

实时计算平台项目交流文件

目录

一、	公司简介.....	2
二、	代表性案例.....	3
1.	郑州银行案例.....	3
	项目背景.....	3
	需求分析.....	3
	解决方案.....	4
2.	中信案例.....	5
	项目背景.....	5
	需求分析.....	5
	解决方案.....	6
3.	百果园案例.....	8
	项目背景.....	8
	需求分析.....	8
	解决方案.....	9
三、	阿里云和流计算相关说明.....	10
	阿里云.....	10
	流计算.....	10
四、	实时计算平台项目技术方案.....	11
	架构图.....	11
	方案介绍.....	11
五、	核心风险关注点.....	13
1.	风险点列举.....	13
2.	风险点分析与解决方案.....	13
六、	建设周期/计划.....	15
七、	专业资质.....	16
	CMMI3.....	16
	企业质量管理体系认证证书.....	17
	发明专利（4项）.....	18
	软件著作权（部分）.....	22

一、 公司简介

深圳市比一比网络科技有限公司，成立于 2009 年，位于深圳南山科技园，是国家高新技术证书，拥有 ISO 资质证书，CMMI3 证及互联网 ICP 证书；拥有四项大数据国家发明专利，三十多项大数据技术相关的国家软件著作权证书。是深圳市互联网技术协会副会长单位，深圳技术创新单位。

公司直致力于大数据收集、分布式存储、数据分析与挖掘,数据可视化方面的研发。为客户提供一站式的大数据解决方案。主要服务对象为泛金融行业,包括银行,商业保理公司,担保公司,互联网金融,供应链金融公司。公司在深圳总部有约 50 人的研发团队,另外同时在多个银行有实施团队驻点。

公司基本情况	
法人营业执照号码: 9144030068375453XL	组织机构代码: 68375453XL
公司成立时间: 2009 年 01 月 12 日	公司注册资本: 1107.7417 万元
注册地址: 深圳市南山区粤海街道高新技术园中区科苑大道讯美科技广场 1 栋 3 楼 306	办公地址: 深圳市南山区粤海街道高新技术园中区科苑大道讯美科技广场 1 栋 3 楼 306
法人代表: 杜卫红	联系电话: 0755-88321616
经济性质: 民营	是否上市及上市地点: 否
服务机构地域分布: 深圳	主营业务: 软件及数据服务
相关资质: “ISO9001 质量管理体系认证证书” “软件企业认定证书” “软件产品登记证书” “CMMI3” “计算机软件著作权登记证书”等	是否有自主知识产权产品: 是
经营范围: 计算机软硬件、通讯设备、网络的技术开发和销售,技术咨询;企业管理咨询及其它信息咨询(不含人才中介服务、证券及其它限制项目);代订机票;会议策划;从事广告业务(法律、行政法规规定应进行广告经营审批登记的,另行办理审批登记后方可经营);国内贸易(法律、行政法规、国务院决定规定在登记前须经批准的项目除外)	
项目联系人: 张岩 联系人手机: 18675507635 联系人邮箱: zhangyan@beyebe.com	

二、 代表性案例

1. 郑州银行案例

项目背景

在项目实施前，郑州银行外部数据使用均为业务部门自行负责，随着使用部门的不断增加，单项目的外部数据接入平台已经无法满足行内需求，并且有着操作复杂，维护繁琐，缺乏统，重复采购等问题。本项目通过采集与整合海量的外部数据，搭建套完整的“数据中心”，建立统的接入、清洗、存储、管理、应用通道，实现全行外部数据的集中与共享，为授信决策、智能流程、精准营销等方面提供有力的支撑。

