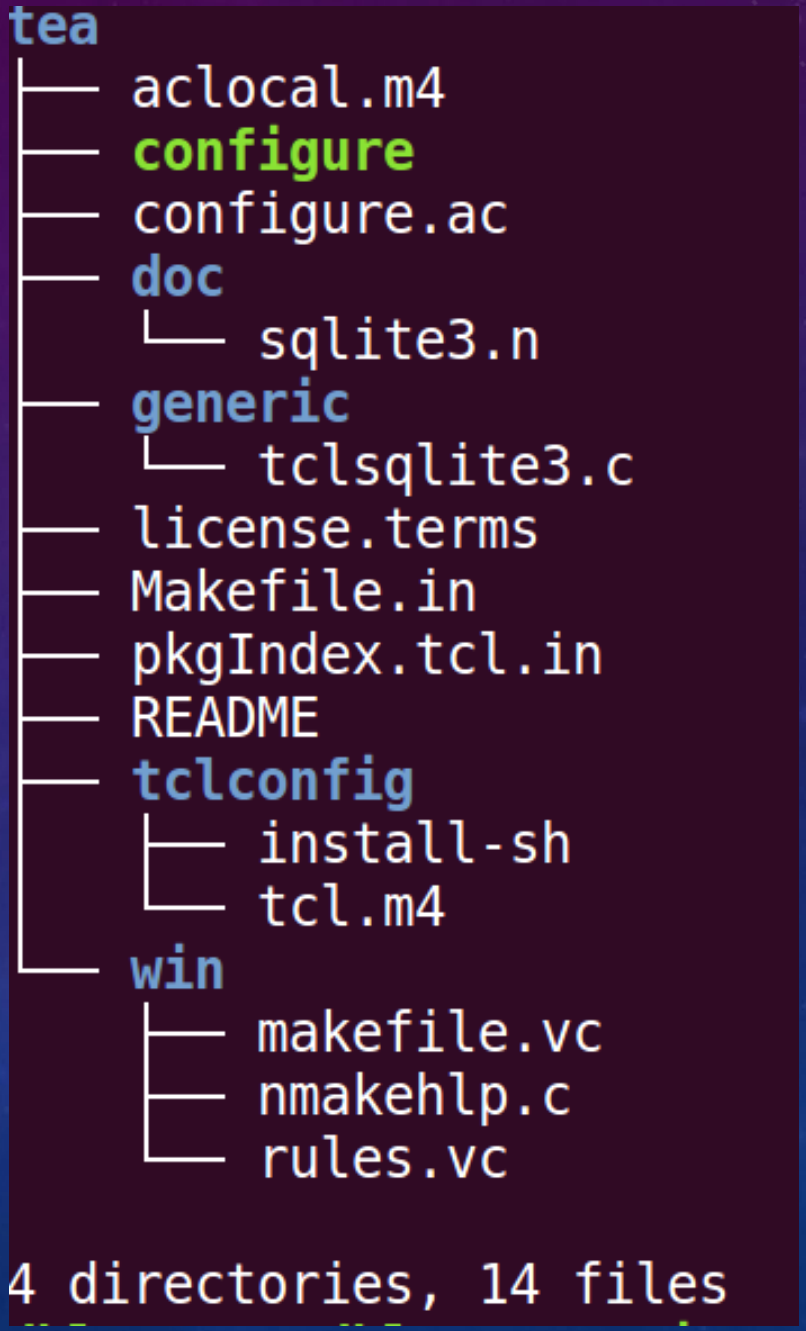
Database System Labs - Project I: Filesystem To DataBase III (Advanced)

李旭东
leexudong@nankai.edu.cn
Nankai University

OBJECTIVES

- Store the info of file system to relational database
  - The structure information of file system
  - Get the structure information of file system
  - Store the info of file system to DB
  - Retrieve the tree info of file system from DB
  - Monitor the operations on specified directory and sync to DB

THE STRUCTURE INFORMATION OF FILE SYSTEM



GET THE STRUCTURE INFORMATION OF FILE SYSTEM USING PYTHON

```
import os
rootDir='/etc/network'
for (dirName, dirs, files) in os.walk(rootDir):
    for fileName in files:
        filePath = os.path.join(dirName, fileName)
        parentFileName=os.path.basename(dirName)
        print(fileName,':',parentFileName,':',os.path.getsize(filePath))
    for dir in dirs:
       print(dir)
print('exit')
```

