



奉献清洁能源
建设和谐社会

互联网+时代下智能电网大数据应用思考

——南瑞瑞中数据
王远

Who ?

江苏瑞中数据股份有限公司 (China Realtime Database Co.,Ltd.) 由南瑞集团公司牵头业内多家高科技公司共同投资组建，注册资金1亿元，是国网内专业的基础软件企业。公司主营业务为通用数据库管理系统、企业级数据中心、大数据解决方案的研究开发、市场营销、工程实施和技术咨询。

瑞中数据立足自有核心技术，大力发展从数据存储、一体化平台、开放型应用到数据维护、增值服务的数据整体解决方案，树立海迅 (HighSoon) 品牌形象，努力营造大数据商业生态系统，实现数据产业商业模式创新，帮助客户实现信息化管理和数据服务最佳体验



1

互联网与智能电网大数据

2

行业相关应用

3

智能电网大数据应用架构

4

技术趋势思考

下一代智能电网

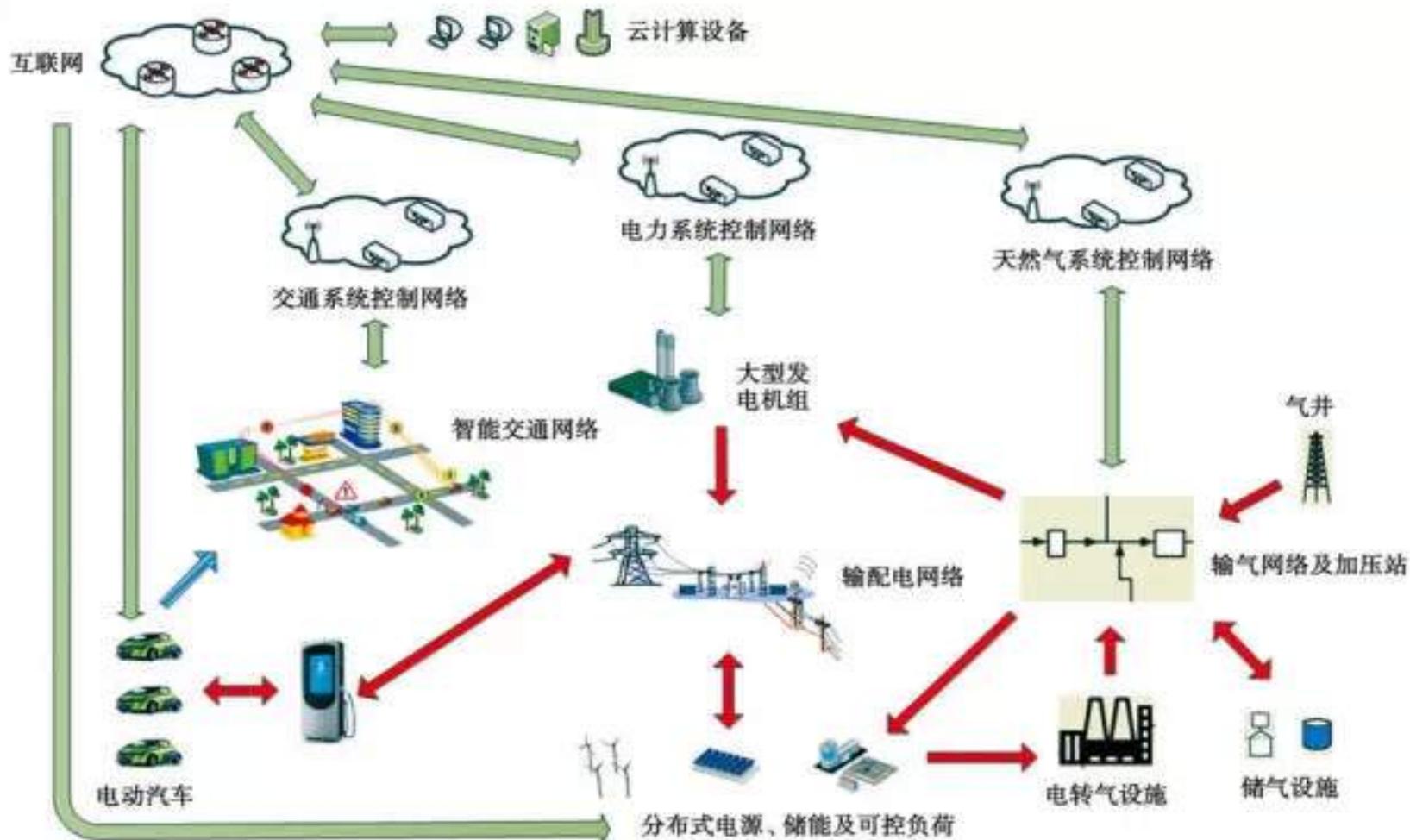
-----能源互联网

- 定义

- 能源互联网是以电力系统为核心，以互联网及其他前沿信息技术为基础，以分布式可再生能源为主要一次能源，与天然气网络、交通网络等其他系统紧密耦合而形成的复杂多网流系统。
- 其目的是实现可再生能源，尤其是分布式可再生能源的大规模利用和共享。

