

CONBY C3 运营平台 技术白皮书

Version 1.1

高性能超大规模业务流程自动化引擎
高性能大规模搜索引擎式数据挖掘引擎
高性能大规模规则式商业智能引擎
高性能大规模规则式社交网络智能引擎
.....

技术代码: C3V3

品牌名称: C3

商标设计:



版权声明

本文档是上海创百信息技术有限公司(*Conby Information Technology Co., Ltd.*)针对 CONBY C3 运营平台 (CONBY C3 Operating Platform) 发布的产品白皮书, 未经上海创百信息技术有限公司同意, 任何组织或个人不得转载、修改、引用, 或做任何未经许可的用途。

上海创百信息技术有限公司保留对所有侵权行为提出法律诉讼的权利。

本文中的“CONBY”, “CONBY C3 运营平台”, “CONBY C3 Operating Platform” 为上海创百信息技术有限公司的专有名称, 其余专门术语和缩写分属相应国际组织和公司。

上海创百信息技术有限公司 平台运营事业部

电话: 020-38293726/606

传真: 020-38296095

邮件: market@conby.com

网址: www.conby.com

地址: 中国广州市珠江新城 CBD 星汇国际大厦东塔 1102 室

目录

一、系统概述.....	4
1、C3 运营市场概述.....	4
2、CONBY C3 运营平台.....	4
3、常用名词解释.....	5
二、系统结构和功能.....	6
1. 系统架构.....	6
2. 公用主框架.....	6
3. C3 SDK.....	6
4. C3 通信协议 – Channel协议.....	7
5. 业务流程自动化.....	7
6. 事件系统.....	9
7. 系统Token.....	10
8. Task任务运行时管理.....	11
9. 平台负荷管理.....	12
10. C3 公用网关管理.....	12
11. 发布Task.....	12
12. 发布Map-reduce Task.....	13
13. SDK下载.....	13
14. Web爬虫应用.....	13
15. 基本API清单.....	17
16. SDK常用模块.....	18
17. 系统模块功能.....	19
18. 自定义编程接口 – 自定义业务流程界面.....	21
19. 自定义编程接口 – 自定义发布界面.....	21
20. 自定义编程接口 – C3 API.....	21
21. 自定义编程接口 – BSP API.....	21
三、系统运行环境.....	22
四、目标客户定位.....	22
五、关联模块 – Web C3.....	23
六、部分获得资本投资的典型成功客户名单.....	25

一、系统概述

1、C3 运营市场概述

随着 Internet 宽带网络进一步改善和无线宽带的普及，网络业务流程自动化服务(BPA)和云计算应用已经成为 Internet 和无线宽带网络上的新一轮投资热点，有投资商开始尝试进入这一领域，这是一个充满变革机会的全新领域。网络业务流程自动化(BPA)和基于云计算的在线应用从出现到现在发展的时间很短，但是发展速度却非常之快，其中有个推动力就是 Open API 出现和广泛普及，国内的例如 Weibo API、腾讯 Open API、Baidu Open API、阿里 Open API 等等，这些方式出现开始抛弃大包揽式的 Web1.0 网络系统模式，而 Web2.0 也开始进一步进化向界外系统扩展，各种网络系统和平台互相交叉已经形成现实；国际上 Twitter，外部通过 API 带来的流量早已超过站内流量。现在，可以说云计算、在线业务流程自动化计算服务已经成为电商、网络社交、娱乐和在线协同工作等等的有效方式。在这个不再严格区分你我网站地盘的 Web3.0 时代即将来临之际，CONBY 经过 4 年的不间断开发和运营实践，为适应下一步市场需要，推出 C3 这个 Web3.0 计算平台。

C3 运营市场的巨大商业机会来自于以下的几个新生市场：

- 基于 Web3.0 的网络概念开始落地和普及，如众多的 Open API 发布，各种网站交叉依赖运行等
- 云计算等在线计算服务普及时代的即将来临
- 基于海量计算的新数据挖掘技术市场，实现针对用户个体一对一的规则式业务流程
- 社交网络的繁荣为规则式社交网络搜索引擎提供新的市场机会
- 传统市场对高性能高并发业务流程自动化系统的需求，以 Web3.0 交叉和兼容的实施方法，可以直接加载在现有任何企业级业务上，而不会对现有业务造成任何负面影响，只会增加整体的自动化程度和商业智能化程度
- 对未实现平台化商业策略的公司，本产品可以提供平台化商业战略的整体支持，从而提高商业竞争的专有和差异化竞争

2、CONBY C3 运营平台

CONBY C3 运营平台是采用完全分布式的即插即用积木式体系架构。包括负载均衡系统、Key-Value 系统、Map-reduce 系统、DEVS 离散事件系统业务流程自动化引擎、API、通道、C3 节点以及支持多层次运营的传统 WEB 管理和维护接口等。具有以下描述：

- 有一个或多个业务 C3 节点启动特定负荷服务，可随意加载和卸下。
- 界外 C3 节点下载并配置 SDK 包，申请账号 ID，然后启动应用程序通过通信通道连接到 C3 云平台。
- 界外 C3 节点系统负责处理平台外的显示和操作界面，具有简单的逻辑处理和计算转移功能。同时负责接收发送与其他节点交互的数据包。
- 平台端系统负责处理云端数据逻辑、数据存储、业务逻辑，业务之间的网络信息传

