

Indice

1. Introduzione

2. UML

- Versione Base
 - Diagramma dei package Client
 - Diagramma delle classi Client
 - Diagramma dei package Server
 - Diagramma delle classi Server
- Estensione
 - Diagramma dei package Client
 - Diagramma delle classi Client
 - Diagramma dei package Server
 - Diagramma delle classi Server

3. Guida Utente

- Requisiti per l'utilizzo
- Configurazione del sistema e del database
- Versione Base
- Versione Estesa

4. Casi di test

- Estensione
- Base

1.Introduzione

Il clustering è una tecnica di analisi dei dati utilizzata per raggruppare insieme elementi simili in base a determinate caratteristiche o attributi. L'obiettivo principale è quello di suddividere un insieme di dati in cluster omogenei, in modo che gli elementi all'interno dello stesso cluster siano più simili tra loro rispetto a quelli in cluster diversi. Tra i vari algoritmi di clustering, l'algoritmo K-means è uno dei più utilizzati e ben noti. K-means è un algoritmo iterativo che assegna punti dati a cluster in modo da minimizzare la somma dei quadrati delle distanze tra i punti dati e i centroidi dei cluster ai quali sono assegnati. I centroidi vengono quindi aggiornati iterativamente fino a quando non viene raggiunta una convergenza. K-means è ampiamente utilizzato in diversi campi, come il riconoscimento di modelli, l'analisi delle immagini e il marketing.