董朝阳, 赵俊华, 文福拴, 薛禹胜

《从智能电网到能源互联网:基本概念与研究框架》



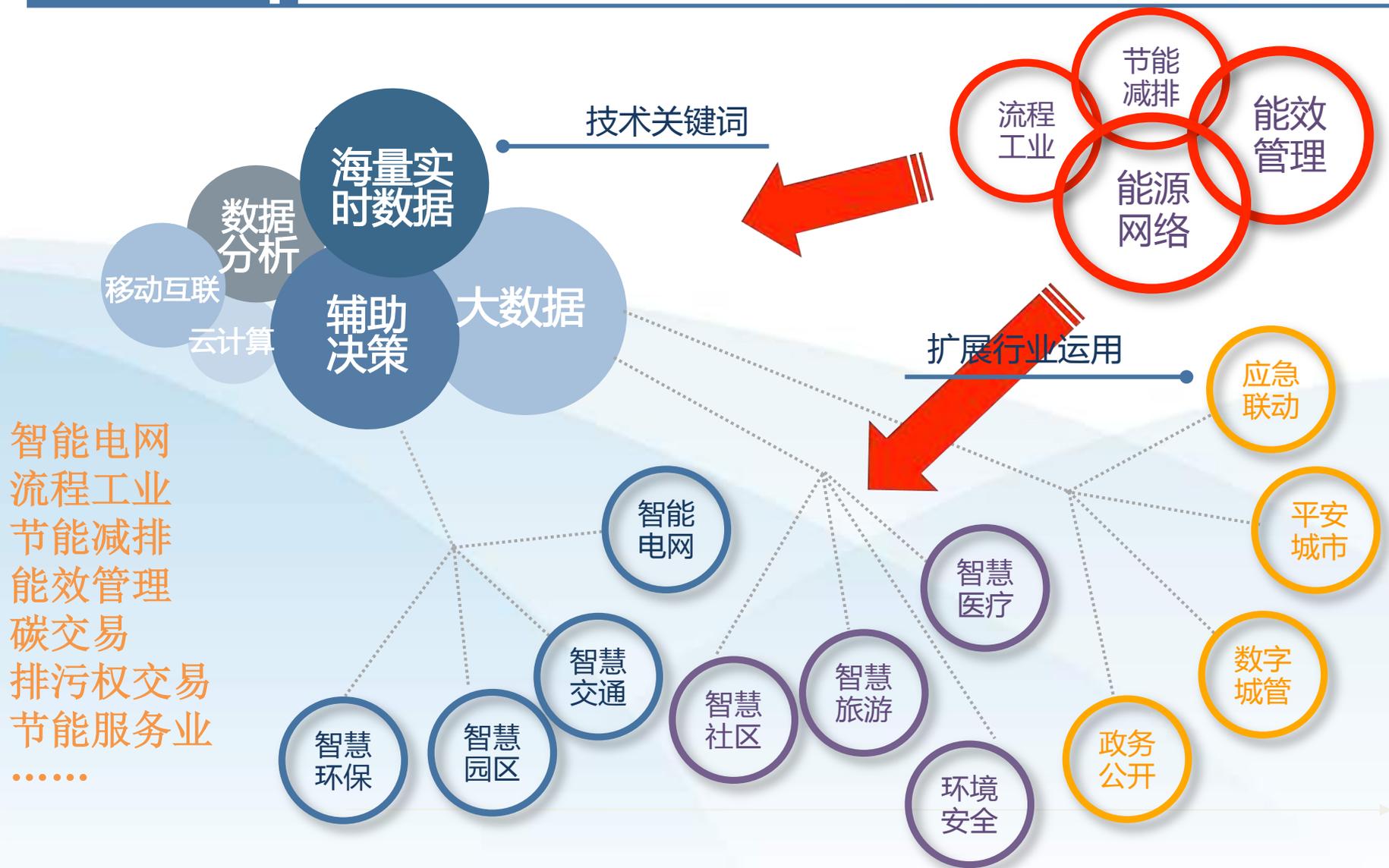
能源互联网的基本架构与组成元素
Basic architecture and components of Energy Internet



基础设施一体化
--能源流
--信息流
--控制流

交互模式社会化
--高度互联
--互通互动
--泛在智能

在互联网+时代下，依托大数据相关相关前沿技术，推动能源及其周边产业智慧化是未来发展的趋势之一



智能电网
流程工业
节能减排
能效管理
碳交易
排污权交易
节能服务业
.....

数据:

