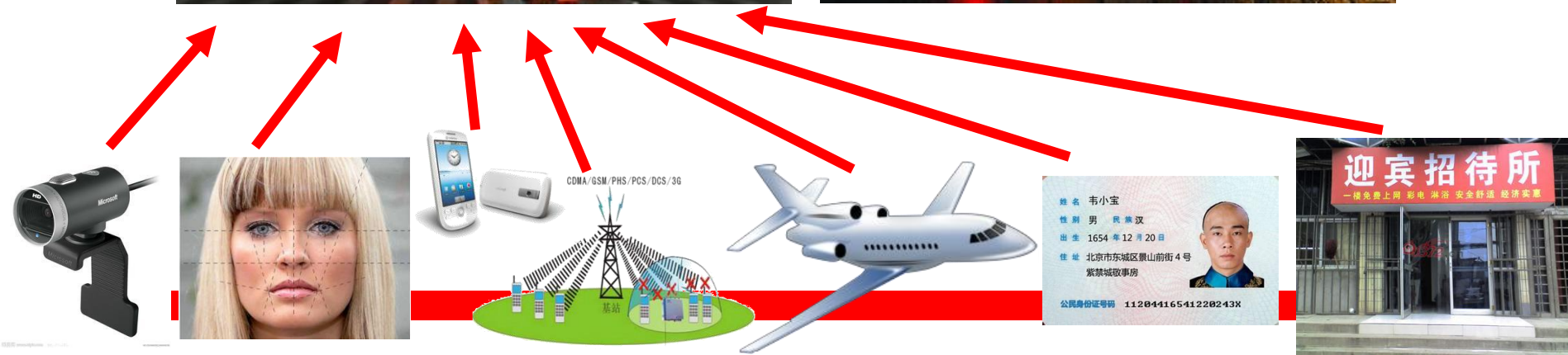


ORACLE®
甲骨文

大数据困境与解决之道

昆明火车站恐怖事件



公安早就很好的利用了大数据：信息化警务创新

信息化引领武汉警务革新 助推破案近八成

2013年05月21日 19:44 来源：中国新闻网

参与互动(0)

0

上海“科技加合作”推进刑侦信息化

作者：吴艺 来源：人民公安报 时间：2012-07-06

字体：大 中 小

据上海市公安局介绍，上海公安机关坚持“科技加合作”的理念与实践，不断强化高科技信息平台建设、深化区域警务合作，显著提升了刑侦信息化水平和打击犯罪能力。截至目前，上海在全市命案、“两抢”案件立案数呈逐年下降趋势的同时，命案破案率始终保持在97%左右，“两抢”案件破案率始终保持在70%以上。

上海市副市长、公安局局长张学兵表示：“实践证明，科技加合作是实现刑侦信息化的重要途径，是破大案要案的主要手段，是有效驾驭社会治安局势的必备法宝。”

上海公安机关大力加强重点案件的现场勘验和物证提取，目前，警方的指纹自动识别系统已累计建库520余万份、生物物证数据库累计存档74万份。

此外，上海公安机关与苏浙皖等地警方正式签订了合作协议，建立了跨区域线索交流、信息警务协作、重大案件协作侦办、重大案（事）件协查快速响应、协作追逃等合作机制。

和作案车辆。先后抓获犯罪嫌疑人金某、詹某等6人，收缴仿64式手枪2支、长柄猎枪1支、钢珠枪1支，手枪子弹2发，猎枪子弹5发。在24小时内成功侦破这起涉枪寻衅滋事案。（完）

公安大数据特点分析

- 数据非常杂
 - 公安自有数据+社会数据+
 - 网络数据+.....
- 数据非常多
 - 亿条规模常见，上百亿条不稀奇
- 数据非常散
 - 数据分散在各个业务系统
 - 缺少完整的全数据平台
- 数据高聚合
 - 人+车+案



