

Unlock the value from BigData

FUJITSU

shaping tomorrow with you

现代企业大数据的实时处理技术

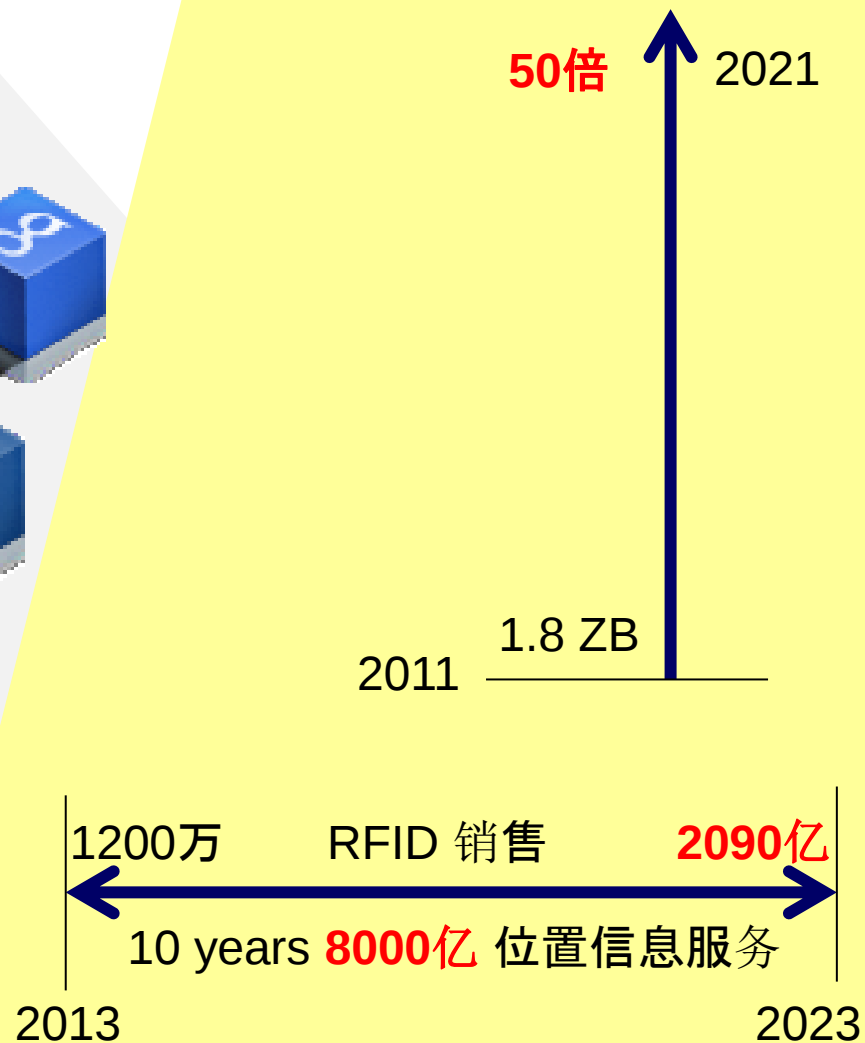
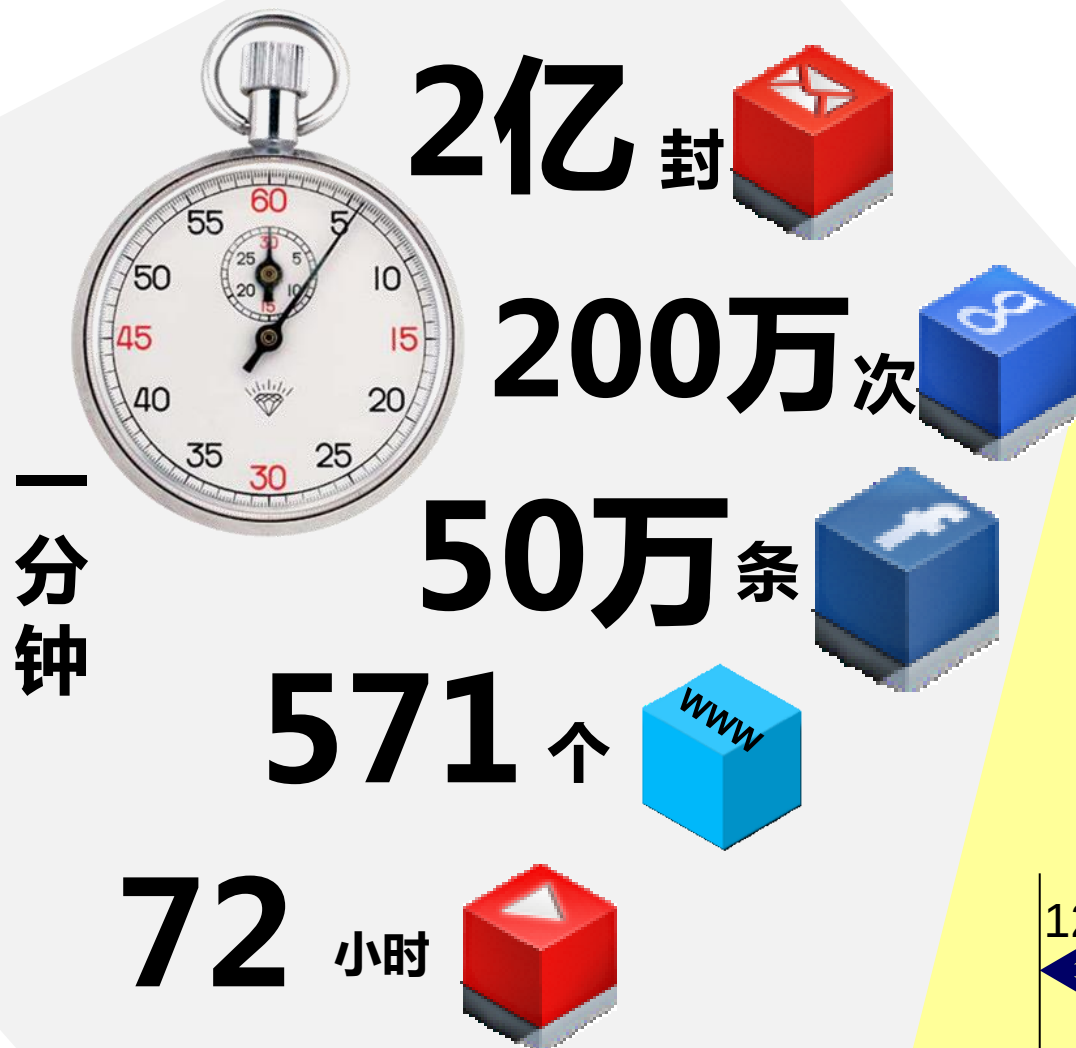
Jan., 20114

