

# Βήματα 2ης Εργασίας

- ▶ Επιπεδοποιώ το αρχείο keywords.csv
- ▶ SQL Server Import
- ▶ Εισαγωγή Δεδομένων στην βάση μου
- ▶ Δημιουργία πρωτευόντων και Ξένων κλειδιών



# Επιπεδοποίηση του αρχείου keywords.csv

movie_id	keywords
862	[{'id': 931, 'name': 'jealousy'}, {'id': 4290, 'name': 'toy'}, {'id': 5202, 'name': 'boy'}, {'id': 6054, 'name': 'friendship'}, {'id': 9713, 'name': 'friends'}, {'id': 9823, 'name': 'rivalry'}, {'id': 165503, 'name': 'boy next door'}, {'id': 170722, 'name': 'new toy'}, {'id': 187065, 'name': 'toy comes to life'}]
8844	[{'id': 10090, 'name': 'board game'}, {'id': 10941, 'name': 'disappearance'}, {'id': 15101, 'name': '"based on children's book"'}, {'id': 33467, 'name': 'new home'}, {'id': 158086, 'name': 'recluse'}, {'id': 158091, 'name': 'giant insect'}]
15602	[{'id': 1495, 'name': 'fishing'}, {'id': 12392, 'name': 'best friend'}, {'id': 179431, 'name': 'duringcreditsstinger'}, {'id': 208510, 'name': 'old men'}]

Η στήλη keywords περιέχει εμφωλευμένη λίστα από λέξεις κλειδιά, η οποία αναπαριστάται με την μορφή Json Συμβολοσειράς.

hasKeyword	
movie_id	key_id
862	931
862	4290
862	5202
862	6054

Keyword	
id	name
931	jealousy
4290	toy
5202	boy
6054	friendship

Σπάω το αρχικό keywords.csv αρχείο σε δύο διαφορετικά αρχεία, το πρώτο απεικονίζει το movie\_id κάθε ταινίας στο key\_id μιας λέξης κλειδιού και το δεύτερο απεικονίζει κάθε id μιας λέξης κλειδί στην αντίστοιχη λέξη κλειδί

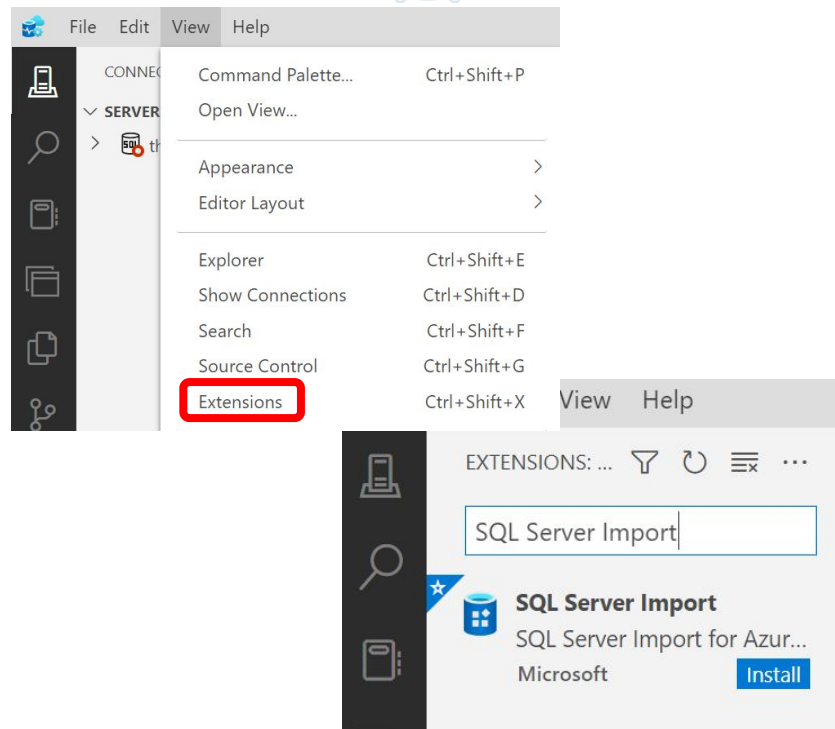
- ▶ Επιπεδοποιώ το αρχείο keywords.csv
- ▶ SQL Server Import
- ▶ Εισαγωγή Δεδομένων στην βάση μου
- ▶ Δημιουργία πρωτεύοντων και Ξένων κλειδιών