2.UML

Versione Base

Diagramma dei package Client



Diagramma delle classi Client

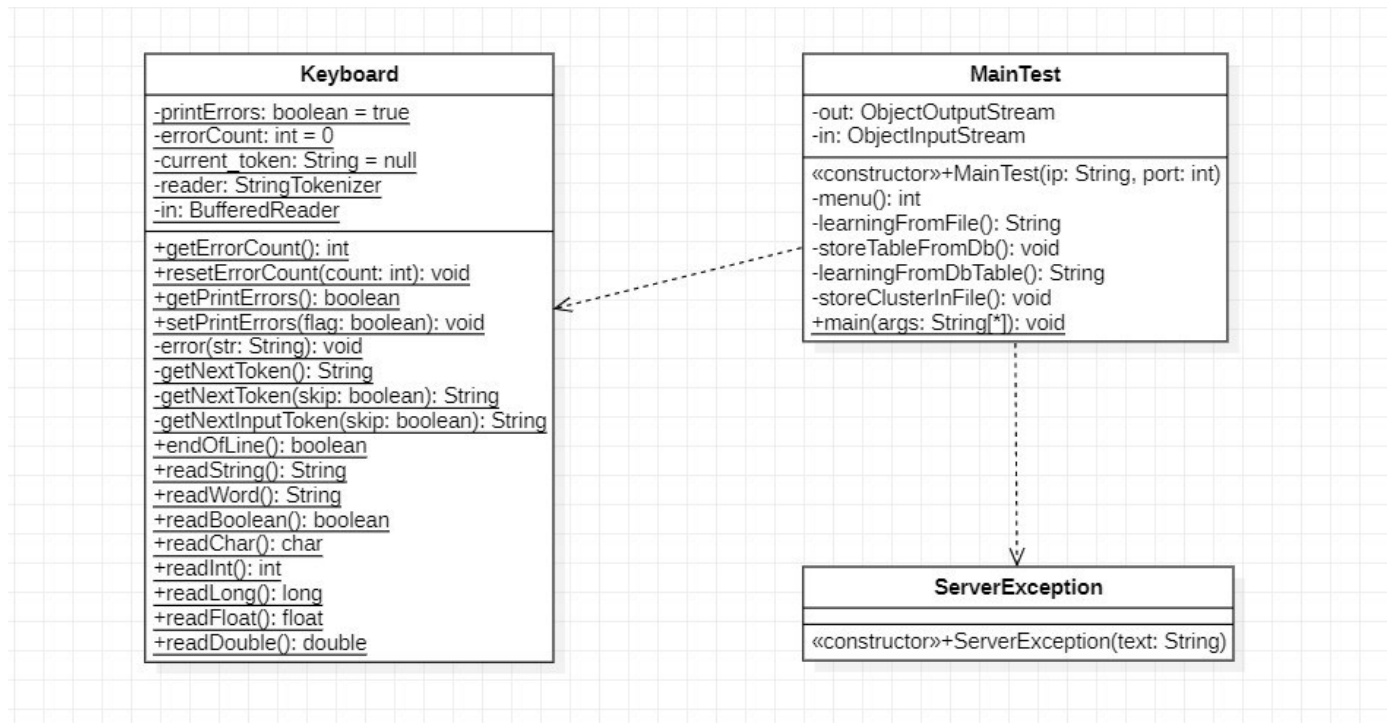