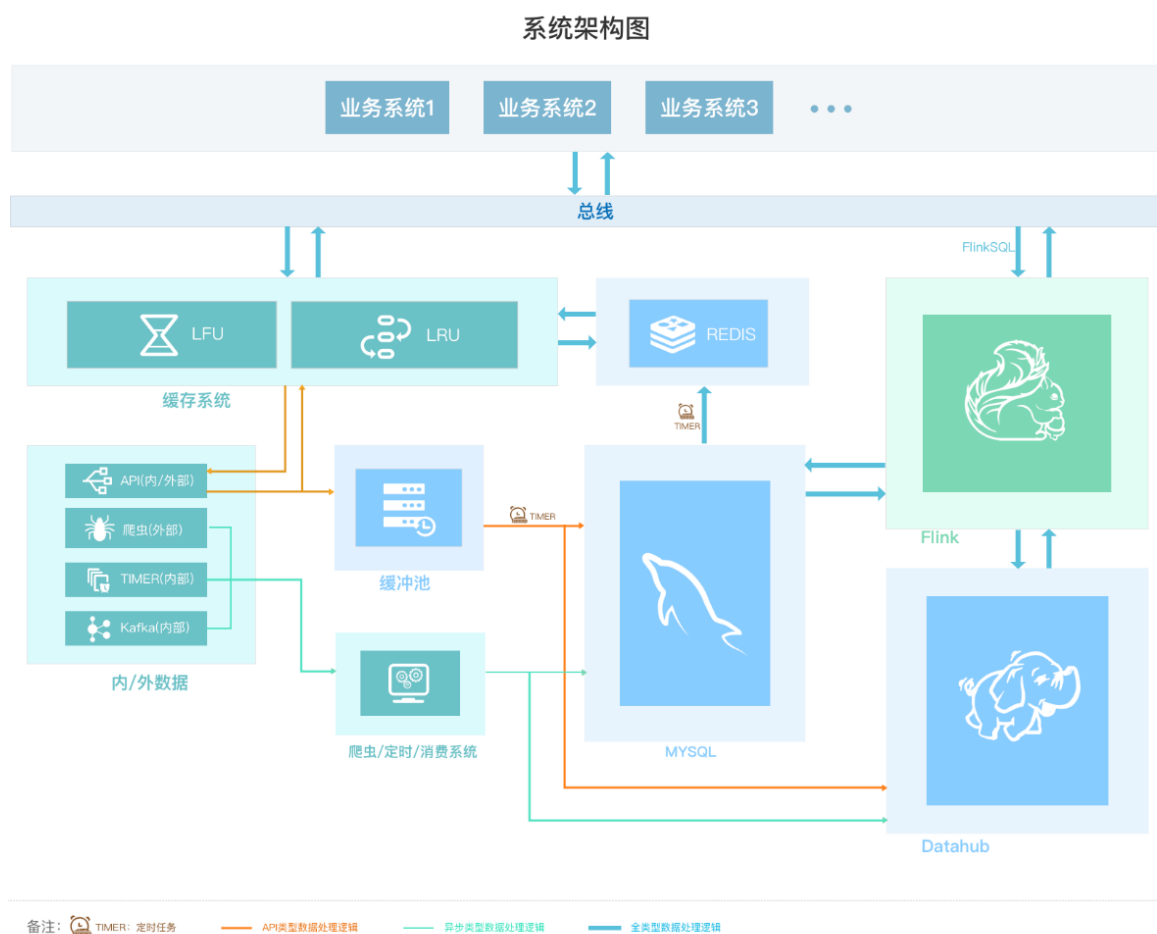
需求分析

数据渠道方面，有同盾、运营商、住房公积金、学信网等。实际上也同时对甲方所涉及到的内部脱敏数据进行了处理，否则达不到银行自动化 BI 的要求。

数据纬度方面，包括用户数据、行为数据、异常日志等，数据行为包括报文解析（比如处理上传的 PDF、EXCAL、日志等）、数据整合（添加外键、建立中间表等）、数据迁移（迁移至自建 mysql、Hadoop 等）。

数据管理方面，银行对数据管理有着严格规定，为了保证不干扰线上服务的正常运行，管理平台建立了健壮、高同步的旁路系统，保证外部数据和内部数据的快速处理。

解决方案



对于旁路系统我们进行了针对性设计。是增加了 Timer 类数据获取，可以自由定义 Timer 的采集时间，从而避免现有系统的二次开发。二是进行完整副本设计，避免任何情况下的二次采集。

对于数据接入场景，我们设计了四种不同的接入方案。使得无论外部数据还是内部数据，都可以做到快速接入系统。并尽量减少对现有系统的干扰和二次开发需求。对于客观场景中的被动接入和主动接入，都可以快速支持。

需求性能方面，要求全部查询都在秒级完成是不现实的。我们针对客户对于效率的需求，分成即时查询和慢查询两种查询方式。对于即时查询的场景，依据架构保证了即时性。对分析类、慢查询的场景，接入分布式计算集群进行异步处理。