# SQL Server Import

# Εγκαθιστώ στο Azure Data studio το extension SQL Server Import

- ▶ Ανοίγω το Azure Data Studio
- ▶ Πηγαίνω στο View→Extensions
- ▶ Επιλέγω το extension SQL Server Import και το κάνω install
- ▶ Ενδεχομένως να χρειαστεί επανεκκίνηση



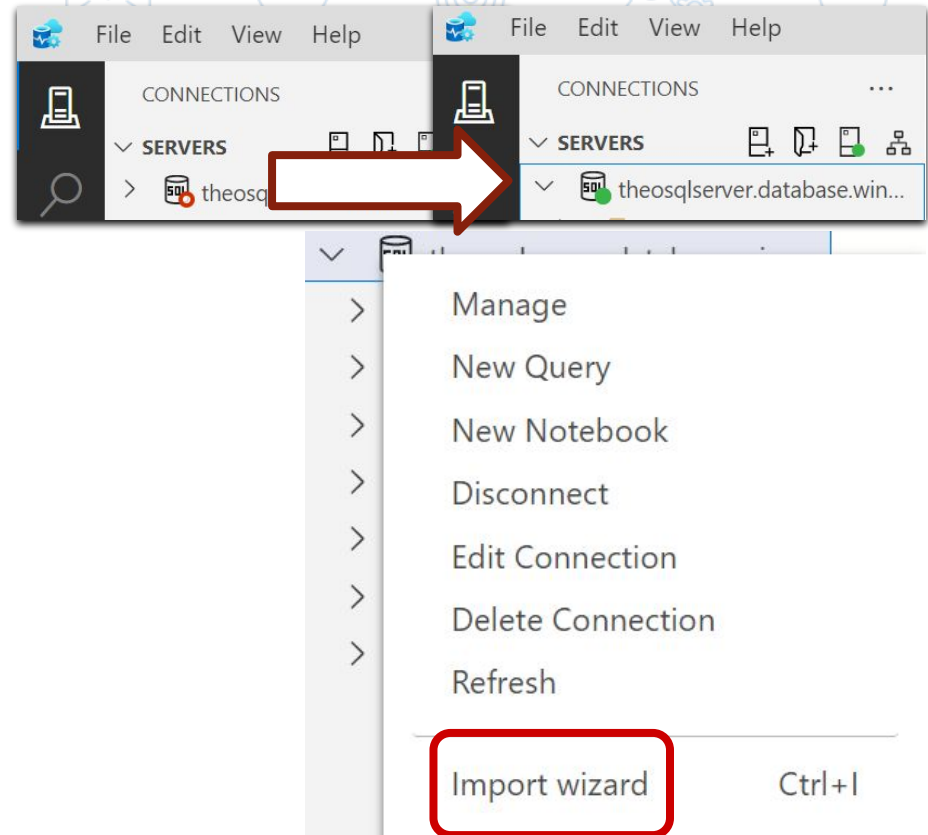
- ▶ Επιπεδοποιώ το αρχείο keywords.csv
- ▶ SQL Server Import
- ▶ Εισαγωγή Δεδομένων στην βάση μου
- ▶ Δημιουργία πρωτευόντων και Ξένων κλειδιών



# Δημιουργία και Αρχικοποίηση Πινάκων

# Χρησιμοποιώ το SQL Server Import για την δημιουργία και την αρχικοποίηση των πινάκων μου

- ▶ Συνδεθείτε στον Server της βάσης
- ▶ Δεξί click στον server και πηγαίνετε στον Import wizard





