

实时计算平台交流



公司介绍

深圳市比一比网络科技有限公司成立于2009年, 公司总部位于深圳市南山区科技园, 现正筹建美国洛杉矶分公司. 公司以机器学习与大数据服务为主营业务, 主要面向金融行业提供大数据和人工智能类解决方案与相关技术服务. 公司拥有30多项软件著作权与4项大数据领域发明专利技术.

公司愿景：
让数据产生
伟大价值！

- 杜卫红 (CEO)



www.shutterstock.com · 181806107

比一比是国家级高新技术企业, 双软认证, 取得互联网信息技术服务证书, ISO 9001 认证证书, CMMI 3 级认证证书等. 同时是深圳市互联网技术协会副会长单位.

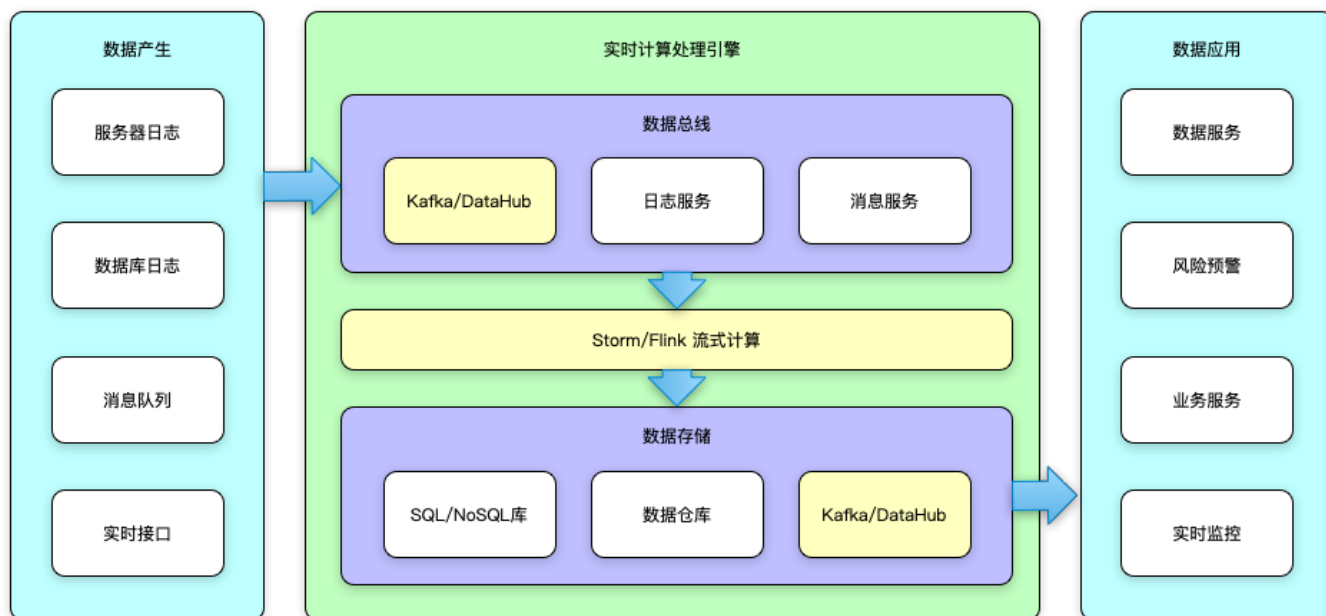
公司一直致力于大数据收集, 分布式存储, 数据分析与挖掘, 数据可视化方面的研发. 为客户提供一站式的大数据解决方案. 主要服务对象为泛金融行业, 包括银行, 商业保理公司, 担保公司, 互联网金融, 供应链金融公司.

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.

