

DBAplus

数据连接未来



# Newsletter 6月版




























## 目录

DB-Engines 数据库排行榜.....	1
数据库排行榜深度解读.....	2
RDBMS 家族.....	4
MySQL 发布高可用方案 InnoDB Cluster.....	4
Percona 发布 MySQL 5.7.18-15 版本.....	5
MariaDB 发布 10.2.6 版本.....	6
SQL Server 发布 2017 CTP2.1 版本.....	6
DB2 发布 LUW 11.1 mod1 fp1 版本.....	7
PostgreSQL 发布 10 beta1 版本.....	7
Greenplum 发布 5.0 Beta 版本.....	8
SQLite 发布 3.19 版本.....	10
NoSQL 家族.....	10
MongoDB 发布 Stitch Backend as a Service.....	10
Redis 发布机器学习应用模块 Redis-ML.....	11
本期新秀: ArangoDB 发布 3.2 beta 版本.....	14
NewSQL 家族.....	18
TiDB 结合 Ansible 简化集群部署, 即将推出 TiSpark.....	18
本期新秀: InfluxDB 发布 V1.2.4 版本.....	19
大数据生态圈.....	21
Hadoop 发布第三个测试版本.....	21
Druid 发布 10.0 版本.....	22
SnappyData 发布.....	25
GPText 2.1 版本发布.....	29
国产数据库概览.....	31
GBase 8t、8s 发布最新版本.....	31
达梦数据库发布最新版本 V7.1.5.186.....	32
人大金仓发布最新版本 V3.1.2.....	33
OceanBase 发布 1.4.2 版本.....	34
星瑞格数据库发布 12.10 FC 8 版本.....	34
推出 DBAplus Newsletter 的想法.....	39
感谢名单.....	40

## DB-Engines 数据库排行榜

以下取自 2017 年 6 月的数据，具体信息可以参考 <http://db-engines.com/en/ranking/>，数据仅供参考。

330 systems in ranking, June 2017

Rank			DBMS	Database Model	Score		
Jun 2017	May 2017	Jun 2016			Jun 2017	May 2017	Jun 2016
1.	1.	1.	Oracle  	Relational DBMS	1351.76	-2.55	-97.49
2.	2.	2.	MySQL  	Relational DBMS	1345.31	+5.28	-24.83
3.	3.	3.	Microsoft SQL Server  	Relational DBMS	1198.97	-14.84	+33.16
4.	4.	 5.	PostgreSQL  	Relational DBMS	368.54	+2.63	+61.94
5.	5.	 4.	MongoDB  	Document store	335.00	+3.42	+20.38
6.	6.	6.	DB2 	Relational DBMS	187.50	-1.34	-1.07
7.	7.	 8.	Microsoft Access	Relational DBMS	126.55	-3.33	+0.32
8.	8.	 7.	Cassandra 	Wide column store	124.12	+1.01	-7.00
9.	9.	 10.	Redis 	Key-value store	118.89	+1.44	+14.39
10.	10.	 9.	SQLite	Relational DBMS	116.71	+0.64	+9.92
11.	11.	11.	Elasticsearch 	Search engine	111.56	+2.74	+24.14
12.	12.	12.	Teradata	Relational DBMS	77.33	+1.00	+3.49
13.	13.	13.	SAP Adaptive Server	Relational DBMS	67.52	-0.23	-4.16
14.	14.	14.	Solr	Search engine	63.61	-0.16	-0.46
15.	15.	15.	HBase	Wide column store	61.87	+2.37	+8.88
16.	16.	 18.	Splunk	Search engine	57.52	+0.82	+12.29
17.	17.	 16.	FileMaker	Relational DBMS	57.08	+0.60	+8.39
18.	18.	 20.	MariaDB 	Relational DBMS	52.89	+1.91	+18.24
19.	19.	19.	SAP HANA 	Relational DBMS	47.50	-1.55	+6.23
20.	20.	 17.	Hive 	Relational DBMS	44.38	+0.91	-2.62

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

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

## 数据库排行榜深度解读

去 0 的浪潮并没有扑灭，Oracle 和 MySQL 的分值之差只有 6 分。从 2016 年 6 月以来，Oracle 数据库的分值减少了近 100 分，微软的 MSSQL 数据库、开源的 PG 和 MongoDB 增长之和 115 分。

云计算的时代已然到来，Oracle 数据库本身作为传统关系型数据库的老大，越来越受到威胁是必然的事情，但 6 月 22 日 Oracle 公司公布其第四财季净利 32.3 亿美元，并且自 2014 财年以来首次实现了年度盈利和营收增长，股价大涨逾 11%。这与分值形成反差。这个反差的主要成因与 DB-engines 分值的 6 大指标中的 3 个（在 Google/Bing/Yandex 上被提到的次数、Google Trends 被搜索的频率、Stack OverFlow 等 IT 网站被讨论的频率）有关，Oracle 数据库被包进 PaaS 解决方案，所以热度降低实属必然。

MySQL 数据库的分值一再上涨，大有盖过 Oracle 数据库之势，我认为也是极有可能的。这个突破点可能会到 MySQL 8.0 正式 GA 的时候，倒序索引以及 MySQL Sharding 的逐渐成熟，势必会引发更多的 MySQL 使用者。类似 DBAplus 社群这样的专业社区组织，纷纷增加的各种开源数据库峰会，在一定程度上也推动了它们分值的上涨。

国内云计算企业的背后支持，使 PostgreSQL 社区走在前列。而相比之下，MSSQL 是一个例外，没有一个专业的社区组织，微软自己似乎也不太在乎它的生态，但抵不住它庞大的使用量基数，让它稳稳占据老三的位置，而且分值还在一路攀升。

MongoDB 则跟前两类都不同，它先是作为一家企业存在，而后再开源，一度在 2016 年冲到了 DB-engines 排行榜第四位。作为一家创业 10 年的数据库厂商，MongoDB

是相当年轻，也是相当成功的，目前 3000 多个客户，800 多个员工，估值几十亿美元。目前 MongoDB 的社区是靠 MongoDB 公司主力去推动的。在 2017 年 5 月 21 日，DBAplus 社群在上海组织了一场 MongoDB 的线下沙龙，国内 TOP4 的云服务商大咖都来了，也谈到 MongoDB 在中国的前景，以及跟这些云服务商合作与竞争关系。MongoDB 是否在未来一年，奋起直追再冲上第四的位置，还是在第五的位置匍匐前行，取决于 MongoDB 是否放开胸怀。

作为大型企业的核心生产数据库，选择国外数据库的话，基本上只会在上述 5 种里选择。其他数据库，暂时还只能在替补梯队里。

值得欣喜的是，国产数据库这两年不仅在市场上攻城略池，在版本发布上进度也快了很多。下一期我们来专门点评一下国产数据库吧。



## RDBMS 家族

### MySQL 发布高可用方案 InnoDB Cluster

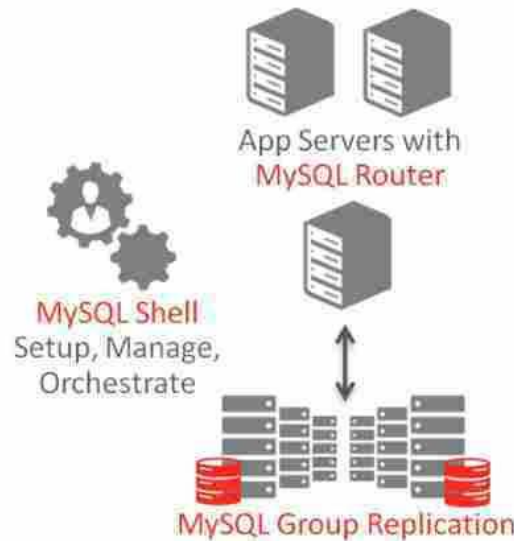
2017 年 4 月中旬，MySQL 官方发布了 InnoDB Cluster 的 GA 版本，InnoDB Cluster 是 MySQL 的一套完整的、全栈的高可用解决方案。

