

# 网易分布式数据库DDB

[bigdata.163yun.com](http://bigdata.163yun.com)

# 網易

- 1997 公司创建
- 2000 在NASDAQ上市（美国上市的第三家中国公司）
- 2006 以互联网应用、服务为主导的杭州基地/杭州研究院成立
- 2015 营收228亿人民币，净利润67.35亿人民币
- 2017 截止2017年2月7日，网易总市值338.14亿美金，每股股价258.67美金

网易总部位于杭州，目前拥有各类  
互联网人才一万多名，产品服务10亿人...



- 01** 为什么需要网易DDB
- 02** 为什么是网易DDB
- 03** 行业案例
- 04** 我们的客户



# **PART 01** 为什么需要 网易DDB

# 单机数据库时代

## 单机数据库瓶颈

磁盘IOPS

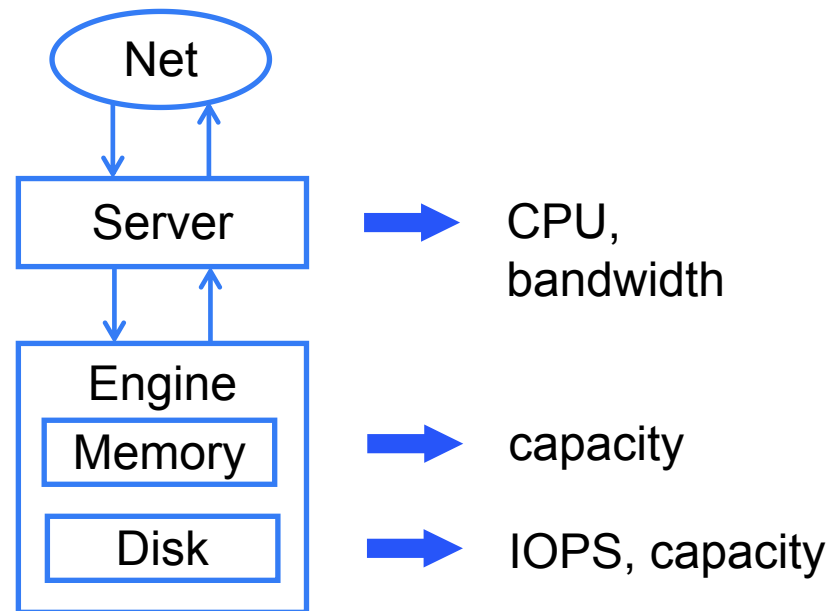
内存容量

处理器性能

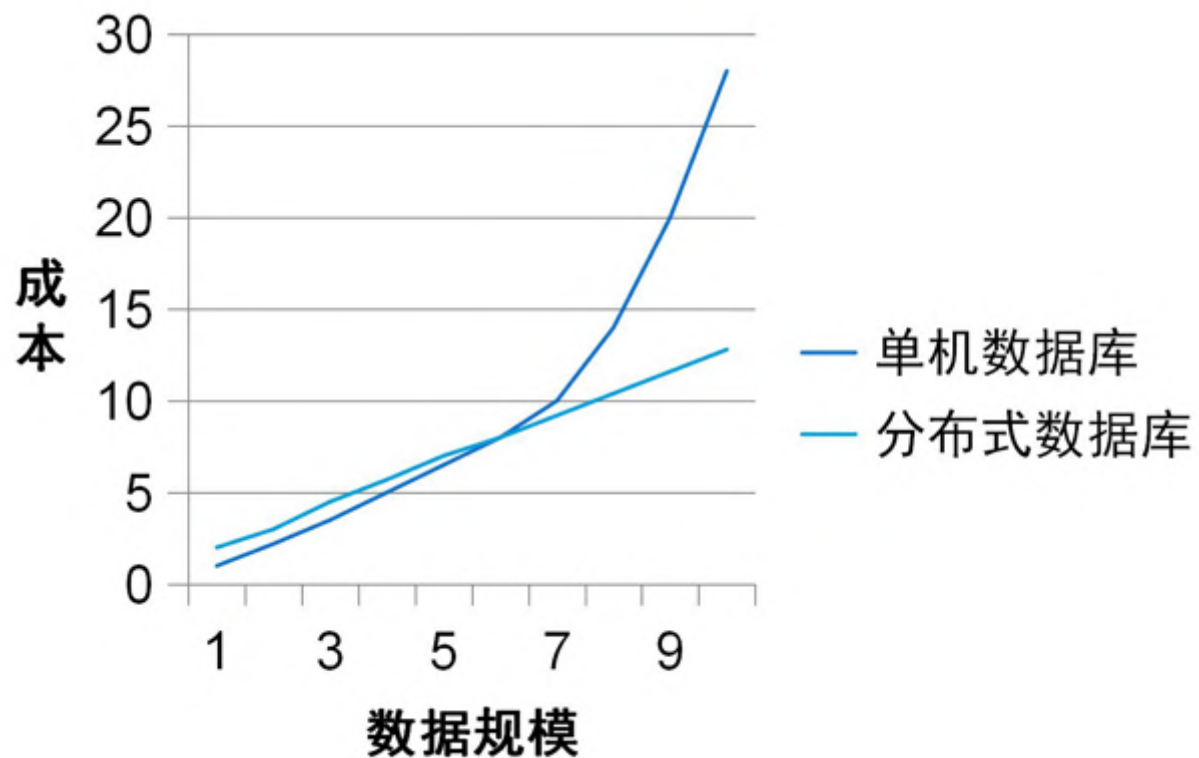
磁盘容量

网络带宽

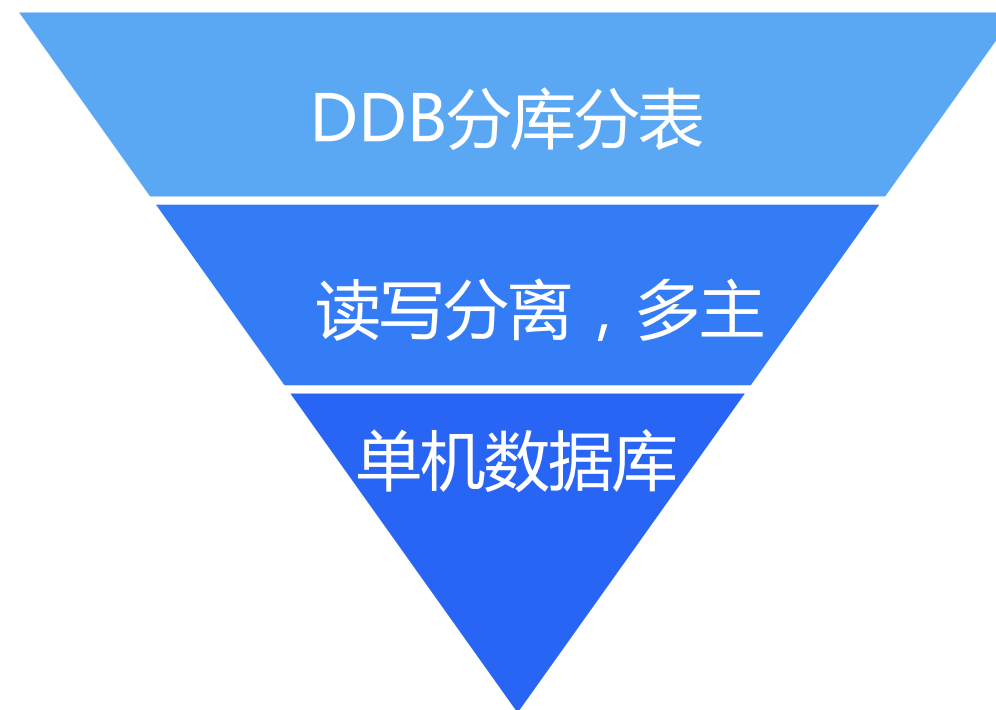
“数据达到一定规模后，单机数据库是在针尖上跳舞。”



