

Σχεδιασμός Βάσεων Δεδομένων

Διδάσκων: Ιωάννης Κωτίδης

Εαρινό εξάμηνο 2019-2020

Εργασία - Μέρος Α

Ανάθεση: **09-04-2020**

Παράδοση: **28-04-2020 Ώρα (23:55)**

Οδηγίες

- Η εργασία είναι ατομική και υποχρεωτική.
- Η υποβολή της εργασίας πρέπει να γίνει στο *eclass*.
- Το παραδοτέο σας θα πρέπει να είναι ένα αρχείο PDF με όνομα *AM.pdf* (όπου *AM* είναι ο αριθμός μητρώου σας, π.χ. "3170001.pdf").
- Πιθανή αντιγραφή θα τιμωρείται με μηδενισμό όλων των εμπλεκομένων.

"Βιβλιογραφική Βάση Δεδομένων" - ΜΕΡΟΣ ΠΡΩΤΟ

Στόχος της εργασίας είναι η πρακτική εφαρμογή των γνώσεων που αποκομίσατε από τις διαλέξεις του μαθήματος σχετικά με την δημιουργία ευρετηρίων και την βελτιστοποίηση των επερωτήσεων SQL. Για την πρακτική σας εξάσκηση θα χρησιμοποιήσετε μια βιβλιογραφική βάση δεδομένων και το DBMS MICROSOFT SQL SERVER. Η βάση δεδομένων περιέχει βιβλιογραφικές εγγραφές που περιγράφουν τα τεκμήρια της συλλογής μιας πανεπιστημιακής βιβλιοθήκης και εγγραφές με τα στοιχεία των δανειζομένων/μελών της βιβλιοθήκης.

Αρχικά θα δημιουργήσετε την βιβλιογραφική βάση και θα φορτώσετε τα δεδομένα στους πίνακες, ακολουθώντας τις παρακάτω οδηγίες. Στη συνέχεια θα απαντήσετε στα ζητούμενα της εργασίας.

1. Οδηγίες για την δημιουργία της βιβλιογραφικής βάσης δεδομένων.

Για να δημιουργήσετε την βάση δεδομένων και να φορτώσετε τις εγγραφές ακολουθείστε **ΠΡΟΣΕΚΤΙΚΑ** τα παρακάτω βήματα:

Βήμα 1: Από το περιβάλλον του Microsoft Sql Server Management Studio δημιουργείτε μια βάση δεδομένων με όνομα **LIBRARY**. Ορίστε το **COLLATION** της βάσης σε **Greek_CI_AS** (όπου **CI** σημαίνει Case Insensitive και **AS** σημαίνει Accent Sensitive). Δείτε αναλυτικές οδηγίες στο **ΠΑΡΑΡΤΗΜΑ** που υπάρχει στο τέλος της εργασίας.

Βήμα 2: Εκτελέστε το SQL script "**CreateLibrarySchema.sql**" που δημιουργεί το λογικό σχήμα της βάσης. Πριν εκτελέσετε το script βεβαιωθείτε ότι η τρέχουσα βάση δεδομένων είναι η βάση **LIBRARY** που δημιουργήσατε στο βήμα 1.

Βήμα 3: Εκτελέστε το SQL script "**LoadLibraryData.sql**" το οποίο θα φορτώσει δεδομένα στους πίνακες της βάσης. Το συγκεκριμένο script περιέχει εντολές της μορφής:

BULK INSERT bibrecs ! Πίνακας στον οποίο θα φορτωθούν τα δεδομένα
FROM 'C:\library\DATA\BIBRECS.txt' ! Αρχείο που περιέχει τα δεδομένα
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|', ROWTERMINATOR = '\n');

ΠΡΟΣΟΧΗ: Αν τοποθετήσετε τα δεδομένα σε φάκελο διαφορετικό από τον "C:\library\DATA" θα πρέπει να τροποποιήσετε ανάλογα το τμήμα **FROM** όλων των εντολών BULK INSERT που περιέχονται στο SCRIPT. Για παράδειγμα αν τοποθετήσετε τα δεδομένα στον φάκελο "C:\MYDATA" η παραπάνω εντολή πρέπει να αλλάξει ως εξής:

BULK INSERT bibrecs
FROM 'C:\MYDATA\BIBRECS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|', ROWTERMINATOR = '\n');

Τα SQL scripts "**CreateLibrarySchema.sql**" και "**LoadLibraryData.sql**" καθώς επίσης και τα αρχεία με τα δεδομένα που θα φορτωθούν στους πίνακες της βάσης θα τα βρείτε στο αρχείο "**Library.zip**".

