

DBAplus

数据连接未来



# Newsletter 4月版

## 目录

DBAplus Newsletter.....	1
推出 DBAplus Newsletter 的想法.....	1
DB-Engines 数据库排行榜.....	2
<b>RDBMS 家族</b> .....	4
Oracle 12c Release 2 所有平台发布完毕.....	4
MySQL 5.7.18 发布.....	7
MySQL 5.7 将要废弃和已经废弃的特性.....	7
MariaDB 10.2.5 RC 版几个实用的特性改进.....	10
SQL Server 新版本即将亮剑.....	11
PostgreSQL 将发布 10.0.....	14
Greenplum Command Center 最近发布 3.2 更新.....	18
本期新秀: SQLite 发布 3.18.0 版本.....	21
<b>NoSQL 家族</b> .....	23
MongoDB 3.4.4 发布.....	23
Redis 3.2.8 版本发布.....	23
<b>NewSQL 家族</b> .....	25
TiDB 近期发布 RC3 版本.....	25
本期新秀: CockroachDB.....	27
<b>大数据生态圈</b> .....	28
Hadoop 发布 2.8.0 版本.....	28
HDB (Apache HAWQ) 2.2.0.0 版本发布.....	30
<b>国产数据库概览</b> .....	34
OceanBase 压测成绩瞩目.....	34
星瑞格多租户数据库.....	37
SequoiaDB 2.8 企业版正式发布.....	41
感谢名单.....	43

## 推出 DBAplus Newsletter 的想法

DBAplus Newsletter 旨在向广大技术爱好者提供数据库行业的最新技术发展趋势，为社区的技术发展提供一个统一的发声平台。为此，我们策划了 RDBMS、NoSQL、NewSQL、大数据、虚拟化、国产数据库等几个版块。

我们不以商业宣传为目的，不接受任何商业广告宣传，严格审查信息源的可信度和准确性，力争为大家提供一个纯净的技术学习环境，欢迎大家监督指正。

至于 Newsletter 发布的周期，目前计划是每两个月左右会做一次跟进，下期计划时间是 2017 年 6 月 12 日~6 月 23 日，如果有相关的信息提供请发送至邮箱：[newsletter@dbaplus.cn](mailto:newsletter@dbaplus.cn)，或扫描以下二维码填表申请加入“Newsletter 信息征集群”。



↑ 扫码填写申请表 ↑  
通过审核后群秘邀你入群

## DB-Engines 数据库排行榜

以下取自 2017 年 4 月的数据，具体信息可以参考 <http://db-engines.com/en/ranking/>，数据仅供参考。

323 systems in ranking, April 2017

Rank			DBMS	Database Model	Score		
Apr 2017	Mar 2017	Apr 2016			Apr 2017	Mar 2017	Apr 2016
1.	1.	1.	Oracle +	Relational DBMS	1402.00	+2.50	-65.54
2.	2.	2.	MySQL +	Relational DBMS	1364.62	-11.46	-5.49
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1204.77	-2.72	+69.72
4.	4.	↑ 5.	PostgreSQL +	Relational DBMS	361.77	+4.14	+58.05
5.	5.	↓ 4.	MongoDB +	Document store	325.43	-1.51	+12.98
6.	6.	6.	DB2 +	Relational DBMS	186.66	+1.74	+2.57
7.	7.	7.	Microsoft Access	Relational DBMS	128.18	-4.76	-3.79
8.	8.	8.	Cassandra +	Wide column store	126.18	-3.01	-3.49
9.	↑ 10.	9.	Redis +	Key-value store	114.36	+1.35	+3.12
10.	↓ 9.	10.	SQLite	Relational DBMS	113.80	-2.39	+5.83
11.	11.	11.	Elasticsearch +	Search engine	105.67	-0.56	+23.09
12.	12.	↑ 13.	Teradata	Relational DBMS	76.56	+3.02	+4.30
13.	13.	↓ 12.	SAP Adaptive Server	Relational DBMS	67.46	-2.67	-5.86
14.	14.	14.	Solr	Search engine	64.37	+0.38	-1.65
15.	15.	15.	HBase	Wide column store	58.47	-0.51	+6.98
16.	16.	↑ 17.	FileMaker	Relational DBMS	57.18	+2.61	+11.07
17.	17.	↑ 18.	Splunk	Search engine	55.50	+1.42	+13.15
18.	↑ 19.	↑ 21.	MariaDB +	Relational DBMS	48.72	+1.84	+17.14
19.	↓ 18.	19.	SAP HANA +	Relational DBMS	48.15	-1.91	+7.80
20.	20.	↓ 16.	Hive +	Relational DBMS	41.65	-2.97	-7.43

DB-Engines 排名的数据依据 5 个不同的因素：

- ✓ Google 以及 Bing 搜索引擎的关键字搜索数量
- ✓ Google Trends 的搜索数量
- ✓ Indeed 网站中的职位搜索量
- ✓ LinkedIn 中提到关键字的个人资料数

✓ Stackoverflow 上相关的问题和关注者数量

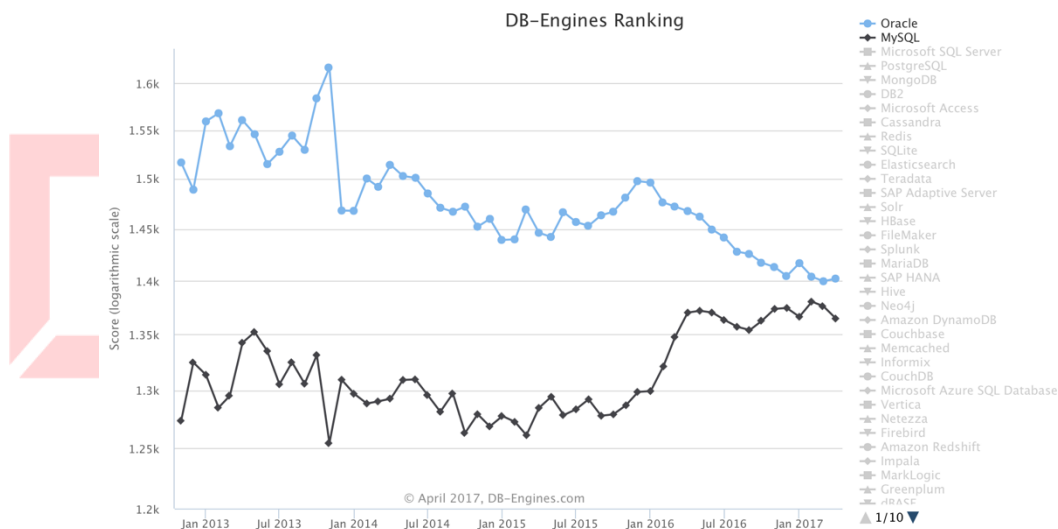
本期值得一提的是 Redis 超 SQLite 成为第 9 名，虽然相对第 10 名优势不到 1 分。

The logo for DBAplus, featuring the letters 'DBA' in large, bold, sans-serif font. The 'D' is pink, the 'B' is light blue, and the 'A' is yellow. To the right of 'DBA' is the word 'plus' in a smaller, green, sans-serif font.

## RDBMS 家族

### Oracle 12c Release 2 所有平台发布完毕

从分值趋势图，我们可以看到，从1月份起 Oracle 数据库正在艰难拉大与 MySQL 分数差距。分值上千的数据库三巨头中，Oracle 公司的 Oracle 和 MySQL 一直稳居状元和榜眼，SQL Server 始终是探花。但从2013年起，MySQL 作为老二，步步紧逼 Oracle，必欲取之而后快，并取得了惊人的成效，最少只距离30分。



那么 Oracle 近 2 个月来，相对于 MySQL 的微弱优势差距增幅来自于哪里呢？

首先是 Oracle 12cR2 的发布。自 3 月份发布了 Exadata 版本、Linux 版本和 Solaris 后，在 4 月 11 日，AIX 版本和 HP-UX 版本也相继发布了。至此，Oracle 12cR2 所有要发布的平台都已经发布，接下来就看具体的装机量了。

## Oracle Database 12c Release 2

(12.2.0.1.0) - Standard Edition 2 and Enterprise Edition

Microsoft Windows x64 (64-bit)	<a href="#">File 1</a> (2.8 GB) <a href="#">See All</a>
Linux x86-64	<a href="#">File 1</a> (3.2 GB) <a href="#">See All</a>
Oracle Solaris (SPARC systems, 64-bit)	<a href="#">File 1</a> (3.1 GB) <a href="#">See All</a>
Oracle Solaris (x86 systems, 64-bit)	<a href="#">File 1</a> (2.8 GB) <a href="#">See All</a>
HP-UX Itanium	<a href="#">File 1</a> (3.7 GB) <a href="#">See All</a>
AIX (PPC64)	<a href="#">File 1</a> (3.1 GB) <a href="#">See All</a>

其次，越来越多的最终用户在下载、安装和测试 Oracle 12cR2，甚至不少行业大客户已经在新的生产环境直接使用这个新版本了。所以，随之而来的各种安装、配置文章、用户问题咨询和搜索冲高了 Oracle 指数。