# 企业使用数据库面临的问题



## 数据库能力金字塔



# DDB：一步到位的海量数据存储方案

网易大数据  
Netease Big Data

PB级结构化数据存储

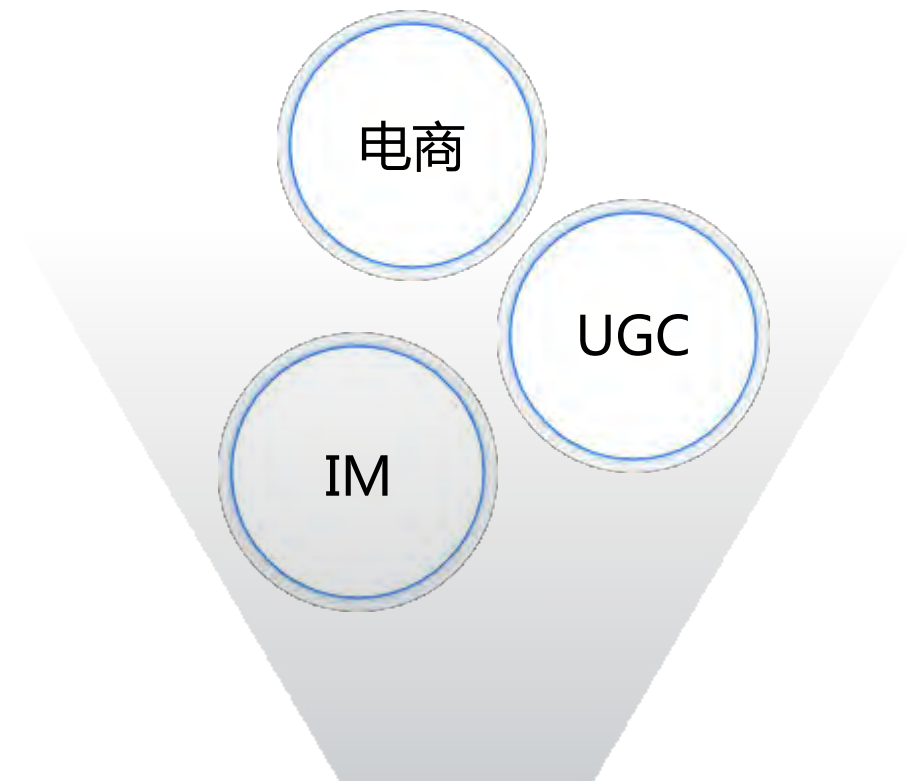
百万级qps

每日GB-TB数据增长

在线扩缩容

管理上千个数据节点

标准化的访问协议



DDB典型应用场景

---

## **PART 02** 为什么是网易DDB



# DDB简介 - 十年一剑

2006年开始，DDB为网易各大互联网产品提供透明分库分表服务。  