MySQL InnoDB Cluster 由下面三个不同产品和技术组成，它包含 MySQL Shell、MySQL Router、MySQL Group Replication。

其中 MySQL Shell 通过内置的 AdminAPI 来创建和管理整个 InnoDB 集群。

MySQL Router 2.1+ 确保客户端请求是负载均衡的，在任何数据库故障的情况下，都会传输到正确的服务器。

而 MySQL Group Replication 是近期发布的一个重大特性，通过 Group Replication 来将数据复制到集群的所有成员，同时提供容错、自动故障转移和弹性扩展等重要特性。



### Percona 发布 MySQL 5.7.18-15 版本

2017 年 5 月 26 日, Percona 发布了 MySQL 5.7.18-15 版本, 可以在 Percona 官网下载, 也可以通过 Docker 站点下载: <https://hub.docker.com/r/percona/percona-server/>

该版本基于 MySQL 5.7.18, 包含一些 bug 修复, 5.7.18-15 是 Percona 分支的最新版本。修复的 bug 信息如下:

- The server would crash when querying partitioning table with a single partition. Bug fixed [#1657941](#) (upstream [#76418](#)).
- Running a query on *InnoDB* table with [ngram full-text parser](#) and a LIMIT clause could lead to a server crash. Bug fixed [#1679025](#) (upstream [#85835](#))

## MariaDB 发布 10.2.6 版本

MariaDB 10.2 是目前稳定的版本。它是 MariaDB 10.1 的演进，具有几个亮点的全新功能，并具有 MySQL 5.6 和 5.7 的后端和重新实现的功能。

需要注意的是 Percona XtraBackup（从 2.4 开始）将不能与 MariaDB 10.2（和 MariaDB 10.1）压缩一起使用。但是，MariaDB 的分支机构 MariaDB Backup 可以进行压缩。它将包含在未来版本的 MariaDB 10.2 中。

具体信息可以参见：<https://mariadb.com/kb/en/mariadb/mariadb-1026-release-notes/>

## SQL Server 发布 2017 CTP2.1 版本

微软去年宣布将发布 SQL Server on Linux 版本，并在 2017 年年中推出正式版。从今年 5 月份发布的 SQL Server 2017 CTP2.1 版本来看，SQL Server 2017 的开发已经接近尾声，估计 7 月份会推出正式版。

在 5 月份发布的 SQL Server 2017 CTP2.1 版本中，微软基本上已经将传统四大件移植到其他非 Windows 平台（SSIS，SSRS，存储引擎，SSAS），在 CTP2.1 版本中还包括了两个新工具：

- mssql-scripter tool：快速生成针对数据库对象的 create 和 insert 语句。
- DBFS tool：更容易在 Linux 上通过 DMV 监控 SQL Server，通过读取文件来获取 SQL Server 的数据。

可以看到 SQL Server 2017 不是简单地将 SQL Server 2016 移植到非 Windows 平台，还增加了一些 SQL Server 2016 SP1 不具备的新特性。例如：

- 高可用方面引入了 Read-scale 可用性组；
- R、python 等机器学习方面功能；
- 支持 Linux 下 Docker 引擎。



## DB2 发布 LUW 11.1 mod1 fp1 版本

DB2 在 V11 版本与之前版本并无太大改善，最新推出的 11.1 mod1 fp1 只是侧重点更偏向于 Share Nothing 的集群架构上的 bug 进行了修复，但与之同时，IBM 也在推进基于 DB2 内核的私有云解决方案 dashDB local。

dashDB Local 是一个以 Docker 容器形式交付的配置数据仓库，同时 dashDB 支持 MPP 分布式架构，以便带来更高效的数据分析，并且可以与 Oracle、Netezza 等数据库兼容。

但是根据官网的描述，并没有从 dashDB local 上发现过多的特征和亮点，并且 dashDB local 和 DB2 在功能上存在很大程度上的重叠，两条独立的产品路线有可能会让用户感到迷惑。

## PostgreSQL 发布 10 beta1 版本

2017 年 5 月 18 日，PostgreSQL 发布 10 beta1 版本，新版本将支持逻辑订阅、分区表、quorum based 多副本同步复制、表分区、协议级多机 failover、多列统计信息、sharding 增强（聚合下推）、多核并行增强（索引扫描、位图扫描、merge JOIN 等）、安全增强（SCRAM 认证模块）等。非常值得期待。详见：

<https://www.postgresql.org/docs/devel/static/release-10.html>

PostgreSQL 积分连续数月增长，同比去年增长 61.94 分。

CrunchyData 开源 PostgreSQL 中间件，详见：

<https://github.com/CrunchyData/crunchy-proxy>

zombodb 开源，整合 Elasticsearch 作为 PostgreSQL 的搜索引擎，详见：

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

## Greenplum 发布 5.0 Beta 版本

相对于之前的 alpha 版本，beta 版最大的更新是支持了 4.x 中所有的扩展组件，也包括很多的 bug 修复和功能的增强，完整的 release notes 后面有介绍。Beta 版本的发布意味着 Greenplum 5.0 正式版中的全部计划的功能都已经基本完成，更意味着正式版很快就会发布。下面对 Beta 版的主要功能做一个大致的介绍。

### Greenplum 5.0 Beta 版的新增功能

Greenplum 支持 PostgreSQL 8.3 的完整功能以及某持 PostgreSQL 高版本的功能，相对 Greenplum 4.x，主要增加的功能如下：

- ‘Order by’ 可以支持 NULL first 或 NULL last；
- 支持新的数据类型及相关函数：Json、UUID、ENUM；
- Postgres 9.1 中的 XML 类型的操作函数；
- 支持任意复杂和组合类型的数组；
- 支持 PostgreSQL 9.1 的 Extension 功能；
- 重新实现了 ANALYSE 函数提升了性能；
- 支持 dblink 扩展；
- 支持 pgcrypto 扩展；
- gprestore 支持了 CASTs；
- 增强的 session 状态监控；
- 管理脚本和 pl/python 支持了 python 2.7；
- 默认启用 ORCA 优化器；
- 备份支持 Veritas NetBackup 7.7.3；

- 内置 Partner connector 支持；
- 支持 Binary COPY；
- COPY 命令支持直接将文件生成在 Segment 节点。

### Greenplum 5.0 Beta 支持的扩展组件

Greenplum 不仅仅是一个 MPP 的数据库，它还提供了丰富的扩展组件来满足各种各样的数据分析和处理的要求，比如扩展的存储过程语言 PL/Language、文本分析工具 GPTEXT、基于浏览器的图形化管理工具 GPCC、用于机器学习的算法库 Madlib 和处理地理信息的 PostGIS 等等。支持的模块包括：

- PL/Language: PL/Java, PL/Perl, PL/R, PL/Perl
- Madlib 1.11
- PostGIS 2.1.5
- Client/Loaders tools 5.0.0
- Greenplum Command Center 3.2.2
- Greenplum Gemfire Connector

Greenplum 5.0 Beta 企业版的支持平台：

- Red Hat Enterprise Linux 64-bit 6.x / 7.x
- CentOS 64-bit 6.x / 7.x
- SuSE Linux Enterprise Server 64-bit 11 SP4

获取 Greenplum 5.0 Beta：

官方的 Greenplum 安装包可以从 [Pivotal Network](#) 官方网站获取，文档在 [Pivotal Documentation](#) 可以找到，代码在 [github](#) 上，中文开发者 blog 在[这里](#)，用户的讨论组为 [greenplum-user](#)，开发讨论组为 [greenplum-dev](#)。欢迎大家使用，反馈以及贡献代码。

