



Oracle技术嘉年华

Oracle Technology Carnival 2015

稳健•高效•云端 - 数据技术最佳实践

主题：新浪MySQL服务自助平台iDB构建实践

讲师：张博森



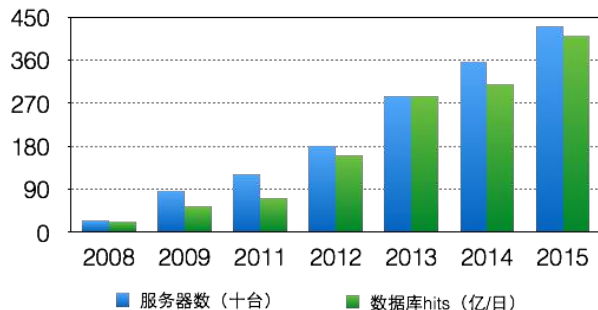
个人介绍

- 张博森
- MySQL DBA
- 13年初加入新浪，新浪微博MySQL运维，自动化运维工具开发
- 微博 [@repls](#)，微信 icer_repls

内容大纲

- 新浪数据库平台概览
- 早期的自动化运维平台构建思路
 - 当初遇到了什么问题
 - 是什么思路解决的
- 后期的MySQL自助服务平台构建实践
 - 为什么还要MySQL自助服务平台
 - 实现思路
 - 提供的哪些自助服务

新浪数据库平台概览



平台规模

- 15个IDC数据中心
- 3000+ 台服务器
- 600+项目在使用
- 数PB级的数据量
- MySQL 400+亿hits/天, 峰值150+万hits/秒

自动化运维平台构建之路

从问题开始

业务快速发展

- 问题
 - 服务器的快速增长、端口数的快速增长
 - 手工干的事情太多
 - Excel工程师、rsync工程师
 - DBA工作低效、苦逼
- 怎么解决大规模服务器、MySQL集群运维？
- 怎样实现DBA的价值？

解决问题思路

我们需要只是一个自动化运维平台而已！



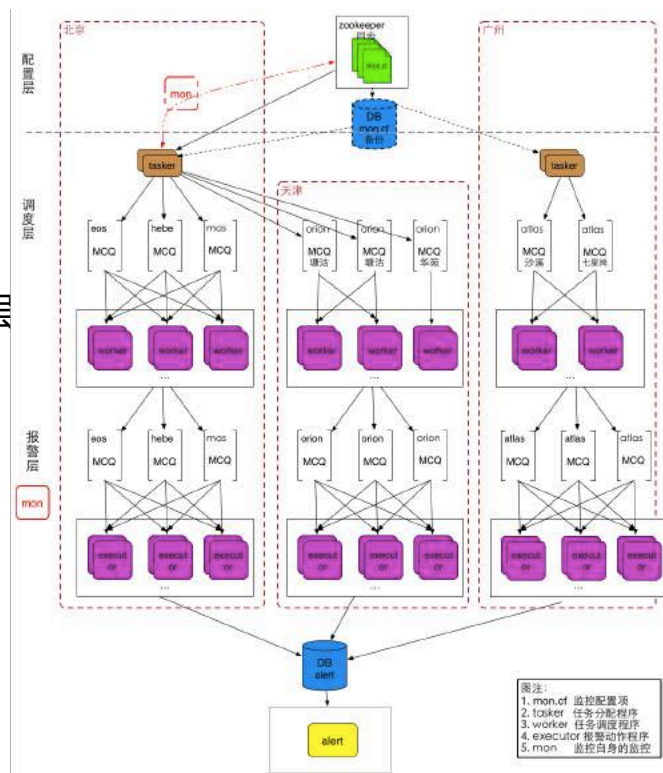
自动化运维平台怎么做？

- 自动化运维平台构建三板斧
 - 监控报警
 - 备份
 - 高可用
- 自动化运维建设思路
 - 标准化、规范化
 - 人工化→脚本化→工具化/系统化
- 技术选型，自研 or 开源？
 - 功能
 - 扩展
 - 维护成本
 - 平台规模



监控报警

- Mon问题：
 - 中心化，单点，调度性能
 - 监控报警逻辑耦合
- MQ解耦tasker-worker-executor三层
- 无状态、分布式部署
- 高度扩展性
 - 支持自定义开发支持其他服务