2. 中信案例

项目背景

中信银行信用卡中心在收集客户信息的过程中，耗时费力，信息不全；客户提交的纸面信息可能会经过篡改，银行花费大量的时间、人力来审核客户提交的信息的真实性。

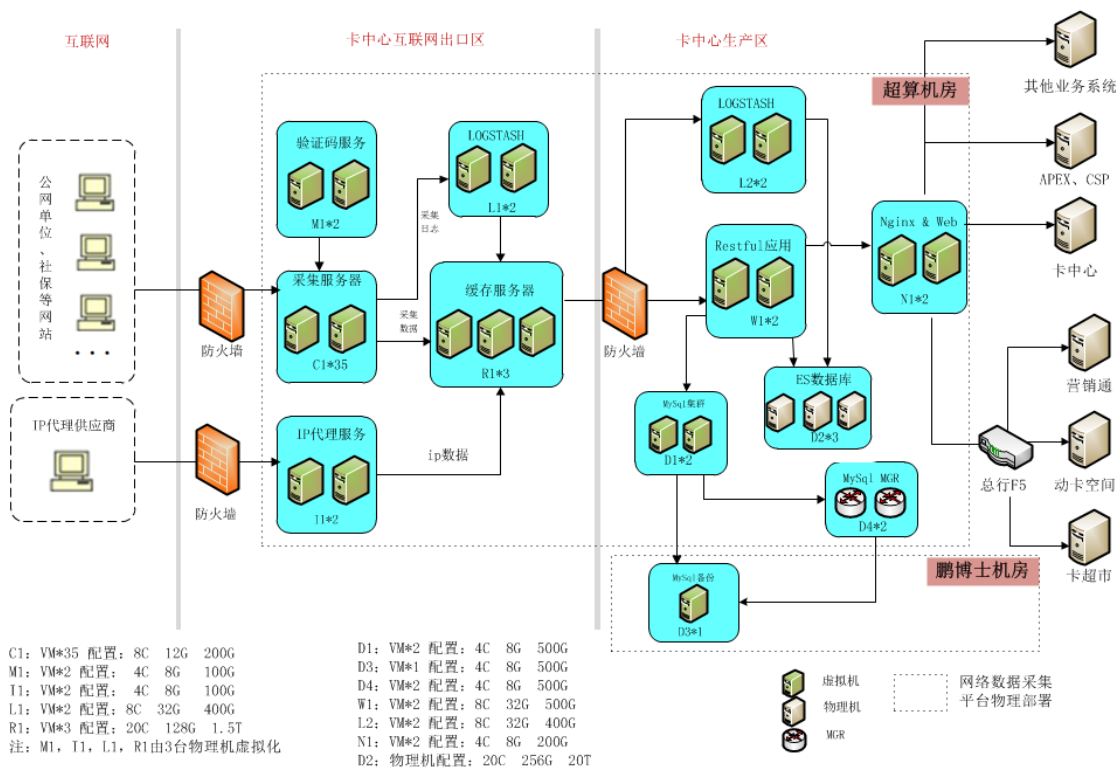
需求分析

中信银行信用卡中心在审核时，客户的社保信息，是以纸质盖章的形式提供。这个方式会引起造假、审核繁琐、用户体验差等问题。通过我们搭建的云采集平台，可以做到在客户授权后，自动化采集社保信息并提交系统审核。通过该服务，优化了客户的服务流程，也杜绝了造假的可能。

采集平台同时也为银行采集招聘网站数据、银行对公数据、房地产数据等提供了可能。对内部数据进行了多维度填充，在多个业务线上提供决策辅助。

解决方案

网络数据采集平台物理部署图



针对行方的高可用的需求，我方通过分布式、多节点部署。保证了单点故障不会引起系统不可用。达到了央行对于“银行业”技术安全规范中对于服务、数据库、接口等设计要求。

针对行方对于系统易管理性的要求，我们采取了动态采集配置。采集结果展示维度丰富。高自由度、安全性的权限设计、达到了行方运营人员的要求。

针对行方的性能要求，该系统通过了行方关于吞吐量、网络带宽、并发数、性能计数、承载能力的系统参数测试，客户端正常、系统运行正常。并且所有模块具备热横纵向拓展能力。

针对行方的安全性要求，我们遵循了以下规范：

- (1) 在日志中不出现口令明文、密钥等敏感信息；
- (2) 针对用户的每次操作，可在后端/前端提供日志及表格供查询审计。
- (3) 在配置文件中不出现口令明文。

- (4) 在数据库中不出现口令明文，涉及相关敏感信息必须加密处理。
- (5) 提供全站 HTTPS 访问服务。

3. 百果园案例

项目背景

2016 年水果销售外卖行业异军突起，百果园作为传统线下水果零售商受到了极大冲击，百果园在线上水果销售中表现不佳。

百果园急迫需要了解在外卖平台上有哪些竞争对手，做何种营销活动，对竞争对手进行产品分析、数据分析、服务分析、地域分析，从而得出百果园的最优线上营销方案。

需求分析

百果园通过数据，需要得到以下信息。

- (1) 区域运营数据：包括指定区域监控、我方产品在指定区域的市场差异，通过 BI 分析，从而优化运营提高销量。
- (2) 渠道分布数据：监控城市为单位的热销分布，获得全渠道市场数据。分析品牌份额及趋势。
- (3) 竞品数据对比：通过了解品牌分布及特征，掌握详细对比差异，监控竞品运营动态。

