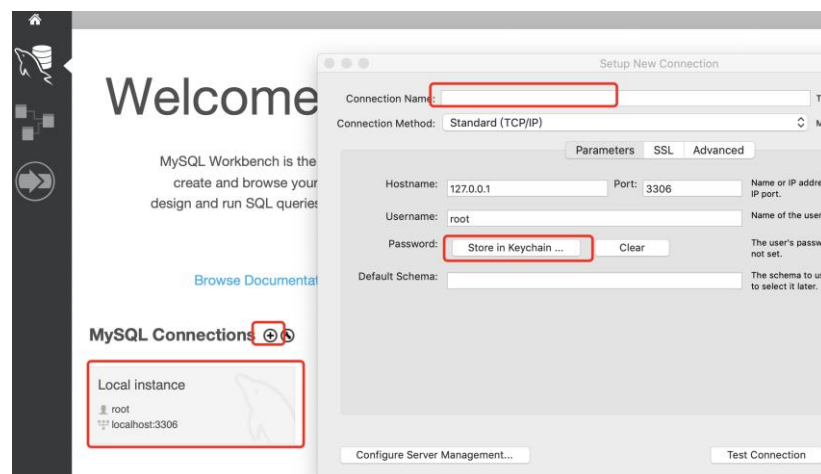


说明：本代码为 2020 年 4 月可运行代码，仅作为操作流程参考，由于平台改版及反爬加强，有需要的同学可自行修改代码进行数据采集。

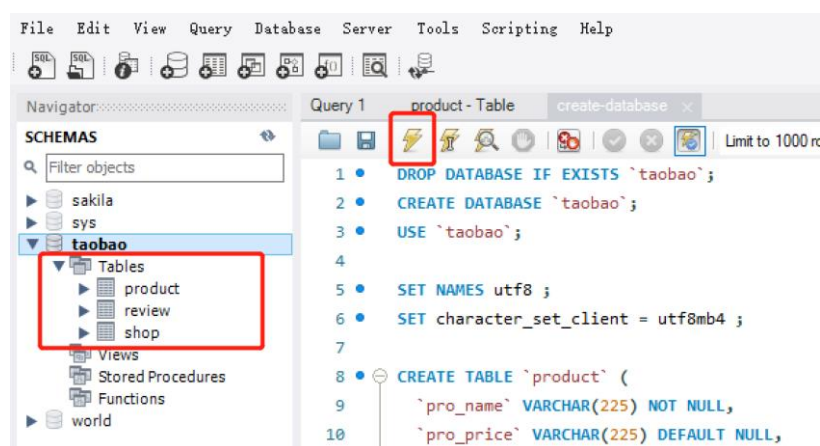
一、淘宝爬取步骤：

1、新增数据库和表单

(1) 进入连接，如果没有连接则需新建连接（点击“+”，输入命名并在“Store in Keychain”里输入密码（新建数据库中设置的密码）即可）



(2) 新建数据库&表：File-Open SQL Script Serve-选择 SQL 文件“create-database”-运行（闪电符号，在左侧刷新后即可显示“taobao”数据库和“product” / “review” / “shop”表）