## SQLite 发布 3.19 版本

SQLite 在 5 月的版本更新比上月频繁。分别在 5 月的 22 日、24 日和 25 日发布了 3.19.0、3.19.1 和 3.19.2 版本。

3.19.0 版本为定期维护版本，此版本中对查询的执行计划进行了多方面的优化，并且修复了 6 个漏洞（详情请参考：

[http://sqlite.org/releaselog/3\\_19\\_0.html](http://sqlite.org/releaselog/3_19_0.html)）。还有一些奇怪的 bug。如果你使用之前的版本没有遇到问题，则不需要升级。

3.19.1 版本修复了 3.19.0 中查询优化器中暴露的一些 bug。从 3.19.0 版本开始，左连接运算符右侧的子查询和视图有时会进行扁平化处理。新优化器在开发者提供的所有案例和其他各种案例中工作良好，但是 3.19.0 发布上线之后，有些用户发现在某些场景中，优化会失败。标号 cad1ab4cb7b0fc344 中包含有相关示例。

由于近期版本在左连接扁平化优化过程中存在很多问题，所以随机发布了 3.19.2 版本，该版本解决了前两个版本中发现的所有问题。

## NoSQL 家族

### MongoDB 发布 Stitch Backend as a Service

MongoDB 的年度开发者大会 MongoDB World 最近在美国芝加哥举行。一个好的开发者大会与新产品的发布密不可分。MongoDB 利用今天的主题演讲发布了两个新产品，一个是 MongoDB 的云托管服务 Atlas 正式支持微软 Azure 和谷歌的 GCP 云平台，另一个是与 Atlas 一样棒的新产品：Stitch，它是一个基于 MongoDB 数据库之上的全新后端服务（Backend as a Service）工具。Stitch 可以帮助开发人员整

合不同的第三方服务，以及 Mongo 图表，一个新的类似于 Tableau-like 的商业智能服务。

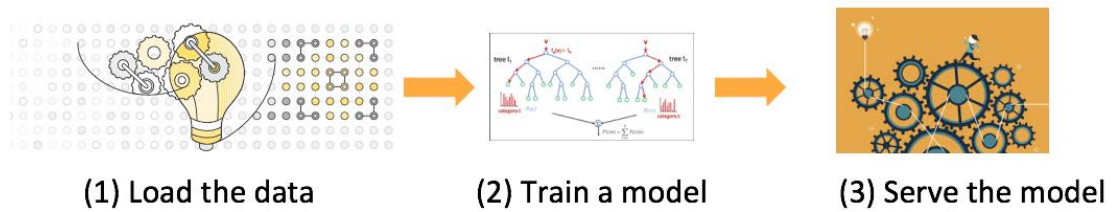
在许多方面，Stitch 标志着 MongoDB 首次涉足到其核心数据库服务之外的业务。MongoDB 首席技术官和联合创始人 Eliot Horiwitz 告诉我，MongoDB 的开发团队意识到 MongoDB 的大多数用户编写的应用程序会使用很多第三方服务的组合。Stitch 允许开发人员连接到这些外部服务和并轻松把数据从 API 服务拉回到 Mongo 的同名数据库里然后进行数据访问。理想情况下，这意味着开发人员不需要构建自己的安全和隐私控制。例如，他们可以用这些服务来专注于构建而非整合他们的应用程序。

Stitch 目前支持 Google、Facebook、AWS、Twilio、Slack、MailGun 和 PubNow 等服务的集成。但是正如 Horowitz 所强调的，开发者们也可以十分容易地将 Stitch 和其他提供 REST API 的服务结合起来。

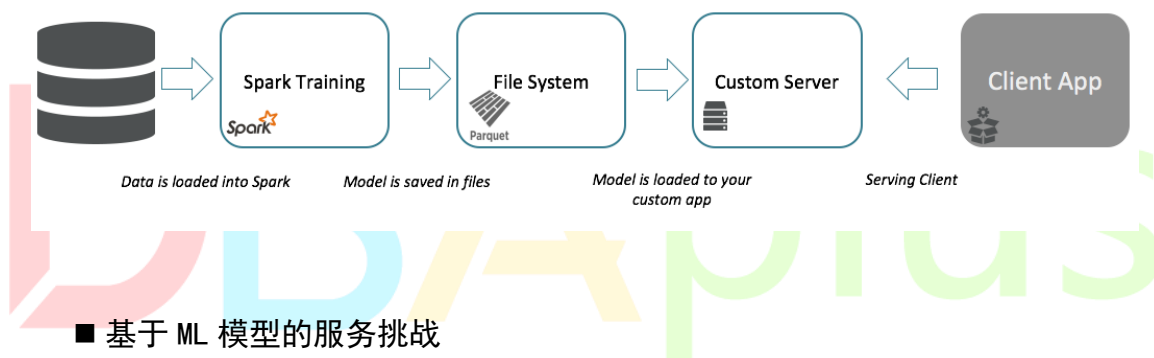
Stitch 现在处于公测阶段，MongoDB Atlas 用户可以免费开始试用。Stitch 的定价目前是基于开发人员需要传输的数据量而定。一旦 Stitch 问世，MongoDB 用户将在一个类似于 Atlas 的价格模型下使用其服务。中国的用户不用担心，虽然目前 Atlas 尚未在大陆开展业务，Stitch 将为有一个本地版本供非云用户试用。Mongo 开发者，让我们拭目以待吧！

## Redis 发布机器学习应用模块 Redis-ML

- Redis-ML - 机器学习模型服务器
  - 机器学习的世界



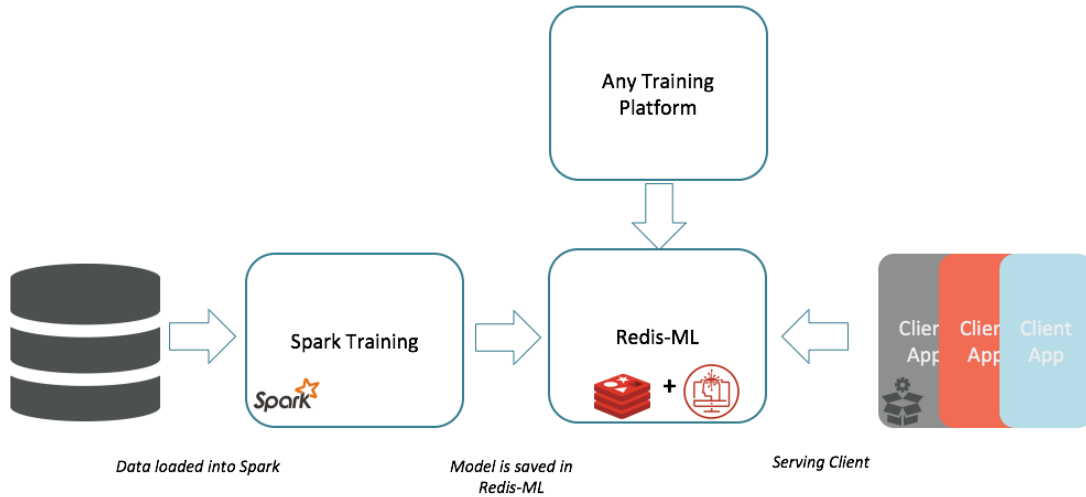
### ■ 传统的基于 Spark-ML 机器学习生命周期



### ■ 基于 ML 模型的服务挑战

1. ML 变得非常流行，但是 Models 却变得很大，很复杂；
2. 需要以更快的速度去分发/部署它们；
3. 速度和大小方面不能很好的扩展，需要保持他们在集群中的一致性以及使得他们持久化；
4. 可靠的服务很难做到 - 需要确保服务的可靠性；
5. 比较昂贵 - 需要投入更多的硬件资源和人力成本。