递，以及数据库之间的数据读取保存工作。同时平台端还要承担平台外应用数据的管理、转发、维护等工作。

- CONBY C3 运营平台提供在线服务模式和 SDK 式系统实施方案，拥有经验丰富的技术工程师，可针对客户的具体要求，提供具体的技术指导和项目开发服务。
- CONBY C3 运营平台拥有流量分流、合作运营的功能，能支持独立 SP 业务模式。十分适合网络运营提供商使用，部分内核节点可由 SP 完全控制运行，支持多层次运营商网络(即一级运营商，二级运营商，三级运营商等等)。
- CONBY 公司拥有完全的自主产权，从通信接入到算法和界面、节点引擎等等，没有使用任何有版权问题的第三方代码(部分基于没有任何版权问题的开源项目，例如 LINUX, HADOOP, MYSQL, PYTHON, PHP, JAVA 等等开源技术)，适合于一些大型企业和互联网运营商使用。特别适合进行大规模的商业运营，无任何法律风险。
- CONBY C3 运营平台拥有成熟的新媒体传播运营商业模式，结合微博、社交网络系统可产生智能化的传播效果。
- CONBY C3 运营平台拥有成熟的分布式全局会话级搜索引擎应用模块(搜索引擎本身为 C3 平台一个普通的应用模块，不属于核心)
- CONBY C3 运营平台拥有成熟的负荷控制和高性能并发计算能力，能轻松地进行大规模密集并发的逻辑运算服务和异步数据服务，大规模海量计算和异步数据挖掘服务等等

目前 C3 有两个版本同时运行，V2.0 的版本基于 Google 云平台开发，V3.0 版本基于 Conby 独立平台开发，有更好的性能、业务吞吐和并发量。

3、常用名词解释

节点 (Node) 通常为一个独立机器，但一个机器也可以运行多个节点

通道 (Channel) 业务逻辑级别的通信通道，一个机器可有多通道

模块 (Module) 业务逻辑级别的程序模块，C3 上均为自动部署和免维护

C3 (CONBY CLOUD COMPUTING) CONBY C3 运营平台产品

XMPP (Extensible Messaging and Presence Protocol) 可扩展消息处理现场协议

SDK (Software Development Kit) 软件开发包

TOKEN (TOKEN) 代表一段定义，类似于专有名词、变量的概念

BPA (Business Process Automation) 业务流程自动化

CMC (CONBY Management Console) CONBY通信平台管理台

DEVS (Discrete Event System Specifications) 离散事件系统规范

RE (Regular expression) 正则表达式技术

KV (Key-Value) 键值型数据库技术

MR (Map-Reduce) 一种并行计算技术

CJG (CONBY Job Gateway) CONBY任务网关

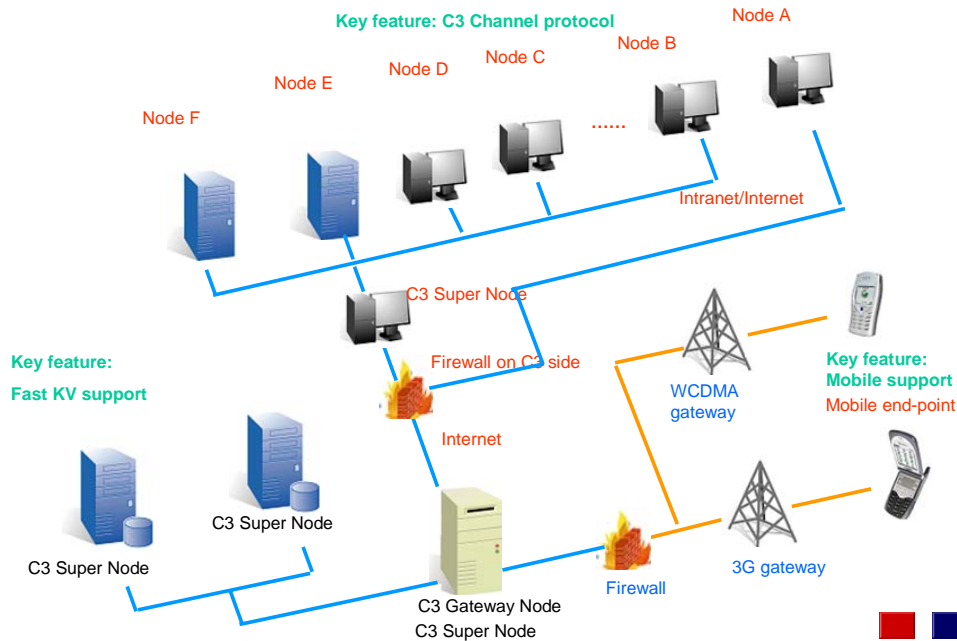
EJC (Enterprise Job Center) 企业任务中心

CJC (CONBY Job Center) CONBY任务中心

二、系统结构和功能

1. 系统架构

System Architecture



2. 公用主框架

CONBY C3 运营平台构造在 CONBY C3 Channel 协议框架之上，建立底层的网络通信和负载均衡机制，利用网络通信连接构建业务节点、界外节点之间的通信连接，然后在该连接的基础上利用自己设定好的方式进行应用端登陆、进入系统、选择情景模式、开始业务应用、通信、计算等操作，而这些操作是 CONBY C3 运营平台的公共部分。在以上协议的基础上，根据不同的业务编写不同的业务逻辑处理模块，在该逻辑处理模块中实现对应业务逻辑。同时在平台端还需要和数据库节点交互，用于读取或保存应用信息（如用户任务、数据、逻辑资料等数据）。

3. C3 SDK

CONBY C3 运营平台预置两个快速易用的 SDK，分别为 Python 和 PHP 的 SDK，可供业务节点在此基础上快速定制出自己所需要的 C3 应用，这些 SDK 均支持平台端模块的自动更新、部署和维护升级，具有免维护特征。一般而言 Python SDK 适合平台级开发，PHP SDK

适合针对企业商业逻辑应用的低成本应用级开发。

4. C3 通信协议 – Channel 协议

CONBY C3 Channel 协议采用成熟的工业级别的消息队列通道模式进行通信和计算转移，技术成熟，性能优异，业务操作方式灵活，扩展性强，可支持超大规模并发分布式计算。

CONBY C3 Channel 协议目前采用的版本为 AMQP(Advanced Message Queuing Protocol) 0.8 版本，在技术形式上为一个独立的 Python 库，当然也支持任何其他 AMQP 兼容库，可以为二次开发提供完全的灵活性和自由度，接入可以如下两种方式：