- 传感器
- 生产设备
- 文档文本
- 交易记录
- 邮件、消息
- 视频、图片
- 运营商数据
- 交通卡口
- 物流
- 金融
- 环境监测
-



1

互联网与智能电网大数据

2

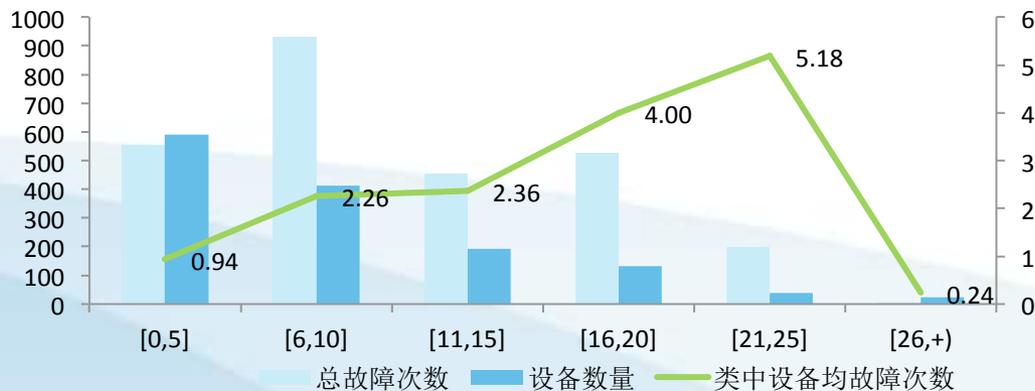
行业相关应用

3

智能电网大数据应用架构

4

技术趋势思考



按5年分类的主变设备故障数量与使用年限关系

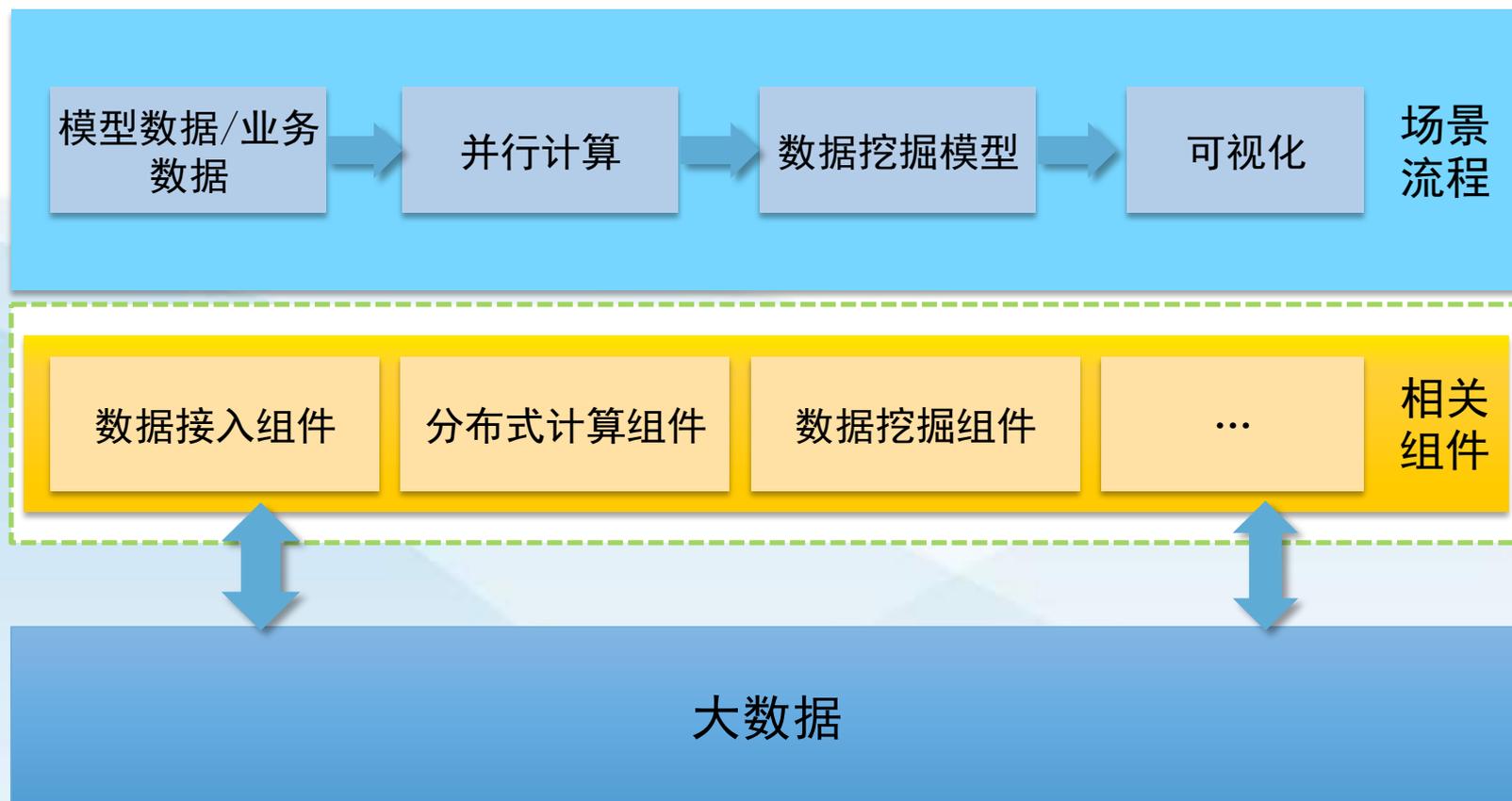


用数据挖掘技术得出的主变设备故障数量与使用年限关系

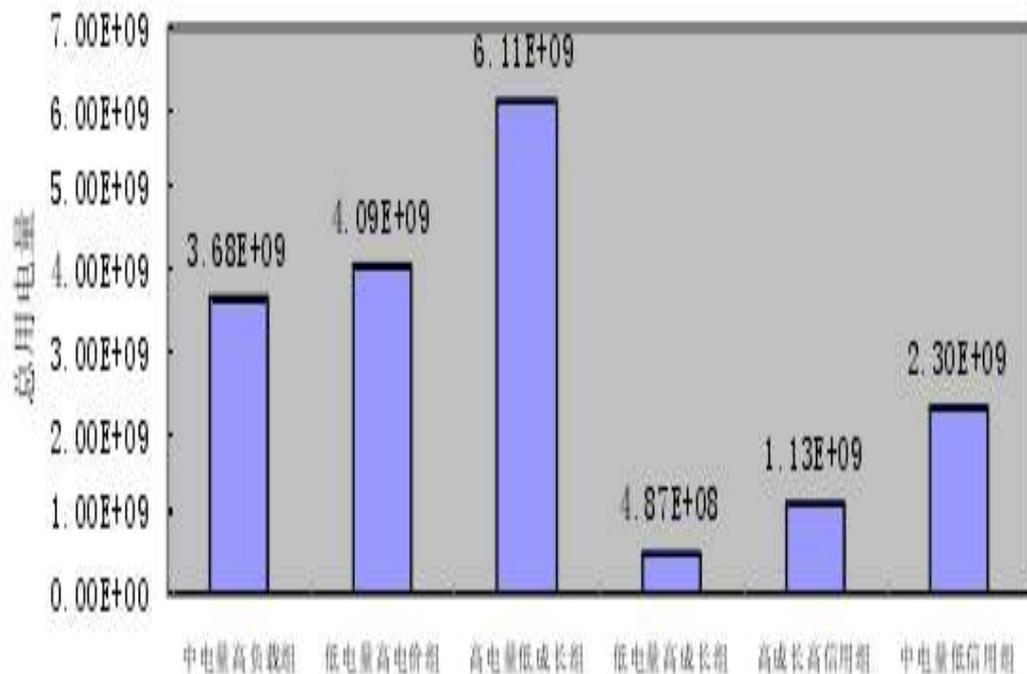
可对电网历史运行过程中任意时间段的历史回放，即具备历史时间段中全部采集数据的追忆能力。