# Δημιουργία του πίνακα ratings, από το αρχείο ratings.csv

- ▶ Location of the file to be imported → Επιλέγεται το αρχείο το οποίο περιέχει τα δεδομένα του πίνακα
- ▶ New Table name → να έχει το ίδιο όνομα με το csv αρχείο
- ▶ Table Schema → αφήστε το default (dbo)

1

2

3

4

## Step 1: Specify Input File

Server the database is in \*

theosqlserver.database.windows.net (azureuser) ▼

Database the table is created in \*

mySampleDatabase ▼

Location of the file to be imported \*

d:\theof\Downloads\dataset(1)\ratings.csv

Browse

New table name \*

ratings

Table schema \*

dbo ▼

# Δημιουργία του πίνακα ratings, από το αρχείο ratings.csv

- ▶ Στο δεύτερο βήμα σας δείχνει την μορφή των δεδομένων που θα εισαχθούν



## Step 2: Preview Data

This operation analyzed the input file structure to c

user_id	movie_id	rating
1	1371	2.5
1	1405	1.0
1	2105	4.0
1	2193	2.0
1	2294	2.0
2	62	3.0
2	110	4.0
2	144	3.0
2	150	5.0

# Δημιουργία του πίνακα ratings, από το αρχείο ratings.csv

- ▶ Στο τρίτο βήμα, ο SQL server import έχει βρει τα ονόματα των γνωρισμάτων σας από την 1<sup>η</sup> γραμμή του csv αρχείου
- ▶ “Μαντεύει” τον τύπο κάθε γνωρίσματος από τις πρώτες εγγραφές
- ▶ Αφήστε το πεδίο Primary keys κενό (θα το φτιάξετε στην συνέχεια)

## Step 3: Modify Columns

Column Name	Data Type	Primary Key <input type="checkbox"/>	Allow Nulls <input type="checkbox"/>
user_id	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
movie_id	int	<input type="checkbox"/>	<input type="checkbox"/>
rating	float	<input type="checkbox"/>	<input type="checkbox"/>

Σε αρκετές περιπτώσεις ο τύπος που έχει μαντέψει μπορεί να είναι λάθος, π.χ., μπορεί να μαντέψει tinyint αντί για int, ή varchar(50) αντί για varchar(300) ή text. Σε αυτήν την περίπτωση θα **επιστραφεί error**. Θα πρέπει να αναπροσαρμόσετε ανάλογα τους τύπους των πεδίων σε πιο γενικούς τύπους.

# Δημιουργία του πίνακα ratings, από το αρχείο ratings.csv

Το συγκεκριμένο error μας λέει ότι η τιμή 256 δεν μπορεί να απεικονιστεί σαν tinyint για το Column 0 που αντιστοιχεί στον user\_id.

## Step 3: Modify Columns

Column Name	Data Type	Primary Key <input type="checkbox"/>	Allow Nulls <input type="checkbox"/>
user_id	tinyint	<input type="checkbox"/>	<input type="checkbox"/>
movie_id	int	<input type="checkbox"/>	<input type="checkbox"/>
rating	float	<input type="checkbox"/>	<input type="checkbox"/>

Σε αρκετές περιπτώσεις ο τύπος που έχει μαντέψει μπορεί να είναι λάθος, π.χ., μπορεί να μαντέψει tinyint

Import status

X Microsoft.SqlServer.Prose.Import.BcpProcessException: Error inserting data into table.

---> System.InvalidOperationException: The given value '256' of type String from the data source cannot be converted to type tinyint for Column 0 [user\_id].

---> System.OverflowException: Failed to convert parameter value from a String to a Byte.