TEST CASE

- Sqlite
  - http://www.sqlite.org/2018/sqlite-autoconf-3230100.tar.gz
  - tar xvzf sqlite-autoconf-3230100.tar.gz

```
/etc/network
|---interfaces : 82
|---interfaces.d:0
|       |---ethtool : 344
|       |---wpasupplicant : 4696
|       |---wireless-tools : 3839
|---if-pre-up.d:3
|       |---upstart : 1483
|       |---ethtool : 1685
|       |---avahi-autoipd : 923
|       |---wpasupplicant : 4696
|       |---avahi-daemon : 484
|       |---000resolvconf : 817
|---if-up.d:6
|       |---upstart : 332
|       |---avahi-autoipd : 1015
|       |---wpasupplicant : 4696
|       |---resolvconf : 256
|---if-down.d:4
|       |---wpasupplicant : 4696
|       |---avahi-daemon : 484
|       |---wireless-tools : 1070
|---if-post-down.d:3
Total Dirs:5    Total Files:17
```

GET THE STRUCTURE INFORMATION OF FILE SYSTEM USING PYTHON

```
import os
rootDir='/etc/network'
for (dirName, dirs, files) in os.walk(rootDir):
    for fileName in files:
        filePath = os.path.join(dirName, fileName)
        parentFileName=os.path.basename(dirName)
        print(fileName,':',parentFileName,':',os.path.getsize(filePath))
    for dir in dirs:
       print(dir)
print('exit')
```

TEST CASE

- Sqlite
  - http://www.sqlite.org/2018/sqlite-autoconf-3230100.tar.gz
  - tar xvzf sqlite-autoconf-3230100.tar.gz

```
|---README.txt : 3558
|---Replace.cs : 7272
|---Makefile.am : 828
|---sqlite3.1 : 8928
|       |---pkgIndex.tcl.in : 167
|       |---license.terms : 257
|       |---configure.ac : 8308
|       |---Makefile.in : 15902
|       |---configure : 280772
|       |---README : 1338
|       |---aclocal.m4 : 147
|               |---tclsqlite3.c : 117731
|       |---generic:1
|               |---sqlite3.n : 494
|       |---doc:1
|               |---rules.vc : 18743
|               |---nmakehlp.c : 17368
|               |---makefile.vc : 13830
|       |---win:3
|               |---tcl.m4 : 134055
|               |---install-sh : 13868
|       |---tclconfig:2
|---tea:18
Total Dirs:5    Total Files:37
```

STORE THE INFO OF FILE SYSTEM TO DB 1/2

```
#!/usr/bin/python
import sys
import mysql.connector
#mysql config
config = {
    'host': 'localhost',
    'user': 'myuser',
    'password': 'mypwd',
    'port': 3306,
    'database': 'dbsclab2018',
    'charset': 'utf8'
}
```

```
try:
    cnn = mysql.connector.connect(**config)
except mysql.connector.Error as e:
    cnn = None
    print('connect fails!{}'.format(e))
if None==cnn:
   sys.exit()
```

STORE THE INFO OF FILE SYSTEM TO DB 2/2

```
cursor = cnn.cursor()
sql= 'SELECT name,dept_name,salary from instructor'
try:
    cursor.execute(sql)
    # display the result
    for name,dept_name,salary in cursor:
        print(name.decode('utf-8'),',',dept_name,salary)
except mysql.connector.Error as e:
    print('query error!{}'.format(e))
finally:
    cursor.close()
    cnn.close()
```

RETRIEVE THE TREE INFO OF FILE SYSTEM FROM DB

- Display the structure of tables
- Display the tree of specified directory
  - File size
  - Subdir count

```
/etc/network
|---interfaces : 82
|---interfaces.d:0
|       |---ethtool : 344
|       |---wpasupplicant : 4696
|       |---wireless-tools : 3839
|---if-pre-up.d:3
|       |---upstart : 1483
|       |---ethtool : 1685
|       |---avahi-autoipd : 923
|       |---wpasupplicant : 4696
|       |---avahi-daemon : 484
|       |---000resolvconf : 817
|---if-up.d:6
|       |---upstart : 332
|       |---avahi-autoipd : 1015
|       |---wpasupplicant : 4696
|       |---resolvconf : 256
|---if-down.d:4
|       |---wpasupplicant : 4696
|       |---avahi-daemon : 484
|       |---wireless-tools : 1070
|---if-post-down.d:3
Total Dirs:5    Total Files:17
```

MONITOR THE OPERATIONS ON SPECIFIED DIRECTORY AND SYNC TO DB

- Python for windows
  - win32file and win32con
- Python for linux
  - pyinotify

QUIZ

- …

LAB 2:

针对某一个硬盘分区（例如E☺，导入到数据库中，并做如下统计：

- 1.一共有多少普通文件，多少目录夹
- 2.所有普通文件及目录共占用了多少字节、所占用空间（数据块单位512字节）
- 3.统计最近1周内修改的普通文件的文件名称、文件大小、以及文件总数
- 4.统计直接包含100普通文件以上的目录的目录名称以及文件数量
- 5.统计目录深度大于10层的目录名称
- 6.统计所有只读的普通文件数量、所占字节数
- 7.统计所有文件的的大小，按照1KB，512KB，1MB，10MB，512MB，1GB，2GB，4GB，以及以上统计文件数量

©LXD

THANKS!

leexudong@nankai.edu.cn