第三个因素，其实是 MySQL 的热度下降了 11 个点，是 Top 20 里下降最多的，原因不明。

经过我们的生产使用，有一些参数建议你在安装配置的时候作为参考（注意：只是参考，在你的环境怎样设置，请以你的专业服务商建议为准。）

参数名	建议值	备注说明
optimizer_adaptive_features	false	optimizer_adaptive_features 主要功能是为了在语句执行过程中实时收集表的统计信息，方便 Oracle 选择更准确的执行计划。  在 OLAP 数据仓库环境中可以获得较好的效果。  在 RAC 环境的 OLTP 系统中非常损耗性能。

<code>_drop_stat_segment</code>	1	12c 新特性 HeapMap 带入的参数，即使不启用该特性，Rman 备份也会有问题。bug 19543595。
<code>_common_data_view_enabled</code>	false	该参数会导致数据字典性能视图查询缓慢，12.1 中受害严重，官方说法是 12.2 已经解决，建议观察。
<code>_optimizer_dsdire_usage_control</code>	0	controls optimizer usage of dynamic sampling directives  和 optimizer_adaptive_features 一起使用。
<code>_optimizer_gather_feedback</code>	false	
<code>_optimizer_enable_extended_stats</code>	false	
<code>_column_tracking_level</code>	0	
<code>_optimizer_ads_use_result_cache</code>	false	
<code>_use_single_log_writer</code>	true	多 LGWR 特性参数，默认值是 adaptive。

参考链接: <https://blog.dbi-services.com/sql-plan-directive-disabling-usage-and-column-groups/>



<https://mikedietrichde.com/2016/03/04/parameter-recommendations-for-oracle-database-12c-part-i/>

## MySQL 5.7.18 发布

2017 年 4 月 11 日, Oracle 发布了 MySQL 5.7.18。新的版本补充了部分特性, 并修正了不少 bug。

例如:

- Windows 构建现在使用默认运行时库 (构建使用 / MD 标志)。 (bug # 25611609)
- 添加了用于使用 Developer Studio 12.6 进行编译的 CMake 支持。 (bug # 25384295)

详细信息参考: <https://dev.mysql.com/doc/relnotes/mysql/5.7/en/news-5-7-18.html>

## MySQL 5.7 将要废弃和已经废弃的特性

本期主要总结一下将要废弃或者已经在 5.7 版本中废弃掉的特性

### 1. 将要废弃

(1) ERROR\_FOR\_DIVISION\_BY\_ZERO/NO\_ZERO\_DATE/NO\_ZERO\_IN\_DATE SQL modes

(2) 使用 grant 来创建用户

(3) explain 不再支持 EXTENDED、PARTITIONS 关键词

(4) server 和 client 段参数: `--skip-innodb/--temp-pool/--ssl-verify-server-cert/--ssl/--ssl-verify-server-cert`

(5) 系统参数:

`log_warnings/binlog_max_flush_queue_time/innodb_support_xa/metadata_locks_cache_size/metadata_locks_hash_instances/sync_frm/character_set_database/collation_database`

(6) 加密函数: `ENCRYPT()/ENCODE()/DECODE()/DES_ENCRYPT()/DES_DECRYPT()`

(7) INFORMATION\_SCHEMA 库中表: `PROFILING/INNODB_LOCKS/INNODB_LOCK_WAITS`

(8) 在语句中使用: `\N` 来代替 NULL

(9) `PROCEDURE ANALYSE()`

(10) `mysqld_safe` 支持 `syslog` 输出

(11) 使用 `mysql_install_db` 来初始化 mysql

(12) C API:

`mysql_kill()/mysql_list_fields()/mysql_list_processes()/mysql_refresh()/mysql_shutdown()`

(13) 对 linux 命令 `replace` 的支持

(14) 对 DTrace 的支持

## 2. 已经废弃

(1) 旧的 pre-4.1 password hashing 格式废弃

(2) 数据类型: YEAR (2) 废弃

(3) 系统参

数: innodb\_mirrored\_log\_groups/storage\_engine/thread\_concurrency/timed\_mu  
texes/innodb\_use\_sys\_malloc/ innodb\_additional\_mem\_pool\_size

(4) alter table 时 ignore

(5) INSERT DELAYED

(6) mysql\_upgrade 参数废弃: --basedir/--datadir/--tmpdir

(7) 参数的简称已经废弃 (比如 --key-buffer 必须写成 --key-buffer-size)

(8) SHOW ENGINE INNODB MUTEX

(9) InnoDB Tablespace Monitor and InnoDB Table Monitor (PS: 可以从  
information\_schema 的相关表中获取)

(10) mysql 命令:

mysql2mysql/mysql\_convert\_table\_format/mysql\_find\_rows/mysql\_fix\_extensio  
ns/mysql\_setpermission/mysql\_waitpid/mysql\_zap/mysqlaccess/mysqlbug/mysq  
lhotcopy

(11) binary-configure.sh

(12) cmake 参数: INNODB\_PAGE\_ATOMIC\_REF\_COUNT

(13) innodb 启动和系统参数废

弃: innodb\_create\_intrinsic/innodb\_optimize\_point\_storage/innodb\_log\_chec  
ksum\_algorithm

## MariaDB 10.2.5 RC 版几个实用的特性改进

### 1、XtraDB 存储引擎由 5.6 升级到 5.7.14

(1) XtraDB 5.7 在支持多处理器和高度并发 CPU 线程的系统上，提供更持续的线性性能和扩展性。实现这一点的关键是通过改进 Oracle InnoDB 存储引擎的效率和并发性，来消除 InnoDB 内核中原有的线程争用和互斥锁定的现象。通过这些改进，MySQL 可以充分利用当今基于 x86 的商用硬件先进的多线程处理能力。

(2) 在线调整 `InnoDB_Buffer_Pool_Size` 不用重启 `mysqld` 进程，可以直接动态设置 `set global innodb_buffer_pool_size = 11274289152*2`; 需要注意的地方，在调整 `Buffer_Pool` 期间，用户的请求将会阻塞，直到调整完毕，所以请勿在白天调整，在凌晨 3-4 点低峰期调整。

(3) 支持在线回收（收缩）undo log 回滚日志物理文件空间。undo log 回滚日志是保存在共享表空间 `ibdata1` 文件里，随着业务的不停运转，`ibdata1` 文件会越来越大，想要回收（收缩空间大小）极其困难和复杂，必须先 `mysqldump -A` 全库的导出，然后删掉 `data` 目录，然后重新初始化安装，最后再把全库的 SQL 文件导入，采用这种方法进行 `ibdata1` 文件的回收。`innodb_undo_log_truncate` 参数设置为 1，即开启在线回收（收缩）undo log 日志文件，支持动态设置。

### 2、增加延迟复制功能

数据可能因为人为失误而遭受毁灭性的破坏，可能有人不小心 DROP 删除了数据库，或者刚上线的新版应用程序有一个严重 bug，把所有数据都变成了垃圾。为了防止这类问题，可以通过命令 `CHANGE MASTER TO master_delay=3600`（单位秒）设置一个延迟的从节点。这样做是有意义的，尤其是你的数据库很大的情况，可以快速追溯到 12 小时之前的快照，从而避免 Gitlab/AWS 事件。

注：之前的版本只能通过 `percona-toolkit` 工具实现。

范例: `perl /usr/local/bin/pt-slave-delay -S /tmp/mysql.sock --user root --password 123456 --delay 3600 --log /tmp/log/delay.log --daemonize`

**3、mysqlbinlog 工具增加--flashback 参数，实现误操作闪回功能，该补丁由阿里巴巴彭立勋贡献。**

之前我是通过 sed 命令实现了同一功能，具体可以参考：

<http://mp.weixin.qq.com/s/9jxcR6iz90WDsLuC8SwJhA>

新版本 mysqlbinlog 没有刻意去研究，感兴趣的朋友请参考手册。

#### **4、支持一个表有多个 INSERT/DELETE/UPDATE 触发器**

最大的好处是：原表已有触发器也可以支持使用 pt-online-schema-change 修改表结构了。



微软 SQL Server 产品人员透露 SQL Server 的新一代产品“vNext”即将面世，在巩固自己在传统 RDBMS 市场优势的同时向大数据、AI 领域开疆拓土。

