

Amoeba 技术介绍

陈思儒

2011/3/26

本文主要对 Amoeba 历史、内部技术以及未来发展进行相关的描述，希望本文档对 Amoeba 架构设计以及使用的人有所帮助。

目录

一、	引言.....	3
二、	Amoeba 简介	3
三、	为什么需要 Amoeba.....	3
四、	Amoeba 框架描述	6
	1. Amoeba 与应用层的关系	6
	2. 内部模块以及技术描述.....	8
五、	如何使用 Amoeba	10
六、	Amoeba 未来的发展	11
	1. 产品发展.....	11
	2. 应用 Amoeba 的领域.....	11
七、	联系我们.....	12

一、 引言

为什么会有 Amoeba 这个产品，在接受 InfoQ 中文站 Feng 采访的时候，也提到这个问题。主要原因是自己在使用 mysql，并且关注 MySQL Proxy 也有一段时间了。MySQL Proxy 的这种想法做的非常棒，它能够根据自己的想法去构造目标的 MySQL Proxy 应用，比如监控 SQL 执行、数据流量、读写分离。但由于我们使用 MySQL Proxy 并不能非常轻易地解决一些问题（读写分离、数据切分、水平切分、负载均衡），而是需要写大量的 Lua Script，这些 Lua 并不是现成的而是自己需要去写。这个工作对于并不熟悉 MySQL Proxy 内置变量、MySQL Protocol 来说是非常困难的。

因此带着这个想法去设想做一个非常容易使用、可移植性非常强的软件。Amoeba 就因此诞生了。为什么叫 Amoeba? 这个想法我是突然想到的，Amoeba 中文意思是"变形虫"，Amoeba 被设想为数据库代理的开发框架，它可以为符合 Amoeba 框架的任何数据库开发代理层。因此也比较象"变形虫"一样能够变成目标数据库的代理层软件。

二、 Amoeba 简介

开源 Amoeba(变形虫)项目,该开源框架于 2008 年 发布一款基于 amoeba 框架的产品 amoeba for mysql 软件而开始,框架自 2.0 以及未来版本 License 在 AGPL 协议下发布。Amoeba for mysql 这个软件致力于 mysql 的分布式数据库代理层，座落在 Client 与服务（数据库服务以及其他服务）之间。对客户端透明，具有负载均衡、高可用性、query filter、读写分离、可路由相关的 query 到目标数据库也可并发请求多台数据库合并结果。在 amoeba 上面你能够完成多数据源的高可用、负载均衡、数据切片等功能。目前在很多企业的生产线上使用。

Amoeba(变形虫)项目的定位：Proxy frameWork，专注后端服务代理开发。Amoeba 本身具备服务管理，网络连接管理、心跳管理器、路由规则控制等、后端服务负载均衡策略。

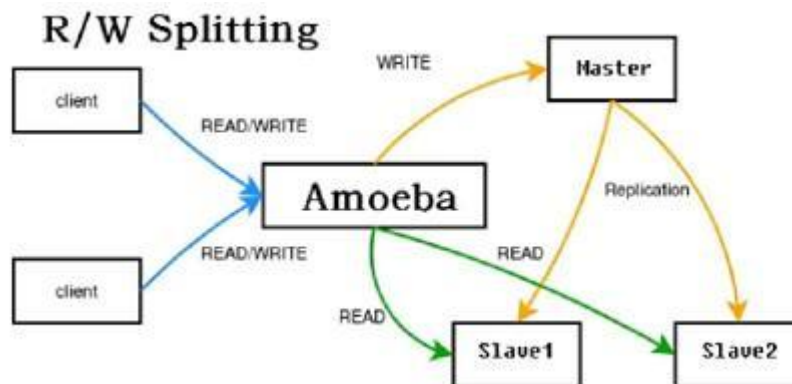
三、 为什么需要 Amoeba

随着传统的数据库技术日趋成熟、计算机网络技术的飞速发展和应用范围的扩充，数据库应用 已经普遍建立于计算机网络之上。这时集中式数据库系统表现出它的不足以及商业分布式方案代价大：