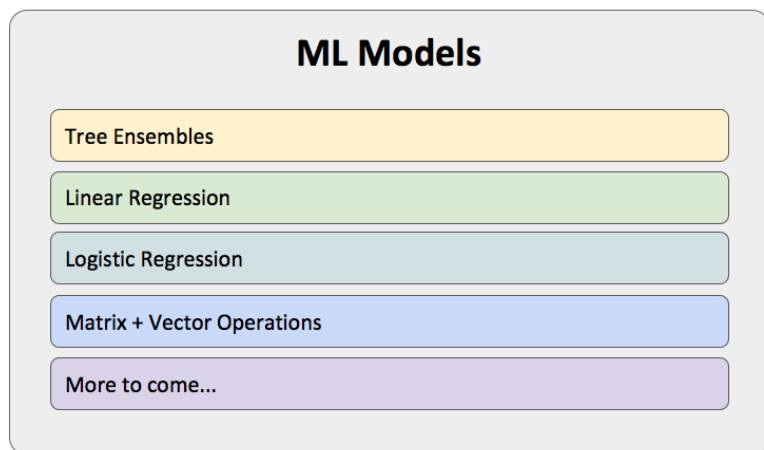
### ● 一个简单的基于 Redis-ML 的机器学习生命周期



- Redis-ML - ML 服务引擎

- 以“热模式”存储机器训练输出
- 直接在 Redis 中进行评估
- 轻松整合现有的 C/C++ ML lib 库
- 可以在运行过程中进行调整
- 享有 Redis 的高性能，可扩展性和 HA

- ML Modules



- 随机森林模型

- 一个决策树集合
- 支持分类和回归
- 分离节点
  - ◆ 按类别（例如：day == “sunday”）
  - ◆ 按数值（例如：age < 43）
- 决策是由大多数决策树所决定的

- Gini impurity - 基尼不纯度准则

了解 Redis-ML 的更多详细信息，可点击查看 DBAplus 社群文章：[《【独家】用 Redis-ML 模块实现实时机器学习！》](#)

## 本期新秀：ArangoDB 发布 3.2 beta 版本

ArangoDB 是一个开源的分布式原生多模型数据库（Apache 2 license）。其 Vision 是：利用一个引擎，一个 query 语法，一项数据库技术，以及多个数据模型，来最大力度满足项目的灵活性，简化技术堆栈，简化数据库运维，降低运营成本。Github: <https://github.com/arangodb/arangodb>

### 为什么叫“原生”？

ArangoDB 从第一行代码开始就是按照[多模型数据库](#)来设计编写，可以根据业务需求灵活地为数据建立模型，ArangoDB 支持 graph 图，document 文档和 key-value 三种数据存储模型。这三种数据模型的操作都基于同一个引擎，并使用统一的 query 语法- AQL 。





而且用户也可以在单次查询过程中混合使用多种模型，在执行查询过程时，无需在不同数据模型间相互“切换”，也不需要执行数据传输过程，并且这三种数据模型均支持水平扩展，这使得 ArangoDB 在速度和性能方面与其他数据库相比具有极大优势，[因此适用于高性能领域需求](#)。

## 什么是 AQL?

AQL- ArangoDB 查询语言 (ArangoDB Query Language)，是一种类似于编码的声明式查询语言。如果你了解 SQL，就能够对 AQL 清晰直观的语法轻松上手。不再需要在多种数据库技术之间互相切换，ArangoDB 让你只维护一个数据库，一个查询语言，来满足各种各样不同的需求。请查阅 [SQL/AQL](#) 的对比。以下是 AQL 的特殊性能：

- Joins
- Graph Traversals
- Aggregations
- Complex, powerful and convenient queries
- Multi-collection and multi-document queries with transactional semantics (single instance)
- Single document transactions (cluster)
- Geo-spatial queries
- Schemaless querying
- Extensible via user-defined functions

SQL:

```
INSERT INTO users (name, gender)
VALUES ("John Doe", "m");
```

AQL:

```
INSERT { name: "John Doe", gender: "m" }
INTO users
```

## 企业级版本的特性

- SmartGraphs

SmartGraphs 提供社区版本之外的更多高级性能， 以及在处理大型分布式图任务时几乎接近单台机器的更快的查询时间。[阅读更多](#)

- 增强加密支持

- Auditing

企业级版本下载: [download ArangoDB Enterprise Edition](#)

## 3.2 beta 版本新特性

- 社区版新特性

- 融合 RocksDB 存储引擎
- Pregal 分布式图形处理
- 有容错机制的 Foxx 在集群模式下根据您的需求扩展数据库
- Geo-cursor，根据“距离”分类数据信息
- Arangoexport 数据导出

3.2 beta 社区版下载: <https://www.arangodb.com/download-technical-preview/>

- 企业级版新特性

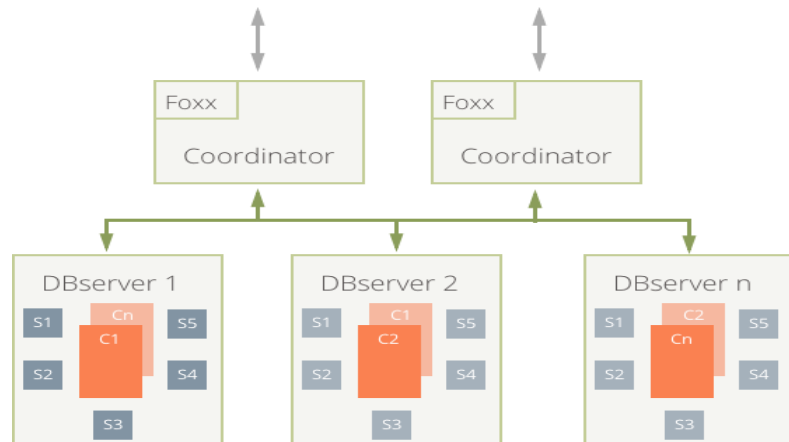
- SatelliteCollections

使用 SatelliteCollections 可以将集合定义为集群，并将集合复制到每台计算机。ArangoDB 查询优化器知道每个分片所在的位置，并将请求发送到所涉及的 DBServers，然后本地执行查询。使用这种方法，可以避免在分片集合的连接操作期间的网络跳数，并且响应时间可以接近于单个实例的响应时间。

具体实例请参阅:

<https://docs.arangodb.com/3.2/Manual/Administration/Replication/Synchronous/Satellites.html>

- 加强版 LDAP
- 闲时数据加密



3.2 beta 企业级版本下载: <https://www.arangodb.com/download-technical-preview-enterprise/>

DBAplus

## NewSQL 家族

### TiDB 结合 Ansible 简化集群部署，即将推出 TiSpark

TiDB 是一款定位于在线事务处理/在线分析处理（HTAP：Hybrid Transactional/Analytical Processing）的融合型数据库产品，模型参考了 Google 最新的分布式数据库 F1/Spanner。除了底层的 RocksDB 存储引擎之外，分布式 SQL 层、分布式 KV 存储引擎（TiKV）完全自主设计和研发。

TiDB 完全开源，兼容 MySQL 协议和语法，可以简单理解为一个可以无限水平扩展的 MySQL，并且提供分布式事务、跨节点 JOIN、吞吐和存储容量水平扩展、故障自恢复、高可用等优异的特性；对业务没有任何侵入性，简化开发，利于维护和平滑迁移。

#### TiDB-Ansible

日前，TiDB-Ansible 已正式在 GitHub 上开源。Ansible 是一款分布式集群配置管理工具。TiDB-Ansible 提供了一套 ansible-playbook 脚本，专门用于安装部署 TiDB 集群，以及完成相关的配置。使用 TiDB-Ansible 可以快速部署一个完整的 TiDB 集群以及配套的运维监控工具（包括 PD、TiDB、TiKV 和集群监控模块）。本部署工具具有如下特性：