10年来不断完善，精益求精，是网易大体量互联网产品的立身之本。

2006年

博客上线

简单SQL兼容  
部分管理功能

2008年

V2.0发布

SQL兼容扩充  
在线扩容功能  
图形化管理工具

2010年

V3.0发布

分布式事务  
在线修改表结构  
SQL兼容扩充  
管理功能完善  
集群规模上千

2012年

V4.0发布

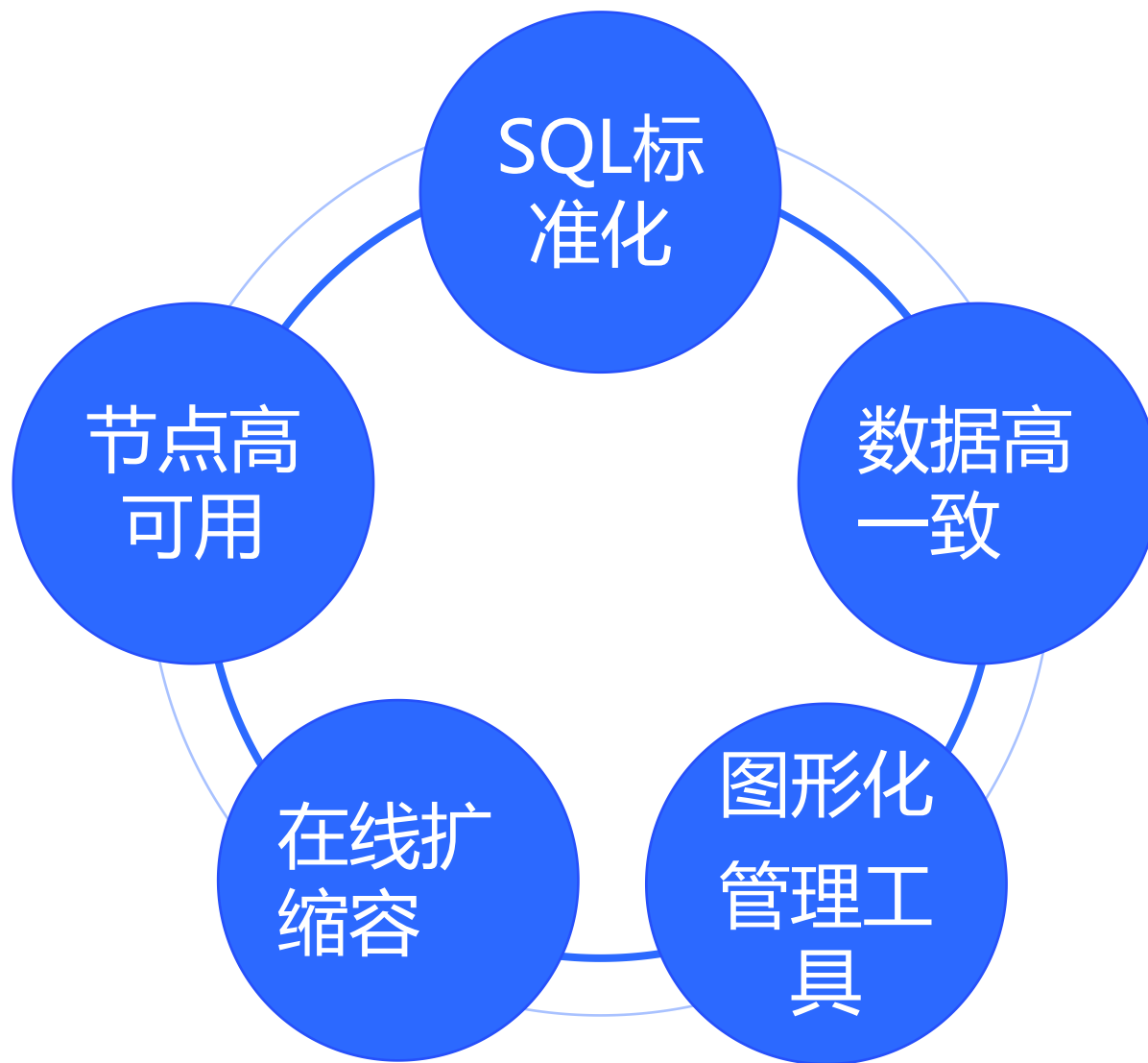
多语言支持  
SQL统计功能  
云计算DDB

2017年

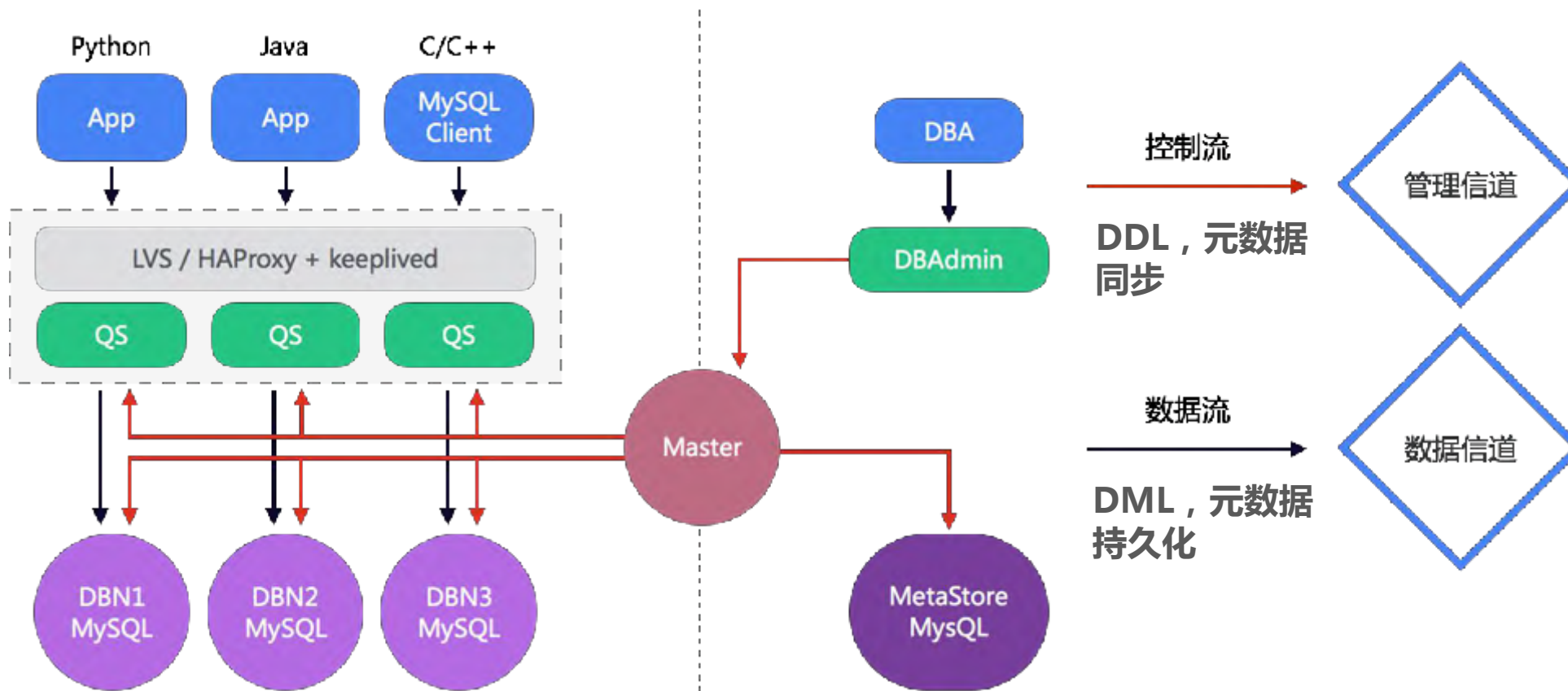
V5.0