监控报警

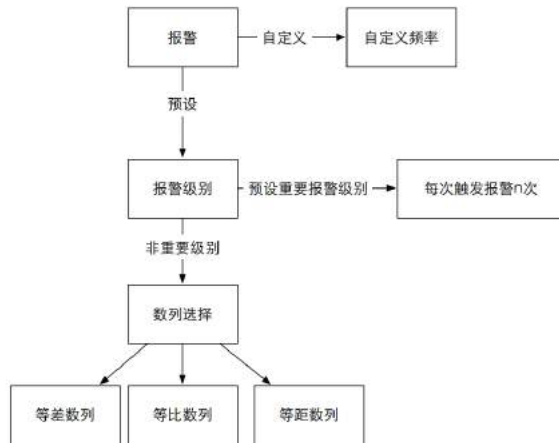
- 数量、频率控制

- “轰炸”
- 收敛

- 时间窗口

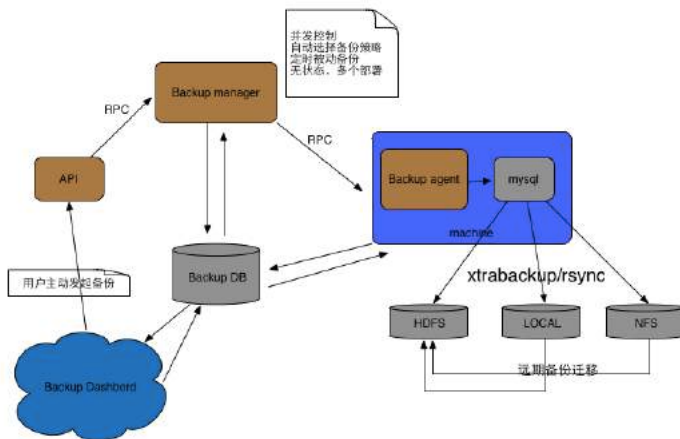
- 自动报警处理

- ELK分析报警信息



MySQL备份

- 备份目的
 - 数据恢复
 - 扩容来源
- 备份工具
 - xtrabackup、rsync
- 备份池
 - NFS、LOCAL、HDFS、S3
- 基于openstack RPC框架
- 丰富的备份方式支持
- 丰富的存储方式支持
- 灵活的备份策略
- 主动、自动智能调度
- 失败自动重试
- 成本节约&效率
 - 近期备份优先放LOCAL，高效
 - 较远期备份自动迁移到HDFS，节约成本
 - 历史备份自动迁移到S3

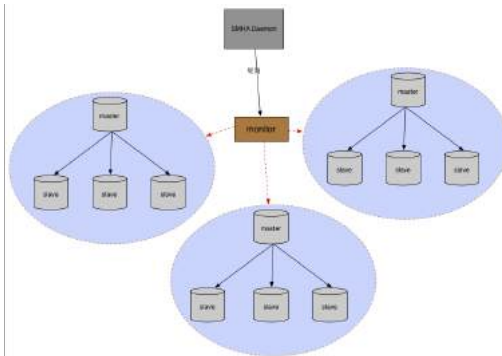
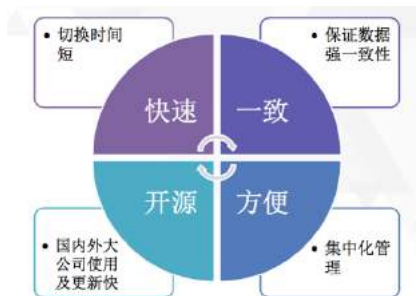