2、修改：

(1) webdriver 改绝对位置（即 webdriver 存储的位置）

（taobaocookie：第 23 行；taobao：第 25 行）

(2) 数据库中的密码

（taobao：第 464 行）

3、新建 url.csv（将需爬取的链接复制在 url.csv 中）和 output.csv

4、运行 taobaocookie（注：不要打开 pickle 文件，否则可能出现乱码）和 taobao

5、爬取过程中可能出现的问题：

(1) 爬取中断（如断网、报错、爬取限制等），都需要在 url.csv 中删除已经爬取的链接，否则会出现重复爬取并报错

(2) 出现验证滑块：关闭网页，半小时后再试

(3) 显示“限制了第三方 Cookie”：

<https://www.cnblogs.com/Summer6/p/11671204.html>，如果还是不行，关闭网页重新运行即可

(4) 产品下架无法显示信息/非淘宝页面：关闭网页即可（该链接会自动存储到 output 文件中），若失效数据多于 5 组须联系助教补数据

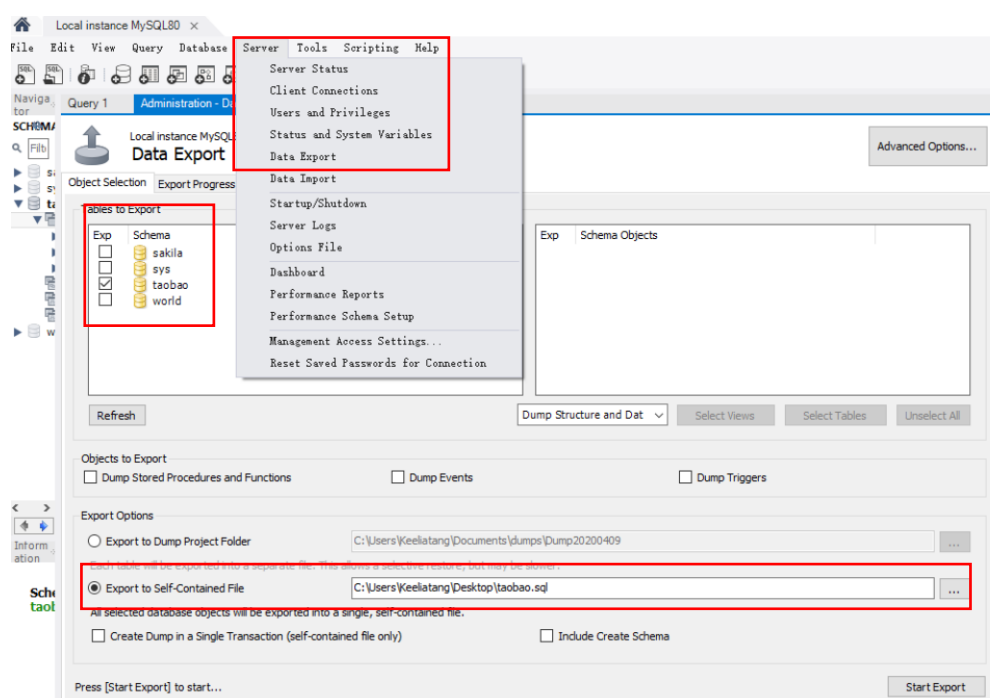
(5) 爬取过慢或过快：修改等待时间（taobao：第 526 行）或者间隔爬取（比如间隔 5 分钟或 10 分钟进行爬取，注意 taobaocookie 有有效时长限制，故需在半小时内进行下一个链接的爬取。这种方式会稍微麻烦一些，但是能够有效避免滑块的出现）或者换账号登录

(6) 爬取时仍提示登录，删除 TaobaoCookies，运行 taobaocookie 重新登录即可（未进行爬取时，cookie 的有效时长约半小时，超过即需重新登录）

(7) 如果报错出现违反“PRIMARY KEY”限制，删除当前爬取的 url 即可

6、转储为 SQL 文件：

Server-Data Export-选择数据库（taobao）-选择 Export to Self-Contained File-Start Export



1、根据 Sheet1 中给出的 50 个链接进行淘宝数据爬取

2、将数据分别以 SQL 格式（爬取数据）和 CSV 格式（output 输出数据）进行存储，并比较两者的不同

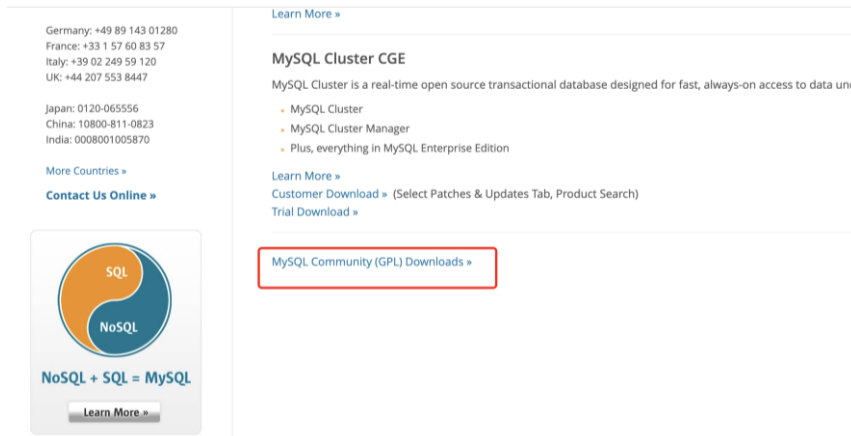
3、根据 Sheet2 中 50 个链接进行淘宝数据爬取，并对 100 组店铺/产品/评论数据（若无效数据多于 5 组须联系助教补数据）进行自由探索分析，形成数据分析报告

