



# 生物信息学编程基础

# 第一节 Linux操作系统

## 一、Linux简介

## 二、Linux常用命令行操作

表 11-1 Linux 与 DOS 常用命令比较

Linux 命令符	DOS 命令符	功能
ls	dir	显示文件及目录
cd	cd	改变当前目录
mkdir	md	新建目录
rmdir	rd	删除目录
rm	del	删除文件
mv	move	移动文件
cp	copy	复制文件
less	type	阅读文件
echo	echo	输出字符串

# 第二节 生物信息学中的编程语言

## 一、Perl

### 1、Perl简介



The screenshot shows the official Perl Programming Language website. The header features the Perl logo (a camel) and the text "The Perl Programming Language" with the URL "www.perl.org". A navigation bar includes links for HOME, LEARN, DOCUMENTATION, CPAN, COMMUNITY, GET INVOLVED, DOWNLOAD, and ABOUT PERL. The main content area has a large heading "Flexible & Powerful" with the subtext "That's why we love Perl 5" and a "Get started" link. A prominent "DOWNLOAD PERL" button is also visible. Below this, a paragraph states: "Perl 5 is a highly capable, feature-rich programming language with over 25 years of development. [More about why we love Perl...](#)". The page is organized into several sections: "Learning Perl 5" (with a graduation cap icon) describing free online books and a large developer community; "The Perl Community" (with a group of people icon) mentioning an active world-wide community with over 300 local groups; "Documentation" (with a document icon) pointing to core documentation, FAQs, and translations; and "Contribute to Perl" (with a gear icon) stating that Perl is actively developed and many ways exist to get involved. On the right side, there are additional sections: "Current Perl version" showing "5.18.1 - download now"; "Find out more" with links to Learn, Documentation, Community, and Events; a "Tip" section about database interfaces using DBIx::Class; and a "Sponsor" section for "phyber" (Networking and Colocation).

The Perl Programming Language www.perl.org

HOME LEARN DOCUMENTATION CPAN COMMUNITY GET INVOLVED DOWNLOAD ABOUT PERL

### Flexible & Powerful

That's why we love Perl 5

[Get started](#) [DOWNLOAD PERL](#)

Perl 5 is a highly capable, feature-rich programming language with over 25 years of development. [More about why we love Perl...](#)



#### Learning Perl 5

With free online books, over 25,000 extension modules, and a large developer community, there are many ways to learn Perl 5.



#### The Perl Community

Perl has an active world wide community with over 300 local groups, mailing lists and support/discussion websites.



#### Documentation

Core documentation, FAQs and translations.



#### Contribute to Perl

Perl is being [actively developed](#). There are many ways to [get involved](#).

#### Current Perl version


[5.18.1 - download now](#)

#### Find out more

[Learn](#)  
[Documentation](#)  
[Community](#)  
[Events](#)

#### Tip

**Database interfaces**  
[DBIx::Class](#) provides an Object Relational Mapper (ORM) to databases (e.g. Oracle, SQL Server, MySQL, Postgress, Access etc)

**Sponsor**  
  
Networking and Colocation

## 2、Perl基础

### 数据类型：

字符串(\$string)、数组(@array)、哈希表(%hash)

### 正则表达式

m//;

### 子程序

```
Sub SUBNAME  
{  
    code;  
}
```

# 包

```
Package package1;
```

---

## 内置函数

abs、sqrt、exp

## 文件处理

```
open(FH, "path../filename");
```

## 模块

LWP::Simple

LWP::UserAgent

DBI

# 3、BioPerl概况

BioPerl是Perl的扩充，专门用于生物信息的工具与函数集。

表 11-3    BioPerl 模块说明


应用	模块
1. 序列操作	
对序列进行统计	SeqStats, SeqWord
确定限制性内切核酸酶位点	Bio::Restriction
识别氨基酸裂解位点	Sigcleave
多样性序列功能	OddCodes, SeqPattern
转换坐标系	Coordinate::Pair, RelSegment
2. 搜索相似序列	
运行远程 BLAST	using RemoteBlast, pm
分析 BLAST 和 FASTA 结果	Search, SearchIO
分析 BLAST 结果	BPlite, BPpsilite, and BPbl2seq
分析 HMM 结果	HMMER::Results, SearchIO
运行本地 BLAST	StandAloneBlast
3. 序列比对	SimpleAlign
4. 在基因组 DNA 寻找基因和其他结构	Genscan, Sim4, Grail, Genemark, ESTScan, MZEF, EPCR
5. 开发机器可读性的序列注释	
表征序列注释	SeqFeature, RichSeq, Location
表征序列注释	Annotation::Collection
表示大的序列	LargeSeq
表示改变的序列	LiveSeq