- 通过配置文件定义集群拓扑，一键完成部署工作：
  - 初始化集群
  - 系统参数检查
  - 性能检查
  - 获取最新代码
  - 部署核心服务

- 部署运维监控模块
- 集群启停控制
- 系统扩容
- 滚动升级
- 数据清理

## TiSpark

TiSpark 是 Spark 和 TiDB 结合的产物，使 Spark 能无缝对接 TiDB，并提供原生 SparkSQL 所没有的优化支持，结合 Spark 生态圈原有的成员，TiSpark 将在同一平台提供超越 SQL 界限的多种 Workload 支持，让大数据分析无需复杂数据搬运。

除了标准的 Spark DataSource 接口的功能之外，TiSpark 通过扩展 SparkSQL 本身的 Planner 和优化器深度对接 TiKV（TiDB 存储层），调整并切分查询计划，使 SparkSQL 得以支持聚合下推，索引读取，KeyRange 下推和基于代价的优化等，从而更聪明地处理数据。而 TiSpark 对 Spark 的扩展是无侵入的，让用户可以通过第三方扩展包方式使用，无需更新现有 Spark 环境。

预计 7 月底，TiSpark 将正式上线。

更进一步文档请阅读：<https://github.com/pingcap/docs-cn>

## 本期新秀：InfluxDB 发布 V1.2.4 版本

InfluxDB 是用 Go 语言写的，专为时间序列数据持久化所开发，由于使用 Go 语言，所以各平台基本都支持，无需外部依赖。其设计目标是实现分布式和水平伸缩扩展。它有三大特性：

1. Time Series（时间序列）：你可以使用与时间有关的相关函数（如最大，最小，求和等）；
2. Metrics（度量）：你可以实时对大量数据进行计算；
3. Events（事件）：它支持任意的事件数据。

## InfluxDB

An open-source, distributed, time series, events, and metrics database with no external dependencies.

InfluxDB 的发布更新相对频繁，可见社区的关注度和热度。

2017 年 3 月 8 日发布了 V1.2.1 版本，3 月 14 日发布了 V1.2.2 版本，4 月 17 日发布了 V1.2.3 版本，5 月 8 日发布了 V1.2.4 版本，还有两个大版本 V1.3, V1.4 暂未发布，其中 V1.3 修复了大量的 bug，加入了更多的特性。

## 大数据生态圈

### Hadoop 发布第三个测试版本

2017 年 5 月 26 日, Apache 基金会发布了 Hadoop 3.0.0 后的又一个 alpha 测试版本, 这是继 2016 年 9 月 Hadoop 3.0.0 版本第一个测试版本以来的第三个测试版本, 是 3.0 版本线中涉及到安全相关的一个版本, 它修复了 alpha2 版本以来的一些安全漏洞。用户如果在使用 3.0 的 alpha1 或者 alpha2 版本的话, 推荐升级到 alpha3 版本。

不过需要注意的是, 由于是 alpha 测试版本, 不能够保证运行的质量, 以及 api 的稳定性。所以不建议在立即投入到生产环境中使用。

3.0.0 版本的主要更新:

<http://hadoop.apache.org/docs/r3.0.0-alpha2/index.html>

中文内容详见 DBAplus Newsletter 2 月版。

此次更新主要分成四个部分:

alpha3 版本更新的主要内容如下:

<http://hadoop.apache.org/docs/r3.0.0-alpha3/hadoop-project-dist/hadoop-common/release/3.0.0-alpha3/CHANGES.3.0.0-alpha3.html>

主要涉及了两个安全相关的问题:

YARN-6278 Enforce to use correct node and npm version in new YARN-UI build

<https://issues.apache.org/jira/browse/YARN-6278>

在新版本的 yarn 的 web UI 中，强制使用正确的 node 和 npm 的版本。

YARN-6336 Jenkins report YARN new UI build failure

<https://issues.apache.org/jira/browse/YARN-6336>

在 yarn 新 UI 中 Jenkins report 报错的修复

### 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，该版本引入了对 Apache Ranger 的支持。Apache Ranger 可以让用户以统一的接口来管理包括 HAWQ 在内的 Hadoop 生态产品的用户授权/审计等操作。基于 HAWQ Ranger REST 插件，HDB 提供了 HAWQ 和 Ranger 之间的策略管理集成，让管理员可以方便地通过 Ranger 来管理不同用户对 HAWQ 资源的访问权限。

另外该版本中还包含其它的一些重要更新：

- PXF 开始全面支持 ORC 文件格式。
- Pivotal HDB 开始提供 RHEL7/CentOS7 平台的支持。
- 在上一个版本基础上进行了 77 个错误修正和改进。

### Druid 发布 10.0 版本

Druid 0.10.0 包含来自 40 多位贡献者的 100 多项性能和稳定性改进以及 bug 修复。主要的新特性包括内置的 SQL 层、数值类型维度、Kerberos 认证扩展、Index Task 的改进、新的 “like” 过滤器、大列的存储、支持 Coordinator 和 Overlord 合并成一个服务、默认参数的性能提升以及 8 个新扩展。



## 内置 SQL 层

Druid 新增加了使用 Apache Calcite 驱动的内置 SQL 层，提供了 JDBC 和 HTTP POST 两种方式的 SQL API。JDBC 方式使开发者轻松地集成已有的应用，但这一版本并不是支持所有的 SQL 特性，并计划在以后的版本中支持更多的 SQL 特性。

使用 SQL 功能，必须先配置文件中设置 `druid.sql.enable=true`，详细文档参见 <http://druid.io/docs/0.10.0/querying/sql.html>

## 数值类型维度

Druid 在查询和数据摄入阶段中支持数值类型维度，用户可以将 `long` 和 `float` 类型的列作为维度摄入（例如开启 `rollup` 以后，在数据摄入阶段把数值类型的列当作分组的 `key` 而不是 `Aggregator`）。此外 Druid 查询可以把任意数值类型的列当成过滤条件或者分组查询的依据。

数值类型和字符类型的列之间有一个性能的权衡，数值类型列在分组聚合（`group by` 查询）要比字符类型列要快，但数值类型列没有索引，所以在 `filter` 过程要比字符类型列要慢。

## Kerberos 认证支持

Druid 增加 Kerberos 认证的扩展，使用户更加轻松使用开启 Kerberos 认证的 Hadoop。

## Index Task 的改进

Index Task（非使用 Hadoop 的离线任务）为了提升性能进行了重写，特别针对跨越多个时段生成多个 `shard` 的任务，`intervals` 可以自动设定而不再必须设定。但是设定了 `intervals` 和 `numShards` 可以减少数据摄入的时间。

此外，增加了“appendToExisting”设置允许在当前版本上增加 shard，而不是创建版本的 Segment 覆盖先前版本的。

### Like 过滤器

Druid 增加了 SQL “like” 方式的过滤器，例如 `foo like 'boo%'`。其实现的性能通常情况下比 `regex filter` 性能好，因此建议尽可能地使用 `like` 而不是 `regex`。经过实践发现，`like` 过滤器下的前缀过滤，类似 `foo%`，其性能显著地比等价的 `regex` 过滤器下的 `^foo%` 好。

### 大列存储

Druid 增加了单个大小超过 2G 的列存储支持。该特性通常不是必需的，因为在官方文档中推荐 Segment 的大小为 500M~1G。但在某些场景下有用，例如某个列的大小明显大于其他（例如大的 sketches）。

该特性对 druid 内置的列不需要任何配置就可以使用，假设开发自定义的 Metric，重写 `ComplexMetricSerde` 的 `getSerialize` 方法以便开启该特性。

### Coordinator 和 Overlord 合并

把 Overlord 合并到 Coordinator，可以简化 Druid 的部署。为此，在 coordinator 的配置文件中设置 `druid.coordinator.asOverlord.enabled` and `druid.coordinator.asOverlord.overlordService` 成相应的值，然后关掉 `overlord`。