ΣΗΜΕΙΩΣΗ: Τα περιεχόμενα των παραπάνω scripts παρατίθενται στο **ΠΑΡΑΡΤΗΜΑ** που ακολουθεί στο τέλος της εργασίας.

2. Περιγραφή των πινάκων της βάσης

Ακολουθεί η περιγραφή των πινάκων και των δεδομένων της βάσης.

DEPARTMENTS: Πίνακας με τα στοιχεία των τμημάτων του πανεπιστημίου. Αριθμός Εγγραφών=10	
depcode	Κωδικός τμήματος
depname	Όνομα τμήματος

PUBLISHERS: Πίνακας με τα στοιχεία των εκδοτών. Αριθμός εγγραφών=3448	
pubid	Κωδικός εκδότη
pubname	Όνομα εκδότη

AUTHORS: πίνακας με τα στοιχεία των συγγραφέων. Τα ονόματα των συγγραφέων είναι μοναδικά. Αριθμός εγγραφών=42018.	
aid	Κωδικός συγγραφέα
author	Όνομα συγγραφέα

STERMS: Πίνακας με τους θεματικούς όρους (Subject Terms) που αποδίδονται στις βιβλιογραφικές εγγραφές. Οι θεματικοί όροι είναι μοναδικοί. Αριθμός εγγραφών=18540.	
tid	Κωδικός θεματικού όρου
term	Θεματικός όρος

BIBRECS: Πίνακας με τα στοιχεία των βιβλιογραφικών εγγραφών που περιγράφουν τα τεκμήρια της συλλογής της βιβλιοθήκης. Αριθμός εγγραφών=66713.	
bibno	Κωδικός βιβλιογραφικής εγγραφής
title	Τίτλος του τεκμηρίου που περιγράφει η βιβλιογραφική εγγραφή
material	Είδος τεκμηρίου (book, serial, audiovisual, paper κ.λπ.)
lang	Γλώσσα τεκμηρίου (gre, eng, fre, κλπ)
place	Τόπος έκδοσης
series	Τίτλος σειράς τεκμηρίου
pubid	Κωδικός εκδότη

BIBAUTHORS: Πίνακας που συνδέει τις βιβλιογραφικές εγγραφές με τους συγγραφείς. Αριθμός εγγραφών=82020.	
bibno	Κωδικός βιβλιογραφικής εγγραφής
aid	Κωδικός συγγραφέα

BIBTERMS: Πίνακας που συνδέει τις βιβλιογραφικές εγγραφές με τους θεματικούς όρους. Αριθμός εγγραφών=396901.	
bibno	Κωδικός βιβλιογραφικής εγγραφής
tid	Κωδικός θεματικού όρου

COPIES: Πίνακας με τα στοιχεία των αντιτύπων των τεκμηρίων. Αριθμός εγγραφών=85530.	
copyno	Κωδικός αντιτύπου.
bibno	Κωδικός βιβλιογραφικής εγγραφής
copyloc	Τριψήφιος Κωδικός που δηλώνει την τοποθεσία του αντιτύπου. Π.χ. "ANA" σημαίνει "Αναγνωστήριο".