二、MySQL & MySQL Workbench 的下载与使用（社区版）

（链接：<https://www.mysql.com/downloads/>）

1、MySQL & MySQL Workbench（Windows 版本）：

（1）进入社区版：滑动到最底端（进入 MySQL Community（GPL）Downloads）



（2）进入 MySQL Community Server

⊙ MySQL Community Downloads

- MySQL Yum Repository
- MySQL APT Repository
- MySQL SUSE Repository
- **MySQL Community Server**
- MySQL Cluster
- MySQL Router
- MySQL Shell
- MySQL Workbench
- MySQL Installer for Windows
- MySQL for Excel
- MySQL for Visual Studio
- MySQL Notifier
- C API (libmysqlclient)
- Connector/C++
- Connector/J
- Connector/NET
- Connector/Node.js
- Connector/ODBC
- Connector/Python
- MySQL Native Driver for PHP
- MySQL Benchmark Tool
- Time zone description tables
- Download Archives

（3）下载 **Recommended Download 版本（务必注意，不要下错版本）**

Recommended Download:

MySQL Installer for Windows

All MySQL Products. For All Windows Platforms. In One Package.

Starting with MySQL 5.6 the MySQL Installer package replaces the standalone MSI packages

Windows (x86, 32 & 64-bit), MySQL Installer MSI [Go to Download Page >](#)

Other Downloads:

Windows (x86, 64-bit), ZIP Archive (mysql-8.0.19-winx64.zip)	8.0.19	187.8M	Download
		MD5: f52c52e7b499958acc5f08ce0a869cab Signature	
Windows (x86, 64-bit), ZIP Archive Debug Binaries & Test Suite (mysql-8.0.19-winx64-debug-test.zip)	8.0.19	406.7M	Download
		MD5: 9b885558e74cc4629aef77cbe75d11631 Signature	

(4) 下载第一个项目

MySQL Community Downloads

MySQL Installer

[General Availability \(GA\) Releases](#) [Archives](#) [Download](#)

MySQL Installer 8.0.19

Select Operating System:

[Looking for previous GA versions?](#)

Windows (x86, 32-bit), MSI Installer (mysql-installer-web-community-8.0.19.0.msi)	8.0.19	18.6M	Download
		MD5: 32043776cb22394b45f4daa864c0a361 Signature	
Windows (x86, 32-bit), MSI Installer (mysql-installer-community-8.0.19.0.msi)	8.0.19	398.9M	Download
		MD5: 1a882015da7f993f20c4717a63b6817c Signature	

(5) 跳过注册即可下载

MySQL Community Downloads

Login Now or Sign Up for a free account.

An Oracle Web Account provides you with the following advantages:

- Fast access to MySQL software downloads
- Download technical White Papers and Presentations
- Post messages in the MySQL Discussion Forums
- Report and track bugs in the MySQL bug system