为什么需要高可用

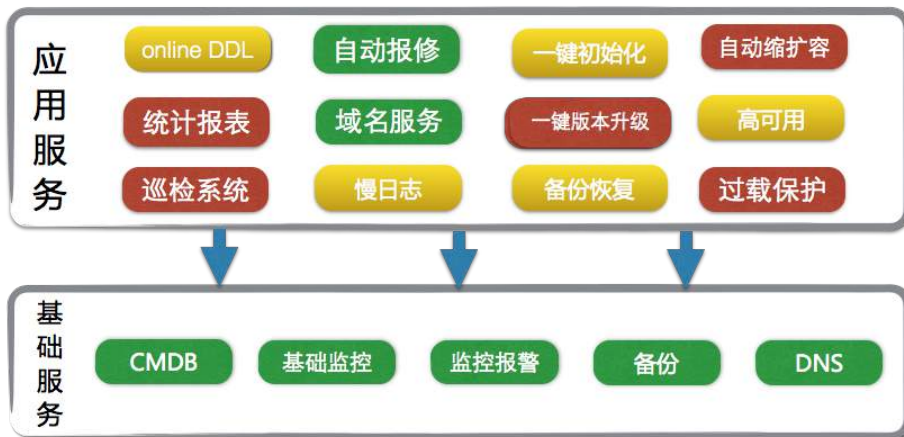


MySQL高可用

- **选型**
 - MMM、MHA
 - PXC
 - DRBD+HEARTBEAT
- **监控**
 - 轮询监控报警DB
- **切换**
 - MHA补binlog
 - save_binary_logs、apply_diff_relay_logs
 - **选举**
 - IDC
 - 主从混跑
 - 服务器负载
 - 服务器类型
- **切换时间平均10s左右**
- **实测并发切20+主库**



自动化运维平台功能总览



v1

v2

v3

MySQL自助服务平台构建之路

为什么还需要MySQL自助服务平台？

DBA的新烦恼

- 紧急上线，改个字段
- 紧急上线，建个表
- 紧急上线，授权一个ip

- 数据库部署在哪个IDC？
- 服务使用了多少服务器资源？
- 服务器类型？



- 查看一下DB访问量
- 数据库有慢查询吗
- 数据库rt是多少

- 数据库有报警吗
- 数据库有优化建议吗

为什么还需要MySQL自助服务平台（一）

- 反思
 - DBA 是不是太“自私”了？
 - 以前专注于规模壮大过程中平台的自动化运维
 - 大规模的服务器
 - 大规模的MySQL集群
 - 自动化运维系统还是不能解决我们开发的问题！