BORROWERS: Πίνακας με τα στοιχεία των μελών της βιβλιοθήκης. Αριθμός εγγραφών=11287. ΤΑ ΣΤΟΙΧΕΙΑ ΔΕΝ ΕΙΝΑΙ ΠΡΑΓΜΑΤΙΚΑ.	
bib	Κωδικός μέλους
depcode	Κωδικός τμήματος
sex	Φύλο (F=female, M=Male)
bname	Ονοματεπώνυμο μέλους

LOANSTATS: Πίνακας με το ιστορικό των δανεισμών. Κάθε φορά που ένα μέλος δανείζεται ένα αντίτυπο, μια εγγραφή που περιγράφει την συναλλαγή του δανεισμού καταχωρείται στον πίνακα. Αριθμός εγγραφών=91640.	
lid	Κωδικός εγγραφής δανεισμού.
copyno	Κωδικός αντιτύπου
bid	Κωδικός μέλους
loandate	Ημερομηνία δανεισμού

3. Ζητούμενα εργασίας

Ακολουθούν τα ζητούμενα της εργασίας. Όπου ζητείται να γράψετε εντολές SQL που απαντούν σε συγκεκριμένα ερωτήματα **ΑΠΑΓΟΡΕΥΕΤΑΙ Η ΔΗΜΙΟΥΡΓΙΑ ΚΑΙ ΧΡΗΣΗ ΟΨΕΩΝ και ΠΡΟΣΩΡΙΝΩΝ ΠΙΝΑΚΩΝ.**

Ζήτημα Πρώτο

1. Να δημιουργήσετε κατάλληλο ευρετήριο που να επιταχύνει την εκτέλεση του παρακάτω ερωτήματος:

```
select title
from bibrecs where title like 'Οικ%'
order by title
```

Να γράψετε την εντολή δημιουργίας του δείκτη.

2. Να γράψετε κατάλληλες εντολές σε γλώσσα SQL ώστε να απαντηθούν τα παρακάτω ερωτήματα:
 - a. Εμφάνισε τον τίτλο των τεκμηρίων που περιέχουν στον τίτλο την λέξη πληροφορική.
 - b. Εμφάνισε τον τίτλο, και το είδος των τεκμηρίων με τίτλο "Economics".
 - c. Εμφάνισε τον τίτλο και το είδος των τεκμηρίων των οποίων ο τίτλος ξεκινάει με την λέξη "Economics".

Για κάθε ένα από τα παραπάνω επερωτήματα να εξετάσετε αν το ευρετήριο που δημιουργήσατε (ερώτημα 1) επιταχύνει την εκτέλεσή του. Σε κάθε περίπτωση να αιτιολογήσετε την απάντησή σας.

Ζήτημα Δεύτερο

1. Να γράψετε κατάλληλες εντολές σε γλώσσα SQL ώστε να απαντηθούν τα παρακάτω ερωτήματα:
 - a. Εμφανίστε τον τίτλο και την γλώσσα των τεκμηρίων που έχει εκδώσει ο 'Κλειδάριθμος'.
 - b. Εμφανίστε έναν κατάλογο με το όνομα κάθε τμήματος και τον συνολικό αριθμό των δανεισμών που έγιναν από τα μέλη του το έτος 2000.

- c. Εμφανίστε έναν κατάλογο με τον τίτλο, την γλώσσα και το όνομα του συγγραφέα (ή των συγγραφέων) των τεκμηρίων στα οποία έχει αποδοθεί ο θεματικός όρος "Databases".
2. Να δημιουργήσετε κατάλληλα ευρετήρια που θα επιταχύνουν την εκτέλεση των παραπάνω ερωτημάτων. Να παραθέσετε τις εντολές δημιουργίας των ευρετηρίων, καθώς επίσης και στοιχεία που να αποδεικνύουν ότι, τα ευρετήρια που δημιουργήσατε επιταχύνουν την εκτέλεση των ερωτημάτων.

Ζήτημα Τρίτο

Θεωρήστε το παρακάτω ερώτημα σε φυσική γλώσσα:

"Εμφανίστε τον κωδικό και τον τίτλο των βιβλίων για τα οποία η βιβλιοθήκη διαθέτει τουλάχιστον ένα αντίτυπο στην τοποθεσία 'ΟΡΑ' **ΚΑΙ** τουλάχιστον ένα αντίτυπο στην τοποθεσία 'ΑΝΑ'.

1. Να γράψετε τουλάχιστον δύο με τρία διαφορετικά επερωτήματα σε SQL που να απαντούν στο παραπάνω ερώτημα. Ποιό επερωτήμα θα επιλέγατε και γιατί;
2. Δημιουργείτε κατάλληλα ευρετήρια που θα επιταχύνουν την εκτέλεση του επερωτήματος που επιλέξατε.