设备状态分析-海量数据知识提取

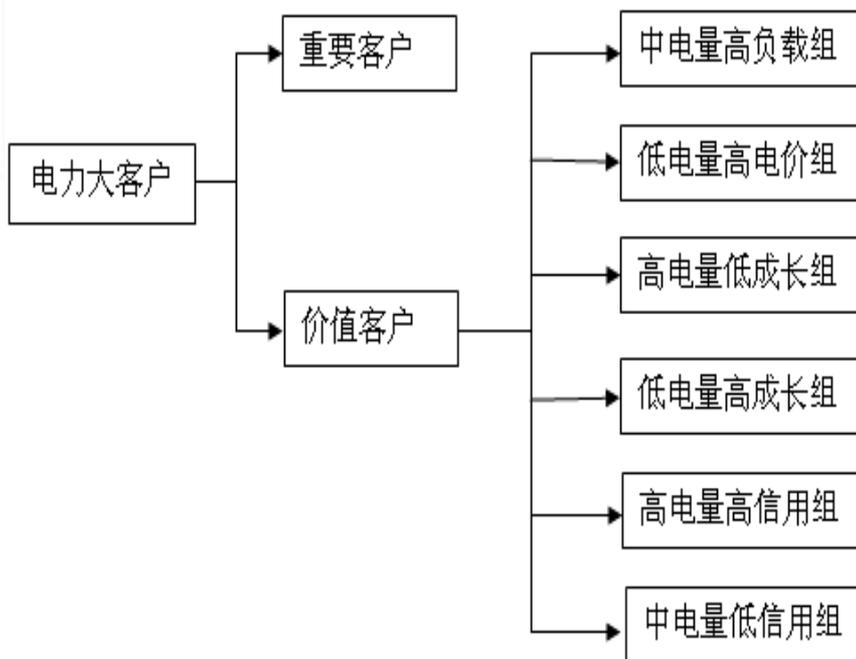


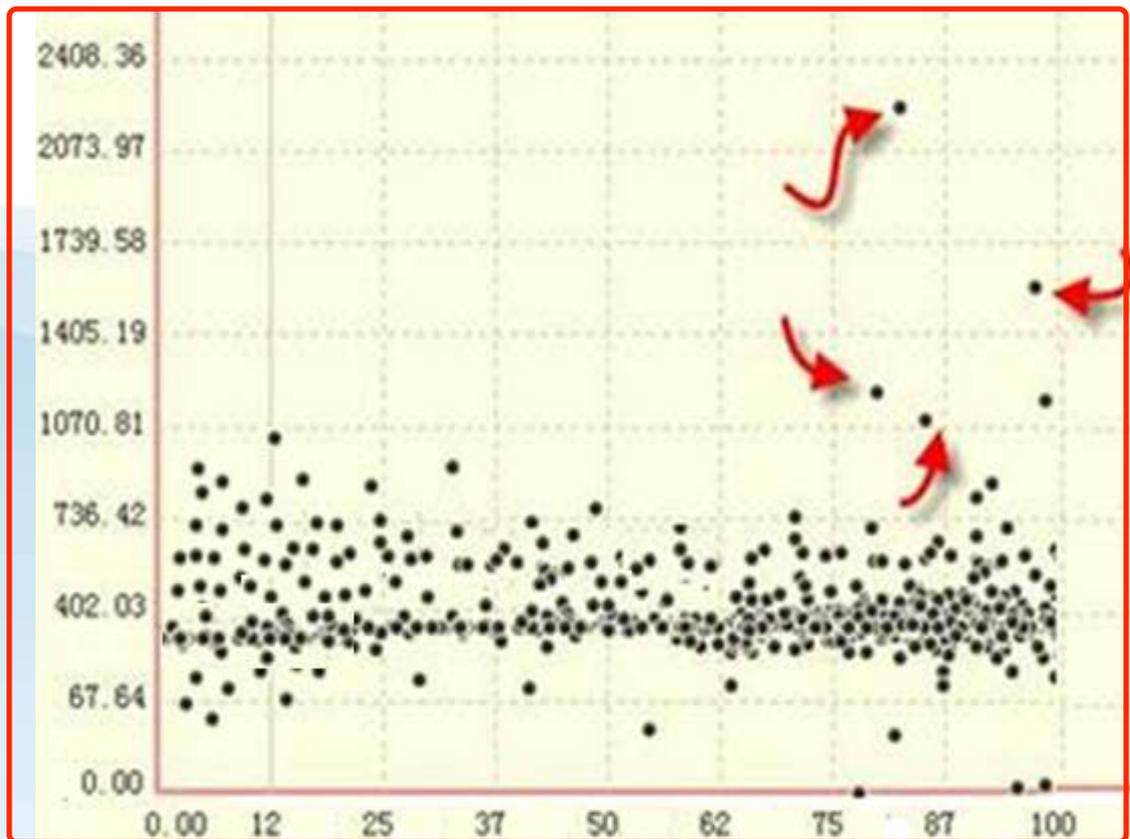
该应用场景特点是统计数据量庞大，时间跨度长、同时需要结合数据挖掘算法进行知识提取。