为什么还需要MySQL自助服务平台（二）

- 云时代的运维思路转变

- 资源服务化

- 资源（MySQL）的快速创建、回收
 - 资源的弹性伸缩（扩容、缩容）
 - 快速响应业务的变更需求
 - 思考业务的需求点

- 用户体检

- 尝试数据运营

- 来自外界RDS的压力



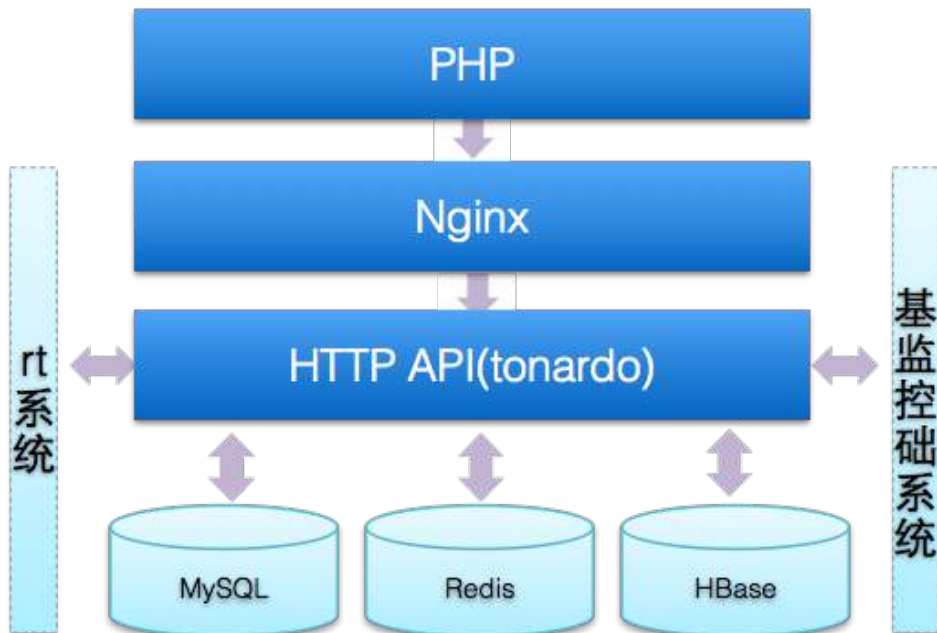
Database as a Service

iDB是什么

- 新浪数据库MySQL自助服务平台
- 核心用户是业务开发，不再是DBA
- 专注于快速、高效、高质量解决业务需求
- DBA不再被业务打断，时间不再碎片化



iDB架构



iDB服务组成

工单系统



- 自动建表、改表
- 自动授权
- 自动托管数据库
- 导入导出

服务分析



- 服务体检
- 智能修复
- 优化建议

信息展示



- 响应时间
- 慢日志
- 拓扑结构
- 元数据信息
- 数据库运行状态

报表分析



- 资源使用
- 关键数据
- 数据趋势

iDB-工单系统

- 几乎业务所有“写操作”
 - 线上数据库申请
 - 自动建表、改表（字段、索引等等）
 - 自动授权
 - 自动测试环境搭建
 - 自动数据导入导出等等
- 与公司RT系统联动，层级审批
- 内置语法审核、基本规范审核
- 后台机器人自动执行
- pt-online-schema-change
 - 不加唯一索引
 - 主从不异构

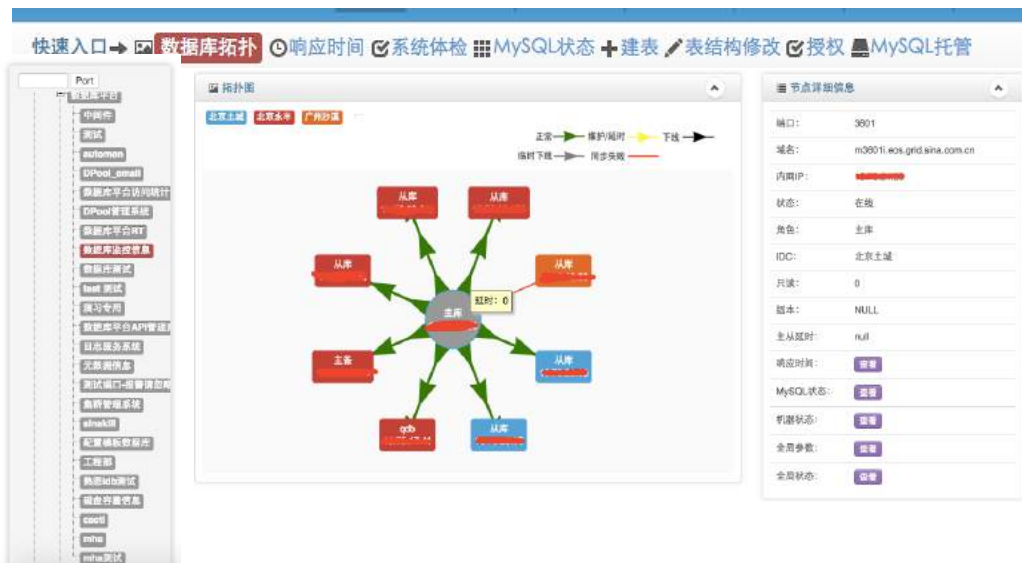


iDB-信息展示

- 设计原则
 - 凡是业务真正需要关心的数据
 - 凡是能减少业务与DBA沟通、交互的数据
 - 凡是能发现问题、分析问题的数据
- 目前已经提供的服务
 - 响应时间
 - MySQL状态
 - 慢日志
 - 拓扑结构
 - 访问统计
 - 元数据