Να τεκμηριώσετε τις απαντήσεις σας.

Ζήτημα Τέταρτο

Η βιβλιοθήκη θέλει η βάση δεδομένων να υποστηρίζει αναζητήσεις ελεύθερου κειμένου (free text search) από συγκεκριμένα πεδία κάθε βιβλιογραφικής εγγραφής, τα οποία καλούνται "Πεδία ελεύθερου κειμένου" (free text fields). Τα πεδία αυτά είναι: **Τίτλος (title)**, **Τίτλος Σειράς (series)**, **Εκδότης (pubname)**, **Συγγραφέας (Author)** και **Θεματικός όρος (term)**. Για τον λόγο αυτό θα χρησιμοποιηθούν οι παρακάτω πίνακες:

- **words (wid, word)** όπου **wid**: κωδικός λέξης και **word** η λέξη.
- **bibwords (wid, bibno)** όπου **wid**: κωδικός λέξης και **bibno**: κωδικός βιβλιογραφικής εγγραφής.

Κάθε φορά που μία νέα βιβλιογραφική εγγραφή καταχωρείται στην βάση, ένα πρόγραμμα που εκτελείται στο παρασκήνιο, εντοπίζει τις λέξεις που υπάρχουν στα "πεδία ελεύθερου κειμένου" και ενημερώνει κατάλληλα τους παραπάνω πίνακες. Με τον τρόπο αυτό η εφαρμογή αναζητήσεων (OPAC) μπορεί να απαντά ερωτήματα της μορφής:

"Εμφάνισε τον κωδικό των βιβλιογραφικών εγγραφών που περιέχουν την λέξη 'οικονομία' και την λέξη 'Ευρώπη'".

ΠΑΡΑΔΕΙΓΜΑ: Αν στο σύστημα καταχωρηθεί η παρακάτω βιβλιογραφική εγγραφή:

Κωδικός βιβλιογραφικής Εγγραφής=100
Τίτλος= Σχεδιασμός Βάσεων Δεδομένων

Εκδότης=Σμπίλιας

Συγγραφέας=Γιαννακουδάκης Εμμανουήλ

Θεματικός όρος=Βάσεις Δεδομένων

το περιεχόμενο των πινάκων **words** και **bibwords** θα είναι:

wid	word	wid	bibno
1	Σχεδιασμός	1	100
2	Βάσεων	2	100
3	Δεδομένων	3	100
4	Σμπίλιας	4	100
5	Γιαννακουδάκης	5	100
6	Εμμανουήλ	6	100
7	Βάσεις	7	100

Ζητείται:

1. Να γράψετε τις κατάλληλες εντολές SQL για την δημιουργία των πινάκων **words** και **bibwords**. Να ορίσετε τα κλειδιά του κάθε πίνακα (πρωτεύοντα και ξένα κλειδιά).
2. Να δημιουργήσετε κατάλληλα ευρετήρια που να επιταχύνουν την εκτέλεση των ερωτημάτων της παραπάνω μορφής.

Ζήτημα Πέμπτο (Bonus 2 μονάδες)

Η δομή που προτείνεται στο Τέταρτο Ζήτημα δεν υποστηρίζει ερωτήματα της παρακάτω μορφής:

*"Εμφάνισε τον κωδικό και τον τίτλο των βιβλιογραφικών εγγραφών που περιέχουν στον **τίτλο** την λέξη "οικονομία" στην **σειρά** την λέξη "ελληνική" και στον **συγγραφέα** την λέξη "Οικονόμου".*

Να προτείνετε μία ή περισσότερες λύσεις ώστε να υποστηρίζονται αποτελεσματικά ερωτήματα της παραπάνω μορφής. Εξυπακούεται ότι στα ερωτήματα της παραπάνω μορφής μπορεί να υπάρχει οποιοσδήποτε συνδυασμός των "πεδίων ελεύθερου κειμένου".

Στην αξιολόγηση θα ληφθεί υπόψη η ποιότητα της προτεινόμενης λύσης.