应用	模块
表示相关联的序列——突变体,多态性	Allele, SeqDiff
在序列注释中加入特征数据	SeqWithQuality
序列 XML 表征——生成和分析	SeqIO::game, SeqIO::bsml
使用 GFF 表示序列	Bio::DB::GFF
6. 操作序列簇	Cluster, ClusterIO
7. 在 BioPerl 中表示非序列数据:结构、树和图谱	
使用 3D 结构对象并读取 PDB 文件	StructureI, Structure::IO
树对象和系统发生树	Tree::Tree, TreeIO, PAML
图谱对象操作遗传图谱	Map::MapI, MapIO
文献对象搜索文献数据库	Biblio
图形对象以图形代表序列对象	Graphics



## 二、Python

### 1、Python简介



python™

ABOUT >>

NEWS >>

DOCUMENTATION >>

DOWNLOAD >>

下載 >>

COMMUNITY >>

FOUNDATION >>

CORE DEVELOPMENT >>

Help

Package Index

Quick Links (2.7.5)

- >> Documentation
- >> Windows Installer
- >> Source Distribution

Quick Links (3.3.2)

- >> Documentation
- >> Windows Installer
- >> Source Distribution

Python Jobs

## Python Programming Language – Official Website

**Python is a programming language that lets you work more quickly and integrate your systems more effectively. You can learn to see almost immediate gains in productivity and lower maintenance costs.**

Python runs on Windows, Linux/Unix, Mac OS X, and has been ported to the Java and .NET virtual machines.

Python is free to use, even for commercial products, because of its OSI-approved [open source license](#).

New to Python or choosing between Python 2 and Python 3? Read [Python 2 or Python 3](#).

The [Python Software Foundation](#) holds the intellectual property rights behind Python, underwrites the [PyCon conference](#), and funds many of the Python community.

[Read more](#), -or- [download Python now](#)

» **Python 3.4.0 alpha 4 has been released**  
The fourth and final alpha for Python 3.4, [Python 3.4.0a4](#), has been released.  
Published: Sun, 20 October 2013, 17:30 -0700

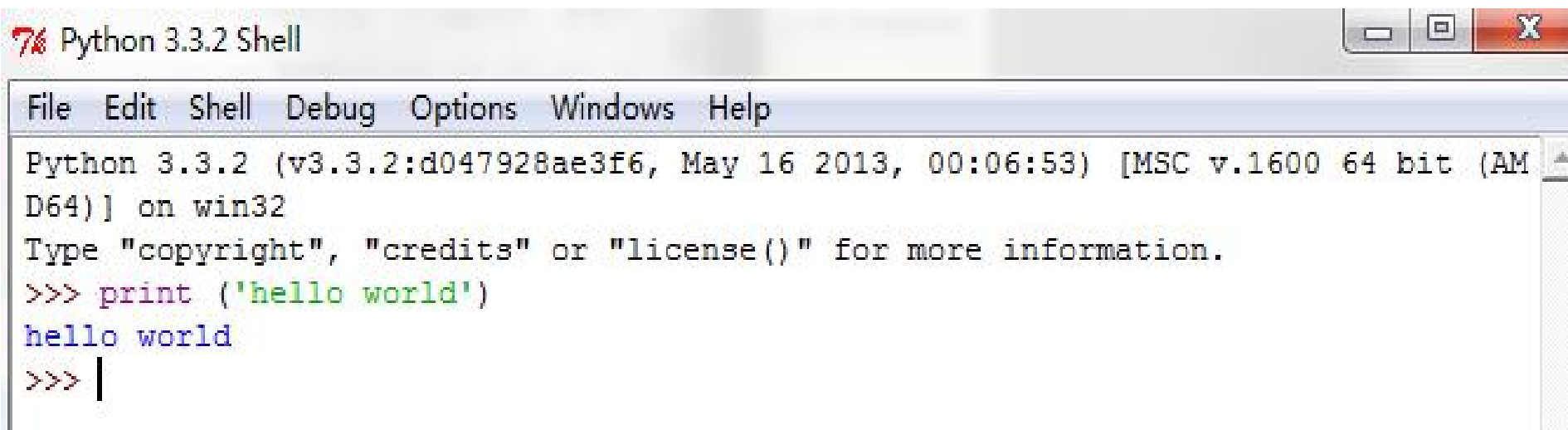
» **Python 2.6.9 release candidate 1 has been released**  
The first release candidate for Python 2.6.9, [Python 2.6.9rc1](#), has been released.  
Published: Mon, 30 September 2013, 21:12 -0400

» **2014 PyTennessee Conference CFP**



## 2、Python安装及开发运行环境的搭建

Windows:



```
Python 3.3.2 Shell
File Edit Shell Debug Options Windows Help
Python 3.3.2 (v3.3.2:d047928ae3f6, May 16 2013, 00:06:53) [MSC v.1600 64 bit (AMD64)] on win32
Type "copyright", "credits" or "license()" for more information.
>>> print ('hello world')
hello world
>>> |
```