- 1) 集中式处理,势必造成性能瓶颈;应用程序集中在一台计算机上运行,一旦该计算机发生故障,则整个系统受到影响,可靠性不高,集中式处理引起系统的规模和配置都不够灵活,系统的可扩展性差。
- 2) 海量数据无法支撑,越来越多的互联网企业在使用着一些关系型数据库以及非关系型数据库,但目前很多产品目前没有一个很好数据库架构来解决海量数据存储以及高访问量问题,通常受限于软件、硬件、网络 IO 瓶颈,它们无法支撑更多的访问量以及数据容量,在目前互联网上很少有一种比较通用的解决手段。
- 3) 一般企业在这方面没有投入大量成本去做相关的技术研究,通常会在客户端做相应的技术改造,而这些方案是在特定语言以及特定的软件框架下形成的,很难有一个通用型,跨语言、开发框架难以复用的问题,也对相关的运维人员(比如 DBA)带来维护上面的困难,因为这个方案 DBA 很难介入。
- 4) 一些商业软件它们的费用非常昂贵,不仅软件成本包括硬件的实施成本也是一笔不小的开销。

引入 Amoeba 会带来什么好处呢?

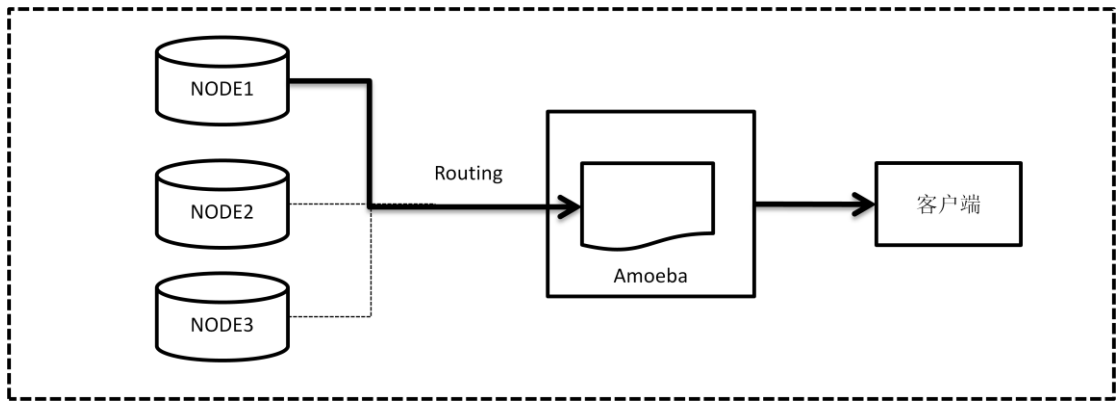
- 1) 引入 Amoeba 的成本很小,熟悉过程需要的时间也很少,目前 Amoeba 的使用文档也相对比较完善
- 2) 能够在很多场景解决单机大数据量问题、大访问量的问题
- 3) 很容易解决读写分离 (Amoeba for mysql)



- 4) 能够轻松的配置后端被代理服务的负载均衡, Amoeba 内部可用构建一个虚拟的服务池,该服务池有多个目标服务构成。他们之间可采用多种负载均衡策略:

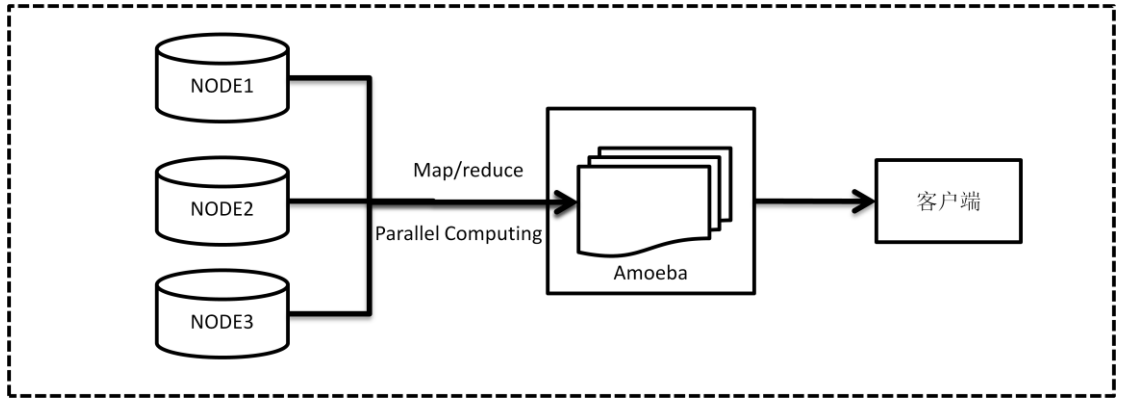
- i. 支持 HA 机制: Pool 中只有一个服务处于 active 状态, 只有 active Pool 不可用的时候才会切换到备用的 Pool
 - ii. 支持活动连接数均衡: amoeba 会优先选择当前正忙于处理业务请求的队列较少的服务进行接待当前的请求
 - iii. 支持权重配置: Amoeba 会根据配置比例进行轮询, 支持 1:1 或者其他比例, 例如 2:3 等
- 5) 数据切分: 强大的规则路由支持, 能够支撑足够想象空间的路由规则, 使得绝大多数数据切分的想法得以实现。
- i. 支持 Query routing

根据路由规则 Amoeba 非常轻松地能够解决这类问题



- ii. 支持 Parallel Computing

Amoeba 采用 NIO 无阻塞机制, 此时 Amoeba 的开销也非常小, 它所消耗的非 Analysis query 与 Merge results, 此时模式比较类似 Map/Reduce 或者 Parallel Computing: 它会将 query 解析以后发送给所有符合目标的节点, 然后每个节点分别计算自己的数据, 然后将结果返回到 Amoeba, Amoeba 接收到所有数据以后进行合并, 然后一起返回给客户端:



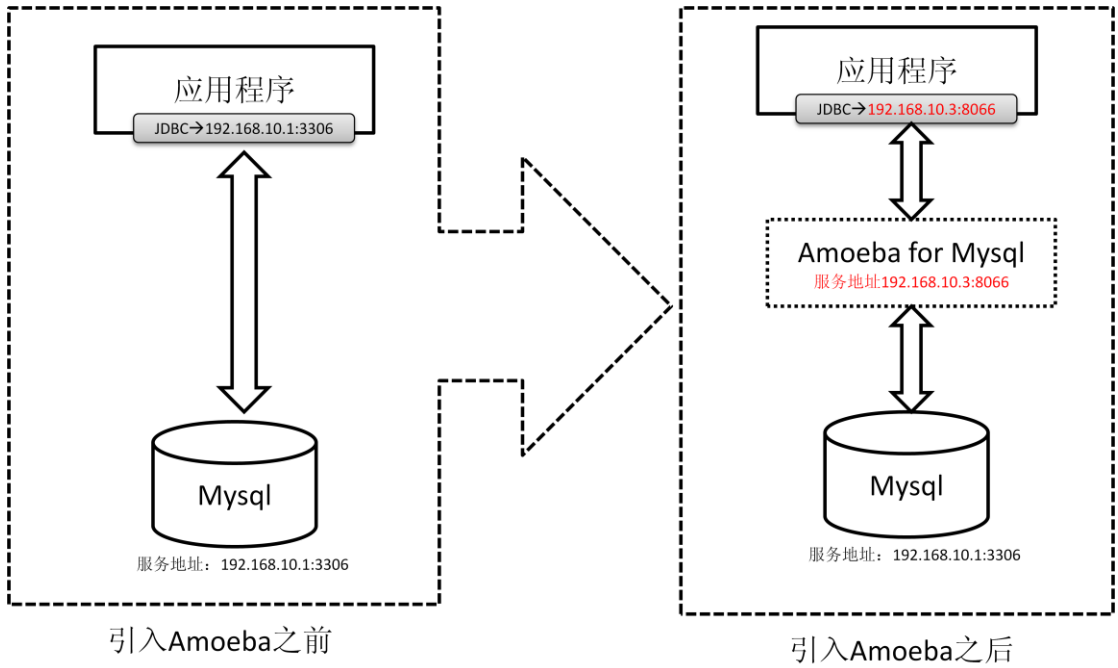
四、 Amoeba 框架描述

1. Amoeba 与应用层的关系

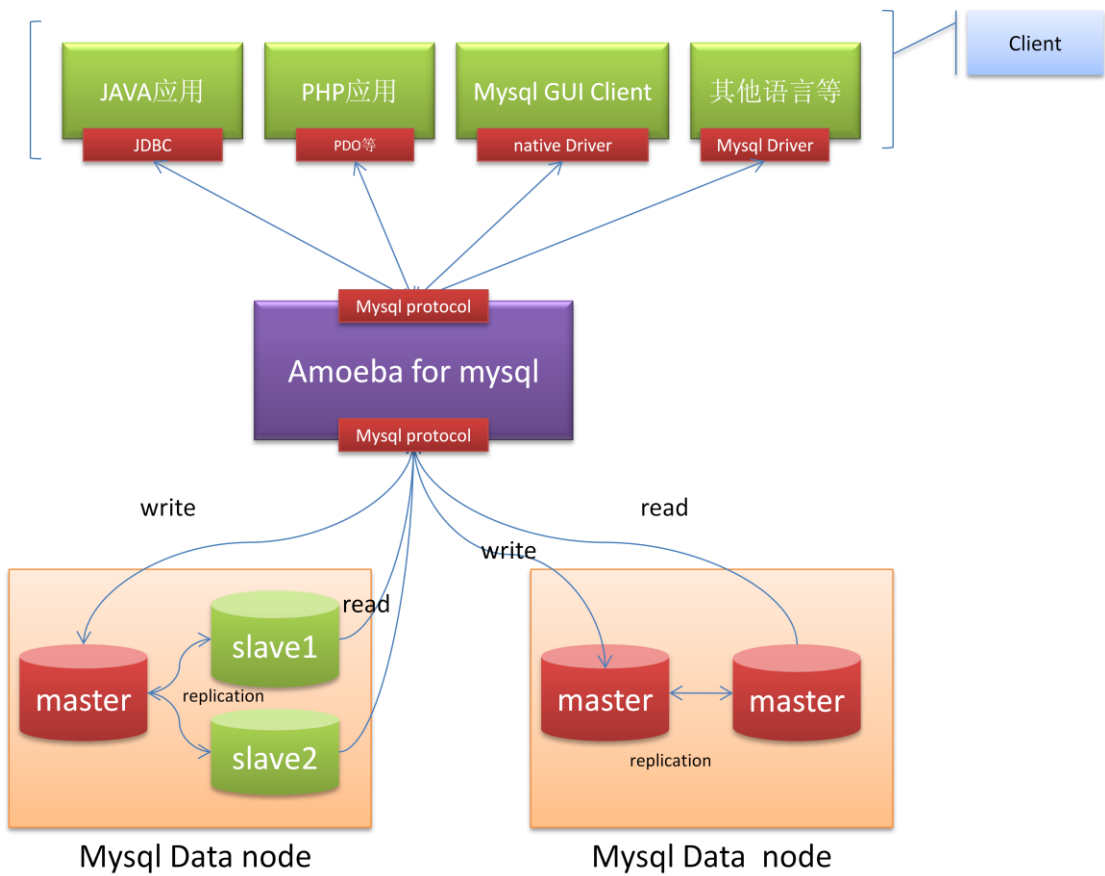
前面简介也提到了 Amoeba 的所处位置，座落在应用层与后端服务或者数据库等中间的一个中间件代理层。它前端连接应用层，后端维护目标服务以及数据库，属于中间件，为了能够让应用层更加透明更加方便使用，或者说引入 Amoeba 框架的代价对一个项目来说几乎可以不用考虑成本问题。因此 Amoeba 需要做的就是保持现有的通讯协议，我们举个简单的例子，就以 Amoeba for mysql 这个产品为例：