**针对传统 RDBMS，新版本的主要新特性如下 (OLTP)**

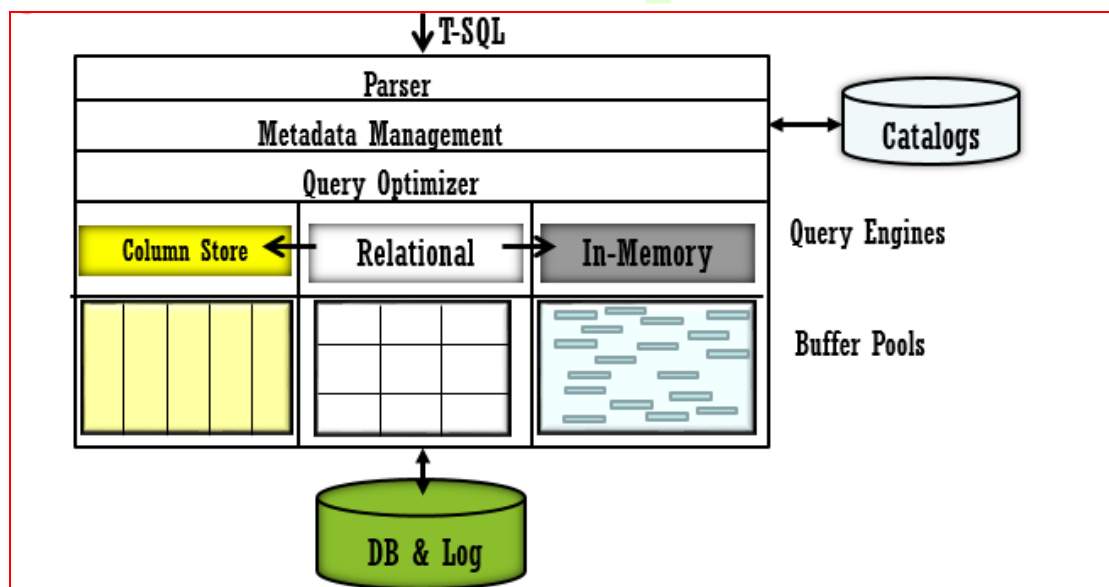
- 1、正式支持 SQL Server on Linux
- 2、支持 Linux 下 Docker 引擎
- 3、基于非 Windows Cluster (Cluster-less) 下的 SQL Server 高可用支持
- 4、部分 SQL Server 高可用特性如 (AG) 可跨 OS 平台工作
- 5、非聚集列式索引的在线创建与重建

此外，据报新版本中对已有的特性如基于内存执行引擎、基于列式存储执行引擎等均有增强，我们有理由有更多期待。

在现有版本 (SQL Server 2016) 中 In-Memory (Hekaton) 执行引擎针对其初始版本 (SQL Server 2014) 就有了显著提升，如包含更多的数据类型支持 (如 Blob 类型)、更大的建议表容量支持 (256G-2T)、触发器支持、基于 “Machine code” 的更多语法支持 (如 Left/Right join, select Distinct)、执行引擎支持并行等等。

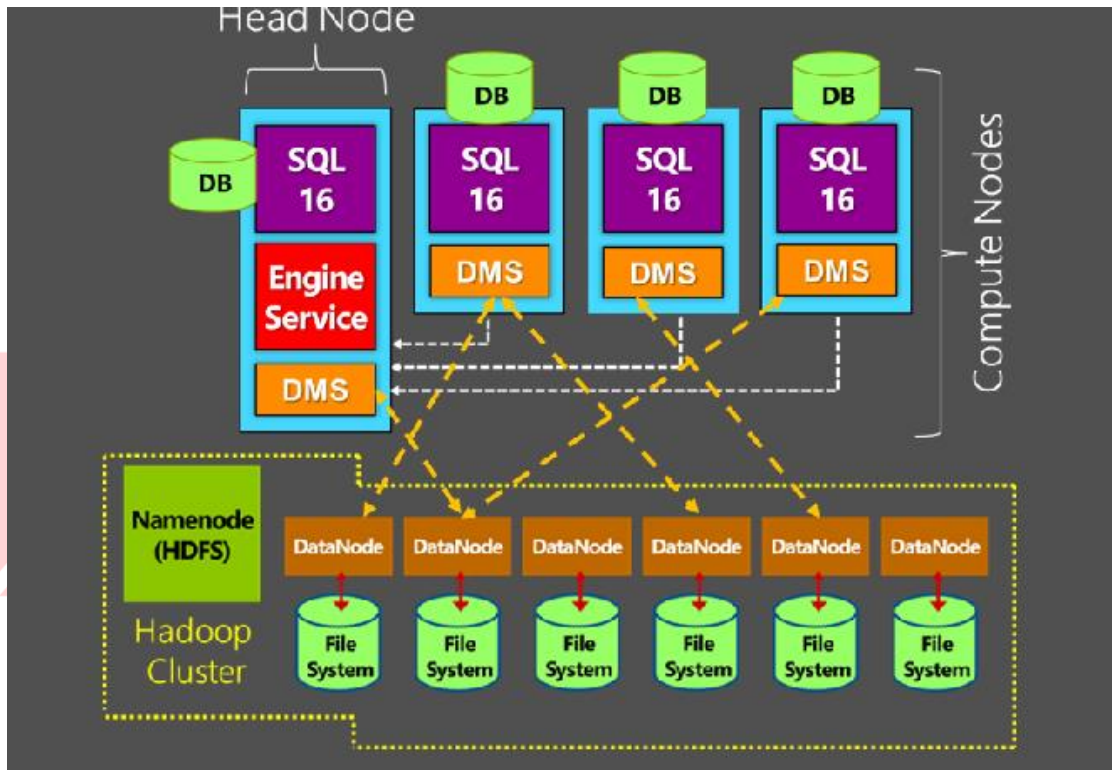
在现有版本 (SQL Server 2016) 中 Column Store (Apollo) 执行引擎针对其初始版本 (SQL Server 2012) 也有了显著提升，如在 SQL 2016 中可对非聚集的 Column Index 进行更新、过滤定义，并支持乐观并发级别，并对列式执行引擎 (Batch mode) 提供更多的操作支持如 Sort、开窗函数等等。

注：SQL Server 中自 SQL Server 2014 版本开始 SQL Server 就支持三类执行引擎基于传统的 Row 模式，基于列式存储的 Batch 模式，基于内存表优化的 native 模式混合运行执行。



针对大数据领域 SQL Server 将推出 SQLGraph，使得应用 SQL 处理图表成为可能。此外，针对 SQL Server 2016 推出的 PolyBase 也将得到加强。

SQL Server Polybase：构建在传统 RDBMS 与大数据平台下的一个桥梁，使得在传统的关系型数据库中分布式处理结构化与非结构化数据更加便捷（如对比 Hive）高效。



此外，针对人工智能领域，微软将利用自己在 AI 领域内的优势，结合现在数据产品，推出新的机器学习工具组件，SQL Server 首当其冲。结合其对大数据的支持，这将使得基于传统数据库领域的从业人员（OLTP、OLAP、SSAS、SSRS）向大数据、人工智能领域迈出可预见的一步。

(SQL Server R services, SQL Server Machine Learning tools, PolyBase, etc.)

除了更多的特性展现和功能支持外，SQL Server 也对当下的硬件设备提供更有力的支持应用。

在 SQL Server 2014 中支持 SSD 构建 BPE (Buffer Pool Extension)，在 SQL Server 2016 中更是紧密支持 persistent Memory (NVDIMM-N) 等硬件，并且基于计算机领域硬件的发展，SQL Server 自身也在进行更适应硬体的架构演化，如自动的 Soft NUMA 支持下有效减少 spinlock 竞争以提升并发、基于 Socket 的线程写入及锁分区等提升整体吞吐性等，这些改变及改进使得 RDBMS 在已经成熟的状态下，极限性能仍能得到有效提升。

SQL Server 开发、运维小知识：

SQL Server 2014 开始，tempdb 在面临最小化日志操作 (Bulk insert) 时，取消主动写 (Eager write) 设定使得 tempdb 面临的 IO 压力减轻，开发的朋友在面临相关应用时可以“更狠”地使用 tempdb；运维的朋友可能需要为实例准备更多的内存，以及合适的数据文件了。



PostgreSQL 将于 2017 年 5 月发布 beta 版本，已提交诸多特性如下：

#### 1、sharding 增强

- postgres\_fdw 多节点异步并行执行
- pushdown 增强
- 支持 Append 节点并行
- 支持分布式事务
- 主动防御 - 可配置是否允许执行不带 where 条件的 update\delete

#### 2、优化器改进



- 不完整索引支持复合排序

### 3、功能增强

- BRIN 索引更新 smooth 化
- CLOG oldest XID 跟踪
- hash index 支持 wal（灾难恢复）
- identify column（SQL 标准：T174）
- JSON 内容全文检索
- OLAP 增强 向量聚集索引（列存储扩展）
- QUERY 进度显示
- slave 提前苏醒
- slave 支持 WAITLSN 'lsn', time;用于设置安全 replay 栅栏
- SQL:2016 标准（之 SQL/JSON）Oracle 12c 兼容
- 内置分区表
- 分区表（hash, range, list）
- 分区表（list default）
- 动态视图 pg\_stat\_activity 新增数据库管理进程信息
- 匿名、自治事务（Oracle 兼容性）（background session）
- 后台运行（pg\_background）

- 唯一约束+附加字段组合功能索引
- 回滚范围可精细控制（事务、语句级）
- 国际化功能增强，支持 ICU (International Components for Unicode)
- 备库支持逻辑订阅，订阅支持主备漂移了
- 客户端 ACL (pg\_hba.conf 动态视图)
- 支持 EUI-64 格式 MAC 地址类型
- 支持分区表 ON CONFLICT .. DO NOTHING
- 新增数十个 IO 等待事件监控
- 流复制延迟评估，动态视图
- 自动预热 shared buffer (auto prewarm)
- 自由定义统计信息维度
- 角色化权限管理 - 新增几个内置角色（统计信息查看、管理用户会话）
- 触发器函数内置中间表
- 逻辑复制支持并行 COPY 初始化数据
- 新特性 - 流式接收端在线压缩 redo
- 变化 - 逻辑复制 pg\_hba.conf 变化，不再使用 replication 条目
- 多核并行增强 - tuplesort 多核并行创建索引
- 多核并行增强 - 并行 hash join 支持 shared hashdata，节约哈希表内存提高