解决方案



利用我们小蜜蜂采集系统，突破美团、饿了么等多渠道反采集限制，通过技术模拟自动全中国各大城市 GPS 定位，实时返回全国所有地区的线上销售信息。

在数据处理分析上，我们设计了基于卷积的深度机器学习模型，问题中传统方式数据过拟合的情况，经分析是因为部分特征过拟合造成的，我们针对这个问题修正了神经网络的传递函数、并进行了部分人为干预，从而使得模型在全局最优和局部最优中达到了平衡。

通过该模型输入了半年的全国的抓取数据记录。通过训练，模型可以做到预测水果销量 90%以上的准确度（正负 10%）。在高维、复杂、非结构化、预测情形大数据处理任务中，获得了甲方的肯定。

三、 阿里云和流计算相关说明

阿里云

我公司部分第三方（或甲方）托管交付的项目，由采购阿里云服务支撑。包括云服务器、CDN、域名、数据库等。在阿里云我们有专属大客户经理，在与阿里云业务沟通中熟悉其工单系统、运维系统等。

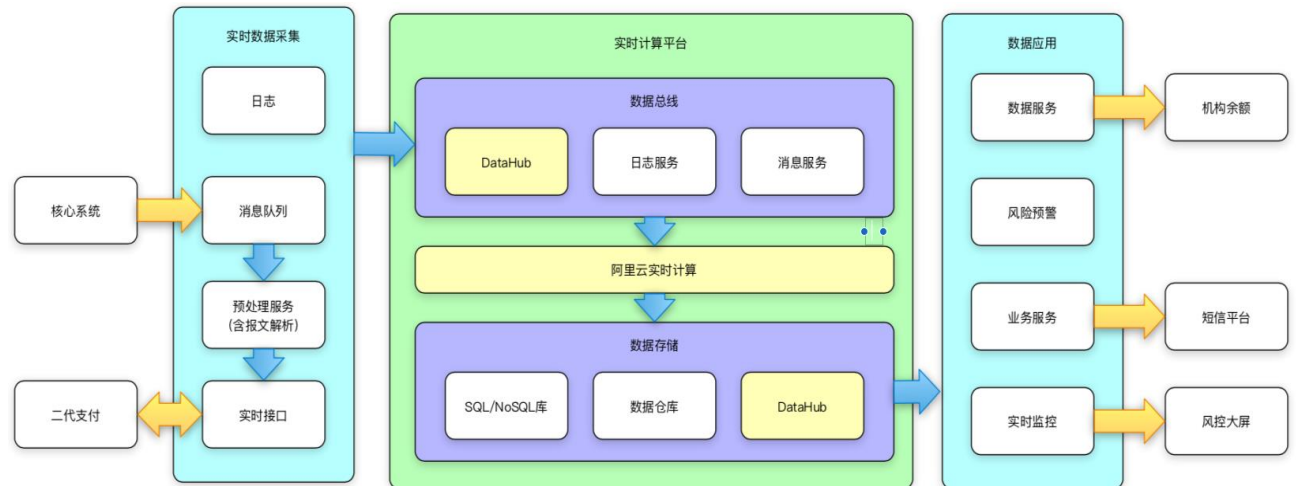
阿里云云服务在我方涉及第三方托管交付的项目中，在稳定性、需求边界确定、运维、定价等方面有着独到的优势。

流计算

我公司未采购过阿里云实时计算服务，主因是阿里云实时计算服务基于 Flink 开发。我公司有完整的 Flink 开发能力，无需额外支付成本采购阿里云 Flink 实时计算服务。在我公司搭建的基于 Flink 站式高性能实时大数据处理平台，在部署运维、作业开发、FlinkSQL 使用方面，都具有丰富的经验。如果贵公司以采购阿里云实时计算作为技术选型的条件下，更能如虎添翼，性能更上个台阶。

四、 实时计算平台项目技术方案

架构图



方案介绍

(一) 我们方案主要关注两个问题：

1、流计算引擎如何保证短信完整性？

首先，我们保证短信发送流程中，在多个对手交易系统、二代支付之间查询的事务性，在此基础上对于异常导致的发送问题，我们推送监控中心，根据运营人员的策略，定时重发。对于的重试仍然失败的情况（比如销号），针对异常情况定义异常模板，数据归集可追溯日志。