公安业务数据处理的问题

业务办理和精准查询

多警种联合
精确查询

模糊查询
多线索碰撞
联合查询
数据量
任...
多...
分析
模式识别
犯罪挖掘

业务办理和精准查询

业务办理和精准查询

安全重点
单位信息
资源库

人口基本
信息资源
库

出入境人
员资源库

吸毒人员
库

银行, 税务
社保等其它
外部信息

网络实名
制库

网聊和社
交媒体记
录库

手机短信/
通话库

娱乐场所
从业人员
库

宾旅馆住
宿人员库

机动车/驾
驶人信息
库

警员基本
信息资源
库

在逃人员
信息资源
库

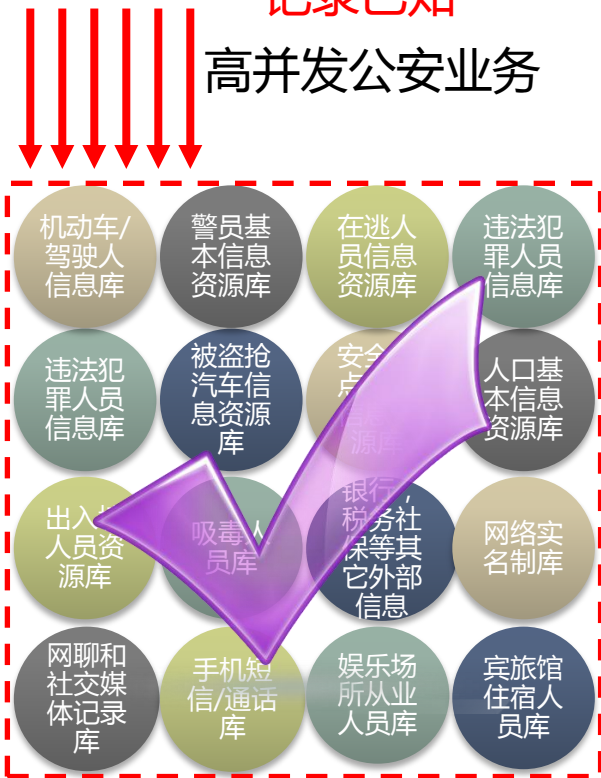
违法犯罪
人员信息
库

被盗抢汽
车信息资
源库

公安业务数据处理的场景

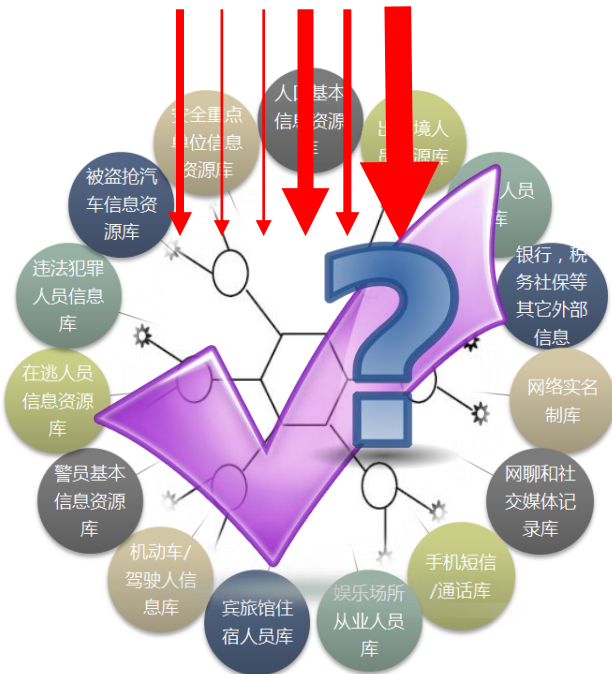
记录已知

高并发公安业务



在已知中发现规律

高效精确查询 & OLAP

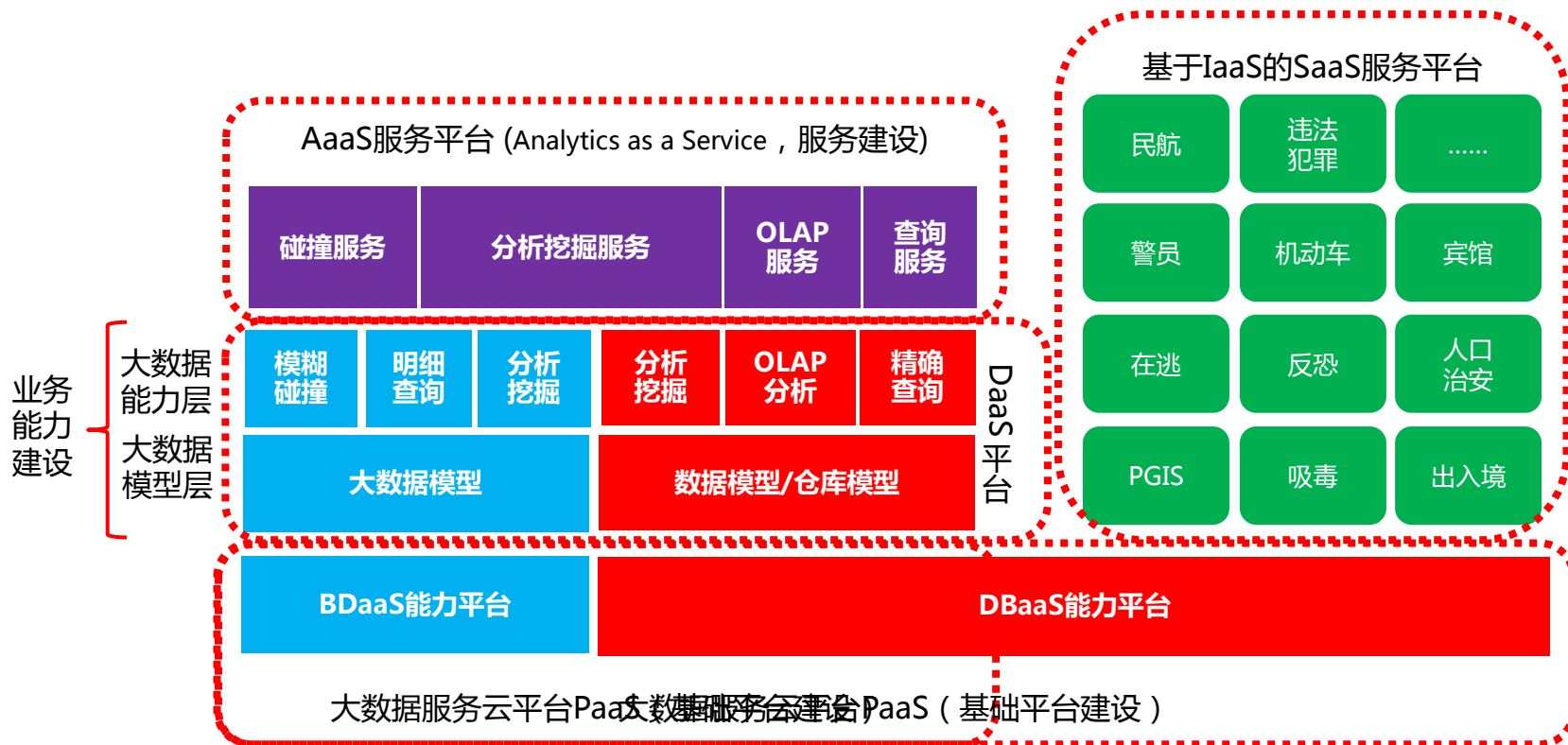


在未知中发现未知

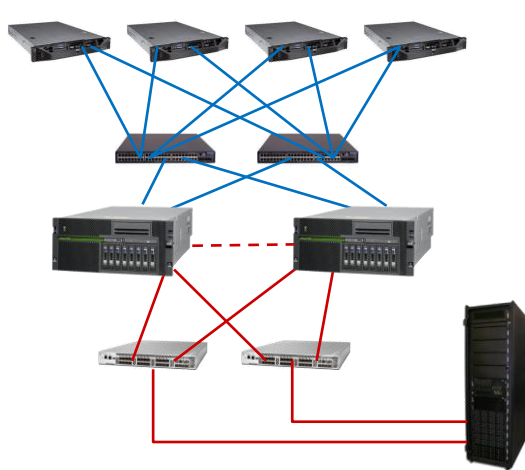
高速模糊碰撞



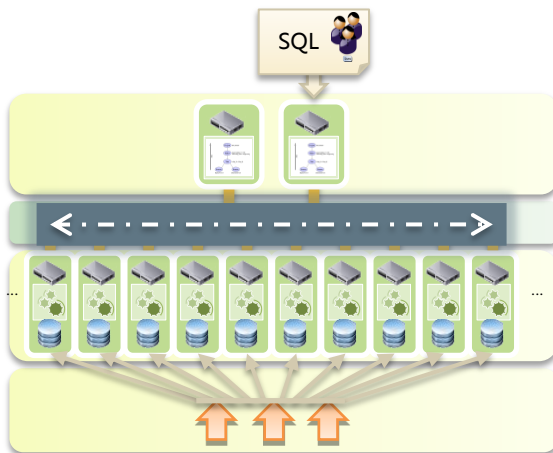
全面的公安大数据云平台



传统的数据平台在大数据时代面临挑战



OR



传统的集中化
RDBMS或者
MPP数据库？



- ✓ 面向结构化数据，固定模式，高密度。
- ✓ 海量数据的处理能力与伸缩性不足。
- ✓ 为传统联机应用与数据分析需求而优化设计。

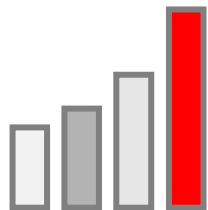
多样、模型不定、松散

PB级别、快速膨胀

探索海量低密度数据价值

传统数据平台的演进以满足新的大数据处理需求

VOLUME



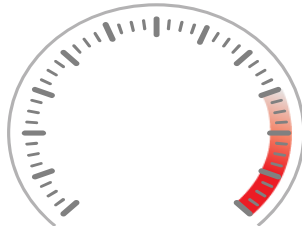
从传统数据库的TB到大数
据时代PB级

VARIETY



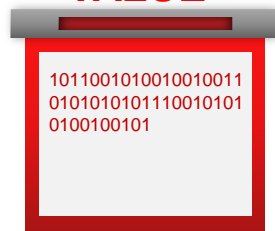
结构化数据、文档、视频、传
感设备、社交媒体

VELOCITY



通信记录、视频监控、
传感数据、微博等

