



第九届中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2017



荣之联大数据平台的应用实践

2017/10

□ 荣之联大数据平台的应用案例介绍

- 商务中心大数据中心建设案例

- 证券交易日志分析案例

- 工业物联网大数据平台案例

□ 荣之联大数据平台产品介绍

- 产品架构及优势

- 产品特色功能介绍

□ 荣之联大数据平台的应用案例介绍

□ 商务中心大数据中心建设案例

□ 证券交易日志分析案例

□ 工业物联网大数据平台案例

□ 荣之联大数据平台产品介绍

□ 产品架构及优势

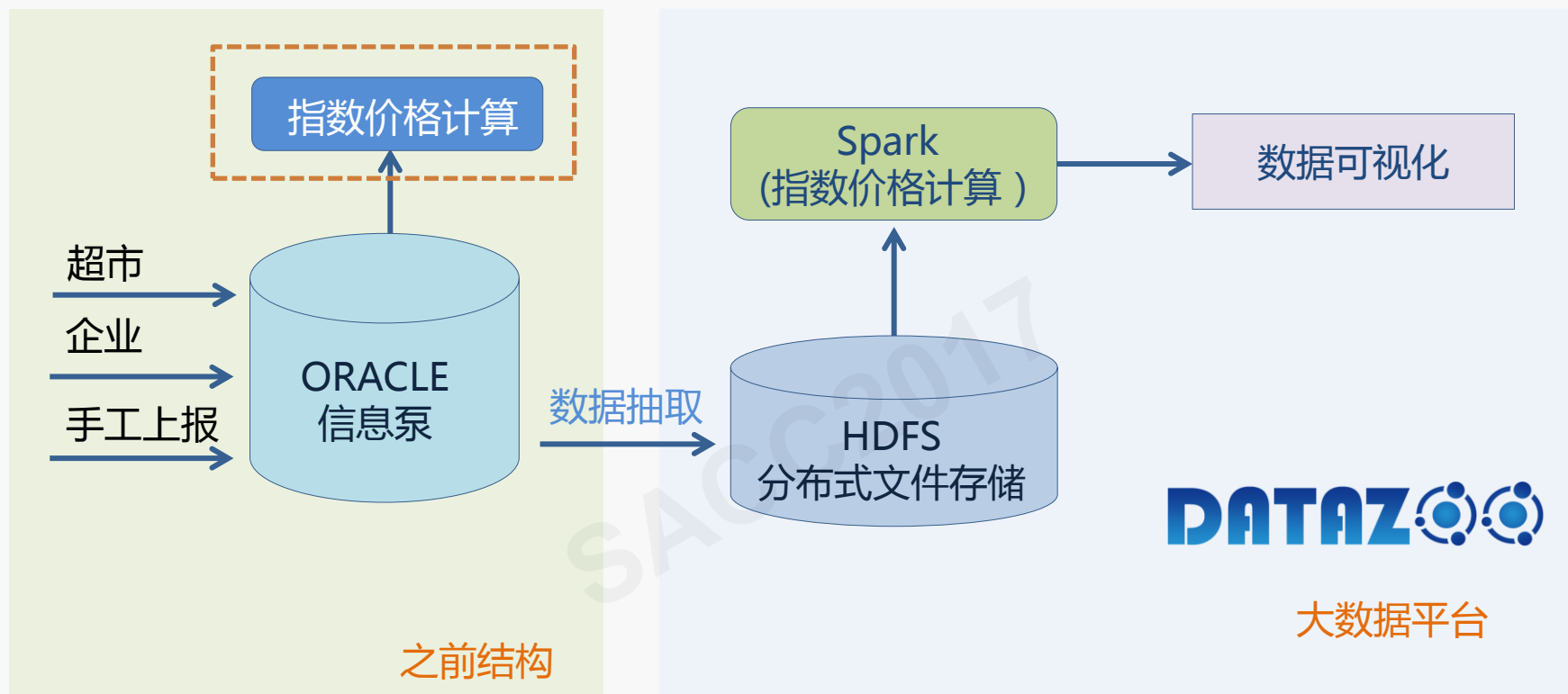
□ 产品特色功能介绍

SACC2017

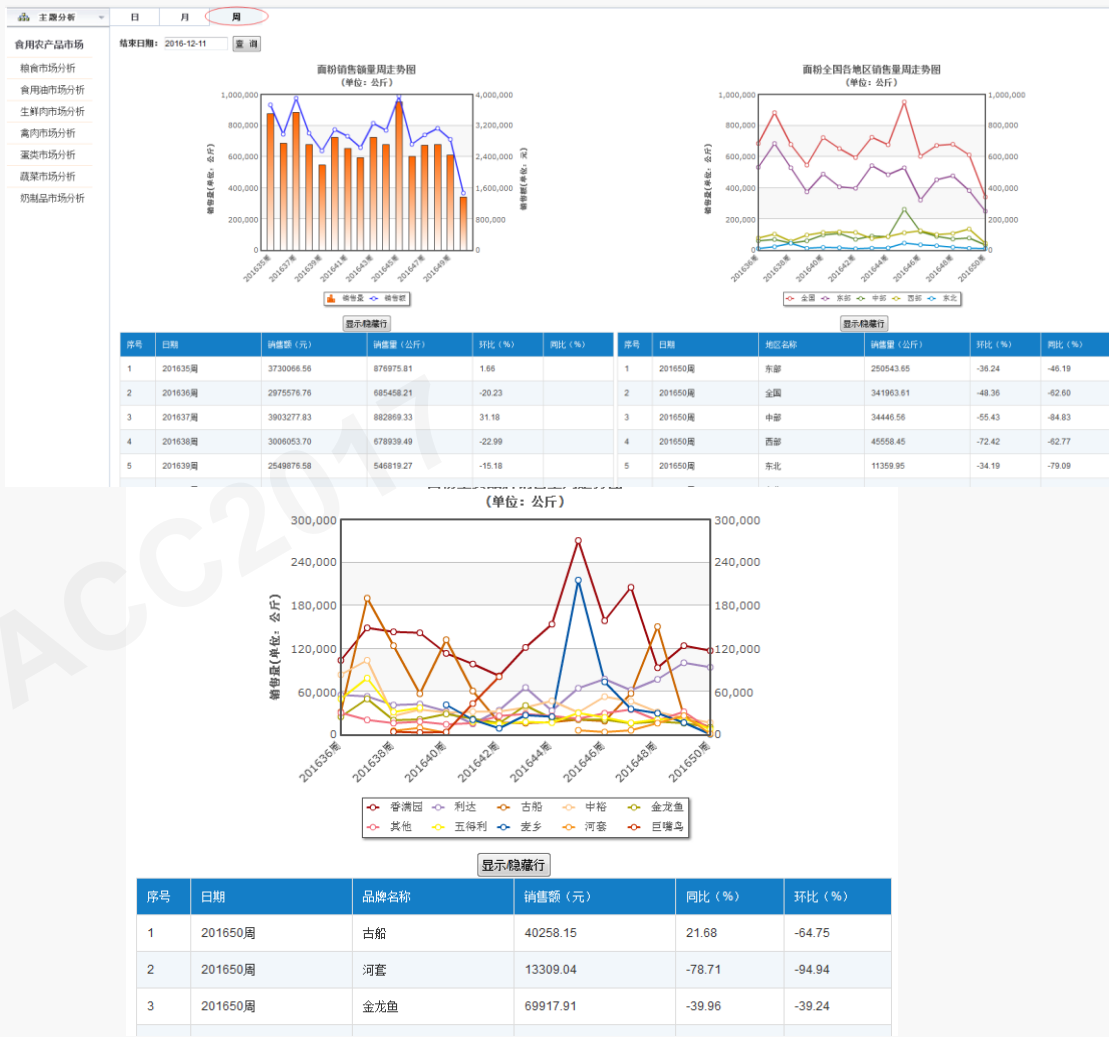
商务中心大数据中心建设需求



商务中心大数据中心建设



- ✓ 构建了大数据平台
- ✓ 拥有处理更大数据量、更复杂格式、更多样化数据的能力
- ✓ 加速数据采集、使用和分析的时间，更快决策
- ✓ 处理的数据量：一个月1-1.3亿条数据，即10-13Gb
- ✓ 原来的处理+计算需要5个小时，现在需要2个半到3小时。



□ 荣之联大数据平台的应用案例介绍

- 商务中心大数据中心建设案例

- 证券交易日志分析案例

- 工业物联网大数据平台案例

□ 荣之联大数据平台产品介绍

- 产品架构及优势

- 产品特色功能介绍

X券商交易系统现状与问题



网上交易



手机委托



电话委托



柜台委托

柜台交易系统

资产管理交易系统

私募交易系统

期货交易系统

.....