大客户经济行为细分结果

客户细分各组总用电量分布图





选取5000个信用同组的电力客户样本100天数据；



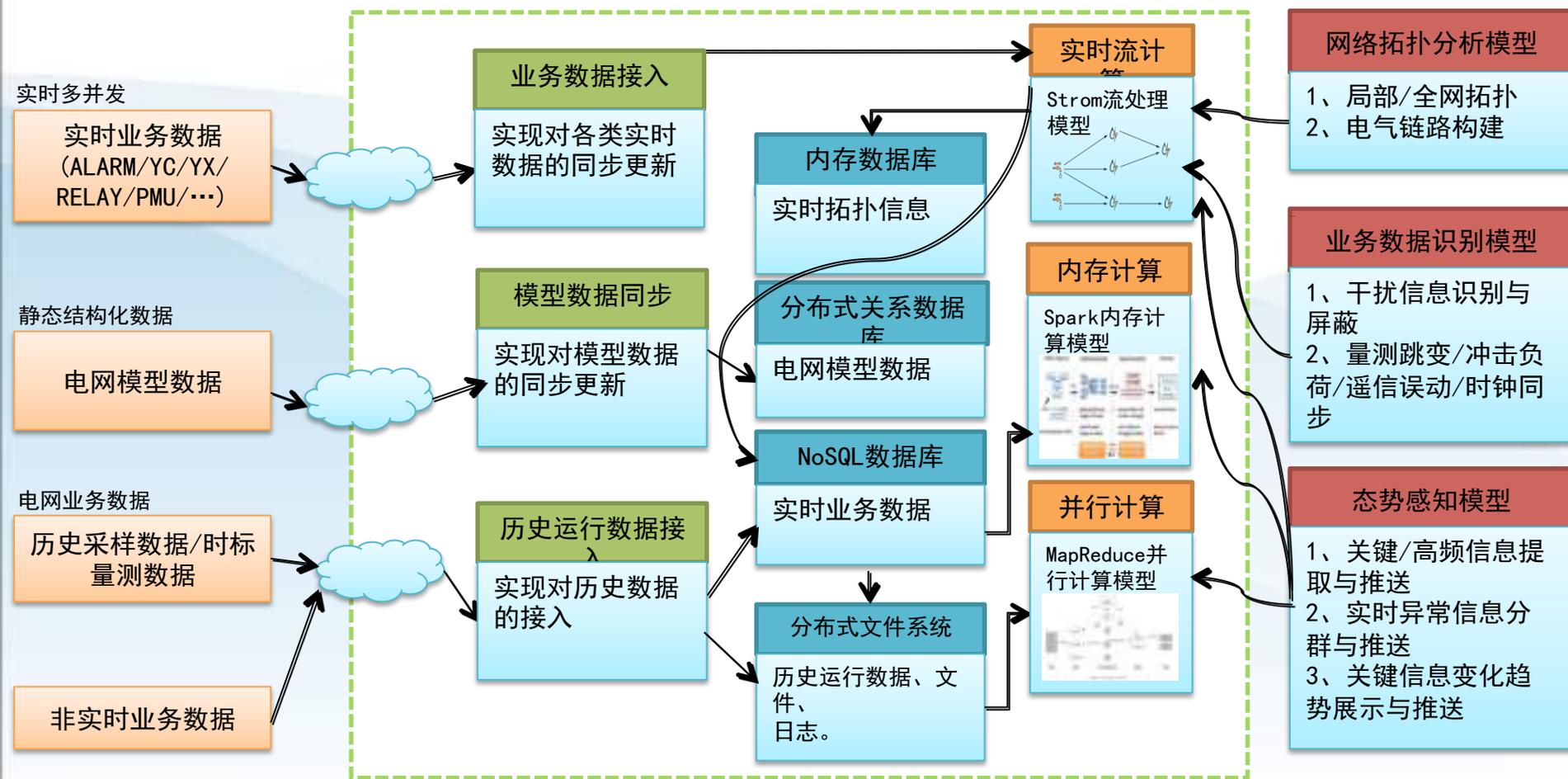
预处理操作后，进行离群点检测，最终检测结果如右；



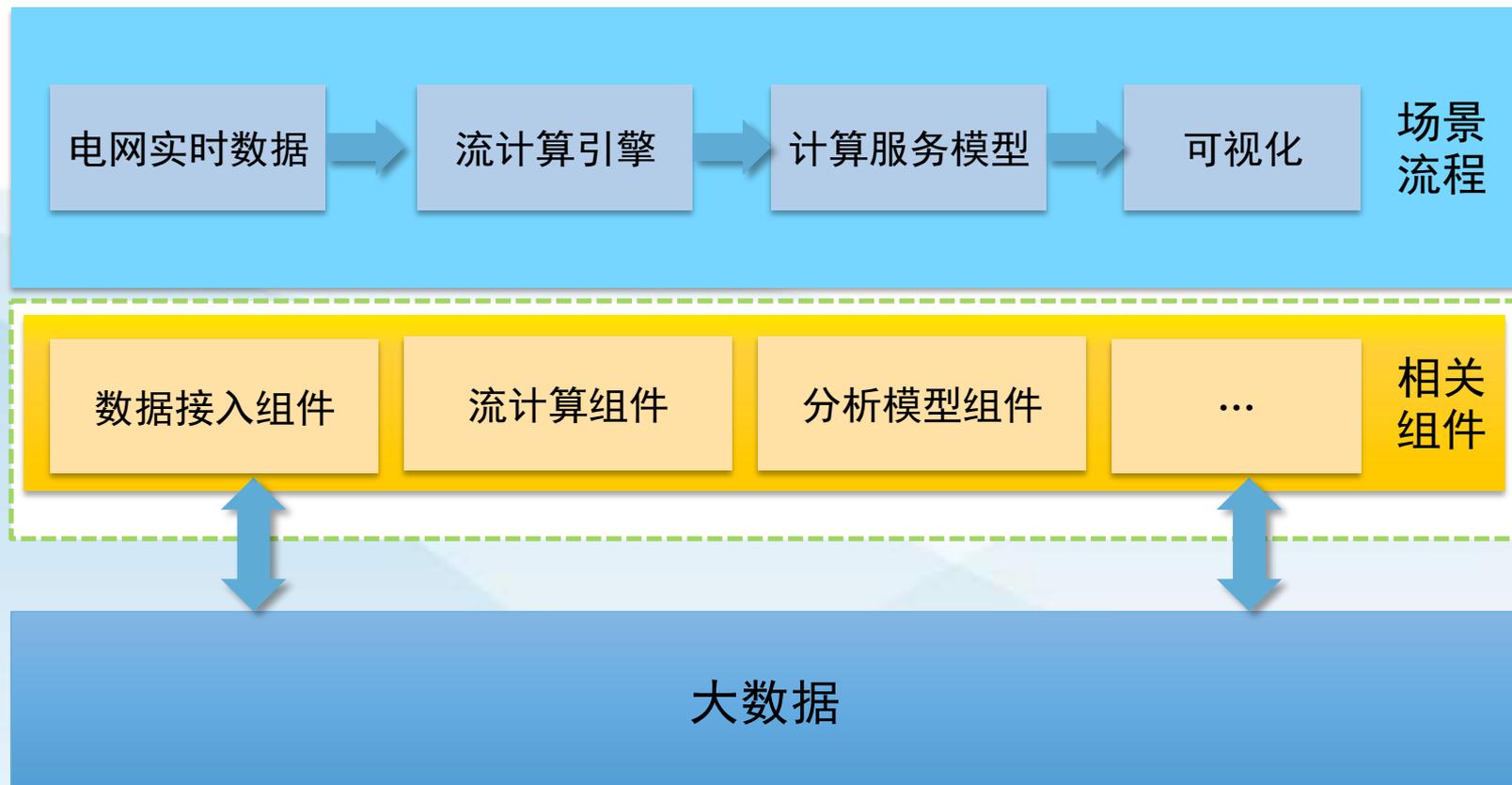
对离群点进行重点检测，防止盗电现象发生；

电力客户离群点分析结果

在电网态势感知应用场景中，大数据技术接入电网的各项业务数据，利用分布式流计算实时处理能力，实现对电网的整体态势感知，使得业务人员全面实时地了解电网运行情况。

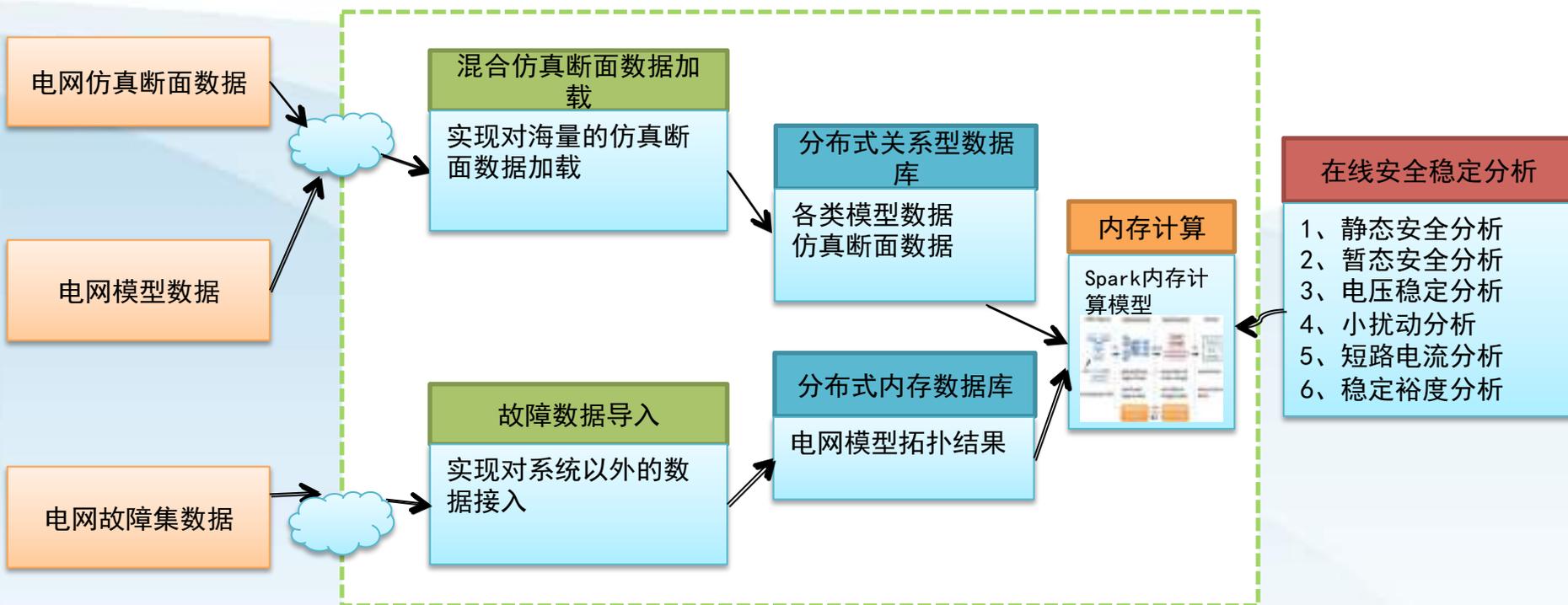


电网实时态势感知

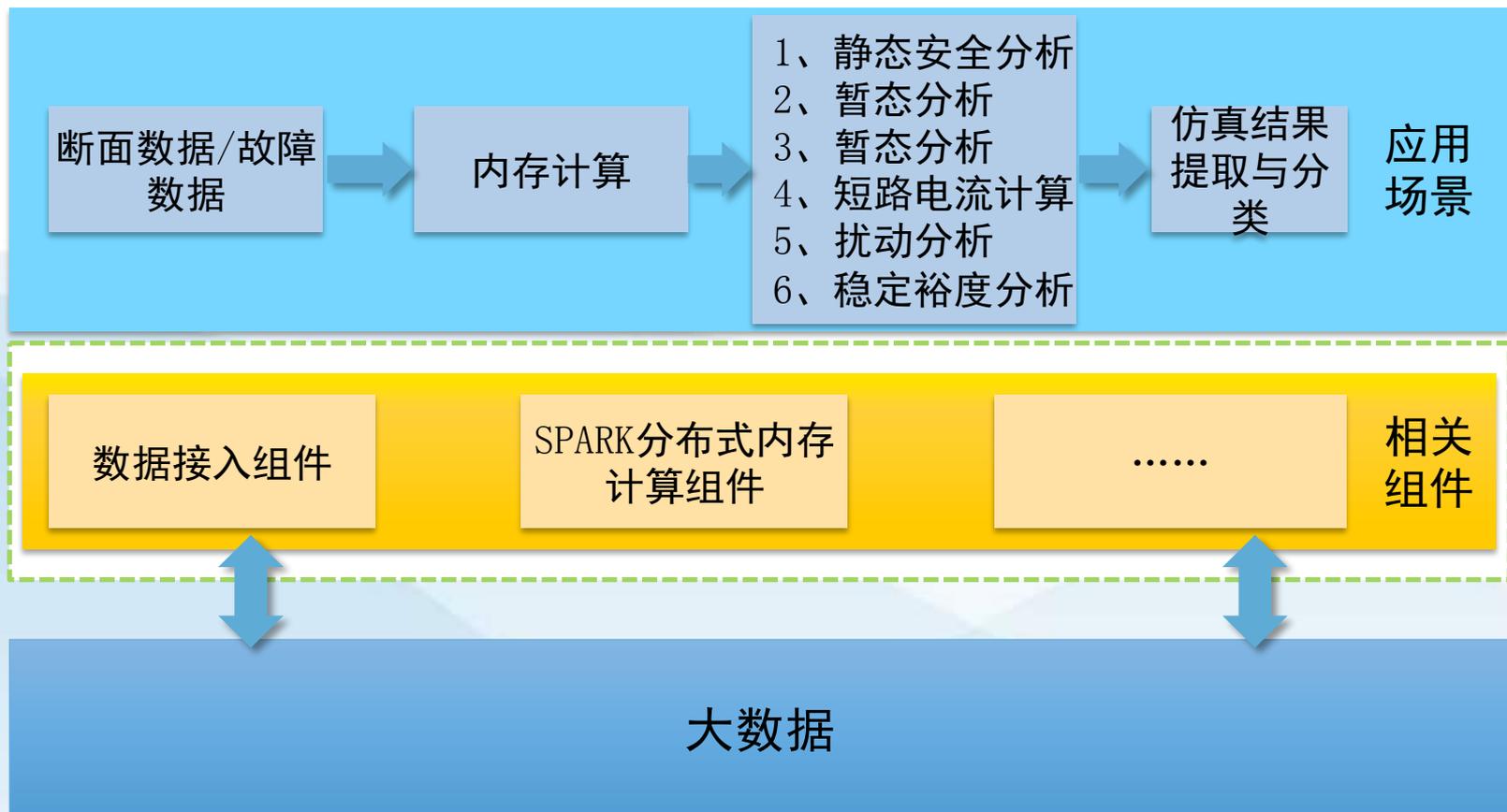


该应用场景的应用特点是数据多并发，计算实时性要求高，并且要求实现关键、高频、异常等信息的定位与推送。

充分利用大数据技术的分布式内存计算能力，实现对于海量断面的混合仿真运算，真正意义上实现大电网的在线安全稳定分析与预警。



混合仿真计算场景



该场景应用特点是计算断面数量多，实时性要求高，单次计算迭代次数多，计算模块之间的依赖性强。

1

互联网与智能电网大数据

2

行业相关应用

3

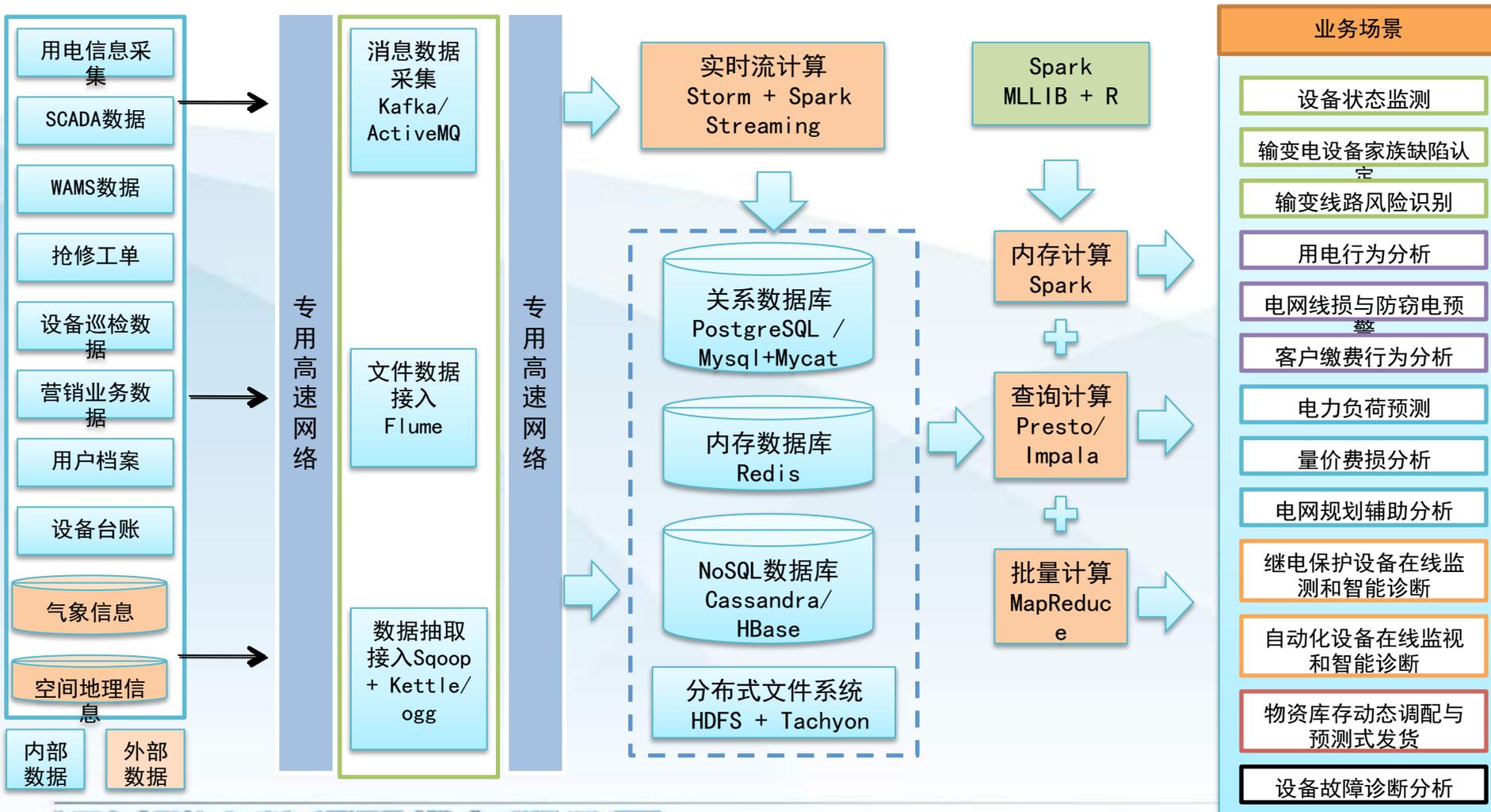