VALUE



松散、低密度、价值隐
蔽但又重要

“分而治之”

传统架构的革命
大规模分布式存储与计算

“灵活多变”

基于普通server的非固定模式数
据的实时或批量存取
(NoSQL\File)

“实时计算”

高并发、低延迟的流式计
算 (Streaming)

“深度挖掘”

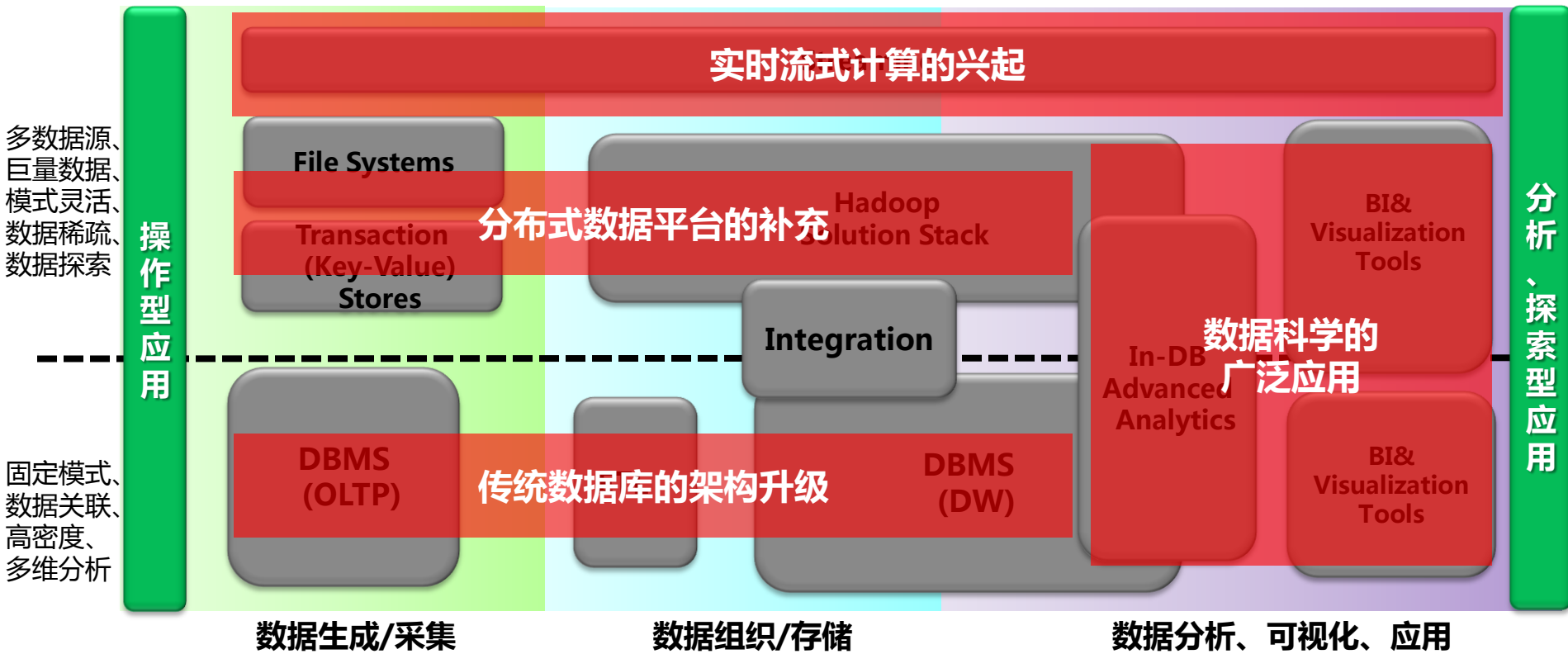
洞察、预测
探索性发现、数据可视化

大数据时代数据平台的进化之四个方面：

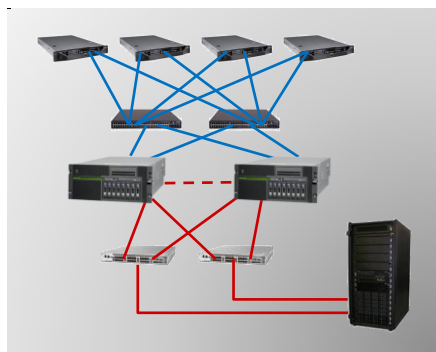
- ① 传统数据库的架构升级
- ② 分布式数据平台的补充

- ③ 实时流式计算的兴起
- ④ 数据科学的广泛应用

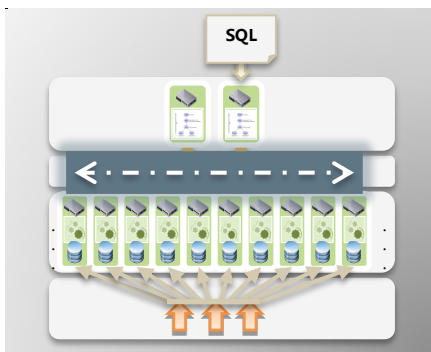
Oracle建议：一个更加完整的平台以满足公安大数据处理需要



1. 传统数据库的架构升级 – 以Oracle为例



传统集中式RDBMS



传统MPP数据仓库



- **Exadata**
- Oracle Database
- In-Memory Option
- In-DB Analytics Option
-



更优秀的实现架构



更简单可控的技术



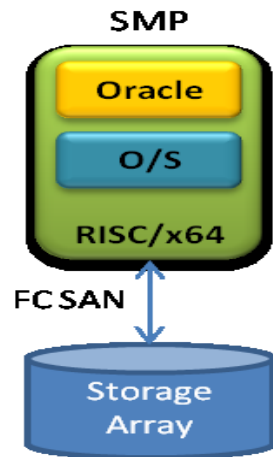
更低的总体拥有成本

ORACLE方法:

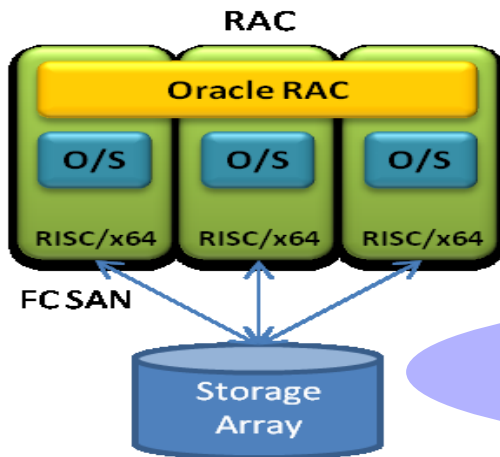
软硬结合以实现内部架构的革命与持续创新，使得数据库具备超越传统的数据处理能力：

- ✓ 极限的性能与存储密度
- ✓ 面向海量数据的水平扩展
- ✓ 保持企业级的高可用性
- ✓ 易于集成与管理
- ✓ 降低TCO

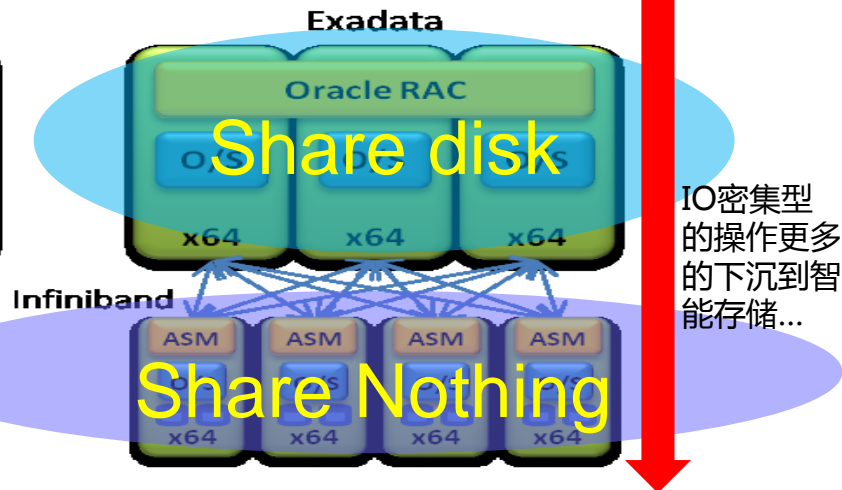
1. 传统数据库的架构升级 – Exadata：不只是软硬件的简单集成



- 混合列压缩
- 内存计算
- MPP存储



- X86服务器
- Infini Band
- 数据库云

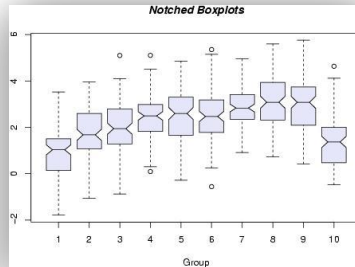
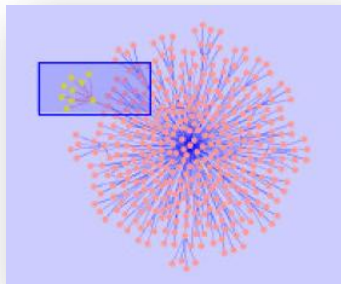