Diagramma dei package Server

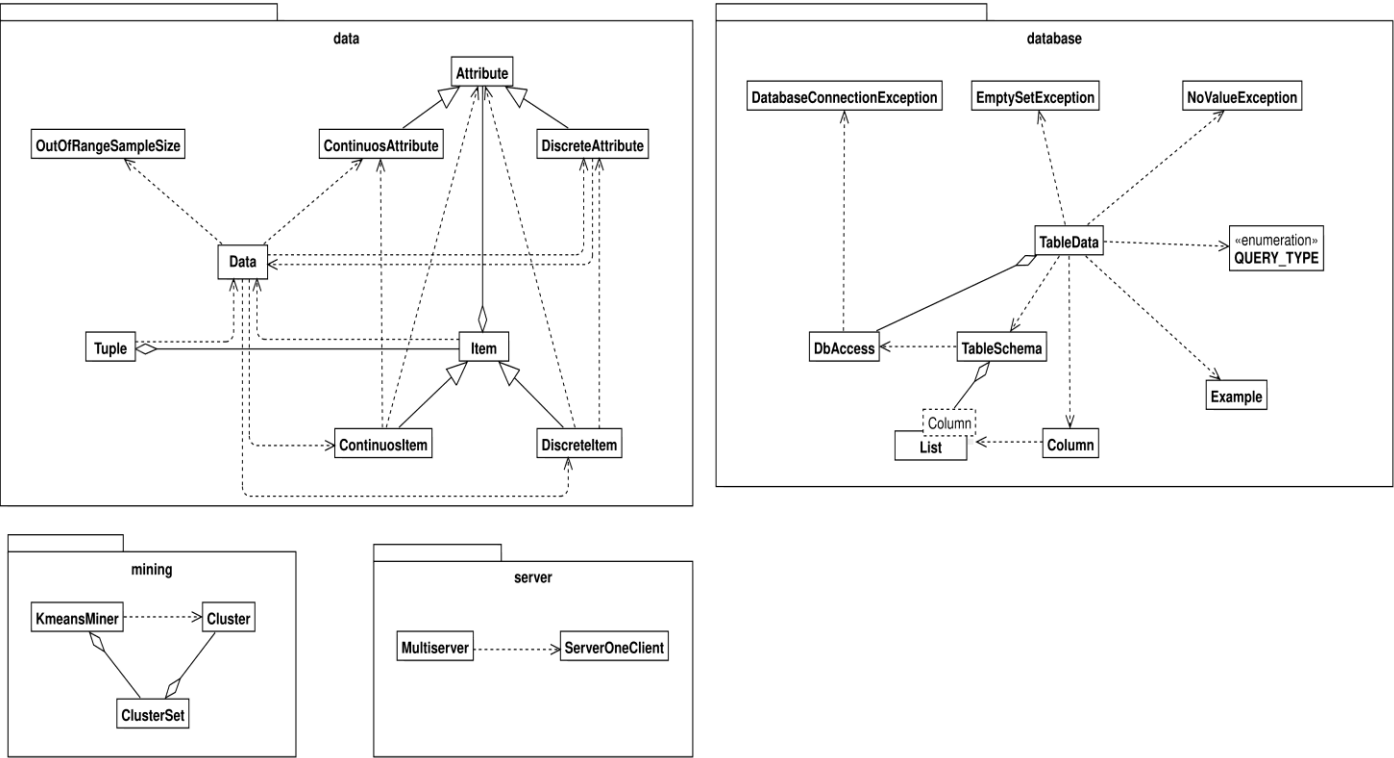
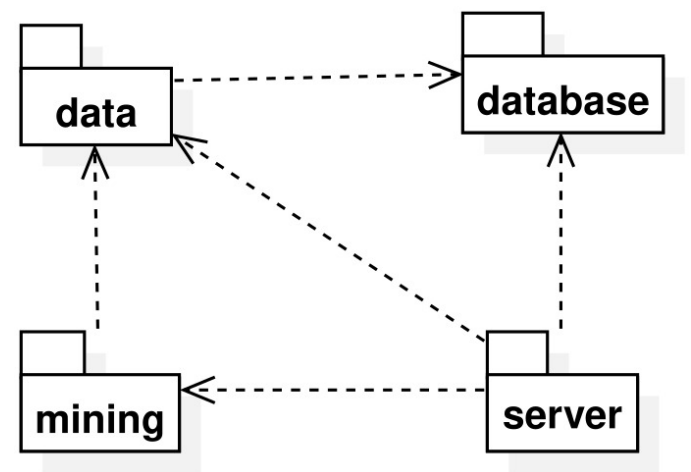
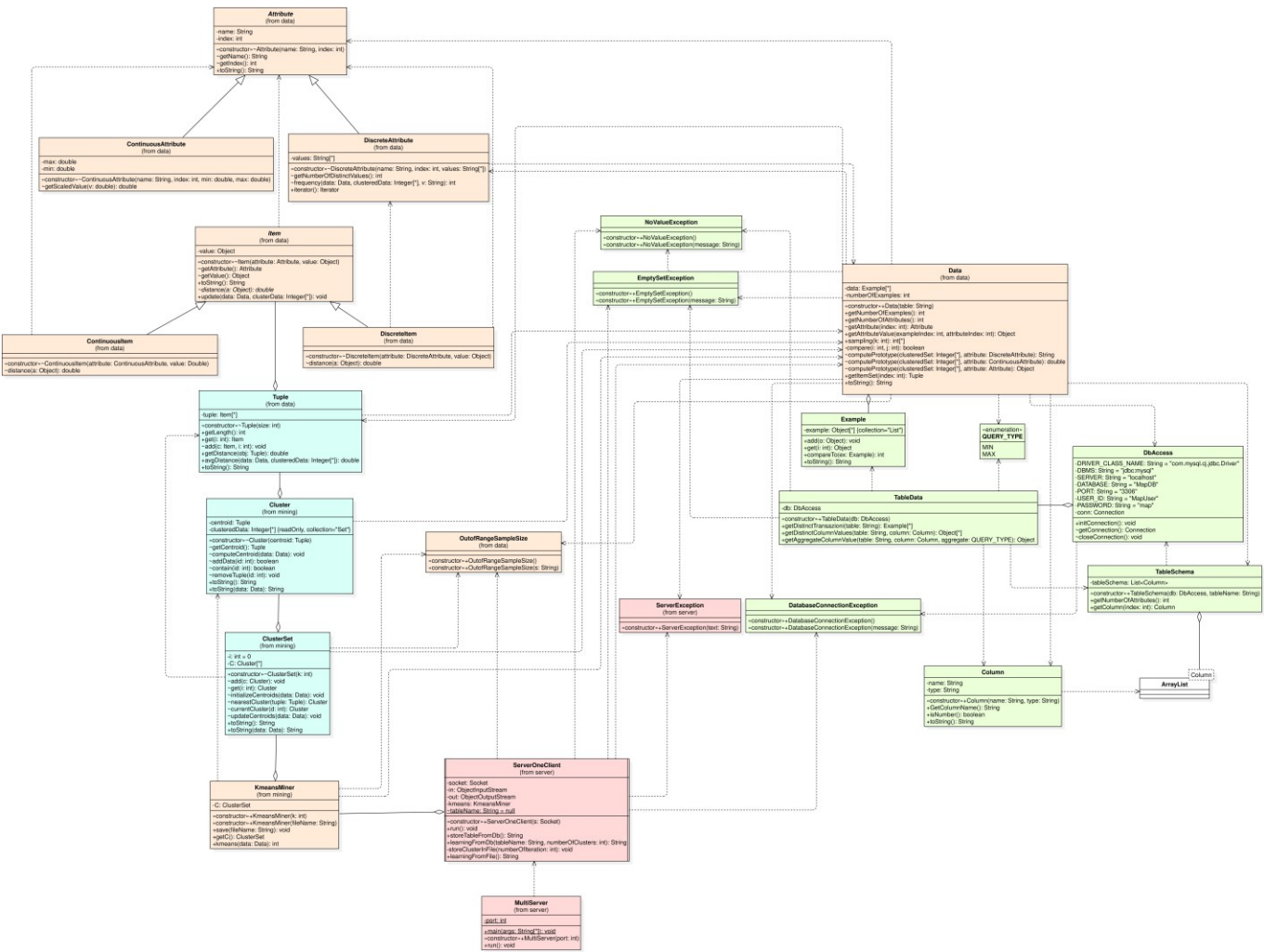