开户系统

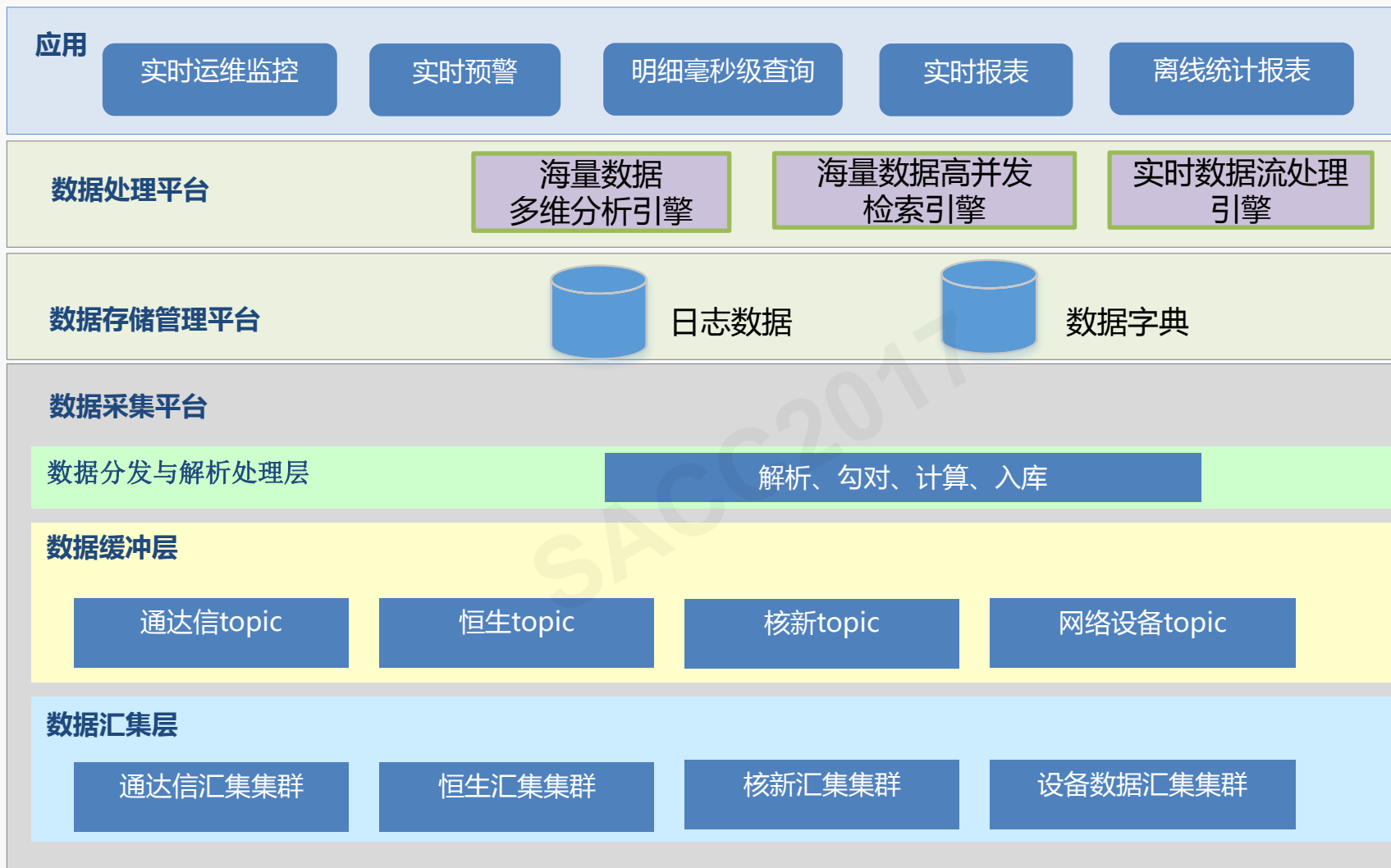
- 全国**22**个站点
- 每个站点约20台左右的网上交易事务机，共计**近400**台网上交易事务机
- 每天网上交易原始日志的日增量为**90-120G**