- SSL 模式

对需要保密的业务应用应该采用 SSL 模式并自行加密关键数据。

- 普通模式

普通模式，以海量计算的方式来处理业务应用。

CONBY C3 Channel 协议使用多重路由机制进行优化，能有效提高系统效能。CONBY C3 Channel 协议将 C3 节点分成三种类型：

- 平台内核节点

专门用于管理和维护各节点之间的必须的联系通信，提供一个经过缓冲的网络快速异步消息，支持事务级别的重试和纠错，具完整分布式系统日志。

- 非内核重要节点

既作为管理节点，同时也作为普通节点参与过程通信，重点对局部网络提供快速缓冲和路由信息，以提高局部网络整体性能和整体网络的稳定性。SP 可在超级节点上提供管理、维护等操作，提供 SP 自身的服务，形成一个相对独立的虚拟内部网络。

- 普通应用节点

根据节点自身用户操作和关联节点的操作，承载相应的业务通信负荷。

5. 业务流程自动化

CONBY C3 业务流程自动化(BPA)为内核核心模块，也是 WEB3.0 逻辑业务的核心，基于 DEVS 离散事件系统规范开发。通过本模块在海量计算和海量数据中进行规则匹配，从而达到整体的业务智能效果，可以通过基本的 Token 界面和规则配置界面进行 TASK 配置。界面简单。包括如下几个基本概念：

- Task 任务

- Runtime 为 Task 的运行时版本，所有的流程全部对应 Runtime，与静态 Task 形成对比

- Condition 条件，支持 OR 和 AND 逻辑
- Action 动作，条件满足后自动执行
- Token 代表一段变量或规则定义
- Payload 一般作为 Spec 的参数
- Spec 一般为 API 或 URL

Home

Create C3 task (Business Process Automation)

Task name *	<input type="text"/>	Maillog	<input type="text"/>
Description	<input type="text"/>	Conditions Type	<input type="text" value="OR"/>

Name *	Type	Spec *	Payload	Eval
<input type="text"/>	<input type="text" value="crontab"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="add Condition"/>				

Name *	Type	Spec *	Payload
<input type="text"/>	<input type="text" value="api"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="add Action"/>			

* indicates required field

基本参考：

<http://www.conby.com/page/c3-free-bpa-service-user-guide>

C3 V2.0 版本地址为：(二级运营商模式)

<http://www.conby.com/page/create-c3-task-business-process-automation>

C3 V3.0 版本地址为：(主运营商模式)

<http://api.conby.com/page/create-c3-task-business-process-automation>

其中 2.0 的版本为 Google 云计算平台支撑，3.0 的版本为 CONBY 自有平台支撑，3.0 版本的性能、并发和业务吞吐远超于基于 Google 云平台的 2.0 版本。建议使用 V3.0 版本。

Task 的内部格式为标准的 XML 文档格式，统一为 UTF-8 编码，故系统支持任何语言。XML 简单例子如：

```
<?xml version="1.0" encoding="UTF-8"?>
<c3task name="task name" type="user" maillog="" description="task description">
<c3conditions type="OR">
<c3condition name="c1" type="due" spec="[NOW]" payload="" eval=""></c3condition>
</c3conditions>
<c3actions>
<c3action name="a1" type="queue" spec="hello" payload="py=c3hello"></c3action>
</c3actions>
</c3task>
```

6. 事件系统

CONBY C3 运营平台基于事件驱动模式运作(DEVS 离散事件系统规范), 从而在设计上可以满足任何超大规模分布式计算的需要。Task 应用层次上提供了主动时钟事件和被动消息驱动 Hook 事件、具有 Timeout 功能的异步 RPC 以及用户条件满足后的 Action 事件。

- 时钟事件以 Linux 的 CRONTAB 格式定义, 常见的有闹钟模式, 周期性重复模式等等, CRONTAB 算法为 CONBY 自主知识产权算法, 基本实现 Linux 版本的完整移植。
- Action 事件为当 Task 中的 Condition 条件满足后自动触发。
- Hook 事件模式是系统的重要部分, 通过将 Action 结果转化为一个主动的 Event, 从而触发其他满足条件的 Condition 或 Action, 形成比较完整的事件模型。Hook 的安全控制分别通过输入码和输出码来进行 task 级控制, 分别为 in_key、out_key 以及权限的超时控制 timeout, 单位为秒(c3_request_reply 的电子邮件请求 timeout 单位为小时, 默认为 12 小时); 传入 session_key 即启动分布式会话事务。通过 addRuleName 和 delRuleName 可实现 Hook 的运行规则, 实现在消息传递时的规则自适应和智能规则处理。

核心层次上有 12 种常态事件定义, 分别为:

- Rule_task 用于处理运行时(runtime)的 task
- Task_job 用于处理运行时(runtime)的 task 的计算和逻辑处理
- C3_public 用于处理进程级接口的业务逻辑, 支持任何语言和环境
- C3_clock 时钟事件, 精确度为秒
- Callback_event 回调事件, 通常用于大规模推送订阅或业务逻辑 Hook
- C3_mail 电子邮件服务, 基于事件驱动
- System_event 系统事件, 控制其他所有事件, 同时处理平台外的请求

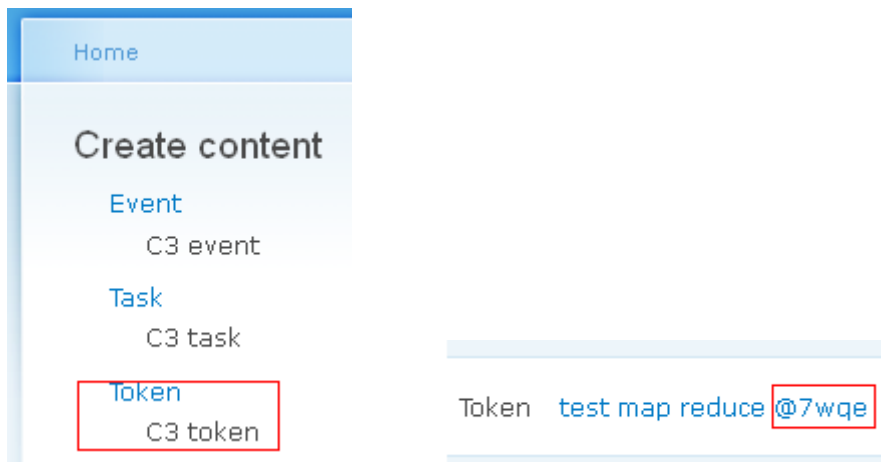
- Request_task 将平台外的请求转换成平台内的 Task
- C3_log 日志事件，处理分布式的日志处理
- Javascript_render JS 脚本渲染事件，用于模拟客户端的 JS 渲染，用于爬虫模块，本功能通过 C3 有 Timeout 功能的异步 RPC 可以实现人机实时交互(例如人工为爬虫识别图形验证码，或其他需要人工干预的事件)
- Hook_event 用于处于异步 Action 的结果
- Page_event 用于搜索引擎，当搜索到符合条件的 Web page 时触发