Diagramma delle classi Server



Estensione

Diagramma dei package Client

org.example.graficDemo.src

Diagramma delle classi Client

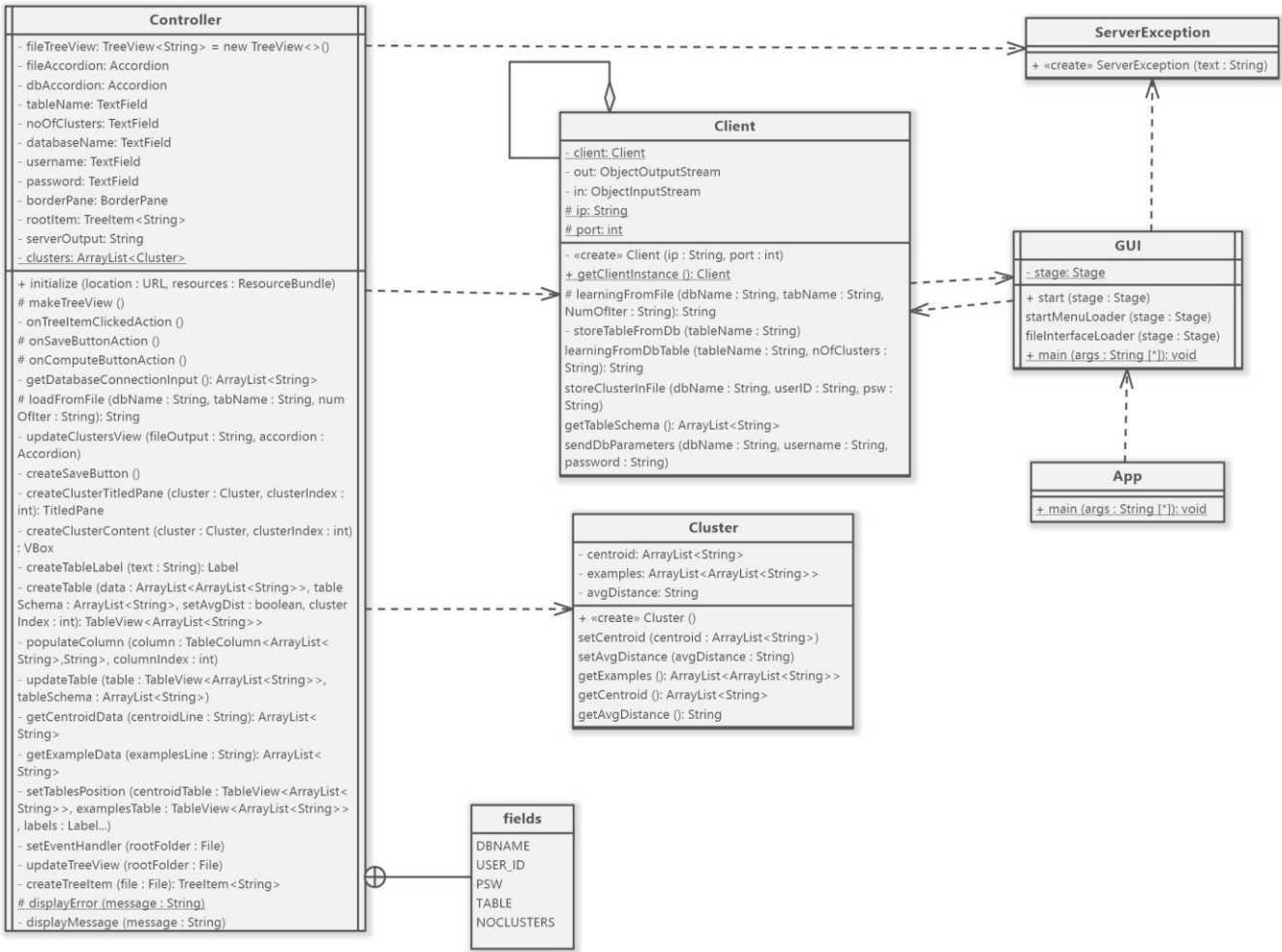
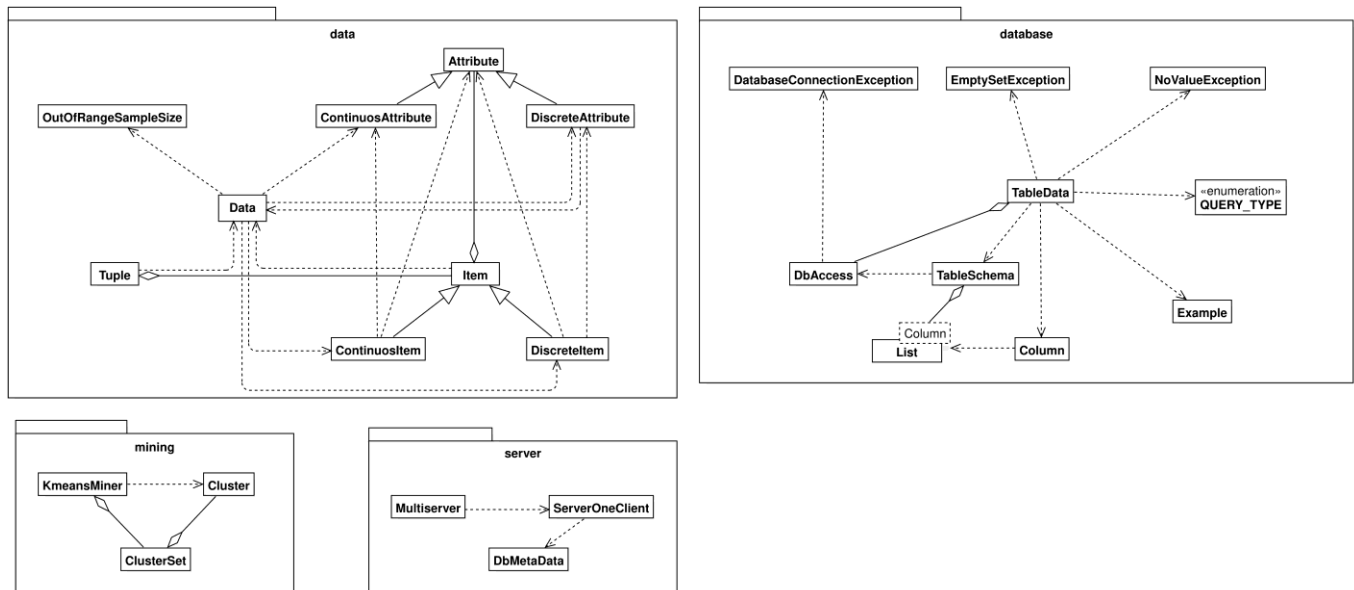
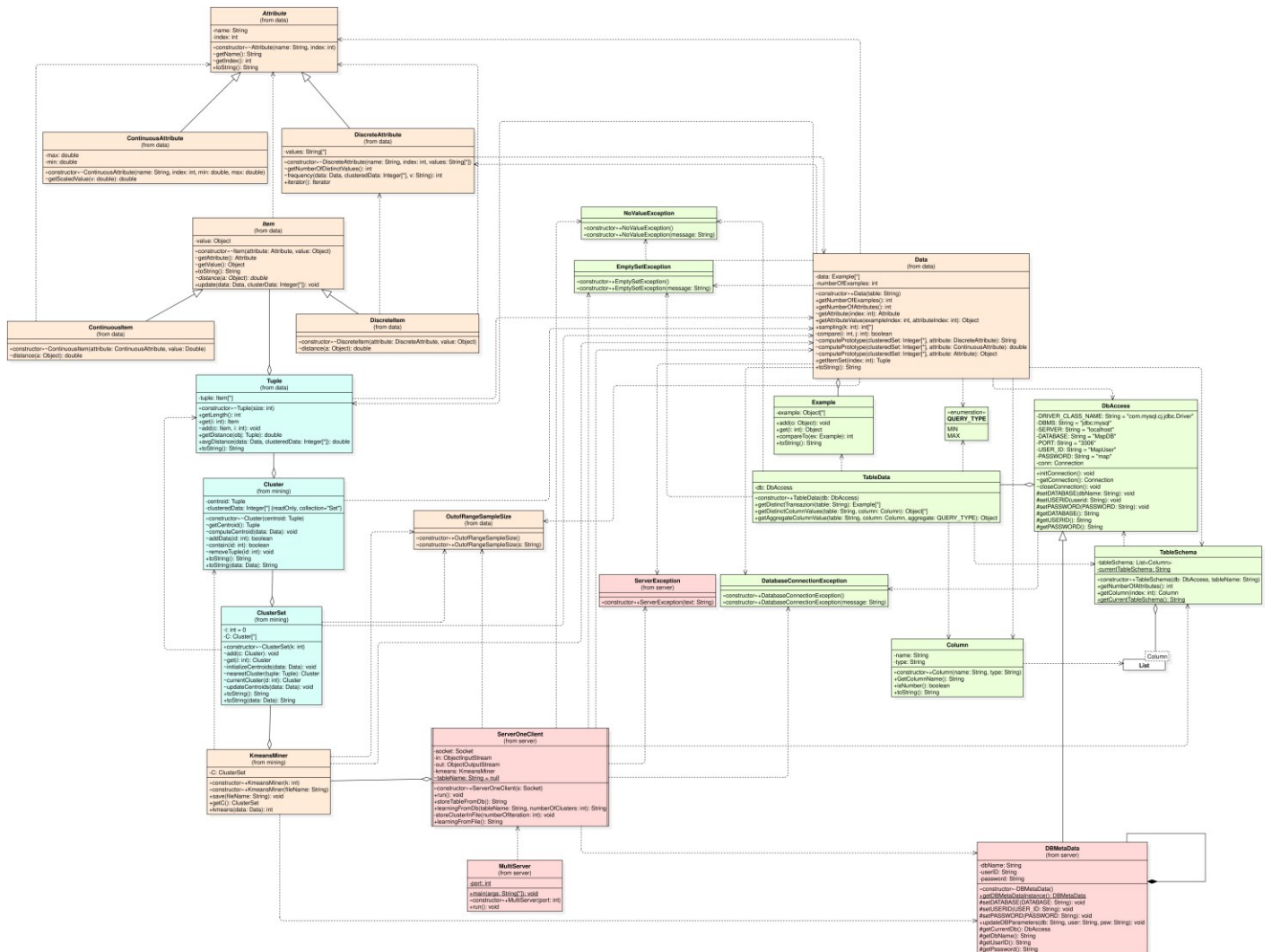