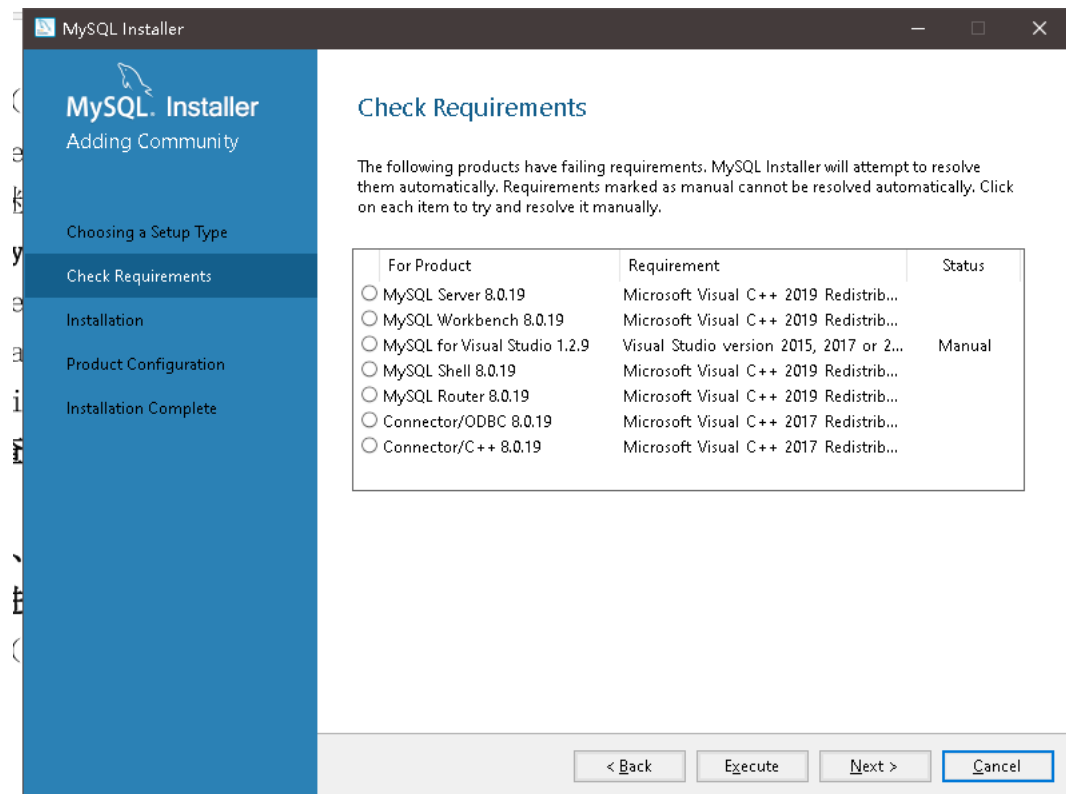
[Login »
using my Oracle Web account](#) [Sign Up »
for an Oracle Web account](#)

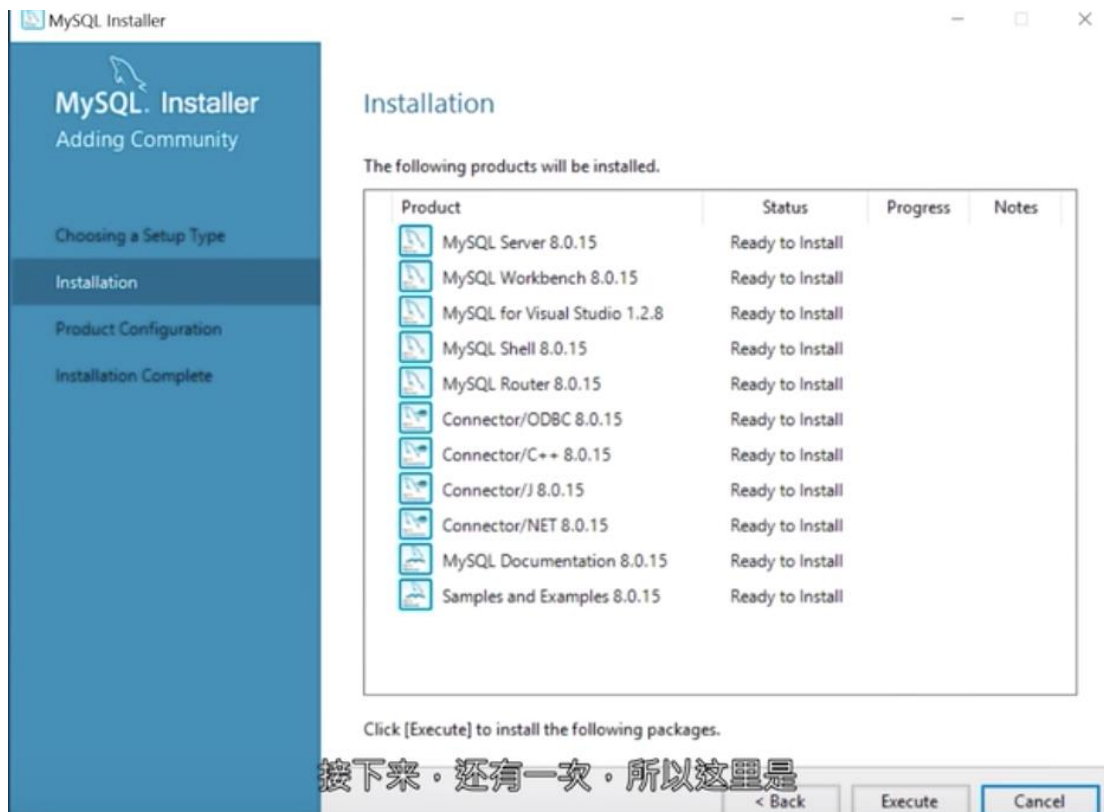
MySQL.com is using Oracle SSO for authentication. If you already have an Oracle Web account, click the Login link. Otherwise, you can signup for a free account by clicking the Sign Up link and following the instructions.