周一平

首席技术官（亚太区）

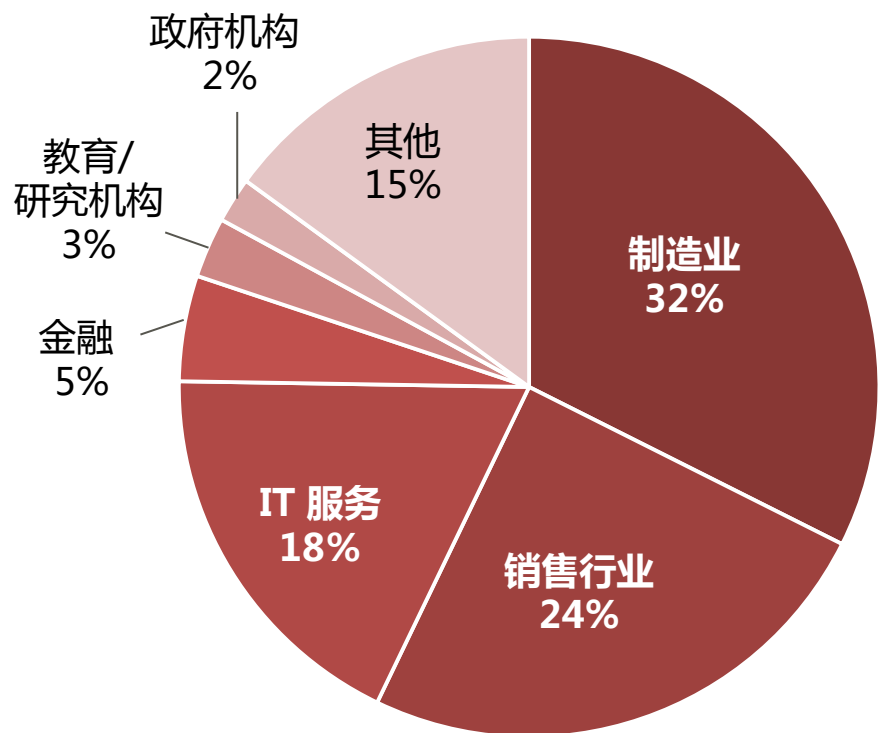
富士通解决方案技术支持整合中心

数字背后的“大数据”

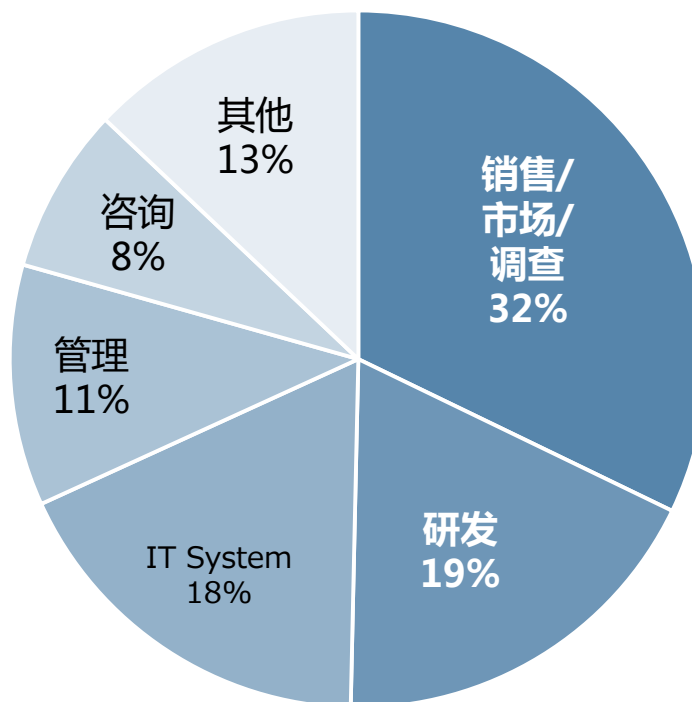


大数据的市场需求分析

行业相关



工作相关



信息来自富士通数据优化创新论坛 (2012年11月)