Diagramma dei package Server





3.Guida Utente

Il software si suddivide in due componenti principali, sia per la versione base che per quella estesa: il client e il server.

Il client ha due funzioni fondamentali: leggere i dati da un file di testo generato da una precedente elaborazione o richiedere al server l'esecuzione di una nuova computazione. Nel secondo caso, il client specifica i parametri necessari per accedere ai dati e per eseguire l'algoritmo.

Il server, d'altra parte, si occupa dell'elaborazione dell'algoritmo K-means e dell'invio dei risultati al client.

Inoltre, è possibile distribuire il server su una macchina separata rispetto al client, specialmente per l'elaborazione di grandi quantità di dati.

Rispetto alla versione base, la versione estesa offre la visualizzazione dei risultati dell'algoritmo K-means attraverso un'interfaccia grafica e la flessibilità di cambiare il database di partenza per una migliore adattabilità a MySQL.

Requisiti per l'utilizzo

Per utilizzare il software, è necessario disporre dei seguenti requisiti:

- Java 8 o superiore
- Una versione aggiornata di Java
- SDK MySQL 5.7 o superiore

Configurazione del sistema e del database

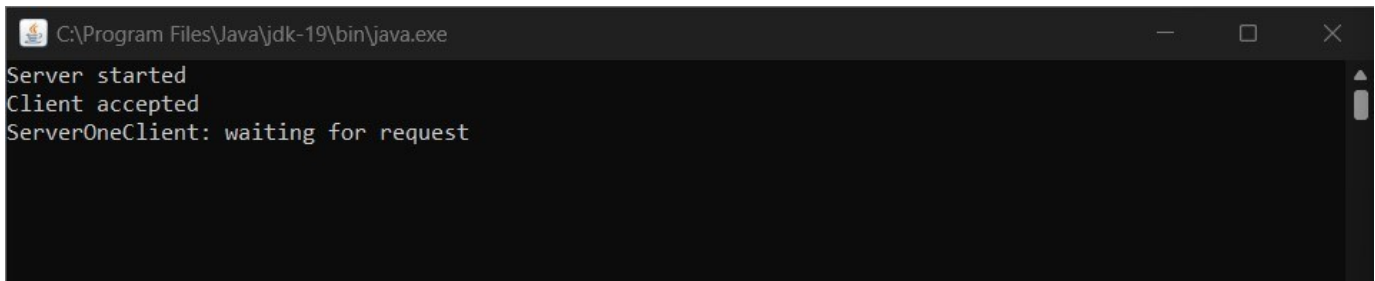
- avviare lo script `createDB.sql`, nella cartella “ conf”, per creare il database di partenza e verificare il corretto funzionamento del software. (opzionale per la versione estesa)
- avviare o verificare che il servizio MySQL sia in esecuzione.

NOTA: entrambi i file `.bat` per eseguire le applicazioni contengono un'istruzione per avviare il servizio mysql in automatico; tuttavia, in tal caso bisognerà eseguire il file come amministratore. In alternativa, è possibile avviare il servizio MySQL manualmente.

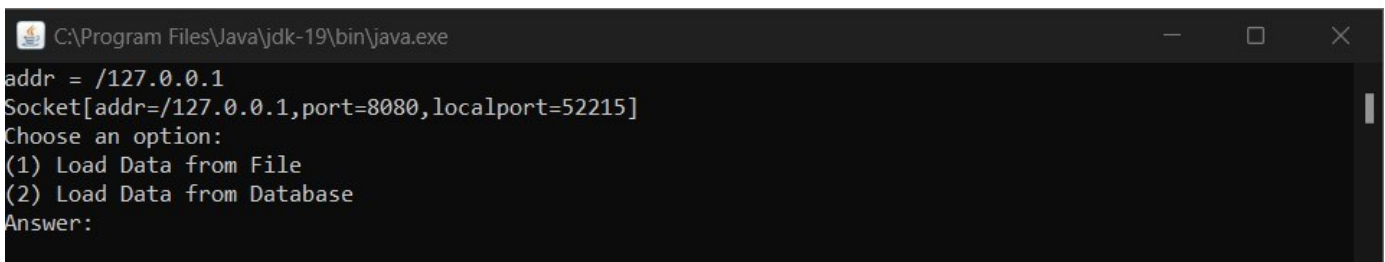
- alla fine di questi passaggi è possibile cliccare sul file `KmeansConsoleApp.bat` per avviare l'applicazione a riga di comando o `KmeansGUI.bat` per avviare l'applicazione con interfaccia grafica.