2、如何把实时采集的数据反馈到大屏上？

我们通过运营人员设置的定时服务需求，提交流计算处理，增量计算所需参数和展示结果。反馈给大屏实时显示。大屏中的参数可以根据监控需求自定义，计算的结果保证一定的冗余性，可以使得大屏的展示形式有更丰富的拓展空间。

(二) 问题应答：我公司通过前期交流，对贵公司关心的些问题进行回答：

问题一：我公司的内部数据处理能力水平

我公司的内部数据处理能力例如高性能运算、OCR 识别、NLP 自然语义分析、分布式计算存储等技术处于业界技术前列。一般来说，外部数据是离散的、不稳定的、非结构化的（如采集数据），除了数据本身的因素，还有许多其他因素（比如外部数据采购）的制约。我方案例合同中往往更侧重于对外部数据的要求。是因为内部数据处理难度往往低于外部数据，属于依赖关系。如果想使得外部数据处理符合客户要求，内部数据的完整处理是外部分析的前提。通过许多案例，可证明我们的大数据处理技术处于业界技术前列。

问题二：如何实现最大化减少现有交易系统的变更

对于旁路系统所遇到的问题，在我方银行业技术方案中（如中信、郑州银行等）和金融业技术方案（金道、联合利丰等）中有着丰富的经验。遵循 BASE 解决方案（基本可用+软状态+最终致性），使得分布式系统在响应和最大化减少现有系统变更两个需求上，得到最完美的平衡。

问题三：实时机构存贷款余额，如何结合我行的数据仓库

封装后打包成数据服务提供给外围系统？

这个问题本质是一个检索问题，不存在流计算需求。性能上首先可以考虑在数据库上添加检索索引、创建中间表等，打包过程可以利用调度中心保证整体事务性，选择合适的打包策略进行提交。流程上可以以 API 的形式提供给外围系统被动调用，或通过 Kafka 推送外围系统主动调用。

五、 核心风险关注点

1. 风险点列举

- a) 对交易系统的影响
- b) 实时性和一致性的矛盾

2. 风险点分析与解决方案

(1) 对交易系统的影响

为对交易系统尽可能小的影响, 实时平台应当遵循 2 个原则:

- a) 全部只读, 保障即使出现系统缺陷与错误也不影响交易系统
- b) 被动接入, 最大化减少现有交易系统的变更

如果能保证这两个原则, 就可以保证影响最低

(2) 实时性和一致性的矛盾

基于分布式系统 CAP 理论, 一致性 C、可用性 A 和分区容错性 P 三者在一套系统中不可能被同时满足, 也就是必须至少牺牲其中一项. 因为实时性需要有可用性做支撑, 因此为了保障完全的一致性(实时一致性)必须完全保障 C 和 A, 只能牺牲 P, 但是牺牲 P 意味着牺牲了可扩展性. (在当前案例中, 牺牲分区容错意味着所有的组件间不可能存在网络分区, 也就意味着物理上约束所有组件连接于同一个交换机甚至部署于单机环境, 这对系统扩容和扩展上是一个极大的约束)

实际上业界对于这种场景, 有基本达到最优解的妥协性解决方案, 称为 BASE 方案, 这个解决方案包含 3 个决策。

- 1) 基本可用 (BasicallyAvailable)

基本可用指不保证完全的可用性保障, 在发生(可能是未预期的)故障时, 受到影响的服务处于等待状态。这可以使得主线程不受干扰, 核心业务不受威胁。

2) 软状态 (Softstate)

软状态又叫柔性事务, 即状态可以有一段时间的不同步。这种柔性同步使得并发提高性能提高。

3) 最终一致 (Eventuallyconsistent)

放弃强一致性, 在一个指定的时间窗口之后达到一致性, 使得锁的分配和设计更为合理。

六、 建设周期/计划

项目周期和计划基于现在的已知需求和信息，无法准确提供。

基于过往案例分析，标 1 类似项目工期大约在 3-6 个月左右，工时在 15-35 个人月附近，标 2 类似项目工期大约在 4-8 个月左右，工时在 25-50 个人月附近。主要取决于内部数据的现状、接入接口的复杂度、对调用性能状态的需求以及上线条件和限制。