- 数据库一体机/仓库一体机
- 智能扫描/存储索引
- 计算能力下沉和水平扩展

■ **Exadata**提供一种混合式的数据库架构，吸取了两种架构的优点，既可以满足高并发高可用的特点，又可以满足批量数据处理的需求；**同时极大的提高的数据处理的性能与存储密度。**

1. 传统数据库的架构升级 – Oracle数据库具有强大的分析能力

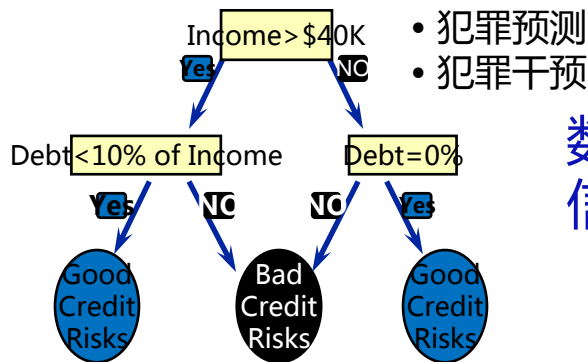
Oracle Advanced Analytics



统计
数据挖掘
文本
图形
空间
语义

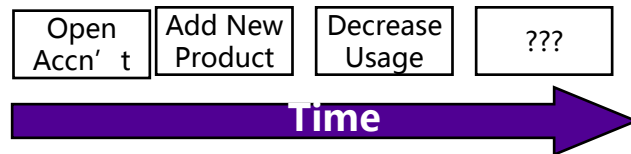
1. 传统数据库的架构升级 – Oracle数据库强大的数据挖掘能力

决策树 Decision Trees



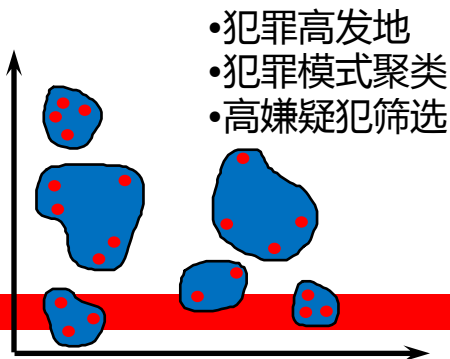
数据挖掘提供更好
信息、更有价值的
洞察和预测

序列分析 Sequence Analysis



- 犯罪模式识别
- 犯罪/案件生命周期
- 犯罪时间序列分析

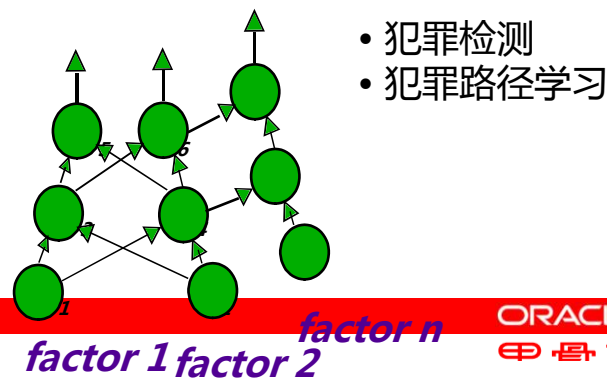
聚类分析 Clustering



关联分析 Association

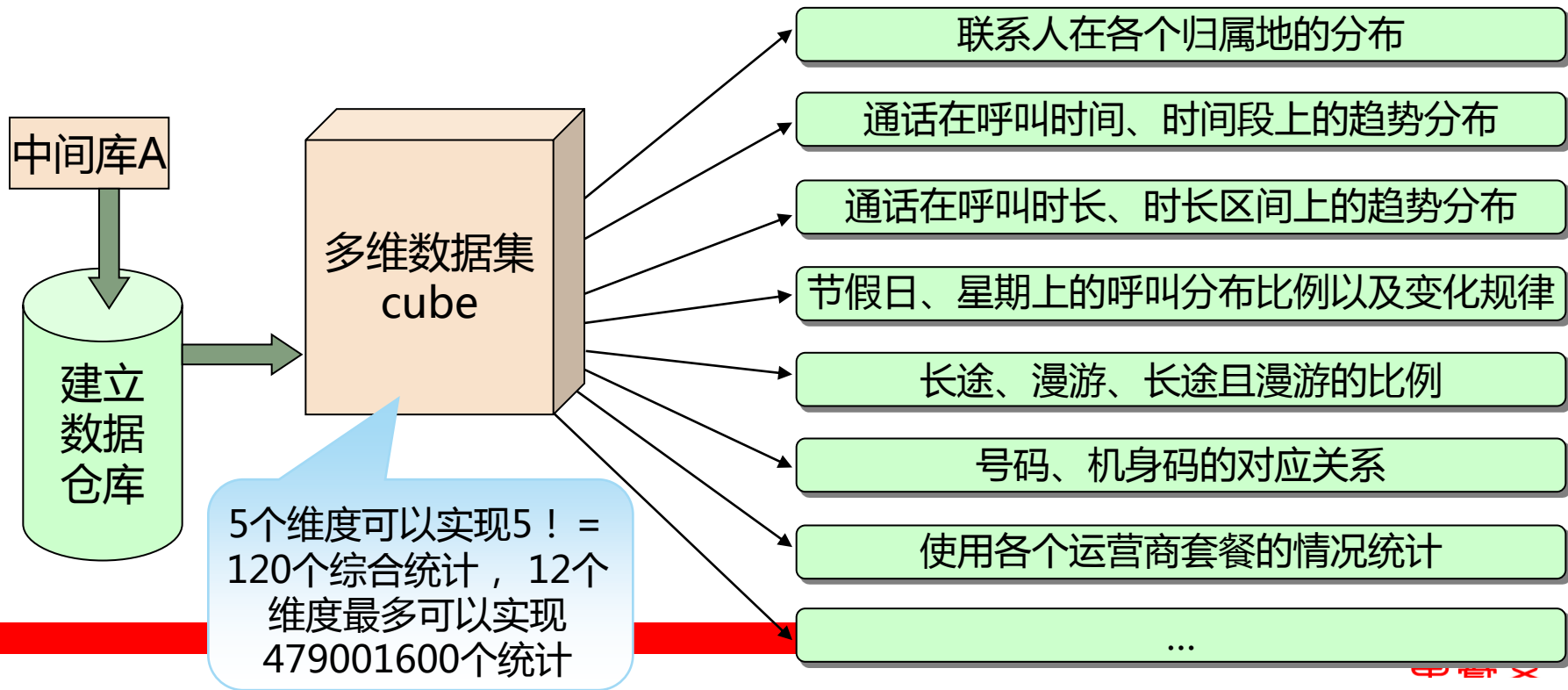


神经网络 Neural Networks

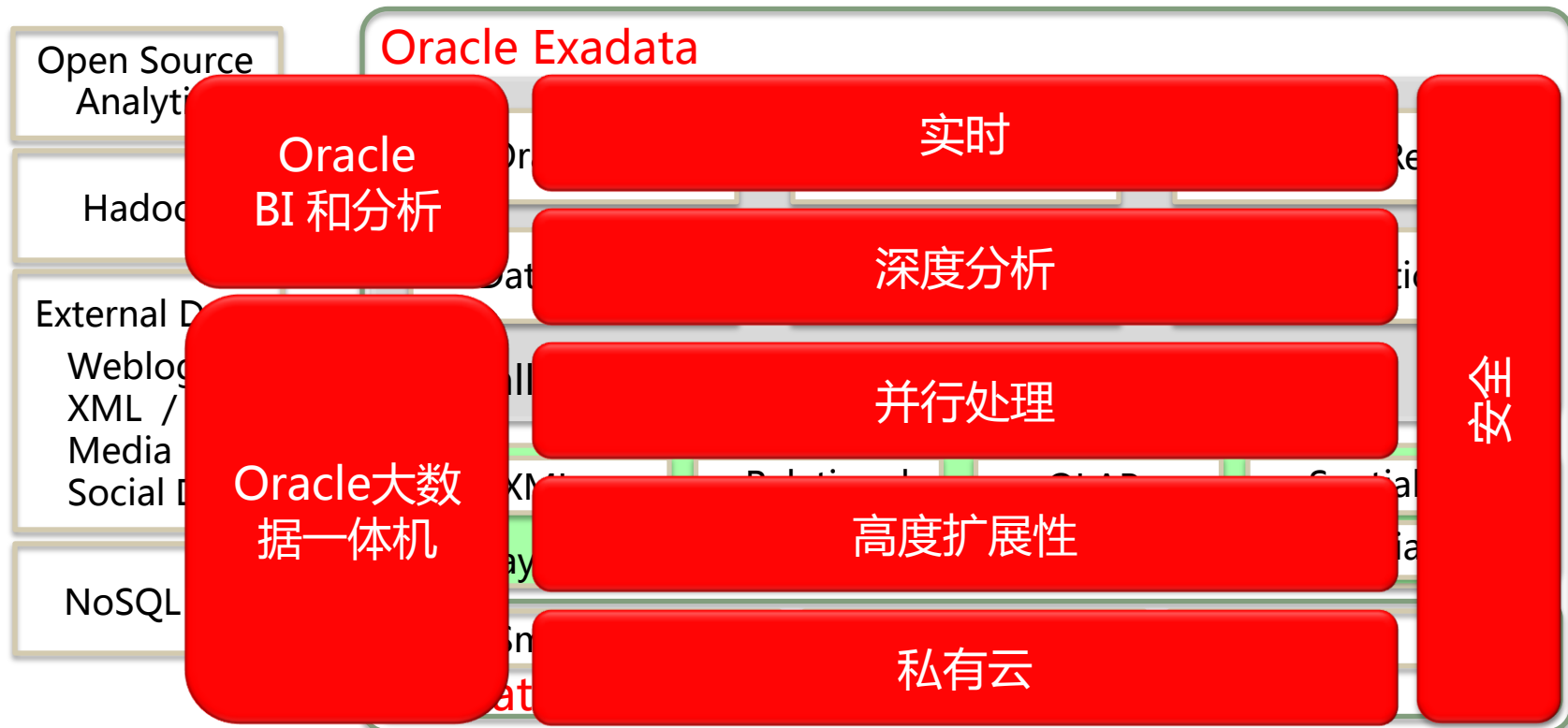


1. 传统数据库的架构升级 – 数据仓库和OLAP多维分析

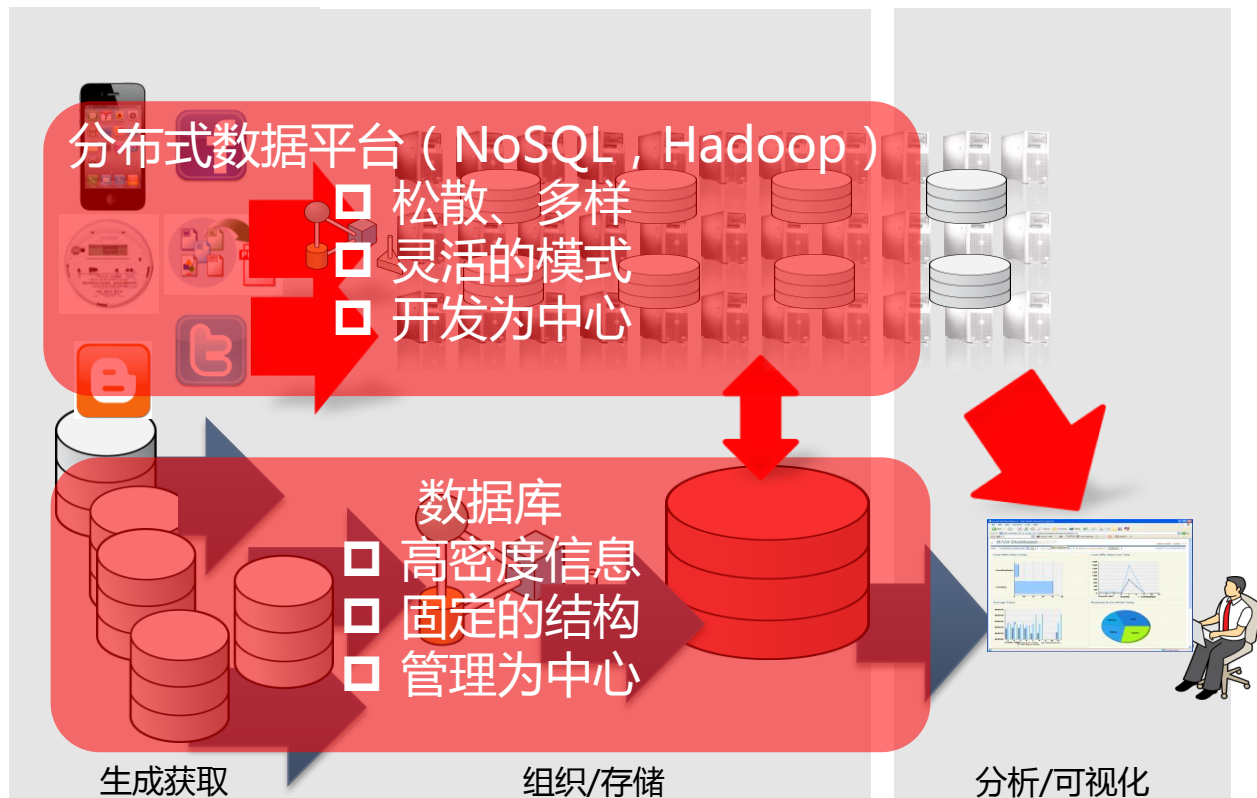
建立数据仓库、多维数据集cube，通过在各个属性维度上聚合统计，利用OLAP快速直观的图表展现来分析A库中犯罪嫌疑人各项详细特征。



1. 传统数据库的架构升级 – Exadata增强数据分析挖掘能力



2. 分布式数据平台的补充 – 满足新型的数据处理需要



- ✓ 新的分布式数据平台作为**有效的补充**，而不是简单的替代
- ✓ 有时无绝对的分割标准，两个平台也会存在**相互的重叠与技术渗透**。

2. 分布式数据平台的补充 – 大数据时代公安如虎添翼

核心思想：分布式、经济、数据就地计算

分布式存储
分布式计算



Hadoop：分布式，大数据集，延时，离线计算

分布式灵活存储
分布式实时简单数据存取



NoSQL：分布式，小数据集，灵活，实时，快速读写

2. 分布式数据平台的补充 – Hadoop的能力和在公安的定位

- 明确的需求驱动
- 支持条件查询与统计，索引优化
- 复杂数据关系应用
- 支持在线、低延迟、交互分析
- 超大规模数据量计算困难
- 数据库为中心
- 通过SQL交互
- 结构化数据为主、静态模式
- 通常在TB级，高密度
- 传统RDBMS
- 一定的扩展性与并行能力
- 相关工具较丰富



- 数据价值往往不确定
- 全量数据应用为主
- 简单的数据关系应用
- 离线应用为主
- 支持巨量数据计算水平扩展
- 开发为中心
- 通过API/脚本/类SQL交互
- 非结构化为主、动态模式
- 通常在PB-EB级，稀疏
- Hadoop生态系统
- 分布式、极度的扩展性与并行能力
- 相关工具完善中

✓ 数据库平台

- ✓ 高价值、富模型、强关系数据的存储
- ✓ 提供精准查询、交互式分析和报表
- ✓ 提供多维分析和挖掘

✓ Hadoop平台

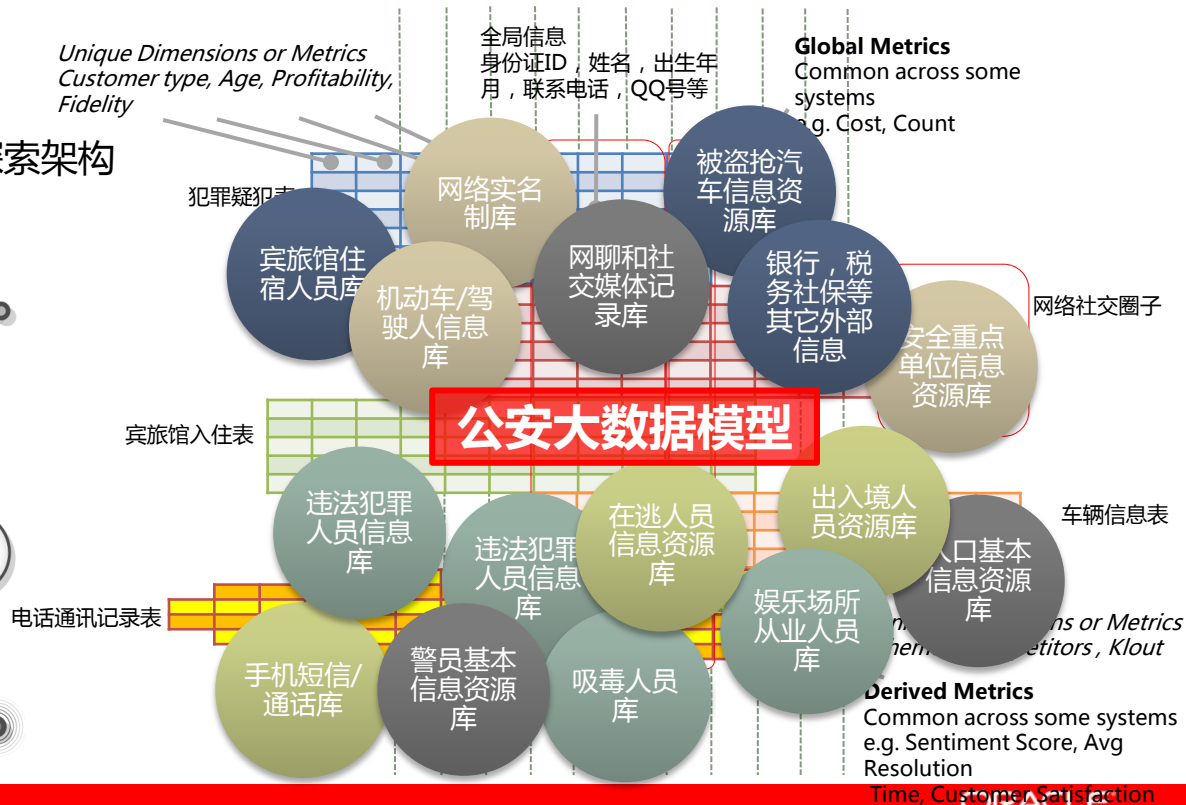
- ✓ 公安多源松散数据的建模和存储
- ✓ 并行计算平台，承担巨量计算工作
- ✓ 高效快速的碰撞业务
- ✓ 基于大数据的分析、挖掘和探索