7. 系统 Token

CONBY C3 运营平台的 Token 分为三种，分别如下：

- 静态 Token

静态 Token 指这个 Token 一旦首次定义后，其内容不会发生变化。静态 Token 通常通过登陆 Web 管理界面后，点击“Create Token”菜单进行定义，保存后在标题的右半部分(@符号后的部分)即为可在 BPA 配置界面中可引用的 Token，使用中括号进行引用。如图：



在 BPA 配置界面中，可按如下格式引用：

[@7wqe] 针对 task key 或 runtime 同时亦生成一个基于用户名的别名串，例如用户为 demo，标题为 test，则别名为 test@demo，可用在 hook 的 spec 或 update_hook 的 task_key 中表示 task key 或 runtime

- 动态 Token

动态 Token 指由系统自动赋予相应的值，例如通常在 Due Condition 中使用的 Token:

- 1 、 [NOW] 表示现在
- 2 、 [NOW+60] 表示一个小时后

3、[result] 表示 Hook 事件中上一 Action 的数据结果，用于 EVAL 字段。同时也可在 EVAL 中使用 postCallBack() 获得灵活的 POST 交互数据

- 运行时 Token

运行时 Token 通常为一个 Python 或 Json 对象，可按相应的计算机语言进行使用，例如系统对每个 Condition 默认生成两个 Token，假设 Condition 的名字为 c1，那么对应的 Token 为：

1、[c1]

2、[c1_json]

第 2 个 Token 仅在内容为 Json 格式时才会有效。另外对每个搜索引擎的 Page、每个邮件和每个图象都会生成一个 Page Token，例如，如果在规则中定义了 title 的 P 别名，那么在 c1 条件中可以这样使用：

■ [c1.first.0.name]

Json 格式可这样使用：

■ [c1_json.first.0.name]

在 Python 语言接口中（一般为使用 Python 编写应用 API 时），可这样使用：

■ c1_json['first'][0]['name']

8. Task 任务运行时管理

CONBY C3 运营平台的每个静态 Task 可对应 N 个运行时版本(runtime)，这样不同用户对一个 Task 的运行可因上下文环境不同而生成各自需要的结果，下面方式启动一个 Task 的 Runtime：

- Task 发布后，自动启动一个 runtime 进行运行

这种情况下会生成一个全新的 Task Runtime

- 使用 API C3_request_task 刷新 Task 或 runtime (不修改内容, POST 方式)

如果 key 为 task, 这种情况下会生成一个全新的 Task Runtime。可直接在 BPA 配置界面中使用 queue API c3_request_task 直接触发另外一个 task 或 runtime, 只需要传入一个 key 参数即可

- 平台内核通过消息传入 runtime 参数启动一个 task runtime

这种情况一般是事件触发或条件满足时发生

- 外部调用可通过 PHP 开发包中 module.php 接口实现 GET 方式的调用，格式如：

Module.php?m=request_task&k=xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

其中 k 为 task 的 key，32 位长度；m 为模块名

可通过 API C3_update_task 终止一个 Task, 当传入的 XML 长度很小时, Task 将被终止(注意, 不是 runtime 终止, 已经启动的 Runtime 会继续运行), 每个 Task 在有异常或错误情况下最多自动 retry 10 次, 随后自行终止。

9. 平台负荷管理

CONBY C3 运营平台构造在先进的分布式架构上, 可任意启动和卸下各模块, 从而达到负荷管理优化目的。当某部分核心事件负荷较大时, 则增加该事件的处理模块, 从而该事件的处理能力会相应增加。启动模块很简单, 直接启动该模块即可。当监控部分事件负荷很小时, 即可以撤掉不稳定或效能不高的模块, 以优化系统的稳定性和减少 retry 数。卸下模块时必须保持最少有一个相同的模块还在运行, 使用 systemtool 工具发布命令:

```
systemtool -r 事件 -e cancel_事件
```

整个过程是完全平滑, 平台不会有任何影响。优化平台过程中, 主要分为以下级别:

- 门户级接口(VPS512)
使用 HTTP 轮循技术, VPS512 可轻松支持 5 万并发
- 服务接入接口(VPS512)
使用 HTTP 轮循技术, VPS512 可轻松支持 5 千到 1 万并发, 扩展内存和相应资源后无上限
- 应用 API 接口
全套单机下 VPS512 可支持 50 到 200 并发, 扩展内存和相应资源后无上限
- 远程 WEB API 接口或 Timeout 式异步 RPC
视远程硬件配置和并发能力而定

10. C3 公用网关管理

CONBY C3 运营平台对外的网关采用 HTTP 协议的 web 服务器, 为 CONBY 自行开发的专用高性能线程池式 Web 服务器, 只支持写入操作和 POST 请求, 同时在网关内维持了长连接到平台消息核心, 保证了系统的高性能, 不会每次都建立一个连接到平台核心而影响性能。

CONBY C3 运营平台公用网关采用 HTTP 轮循的大规模并发处理技术, 从而可以有效满足大规模的并发请求操作, 另外一方面可以在轮循节点阵列中进行平滑的维护和优化管理操作, 不会中断平台的整体服务。

