

# pandas数据加载

Text  
CSV  
Excel  
HTML  
MySQL  
MongoDB

## Text

```
In [ ]: pd.read_table('a.txt')
```

```
In [ ]: pd.read_table('a.txt', sep=':')
```

```
In [ ]: pd.read_table('a.txt', sep=':', header=None)
```

```
In [ ]: pd.read_table('a.txt', sep=':', names=['name', 'pwd', 'uid', 'gid', 'local', 'home', 'shell'])
```

## CSV

```
In [ ]: pd.read_csv('data.csv')
```

## Excel

```
In [ ]: pd.read_excel('data.xlsx')
```

在使用上面这条语句的时候，需要先安装 lxr 模块: `pip3 install lxr`

## HTML

```
In [ ]: tables = pd.read_html('data.html', head=0)    # 其中的 header=0指用第一行做为head头
```

```
In [ ]: tables[0]
```

```
In [ ]: len(tables)
```

使用 read\_html() 前需要安装 lxml 模块: `pip3 install lxml`

## MySQL

使用之前需要安装 pymysql

```
In [3]: import pymysql
```

```
In [ ]: conn = pymysql.connect(host='159.65.130.172', user='root', password='root', db='test', 'charset'='utf8') #这里charset不要写utf-8, 而是utf8
```

```
In [ ]: sql = 'select * from product'
data = pd.read_sql(sql, conn)
conn.close()
```

```
In [ ]: data
```

## MongoDB

使用之前需要安装 pymongo

```
In [4]: import pymongo
```

```
In [5]: client = pymongo.MongoClient('127.0.0.1', 27017) # mongodb的默认端口号为 27017
```

```
In [6]: client
```

```
Out[6]: MongoClient(host=['127.0.0.1:27017'], document_class=dict, tz_aware=False, connect=True)
```

```
In [ ]: db = 'film'
table = 'film'
cursor = client[db][table].find()
```

```
In [ ]: data = pd.DataFrame(list(cursor))
```

In [ ]: data