加速企业挖掘大数据价值，为创新提供最佳平台

业务创新

企业创新

富士通大数据业务方向

总体解决方案

大数据应用

大数据分析

大数据处理引擎

大数据处理平台

采购

生产

库存

销售

内部数据

社区

地理

方法

交通

能源

政府

外部数据

富士通关注企业大数据处理的实时性

大数据实时处理的服务要求SLA

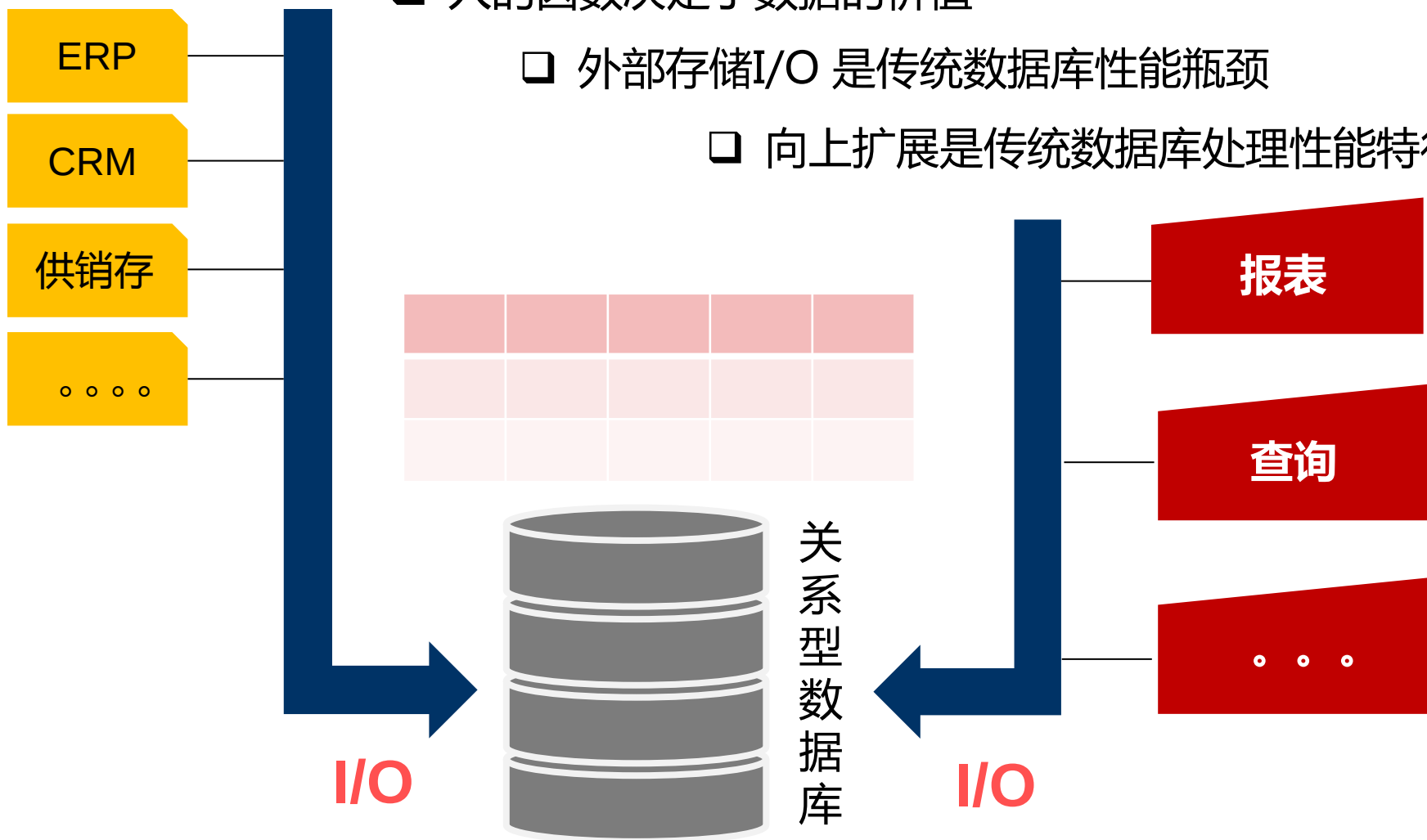