智能电网大数据应用架构

4

技术趋势思考

智能电网大数据应用架构

引入大数据组件24个(含安全与运维3个, 不含基础支撑组件如Yarn、Zookeeper等), 除此之外, 实际应用过程中需要根据业务需求, 进行相应的组件之间的串联研发、开源组件优化、无开源组件而需实现需求的研发。



我们的主要工作

智能电网生产数据存储模型

生产数据的高速写入与实时计算

异构数据源统一访问

智能电网生产数据存储模型

- 数据基本情况

- $\langle \text{DeviceID}, \text{TimeStamp}, \text{Value}, \text{QC} \rangle$

- $3500 \text{万} \times 10 \times 24 \times 4 = 336 \text{亿条/天}$

- $0.1\text{K} \times 336 \text{亿} = 3\text{T/天}$

- 访问模式

- 批量访问

- 以时间为“轴”，访问某一测点一段时间内的数据

- 断面访问

- 以时间节点为截面，访问某一时间点一批测点的数据

基于HBase的数据双写模式（1）

RowKey	偏移1	偏移M	偏移N
...	30	30	30	30	30
...	35	35	35	35	35
...	50	50	50	50	50
...	55	55	55	55	55

RowKey: <指标, 基准时间戳, 测点ID>

注: 指标就是电流、电压等

基于HBase的数据双写模式（2）

RowKey	后缀1	后缀M	后缀N
...	30	30	30	30	30
...	35	35	35	35	35
...	50	50	50	50	50
...	55	55	55	55	55

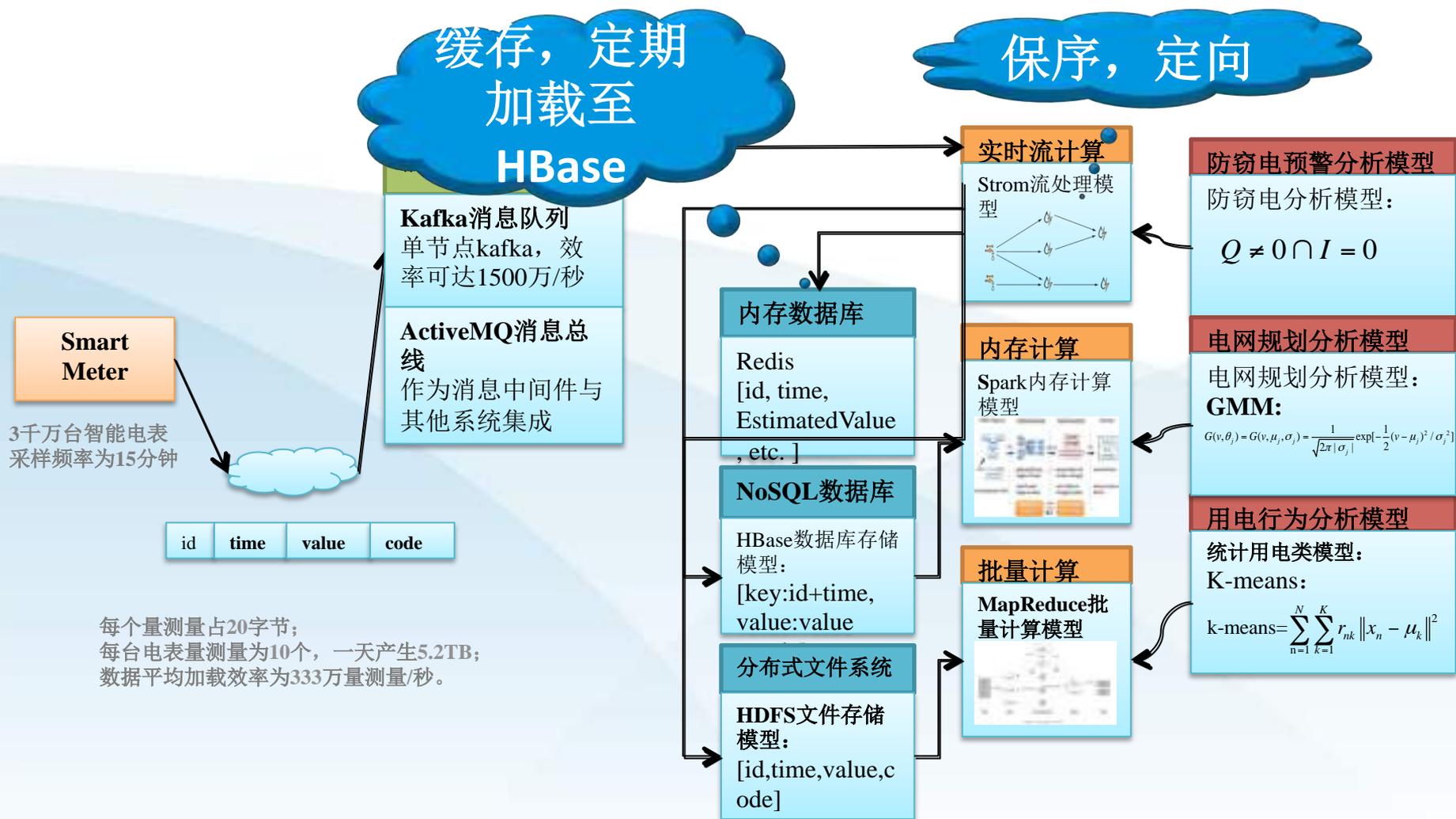
RowKey: <时间戳, 指标, 前缀>

注: 指标就是电流、电压等, 测点ID=前缀. 后缀

小结

- 核心思想：让经常一同访问的数据在物理上存在一起
- 加载时避免热点效应，同时依据region分布欲分割数据
- 效率
 - 加载：8万条/秒/CPU Core/Hard Disk/8G Mem
 - 读取：28万条/秒/CPU Core/Hard Disk/8G Mem
 - 对于标配的机器：
 - CPU：2路6核
 - 硬盘12块，15000转 SAS
 - 内存96G
 - 每秒可加载约100万条记录

生产数据的高速写入与实时计算



异构数据源统一访问

数据管理与展示

数据共享与ETL

异构数据源管理

数据可视化组件

自主开发
Or
定制修改

分布式数据处理

Presto

Spark (Streaming/SQL/R)