分散在全国的各个站点和服务器的状态如何监控？出现宕机或假死状态时能否及时判断和处理？

面对证监会的监管要求，或者客户的查询请求，如何对原始日志进行快速、精确的查询？

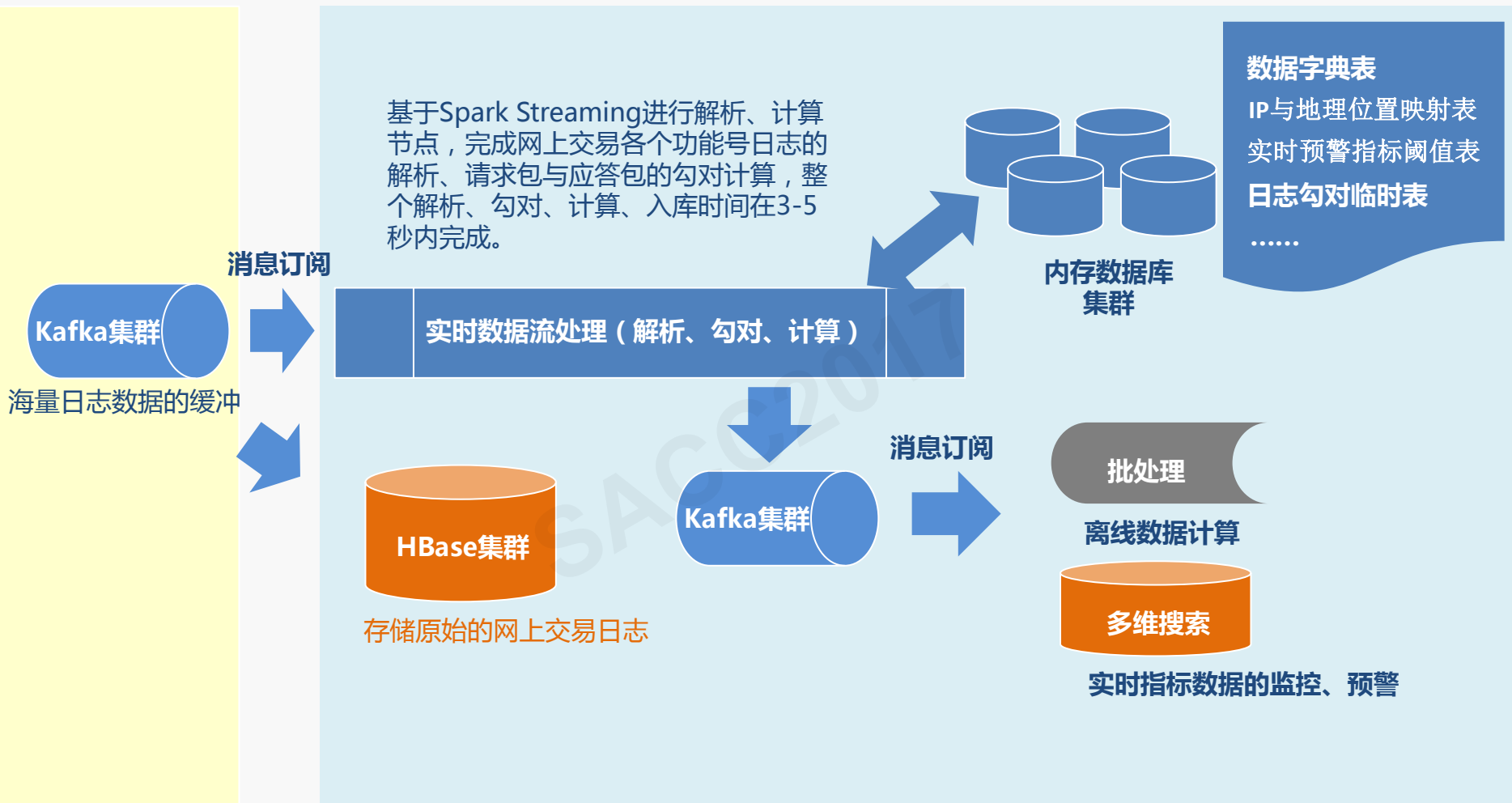
基于历史数据的统计分析报表的计算效率是否需要提高？

券商应用日志分析系统架构

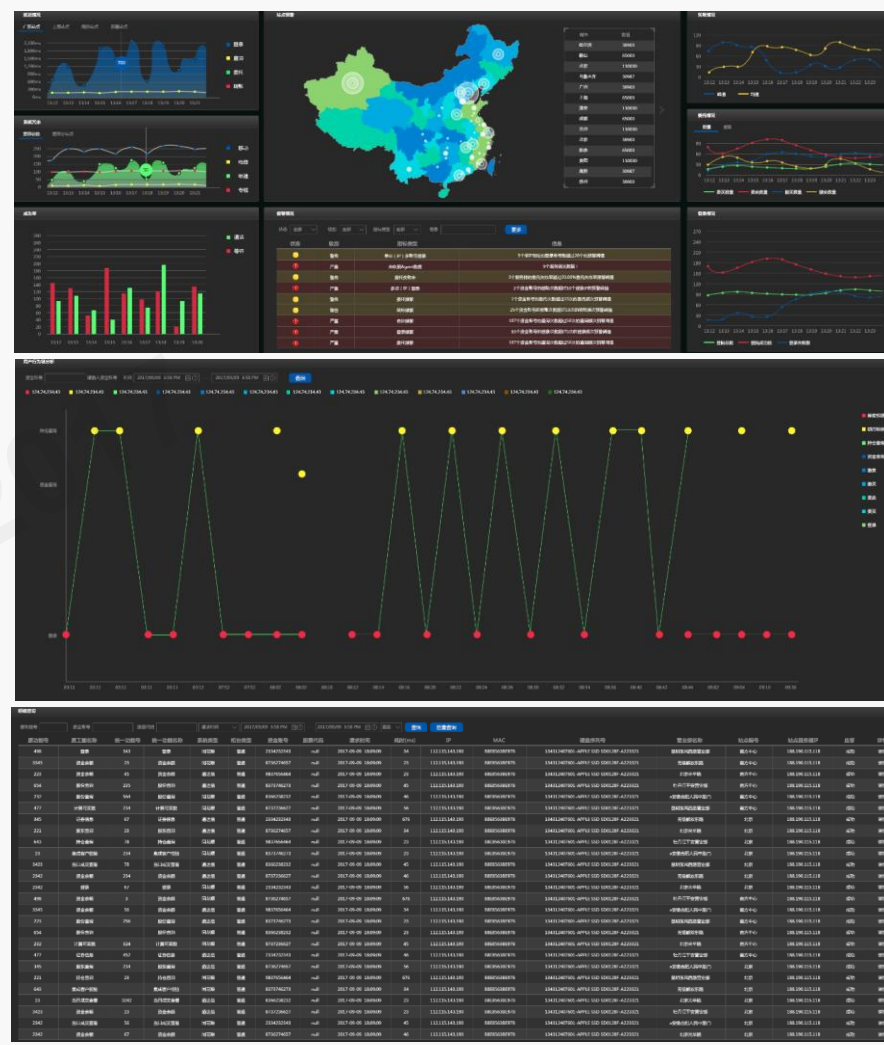


监控及运维管理

券商应用日志分析系统 – 实时计算



- ✓ 为信息技术部**提供更完善的运维管理支持**，对站点/交易事务机的各技术指标做到实时监控，针对交易事务机的假死、宕机、恶意攻击等异常情况做到及时有效应对，防止影响到正常网上交易业务的正常运行。
- ✓ 针对资金账号的各种**异常情况的实时预警**，协助及时发现异常的资金账号，做出更加有效的管控和处理。
- ✓ **秒级的日志查询**，面对监管方的日志查询要求或客户请求，快速响应。
- ✓ 依托大数据技术的日志实时**采集与分析平台的搭建**，为某证券公司未来实现全系统、各交易品种的日志接入和大集中管理、基于海量数据的业务分析做了平滑的铺垫。