拓扑结构

- 核心入口
 - 产品树的维度
 - MySQL服务的维度



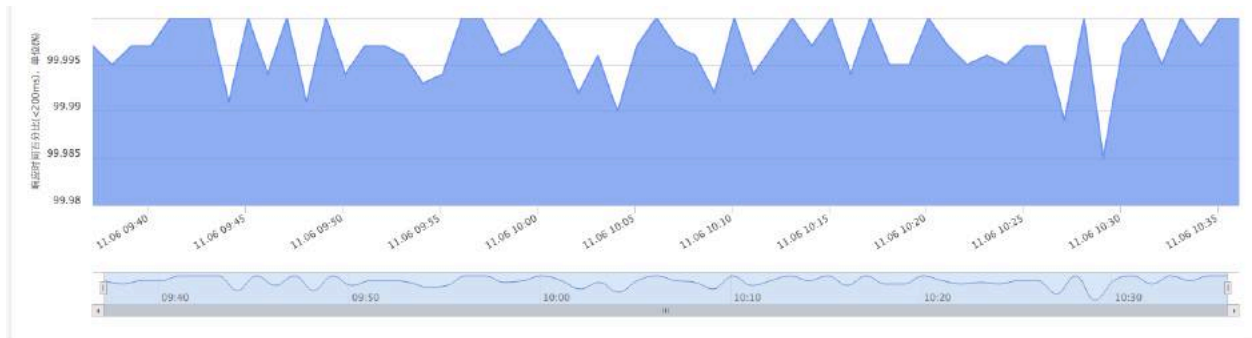
MySQL响应时间

- 衡量查询性能重要指标
- 异常数据驱动优化
- SLA

- tcpdump
 - 抽样采集，性能开销与数据有效性的折衷
- sina-query-digest
 - 分析tcpdump抓包数据
 - 基于pt-query-digest二次开发
 - 调整rt区间划分策略
 - 增加查询所属db信息记录
 - 过滤无意义的语句（select @@，show，set）

```
# -----  
# Count      31      189  
# Exec time   35      83ms 107us      8ms 437us 690us      1ms 176us  
# Rows affecte 61      189      1      1      1      1      0      1  
# Query size   57 122.04k 307 1.97k 661.22 1.20k 302.62 563.87  
# Warning coun 0      0      0      0      0      0      0      0  
# String:  
# Databases  
# Hosts      2 (17/8%), (15/7%)... 20 more  
# Users      openapi  
# Query_time distribution  
# 1us      0  
# 10us     0  
# 100us    ##### 184  
# 1ms      0  
# 5ms      # 5  
# 10ms     0  
# 20ms     0  
# 30ms     0  
# 50ms     0  
# 100ms    0  
# 200ms    0  
# 300ms    0  
# 500ms    0  
# 800ms    0  
# 1s       0  
# 10s+    0  
# Table  
# SHOW TABLE STATUS FROM `status_14` LIKE `status_151108`\G  
# SHOW CREATE TABLE `status_14`.`status_151108`\G  
insert into status_14.status_151108 (id, content) values(3906965388275619, x'A806A3DF)
```

MySQL 响应时间



端口平均响应时间统计

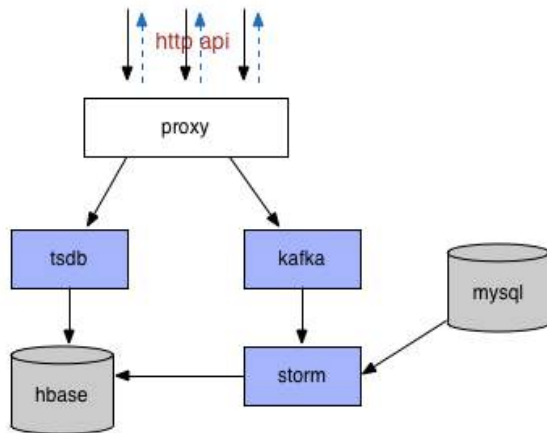
	1us	10us	100us	1ms	5ms	10ms	20ms	30ms	50ms	100ms	200ms	300ms	500ms	800ms	1s	10s+
执行次数	0	3	137	15520	19358	19421	19424	19424	19424	19424	19424	19424	19424	19424	19424	19425
RT SLA	0%	0.015%	0.705%	79.897%	99.845%	99.979%	99.995%	99.995%	99.995%	99.995%	99.995%	99.995%	99.995%	99.995%	99.995%	100%
执行比率	0.015%	0.68%	79.192%	19.746%	0.335%	0.015%	0%	0%	0%	0%	0%	0%	0%	0%	0.005%	-

Top SQL

IP	Query ID	总执行次数	执行比率/%	RT/ms	MinRT/ms	MaxRT/ms	总耗时/ms	SQL 样例
[REDACTED]	14836071049966305293	1	100%	0.13	0.13	0.13	0	insert into status (rt,status,rt1,rt2) values (rt,status,rt1,rt2) where id in (3554430074101310) Last
[REDACTED]	3225642287529924420	6306	91.65%	0.75	0.08	15.58	4865	select id, content from status where id in (3554430074101310)
[REDACTED]	7144874373204763743	583	8.35%	0.39	0.07	2.19	226	select id, content from status where id=3685145230177917