11. 发布 Task

CONBY C3 运营平台可通过如下方式发布或运行 Task:

- SDK API 直接通过 SDK 调用 API 进行发布
- XMPP client 通过 xmpp 的客户端，如 gtalk，以即时聊天的方式进行发布，在客户端界面中输入 /new event 即可看到响应的帮助提示信息
- BPA C3 Task 网页 具体网址可参考前面 BPA 部分
- BPA 用户页 登陆进入后可以直接以 XML 方式直接发布
- email(subject, body) 通过电子邮件方式，标题格式：title – [key.spec][in_key]
- HTTP POST 通过 HTTP POST 方式，具体可参考 API 的调用细节

12. 发布 Map-reduce Task

CONBY C3 运营平台可支持大规模的 Map-reduce 运算，PHP 和其他 Web Services 语言版本 API 在发布前需要自行开发相应的 Map-reduce 应用级 API，同时向 CONBY 提交，经审核发布后再进行操作；Python 版本针对数组列表类型的数据已经有易用的模块：C3dispatch，开发人员不需要了解 Map-reduce 计算模型的框架可细节，简单 2 个参数即可开发高并发的 Map-reduce 应用 Task。详细细节参考 C3dispatch 模块和相关入门教程。

Map API 目的是将计算的工作量按集合分散成 N 个小 task，Reduce API 的目的是将前面 N 个小 Task 运行的结果汇总起来，所以必须在相应的应用 API 自行处理逻辑。C3 平台负责自动分发、运算和协调及事务级控制。

13. SDK 下载

CONBY C3 运营平台 目前提供两种语言的开发包，分别为 PHP 和 Python，下载地址为：

中文：http://www.conby.com/products_ch.html

英文：<http://www.conby.com/products.html>

14. Web 爬虫应用

CONBY C3 运营平台内置分布式全局会话级搜索引擎支持，主要是通过简单地配置正则表达式、基于自定义图像特征组技术的图像识别以及 Javascript 渲染模拟技术来驱动搜索引擎模块工作。可调用如下 API：

- HTTP FORM

API c3_http_form 模拟单个 HTTP FORM POST 操作

API c3_http_form_session 用于生成 login 后的 cookie 数据，用于后续 session 方式的访问

参数：在 BPA 界面的 payload 输入 form=j3k2，其中 form 后面的字串为 token,可登录 <http://api.conby.com/> 然后通过 Create Token 生成

详细格式如下：

[action]

@url=http://www.abc.com/hello.html 表示启始的 url

@method=POST 表示 HTTP GET 或 POST

@encoding=y 表示将 data 中数据转为 utf-8

@json=y 表示返回的数据格式，存在则返回 json 格式，否则按行按本 INI 格式返回数据(无 section 部分)

[data] 表示 HTTP POST 的数据

@aa=message

@bb=message

...

[header] 表示 HTTP POST 的头信息

@aa=message

@bb=message

...

[cookie] 表示 HTTP POST 的 cookie 信息

@uid=message

@pwd=message

...

- C3_page

API c3_page 对一个或几个 URL 进行爬虫操作

参数：在 BPA 界面的 payload 输入 page=j3k2&w=200&d=5&t=2 或 page=[@j3k2]，当使用后面的格式的时候，token 中可包含嵌套的变量，例如表达 Condition 结果可以用 {@c1}，实际相当于 [c1]。

page 后面的字串为 token,可登录 <http://api.conby.com/> 然后通过 Create Token 生成

w 为爬虫的宽度，默认 200

d 为爬虫的深度度，默认 5

t 为爬虫的线程并发数，默认无，即单线程

page token 格式如下:

[default-match] 默认的，一般 match 匹配 URL

@m1=/auto_scrollnews/ 正则表达式

@m2=/auto_scrollnews/ 正则表达式

[default-pattern] 默认的，一般 pattern 匹配内容

@aa=^(P<title>+)</h3>/ 正则表达式，含别名变量定义是需要输出的数据

@aa2=^(P<title>+)</h3>/ 正则表达式，含别名变量定义是需要输出的数据

[event-match]

@m3=/C3EVENT/

[event-pattern]

@aa2=^((P<key>+))<a href="(P<link>+)">(P<linktitle>+)/

@aa3=^((P<key>+))<a href="(P<link>+)">(P<linktitle>+)/

[DATE-format]

@date=/(d{4})-(d{2})-(d{2}) (d{2}):(d{2})/

@format=1-2-3 4:5

[login-form] -form 结尾，用于定义自动 Form 提交，可实现自动发帖人机交互类功能

@url=(.+?)+.html\$ 当前 webpage 的 url，正则表达式

@action=(.+?)+.php\$ form 里面 action 的地址，正则表达式

@javascript_render=submit 用于提示模拟点击指定的按钮，进行 JavaScript 渲染

@uid=hello 数据项，文本，可以使用 Condition 中产品的变量，如[c1]

@pwd=123456 数据项，文本，可以使用 Condition 中产品的变量，如[c1]

[page] (名称为 mypage1-5-3-0 格式可直接指定爬虫宽度、深度和线程并发数)

@index=http://auto.163.com/special/00083TL0/auto_scrollnews.html

@form=j3k2 (比上面的@index 优先，覆盖@index，格式按 c3_http_form 定义)

@mypage1=http://www.abc.com/hello.html 自定义 URL

@mypage2=http://www.abc.com/hello.html 自定义 URL

[proxy] HTTP 代理服务器列表，如空则不使用代理服务器访问

@server1=http://121.121.121.121:8080/

@server2=http://121.121.121.121:8080/

[useragent] 模拟浏览器的 UA 列表，如空则随机选择一个 PC 版本的 WEB 浏览器

@ua1=http://121.121.121.121:8080/

@ua2=http://121.121.121.121:8080/

说明：

1 输出的 json 格式可参考：http://api.conby.com/c3/c3_page_json.js

2 系统自动对每个 page 进行生成 C3HREF C3EVENT 事件

3 pattern 如果为 C3_PATTERN，则自动提交到 C3 平台的 pattern 库自动查询

4 -match 部分和-pattern 部分为成对的定义，缺一不可

5 正则表达式的定义需要尽量规范，这些表达式将在 Python,PHP 和 JAVA 开发的系统中使用，不规范定义可能导致兼容问题，一般可同时定义 Python 和 PHP 两种格式的正则式。