□ 荣之联大数据平台的应用案例介绍

- 商务中心大数据中心建设案例

- 证券交易日志分析案例

- 工业物联网大数据平台案例

□ 荣之联大数据平台产品介绍

- 产品架构及优势

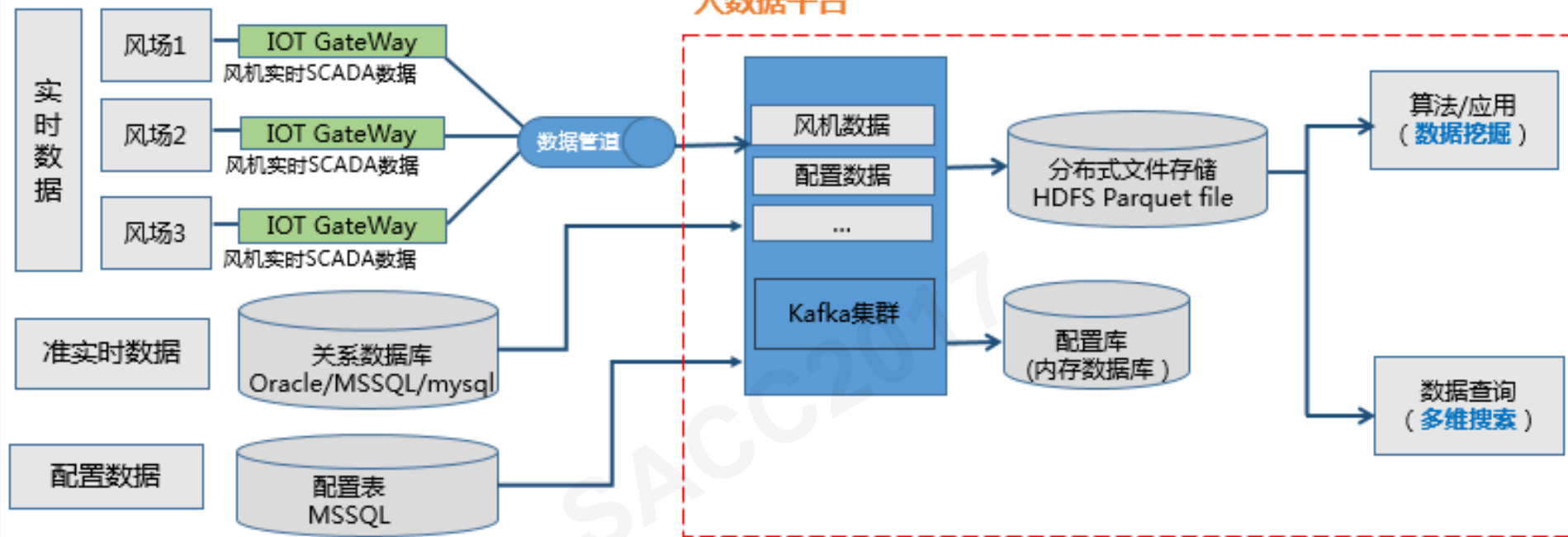
- 产品特色功能介绍

风机数据资源现状

数据应用	数据管理	数据平台
<ul style="list-style-type: none"> • 数据存储分散：统计数据、历史全量数据是云存储，本地存储载荷分析相关的数据 • 风资源数据分散：散落到不同的业务系统中 	<ul style="list-style-type: none"> • 子系统多而独立：信息重复、信息堡垒和信息孤岛等问题出现 • 不能快速抓取到数据：跨部门的业务及新业务需求需要数据支撑时，不能快速的抓取到所需要的数据（风机数据，业务数据，水务数据等等） • 集团级统计数据慢：数周时间才能获得集团级统计分析数据 	<ul style="list-style-type: none"> • 数据平台试验环境：目标是生产环境 • 改进现有的数据处理业务：云平台存在系统安全隐患及其他未知的风险