## 效率

- 多核并行增强 – 控制集群并行度
- 多核并行增强 – 索引扫描、子查询、VACUUM、fdw/csp 钩子

## 4、安全增强

- SASL 认证方法 之 scram-sha-256 安全认证机制
- 任意 wal 副本数，金融级高可用与可靠性并存需求

## 5、性能增强

- 2PC 事务恢复阶段性能提升
- CLOG group commit
- GIN 索引 vacuum 锁降低
- hashed aggregation with grouping sets (多维分析) 更快，更省内存
- hash index metapage cache、高并发增强
- libpq 支持 pipeline batch 模式减少网络交互提升性能
- mergesort (Gather merge)
- OLAP 提速框架，Faster Expression Evaluation Framework (含 JIT)
- pg\_xact align (cacheline 对齐)
- Write Amplification Reduction Method (WARM) 消除某些情况下的 INDEX 写放大

- 分区表子表元信息搜索性能增强
- 推出 JIT 开发框架（朝着 HTAP 迈进）
- 间接索引（secondary index）

#### 6、流复制增强

- 支持可配置的 wal send max size

#### 7、逻辑复制

- 原理与最佳实践

### Greenplum Command Center 最近发布 3.2 更新

Greenplum Command Center 最近发布了 3.2 更新，相比于去年发布的 3.0，3.1 和 3.2 新增了 Kerberos 单点登录支持、用户权限管理、pg\_hba.conf 的图形化编辑、DCA 监控页面等新特性。

**Pivotal Greenplum Command Center** FEEDBACK DOCS SUPPORT Server: testtwins\_display

**Permission Levels for GPCC Access** Controls access to various aspects of GPCC on this server

**GPCC Access** [Change Selected to...](#)

<input type="checkbox"/>	Role Name ▼	Permission Level
<input type="text" value="Filter by..."/>	<input type="text" value="Filter by..."/>	
<input type="checkbox"/>	tpch_4	Self Only ▼
<input type="checkbox"/>	tpch_1	Self Only ▼
<input type="checkbox"/>	testuser	Operator ▼
<input type="checkbox"/>	readonly	Self Only ▼
<input type="checkbox"/>	gpadmin/kerberos-gpdb	Basic ▼
<input type="checkbox"/>	demo	Basic ▼
<input type="checkbox"/>	cc_user_krb2/test0	Self Only ▼
<input type="checkbox"/>	cc_user_krb2	Basic ▼
<input type="checkbox"/>	cc_user_krb	Basic ▼
<input type="checkbox"/>	cc_user	Self Only ▼
<input type="checkbox"/>	cc_test	Self Only ▼
<input type="checkbox"/>	cc_selfonly	Self Only ▼
<input type="checkbox"/>	cc_operator_readonly	Operator Basic ▼
<input type="checkbox"/>	cc_operator	Operator ▼
<input type="checkbox"/>	cc_basic	Basic ▼
<input type="checkbox"/>	cc_admin	Admin (superuser) ▼

**Permission Levels**

*See Queries & Metrics only* *See all screens*

**Self Only (default)**  
See/Cancel own queries  
group: none

**Operator Basic**  
See all queries, cancel own  
Cannot perform admin tasks  
group: gpcc\_operator\_basic

**Basic**  
See all queries, cancel own  
group: gpcc\_basic

**Operator**  
See all/cancel any query  
Cannot perform admin tasks  
group: gpcc\_operator

**Admin (superuser)**  
See all/cancel any query  
Perform admin tasks  
group: none  
role attribute: superuser

**Users must also have entry in pg\_hba.conf**  
(see [Authentication](#)) to sign in to GPCC

用户管理界面，可以让管理员方便地管理 GPCC 的使用权限

**Pivotal Greenplum Command Center** FEEDBACK DOCS SUPPORT Server: testtwins\_display Welcome, cc\_admin

**Host-Based Authentication** Controls user access for all GPDB activity on this server

**pg\_hba.conf** This version 2017-04-12, 20:29 [Load version...](#)

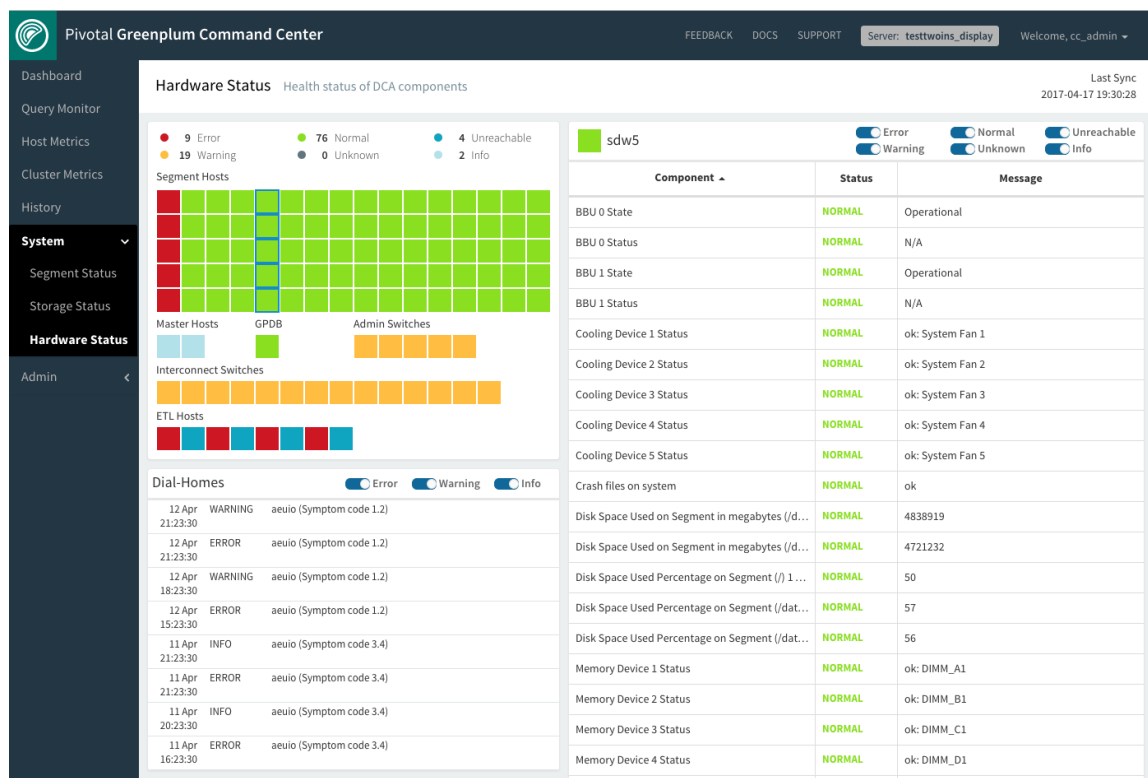
Type	Database	User	Method	Comment
local	all	cc_user	trust	# 1
host	all	cc_user	reject	# 2
host	all	gpadmin	10.152.10.134/32	trust # 3
host	replication	gpadmin	10.152.10.134/32	trust # 4
host	all	gpadmin	10.152.10.248/32	trust # 5
host	replication	gpadmin	10.152.10.248/32	trust # 6
host	all	gpadmin	192.168.122.1/32	trust # 7
host	replication	gpadmin	192.168.122.1/32	trust # 8
local	all	gpadmin	ident	# 9
host	all	gpadmin	10.152.10.149/32	trust # 10
host	all	gpadmin	192.168.122.1/32	trust
host	all	gpadmin	::1/128	trust # abcde
host	all	gpadmin	fe80::250:56ff:fea0:f439/128	trust
host	all	gpadmin	fe80::5054:ff:fe22:c22c/128	trust
host	all	all	10.152.10.128/32	trust #test4
host	all	all	10.152.10.154/32	trust #test6
host	all	all	10.152.10.255/28	md5
local	replication	gpadmin	ident	# test
host	replication	gpadmin	sameinet	trust
local	all	normal	md5	
local	gpperfmon	normal	md5	
host	all	normal	127.0.0.1/28	md5
host	all	normal	10.152.10.1/16	md5
host	all	normal	::1/128	md5
host	gpperfmon	normal	::1/128	md5

Options:      [Done](#) [Cancel](#)

**For guidance on managing pg\_hba.conf** consult the [postgres documentation](#)

*pg\_hba.conf* 的编辑界面，告别过去呆板的文本编辑器

带有输入检查功能和版本管理控制



DCA 硬件监控界面，硬件问题一目了然

除了以上这些，GPCC 还可以查看数据库的存储状况，各个 segment 的系统特性等等。另外，GPCC 也支持查看多个 GPDB 集群的状态。

在接下来的 3.3 中，我们会加入 Greenplum Workload Manager 的管理页面，敬请期待。

这些还不能满足你？想要其他新的功能？欢迎发邮件到：[gpcc-feedback@pivotal.io](mailto:gpcc-feedback@pivotal.io) 提出宝贵建议。