[No thanks, just start my download.](#)

(6) 安装：除了两次输入密码，基本点击继续即可完成安装

Next (Developer Default) -Next-~~(Check Requirements: 如果存在警告, 点击继续即可; 如果没有则无需这一步)~~-Execute (**Installation: 检查是否存在 MySQL Workbench**) -Next-Next (Group Replication: Standalone MySQL) -Next (Type and Networking) -Next (Authentication Method: Use Strong Password Encryption) -**输入密码**-Next (Windows Service) -Execute-Finish-Next-Finish-**输入之前设置的密码 (Connect To Serve)** -**Check (可检查是否连接成功)** -Next-Execute-Finish-Next-Finish

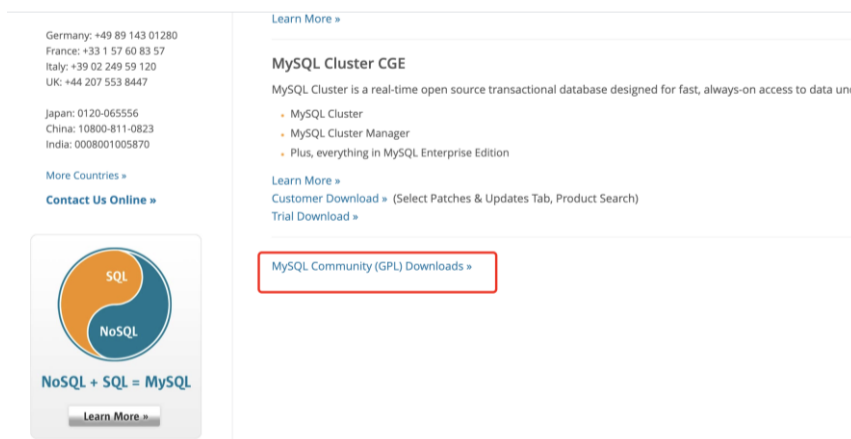




2、MySQL (MAC 版本):

进入 downloads 链接 (<https://www.mysql.com/downloads/>)

(1) 进入社区版: 滑动到最底端 (进入 MySQL Community (GPL) Downloads)



(2) 进入 MySQL Community Server

MySQL Community Downloads

- MySQL Yum Repository
- MySQL APT Repository
- MySQL SUSE Repository
- MySQL Community Server**
- MySQL Cluster
- MySQL Router
- MySQL Shell
- MySQL Workbench
- MySQL Installer for Windows
- MySQL for Excel
- MySQL for Visual Studio
- MySQL Notifier
- C API (libmysqlclient)
- Connector/C++
- Connector/J
- Connector/NET
- Connector/Node.js
- Connector/ODBC
- Connector/Python
- MySQL Native Driver for PHP
- MySQL Benchmark Tool
- Time zone description tables
- Download Archives

(3) 下载第一个项目

General Availability (GA) Releases Archives

MySQL Community Server 8.0.19

Select Operating System: **macOS** [Looking for previous GA versions?](#)

ⓘ Packages for Catalina (10.15) are compatible with Mojave (10.14)

macOS 10.15 (x86, 64-bit), DMG Archive (mysql-8.0.19-macos10.15-x86_64.dmg)	8.0.19	376.8M	Download
macOS 10.15 (x86, 64-bit), Compressed TAR Archive (mysql-8.0.19-macos10.15-x86_64.tar.gz)	8.0.19	159.0M	Download
macOS 10.15 (x86, 64-bit), Compressed TAR Archive Test Suite (mysql-test-8.0.19-macos10.15-x86_64.tar.gz)	8.0.19	221.2M	Download

(4) 跳过注册即可下载

MySQL Community Downloads

Login Now or Sign Up for a free account.

An Oracle Web Account provides you with the following advantages:

- Fast access to MySQL software downloads
- Download technical White Papers and Presentations
- Post messages in the MySQL Discussion Forums
- Report and track bugs in the MySQL bug system

[Login »](#)
using my Oracle Web account

[Sign Up »](#)
for an Oracle Web account

MySQL.com is using Oracle SSO for authentication. If you already have an Oracle Web account, click the Login link. Otherwise, you can sign up for a free account by clicking the Sign Up link and following the instructions.

