mantenuto	NON mantenuto	NON mantenuto
Viste	Mantenuto in parte	NON mantenuto	NON mantenuto	NON mantenuto	NON mantenuto

Legenda:

- Mantenuto
- Mantenuto in parte
- NON mantenuto



---

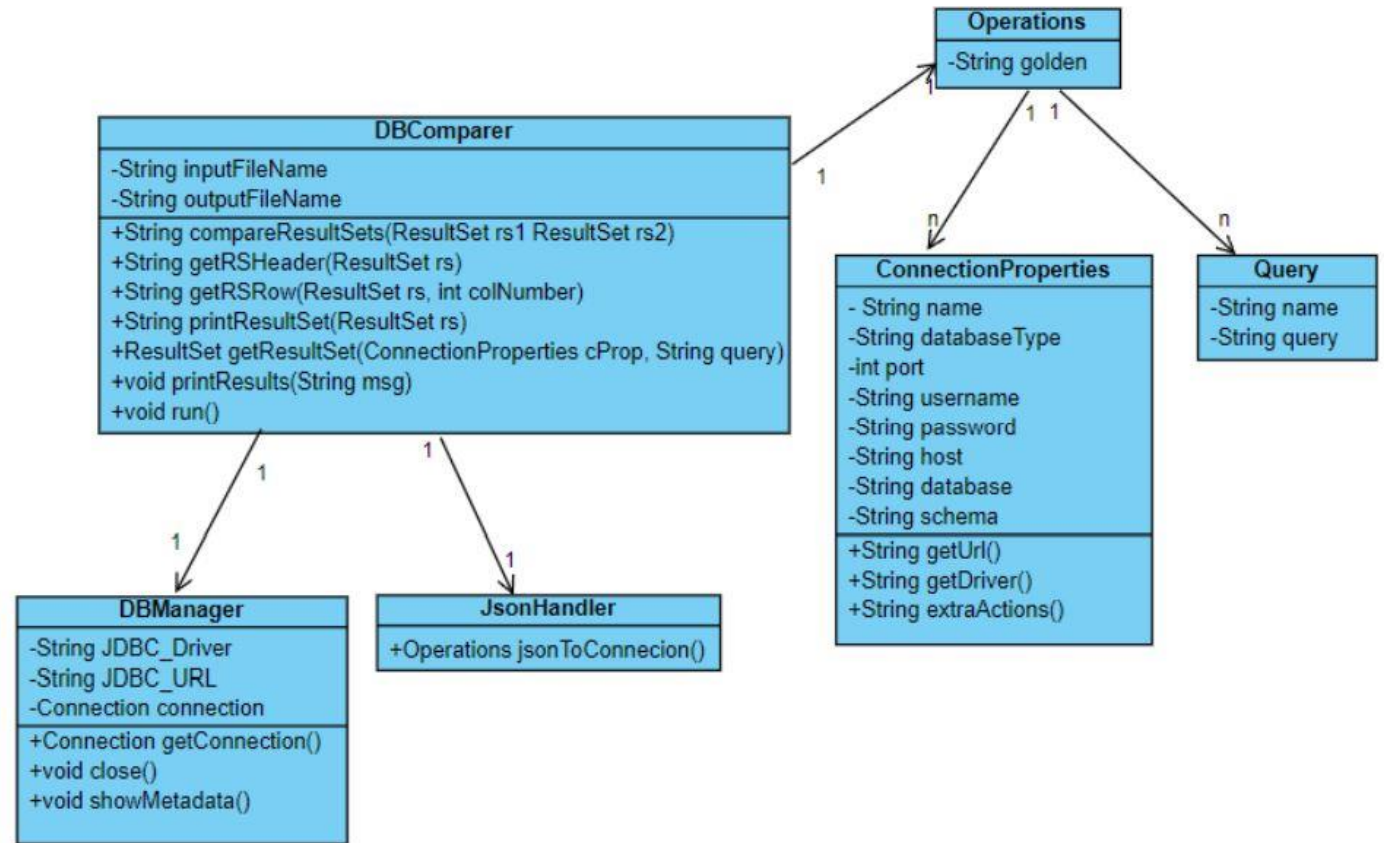
# Software DBComparer

---

Lo scopo del software è quello di avere un tool per confrontare in modo veloce due database, in particolare con l'obiettivo di verificare eventuali differenze nei risultati di specifiche query SELECT. Esso è stato creato per verificare il rispetto dei principi di normalizzazione del formato SIARD.

# DBComparer: design

Class Diagram UML per  
descrivere la struttura  
generale del software e le  
classi di cui è composto:





# Conclusioni sul formato SIARD

Visti i risultati dei test effettuati si può concludere che il sistema di normalizzazione SIARD, usato tramite il corrispettivo software di conversione SIARDGui, rispetti in toto le caratteristiche prescritte dai criteri di conservazione di database.

Sono però state riscontrate criticità per quanto riguarda il tool SIARDGui, che non risolve le incompatibilità presenti fra i DBMS

Grazie per  
l'attenzione