## 本期新秀：SQLite 发布 3.18.0 版本

3 月 30 日，SQLite 发布了 3.18.0 定期维护版本，此版本的功能初步实现了“PRAGMA optimize”命令。该命令可以在有需求的地方执行调用 ANALYZE 分析命令。应用程序需要在关闭数据库连接之前调用“PRAGMA optimize”。“PRAGMA optimize”语句在未来的版本中可能会在数据库维护的更多领域得到增强。

之前的版本，用来管理 SQLite 项目的控制系统中的 SHA3-256 哈希算法已经被 SHA1 更新替代。

以下是 3.18.0 版本详细的增强和优化记录：

- 1、增加了 PRAGMA optimize 命令；
- 2、SQLite 中 `sqlite_source_id()` 函数和 `sqlite3_sourceid()` C API 还有 `SQLITE_SOURCE_ID` 宏中的返回标识从以前的 40-digit SHA1 改为了 SHA3-256 哈希算法；
- 3、将 `json_patch()` SQL 函数加入了 JSON1 扩展中；
- 4、增强了 LIKE 语句优化；
- 5、添加 `sqlite3_set_last_insert_rowid()` 接口并 `fts3`，`fts4`，和 `fts5` 扩展组件上使用新接口来确保 `sqlite3_last_insert_rowid()` 接口总是返回合理值；
- 6、加强了 PRAGMA integrity\_check 和 PRAGMA quick\_check 所以他们可以验证检查约束了；
- 7、增强了对于 join 的查询计划，可以尽早的发现空表并且停止，以避免不必要的消耗；
- 8、增强了 `sqlite3_mprintf()` 类函数的接口并且打印函数的时候使用都好分隔

成千上万的整数标识，如果 “,” 格式标识用在 “%” 和 “d” 之间（例：“%, d”）；

- 9、增加了 `SQLITE_MAX_MEMORY=N` 编译时的 `-D` 选项；
- 10、给命令行模式增加了 `.sha3sum` 和 `.selftest` 两个命令；
- 11、开始增强 `SQLITE_LIMIT_VDBE_OP`，可以用来避免非信任用户在系统中执行大量的语句操作；
- 12、各种性能改进。

漏洞修复：

- 1、确保正确处理带排序序列的索引表达式，在 eb703ba7b50c1a5 中修复；
- 2、修复了日期和时间函数中 “start of...” 标识的 bug；
- 3、修复了递归触发器中一个复杂的 bug，在 3.15.0 版本中优化过程中遇到 `OP_Once` 时引起错误的 bug；
- 4、在 RBU 扩展中，加入额外的同步操作，以避免断电后数据丢失的可能；
- 5、嵌套语句在 `sqlite3_trace_v2()` 函数的输出总是带有 “--” 注释标识。



## NoSQL 家族

### MongoDB 3.4.4 发布

MongoDB 3.4.4 于 2017 年 4 月 22 日发布。这是个新版本发布，其中包括了一些 bug fix 以及一些聚合框架的功能完善。MongoDB 的版本发布大致为每年一个大版本（第二位数增大即为大版本），基本每个月一个小版本（第三位数）。一般小版本到 2 或者 3 就可以考虑从老版本进行升级。官方建议大家升级到 3.4.4，其中有一些对分片集群下数据块移动的问题有些修正，另增加了 `objectToArray` 及 `arrayToObject` 的聚合功能。

### MongoDB Atlas 云服务大规模扩张

MongoDB 最近宣布其从 2016 年中开始运行的 Atlas 云服务最近大力扩张。从本来只是在 AWS 的 5-6 个地区，扩展到 AWS 的所有地区，以及 Microsoft Azure 和 Google Computing Platform。和云服务商自己主推的 DBaaS 有所不同，Atlas 的口号是一个“云独立”的 MDBaaS，意思是您可以在任意一个云上部署，不必有长期被 lock-in 的顾虑。Atlas 更于近期推出了类似于亚马逊的 Free Tier，福利在这里：<https://www.mongodb.com/cloud/atlas>

### Redis 3.2.8 版本发布

2017 年 2 月 12 号发布了 Redis 3.2.8 版本。新版本有两个重要的 bug fix，其中第一个很关键：

1、Jemalloc 4.4.0 可能会在特定条件下发生死锁，bug 简单描述如下：

用户升级 redis 3.2.7 版本后，出现 cpu 使用 100%、无日志输出、rdb file 没有更新、所有连接失效等问题，经过定位发现是 Jemalloc 升级导致的问题。

详情请参阅 <https://github.com/antirez/redis/issues/3799>

因此 3.2.8 版本还原到以前使用的 Jemalloc 版本，并计划在获取更多有关错误原因的信息后再次升级 Jemalloc。

2、该版本修复了 MIGRATE 可能会导致服务器崩溃的问题，bug 简单描述如下：

redis cluster 在对 nodes 进行 resharding，迁移一些 big keys 时遇到：[ERR]  
Calling MIGRATE: Connection timed out

redis log 中出现 Redis 3.2.3 crashed by signal: 11, 只能重启修复。

详情请参阅 <https://github.com/antirez/redis/issues/3796>



## NewSQL 家族

### TiDB 近期发布 RC3 版本

4 月 27 日, TiDB 正式发布 RC3 版。该版本对 MySQL 兼容性、SQL 优化器、系统稳定性、性能做了大量的工作。对于 OLTP 场景, 重点优化写入性能。另外提供了权限管理功能, 用户可以按照 MySQL 的权限管理方式控制数据访问权限。

#### TiDB:

- SQL 查询优化器
  - 统计信息支持自动增量更新
  - 支持 Merge Sort Join 算法
  - 支持 Index Lookup Join 算法
  - 支持 Optimizer hint 语法
  - 优化 Scan, Join 和 Aggregation 算子的内存消耗
  - 优化 CBO 框架
  - 重构 Expression
- 支持更完整的权限管理
- 支持 HTTP API 获取表的数据分布信息
- 支持多个控制并发度的全局系统变量
- 新增大量 MySQL 内建函数
- 在大事务限制下, 支持通过变量控制大事务自动拆分小事务提交
- 修复 Bug
- 修复内存泄漏问题

#### PD:

- 支持 gRPC
- 支持 GC 掉过期数据
- 更高效的数据 balance
- 支持热点 region 的调度, 使机器负载均衡, 提升数据导入速度
- 性能优化
  - 加速 client TS0 获取
  - 提升 region heartbeat 处理效率
- pd-ctl 支持更多功能
  - 动态更新 replica 配置
  - 获取 TS0
  - 通过 ID 获取 region 信息

#### TiKV:

- 支持 SST format snapshot, 提升集群 balance 速度
- 支持 profile 内存使用
- 支持 SSE, 加速 CRC32 的计算
- 加速 transfer leader, 能更快的 balance
- batch apply, 减少 cpu 使用, 提高写入性能
- 优化 coprocessor 线程池的调度, 减少大查询对点查的影响
- 新版 loader 支持表级别并发导入, 也支持将大表划分成多个逻辑块并发导入, 提升数据导入速度

目前, TiDB 测试集群规模已突破 200T, 完成 500 个以上节点的测试, 单表最大 120T; 已有上线用户数十家, PoC 200 余家。在此感谢社区的小伙伴们长久以来的参与和贡献。源码地址: <https://github.com/pingcap/tidb>

## 本期新秀：CockroachDB

CockroachDB 的思路源自 Google 的全球性分布式数据库 Spanner。其理念是将数据分布在多数据中心的多台服务器上，实现一个可扩展、多版本、全球分布式并支持同步复制的数据库。

中文翻译为蟑螂数据库，也许被用在软件命名中很违和，但联合创始人 Spencer Kimball——一名前 Google 工程师则认为只有小强才能配得上他们的项目。

**“这个名字代表了项目最重要的两大特征：首先不用说，是生存性，其次是一种近乎于自主的散布性”。**

CockroachDB 是在 2014 年下半年推出了 Alpha 版，只实现了规模很小的一个功能子集。但 2016 年宣布获得融资的同时，CockroachDB 也正式推出了 beta 版，beta 除了实现当初设计的所有功能以外，还新增了对 SQL 的支持。开发者希望 beta 的推出能够在用户帮助下发现更多的 bug 和获得改进的建议。

## 大数据生态圈

### Hadoop 发布 2.8.0 版本

2017 年 3 月 17 号, Apache 基金会发布了 Hadoop 2.8.0 版本, 这是继 2015 年 4 月 Hadoop 2.7 版本以后的又一次重大的版本变更, 总共涉及 2917 个问题修复及新功能。

不过需要注意的是, 由于有些严重的问题还在修复, 测试过程中, 2.8.0 版本不建议立即投入到生产环境中使用, 生产用户应该等待 2.8.1/2.8.2 等后续版本。

版本更新的主要内容如下:

<http://hadoop.apache.org/docs/r2.8.0/index.html>

全部更新列表:

<http://hadoop.apache.org/docs/r2.8.0/hadoop-project-dist/hadoop-common/release/2.8.0/RELEASENOTES.2.8.0.html>

此次更新主要分成四个部分

通用部分:

- 1、在异步的分布式文件系统中, 支持异步调用的重试和故障转移。
- 2、新版本通过的 servlet 过滤器, 可以为 Hadoop 的 UI 提供对 XFS (Cross FrameScripting, 跨框架脚本攻击) 的防护。
- 3、S3A (Amazon Simple Storage Service) 改进: 除了使用 XML 配置文件的方式, 还可以使用 Amazon 的 STS 临时认证体系, 通过 Hadoop 的认证的 api 接口, 能够去读取 S3A 的认证, 从而提供了能够接入到 AWS 认证接口的能力。

- 4、Azure Blob Storage file system (WASB) 增强：添加了额外的 API 支持。
- 5、构建的增强：2.8 版本使用 Yetus 取代了 wrapper 的版本方式方式。并且提供了一个基于 Docker 的解决方案，用于快速的构建测试的环境。
- 6、在 LDAP 组映射服务中添加对 posixGroup 的支持。
- 7、提供了对微软 Azure ADL 的集成，使之可以作为 Hadoop 底层的文件系统。

#### **HDFS：**

- 1、WebHDFS 的增强：在 WebHDFS 中整合了对 CSRF (Cross-site Request Forgery) 伪造跨站请求的检查，在 WebHDFS 中，支持了 OAuth2，允许创建快照。
- 2、对于长时间运行的 balancer，允许使用 keytab 认证进行登录。
- 3、增加 ReverseXML 处理程序使其可从 XML 文件重建 fsimage。它能够帮助我们很容易的创建用于测试的 fsimage 并且能够手工的修复损坏的 fsimages。
- 4、支持多层次的嵌套的加密机制，Hadoop 从 2.7 版本开始支持一种文件系统级加密措施称为 encryption zone (简称 EZ)，用法是指定一个目录为 EZ 后该目录下所有文件的存取都会相应加解密，但对该目录下的子目录加密不会生效，除非再次配置。这就造成一些问题，比如一个目录是加密的，将其删除后默认会进 Trash，不管 Trash 本身是否加密，该文件夹移进 Trash 后就不加密了，这显然存在安全隐患。这个特性就是为了解决此类问题而使 EZ 可以嵌套。
- 5、该版本的重要改进之一，引入了一种新的 DataNode 状态上报协议，试图解决旧的基于心跳的协议在高负荷下的各种问题。其与旧协议的区别主要有三点：1) 处理 lifeline 消息开销较低；2) 其不会和 block 报告和用户操作争用名字空间的锁；3) 其使用了一批独立的 RPC 处理线程，使其与繁重的用户操作线程隔离。通过使用这种协议，理论上可消除集群繁忙时心跳处理的延迟和状态的误报。

6、能够记录 HDFS 操作调用的上下文日志到审计日志中。

7、为 dfsadmin 命令提供了一个新的子命令 evictWriters，可立即终止该 DN 上的所有写块操作。一般用于 DN 退服时需要很慢的 writer 使退服操作卡住的情况。

#### **YARN:**

1、新版本支持了在 windows 下监控 nodemanager 的 cpu 资源。

2、NodeManger 更加优雅的关闭：在老版本中 NodeMnager 只有在连接超时 10 分钟以后才能够被 ResourceManager 所感知，才能进行重连。在新版本中，NodeManager 可以通过取消注册的方式快速脱离。

3、当 AM 卡住的时候，新版本可以主动的使它失败。

4、进一步完善审计的功能。

5、Hadoop 内的 ApplicationTimeline Service (ATS) 是用来统计性能指标的，需要一个配置项以指定 TimelineService 的版本。

#### **MAPREDUCE:**

1、允许在提交 MR 任务的时候，指定节点的标签。

2、创建一个工具，以将聚集的 log 打成 HAR 包。

### **HDB (Apache HAWQ) 2.2.0.0 版本发布**

Pivotal HDB 2.2.0.0 企业版于 2017 年 4 月正式发布，主要包括几个功能：

1、Apache Ranger 集成；

2、RHEL/CentOS 7 支持；



- 3、PXF 对 ORC 外部表的正式支持；
- 4、包含查询执行器、容错、管理工具等重要模块错误的共计 76 个修正。

Apache HAWQ 2.2.0.0-incubating 版本与 Pivotal HDB 2.2.0.0 平行发布，已经通过 PMC 的投票，正在进行 IPMC 投票阶段。它是 Apache HAWQ 的第一个二进制发布，包含了源码和安装包。从开源社区用户使用的角度来讲，它提供以下方便性和灵活性：

- 1、用户可以方便地安装和部署 Apache HAWQ，省去编译、打包等繁琐过程；
- 2、用户可以按照需要配置 Apache HAWQ，利用改进的编译和打包机制，灵活定制自己的二进制安装包。

具体安装和打包方法可以参考：

<https://cwiki.apache.org/confluence/display/HAWQ/Build+Package+and+Install+with+RPM>

### Apache Ranger 集成

Apache Ranger 是 Hadoop 生态系统中安全访问的重要部件，提供了用户信息同步、用户权限管理、以及用户行为审计等功能。通过中央可视化界面，Ranger 可对 HDFS、Hive、HBase、Yarn 及 Hadoop 系统中多个其他部件进行访问控制。通过 HAWQ 与 Ranger 的集成，用户可以直接在 Ranger 中进行 HAWQ 用户权限的管理，省去单独登陆到 HAWQ 中进行权限管理的步骤。

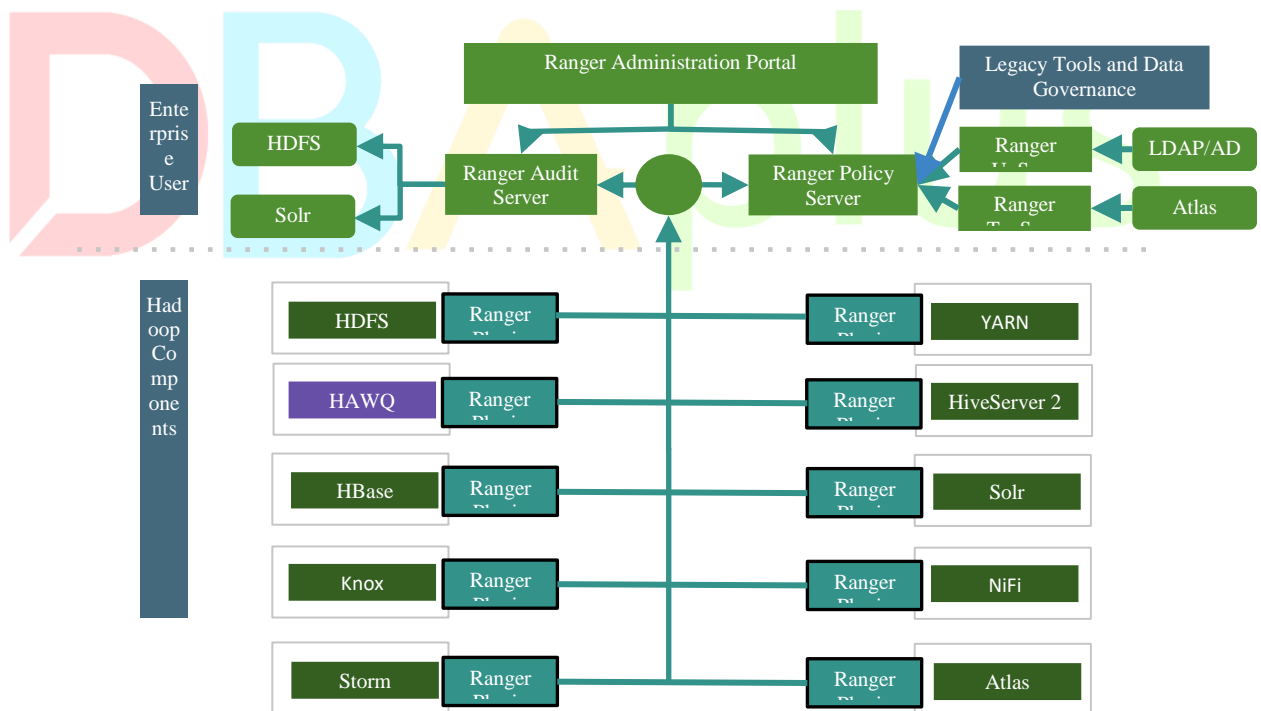
Ranger 的集成为企业用户提供了更加方便和统一化的用户安全管理功能，从而使企业用户可以使用以下部署：

- 1、LDAP 作为用户和组管理工具；
- 2、Ranger 作为用户权限管理工具，可自动从 LDAP 中同步用户信息，可以直接定义用户在不同 Hadoop 组件中的访问权限；

- 3、HAWQ / Hive / HBase 等各部件从 LDAP 中同步用户信息，在收到查询请求时将请求发送给 Ranger 进行验证；
- 4、用户访问的请求信息存储在 HDFS 或者 Solr 上，用作审计。