能源物联网大数据平台搭建

大数据平台



数据采集与处理平台

数据采集代理
数据流处理
数据入库

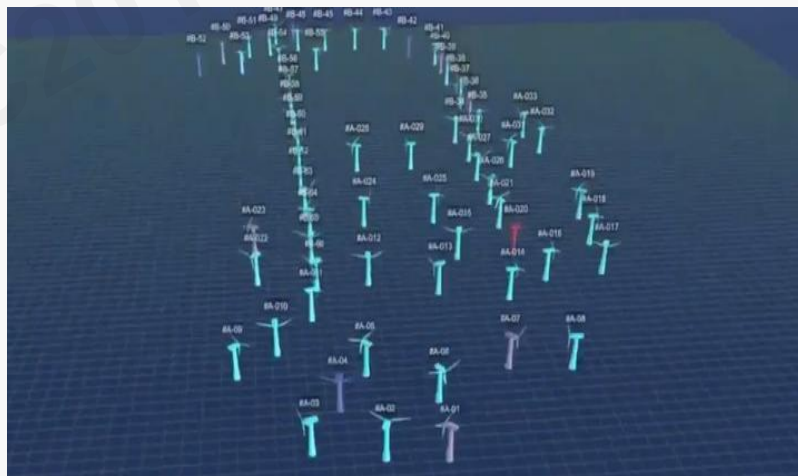
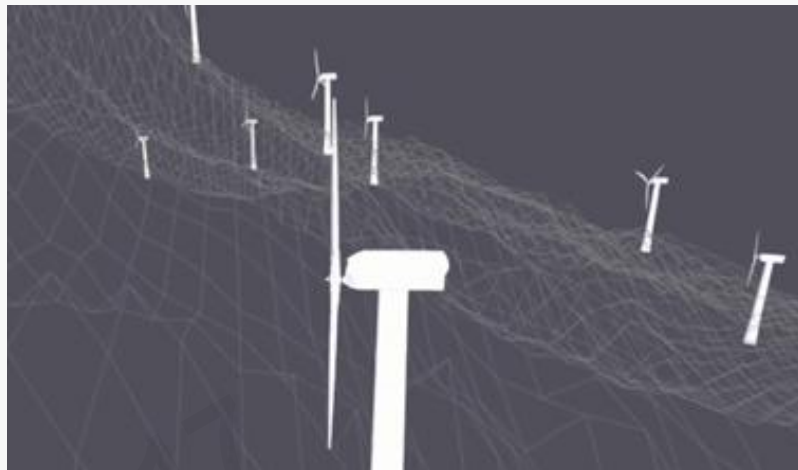
数据存储

分布式文件存储
采用Parquet 列存储
优化存储模型

计算优化

优化数据读取性能
算法本身优化
采用数据科学开发工具

- ✓ 可弹性伸缩的大数据平台
- ✓ 数据集中存储，推动企业**数字化转型**
- ✓ 有效支撑**物联网**下大数据的应用
- ✓ 数据资产**集中化管理**，包含：数据采集和存储、数据查询、数据应用等
- ✓ 快速为管理层和业务层提供**数据服务**



□ 荣之联大数据平台的应用案例介绍

- 商务中心大数据中心建设案例

- 证券交易日志分析案例

- 工业物联网大数据平台案例

■ 荣之联大数据平台产品介绍

- 产品架构及优势

- 产品特色功能介绍

荣之联DataZoo大数据平台

工具
套件

数据采集工具

数据分析工具

数据挖掘工具

数据可视化



Ambari Manager 运维管控

批处理

实时流处理

多维搜索

图计算

数据挖掘



Hadoop 生态系统

协同服务
Zookeeper

批计算 MapReduce2	内存计算 Spark	SQL Hive	NoSQL HBase	消息系统 Kafka	数据传输 Flume	数据传输 Sqoop
YARN资源管理系统						
HDFS分布式文件系统						



Security
安全权限

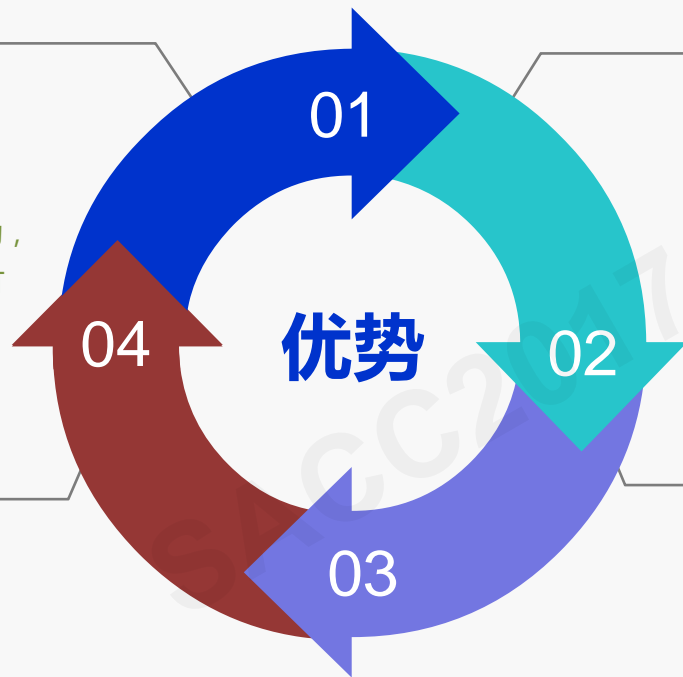
Linux Operation System 操作系统

完整的数据处理能力

- 新一代大数据平台，不仅仅是Hadoop生态的集成
- 具备数据分析生命周期各方面能力，从数据采集、分析、挖掘到数据可视化都有相应的组件能力
- 平台自有特色功能