2. 分布式数据平台的补充 – 大数据时代的公安大数据模型特性

灵活的、非预定义的、非固定模型的、垂直水平双向扩展的

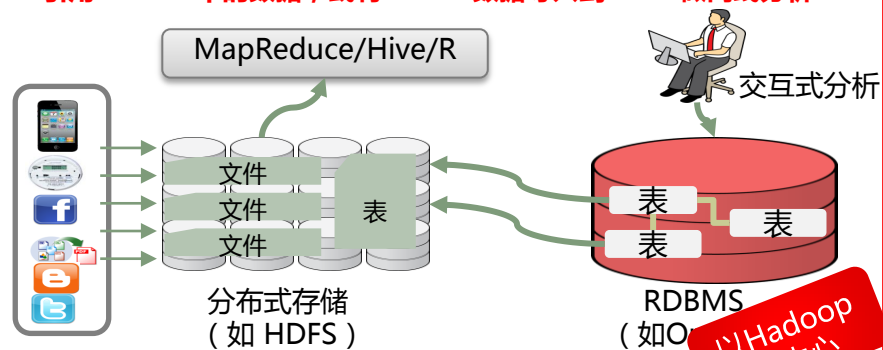
Table-free

=
不需要过度架构、自适应、灵活的数据探索架构

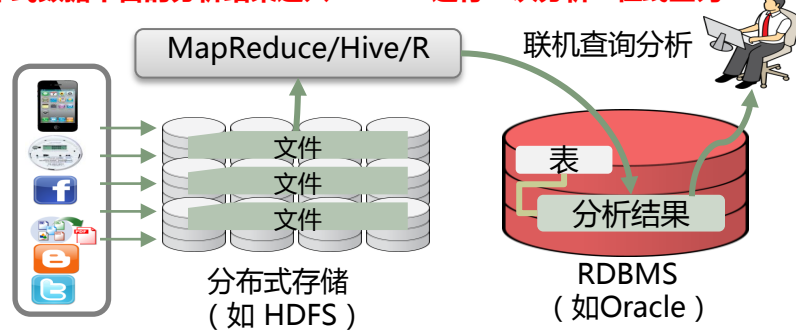


2. 分布式数据平台的补充 – 两种平台之间通常会互相集成

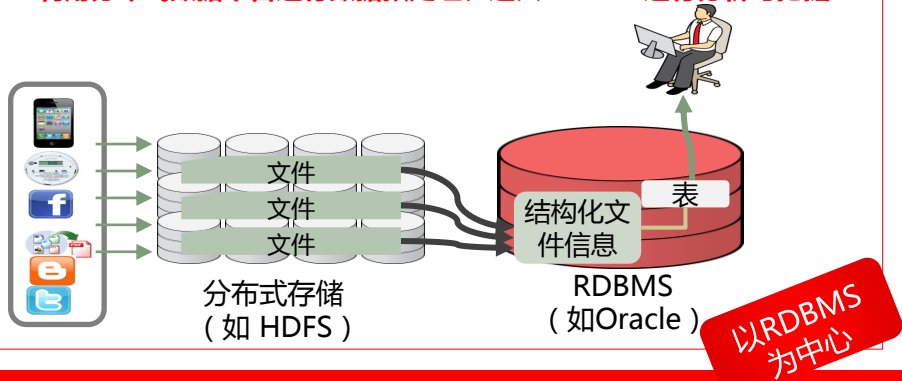
引用RDBMS中的数据；或将RDBMS数据导入到HDFS做离线分析



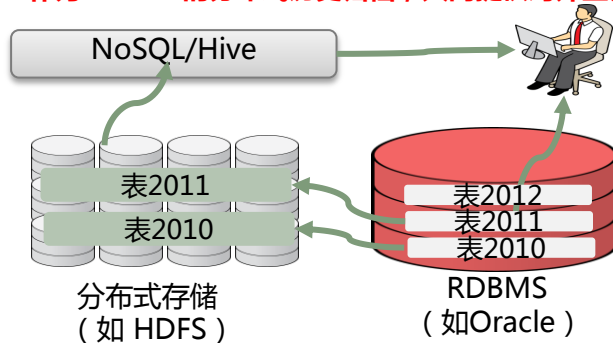
分布式数据平台的分析结果进入RDBMS进行二次分析&在线查询



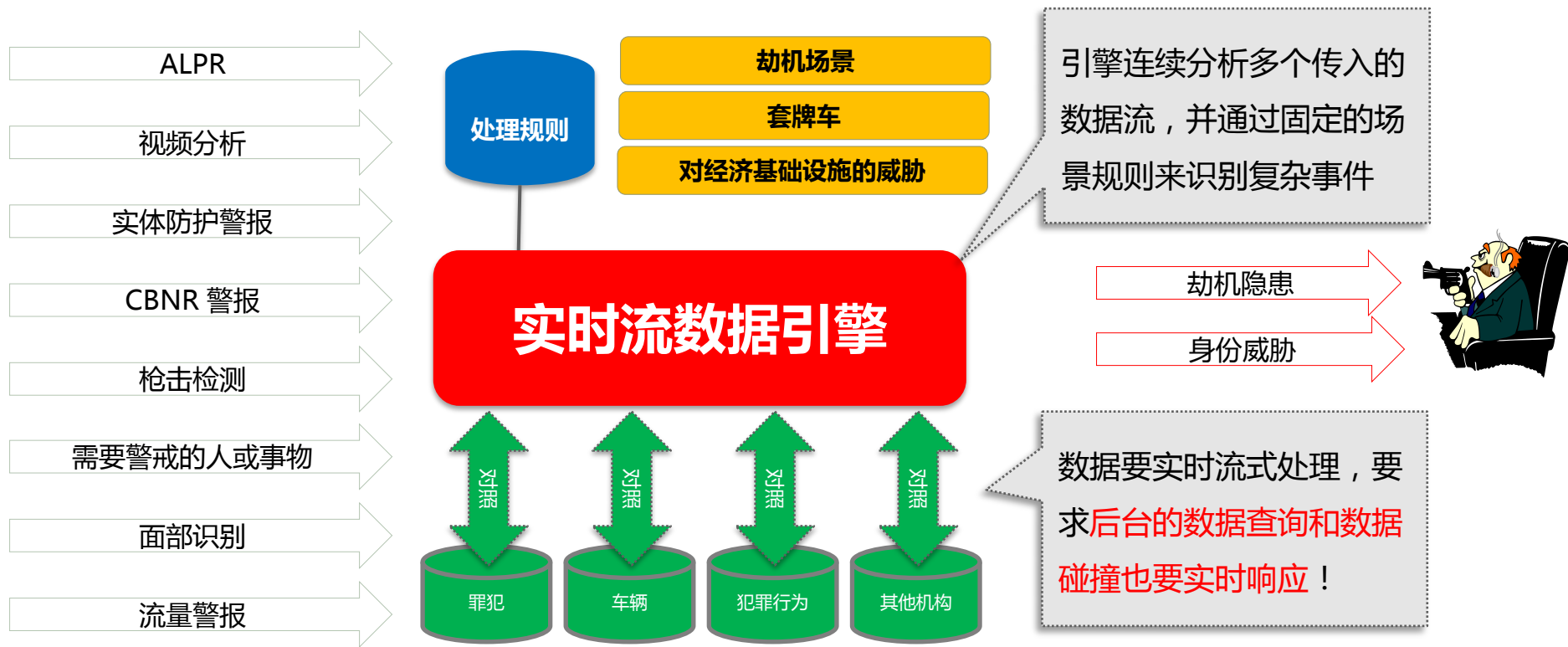
利用分布式数据平台进行数据预处理、进入RDBMS进行分析与挖掘



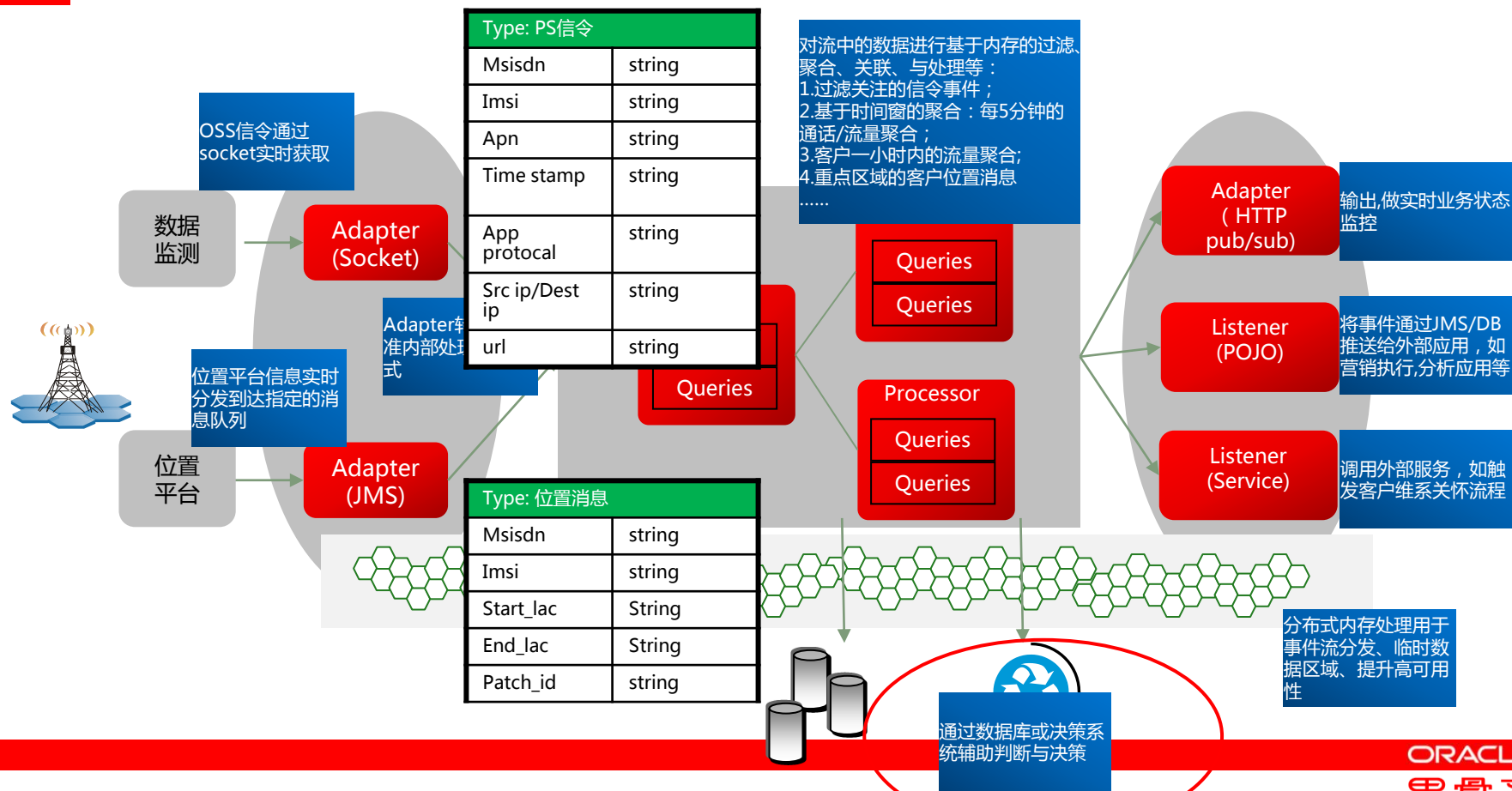
HDFS作为RDBMS的分布式历史归档，共同提供对外查询



3. 实时流式计算的兴起 – 实时公共安全问题监控、识别与处理



3. 实时流式计算的兴起 – 海量实时事件处理能力



3. 实时流式计算的兴起 – 机动车管理

实时数据匹配 (Complex Event Processing)

受怀疑车辆

川P73617 京GF9473
京W38240 苏TW8347
津X97936 京RV4482

朝阳区工体北路
盈科中心被人炸了！
可疑分子坐了
蓝色本田雅阁
离开现场



实时数据匹配
Complex Event Processing

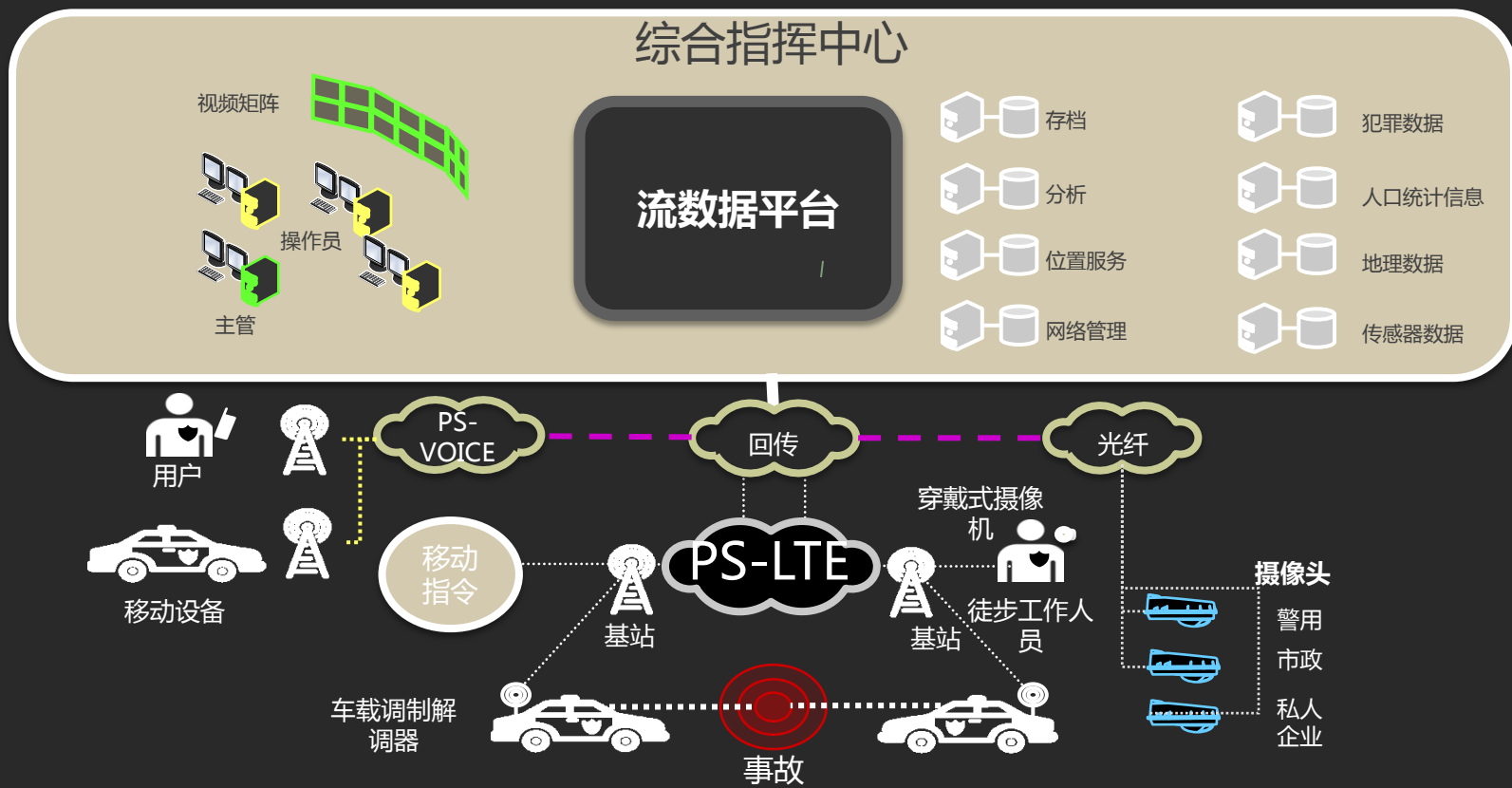
自动车牌识别系统

京GK7402 京E64965 津JD9349
京B54720 川P73617 京XL4825
苏VD1234 京GF9473 京AZ9999
京W38240 京K52505 京F93375
京GH4921 川M13840 京C97451
京A68888 京EF4154 京B73954
津X97936 京G76363 苏TW8347
京HK9012 京F35623 京RV4482