企业质量管理体系认证证书



质量管理体系认证证书

兹证明

深圳市比一比网络科技有限公司

统一社会信用代码: 9144030068375453XL

注册地址:中国广东省深圳市南山区粤海街道高新技术园中区科苑大道讯美科技广场1栋3楼306

审核地址:中国广东省深圳市南山区粤海街道高新技术园中区科苑大道讯美科技广场1栋3楼306

建立的管理体系经审核符合以下标准要求:

GB/T19001-2016/ISO9001:2015

认证范围

计算机软件开发及服务

证书编号: ZZW17Q14039R0S

发证日期: 24/11/2017 有效期至: 23/11/2020

注:组织须按规定接受每年监督审核,并遵守相关认证规定,认证证书有效性才能保持有效。
认证注册范围不包括未获得有效的国家规定的行政许可,资质许可和3C认证的产品服务范围

证书签发人:

总经理



注:按照认证要求,每次监督审核时间与上次现场审核时间间隔不得超过12个月。

上海中致威认证有限公司 电话: 021-39528556 传真: 021-64267178

地址: 上海市嘉定区鹤旋路26弄29号(万达广场9号楼)507室

证书有效性查询方式: 中致威公司网站 www.zzwz.com; 国家认证认可监督管理委员会网站 www.cnca.gov.cn

发明专利（4项）

HTML 文档信息抽取表达式的方法及系统

证书号第 2735424 号



发明专利证书

发明名称：一种 HTML 文档信息抽取表达式的方法及系统

发明人：杜卫红; 谢立欧

专利号：ZL 2014 1 0681853.6

专利申请日：2014 年 11 月 24 日

专利权人：深圳市比一比网络科技有限公司

授权公告日：2017 年 12 月 12 日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年 11 月 24 日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。