---> System.OverflowException: Value was either too large or too small for an unsigned byte.

at System.Number.ThrowOverflowException(TypeCode type)

# Δημιουργία του πίνακα ratings, από το αρχείο ratings.csv

- ▶ Στο τελευταίο βήμα έχει ολοκληρωθεί η δημιουργία και η αρχικοποίηση του πίνακα, οπότε πατάμε το κουμπί done

Done

1

2

3

4

## Step 4: Summary

### Import information

Object type	Name
Server name	theosqlserver.database.windows.net
Database name	mySampleDatabase
Table name	ratings
Table schema	dbo
File to be imported	d:\theof\Desktop\CreateDatabase\ratings.csv

### Import status

✓ You have successfully inserted the data into a table.

- ▶ Επιπεδοποιώ το αρχείο keywords.csv
- ▶ SQL Server Import
- ▶ Εισαγωγή Δεδομένων στην βάση μου
- ▶ Δημιουργία πρωτευόντων και Ξένων κλειδιών



# Δημιουργία Πρωτευόντων και Ξένων Κλειδιών

# Δημιουργία Πρωτεύοντων Κλειδιών

- ▶ Δημιουργώ για τους πίνακές μου τα πρωτεύοντα κλειδιά:
  - ▶ Π.χ. για να **δημιουργήσω πρωτεύον κλειδί** για τον πίνακα ratings στα γνωρίσματα (user\_id, movie\_id) τρέχω την εντολή
  - ▶ Σε περίπτωση που **δεν με αφήνει να δημιουργήσω πρωτεύον κλειδί γιατί το γνώρισμά μου επιτρέπει Null τιμές**, κάνω τις αντίστοιχες αλλαγές, έτσι ώστε να μην επιτρέπονται null τιμές.

```
ALTER TABLE ratings
ADD CONSTRAINT ratings_pk
PRIMARY KEY (user_id, movie_id);
```

Ονομασία για το  
πρωτεύον κλειδί

Το πρωτεύον κλειδί  
του πίνακα ratings

```
ALTER TABLE ratings
ALTER COLUMN user_id int NOT NULL;
ALTER TABLE ratings
ALTER COLUMN movie_id int NOT NULL;
```

# Δημιουργία Ξένων Κλειδιών

- ▶ Δημιουργώ για τους πίνακές μου τα ξένα κλειδιά:
  - ▷ Προσοχή για να είναι το **movie\_id** ξένο κλειδί που αναφέρεται στο **γνώρισμα id του πίνακα Movie**, θα πρέπει ο πίνακας Movie να έχει σαν πρωτεύον κλειδί το id.

```
ALTER TABLE ratings  
ADD CONSTRAINT FK_movieid FOREIGN  
KEY (movie_id)  
REFERENCES movie(id);
```

Ονομασία για το ξένο  
κλειδί

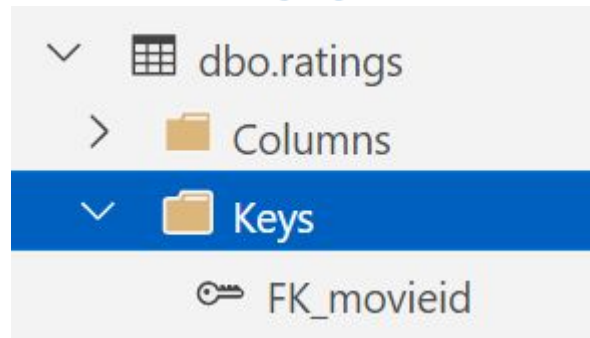
Το γνώρισμα  
movie\_id είναι  
δευτερεύον κλειδί  
που αναφέρεται στο  
γνώρισμα id του  
πίνακα movie



# Διαγραφή Κλειδιών

**Αν χρειαστεί** να διαγράψω ένα πρωτεύον ή ξένο κλειδί:

- ▶ πρώτα το βρίσκω στο azure data studio και
- ▶ μετά το κάνω drop



```
ALTER TABLE ragins  
DROP CONSTRAINT FK_movieid;
```