架构简化  
服务拆分  
SQL兼容度进  
一步扩充

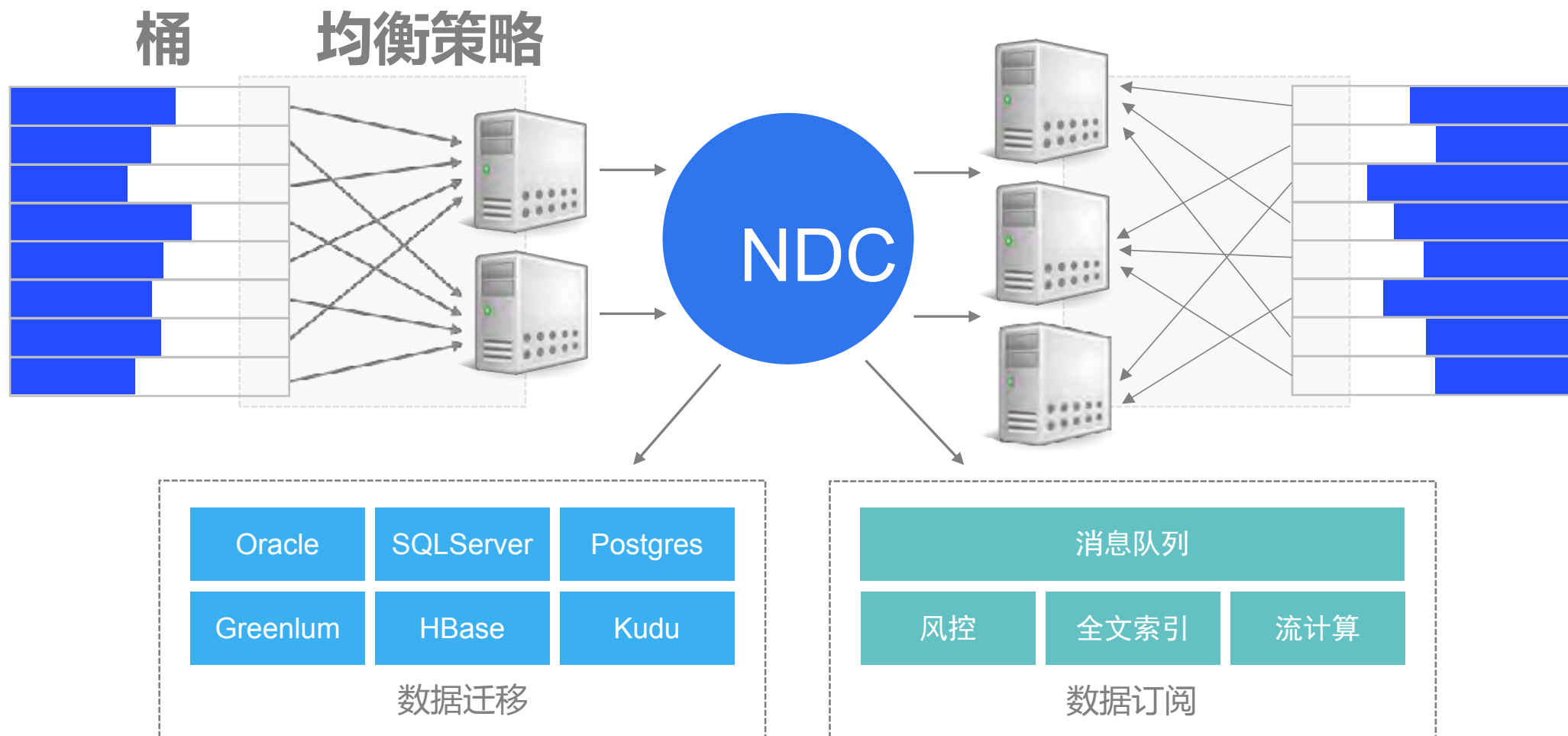
# DDB简介 - 核心优势



# DDB系统架构



# DDB系统架构



# DDB功能特性——数据信道

关键字：标准化，MySQL-like，数据高一致，SQL统计

## 数据分布

- 两级映射
- 自定义哈希函数

## 标准化

- SQL92 高兼容
- 全局自增ID
- 支持explain
- 数据导入导出
- 兼容MySQL通信协议

## 分布式事物

- 实现2PC协议
- 数据高一致
- 用户透明