该特性目前是实验性，默认关闭，但考虑在未来的版本中把它作为默认。

### 默认参数的性能提升

在这次的发布版本中修改了两项默认配置以便提升性能：

- `buildV9Directly`, 该项配置在 0.9.2 版本引入, 将其默认开启。它可以提升索引的性能, 直接构建 v9 版本, 而不是先构建 v8 然后再转成 v9。如有需要, 设置成 `false` 关闭该功能。
- V2 GroupBy 查询引擎也是在 0.9.2 版本引入, 将其默认设置为 v2。新的查询引擎在查询性能和内存管理方面有所改进提升。如果继续想使用老的引擎的话, 可以把 `runtime` 配置中 `druid.groupby.query.defaultStrategy`, 或者获取查询上下文中的 `groupByStrategy` 修改为 v1。前者适用所有的查询, 后者只适用当前查询。

完整文档详见 <https://github.com/druid-io/druid/releases/druid-0.10.0>

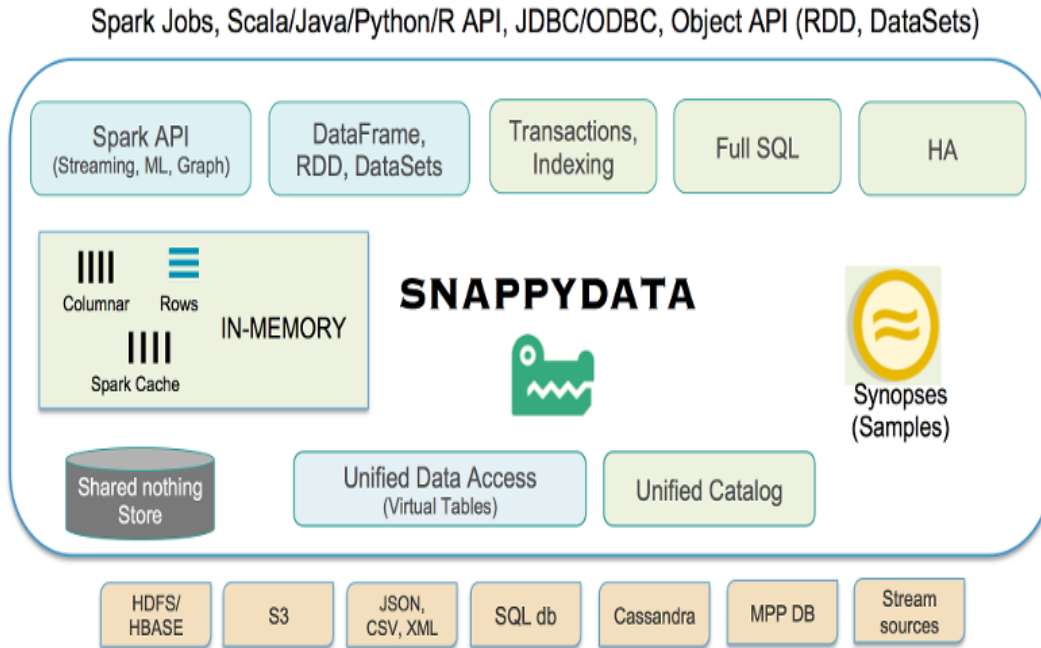
## SnappyData 发布

GemFire 全面升级版本 Snappy 面世。Snappy 将 Apache Spark 与内存数据库融合, 以提供能够在单个集群中处理流, 事务和交互式分析的数据引擎。

Snappy 是混合工作负载应用程序的高性能内存数据平台。基于 Apache Spark, SnappyData 为单个集群中的流媒体, 事务, 机器学习和 SQL Analytics 提供统一的编程模型。SnappyData 将实时数据源与具有 Spark 连接器的外部数据源进行统一。

SnappyData 的环境也比较适合尝试使用 Apache Spark 新特性。比如 [Spark 2.0](#) 将在今年发布, 届时将重构内存管理和[流系统](#), 拉取流数据将更加简单。

SnappyData 整体架构如下:



## 全面改善 Spark 数据存取性能

Spark 的数据源往往是外部的，比如 hdfs. 为了更好的响应时间查询不变的数据集，Spark 建议将外部数据源中的数据缓存为 Spark 中的缓存表。

在存储和取回数据方面，SnappyData 使用了一个分布式的数据存储，叫做 Snappy-Store，起源于 GemFire 的技术。它要么使用自己的数据存储，要么使用一种异步的回写式高速缓冲存储器连接另一个数据库，比如 [Hadoop](#) 或者 HDFS。这意味着，原来的数据可以直接使用 SnappyData，而不需要一些正式的数据迁移操作。

优化体现在：

- SnappyData 的存储层现在允许分区表的并置。
- 支持计划缓存，以避免查询解析、分析、优化、策略和准备阶段。
- 列级别统计信息的增强功能允许基于所述范围的批处理。从而大大提升了对这些查询的时间序列数据。

- 添加了替代哈希聚合和散列连接运算符，它们已经针对 SnappyData 存储进行了精细优化，以充分利用存储布局与矢量化，字典编码，提供了比 Spark 的默认实现更多的性能优势。
- 添加了一个有效的集合版本的 Kryo serializer，现在默认使用数据，关闭和较低级别的 netty 消息。这与前面提到的改进一起，大大减少了短期任务的总体延迟（从近 100ms 到几 ms）。这些还可以减少驱动程序消耗的 CPU，从而增强任务的并发性，特别是更短的任务。

### Snappy 统一实时和静态数据处理

Snappy 通过与 Spark 的无缝衔接，可以与不同的数据来源（如 Parquet、Hive、JSON / XML 文档数据库、NoSQL 数据库、关系数据库、流源和文本文件）之间提供高效的加载和同步。

### 速度比 Apache Spark 快 20 倍

经过测试，SnappyData 的性能超过了 Apache Spark 12-20 倍。

它使用了 apache spark 的内存数据分析引擎，在静态数据和流数据中实时分析 SQL。

存储和取回数据方面，SnappyData 使用了一个分布式的数据存储，叫做 Snappy-Store，起源于 GemFire 的技术。

### SNAPPYDATA 极大加强 Spark 开发能力

SnappyData 的核抽象与心编程 Spark 相同。它将所有数据公开为 Spark DataSet 或 DataFrame (Spark's Resilient Distributed DataSet)。使用这个简单的扩展，Spark 可以捕获以前需要单独引擎的各种处理工作负载，包括 SQL，流式处理，机器学习和图形处理。

与 Spark 开发的一致性有几个重要的好处。首先，应用程序由于使用统一的 API 而更容易开发。第二，组合处理任务更有效；而先前的系统需要将数据写入存储器以将其传递给另一个引擎，Spark 可以通常在内存中的相同数据上运行多种功能。

### SNAPPYDATA 极大加强 Spark 数据处理能力

- 与 Spark 兼容 100%：使用 SnappyData 作为数据库，还可以使用任何 Spark API --ML, Graph 等。
- 内存中的行和列存储：运行存储在 Spark 执行程序或其自己的进程空间中的存储（即计算集群和数据集群）。
- SQL 标准合规性：Spark SQL +几个 SQL 扩展：DML, DDL, 索引, 约束。
- 用于流处理的基于 SQL 的扩展：使用本机 Spark 流, DataFrame API 或声明式指定流以及您希望处理的流。在并行处理时，您不需要学习 Spark API 来处理流处理或其细微之处。
- 不唯一的 SQL：使用 SQL 数据库或使用 JSON 或甚至任意的应用程序对象。本质上，任何 Spark RDD / DataSet 也可以持久保存在 SnappyData 表（类型系统与 Spark DataFrames 相同）中。
- 使用概要数据引擎（SDE）的交互式分析：我们通过数据结构（如计数最小草图和分层抽样）介绍多个简介技术，以显着减少内存空间需求，并为分析查询提供真正的交互式速度。这些结构可以由几乎没有统计学背景的开发人员创建和管理，并且可以对运行查询的 SQL 开发人员完全透明。错误估计器也与简单的机制集成，通过内置的 SQL 函数来获取错误。
- 在 Spark 中对数据进行突变，处理：可以使用 SQL 在表中插入，更新，删除数据。我们还提供 Spark 的上下文的扩展，以便您可以在 Spark 程序中更改数据。SnappyData 中的任何表都可视为 DataFrames，而无需维护数据的多个副本：在 Spark 中缓存的 RDD，然后在数据存储中单独存储。
- 优化 - 索引：您可以对 RowStore 和 GemFire SQL 优化器进行索引，该优化器会在可用时自动使用内存中的索引。

