

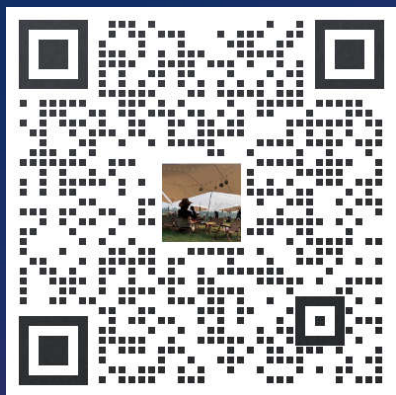


# 数据库系统 | Database Systems

## 第一课：数据库及数据库管理系统

Lecturer: Harbour

Date: 2021.11.10



Scan CC WeChat to Join the Community

添加CC好友，接受进群邀请

Welcome to follow the GitHub repo

欢迎关注我们的代码仓库

<https://github.com/cnosdb/cnosdb>



## 必讲

DB/DBMS

关系模型与关系代数

数据库存储

散列索引（哈希）

B+树

查询处理

并发控制

## 如果有兴趣

SQL

查询优化

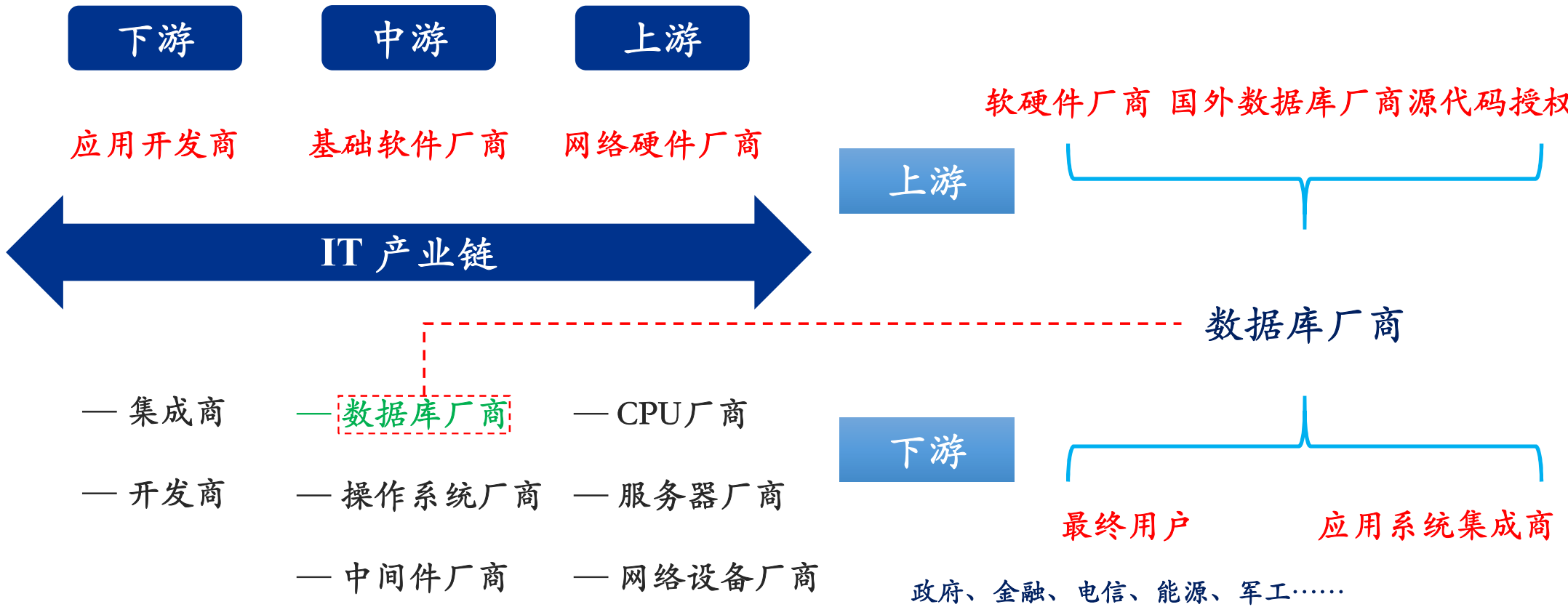
恢复系统

分布式OLTP/OLAP

- **Organized** collection of **inter-related** data that models some aspect of the **real-world**.

- 对现实世界的某些方面进行建模的相互关联的数据的有组织的集合。

数据库在产业链中位置



- 建立一个数据库，为在线音乐建模，用来记录歌手和专辑。

需要存储：

- 歌手信息
- 歌手所发行的专辑信息

- 创建数据库（平面文件，CSV文件）

歌手Artist

（名字name，出道时间year，国家country）

专辑Album

（名字name，歌手artist，年份year）

“王菲”，1985，“中国”  
“陈奕迅”，1996，“中国”  
“周杰伦”，2000，“中国”

“我愿意”，“王菲”，1997  
“只爱陌生人”，“王菲”，1999  
“陈奕迅”，“陈奕迅”，1996  
“范特西”，“周杰伦”，2001

- 我想知道周杰伦哪一年发行的第一张专辑？

歌手Artist

(名字name, 出道时间year, 国家country)

“王菲”，1985，“中国”  
“陈奕迅”，1996，“中国”  
“周杰伦”，2000，“中国”

```
for line in file:  
    record = parse(line)  
    if “周杰伦” == record[0]:  
        print int(record[1])
```