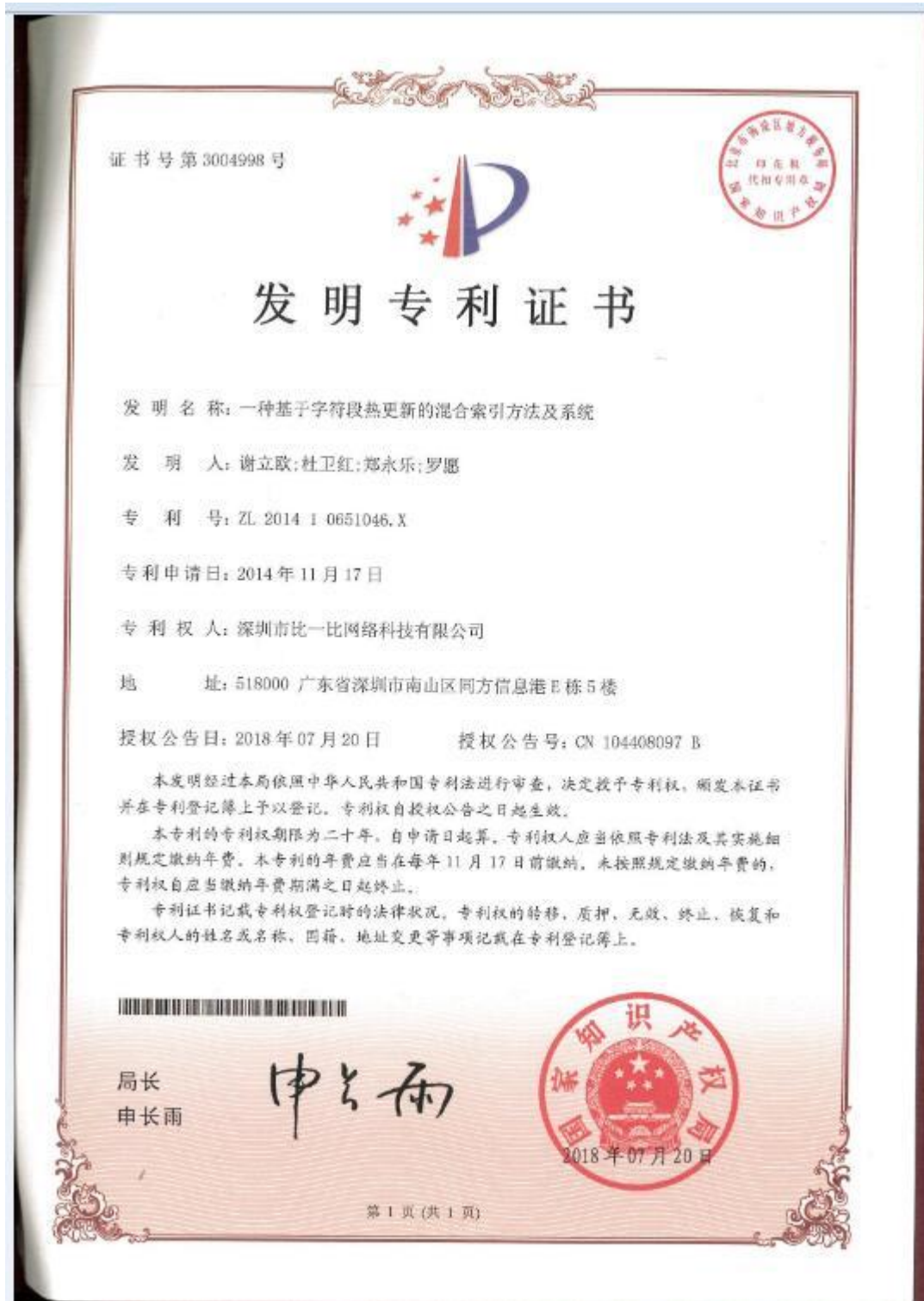
局长
申长雨



2017 年 12 月 12 日

第 1 页 (共 1 页)

基于字符热更新的混合索引方法及系统



基于旋转投影宽度的拉丁字符倾斜纠正方法及系统

证书号第 3217704 号



发明专利证书

发明名称：一种基于旋转投影宽度的拉丁字符倾斜纠正方法及系统

发明人：杜卫红;谢立欣;蒋立成

专利号：ZL 2014 1 0651047.4

专利申请日：2014 年 11 月 17 日

专利权人：深圳市比一比网络科技有限公司

地址：518000 广东省深圳市南山区同方信息港 E 栋 5 楼

授权公告日：2019 年 01 月 15 日 授权公告号：CN 104408452 B

国家知识产权局依照中华人民共和国专利法进行审查，决定授予专利权，颁发发明专利证书并在专利登记簿上予以登记，专利权自授权公告之日起生效，专利权期限为二十年，自申请日起算。