Linux:

```
galaxy@mchen96G:~$ python
Python 2.7.3 (default, Jul 5 2013, 08:39:51)
[GCC 4.6.3] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>> print 'hello world'
hello world
>>> █
```

# Python集成开发环境简介

---

IDLE

Ipython

Eclipse+pydev

Ulipad

Pyscripter

Komodo和Komodo Edit

WingIDE

### 3、Python在生物信息学中的应用

---

#### (1) Python中生物信息学常用的模块

NumPy、SciPy、Rpy、Matplotlib

#### (2) 基于Python语言编写的生物信息常用程序

PyMOL、Modeller

### 三、Java

---

java是一种面向对象的、分布式的、健壮的、平台无关的、安全可靠的、解释的、高性能的、多线程的、动态的语言。

# 第三节 SQL及数据库编程

---

## 一、数据库和数据库技术

### 1、数据库

### 2、生物数据库

NCBI、KEGG、PDB...

### 3、数据库技术支持

---

#### 技术要素：

RDBMS数据库程序（如MS Access、SQL Server、MySQL）

服务器端脚本语言（如PHP或ASP）

SQL

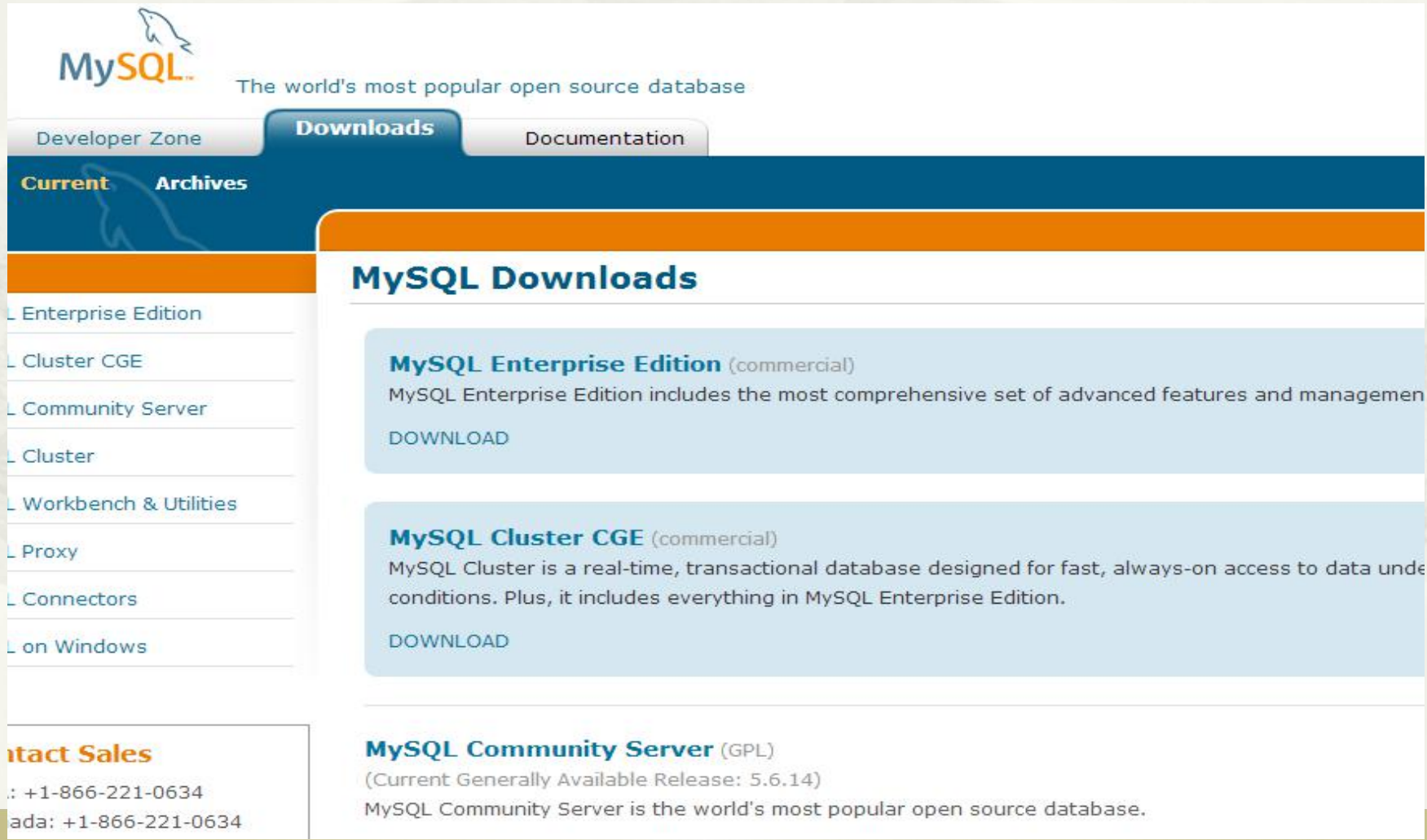
HTML/CSS



## 二、MySQL

### 1、MySQL安装和客户端

开源 (<http://www.mysql.com/>)