Storm

集成开源

异构数据库统一访问中间件

数据存储

各类关系库

ElasticSearch

cassandra

HDFS/HBase/HIVE

万亿级大数据平台

- 数据集成治理
- 高效可靠存储
- 高性能计算处理
- 智能分析挖掘
- 洞察与决策

数据集成与共享

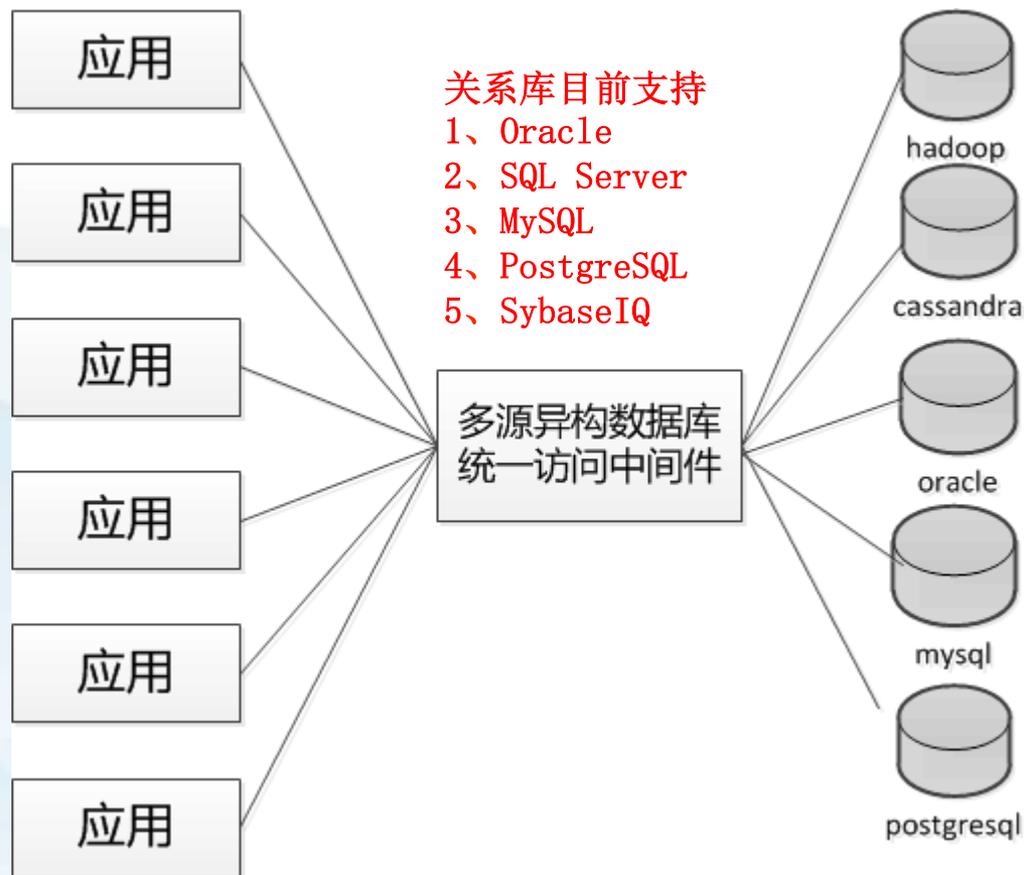
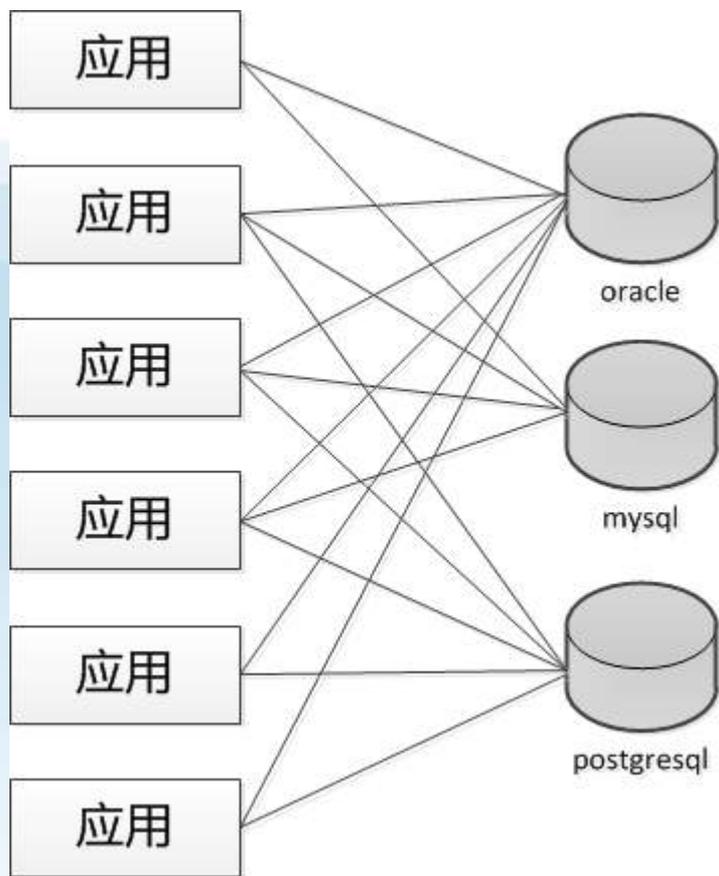
企业数据总线

数据库数据抽取
Sqoop

日志文件汇聚
Flume

分布式消息队列
KAFKA

异构数据源统一访问



1

互联网与智能电网大数据

2

行业相关应用

3

智能电网大数据应用架构

4

技术趋势思考

数据量大到一定程度以后，没有任何数据处理软件可以直接使用，满足企业的大数据处理需求，数据处理架构已从单一产品向整体解决方案演变。

- 1、基础软件必须由纵向扩展向横向扩展转变
- 2、各孤立的软件、系统必须走整合、平台化转变
- 3、数据处理的领域必须覆盖全域（任意类型、任意大小的数据）
- 4、价值为王（帮助客户发现数据价值）



一站式大数据平台

大数据处理时代没有银弹！

特点：

- 全面的数据处理技术
- 开源及兼容性最好的数据处理解决方案
- 支持当前现状向未来的过渡
- 融合企业现有系统
- 灵活的组件选择方式



结构化/半结构化/非结构化共存

面向业务的定制优化

心得：在某一应用下，数据量大到一定程度以后，没有任何一个开源数据组件可以直接使用。

在开源软件的基础上，结合特定的业务应用场景，对组件关键处理流程在源代码级别进行定制优化，以满足业务应用需要。





奉献清洁能源
建设和谐社会



国家电网
STATE GRID

国网电科院江苏瑞中数据股份有限公司
CHINA REALTIME DATABASE CO.,LTD.,SGEPRI

谢谢!