- 优化 - 并置：SnappyData 实现了几个优化，以提高数据的局部性，并避免对分区数据集上的查询进行混洗数据。所有相关数据可以使用声明式自定义分区策略进行并置（例如，普通共享业务密钥）。当表不能共享公共密钥时，引用数据表可以被建模为复制表。副本总是一致的。
- 高可用性不仅仅是容错：数据可以立即复制（一次一个或一次一个批处理）到集群中的其他节点。它与基于会员的分布式系统深入集成，以检测和处理故障，即时为 HA 提供应用程序。
- 耐用性和恢复性：数据也可以在磁盘上进行管理并自动恢复。用于备份和还原的实用程序是捆绑的。

### 统一的 Data Sources API 访问 Spark SQL

Data Sources API 提供了一种用于在 Spark SQL 中访问结构化数据的可插拔机制。数据源不仅仅是将数据转换成 Spark 的简单管道。此 API 提供的紧密优化器集成意味着在许多情况下，过滤和列处理可以一直推送到数据源。这种集成优化可以大大减少需要处理的数据量，从而可以显著加快处理速度。

Data Sources API 的另一个优势在于，它可以让用户以 Spark 支持的所有语言操作数据，无论数据来源如何。例如，在 Scala 中实现的数据源可以由 pySpark 用户使用，无需任何额外的工作。此外，Spark SQL 可以使用单一界面轻松加入来自不同数据源的数据。

### GPText 2.1 版本发布

GPText 是 Pivotal 公司自主研发的内置 Greenplum 数据库的全文检索和文本分析引擎。具有易用性，分布式，高可用，可扩展，易维护，模块可定制化等特性。对大数据中非结构化的数据检索能提供毫秒级的响应，同时所支持的全文检索种类和语法非常丰富。

索引数据分布式存放在索引节点上，并复制拷贝提高高可用性。数据索引和检索都通过 Greenplum segment 服务器并行实现，支持通用的 search，词(term) 相关 search，facet 聚合 search 等常见的全文检索场景。

在最新的 GPText 2.1 的发布中，我们主要包含了以下更新：

- 针对 Greenplum 数据库中分区表 (Partition Table) 进行了优化。用户不需要为每个分区子表建立各自对应的 GPText 的索引，而可以共享同一个索引，从而避免了索引过多占用太多的内存资源。检索的时候，用户可以针对特定分区表的集合进行高效检索。同时提供了 partition\_status 和相关的管理脚本对分区表的索引进行支持。
- 提供了自动化升级工具 gptext-upgrade。让用户可以做到一键升级，如果升级过程发生了中断，用户可以重新允许升级程序即可。升级程序会自动检测上次升级的断点，并从断点处重新执行。除此之外，在安装过程中，如果检测到旧版 GPText 存在，用户也可以选择直接升级。



## 国产数据库概览

### GBase 8t、8s 发布最新版本

GBase 8t 近日发布最新版本 V8.7，以下是新功能一览：

（更多内容可参考官网及社区 <http://www.gbase.cn>；

<http://www.informixchina.net/club/>；）

- 高性能
  - 动态共享内存技术
  - 虚处理器技术
  - 磁盘技术
- 高可用性
  - HDR (High availability Data Replication, 高可用性数据复制)
  - RSS (Remote Standalone Secondary, 远程实时容灾)
  - SDS (Shared Disk Servers, 共享磁盘数据库集群)
  - ER (Enterprise Replication, 企业级复制)
- 高易用性
  - 监控运行状况
  - 管理高可用功能
  - 日志查看
  - 性能分析
  - 服务器管理
- 特殊功能
  - 时间序列
  - 内存加速

➤ 空间数据库

GBase 8s 近日发布最新版本 V8. 4，以下是新功能一览：

（更多内容可参考官网及社区 <http://www.gbase.cn>；）

● 高安全

- 完全达到国标（GB/T 20273—2006 信息安全技术 数据库管理系统安全技术要求）第四级的安全模型和安全要求（已通过公安部的等级评测）

● 高易用性

- 监控运行状况
- 管理高可用功能
- 日志查看
- 性能分析
- 服务器管理



**达梦数据库发布最新版本 V7. 1. 5. 186**

达梦数据库管理系统 7.0 版本，简称 DM7，是采用类 JAVA 的虚拟机技术设计的新一代数据库产品。DM7 基于成熟的关系数据模型和标准的接口，是一个跨越多种软硬件平台、具有大数据管理与分析能力、高效稳定的数据库管理系统。目前官网最新的可下载版本是 V7. 1. 5. 186，近期主要更新包括：

- 增加操作 ASM 文件系统的 jni 接口
- 改造 HUGE 表插入功能，大幅提升 HUGE 表插入性能
- 支持 QT4. X 和 5. X，可以基于 QT 直接访问 DM7
- dmPython 驱动支持 django 的 SQL 参数占位符
- dmflidr 语法上支持多字节、十六进制的分隔符和封闭符

- 扩充并行执行计划适用范围
- 故障重启 redo 性能优化
- 层次查询性能优化
- 系统视图性能优化
- 优化 distinct 操作性能
- 支持 to\_char 日期时间类型批量优化
- 提升 FAGR 操作符并发性能
- 提升获取分区表模板信息的性能
- 完善统计信息，使得统计信息更加精确
- 修复部分已知缺陷

### 人大金仓发布最新版本 V3.1.2

人大金仓新近推出的分析型数据库 KingbaseAnalyticsDB（以下简称：KADB）产品，包含大规模并行计算技术和数据库技术最新的研发成果：包括无共享、大规模并行处理技术（MPP）、按列存储数据库、数据库内压缩、MapReduce、在线扩容、多级容错等特点。

KADB 提供多种高级语言接口，方便外部应用服务访问，拥有良好的用户交互界面以及检测界面，简单易用，方便管理。

最新版本发布做了如下增强，旨在提高系统的可用性，以及覆盖更多应用场景：

- 审计功能支持：数据库可以对用户的操作进行审计，同时也提供了 SQL 命令对审计事件进行配置。数据库会自动将用户的操作事件进入到系统里，审计员可以在合适的时候对审计记录进行查询，确认是否有潜在威胁。
- 加密支持：数据库提供了 SQL 命令接口，用户可以通过这些接口定义和扩展数据库所能支持的加密算法。增强了系统的扩展性。

- 集群在线扩展：KADB 数据库增强了数据库扩展功能，在系统处理能力不足的，并且不中当前业务的情况下，对集群数据库进行扩展。
- MADlib 和 R 支持：用户可以直接在数据库层面进行统计分析。
- 与 Hadoop 交互：KADB 数据库可以直接访问 Hadoop 的数据。