The screenshot shows the MySQL website's 'Downloads' section. At the top, the MySQL logo is followed by the tagline 'The world's most popular open source database'. Below this are three tabs: 'Developer Zone', 'Downloads' (which is active), and 'Documentation'. Under the 'Downloads' tab, there are two sub-tabs: 'Current' and 'Archives'. A left sidebar lists various MySQL products: Enterprise Edition, Cluster CGE, Community Server, Cluster, Workbench & Utilities, Proxy, Connectors, and on Windows. The main content area is titled 'MySQL Downloads' and features three product cards. The first card is for 'MySQL Enterprise Edition (commercial)', describing it as having the most comprehensive set of advanced features and management tools, with a 'DOWNLOAD' link. The second card is for 'MySQL Cluster CGE (commercial)', describing it as a real-time, transactional database designed for fast, always-on access to data under various conditions, also with a 'DOWNLOAD' link. The third card is for 'MySQL Community Server (GPL)', noting the current release is 5.6.14 and stating it is the world's most popular open source database. At the bottom left, there is a 'Contact Sales' section with phone numbers for the US (+1-866-221-0634) and Canada (+1-866-221-0634).

MySQL  
The world's most popular open source database

Developer Zone Downloads Documentation

Current Archives

MySQL Downloads

**MySQL Enterprise Edition** (commercial)  
MySQL Enterprise Edition includes the most comprehensive set of advanced features and management tools.  
DOWNLOAD

**MySQL Cluster CGE** (commercial)  
MySQL Cluster is a real-time, transactional database designed for fast, always-on access to data under various conditions. Plus, it includes everything in MySQL Enterprise Edition.  
DOWNLOAD

**MySQL Community Server** (GPL)  
(Current Generally Available Release: 5.6.14)  
MySQL Community Server is the world's most popular open source database.