丰富的行业应用模型

- 驾驶行为分析
- 碰撞分析模型
- 用户画像
- 推荐模型
- 文本分析（分词、情感分析）



高性能的多维分析引擎

- 超低时延：分析结果一触即发
- 分析特性：任意维度组合分析
- 存储特性：嵌套列存储，计算过程不加载多余数据
- 强扩展性：支持横向纵向任意扩展
- 数据时效性：实时+离线数据

便捷的开发管理工具

- 可视化管理、监控系统
- 统一的配置管理
- 从数据采集、预处理到数据分析与挖掘的工具套件

□ 荣之联大数据平台的应用案例介绍

- 商务中心大数据中心建设案例

- 证券交易日志分析案例

- 工业物联网大数据平台案例

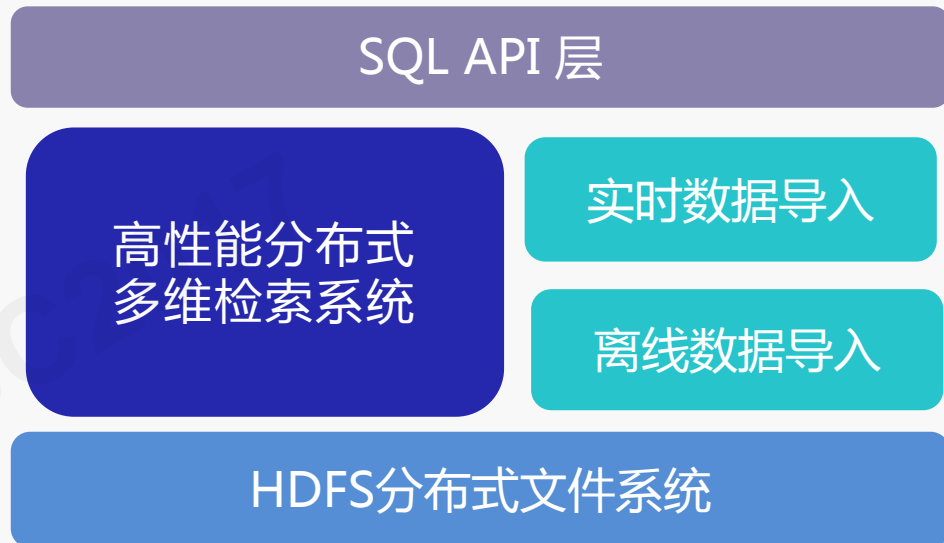
□ 荣之联大数据平台产品介绍

- 产品架构及优势

- 产品特色功能介绍

特色能力-多维搜索

- 将数据存储存储在HDFS之上，基于HDFS做了磁盘与网络做了读写控速逻辑
- 与Spark深度集成，Spark对检索结果集直接分析计算，同样场景让Spark性能加快百倍
- 数据即可离线导入也可实时导入，索引即时生成，通过索引高效定位到相关数据