6 输出 JSON 数据在 HOOK 回调中，可以用 maillog 进行监控，或专门定义 HOOK TASK 实时监控。

7 form 定义中，如果 form 名包含 PROFILEID，则表示爬虫将启用会话保存功能，一次登录后，后续没有关联的独立 Task 运行在会话没有超时前不再需要登录之类的操作。

C3 平台还有一个自营的社交媒体搜索引擎和社交媒体智能业务，但不包括在出售的商业包中，为 CONBY 自营业务，与传统 Web 搜索引擎相比较，即把人与人之间的关系作为“链接”，每个人作为一个“Web 页面”，技术实现基本雷同。在此不作更多介绍。

15. 基本 API 清单

CONBY C3 运营平台定位于基础平台，通过实践逐步积累一些商业 API。PHP 开发包的基本 API 可参考：

<http://www.conby.com/page/basic-c3-platform-api-function-list>

以及

<http://www.conby.com/page/c3-service-api-online-document>

Python 的 API 主要通过用户自定义的应用模块，内置的有：

- C3tool 模块 `c3_import(模块名)` 可自动引用平台上任何有权限的模块
- C3spider 模块 `crawl(page,搜索宽度, 搜索深度, 搜索并发线程数)` 使用方法：

在配置界面中 Spec 填 Crawl, Payload 中填 `py=c3spider&page=lfmy&w=1&d=1&t=1`

Py 表示使用 c3spider 模块，即搜索引擎模块

Page 表示使用 lfmy 这个 token 定义的搜索引擎规则

W 表示搜索引擎的搜索宽度

D 表示搜索引擎的搜索深度

T 表示搜索引擎的搜索并发线程数

本模块支持带 Timeout 的异步 RPC，可实现人机实时交互。例如提示 30 秒内工作人员帮忙识别某些图形验证码之类的功能。

- C3hello 模块 `hello(msg)` 用于常规测试

在配置界面中 Spec 填 hello, Payload 中填 `py=c3hello`

- C3echo 模块，用于回显数据，亦可作为通道使用

在配置界面中 Spec 填 echo, Payload 中填 `py=c3echo&msg=你的数据`

- C3cvpr 模块，即计算机视觉模式识别，用于进行图像智能识别，基于自定义图像特征组技术，准确性高，可用于进行大规模的图像特征组搜索技术

在配置界面中 Spec 填 detect, Payload 中填 `py=c3cvpr&i=BASE64 编码的图像数据`

- C3email 模块，电子邮件处理解析模块

在配置界面中 Spec 填 parse, Payload 中填 py=c3email&i=原始的邮件文本, 返回或一般取 json 对象中的 body_json 的值, 语法规则与 c3_page 中相同

- C3Dispatch 模块, 用于 Map-Reduce 计算的 Map 分发, 使用 List 型数据(即数组类型的数据)

在配置界面中

Action 类型选择 map-reduce,

Spec 填 dispatch-a2 (a2 表示另外一个 Action 的名字),

payload 中填 py=c3dispatch&map=数组类型的数据 (通常使用 json 变量表示), 如 [c1_json.bb]

上述 Python 模块全部支持二级运营商(SP)独立运营功能, 加 sp 参数即可将全部任务流转完全独立的运营节点。

16. SDK 常用模块

同上节所述, Python SDK 主要使用几个基本的模块, c3tool、C3Dispatch 和 c3spider 等, PHP 开发包情况如下:

- 应用 API 开发

```
tool
service_tool
application_services
application_services_tool
```

- 业务端应用

```
tool
user_tool (用于回调 callback, 钩 hook 等)
```

- Map-reduce 应用

PHP SDK 请参考 c3-rss-map-reduce-api.zip 示例包

<http://conby.com/download/c3-rss-map-reduce-api.zip>

Python SDK 请参考 C3Dispath 模块

异步实时人机交互技术请使用 rpc_call_with_timeout API

其他语言可使用 Web Services 方式进行 Map-Reduce 运算, 没有任何限制

17. 系统模块功能

系统具体功能如下所述：

System Layer Component



第一层是传输层：传输层主要实现通信和网络应用的传输协议，支持各个通信服务提供商的协议，如 AMQP 协议、HTTP 协议和 SSL 协议等，然后再经过通信服务提供商的网关，和平台通信中心相连。

第二层是平台层：平台层是系统的核心部分，系统根据通信服务项目的特性、预约、定制、运行等服务要求和业务逻辑，智能控制通信服务在系统中的交互过程。流程管理主要是用户每一次服务时的对话管理，当用户发起通信请求到系统时，系统为这次服务建立一个对话，控制系统和用户对话的交互过程和业务逻辑。当系统处理比较忙时，系统会自动启动性能调度管理功能，通过在系统线程池中增加一定的线程队列来分担繁忙的业务处理，维持系统的性能。

第三层是业务层：业务层主要实现用户业务的承载功能。平台可以承载各种不同的业务，系统主要考虑如下：Task 的更新，Task 定制，规则定制等的交互式服务接口定义，系统的实时监控，统计等。

该层主要由用户按照自己的使用要求进行扩展，系统提供外部接口，允许业务层调用，实现通信控制处理和系统维护。

■ 数据库模块

CONBY C3 运营平台的数据库模块为用户对不同数据库操作提供统一接口而不需要关心具体使用的是什么数据库，而且可以支持多种数据库系统，如：MySQL, ORACLE, Informix, SYBASE, SQL Server, DB2, DBF, FoxPro, Access 等。但 CONBY 推荐使用 Key-value 型快速大型数据库，目前使用 MySQL、HBase 和 Google 的云数据库模拟成系统的 Key-value 数据库。