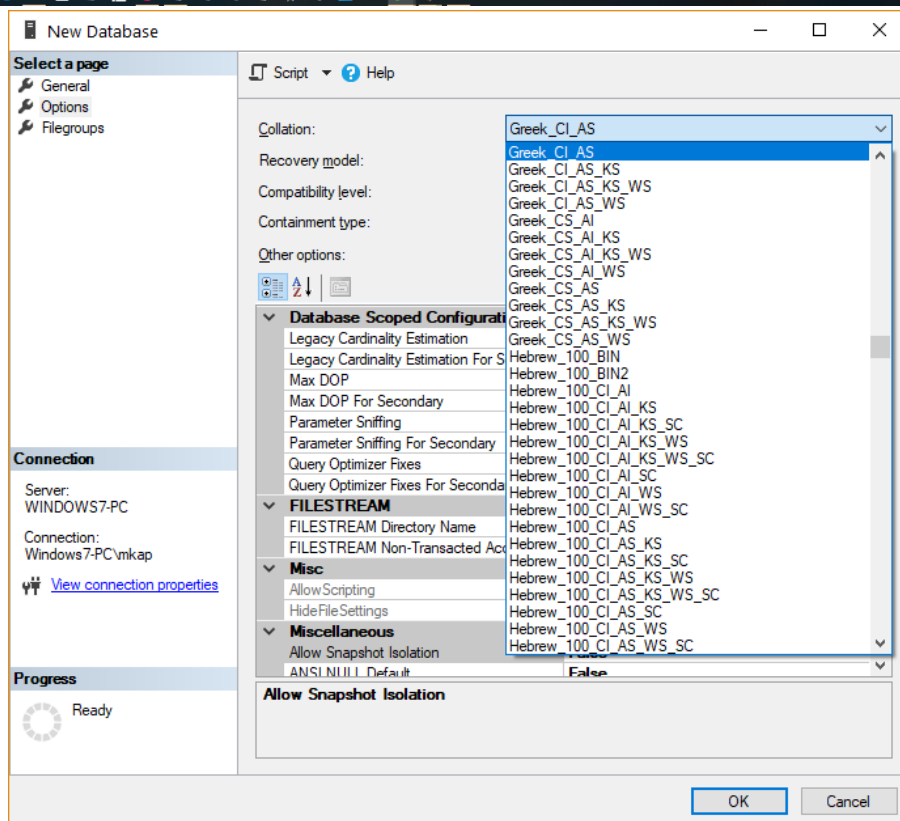
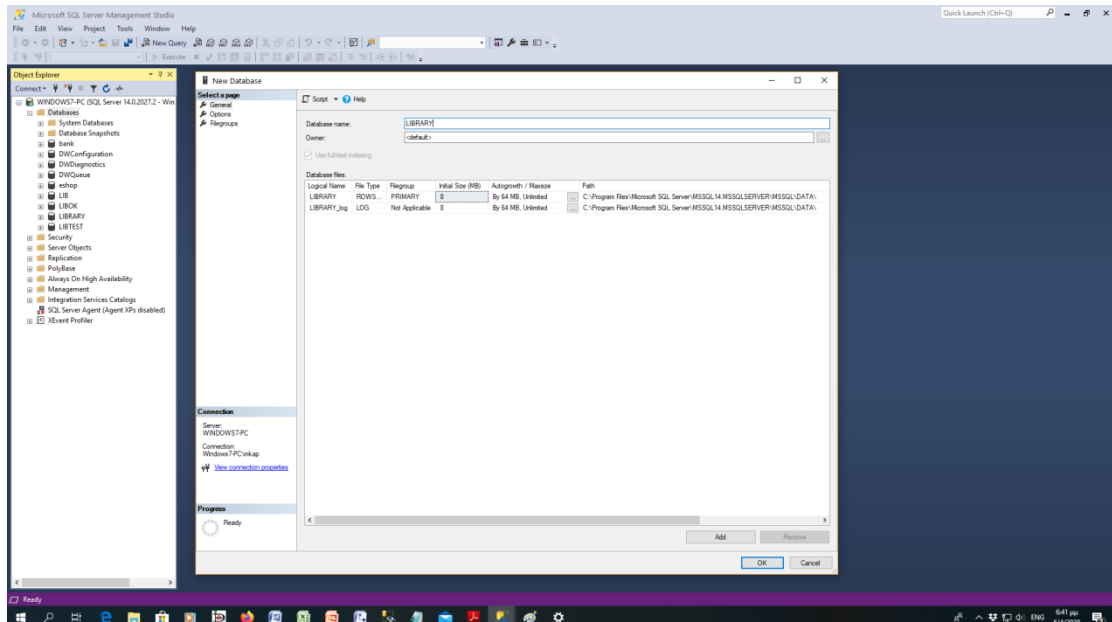
ΣΗΜΕΙΩΣΕΙΣ