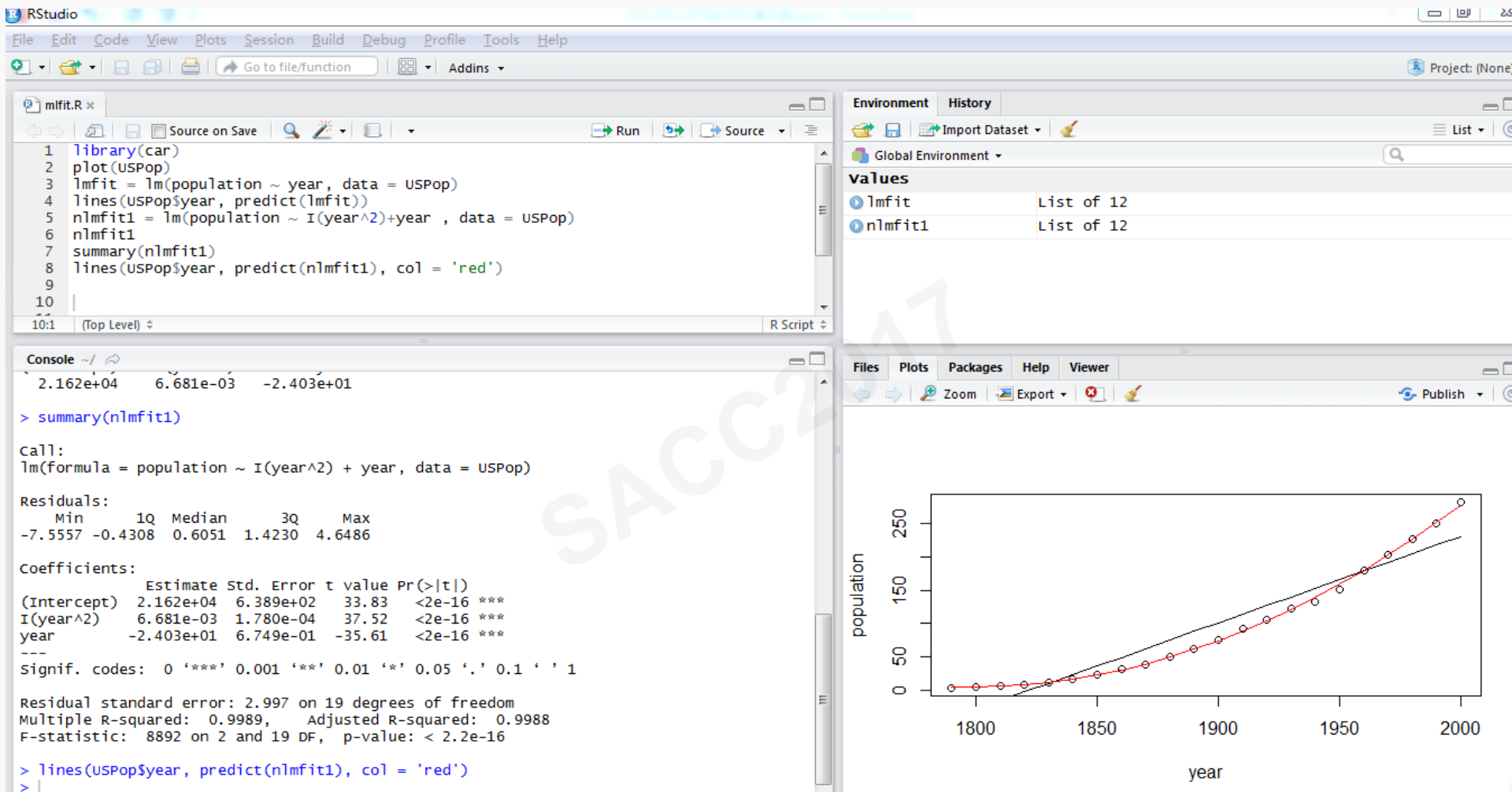


特色能力-行业模型

驾驶行为模型	用户画像	文本分析-分词
识别用户风险，改善车主驾驶行为，降低车险的赔付成本。	标签化用户模型，提供360度客户视图。	构建大量文本的切词模型，是文本分析的基础模型。
碰撞分析模型	推荐模型	文本分析-情感分析
通过识别事故真相减少车险的欺诈赔款，并提供更好的理赔服务。	基于用户的行为数据，做出精准式营销。	基于上下文语义，语句情感模型，应用于舆情监控、商品评估等场景。

Predictive 预测	Segmentation 分类	Associations 关联	Correlations 相关性	Similarities 相似度	分词 模型	情感分析 模型
Tress 树归纳法	DBSCAN	Apply Association Rule 应用关联规则	ANOVA Matrix 矩阵	Cross Distances 交叉距离	Conditional random field 条件随机场 CRF	Long Short Term Memory (长期短期记忆) 本模型为神经网络模型
Neural Nets 神经网络训练	Top Down Clustering	Create Association Rule 创建关联规则	Correlations Matrix 相关矩阵	Data to Similarity		
Functions 函数拟合	K-Means k-均值	FP-Growth	Covariance Matrix 协方差矩阵	Data to Similarity data		
Logistic Reg 逻辑回归	K-Medoids k-Medoids聚类	Generalized Sequential Patterns	Mutual Information Matrix	Similarity data		





文件(F) 编辑 视图 执行 工具 帮助

Perspective: Data Integration

主对象树 核心对象

Entries

- 通用
- 邮件
- 文件管理
- 条件
- 脚本
- 批量加载
- Big Data
- XML
- 应用
- 资源库
- 文件传输
- 文件加密
- Palo
- Deprecated

dbincr

100%

START

显示消息对话框

执行结果

历史 日志 作业度量 Metrics

任务 / 任务条目	注释	结果	原因	文件名
test111				
任务: test111	开始执行任务		开始	
START	开始执行任务		开始	
START	任务执行完毕	成功		
显示消息对话框	开始执行任务		Followed无条件的链接	
显示消息对话框	任务执行完毕	成功		
START	开始执行任务		开始	
START	任务执行完毕	成功		
显示消息对话框	开始执行任务		Followed无条件的链接	
显示消息对话框	任务执行完毕	成功		
START	开始执行任务		开始	
START	任务执行完毕	成功		



THANKS