■ 业务流程解析模块

业务流程解析模块负责各个具体业务流程的解析，通过对其他模块的调用，完成业务流程。该模块和业务任务管理以及性能调度管理结合构成整个平台的核心。

■ 业务模块扩展/调用模块

该模块允许用户自行扩展定制接口功能，接口规范简单清晰，易于开发。在运行时，该模块负责调度外挂的业务模块接口，实现特定要求的业务需求。

■ 统计模块

统计系统的作用在于设置各种业务模块的使用率，成本方式，限制政策，对各业务模块的使用情况、频率、负荷、瓶颈进行实时统计，供运营商决策分析。另外还可以依据不同的经营策略采用不同管理标准（如按次、按服务、按时间、包月、套餐等等），提供各种形式的报告单。本模块目前尚未开发完成。

■ 用户管理模块

用户管理模块是对注册用户资料的维护管理，管理员可根据自定义的用户分组策略，进行用户资料管理。

■ 日志管理模块

日志管理模块主要提供详细的系统日志管理功能。另外和平台的过滤模块结合，产生黑名单列表，可以对黑名单用户进行监控。

- ◆ 信息及服务栏目变更日志
- ◆ 系统运行情况日志
- ◆ 系统维护日志
- ◆ 系统管理日志等

■ 性能调配模块

性能调配模块能根据系统当前的性能压力，自动调配系统的执行性能，实现负载均衡。

■ 网络监控模块

网络监控模块实时监测网络连接状态，一旦网络连接中断，会立即警告，并将发送的关键数据缓存，不会丢失数据。一旦网络恢复连接，再发送数据。

18. 自定义编程接口 – 自定义业务流程界面

CONBY C3 运营平台针对 Internet 运营商的特点，为减轻运营商的技术维护投入和技术门槛，为企业和运营商提供了一套基于 WEB 技术的自定义编程开发接口，用户可根据 C3 SDK 轻松地开发符合自己风格和业务需求的界面和业务逻辑。

19. 自定义编程接口 – 自定义发布界面

CONBY C3 运营平台针对 Internet 运营商的特点，为减轻运营商的技术维护投入和技术门槛，为企业和运营商提供了一套基于 WEB 技术的自定义编程开发接口，用户可根据 C3 SDK 轻松地开发符合自己风格和业务需求的界面和业务逻辑。

目前 C3 推荐两种技术的编程接口，一种是基于 PHP SDK 技术的编程接口，另外一种是基于 JSON 的 Javascript 编程接口。同时，C3 可以根据客户需要快速地提供其他语言的自定义编程接口。

20. 自定义编程接口 – C3 API

CONBY C3 运营平台针对部分 Internet OEM 运营商的特点，为减轻运营商的技术维护投入和技术门槛，提高运营商的 OEM 能力和自行定制能力，为企业和运营商提供了一套基于 Python 和 PHP 的跨平台 API 接口技术的自定义编程开发接口，用户可根据 C3 SDK 轻松地开发符合自己风格和业务需求的界面和业务逻辑。

- 标准 Python 模块 API 接口标准
- 标准 HTTP 回调通知机制
- 其他语言开发的独立进程级模块(兼容 C3 对 stdin 和 stdout 的要求即可)
- 支持应用端多线程调用
- 支持多任务并发运行机制
- 支持自定义内核优化，并自行建立 BPA 引擎内核(需要付费)

21. 自定义编程接口 – BSP API

CONBY C3 运营平台针对部分 Internet OEM 运营商需要灵活控制计费系统和运营系统的特点，为减轻运营商的技术维护投入和技术门槛，提高运营商的 OEM 能力和自行定制能力，为企业和运营商提供了一套基于 PHP 和 Python API 接口技术的自定义编程开发接口，用户可根据 C3 SDK 轻松地开发符合自己风格和业务需求的界面和业务逻辑。

- 可配置的内核节点优先级
- 可配置的应用负荷值
- 可配置的多运营商共存机制，独立经济系统，理论上支持无限层次的多层次运营商标体系，配置非常简单，使用 C3 快速业务运营系统包只需要修改上一级运营商主机网址即可。

三、系统运行环境

CONBY C3 运营平台中心的运行环境如下：

- 硬件环境

主节点最低配置 512M 的 VPS 或服务器，支持最少 2 万门户并发，最少 3 千 API 并发，以及日运行最少 200 万个搜索引擎运行应用。对硬件的要求非常低，可实现低成本运营。普通节点可使用办公室的普通 PC 电脑，操作系统不限制(Linux, Win, Mac...)

- 软件环境

主节点 Linux，建议 CentOS, MySQL, AMQP 兼容消息阵列, Nginx, Google App Engine。普通节点可使用办公室的普通 PC 电脑，操作系统不限制(Linux, Win, Mac...)

- 数据库

建议选用 MySQL/SQL-SERVER/Oracle/HBase/PIG 等数据库。

- 开发环境

平台核心组件采用 GCC 编译器(Windows 平台为 VC)优化，逻辑开发为 Python 和 PHP，平台组件为 C 编译后的 SO 或 DLL 文件；PHP 模块为使用 C3 专用加密算法加密保护。整体系统有完整知识产权和商业运营保护。

建议：对于特大规模的 C3 运营平台，数据库最好装在通信良好另外的服务器上，装在一起对平台处理性能会有一些影响。

四、目标客户定位

- 互联网运营商

如综合门户网站、行业门户网站等等，可以通过本平台营销广告、行业信息、供求信息、业务流程自动化等等载体。

- 文化传媒企业

如电视台、文化传媒机构、纸媒、网媒等，可以通过本平台进行高效新媒体传播等业务。

- 大型商业企业

可以作为内部业务系统的平台，在继 Web 等技术架构后以一种全新的 API 服务架构运作严肃的企业业务应用，同时利用高性能的业务流程自动化模块可提高实时业务协作能力，是一种即将风靡全球的企业业务运营技术平台。