**2000**

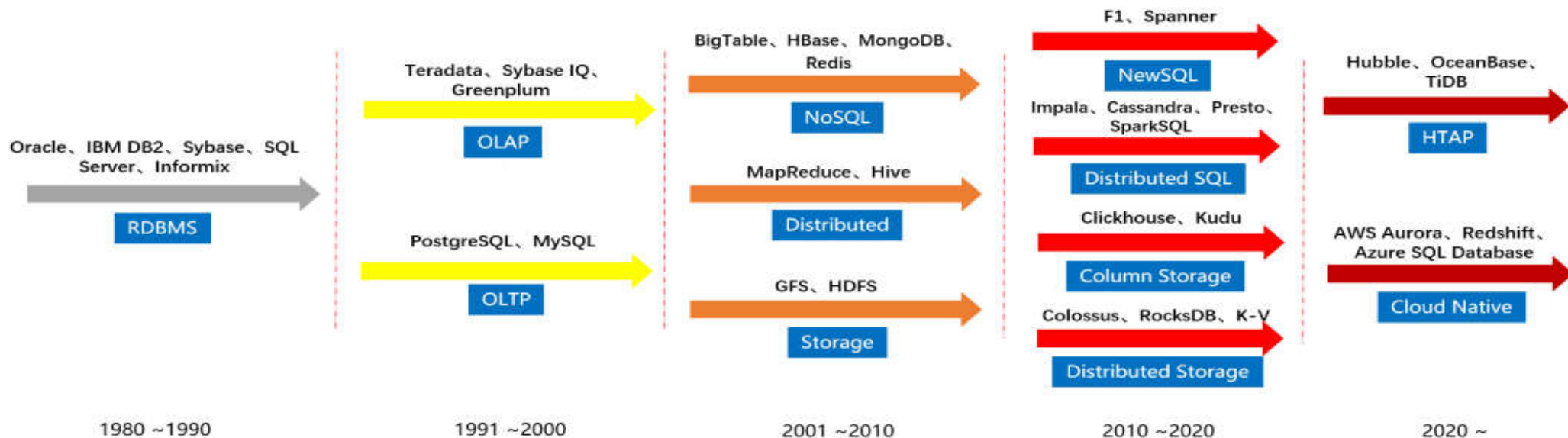
- 我们怎么确保每个专辑条目的歌手都是相同的？如果有拼写错误怎么办？比如把“王菲”写成“王飞”，把“周杰伦”写成“周杰林”。
- 怎么确保如果我们改动他们的出道时间，改成了不正确的字符，比如把1995改成1gg5，系统可以处理？
- 如果一张专辑有多位歌手参与创作，如何存储呢？比如，2011年的专辑《Stranger Under Skin》，《因为爱情》是陈奕迅王菲共同创作的。



- 我们如何找到一个确定的记录？
- 我们如果创建一个新的应用，但要使用原来的数据库，而这个应用可以是不同的编程语言。
- 如果两个线程在同一时间写一个文件，应该如何操作？
- 如果当程序更新数据记录的时候系统崩溃了怎么办？
- 如果我们想在多台机器上复制数据库以获得高可用性怎么办？

- Oracle, SQL-server, MySQL, Neo4J, CnosDB .....
- DBMS，数据库管理系统，是一个软件系统，允许应用在数据库中存储和分析信息。
- 定义、创建、查询、更新、管理数据库。

# 数据库管理系统40年发展



商业关系型数据库起步阶段，关系型数据库模型是把复杂的数据结构归结为二维表格形式。在关系数据库中，对数据的操作几乎全部建立在一个或多个关系表格上，通过对这些关联的表格分类，合并，连接或者选取等运算实现数据的管理（结构化查询语言，sql语句来对数据进行处理）。

如：Oracle、IBM DB2、Sybase、SQL Server、Informix

开源数据库展露头角，同时出现了一些分析型数据库，因为之前都是OLTP数据库。随着大数据的出现，需要对数据进行分析，为了避免读写冲突，就需要建立专门的分析型数据库系统，

如：Teradata、Sybase IQ、Greenplum

以谷歌为代表的互联网公司逐渐推出了面向BigData的数据库计算框架及存储系统。

谷歌新技术退出后，业界很快的衍生了一个新的领域叫NoSQL (Not Only SQL)，就是针对非结构化、半结构化的海量数据处理系统。

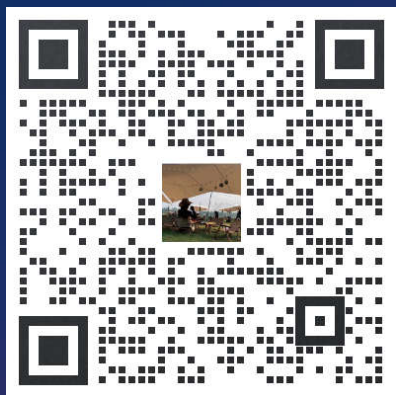
如：MongoDB、Redis

新的10年，数据从结构化数据在线处理到海量数据分析，从SQL+OLTP的RDBMS到ETL+OLAP的Data Warehouse，再到今天NoSQL+Data Lake的多源异构数据类型的发展历程。Google、Amazon、阿里巴巴这些云计算厂商成了这个时期数据库发展的主要源动力。

如：AWS Aurora、Redshift、Azure SQL Database、Google F1/Spanner、Polar DB、AnalyticDB



# Q&A



Scan CC WeChat to Join the Community

添加CC好友，接受进群邀请

Welcome to follow the GitHub repo

欢迎关注我们的代码仓库

<https://github.com/cnosdb/cnosdb>