用户在使用 mysql 的时候通常会采用与开发语言相关的驱动程序去连接 mysql，比如 PHP 会采用 PDO，ASP 通常会采用 ODBC，java 语言会采用 mysql 出品的 Connector-J 的 JDBC 驱动。那么引入 amoeba for mysql 需要客户端在代码层面上如何改动呢？唯一需要改动的是，将 mysql 地址修改成 amoeba for mysql 的服务地址，当然 Amoeba 内部还是需要经过一系列配置的。

下面用一幅图来描述 java 程序引入 Amoeba for mysql 与未引入的区别：



其它语言与工具同样可以与 Amoeba 进行通讯，只要采用 Mysql 通讯协议即可，各种语言、客户端与 Amoeba 通讯示意图：

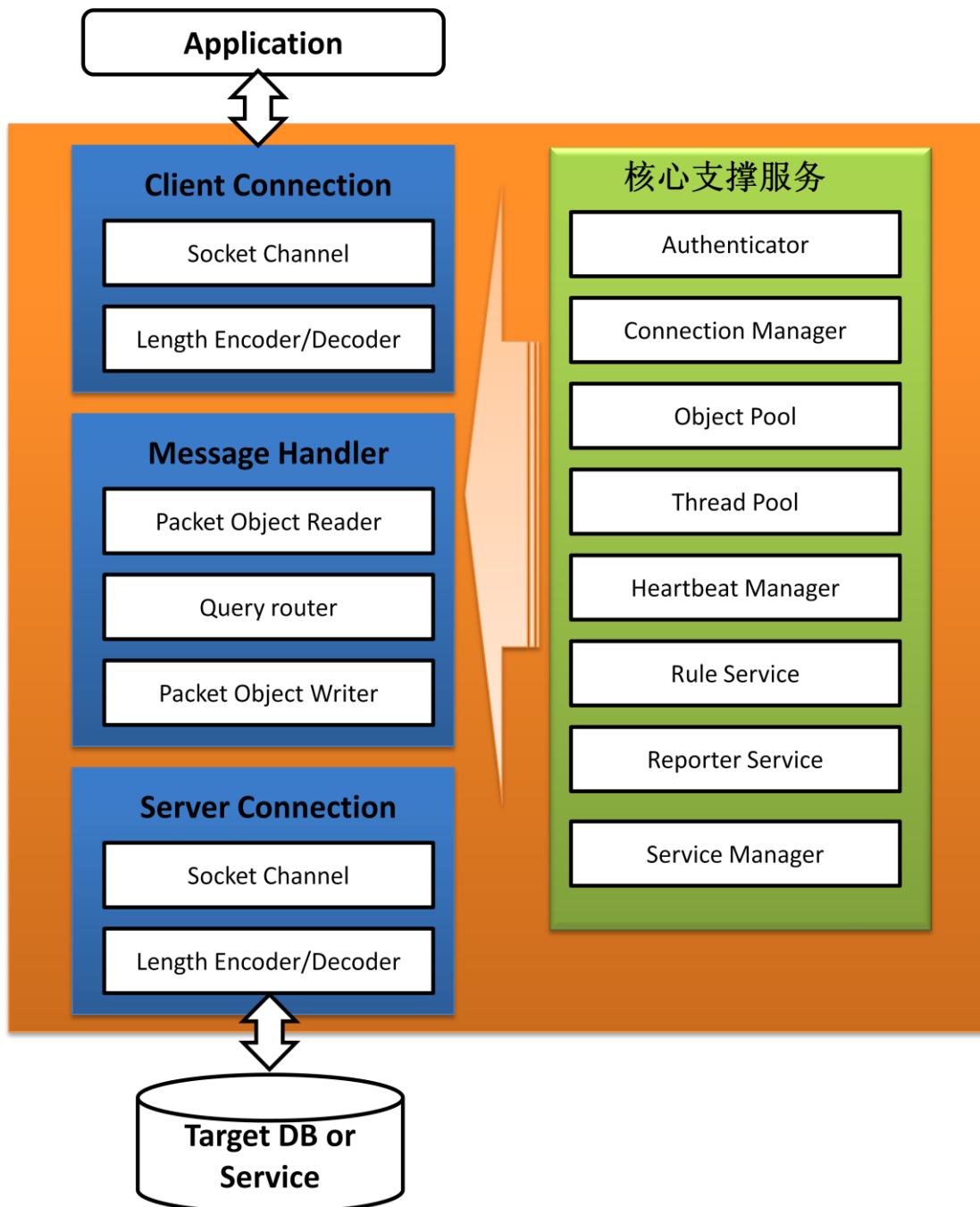


2. 内部模块以及技术描述

Amoeba 框架不仅仅代码上面作为其他产品的开发框架，也提供了启动方面的框架，仅仅修改 amoeba 的配置就能够让他变成另外一个产品。

Amoeba 的配置分成几个文件：amoeba.xml、dbServer.xml、rule.xml、log4j.xml，其中不同产品有所变化的文件是：amoeba.xml、dbServer.xml。