主讲介绍

郑永乐(Syeerzy Zheng), 比一比解决方案高级顾问, 大数据与机器学习专家, 同时持有国家科技部和劳动部联合颁发的分析师(高级)职业资格证书, 超过 15年 IT 行业管理与顾问经验, 自2010年至今在比一比公司主持过多个人工智能与数据处理大型项目. 同时积极参与大型数据类开源项目, 是百度分布式存储系统 BFS 核心贡献者, 和 F# 编程语言核心库贡献者.

实时计算平台总体架构



数据产生

日志

服务日志

数据库日志

消息队列

互联网爬虫

业务系统异步推送

实时接口

业务系统同步调用

外部同步 API 接口

实时数据处理

数据总线

日志服务

消息服务

Kafka / DataHub

流式计算

Storm

Flink / 云流计算

数据存储

关系数据库

Kafka / DataHub

HBase / MongoDB / Redis

数据应用

数据服务

机构余额

灵活查询

业务服务

短信通知

客服平台

预警监控

风险预警

大屏监控

实施风险点与难点

核心风险关注点:

- (1) 对交易系统的影响
- (2) 实时性
- (3) 一致性

难点分析与方案:

为对交易系统尽可能小得影响, 实时平台应当遵循2个原则:

- (1) 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
- (2) 被动接入, 最大化减少现有交易系统的变更

同时, 基于分布式系统 CAP 理论, 一致性C / 可用性A / 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, 有比较标准的妥协性解决方案, 称为BASE方案, 包含3个决策:

- (1) 基本可用 (Basically Available)

基本可用的意思就是放弃了完全的可用性保障, 在发生(可能是未预期的)故障时, 受影响的服务处于等待状态.

- (2) 软状态 (Soft state)

软状态又叫柔性事务, 即状态可以有一段时间的不同步.

- (3) 最终一致 (Eventually consistent)

放弃强一致性, 在一个指定的时间窗口之后达到一致性.

任务作业开发流程

1. 创建项目
2. 准备上下游数据 (含测试数据)

- 2.1. Kafka topic
- 2.2. DataHub 源表
- 2.3. 数据库结果表
- 3. 数据开发
 - 3.1. 新建作业
 - 3.2. 引入上下游数据存储
 - 3.2.1. 引入上游Kafka/Datahub Topic
 - 3.2.2. 注册结果表
 - 3.2.3. 定制的 API 开发 (选)
 - 3.3. 编写业务脚本 (如 SQL)
 - 3.4. 在流计算平台调试
- 4. 作业上线
 - 4.1. 资源配置
 - 4.2. 数据检查
- 5. 生产运维
 - 5.1. 后台启动业务
 - 5.2. 设置启动时间
 - 5.3. 运行
 - 5.4. 查看结果
- 6. 业务运维
 - 6.1. 作业状态监控
 - 6.2. 数据曲线
 - 6.3. 异常监控
 - 6.4. 状态快照
 - 6.5. 任务管理
 - 6.6. 属性参数设置

公司相关实施方案案例

- 1. 中信银行
 - 1.1. 灵活查询平台
 - 1.1.1. 实时查询
 - 1.1.2. 海量数据
 - 1.1.3. 模板化
 - 1.1.4. 数据服务
 - 1.2. 云架构升级采集平台
 - 1.2.1. 信用卡核卡
 - 1.2.2. 企业基本信息

1.2.3.舆情

1.2.4.地址库

2. 平安集团

2.1.财险大数据

2.2.同业理财

2.3.上市公司公告智能分析应用

3. 郑州银行 / 广州农商

3.1.数据管理平台

4. 香港金道集团

4.1.实时财讯大数据中台

4.2.智能数据报表平台

5. 银达担保

5.1.工程担保风控平台

6. 银泰保理 / 联合保理

6.1.智能保理解决方案

7. 联合利丰

7.1.供应链金融智能中台

8. 百果园 / 天虹

8.1.零售大数据平台

9. 中车集团

9.1.二级涉密项目

9.1.1.轨道安全

9.1.2.行车安全

场景结合