- 自动识别

## Hint功能

- 读写分离
- SQL路由

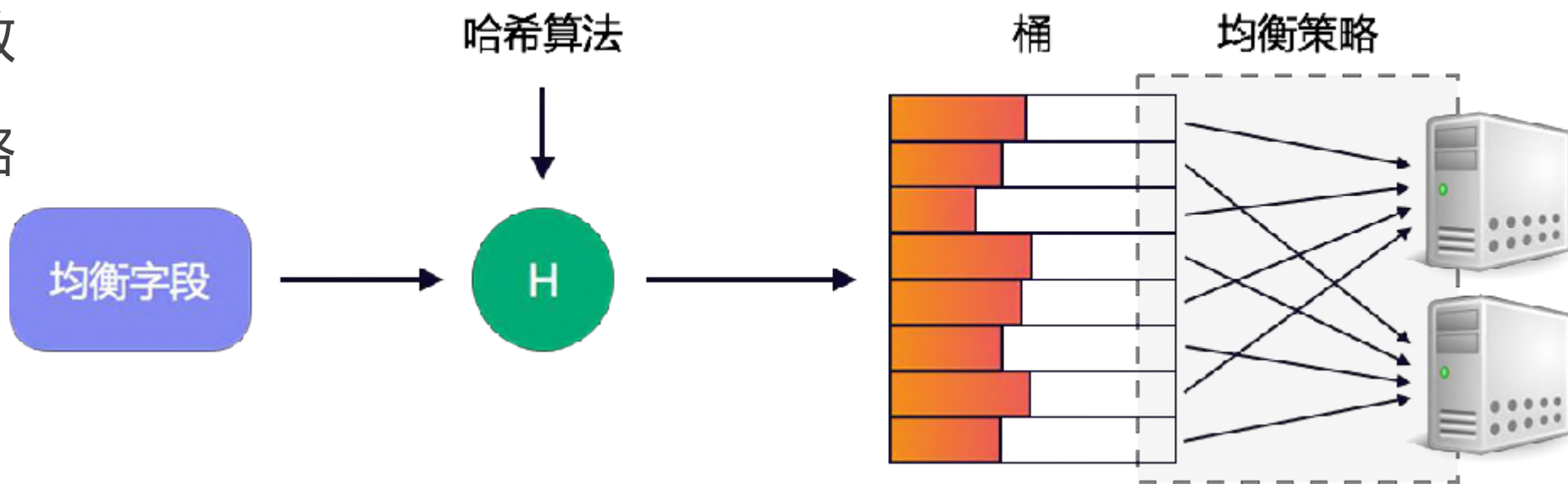
## SQL统计

- SQL模式统计
- SQL频度统计
- 慢SQL统计
- 多维度QPS统计

# DDB数据分布

- 两级映射

- 第一级：哈希函数
- 第二级：均衡策略
- 均衡性 + 单调性



- 自定义哈希函数

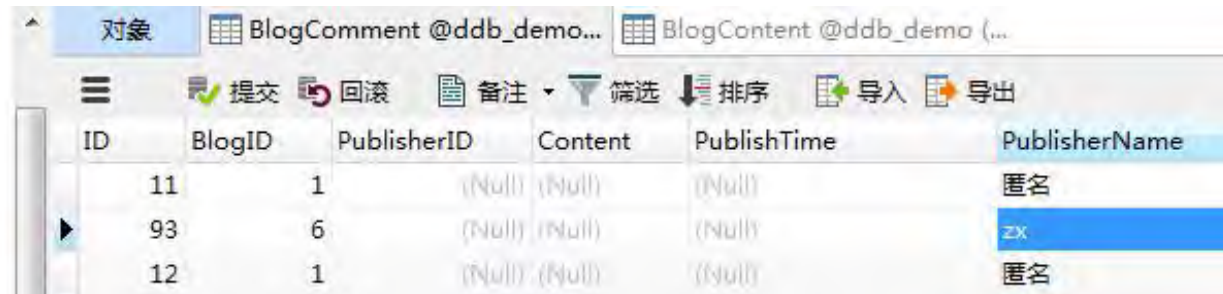
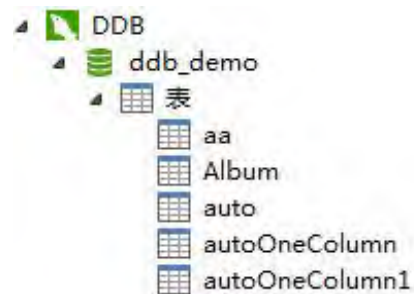
- range分区
- list分区