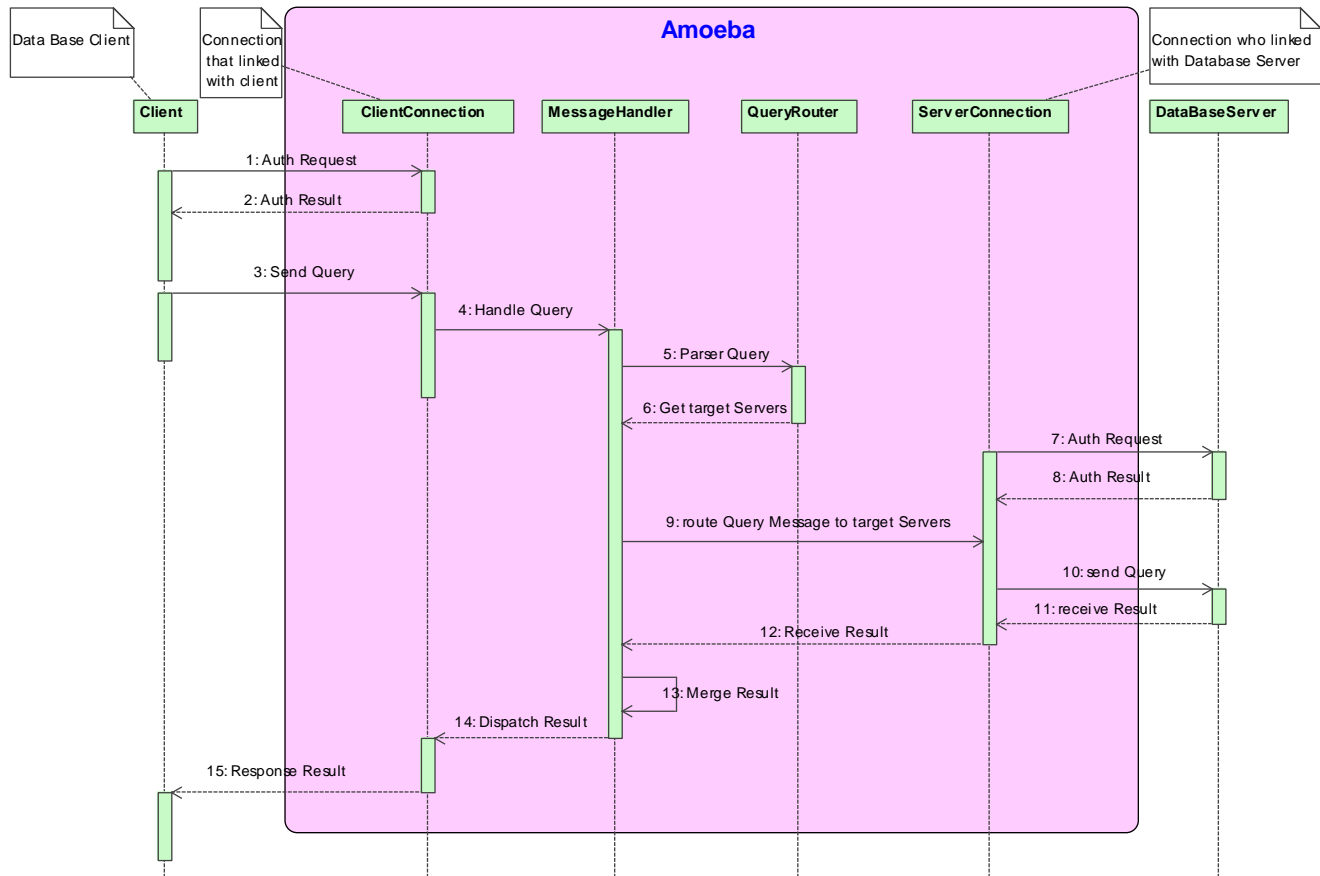
模块相关示意图



相关模块的描述说明:

- 1) **Built on Java NIO:** 采用 java NIO 框架无阻塞模式, 不像传统的 Socket 编程在大量并发的情况非常浪费系统资源, 可以支撑数万个并发 Socket 连接, 这块由 **Connection Manager** 负责管理
- 2) **线程池支撑:** Amoeba 网络 IO 与内部逻辑执行线程分离, 减少线程资源竞争与线程资源的利用率, 充分利用多 cpu 多核的硬件资源
- 3) **Heartbeat Manager** 负责管理每个心跳实体的生命周期以及心跳实体的任务调度, 心跳实体—这儿可泛指需要进行心跳检测服务的一些资源, 比如数据库、后端的服务等等
- 4) **规则表达式的服务**—提高了强大的类 sql 的支持, 它可以实现非常复杂的表达式, 也可以让用户参与规则函数的自定义, 从而扩展规则的功能
- 5) **Report Service** —负责产生一些系统运行情况的日志