MySQL状态

- 查看数据库运行状态
 - 异常问题定位
 - 数据统计来源
-
- opentsdb, tcollector采集
 - Hbase分布式存储, 规模几乎不受限制
 - 极其灵活扩展新metric
 - Storm聚合历史数据
 - Highchartsweb端展示, 个性化定制
-
- 几个优化点:
 - metric数据存储格式, 区分度大的用作metric name
 - 原生tcollector还存在一些细节不完善的地方, 需要自行fix



MySQL状态



MySQL Slowlog

10 每页显示条数

Port	产品单元	平均返回行数	平均响应时间	总执行次数	总时间	索引命中率(%)	返回行数	总扫描行数	总等待时间	详情
3588	数据库监控信息	1.00	1.07	4	4.28		4	0	0.00	详情
3005	数据库监控信息	0.00	1.08	3	3.25		0	0	0.00	详情
3588	数据库监控信息	220.00	1.21	1	1.21	100.00	220	220	0.01	详情

当前状态

Query : 3583802290405043739

First Seen: 2014-04-21 13:57:32 Last Seen: 2015-11-06 06:27:17

Fingerprint

Last Sample on host: 10.75.18.75/4442 at 2015-11-06 06:27:17

SELECT table_name, table_rows, ENGINE, table_collation, data_length, index_length AS table_size, create_time, update_time FROM information_schema.tables WHERE table_schema = ?

SELECT table_name, table_rows, ENGINE, table_collation, data_length, index_length AS table_size, create_time, update_time FROM information_schema.tables WHERE table_schema = 'admin_conf'

诊断信息

mysql版本: 5.5.31-log long_query_time: 1.000000 库名: tables 表名: information_schema.tables

查看explain结果 查看表结构 查看表状态

```
***** 1. row *****
id: 1
select_type: SIMPLE
table: tables
type: ALL
possible_keys: NULL
key: TABLE_SCHEMA
key_len: NULL
ref: NULL
rows: NULL
Extra: Using where; Open_full_table; Scanned 1 database
```

iDB-体检系统

数据库服务健康体检功能，包括**体检、智能修复、体检报告、趋势分析**等功能。其中体检综合了数据库报警、流量趋势、运行状态等**多维度的运维数据**，旨在让开发**全面**的了解自己的服务健康程度。

- 体检四大维度
 - 可用性
 - 性能
 - 安全性
 - 其他
- 每天定时自动体检，发送体检报告
- 主动/被动一键修复，快速解决问题
- 针对不同业务，个性化定制体检项、权重

iDB-体检系统



体检结束，您的业务存在一些问题，您已击败全新浪90%的业务

体检完成

查看本次体检报告

一键修复

1

可用性检测

体检端口:3326

2

性能检测

3

安全性检测

4

其他检测

监测结果

共检查了30项，以下5项有问题

- 【可用性检测】 自增起始
- 【性能检测】 慢日志数量
- 【安全性检测】 old_passwords
- 【其他检测】 日访问量 (万/天)
- 【其他检测】 qcache异常

正常情况下，必须是1

存在慢查询，请联系DBA处理

建议关闭，以兼容高版本的MySQL连接客户端

端口访问量过大，请及时关注

qcache超过1kw天，建议前端增加缓存

有问题，点击查看

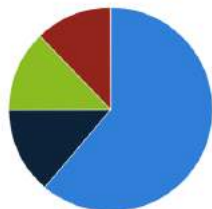
有问题，点击查看

有问题，点击查看

有问题，点击查看

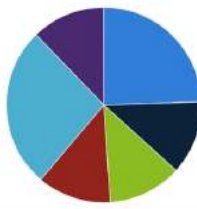
有问题，点击查看

体检问题分布概览



可用性检测 安全性检测 性能检测 其他检测

体检问题TOP 5



可用性检测 安全性检测 性能检测 其他检测 其他检测 qcache异常

iDB-报表分析

- 是不是有这样的疑问：
 - 某个服务使用了多少服务器资源？
 - 数据库平台每天读写是多少？
 - 某个端口一天的读写量是多少？
 - 备份成功率是多少？都有哪些端口备份失败了？
 - 最近平台的报警情况是怎样的，什么类型的报警减少了？什么类型报警增加了？
 - 各种数据的最近的趋势是怎样的？