# DDB标准化——MySQL-like

- SQL兼容性90%
  - 支持所有简单SQL
  - 支持order by, group by, limit
  - 支持标量函数和大部分聚合函数
  - 支持部分特殊MySQL语法
  - 支持跨库join. etc
- 兼容MySQL通信协议
  - 支持任何语言MySQL客户端
  - 应用可使用任意ORM框架

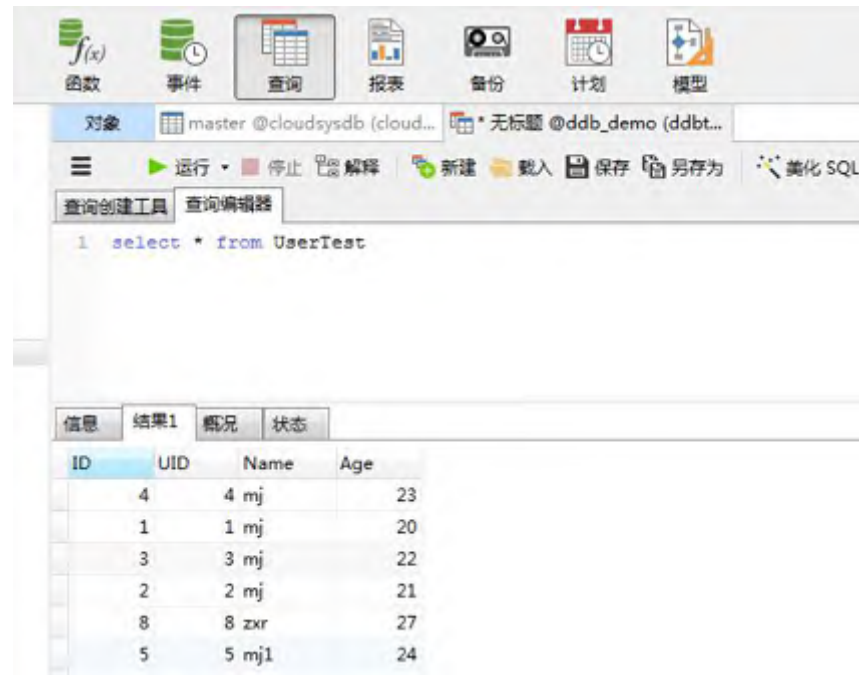
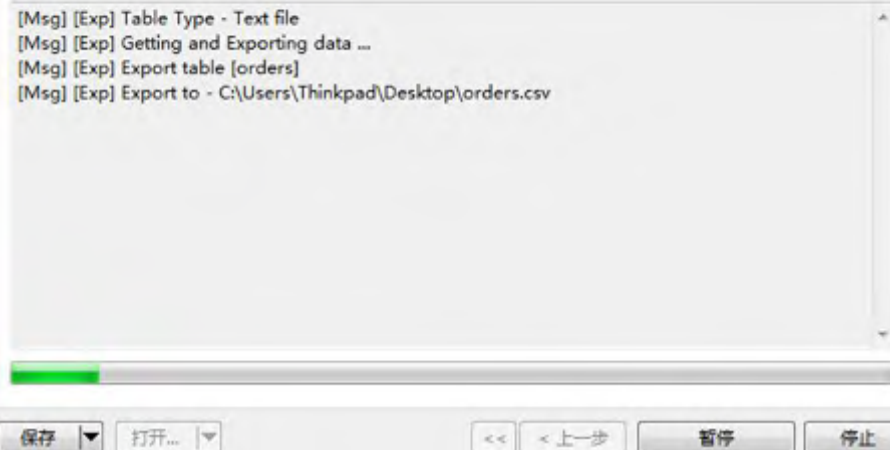
```
mysql> explain select avg(age) from UserTest group by name limit 10,10;
+-----+
| PLAN
+-----+
| LIMIT/OFFSET
  This plan will be dynamically set disable/enable while running based on the underlying plan.
  ^
 /|\
  ||
  AGGREGATE
  Do:
  ^
 /|\
  ||
  PROJECT
  Project record to: SUM(age),COUNT(age),
  ^
 /|\
  ||
  GROUP
  Group By: name,
  ^
 /|\
  ||
  MERGE-SELECT
  SQL: SELECT SUM(age), COUNT(age), name FROM UserTest GROUP BY name ORDER BY name ASC
  Dest Node:
  dbn1[jdbc:mysql://10.120.146.129:3306/dbn1]
  dbn2[jdbc:mysql://10.120.146.129:3306/dbn2]
  dbn4[jdbc:mysql://10.120.146.130:3306/dbn4]
  dbn3[jdbc:mysql://10.120.146.130:3306/dbn3]
  Order by: name ASC, with merge sort.
+-----+
```