- 6) **Object Pool** —它负责管理后端数据库与服务的对象池, 该对象池可设置对象初始化数量、最大连接数、最大空闲连接数以及最小连接数、连接池收缩周期、以及过载保护措施。

Amoeba Sequence



五、 如何使用 Amoeba

大家可以阅读在线文档，共分为 6 个章节。文档地址：[Amoeba 2.0 Use Guide 在线文档](#)

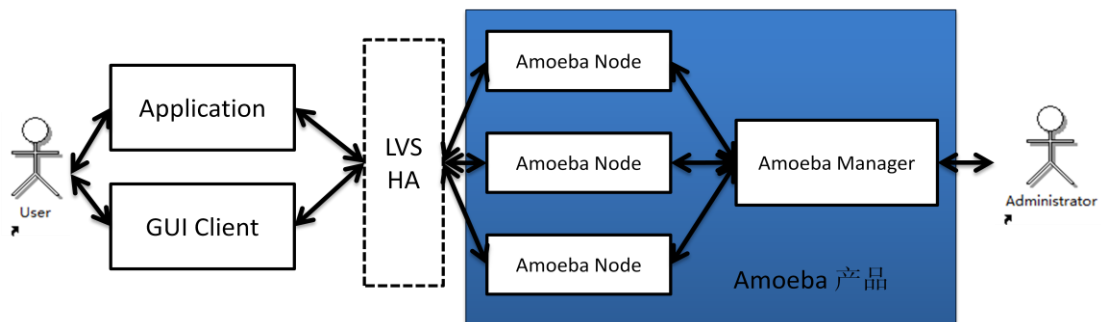
- 第一章：较为粗略地介绍了 Amoeba，包括与其他产品的横向比较
- 第二章：较为详细地描述了 Amoeba 的下载和安装过程
- 第三章：通过动手成功启动一个 Amoeba 实例以及尝试使用简单的切分功能来方便用户的手上
- 第四章：amoeba 的高级使用指南，这章讲述如何垂直、水平切分，以及规则的使用，负载均衡策略配置
- 第五章：性能调优，amoeba 的内存以及线程配置
- 第六章：Amoeba 提供的 Benchmark 工具的使用