Contact Sales  
US: +1-866-221-0634  
Canada: +1-866-221-0634

```
管理员: C:\Windows\system32\cmd.exe - mysql -uroot -p

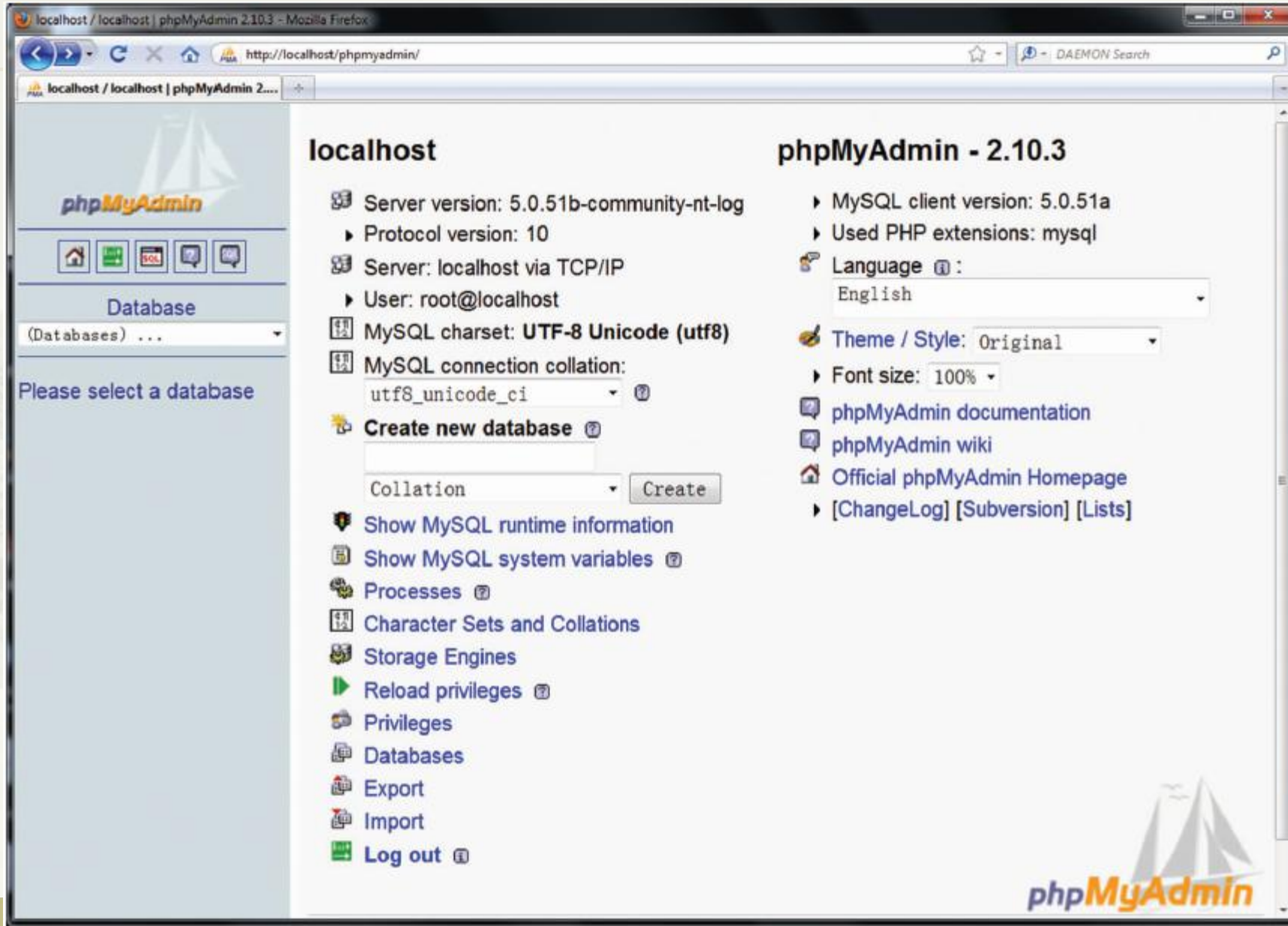
C:\Users\Junjie>mysql -uroot -p
Enter password: 
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 12
Server version: 5.1.37-community MySQL Community Server (GPL)

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

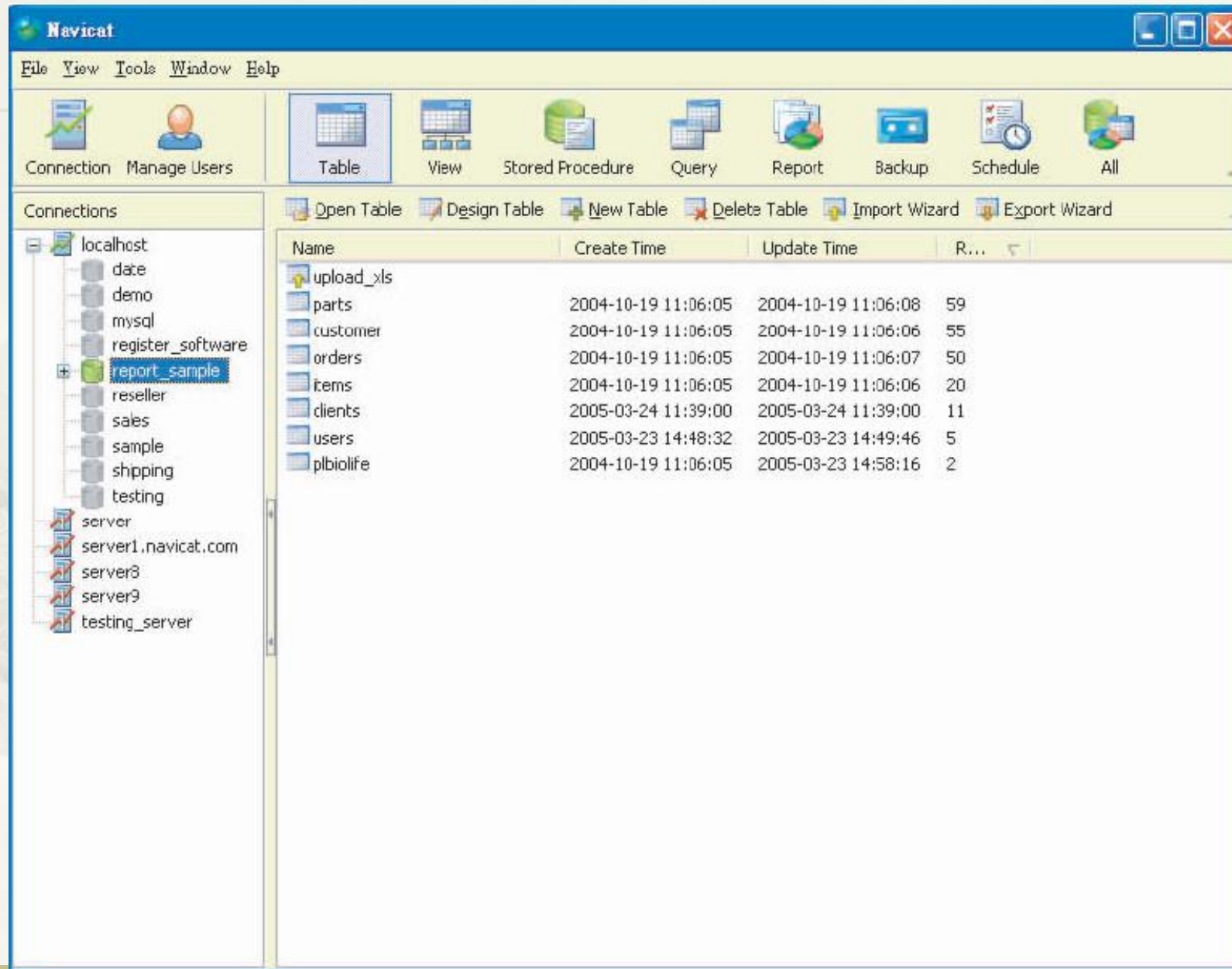
mysql> _
```