## OceanBase 发布 1.4.2 版本

OceanBase 发布 1.4.2 版本，功能持续完善，性能进一步提升。

- 实现了转储功能，解决了大量数据导入或更新过程中内存不足的问题。在未实现转储功能前，只能通过每日合并来释放内存，对数据库服务有一定的影响。开启转储功能后，单个 ObServer 可根据自身的内存使用情况，将 memtable 中的一个或者多个分区转储到持久化存储中，无需考虑转储的分区是主还是备、也无需考虑主备之间的同步，系统可维护性进一步增强。
- 层次查询功能。Oracle 同款的层次查询功能，在新版本的 OceanBase 上也可以使用了，进一步简化了 Oracle 应用迁移到 OceanBase 的过程。
- 并行索引创建技术。在单个分区内，也可以并行创建索引了，索引创建时间缩短到原先的近 1/4，再也不用因为建索引慢影响每日合并的速度了。
- 批量数据插入的性能提升一倍。有效提升了大量数据装载的效率和用户体验。
- 常用表达式编译执行，性能提升 10 倍。目前编译执行支持 add/equal/and/case when 等常规表达式，正持续增加支持的表达式和数据类型种类。

## 星瑞格数据库发布 12.10 FC 8 版本

星瑞格数据库于近期正式发布 SinoDB 12.10 FC 8 版本，在最新的 12.10 FC 8 release 版本中，SinoDB 重点针对万物互联网应用场景下，数据库的技术支撑

特性进行了优化和整合，将适合边缘计算的嵌入式特性、时间序列特性、NoSQL支持特性进行加强，发布的SinoDB 12.10 FC8版本数据库能够满足物联网中所有对数据库的要求。我们知道，ARM在智能手机占有超过95%的市场份额，在平板电脑、机顶盒、网络等各种设备中占有主导性的市场份额。SinoDB支持ARM，是ARM设备中唯一嵌入的企业级数据库。此外，SinoDB小身材大实用，内存需求小，适合安装在物联网的许多设备中；每千兆字节可以保存上百万的记录；作为一款混合数据库，SinoDB在同一数据库中同时支持JSON/BSON和SQL应用；对时间序列（Time Series）和空间地理（Spatial/GIS）数据提供专门的高性能支持，每秒钟持续加载上百万条记录，分析速度比其它数据库快一个数量级。

### SinoDB 12.10 FC 8 物联网特性

- SinoDB 完美符合物联网世界对一个数据库的要求
  - SinoDB 提供基于 ARM 设备的嵌入式版本，提早布局物联网。
  - 较小的安装包大小（小于 100M）以及对运行内存要求不高，可以满足大部分边缘端设备的条件。
  - 高压缩能力减少存储空间。
  - 提供的分析功能可以助力边缘计算实现。
  - 特别支持物联网的时序数据，并提供 SQL 和 NOSQL 的应用接口以应对多种多样的应用需求。
  - 不管在 ARM 设备上还是在服务器端都提供企业级的能力和服务。
- 边缘计算的意义
  - 过滤或者聚合传感器数据。
  - 本地处理以及现场决策。
  - 降低处理时延，减少云平台网络带宽以及存储成本。
  - 80%的简单操作可以本地完成。
  - 更安全地完成实时计算。
  - 动态数据实时分析。
- SinoDB 的嵌入式特性

- 部署工具支持：为完成配置和调优的 SinoDB 数据库以及实例创建镜像，然后简单的复制部署到其他环境中。
- SQL Admin API：应用程序可以通过 SQL 接口来执行 DBA 的任务，可以通过编程的方式自主调控数据库。
- 预定义的 DB 定时任务：可以用来完成重要的 DBA 任务，比如自动统计信息评估和刷新。
- 预先设置的事件告警：极大地减少了应用程序去监控系统状况的开销，并能够对相应的事件做出响应。

## 独特的时间序列数据处理方式

### ● 时间序列介绍

- 众多的系统产生大量时间特性数据
  - ✓ 数据按时间有序排列。
  - ✓ 例如股票交易、智能电表、网络设备、智能设备等。
- 时间序列数据的特点：
  - ✓ 数据是依时间轴变化的，数据具有时间先后顺序。
  - ✓ 需要对数据按时间维度进行查询和分析。

### ● SinoDB 处理时序数据特点

- 单个设备所有的数据存储在一行中，数据在行末尾进行追加 (append)。
- 数据无需索引，只需在设备 ID 字段上创建索引。
- 按设备 ID 进行聚集存储在磁盘上，同时按时间顺序进行管理。
- 设备 ID 只需存储一次，而不是每一个记录都存储设备 ID。
- 磁盘上不存储时间戳 (Timestamps)，而是时间序列的相对偏移位置值。
- 数据缺失点使用占位符标记。

### ● SinoDB 时序数据处理优势

- 更少的存储空间。

比传统的存储方式减少 2/3 的空间。
- 更快的执行效率。

独特的存储优化技术提高了单位内存中的数据量，进而提高了执行效率。

- 专门定制的传感器数据流自动装载。

自动定期装载传感器数据并执行分析和聚合算法。

- 同时支持 SQL 和 NOSQL 的数据存储。

支持 REST/ODBC/JDBC/JSON 等应用接口。

- 预定义了 100 多个时序数据处理函数。

更便利地让应用构建自己的数据分析逻辑。

## SinoDB 支持 NoSQL

- NoSQL - Not Only SQL

- 什么是 NoSQL

NoSQL，指的是非关系型的数据库。是对不同于传统的关系型数据库的数据库管理系统的统称。NoSQL 用于超大规模数据的存储。这些类型的数据存储不需要固定的模式，无需多余操作就可以横向扩展。

- 为什么使用 NoSQL

用户的个人信息，社交网络，地理位置，用户生成的数据和用户操作日志已经成倍的增加。我们如果要对这些用户数据进行挖掘，那 SQL 数据库已经不适合这些应用了，NoSQL 数据库的发展也却能很好的处理这些大的数据。

- RDBMS vs NoSQL

	RDBMS	NOSQL
数据格式	高度组织化结构化数据	没有预定义的模式
查询语言	结构化查询语言（SQL）	没有声明性查询语言
存储方式	数据和关系都存储在单独的表中	键-值对存储，列存储，文档存储，图形数据库
一致性	严格的一致性	最终一致性，而非 ACID 属性

- SinoDB 处理 SQL 和 NOSQL 数据

- SinoDB 可以同时在一个数据库中处理结构化和非结构化数据产品。

- ✓ 将你设备的元数据存储为 JSON（例如：制造业）。

- ✓ 将设备的历史数据存储为时间序列数据（例如：温度）。

- SinoDB 同时处理结构化和非结构化数据。

- ✓ 允许 SQL 应用访问非结构化数据，允许 Mongo 应用访问结构化数据。

- ✓ 结构化和非结构化数据可以同时在一个查询中应用。

- “展示由制造商 X 制造的传感器在 1 月 1 日到 2 月 1 日的所有温度数据”。

- NoSQL 数据的访问方式

- REST 客户端可以发送任何程序的 HTTP 请求来访问 SinoDB 的数据。

- 开发者可以使用 MongoDB 驱动编写 MongoDB 应用访问数据。

- 支持 SQLI、DRDA 协议、NoSQL 与 SQL 的互译，可以访问关系型表、MongoDB 集合以及时间序列数据。



## 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 8 月 14 日~8 月 25 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

## 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	
PingCAP CTO 兼联合创始人	黄东旭	TiDB
MongoDB 中文社区发起人	唐建法	MongoDB
互联网游戏公司 DBA 主管	林勇桦	SQL Server
PGer	德哥@Digoal	PostgreSQL
中国移动苏州研发中心	姚昕	Hadoop
人大金仓产品经理	周文婷	kingbase
新炬网络开源技术专家	高强	SQLite
ArangoDB 中国区负责人	Shelly@Germanwifi	ArangoDB
Redis 中国用户组主席	张冬洪	Redis
Bloom Software CTO	杨旭钧	Spark
Pivotal 开发经理, 产品经理	雷达、李阳、杨瑜	HAWQ、GPDB、GPTxt
南大通用华东区销售总监	李宁	GBase
星瑞格售前技术总监	杨建鑫	星瑞格数据库
优酷土豆技术专家	张海雷	Druid
哈尔滨银行数据库架构师	洪烨	DB2
武汉达梦数据库有限公司		达梦
蚂蚁金服 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