六、 Amoeba 未来的发展

1. 产品发展

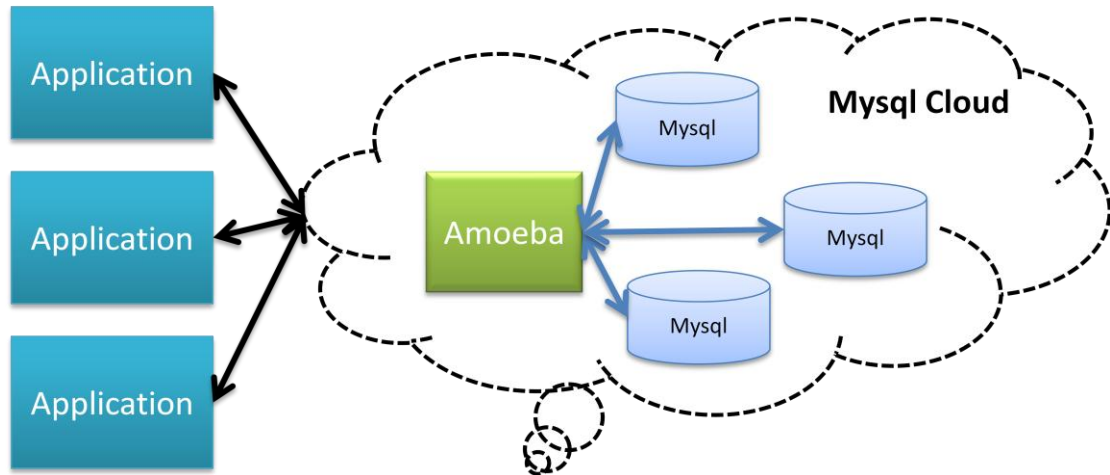
Amoeba 未来会发展成拥有分布式管理与监控体系的产品，子产品将会有多种，目前已经有 3 个基于 Amoeba 框架的产品（Amoeba for Mysql、Amoeba for Aladdin、Amoeba for Mongodb），这些产品将会通过 Amoeba Manager 来负责管理这些 Amoeba Node，努力朝着“Keep simple and stupid”目标发展。Manager 提供了 WEB UI 界面供操作人员进行简单的操作，它的职责有：

- 1) 配置管理——相关服务以及规则等配置管理。
- 2) Amoeba Node 物理节点管理
- 3) Amoeba Node 内部服务配置与控制管理（配置服务，启动、停止服务等控制）
- 4) Amoeba Node 健康监控与报警，包括内部的服务健康状况、网络流量以及当前的节点的内存使用情况等等



2. 应用 Amoeba 的领域

Amoeba 除了在目前项目中作为内部服务代理层中间件以外，其他公司诸如一些提供云计算服务的公司用来作为云服务的一部分对外提供服务。就拿 Amoeba for Mysql 来说吧，它对外提供了 Mysql 协议，还提供了一个简单的认证，可以想象要是能够丰富 Amoeba 认证与授权体系、流量控制、监控等，那么让 Amoeba for Mysql 实现 Simple Mysql Cloud 不是难题。



七、 联系我们

Amoeba 在线使用文档: <http://docs.hexnova.org/amoeba> 此网页会定期更新, 我们对本书的修正会定期发布。

如果对本书有技术相关的问题或对我们有更好的建议, 请邮件: amoeba.meidusa@gmail.com。关于 Amoeba 的更多信息, 请访问: <http://amoeba.meidusa.com>