Versione Base

Una volta avviata l'applicazione, si apriranno due finestre di terminale: una per il client e una per il server.

A screenshot of a Windows terminal window with a dark background. The title bar shows the file path 'C:\Program Files\Java\jdk-19\bin\java.exe'. The terminal output consists of three lines: 'Server started', 'Client accepted', and 'ServerOneClient: waiting for request'.

```
C:\Program Files\Java\jdk-19\bin\java.exe
Server started
Client accepted
ServerOneClient: waiting for request
```

A screenshot of a Windows terminal window with a dark background. The title bar shows the file path 'C:\Program Files\Java\jdk-19\bin\java.exe'. The terminal output shows the address 'addr = /127.0.0.1', the socket configuration 'Socket[addr=/127.0.0.1,port=8080,localport=52215]', a prompt 'Choose an option:', two numbered options '(1) Load Data from File' and '(2) Load Data from Database', and a prompt 'Answer:'.

```
C:\Program Files\Java\jdk-19\bin\java.exe
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52215]
Choose an option:
(1) Load Data from File
(2) Load Data from Database
Answer:
```

Il client chiederà all'utente di scegliere tra due opzioni: leggere i dati da un file di testo o eseguire l'algoritmo K-means.

Se si sceglie la prima opzione, il client chiederà all'utente di inserire il nome della tabella da cui leggere i dati e il numero di cluster della precedente computazione, poiché il file viene salvato come `nomeTabella_numeroCluster.dat`.

```
C:\Program Files\Java\jdk-19\bin\java.exe
addr = /127.0.0.1
Socket[addr=/127.0.0.1,port=8080,localport=52215]
Choose an option:
(1) Load Data from File
(2) Load Data from Database
Answer:1
Table Name:playtennis
Number of clusters:4
0:Centroid=(overcast 30.15 high weak no )
Examples:
[sunny, 30.3, high, weak, no ] dist=1.004950495049505
[overcast, 30.0, high, weak, yes ] dist=1.004950495049505
AvgDistance=1.004950495049505

1:Centroid=(sunny 18.599999999999998 high strong no )
Examples:
[sunny, 30.3, high, strong, no ] dist=0.38613861386138626
[rain, 12.5, high, strong, no ] dist=1.2013201320132012
[sunny, 13.0, high, weak, no ] dist=1.1848184818481848
AvgDistance=0.9240924092409241

2:Centroid=(rain 9.544285714285715 normal weak yes )
Examples:
[rain, 13.0, high, weak, yes ] dist=1.114049976426214
[rain, 0.0, normal, weak, yes ] dist=0.314992927864215
[rain, 0.0, normal, strong, no ] dist=2.314992927864215
[sunny, 0.1, normal, weak, yes ] dist=1.3116925978312117
[rain, 12.0, normal, weak, yes ] dist=0.08104667609618099
[sunny, 12.5, normal, strong, yes ] dist=2.0975483262611974
[overcast, 29.21, normal, weak, yes ] dist=1.6490334747760491
AvgDistance=1.2690509867313262

3:Centroid=(overcast 6.3 high strong yes )
Examples:
[overcast, 0.1, normal, strong, yes ] dist=1.2046204620462047
[overcast, 12.5, high, strong, yes ] dist=0.20462046204620463
AvgDistance=0.7046204620462047

Do you want to choose another menu option?(y/n)y
```

Se si sceglie la seconda opzione, il client chiederà all'utente di inserire il nome della tabella da cui leggere i dati e il numero di cluster desiderati.

```
Choose an option:
(1) Load Data from File
(2) Load Data from Database
Answer:2
Table Name:playtennis
Number of clusters:4
Clustering output:
Number of iterations: 4
Centroid=(overcast 18.33333333333332 high strong yes )
Centroid=(sunny 24.53333333333333 high weak no )
Centroid=(overcast 0.05 normal strong no )
Centroid=(rain 11.135 normal weak yes )

0:Centroid=(overcast 18.33333333333332 high strong yes )
Examples:
[overcast, 30.0, high, weak, yes ] dist=1.3850385038503852
[overcast, 12.5, high, strong, yes ] dist=0.19251925192519248
[rain, 12.5, high, strong, no ] dist=2.1925192519251926
AvgDistance=1.2566923359002569

1:Centroid=(sunny 24.53333333333333 high weak no )
Examples:
[sunny, 30.3, high, weak, no ] dist=0.19031903190319044
[sunny, 30.3, high, strong, no ] dist=1.1903190319031904
[sunny, 13.0, high, weak, no ] dist=0.38063806380638054
AvgDistance=0.5870920425375871

2:Centroid=(overcast 0.05 normal strong no )
Examples:
[rain, 0.0, normal, strong, no ] dist=1.0016501650165017
[overcast, 0.1, normal, strong, yes ] dist=1.0016501650165017
AvgDistance=1.0016501650165017

3:Centroid=(rain 11.135 normal weak yes )
Examples:
[rain, 13.0, high, weak, yes ] dist=1.0615511551155115
[rain, 0.0, normal, weak, yes ] dist=0.3674917491749175
[sunny, 0.1, normal, weak, yes ] dist=1.3641914191419142
[rain, 12.0, normal, weak, yes ] dist=0.0285478547854785
[sunny, 12.5, normal, strong, yes ] dist=2.0450495049504953
[overcast, 29.21, normal, weak, yes ] dist=1.5965346534653466
AvgDistance=1.0772277227722773

Clustering saved on file
Do you want to compute again?(y/n)
```

Dopo la visualizzazione dei risultati, viene segnalato all'utente che il file contenente la clusterizzazione è stato salvato. L'utente, quindi, può ripetere il processo scegliendo nuovamente il numero di cluster da generare.