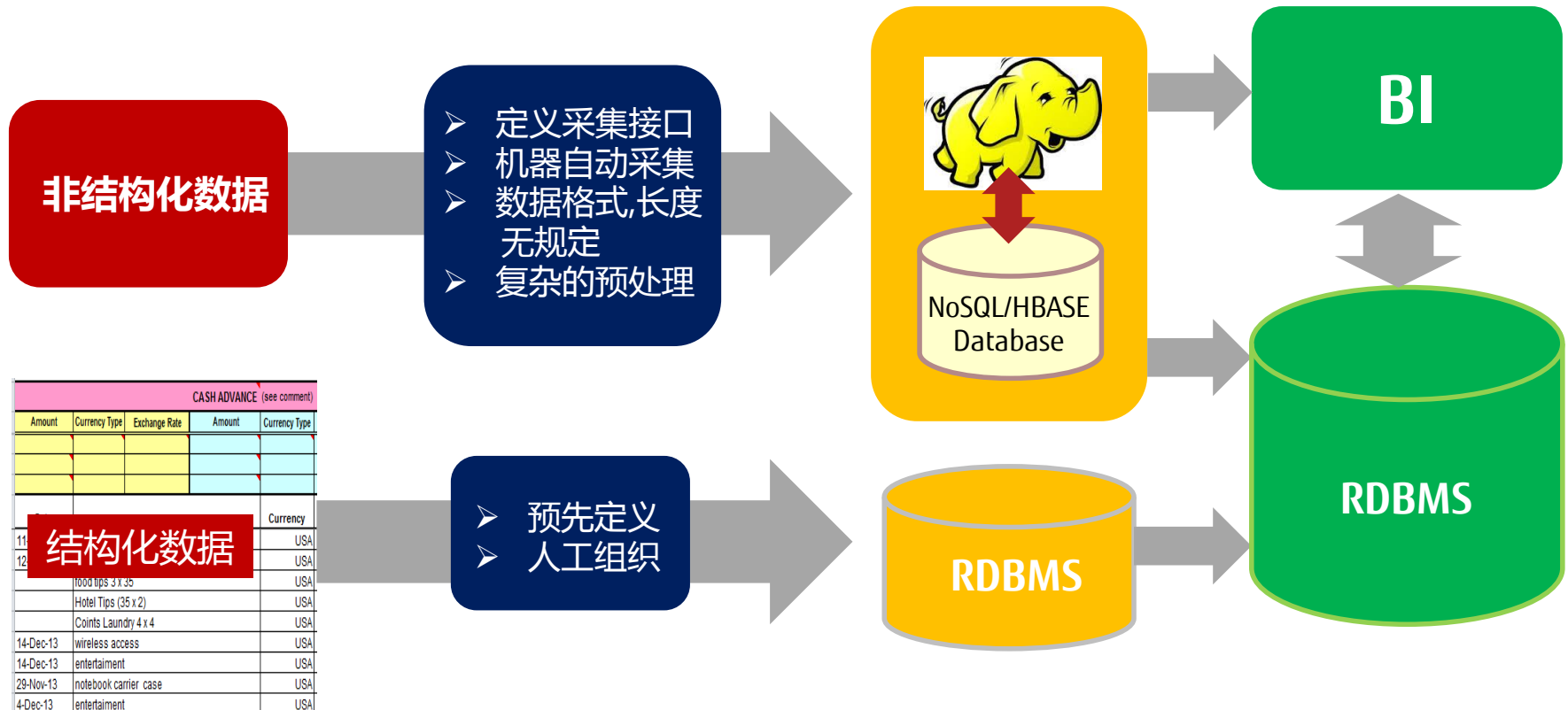
- ❑ 数据采集,整理,直接存为结构化数据

- ❑ 人的因数决定了数据的价值

- ❑ 外部存储I/O 是传统数据库性能瓶颈

- ❑ 向上扩展是传统数据库处理性能特征





■ 非结构化大宗数据的并行处理技术

- Hadoop 分布式文件系统 (HDFS)
- 并行分布式处理 (MapReduce)