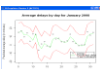
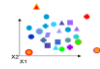
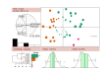
交管车辆纪录

所有识别到
车辆的一确
纪录

3. 实时流式计算的兴起 – 案例



4. 数据科学的广泛应用 – 帮助回答更高级的业务问题



简单统计

即席查询

多位分析

统计分析

预测分析

图片视频文档分析

数据挖掘

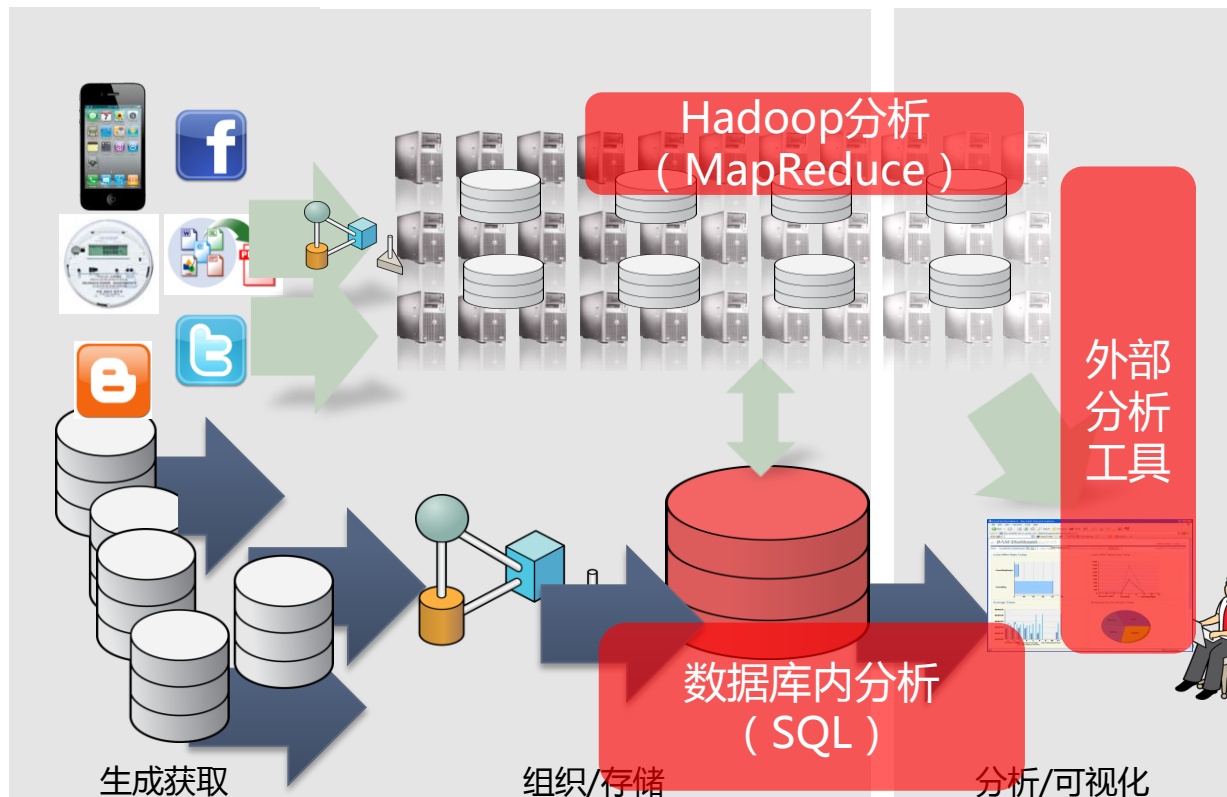
文本/语义分析

.....

- 人员活动轨迹分析
- 机动车轨迹分析
- 监控视频识别抽取
- 案件分类特征研究
- 情报信息关联分析
- 犯罪模式与趋势
- 110警情预测
- 社交信息与舆情监测
-



4. 数据科学的广泛应用 – 数据分析在哪里进行



MapReduce Analytics :

- ✓ 分布式并行
- ✓ DIY模式
- ✓ 不是所有分析都适合

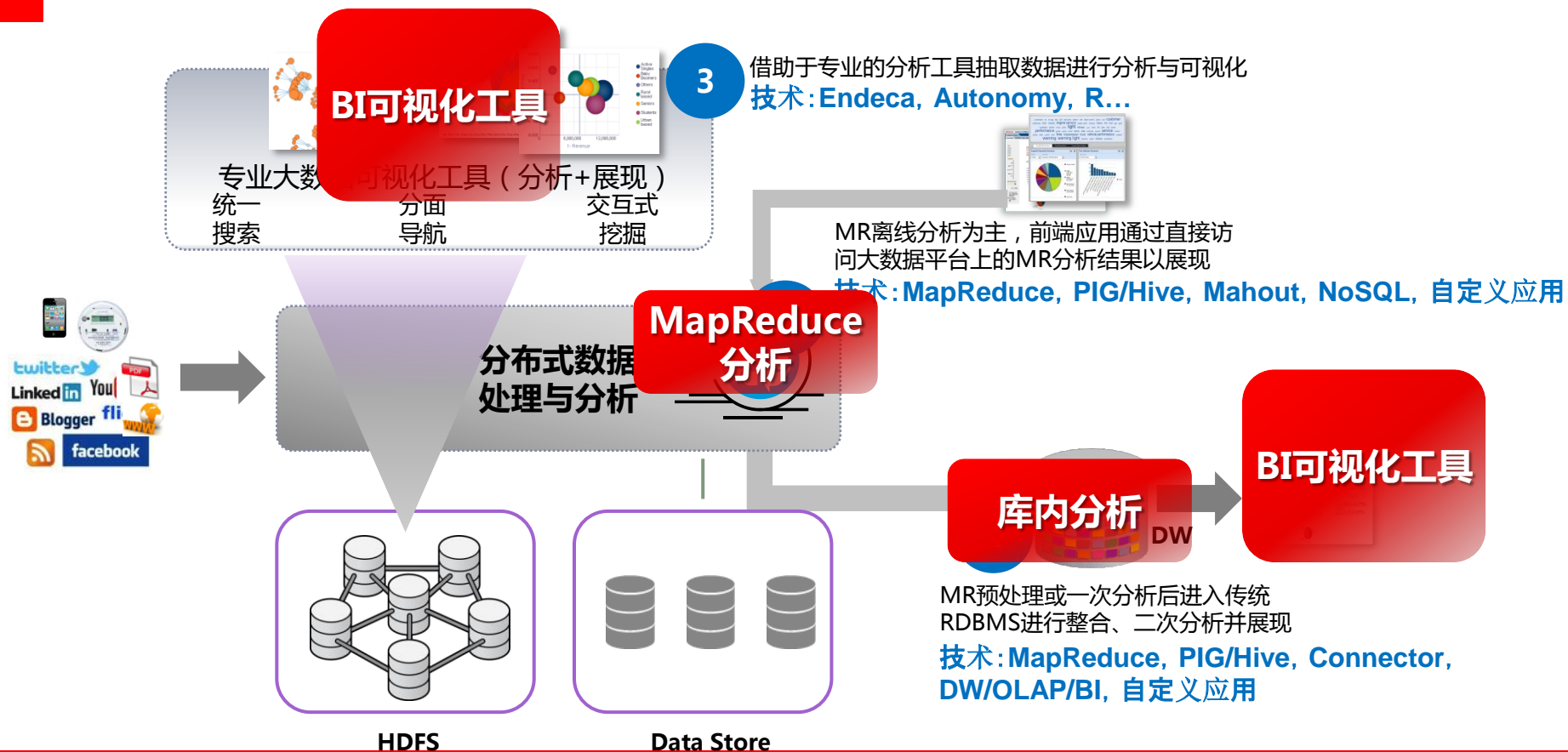
In-Database Analytics:

- ✓ DW提供的库内分析能力
- ✓ 充分利用数据库能力
- ✓ 基于SQL

BI & Visualization Tools:

- ✓ 商业智能模型&工具
- ✓ 灵活性
- ✓ 受限于数据规模

4. 数据科学的广泛应用 – 数据分析在哪里进行



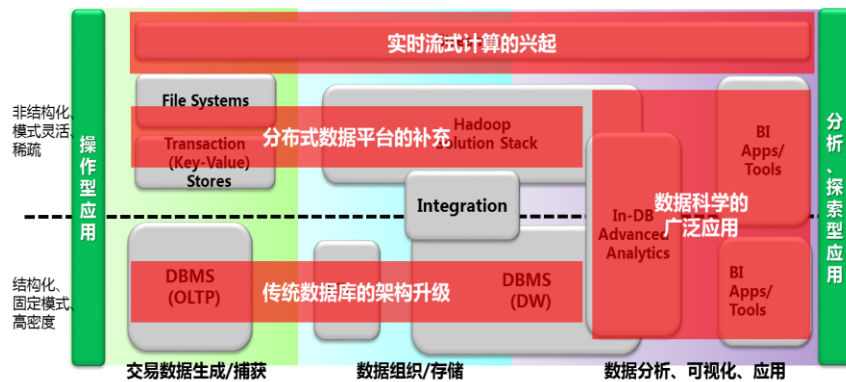
总结 - 大数据应用的IT平台整体架构

挑战

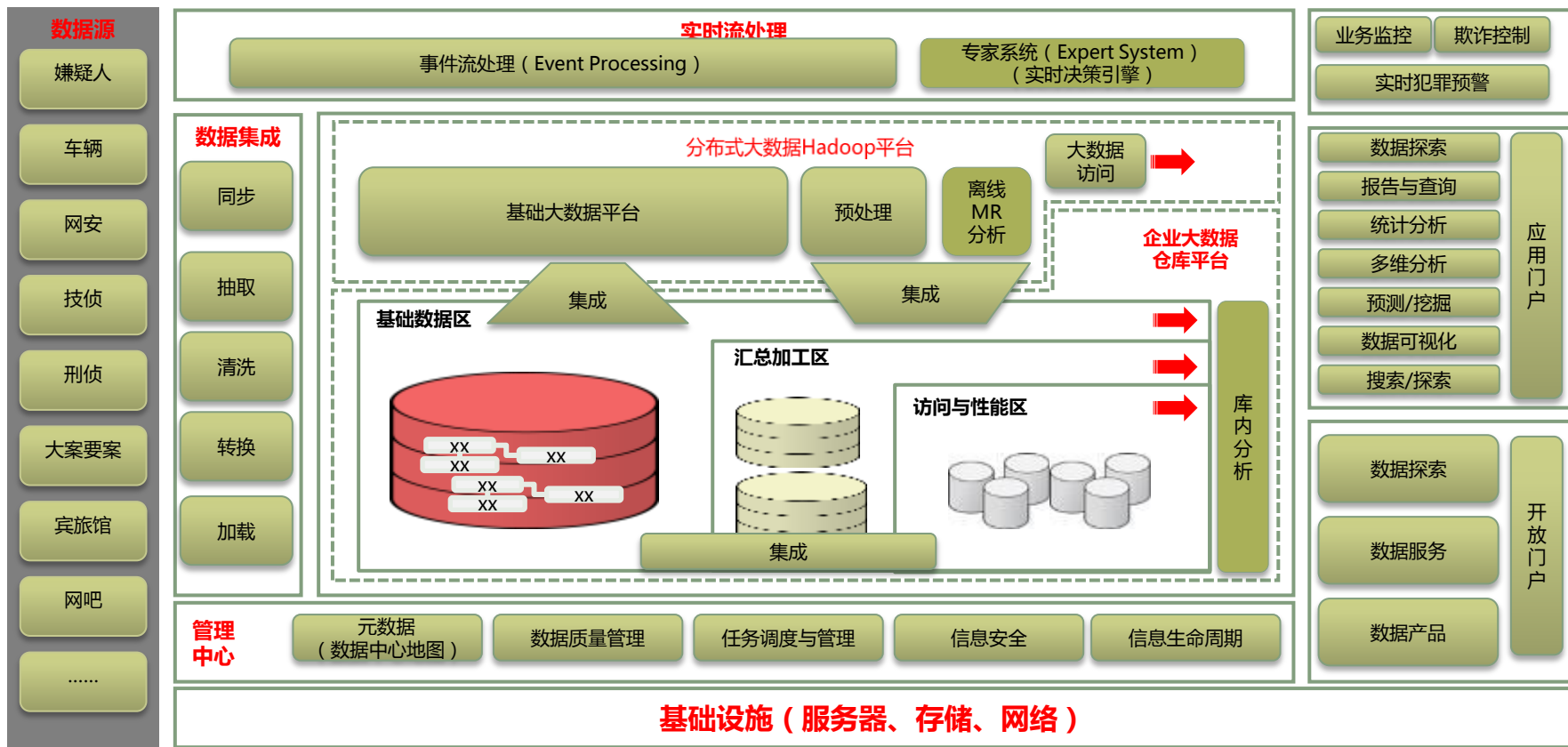
- 海量的数据，极高的性能要求
- 数据多样性与复杂性
- 快速产生，实时事件的需求
- 低密度、价值不明显