ATTENZIONE:

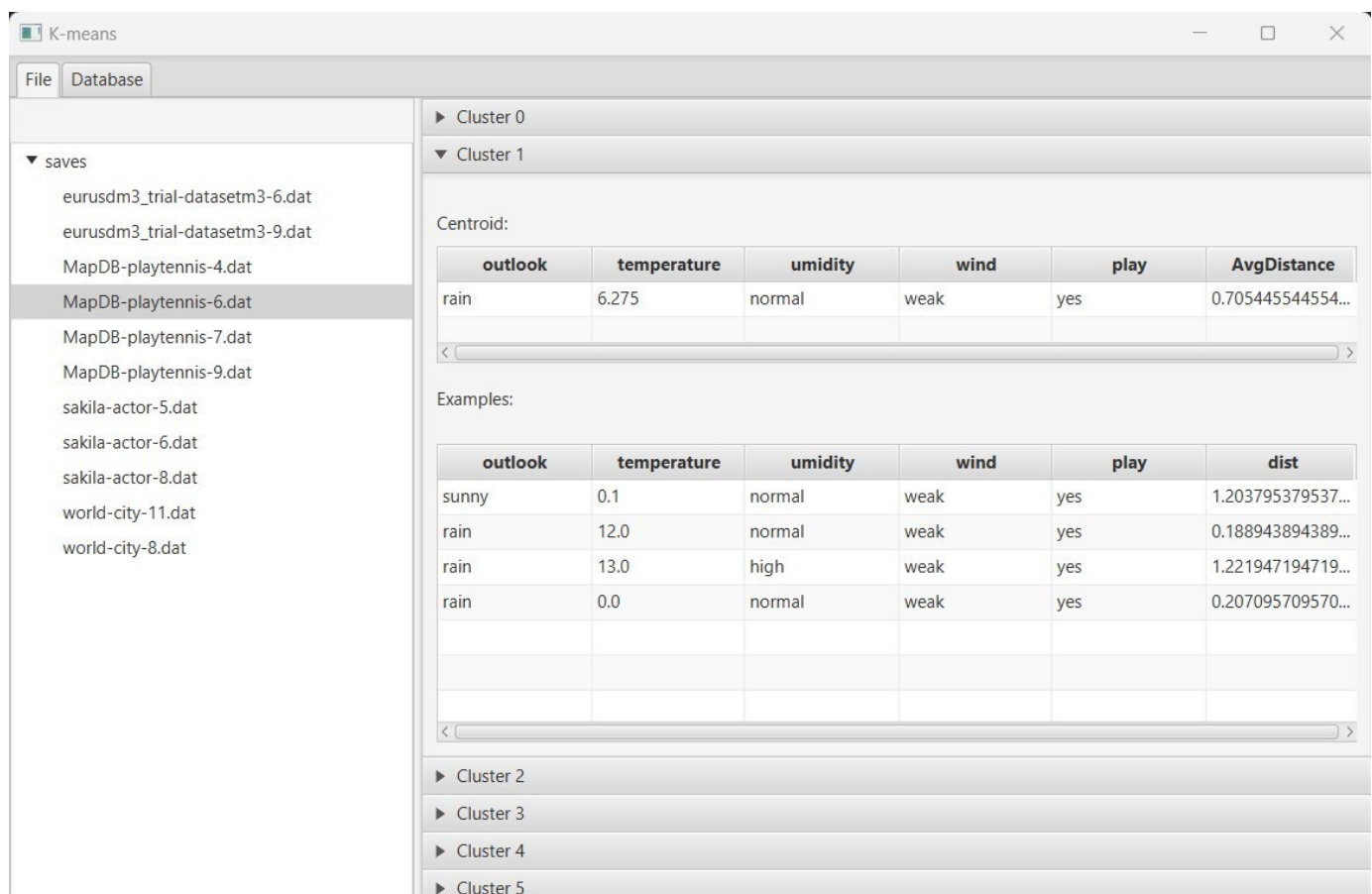
La versione base è un'applicazione a riga di comando che consente di eseguire l'algoritmo K-means su un database di partenza non modificabile; pertanto, se si ha la necessità di eseguire l'algoritmo su un dataset diverso, è necessario aggiungere una tabella al database di partenza.

Versione Estesa

Una volta avviata l'applicazione, si apriranno due finestre di terminale e una finestra di interfaccia grafica.

l'interfaccia è composta da due tab: una per la visualizzazione dei dati da File e una per la visualizzazione dei dati da Database.

la tab per la visualizzazione dei dati da File permette selezionare dall'apposito menù a tendina il file desiderato e di visualizzarne il contenuto.



la tab per la visualizzazione dei dati da Database contiene nella parte laterale una serie di campi di testo per inserire i parametri di connessione al database (Nome database, Nome utente, Password) e altri due campi per inserire il nome della tabella e il numero di cluster desiderati.

The screenshot shows the 'K-means' application window with the 'Database' tab selected. The interface is divided into a left sidebar for configuration and a main area for results.

Configuration Fields (Left Sidebar):

- Database Name:** MapDB
- Username:** MapUser
- Password:** map
- Table Name:** playtennis
- Number of Clusters:** 6
- Compute Button:** A button to execute the K-means algorithm.