专利证书记载专利权登记时的法律状况，专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。



局长
申长雨

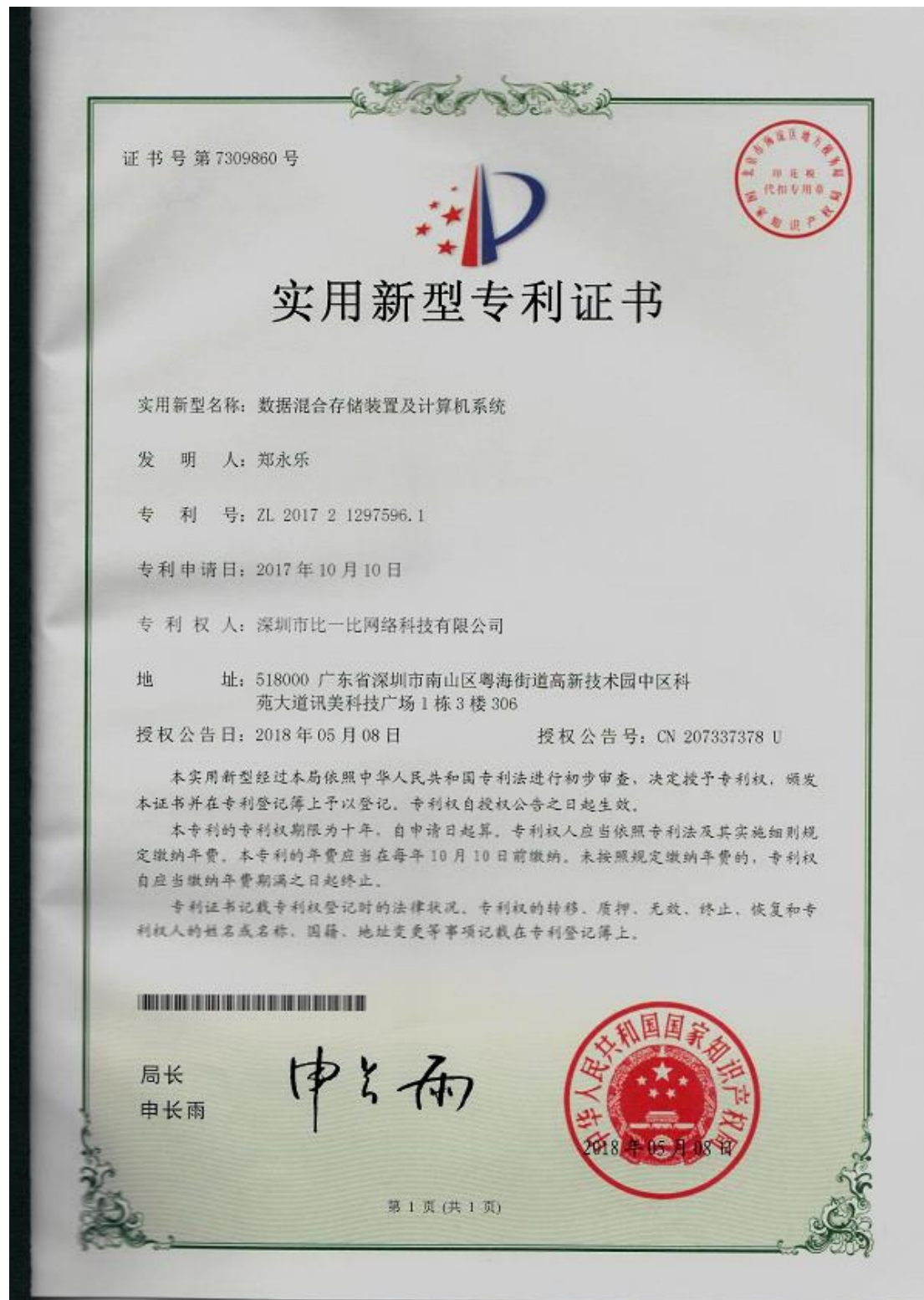


2019 年 01 月 15 日



第 1 页 (共 2 页)

其他事项参见背面

数据混合存储装置及计算机系统



软件著作权（部分）

中华人民共和国国家版权局	
计算机软件著作权登记证书	
证书号： 软著登字第0829428号	
软件名称：	比一比混合索引分析器系统软件 [简称：混合索引分析器系统] V1.0.0
著作权人：	深圳市比一比网络科技有限公司
开发完成日期：	2012年12月24日
首次发表日期：	未发表
权利取得方式：	原始取得
权利范围：	全部权利
登记号：	2014SR160191
根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。	
	
No. 00564439	 2014年10月24日

中华人民共和国国家版权局
计算机软件著作权登记证书

证书号： 软著登字第0829490号

软件名称： 比一比日志管理系统软件
[简称： 日志管理系统]
V2.1

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2012年06月16日

首次发表日期： 2012年06月18日

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2014SR160253

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。



No. 00564445



2014年10月24日

中华人民共和国国家版权局
计算机软件著作权登记证书

证书号： 软著登字第0829488号

软件名称： 比一比数据控制中心系统软件
[简称： 数据控制中心系统]
V2.0

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2013年07月30日

首次发表日期： 未发表

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2014SR160251

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。



No. 00564444

2014年10月24日

中华人民共和国国家版权局
计算机软件著作权登记证书

证书号： 软著登字第0829252号

软件名称： 比一比Web文档爬取系统软件
[简称： Web文档爬取系统]
V5.0

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2014年07月30日

首次发表日期： 未发表

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2014SR160015

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。



No. 00564400

2014年10月24日

中华人民共和国国家版权局
计算机软件著作权登记证书

证书号： 软著登字第0828259号

软件名称： 比一比Web文档信息抽取系统软件
[简称： Web文档信息抽取系统]
V2.0

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2013年07月30日

首次发表日期： 未发表

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2014SR159022

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。



No. 00563706



中华人民共和国国家版权局
计算机软件著作权登记证书

证书号： 软著登字第2156126号

软件名称： 比一比云采集数据中心系统
V1.0

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2016年12月20日

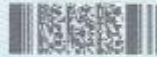
首次发表日期： 2016年12月26日

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2017SR569841

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的
规定，经中国版权保护中心审核，对以上事项予以登记。



No. 02025295



210

中华人民共和国国家版权局 计算机软件著作权登记证书

证书号： 软著登字第2153966号

软件名称： 比一比财讯管理系统
V1.0

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2017年05月10日

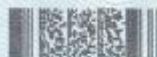
首次发表日期： 2017年05月20日

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2017SR568582

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。



No. 02025186

2017年10月16日

15/10

中华人民共和国国家版权局 计算机软件著作权登记证书

证书号： 软著登字第2151105号

软件名称： 比一比云采集操作系统
V1.0

著作权人： 深圳市比一比网络科技有限公司

开发完成日期： 2016年12月20日

首次发表日期： 2016年12月26日

权利取得方式： 原始取得

权利范围： 全部权利

登记号： 2017SR565821

根据《计算机软件保护条例》和《计算机软件著作权登记办法》的规定，经中国版权保护中心审核，对以上事项予以登记。



No. 02020521