方法

- ① 传统数据库的架构升级 – 软硬一体、深度优化
- ② 分布式数据平台的补充 – Hadoop、NoSQL、M/R
- ③ 实时流计算的兴起 – Storm、CEP、内存分析
- ④ 数据科学的广泛应用 – 高级分析、R语言、可视化工具



公安大数据平台总体架构



大数据时代，Oracle的解决方案 – 完整、集成、企业级

多数据源、
巨量数据、
模式灵活、
数据稀疏、
数据探索

操作型应用

固定模式、
数据关联、
高密度、
多维分析

分析、探索型应用

Storm

OEP

Oracle Big Data Appliance



Oracle
Endeca
Info
Discovery

Big Data Connector

Text

Oracle Exadata



Oracle Exalytics



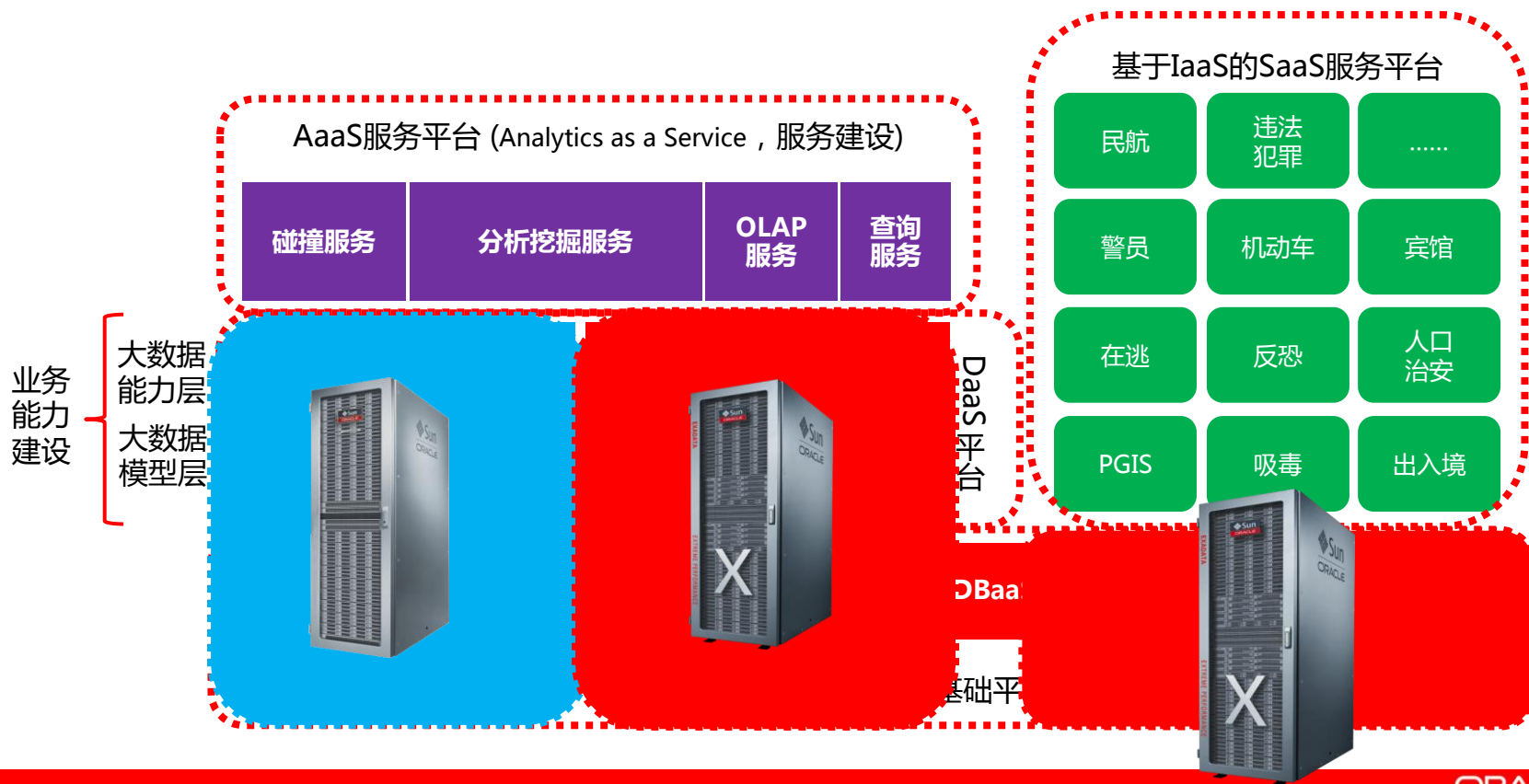
交易数据生成与捕获

数据组织与存储

数据分析、可视化、应用

Enterprise Manager (从磁盘到以应用的端到端运维管理)

Oracle助力全面的公安大数据云平台



大道至简---软硬一体极限性能的Oracle大数据平台

加快端到端的解决方案上市时间和降低风险，简化IT，聚焦业务

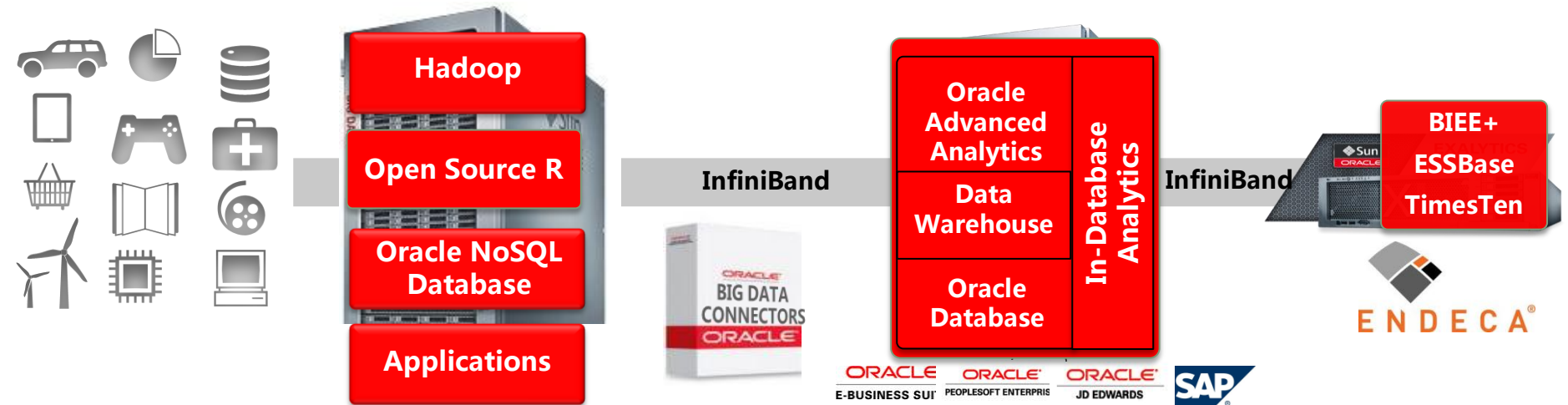
流

获取

组织

分析

决策



Oracle是数据库和信息管理行业的领导者。
Oracle提供的所有组件，你需要从你的大数据平台得到真正的结果

Big Data Appliance X4-2 -- 稳定可靠的生产环境就绪大数据平台

- 18 Sun X4-2L PC服务器
- 每台2CPU*8核，总288核
- 每台64 GB内存，总1152G
- 每台12块4TB SAS盘，总864T
- 3台40Gb InfiniBand交换机
- Oracle Linux / Oracle JDK
- Cloudera Hadoop Distribution
(HDFS/YARN/MR/Hbase/Hive/Impala/Search/BDR/Sentry)
- Cloudera Manager EE
- Oracle NoSQL Database
- Open-source R distribution
- Oracle Enterprise Management
- Oracle Big Data Connector



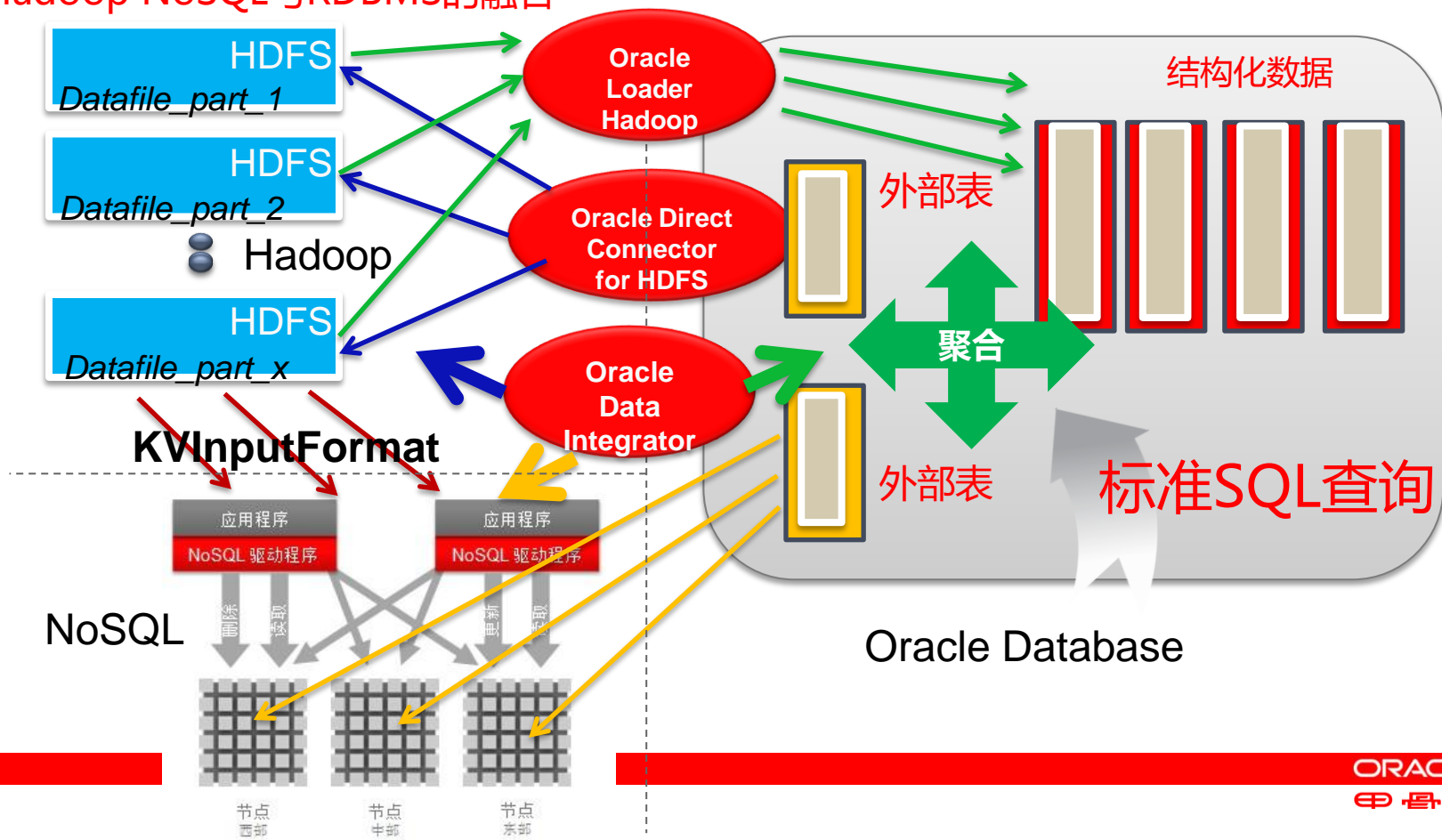
- **软硬一体化**集成设计，规避开源自建高风险，性价比高
- **全栈调优**、最佳实践、性能卓越，性能远超出自建系统
- **全开源全开放**Hadoop/NoSQL软件
- **完整的**非结构化数据**存储/计算/管理/分析/挖掘/整合**的平台
- **快速上线部署**，缩短项目周期
- **全线商业支持**，包括开源软件Hadoop和NoSQL的所有系统软件，硬件和网络
- 支持数据**增量同步**到Hadoop，支持**标准SQL**访问Hadoop，支持高速数据加载
- 支持**非结构化数据即时分析**
- 绿色环保，高密度存储，占地面积小