- 电商企业

- 咨询顾问企业

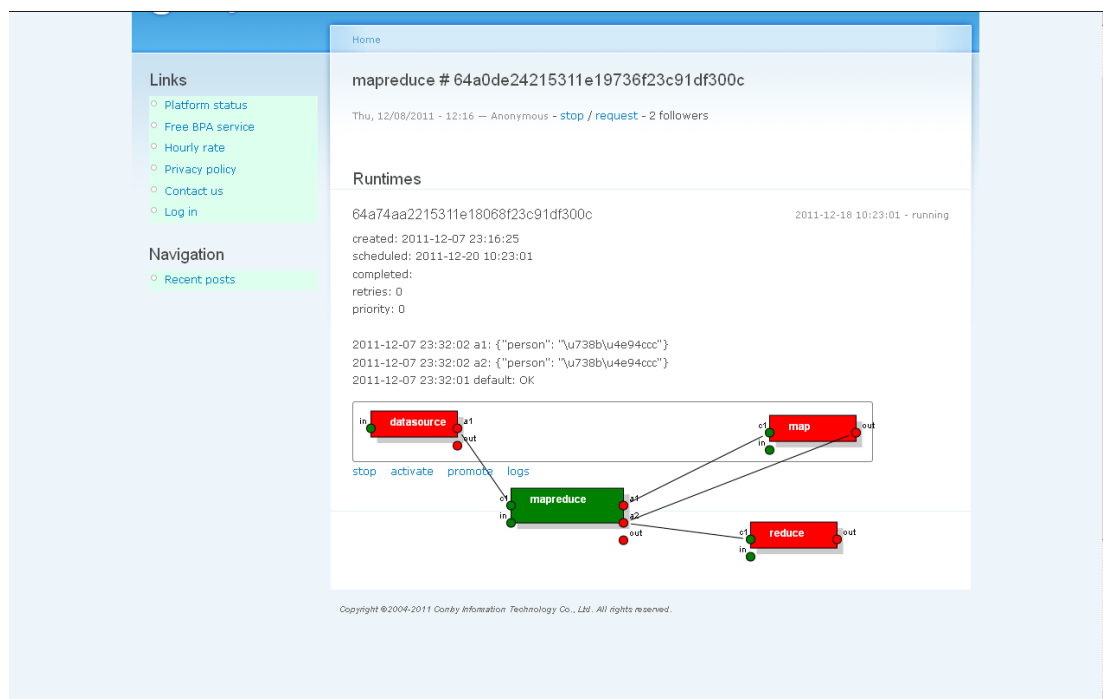
五、关联模块 – Web C3

Web C3 (<http://api.conby.com>) 是 C3 平台的多层次运营商 UI 辅助工具，除了可以支持无限层次的运营商网络结构外，同时又有如下特点：

- API 源码管理，及加密服务，加密下载后仍可运行，在开源基础上进行商业运作
- 方便 Task, Event, Token, runtime, 流程图形化的管理和跟踪维护，实时监控核心指标



The screenshot displays a web interface for monitoring a runtime. At the top, there is a navigation bar with a 'Home' link. The main content area shows a header with a long alphanumeric string: 'ccrrroonnn to yityihgh2 to gjfg to me3 # 995b549e0a1711e192a0f23c91df300c'. Below this, it indicates the runtime was started on 'Tue, 11/08/2011 - 22:40' by an 'Anonymous' user, with a 'stop / request' link and '0 follower'. A section titled 'Runtimes' lists a specific runtime with ID '995ca1aa0a1711e1beb1f23c91df300c', which is currently 'running' as of '2011-11-08 21:53:02'. It provides creation and scheduling timestamps, and shows that it has not completed or retried. Log entries show successful status checks: '2011-11-08 09:41:02 default: OK' and '2011-11-08 09:41:02 bfgjhfgj3: OK from c3hello.hello'. At the bottom, there are links for 'stop', 'activate', and 'logs'.



上图为简单的 Map-reduce 计算在 DEVS 离散事件流程引擎实现，鼠标双击即可跟踪关联流程

- ◆ Task 下面的 stop 停止 Task
- ◆ Task 下面的 request,启动该 Task 一个新的 runtime
- ◆ Task 下面的 follower, 表示有多少个 Task 订阅了本 Task 的运行结果
- ◆ Runtime 下面的 Stop,停止该 Runtime
- ◆ Runtime 下面的 Activate,激活该 Runtime
- ◆ Runtime 下面的 Promote,提高该 Runtime 的优先级
- ◆ <http://api.conby.com> 监控界面各参数含义:

C3 Core - V3.0	核心版本
(
[rule_task] => 0	待处理数
[task_job] => 0	待处理数
[c3_public] => 0	待处理数
[c3_clock] => 0	待处理数
[callback_event] => 0	待处理数
[c3_mail] => 0	待处理数
[system_event] => 0	待处理数
[request_task] => 0	待处理数
[c3_log] => 0	待处理数
[hook_event] => 0	待处理数
[page_event] => 0	待处理数
[task] => 0	待处理数
[javascript_render] => 0	待处理数 (JS 渲染技术)
[max_callbackid] => 365208	已处理数
[last_runtime_updated] => 2010-11-17 10:26:50	最新更新

```
[max_hookid] => 2999527 / 2010-11-17 10:26:50 已处理数
[max_pageid] => 6656 / 2010-11-17 10:23:26 已处理数
[max_logid] => 13917727 / 2010-11-17 10:26:50 已处理数
[max_taskid] => 1888480 / 2010-11-17 10:26:50 已处理数
[fastcgi_request] => 1 / 6 请求进程数/总进程数
[c3web_request] => 919 当前请求并发数
[load_average] => 0.50, 0.67, 0.80 系统 CPU 负荷
[time_zone] => Eastern Standard Time 系统时区
)
```

- 方便的回调事件定义，以及 WEB Rule 的定义。
- 有较多操作功能来扩展 Web UI 功能，与手机服务进行集成，方便手机访问
- 可以丰富的权限控制
- 支持 XMLRPC REST API 等接口
- 以极少的软硬件资源支撑一个的网络业务运营系统
- 通过 C3 技术建立在线协同商业的应用系统

六、部分获得资本投资的典型成功客户名单