Results Area (Main Panel):

- Cluster List:** A list of clusters from 0 to 5. Cluster 2 is currently expanded.
- Centroid Table:** A table showing the centroid for Cluster 2.
- Examples Table:** A table showing four data points (examples) for Cluster 2.
- Save Button:** A button at the bottom right to save the results.

Centroid Data (Cluster 2):

outlook	temperature	umidity	wind	play	AvgDistance
rain	6.275	normal	weak	yes	0.705445544554...

Examples Data (Cluster 2):

outlook	temperature	umidity	wind	play	dist
rain	13.0	high	weak	yes	1.22194719471...
rain	0.0	normal	weak	yes	0.20709570957...
sunny	0.1	normal	weak	yes	1.20379537953...
rain	12.0	normal	weak	yes	0.18894389438...

Una volta inseriti i parametri, è possibile cliccare sul pulsante "Compute" per avviare l'algoritmo K-means. Se i dati inseriti sono corretti, verrà visualizzata una finestra di dialogo con i risultati dell'algoritmo. Subito sotto ci sarà un pulsante "Save" per salvare i risultati in un file di testo.

Una volta salvati i risultati, è possibile ritrovare il file appena salvato, nel menu a tendina della tab "File" sotto il nome di **nomeDataBase-nomeTabella-numeroCluster.dat**.



NOTA: è stato necessario, sia includere nel nome del file il nome del database per poter identificare correttamente il file, sia adottare un nuovo carattere separatore poiché il carattere `_` è comunemente utilizzato nei nomi delle tabelle di un database, creando confusione con il separatore di default.

4.Casi di test

Estensione

Descrizione	Input	Output
Servizio MySql non attivo	Click su “Compute”	Messaggio di errore: “Error while connecting to Database. Try to start Mysql57”
Selezione di una delle due opzioni nella schermata iniziale	Click su “File”	Visualizzazione della schermata dedicata alla sezione da locale del file
Selezione di una delle due opzioni nella schermata iniziale	Click su “Database”	Visualizzazione della schermata dedicata alla generazione dei clusters da DB
Inserimento del nome della tabella non valido	“taberrata”	Visualizzazione del messaggio di errore “tabella non esistente”
Inserimento del nome della tabella corretto	“playtennis”	Visualizzazione del messaggio di errore per il mancato inserimento dei dati (numero cluster mancante)
Inserimento del numero di cluster corretto	7	Visualizzazione del messaggio di errore per il mancato inserimento dei dati (nome tabella mancante)
Inserimento del numero di cluster non corretto	20	Messaggio di errore: Il numero dei cluster è più grande del numero degli esempi
Inserimento del nome della tabella non valido + numero cluster corretto	taberrata + 7	Visualizzazione del messaggio di errore “tabella non esistente”
Inserimento del nome della tabella corretto + numero cluster non valido	playtennis + str	Visualizzazione del messaggio di errore per l’errato inserimento dei dati (numero cluster non valido)
Inserimento del nome della tabella corretto + numero cluster valido	playtennis + 7	Generazione corretta del risultato
Inserimento del nome della tabella corretto + numero cluster non valido	playtennis + 20	Messaggio di errore: Il numero dei cluster è più grande del numero degli esempi
Inserimento del nome della tabella corretto + numero cluster non valido	playtennis + 0	Messaggio di errore: Il numero di cluster deve essere > 0
Selezione di un file nella sezione ‘File’	Click su un nome del file	Visualizzazione dei cluster del file selezionato
Selezione di un file nella sezione ‘File’ con servizio MySql non attivo	Click su un nome del file	Messaggio di errore: “Error while connecting to Database. Try to start Mysql57”

Base

Descrizione	Input	Output
Avvio dell'applicazione con il servizio MySql non attivo	∅	Visualizzazione del messaggio di errore per il mancato avvio del servizio MySql
Avvio dell'applicazione con il servizio MySql attivo	∅	Corretta esecuzione del client
Inserimento nel menu di un valore accettato	2	Inserimento del nome della tabella del database
Inserimento nel menu di un valore non accettato (intero)	3	Nuovo inserimento nel menu di un valore
Inserimento nel menu di un valore non accettato (stringa)	"test"	Errore [Error reading int data, MIN_VALUE value returned] e nuovo inserimento nel menu di un valore
Inserimento di una tabella esistente e popolata in mapDB	"playtennis"	Prossimo inserimento: numero cluster
Inserimento di una tabella non esistente	"inesistente"	Messaggi di errore: tabella non esistente
Inserimento di un numero di cluster compreso nel range	7	Numero iterazioni + clusters generati
Inserimento di un numero di cluster compreso nel range – Caso limite inferiore	1	Numero iterazioni + clusters generati
Inserimento di un numero di cluster compreso nel range – Caso limite superiore	14	Numero iterazioni + clusters generati
Inserimento di un numero di cluster non compreso nel range	0	Messaggio di errore: numero di cluster non valido
Inserimento di un numero di cluster non compreso nel range	18	Messaggio di errore: numero di cluster non valido
Inserimento di una stringa nel campo cluster	"test"	Messaggio di errore [Error reading int data, MIN_VALUE value returned]
Inserimento di una tabella esistente e popolata + numero cluster (file esistente)	playtennis + 7	Risultato lettura da file playtennis_8
Inserimento di una tabella esistente + numero cluster (file non esistente)	playtennis + 7	Messaggio di errore: file non trovato
Inserimento di una tabella non esistente + numero cluster	inesistente + 7	Messaggio di errore: la tabella non esiste