# DDB标准化——兼容Navicat等工具



我们已收集向导导出数据时所需的全部信息。点击 [开始] 按钮开始导出。(5/5)

源表: orders  
总计: 8523329  
已处理: 885837  
时间: 45.173s





# SQL统计——模式统计

时间范围: 2016-01-06 12:00 ~ 13:00 确定

sql_pattern	avg_exec_time	exec_count	min_exec_time	max_exec_time	dbn_count	avg_row_count	max_row_count	avg_result_size	max_result_size	nestloop
select * from PRIS_BookSourc...	2	161709	1	1226	4	0	1000	12849	1554567	false
select * from PRIS_Approve w...	0	68452	0	339	1	0	0	629	630	false
select * from PRIS_Approve w...	0	57342	0	1236	1	0	1	629	732	false
select * from PRIS_ArticleTem...	0	45100	0	854	1	0	1	788	984	false
select * from PRIS_Account w...	2	41516	0	1654	1	0	1	585	649	false
select * from PRIS_PopUp wh...	0	29794	0	1005	1	1	1	819	837	false
select * from PRIS_TradeDeta...	1	27249	0	656	1	0	1	2397	2637	false
select * from PRIS_Article wh...	1	24090	0	1412	1	0	1	5054	6232	false
select * from PRIS_SourceSco...	0	16990	0	93	1	0	1	725	800	false
select * from PRIS_SourceCo...	1	15172	0	855	4	0	1	2826	9190	false

第1页,共93页

sql\_pattern详情

```
select * from PRIS_BookSource where BookId >= # and BookId <#
```

# DDB功能特性——数据信道

关键字：标准化，在线数据迁移，高可用，集群管理监控

## 集群管理

- 配置管理
- 连接池管理
- 元数据管理和同步

## 表/策略管理

- 创建/删除/更新
- 在线修改表结构
- 支持show/desc等
- 兼容MySQL管理语法

## 用户管理

- 创建/删除/更新
- 支持常用授权操作
- 支持白名单操作
- 管理与访问权限分离

## 在线数据迁移

- 在线策略迁移
- 在线扩/缩容
- 更改均衡字段

## 高可用

- 中间件节点高可用
- 数据节点高可用
- 数据节点手动切换

## 扩展功能

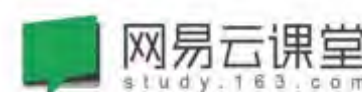
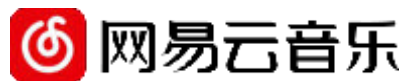
- 数据节点报警
- 中间件节点报警
- 悬挂事务报警
- 定时任务

# DDB使用案例

100+  
产品实践

10000+  
数据节点

1000000+  
QPS



# 实施与服务

- 解决方案，提供一体化方案
- 定制开发，满足个性化需求
- 培训辅导，一对一上门服务
- 技术咨询，协助建立业务主题
- 版本升级，保持行业领先
- 技术支持，解决后顾之忧



---

## **PART 03** 行业案例

# DDB应用案例——IM



业务表



业务规则

表名	数据规模	每日数据增长
用户表	200GB +	100MB +
朋友表	1TB +	800MB +
消息表	10TB +	10GB +

- 用户可通过手机号和账号名登陆
- 查询“我的朋友”和“我是谁朋友”各占一半
- 聊天记录3个月定期归档，以冷热分离

# DDB应用案例——IM

## 建表

- 1.设置用户表自增ID为userId
- 2.选择三张表的均衡字段均为userId
- 3.保障业务中80%以上SQL包含userId的判等条件

(在建表语句后指定hint: /\*BF=userId\*/)

## 业务1 & 2

- 1.为用户表建反查表phone2user
- 2.为朋友表建立冗余表friend\_reverse
- 3.插入用户数据和朋友数据时，同时插入反查表和冗余表
- 4.用电话登陆时反查userId，查询“我是谁的朋友”用冗余表

## 业务3

- 1.使用DBA管理工具建立归档的存储过程
  - 2.使用DBA管理工具建立定期任务每周执行归档存储过程
- ( 归档：按照一定条件将线上表中的数据迁移到冷表中 )



# THANKS

[bigdata.163yun.com](http://bigdata.163yun.com)

[bigdata-bd@hz.netease.com](mailto:bigdata-bd@hz.netease.com)