HAWQ 与 Ranger 的集成包括三部分：

- 1、HAWQ Ranger Plugin，包括在 Ranger Server 上注册的提供 HAWQ 资源查找的部件，以及 HAWQ 自身对 RangerBasePlugin 的扩展；
- 2、Ranger Plugin Service(RPS)，对 HAWQRangerPlugin 进行封装，并通过 REST 的接口对外提供服务；
- 3、HAWQ 内核端权限管理实现。在 HAWQ 核心代码端对用户提交的查询请求进行封装，并一并传递给 RPS 进行检查，并对用户进行相应的反馈。



重要错误修正如下，详细的修正可参考：

<https://issues.apache.org/jira/secure/Dashboard.jspa?selectPageId=123305>

- HAWQ-1347: QD should check segment health only
- HAWQ-1371: Fix QE process hang in shared input scan
- HAWQ-1385: hawq\_ctl stop failed when master is down
- HAWQ-1408: PANICs during COPY ... FROM STDIN

DBAplus

## 国产数据库概览

### OceanBase 压测成绩瞩目

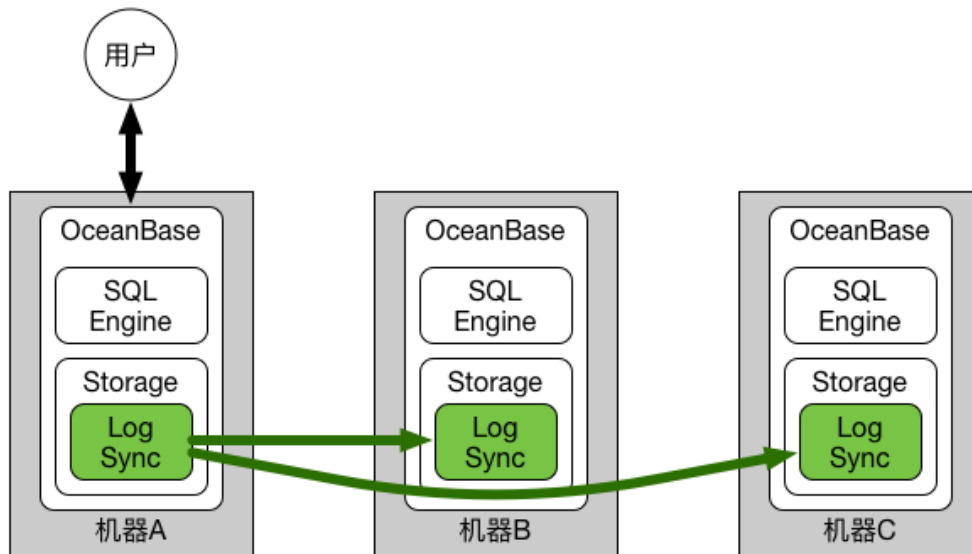
为进一步发挥 OceanBase 数据库对业务的价值，除高可用、高可扩展、高兼容性等既有优势外，OceanBase 在新版本中大幅提升了产品性能。高性能的 OceanBase 使得同样的硬件可以支撑更大规模的业务，不仅减小了系统频繁扩容的压力，更大幅度降低了成本。

测试 OceanBase 集群由 3 台服务器组成，数据表有 3 个副本，副本之间通过 Paxos 协议进行日志同步，确保任何单机故障不丢失数据；另有 3 台相同配置的服务器作为测试客户端。

在测试过程中，OceanBase 集群中只有一台服务器对外提供读、写服务，其他两台服务器不对外提供服务，仅进行日志同步和应答 (Paxos 协议)，以满足高可用的需求。3 台测试客户端只连接集群中对外提供服务的数据库服务器进行测试。

测试程序采用 sysbench 0.5 版本，测试结果：INSERT 场景突破百万 TPS，是全球首次三机同步下单机写入性能过百万的系统；SELECT 场景超过 250 万 QPS；OLTP 场景超过 60 万 TPS。

#### 测试架构



测试 OceanBase 集群由 3 台服务器组成，数据表有 3 个副本，副本之间通过 Paxos 协议进行日志同步，确保任何单机故障不丢失数据；另有 3 台相同配置的服务器作为测试客户端。

在测试过程中，OceanBase 集群中只有一台服务器节点(上图中的机器 A)对外提供读、写服务，其他两个节点不对外提供服务，仅进行日志同步和应答(Paxos 协议)，以满足高可用的需求。3 个测试客户端只连接集群中对外提供服务的数据库节点(机器 A)进行测试。

测试程序采用 sysbench 0.5 版本。

#### 机器配置：

CPU: Intel (R) Xeon (R) CPU E5-2682 v4 2.50GHz \* 2 ( 32 Core/64 Thread)

Memory: 512GB (16 \* 32 GB, 2133 MHz)

Disk: PCIe SSD ( 6.4TB \* 3)

NIC: 10Gb \* 2

上述配置的机器共 6 台，其中 3 台运行 OceanBase 服务；3 台运行测试客户端程序。

### 网络环境：

客户端和 OceanBase 服务器之间的 RTT

rtt min/avg/max/mdev = 0.047/0.052/0.069/0.004 ms

三台 OceanBase 服务器之间的 RTT

rtt min/avg/max/mdev = 0.050/0.054/0.073/0.010 ms

### 测试工具及参数设置：

sysbench 0.5 版本 (<https://github.com/akopytov/sysbench/tree/0.5>)

```
sysbench --mysql-host=***.***.***.*** --mysql-port=*** --mysql-db=*** --  
test=insert.lua --oltp-tables-count=10 --oltp-table-size=0 --max-  
time=180 --max-requests=0 --report-interval=1 --num-threads=1000 --  
mysql-user=*** --db-ps-mode=auto --oltp-read-only=off
```

```
sysbench --mysql-host=***.***.***.*** --mysql-port=*** --mysql-db=*** --  
test=select.lua --oltp-tables-count=10 --oltp-table-size=1000000 --max-  
time=180 --max-requests=0 --report-interval=1 --num-threads=1000 --  
mysql-user=*** --db-ps-mode=auto --oltp-read-only=off
```

```
sysbench --mysql-host=***.***.***.*** --mysql-port=*** --mysql-db=*** --  
mysql-user=*** --report-interval=1 --test=oltp.lua --db-ps-mode=auto --
```

```
num-threads=2000 --oltp-tables-count=10 --oltp_table_size=10000000 --  
max-time=180 --max-requests=0 --rand-type=uniform run
```

主要测试参数含义如下：

三种测试场景表数量都是 10 张，测试时间均为 180 秒；

INSERT、SELECT 两个场景的客户端线程数是 3000 个（1000 \* 3），OLTP 场景的客户端线程数是 6000 个（2000 \* 3）；

INSERT 场景，测试前表中无数据，测试完成时表的总行数超过 1 亿 8 千万；

SELECT 场景单表行数为 100 万；OLTP 场景，测试前单表行数为 1000 万；

性能结果：

	<u>insert.lua</u>	<u>select.lua</u>	<u>oltp.lua</u>
三副本性能值	1,071,932 (2.70ms/4.03ms)	2,597,304 (1.08ms/1.96ms)	627,607 (57.34ms/96.51ms)

数值格式：括号前面数字是 QPS/TPS；括号内第一个时间是平均 RT；括号内第二个时间是 95% RT；OLTP 的 RT 是整个事务的，包含了 18 条语句。

## 星瑞格多租户数据库

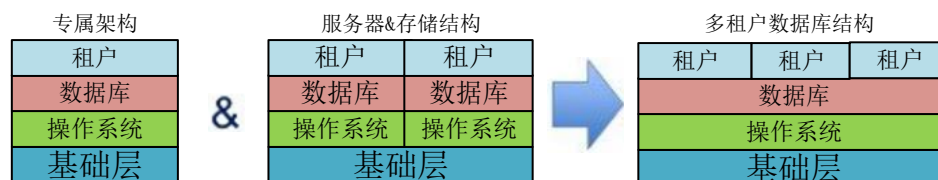
进入云计算的时代，“多租户”是一个基本的概念，无论是公有云还是私有云，多租户的支持是必须的，可以说多租户是云的基本技术。那什么是多租户呢？

多租户是指软件架构支持一个实例服务多个用户（Customer），每一个用户被称之为租户（tenant），软件给予租户可以对系统进行部分定制的能力，如用户数据的管理及应用规则，但是他们不能定制修改软件的代码。

那么在云时代多租户的应用数据实施中，星瑞格数据库一个重要考量点就是如何对应用数据进行设计，以支持多租户环境，而这种设计的思路，是要在数据的共享、安全隔离和性能间取得平衡，以实现云时代下多租户数据库以下特点及好处：

- 1、最大化利用资源，降低成本，数据层虚拟化
- 2、快速迁移，克隆，场合：测试，二次开发，数据迁移
- 3、易于管理，备份恢复、性能调节
- 4、简化升级，少打补丁，把多个数据库当主一个数据库来管理
- 5、资源动态调整，支持横向扩展
- 6、数据天然隔离，逻辑结构，物理结构，用户体系

星瑞格多租户环境减少了数据库数量，有效管理了数据库资源。见下图：



传统的数据库应用模式，仅服务于单个租户，数据库基本部署在封闭网络环境中，对于数据拥有者来说，这些数据是自己“私有”的，它符合自己所定义的全部安全标准。而在云计算时代，随着应用本身被放到云端，导致数据层也经常被公开化，但租户对数据安全性的要求，并不因之下降。同时，多租户应用在租户数量增多的情况下，会比单租户应用面临更多的性能压力。以下即对星瑞格数据库就这个主题进行探讨：多租户在数据层框架如何在共享、安全与性能间进行取舍。