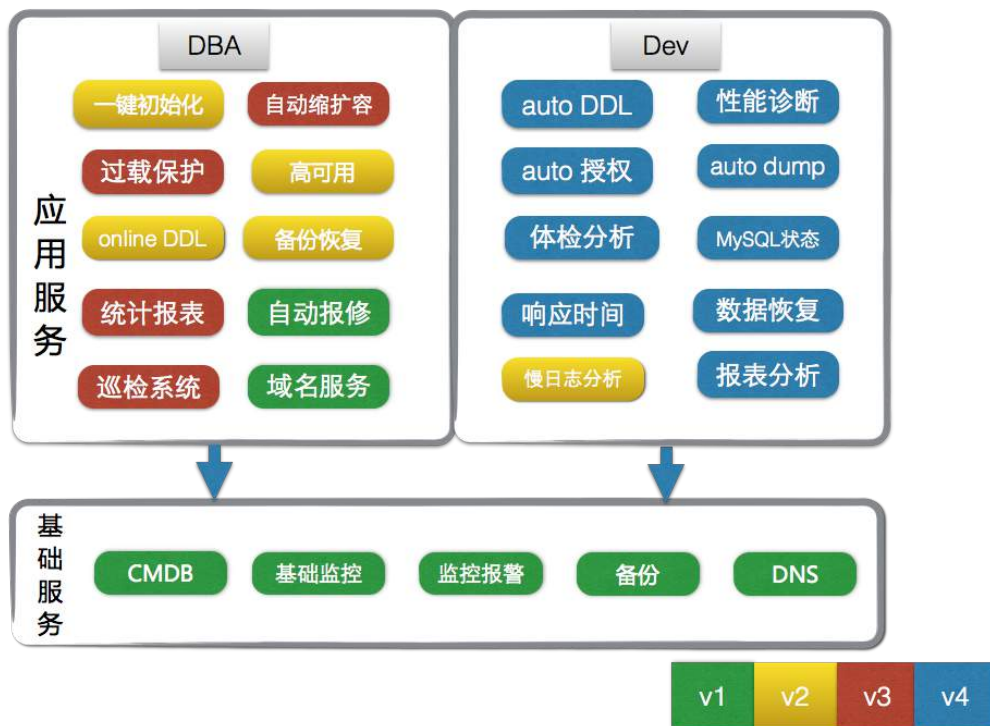
报表分析系统统统告诉你答案！

iDB-报表分析

- 多维度数据统计报表分析
- 发现数据内在的价值
 - eg 同比环比报警，发现异常趋势



数据库平台服务总览



关注微信公众号 获取文档和更新



云和恩墨

(支持ORA错误自动查询)



恩墨学院



Oracle新闻



Z3 – SQL审核工具提升SQL质量

• 独特的 SQL 视图

- SQL生命周期管理 – 捕获, 分析, 归因, 形成SQL全周期管理平台;



z3 – SQL审核



zData – 高性能弹性分布式存储解决方案

• 大数据整合与集中面临的平台压力

- 去“肥”留“瘦”：通过Virtual SAN替代FC SAN
- 同样的成本获得 20倍+ 的IO性能
- 动态扩展，高性能的存储解决方案



zData – 分布式存储



Announcing : BayMax 自动化巡检 即将开放云服务

• 自动化巡检 – 让DBA去完成20%最有价值的工作



BayMax自动化巡检

特别感谢 合作伙伴



云和恩墨
ENMOTech



Shannon Systems
宝 存 科 技

ACOUg
All China Oracle User Group
中国 Oracle 用户组

恩墨学院
ENMOEDU

Data Driven World
EN•CORE
恩核(北京)信息技术有限公司

OCMU
OCM 联盟

SDOUG
Shandong Oracle User Group
山东 Oracle 用户组

ITPUB

ChinaUnix

IT168.com
www.it168.com

BI 会议

Mellanox
TECHNOLOGIES
Connect. Accelerate. Outperform.™

乐导科技
alidao.com

慕课网
imooc.com



清华大学出版社

Broadview®
www.broadview.com.cn

TURING
图灵教育

华章科技

The image features a solid red background. In the top-left corner, there is a faint, light-red decorative swirl. In the bottom-right corner, there is a large, vibrant, multi-colored decorative swirl with shades of yellow, orange, pink, and blue. The word "THANKS" is centered in the middle of the image in a white, sans-serif font.

THANKS