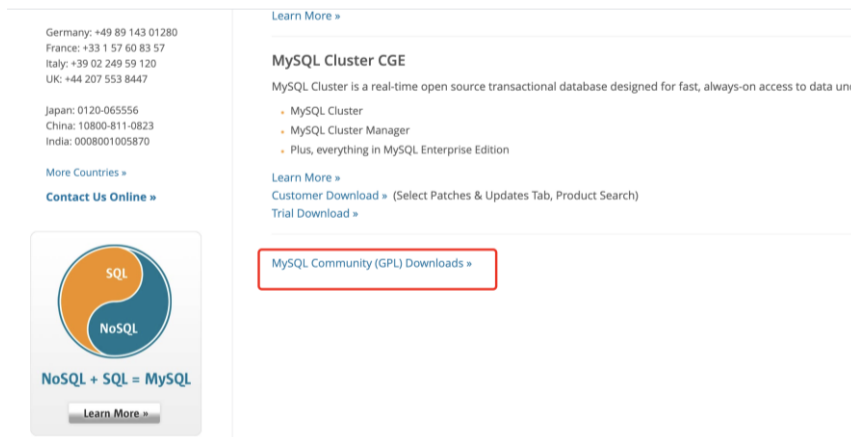
[No thanks, just start my download.](#)

(5) 安装：一直点击继续即可安装

(6) 配置（仍然在安装界面，这里需要设置 MySQL 的密码）：选择 Use Strong Password Encryption-设置密码即可

3、MySQL Workbench (MAC 版本)：需返回到初始链接的 downloads 界面

(1) 进入社区版：滑动到最底端（进入 MySQL Community (GPL) Downloads）

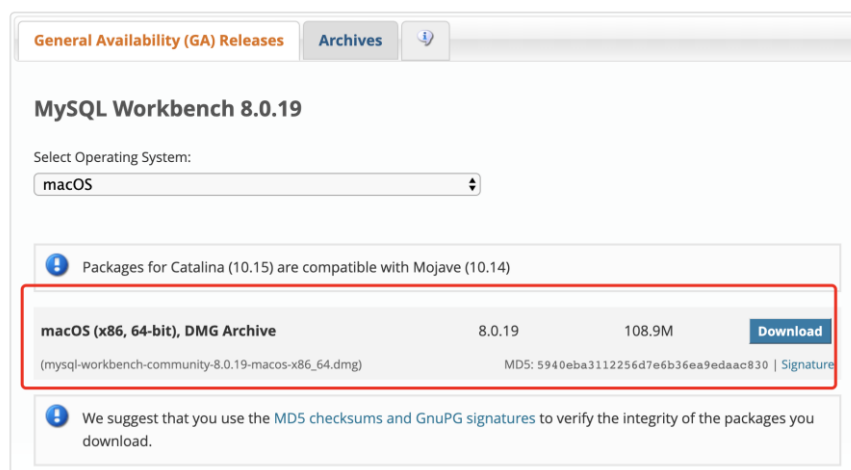


(2) 进入 MySQL Workbench

MySQL Community Downloads

- MySQL Yum Repository
- MySQL APT Repository
- MySQL SUSE Repository
- MySQL Community Server
- MySQL Cluster
- MySQL Router
- MySQL Shell
- MySQL Workbench
- MySQL Installer for Windows
- MySQL for Excel
- MySQL for Visual Studio
- MySQL Notifier
- C API (libmysqlclient)
- Connector/C++
- Connector/J
- Connector/NET
- Connector/Node.js
- Connector/ODBC
- Connector/Python
- MySQL Native Driver for PHP
- MySQL Benchmark Tool
- Time zone description tables
- Download Archives

(3) 下载第一个项目



(4) 跳过注册即可下载，下载完成后将其复制到应用程序中即下载完成

MySQL Community Downloads

Login Now or Sign Up for a free account.

An Oracle Web Account provides you with the following advantages:

- Fast access to MySQL software downloads
- Download technical White Papers and Presentations
- Post messages in the MySQL Discussion Forums
- Report and track bugs in the MySQL bug system

Login »

using my Oracle Web account

Sign Up »

for an Oracle Web account

MySQL.com is using Oracle SSO for authentication. If you already have an Oracle Web account, click the Login link. Otherwise, you can sign up for a free account by clicking the Sign Up link and following the instructions.

No thanks, just start my download.