半:

## (2) phpMyAdmin



### (3) Navicat



## 2、基本概念

### (1) 表、行、列与列类型

---

### (2) 键

键是特殊的列，通过它可以实现实体间的链接。主键（primary key）是某些规则必须遵守的唯一标识符。

### (3) 表关系

one-to-one、one-to-many、many-to-many



### 3、数据库设计与范式

---

数据库设计，就是建立数据库的结构，又称数据建模（data modeling）。

规范化，就是在数据库设计当中要遵循某些规则，这些规则被称为范式（normal form）。

1NF、2NF、3NF



## 4、数据库操作

Create

Insert

Select

Drop

Delete

Update

Alter

```
管理员: C:\Windows\system32\cmd.exe - mysql -uroot -p

mysql> CREATE DATABASE db_paper;
Query OK, 1 row affected (0.06 sec)

mysql> USE db_paper;
Database changed

mysql> CREATE TABLE journal (
  -> id INT UNSIGNED NOT NULL AUTO_INCREMENT,
  -> journal_title VARCHAR(50) NOT NULL,
  -> issn VARCHAR(20) NOT NULL,
  -> total_cites MEDIUMINT NOT NULL,
  -> impact_factor FLOAT NOT NULL,
  -> PRIMARY KEY (id)
  -> );
Query OK, 0 rows affected (0.47 sec)

mysql> CREATE TABLE paper(
  -> id INT UNSIGNED NOT NULL AUTO_INCREMENT,
  -> title VARCHAR(100) NOT NULL,
  -> journal_id INT UNSIGNED NOT NULL,
  -> PRIMARY KEY (id)
  -> );
Query OK, 0 rows affected (0.12 sec)

mysql> SHOW TABLES;
+-----+
| Tables_in_db_paper |
+-----+
| journal             |
| paper               |
+-----+
2 rows in set (0.02 sec)


mysql> EXPLAIN journal;
+-----+-----+-----+-----+-----+-----+
| Field      | Type                | Null | Key | Default | Extra          |
+-----+-----+-----+-----+-----+-----+
| id         | int(10) unsigned    | NO   | PRI | NULL    | auto_increment |
| journal_title | varchar(50)         | NO   |     | NULL    |                |
| issn       | varchar(20)         | NO   |     | NULL    |                |
| total_cites | mediumint(9)        | NO   |     | NULL    |                |
| impact_factor | float              | NO   |     | NULL    |                |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.06 sec)

mysql> _ 半:
```

# 三、PHP

## 1、PHP简介

实用性、可选择性、开源与低成本



[downloads](#) | [documentation](#) |

---

### What is PHP?

PHP is a widely-used general-purpose scripting language that is especially suited for Web development and can be embedded into HTML. If you are new to PHP and want to get some idea of how it works, try the [introductory tutorial](#). After that, check out the online [manual](#).

Ever wondered how popular PHP is? See the [Netcraft Survey](#).

### Thanks To

[easyDNS](#)  
[Directi](#)  
[pair Networks](#)  
[Server Central](#)

Upcoming conferences: [Ski PHP 2014](#) [Madison PHP Conference](#) [CodeConnexx 2013](#) [International PHP Conference](#)  
Calling for papers: [Ski PHP 2014](#)

---

### PHP 5.4.21 Released

*17-Oct-2013* The PHP development team announces the immediate availability of PHP 5.4.21. About 10 bugs version.

For source downloads of PHP 5.4.21 please visit our [downloads page](#), Windows binaries can be found on [win ChangeLog](#).

### PHP 5.5.5 has been released

*16-Oct-2013* The PHP development team announces the immediate availability of PHP 5.5.5. This release fixes build system. All PHP users are encouraged to upgrade to this new version.

For source downloads of PHP 5.5.5 please visit our [downloads page](#), Windows binaries can be found on [win ChangeLog](#).