Oracle Big Data Appliance 提供更多能力



Big Data Connector

Hadoop NoSQL与RDBMS的融合



Exadata是Oracle数据库的软硬件一体化平台

完全为Oracle 数据库优化的服务器 & 存储平台:

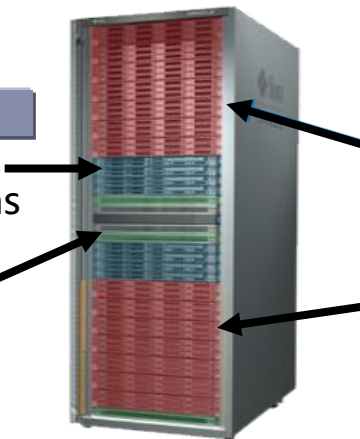
- ❑ 用来解决传统“多用途”服务器的I/O限制
- ❑ 利用通用组件，提供更低的总拥有成本（TCO）
- ❑ 通过预装和配置，提供“开箱即用”的极限性能
- ❑ 自动的存储和运行管理，降低营运成本并能始终提供优越的性能

RAC Database Server Grid

- 8 个高性价比服务器
- 每台机器配2个 Intel 12核 Xeons Processors

InfiniBand Network

- 40 Gb/秒 Infiniband 链路



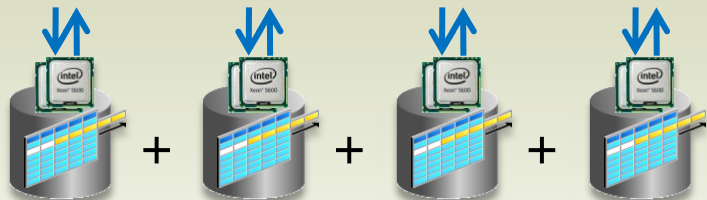
Exadata Storage Server Grid

- 14 个高性价比存储机
- 200 TB 高速 磁盘存储
或
672TB 高容量磁盘存储
- 44.8TB 闪存存储!

技术创新——Exadata智能存储服务器软件

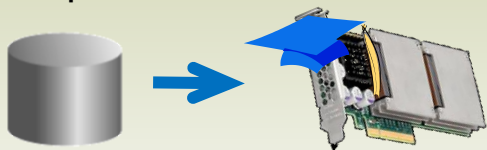
- 智能存储

- Scale-out InfiniBand storage
- Smart Scan query offload



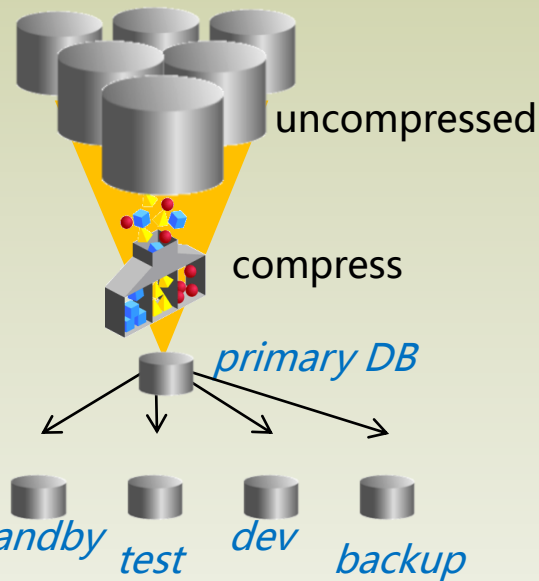
- 智能闪存

- Accelerates random I/O up to 30x
- Triples data scan rate



- 混合列压缩

- 10x compression for warehouses
- 15x compression for archives



公安大数据
战略

更精确甄别犯罪嫌疑人

更强更高效的破案能力

提供更好的社会服务

更早更快的防患于未然

公安大数据
能力和服务



计算服务



数据服务



模型服务

公安大数据
模型和平台



嫌犯洞察



案件洞察



车辆洞察

公安大数据
应用



刑侦



科信



网安



技侦



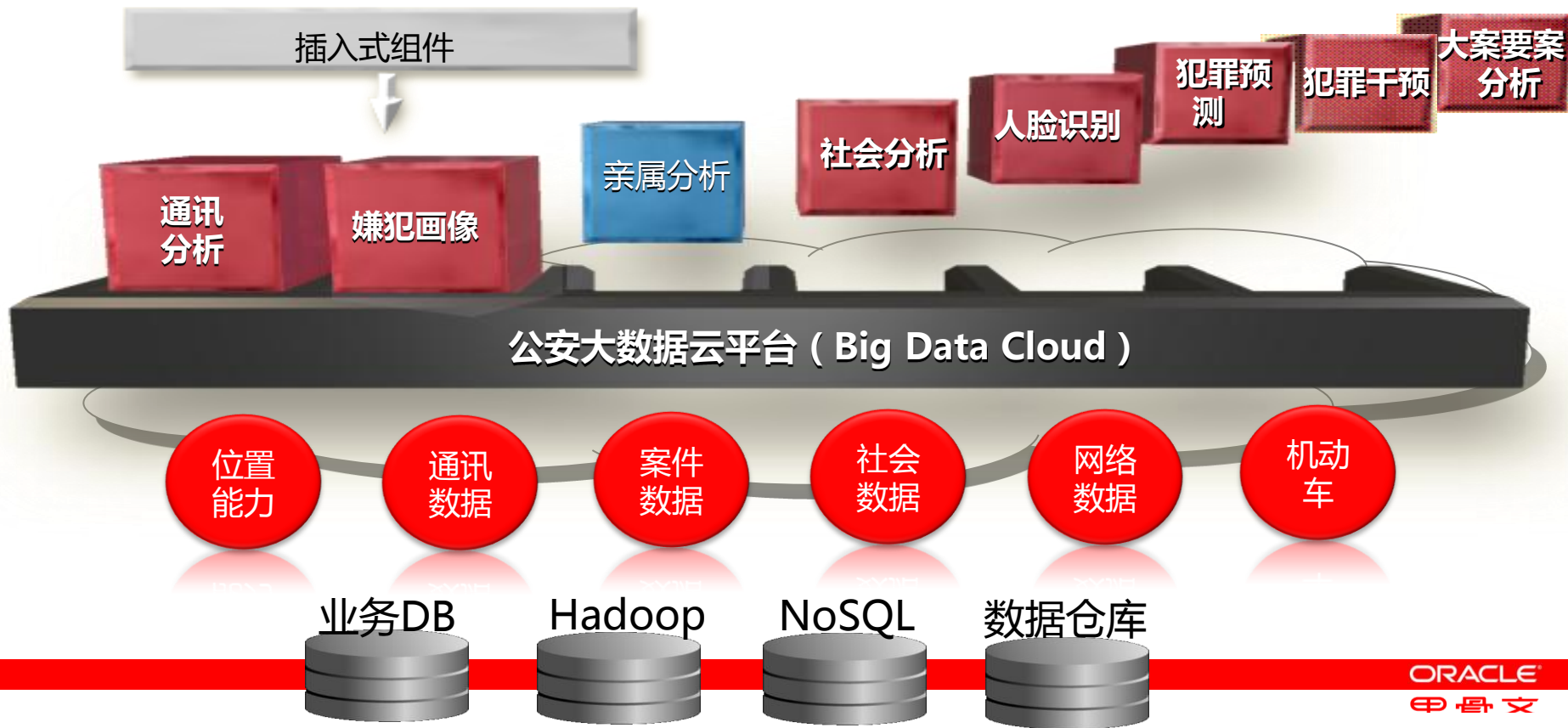
大情报



交管

构建与App Store类似的Data Store运营模式

标准化组件提供、多租户应用开发、插件式管理

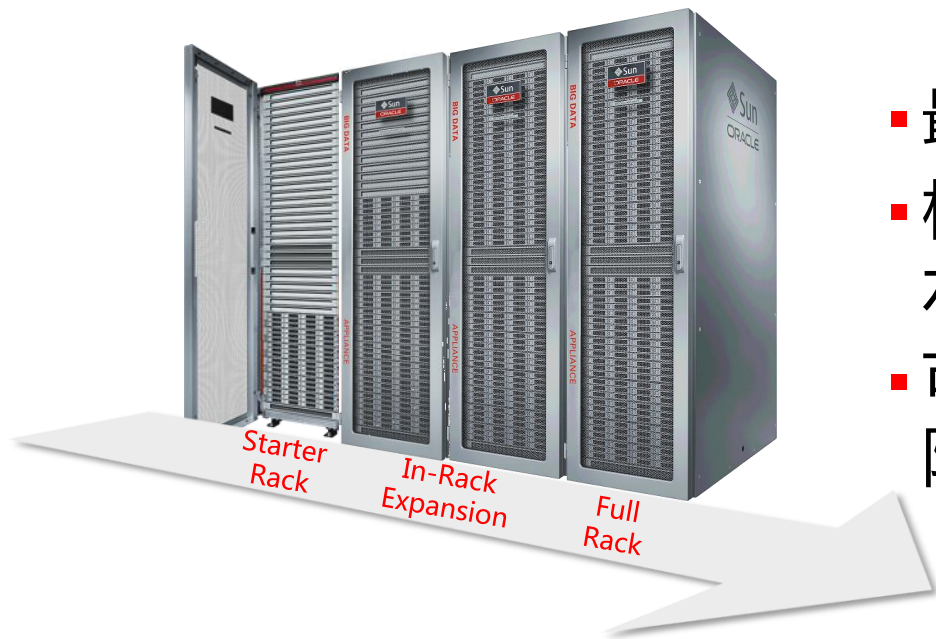




ORACLE®

甲骨文

Big Data Appliance 产品家族



- 最低配1/3配 6个服务器节点
- 柜内可以按照6个节点为单位水平扩展
- 可以通过Infiniband水平无限扩展