1. Μην περιοριστείτε στις δυνατότητες που παρέχει ο SQL SERVER.
2. Η λύση που θα προτείνετε **δεν πρέπει υποχρεωτικά** να βασίζεται (ή να λαμβάνει υπόψη) τους πίνακες **words** και **bibwords**.
3. Θα πρέπει να τεκμηριώσετε την λύση που θα προτείνετε. Αν η λύση δύναται να υλοποιηθεί στον SQL server (ή σε κάποιο άλλο DBMS) να παρουσιάσετε την υλοποίηση.

ΠΑΡΑΡΤΗΜΑ

Δημιουργία της Βάσης LIBRARY

1. Κάνουμε δεξί click στην επιλογή "Databases" του "Object Explorer" και επιλέγουμε "New Database". Στο παράθυρο που εμφανίζεται με τίτλο "New Database" πληκτρολογούμε LIBRARY στο πεδίο "Database name".
2. Στο ίδιο παράθυρο επιλέγουμε "Options" και εμφανίζεται νέο παράθυρο στο οποίο το πρώτο πεδίο είναι το "collation". Επιλέγουμε "Greek_CI_AS".



Περιεχόμενα των SQL scripts "CreateLibrarySchema.sql" και "LoadLibraryData.sql"

---CreateLibrarySchema.sql

```
create table departments
(decode int primary key,
 depname varchar(30)
);

create table publishers
( pubid int primary key,
  pubname VARCHAR(100)
);

create table authors
(aid int primary key,
 author varchar(50)
);

create table sterms
(tid int primary key,
 term varchar(60)
);

create table bibrecs
(bibno int primary key,
 title varchar(200),
 material varchar(30),
 lang char(3),
 place varchar(40),
 series varchar(200),
 pubid int foreign key references publishers(pubid),
);

create table bibauthors
(bibno int foreign key references bibrecs(bibno),
 aid int foreign key references authors(aid),
 primary key(bibno, aid),
);

create table bibterms
(bibno int foreign key references bibrecs(bibno),
 tid int foreign key references sterms(tid),
 primary key (bibno, tid)
);

create table copies
( copyno char(8) primary key,
  bibno int foreign key references bibrecs(bibno),
  copyloc char(3)
);

create table borrowers
(bid int primary key,
 decode int,
 sex char(1),
 bname varchar(60),
);
```



```

CREATE TABLE loanstats
(lid int primary key,
 copyno char(8) foreign key references copies(copyno),
 bid int foreign key references borrowers(bid),
 loandate date
);

---LoadLibraryData.sql

BULK INSERT departments
FROM 'C:\library\DATA\DEPARTMENTS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT publishers
FROM 'C:\library\DATA\PUBLISHERS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT authors
FROM 'C:\library\DATA\AUTHORS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT sterms
FROM 'C:\library\DATA\STERMS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT bibrecs
FROM 'C:\library\DATA\BIBRECS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT bibauthors
FROM 'C:\library\DATA\BIBAUTHORS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT bibterms
FROM 'C:\library\DATA\BIBTERMS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT borrowers
FROM 'C:\library\DATA\BORROWERS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

BULK INSERT copies
FROM 'C:\library\DATA\COPIES.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

SET DATEFORMAT dmy;
BULK INSERT loanstats
FROM 'C:\library\DATA\LOANSTATS.txt'
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',
ROWTERMINATOR = '\n');

```