### PHP 5.4.20 released

## 第四节 并行计算

---

并行计算 (parallel computing) 是一种许多计算同时进行的计算形式，其基于大问题常可以分割成较小的问题并同时（并行）解决的原理。



# 一、概述

## 二、硬件



图 11-9 Cray-1

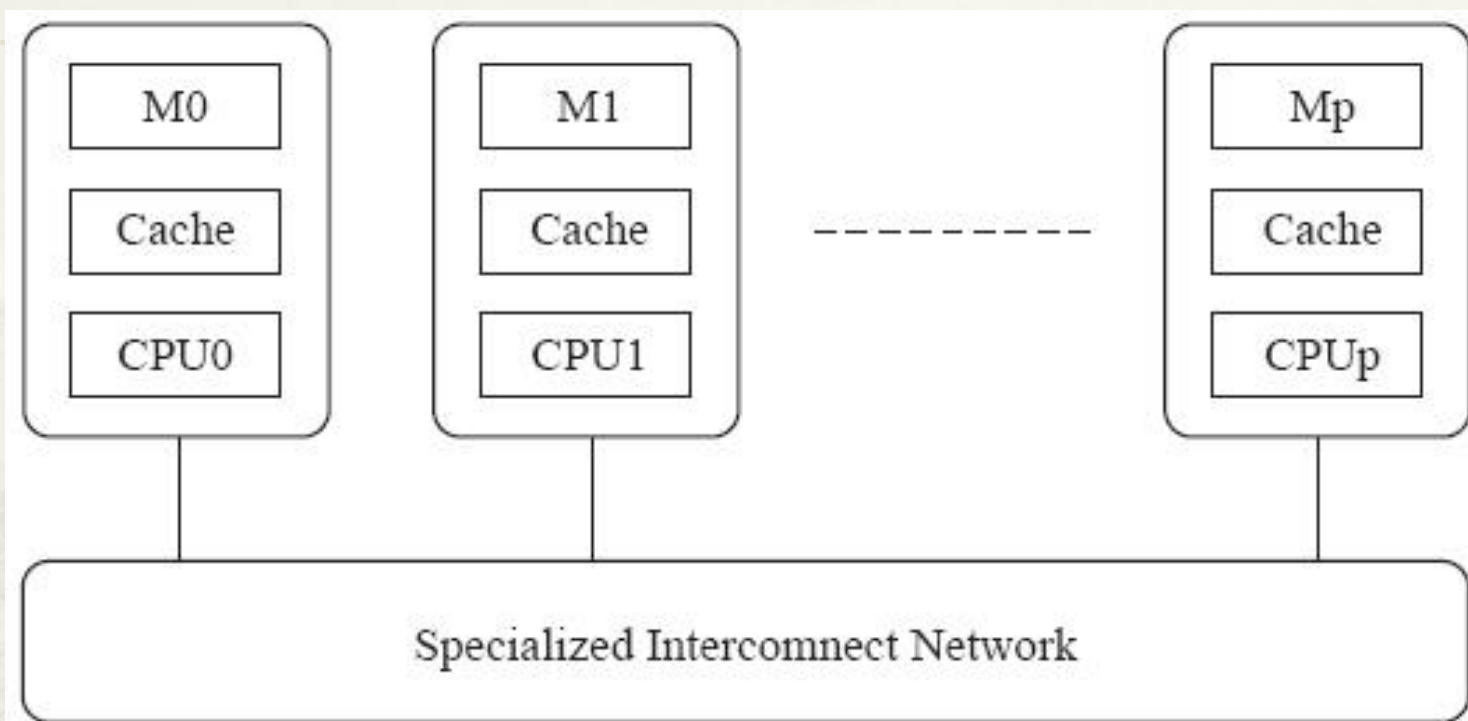


图 11-10 大规模并行处理机的系统结构

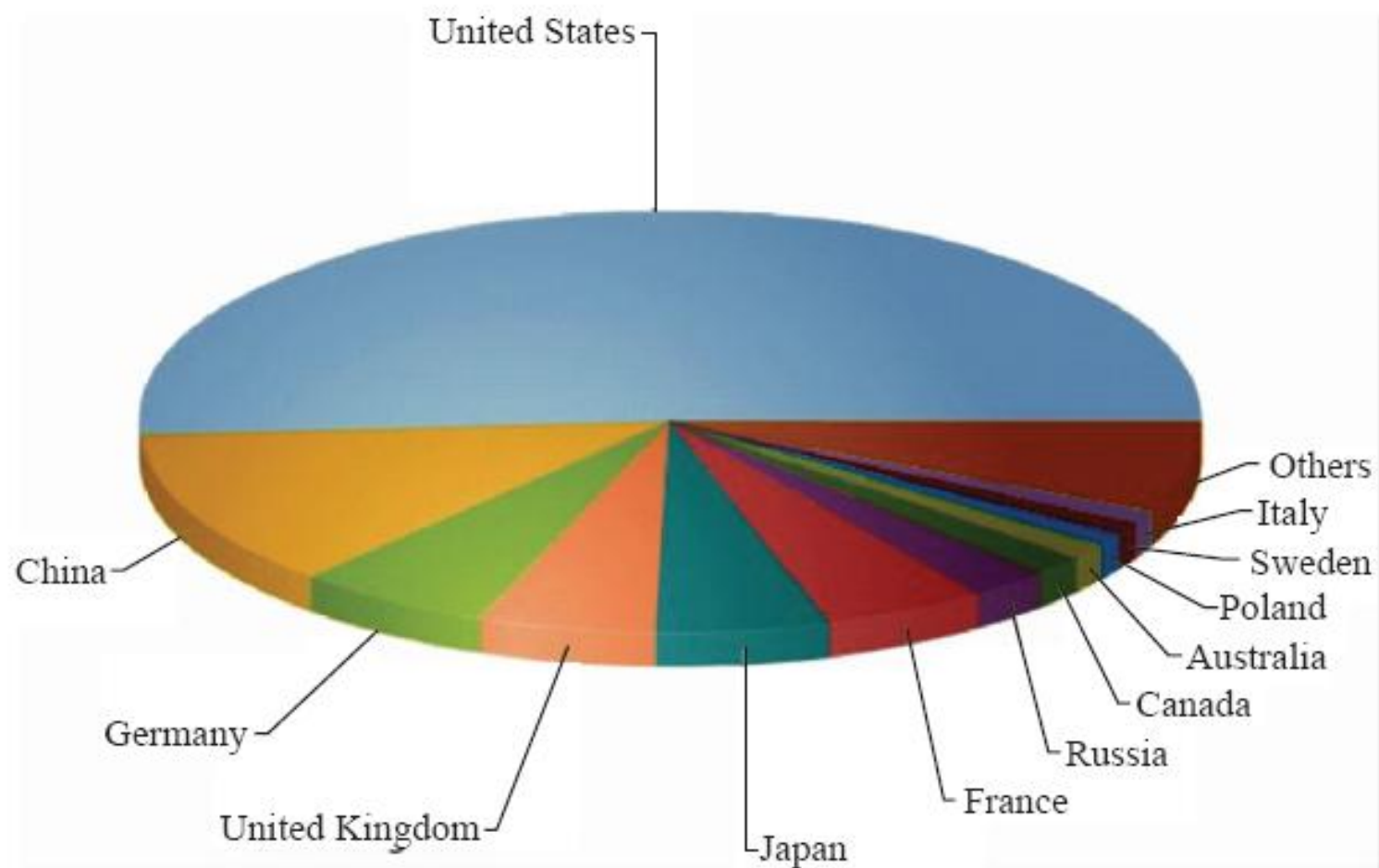


图 11-11 Top 500 按国家分布(资料来源: [www.top500.org](http://www.top500.org))