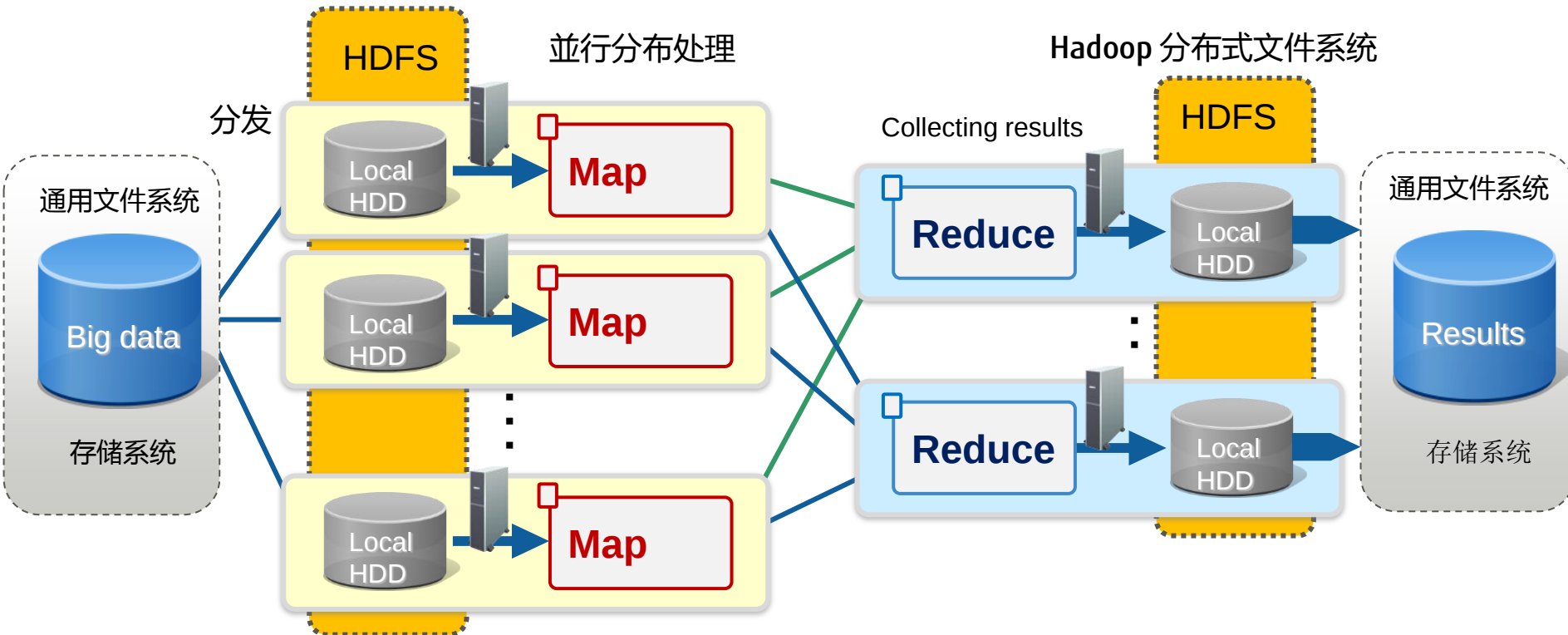


Hadoop 分布式文件系统

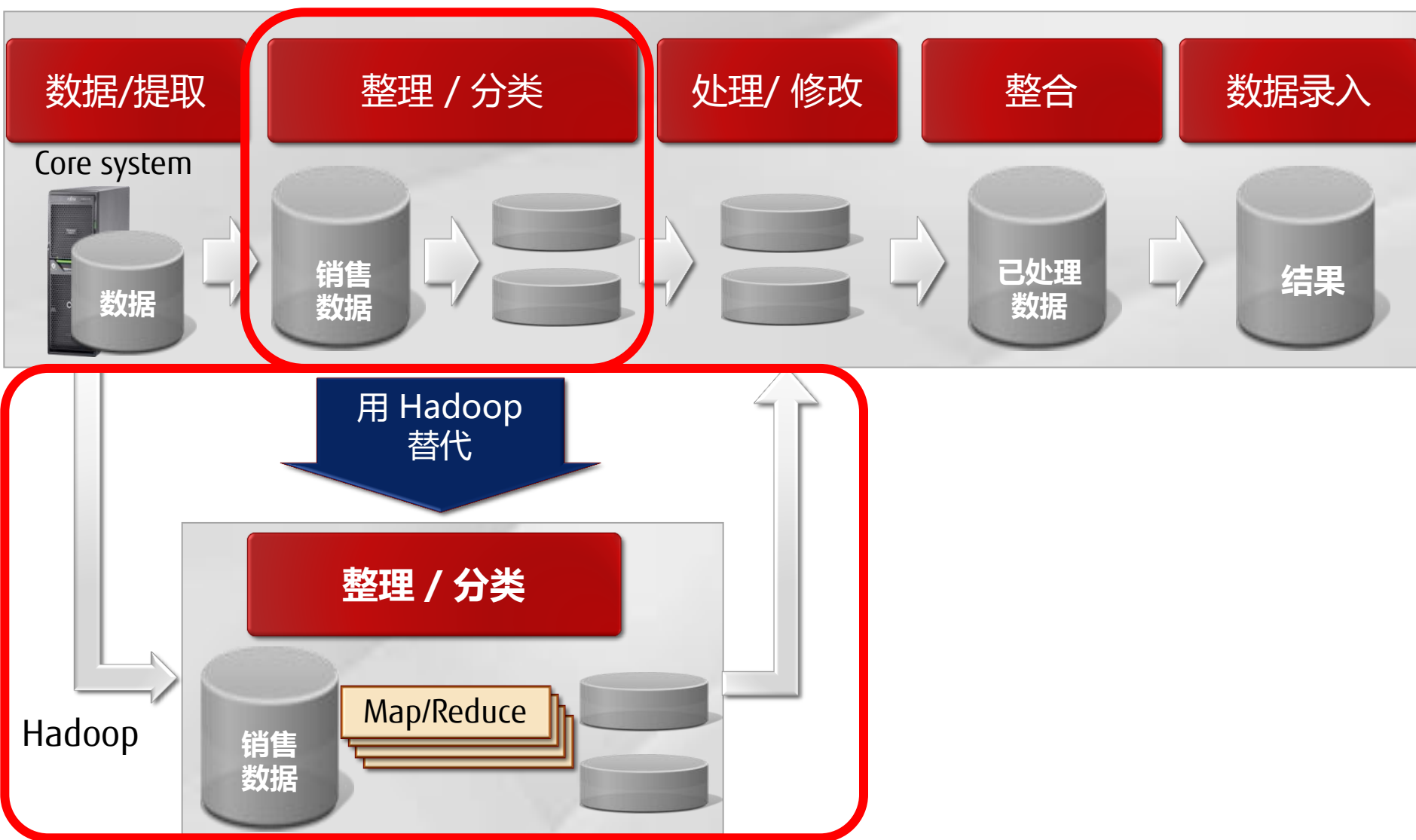
并行分布处理

Hadoop 分布式文件系统

Collecting results



Apache Hadoop 大数据处理的适用性

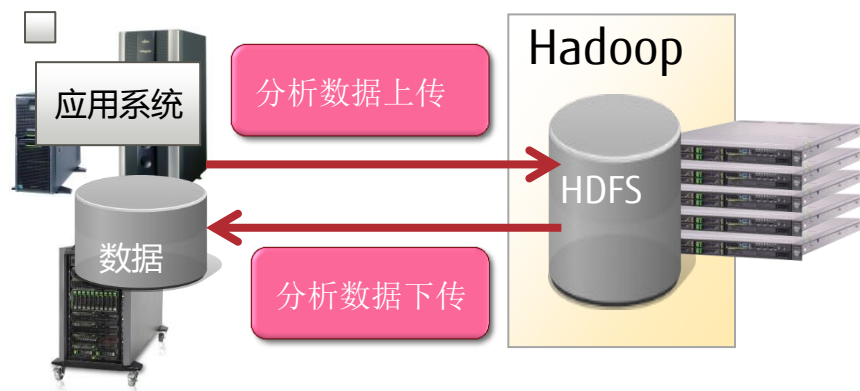


基于开源 Apache Hadoop 并行分布处理技术的解决方案

- 内置富士通集群技术的高可靠主控服务器 (Master Name Server)