常见三种数据层应用模式：

- 1、 独立数据库
- 2、 共享数据库，独立 schema
- 3、 共享数据库，共享 schema，共享数据

独立数据库顾名思义是一个单租户数据库系统，这个租户独享数据库实例提供的资源，属于最强的隔离度，租户的数据彼此物理隔离，备份与恢复都很灵活；共享数据库、独立 Schema 租户关联不同的数据库 Schema，这些 Schema 共享数据库实例资源，租户间数据彼此逻辑隔离，应用程序通过 Schema 实现和独立数据库一样的功能，备份恢复时需区分不同 Schema；第三种模式则在数据表级别实现共享，它在数据库部署上提供了最低的成本，但数据盘根复杂反而使编程更复杂（程序间数据处理需要用 tenantId 来区分不同租户），备份与恢复也更复杂，甚至是灾难性的。这三种模式的特点可以用下图来概括：



独立数据库	共享数据库 独立schema	共享数据库、 schema、数据表
高	基础设施成本	低
高	隔离度	低
低	资源共享	高
低	复杂度	高

上图展示理论性的结论，在具体实施场景下需要综合考虑才能决定哪种方式是合适的。例如，资源共享及复杂度上，认为独立数据库会低，共享模式就比较高了。但谷歌百度与银行保险对这个问题的思考就不会一样。谷歌百度考虑的就是低成本获得更大的资源以及数据的共享，而成本原因却不是主要考虑对象，对于谷歌百度而言，共享度越高，软硬件资源的利用效率更好，成本也会更低。当然谷歌百度也要解决好租户资源共享和数据共享带来的安全与性能、扩展性等问题。

## 星瑞格多租户的基本原理：

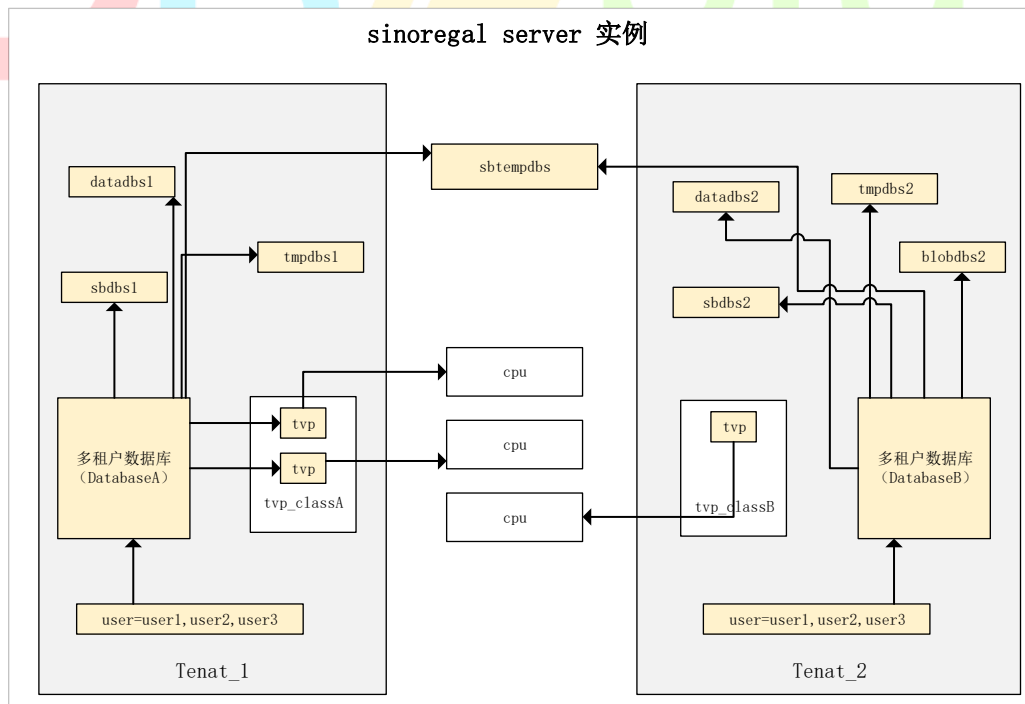
首先，在这种架构下星瑞格数据库只有一个数据库实例，所有多租户数据库共享星瑞格数据库实例的共享内存，包括驻留段、虚拟段、消息段，但多租户使用自己的专属进程，充分利用星瑞格数据库所具备的动态可伸缩性的特征减少 OS 进程资源的消耗。

其次，星瑞格多租户数据库只允许该多租户数据库用户存储管理用户数据，非该多租户数据库用户除管理员外不被允许链接该多租户数据库。

再者，和传统数据库一样，多租户使用星瑞格数据库实例的逻辑日志，物理日志，系统 dbspace 以及临时表空间，临时表空间也可以多租户数据库中设置为专属。

最后，多租户数据库与其他非多租户数据库使用一样的备份恢复工具。

下图展示了星瑞格数据库的逻辑结构：



从星瑞格多租户数据库基本原理可以看出，星瑞格的多租户数据库设计类似于第二种数据层多租户架构，也就是 N 个多租户位于一个数据库实例上，这样减少了数据库实例的管理操作，我们只需对一个数据库实例进行补丁升级操作即实现了对多个数据库的管理，并且星瑞格多租户利用星瑞格数据库的动态可伸缩性的线索处理机制尽可能的在基础架构成本、资源共享、数据独立性安全性做出很好的平衡。

## SequoiaDB 2.8 企业版正式发布

SequoiaDB 2.8 企业版本于近期正式发布，2.8 版本增加了多个新功能特性，特别在非结构化数据处理层面，增加了许多新特性以及大幅的性能、稳定性提升。

### 非结构化数据处理：

SequoiaDB 2.8 版本，对于 SequoiaDB 内置的块存储也就是非结构化数据管理引擎，增加了许多功能特性，也在性能和稳定性得到了大幅提升。

主要非结构化数据管理的增强包括：

2.8 版本，对于分布式块存储，实现元数据和 LOB 文件数据分离管理的机制，更方便用户在非结构化数据管理下使用。

非结构化数据管理，对 LOB 数据的切分功能进一步精细化，提升管理效率，总体性能提升 50%以上。

分布式非结构化数据管理，增强对于海量小文件的支持，减少海量小文件下的冗余数据。

### 数据库高可用：

在数据库高可用和数据安全方面，SequoiaDB 2.8 版本，在高可用机制上进行了大幅度优化，加入了备份数据的精细化运营管理功能，使得数据库的数据恢复中能够避免大量重复的工作，答复提供了企业用户的容灾能力和效率。

- SequoiaDB 提供同步日志归档能力，支持压缩和过期清理
- SequoiaDB 提供定时和定量方式的脏页刷盘能力，并实现异常重启时副本间数据校验能力，减少数据的全量恢复
- SequoiaDB 提供手动刷盘数据和日志的能力

#### 数据库匹配：

- 提供匹配条件(matcher)支持函数的能力
- 提供匹配条件(matcher)支持流水线的处理能力，能够对同一字段进行多次匹配和函数运算
- 提供对匹配的记录进行数组展开多条记录和只返回数组匹配项的能力

#### 可视化界面与监控全面升级：

- 提供节点和分区组信息的 WEB 监控能力
- 提供会话、上下文、事务等资源的 WEB 监控能力
- 提供主机内存、磁盘等信息的 WEB 监控能力
- 提供节点和分区组启停等的 WEB 操控能力
- 基于 WEB 的安装部署支持导入导出配置能力

## 感谢名单

最后要感谢那些提供宝贵信息和建议的专家朋友，排名不分先后。

贡献者单位/职务	贡献者	贡献领域
DBAplus 社群	杨建荣、杨志洪、林林	
大街网数据库&缓存负责人	代晓磊	Redis、MySQL
宜信技术研发中心数据库架构师	韩锋	MySQL、通稿审核
惠普金融 MySQL 专家	贺春昉	MariaDB
PingCAP 的 CTO 兼联合创始人	黄东旭	TiDB
MongoDB 中文社区发起人	唐建法	MongoDB
易车网数据架构师	高继伟	SQL Server
PGer	德哥@Digoal	PostgreSQL
中国移动苏州研发中心	姚昕	Hadoop
新炬网络开源技术专家	高强	SQLite
熊猫直播高级 DBA	杨尚刚	通稿审核
Pivotal 高级产品经理	马丽丽	HAWQ
SequoiaDB 巨杉数据库市场部经理	李方舟	巨杉数据库 SequoiaDB
Pivotal 公司 GPCC 研发	程元	GPCC
星瑞格售前技术总监	杨建鑫	星瑞格数据库
OceanBase 研发	蒋志勇(恒谦)	OceanBase

欢迎提供 Newsletter 信息，发送至邮箱：[newsletter@dbaplus.cn](mailto:newsletter@dbaplus.cn)

欢迎技术文章投稿，发送至邮箱：[editor@dbaplus.cn](mailto:editor@dbaplus.cn)



扫码关注