图 11-12 NBCR 集群

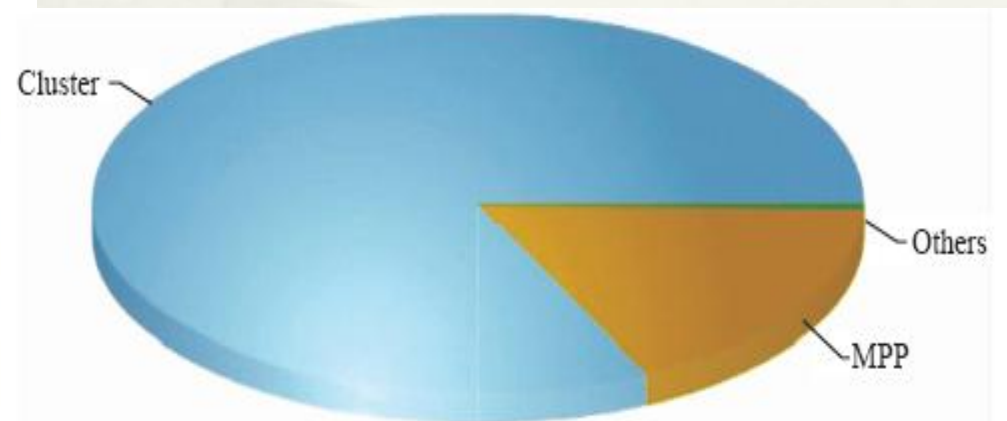


图 11-13 Top500 按体系结构分布(资料来源:www.top500.org)

## 三、软件

---

共享存储、分布式存储、共享分布式存储

## 四、并行算法设计

### 1、基本概念

粒度、加速比、效率、性能

### 2、基本方法

### 3、并行计算的实现

## 五、云计算

云计算是分布式计算技术的一种，是指通过网络将庞大的计算处理程序自动分拆成无数个较小的子程序，再交由多部服务器所组成的庞大系统经搜寻、计算分析之后将处理结果回传给用户。

特点：

超大规模、虚拟化、高可靠性、通用性、高可扩展性、按需服务、及其廉价。