- 富士通分布式文件系统 (P-DFS)
 - 同时支持 Hadoop HDFS 文件接口 和标准 POSIX 文件接口
 - 支持 MR 服务器群共享存储系统
- 优化任务分配和页面缓存的专利技术
 - P-DFS I/O 性能超越 HDFS 近 10 倍
 - 处理时间比较 Hadoop 标准版减少了近 5 倍

[Hadoop 标准文件系统]

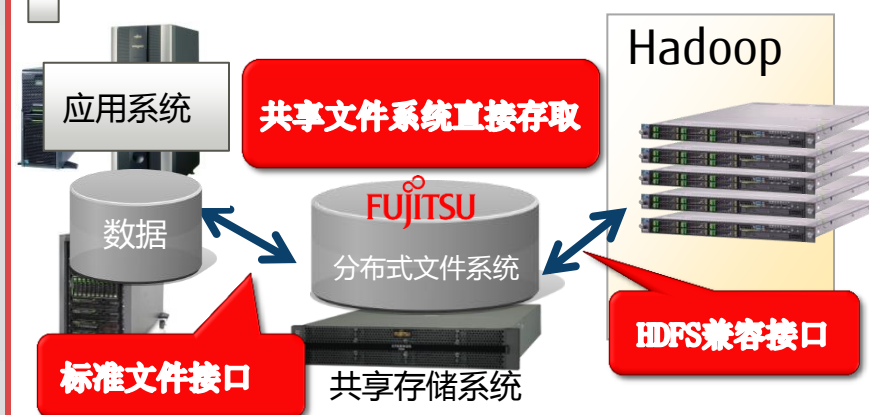
大数据处理系统



Hadoop 必须在数据处理之前将数据上传至 HDFS

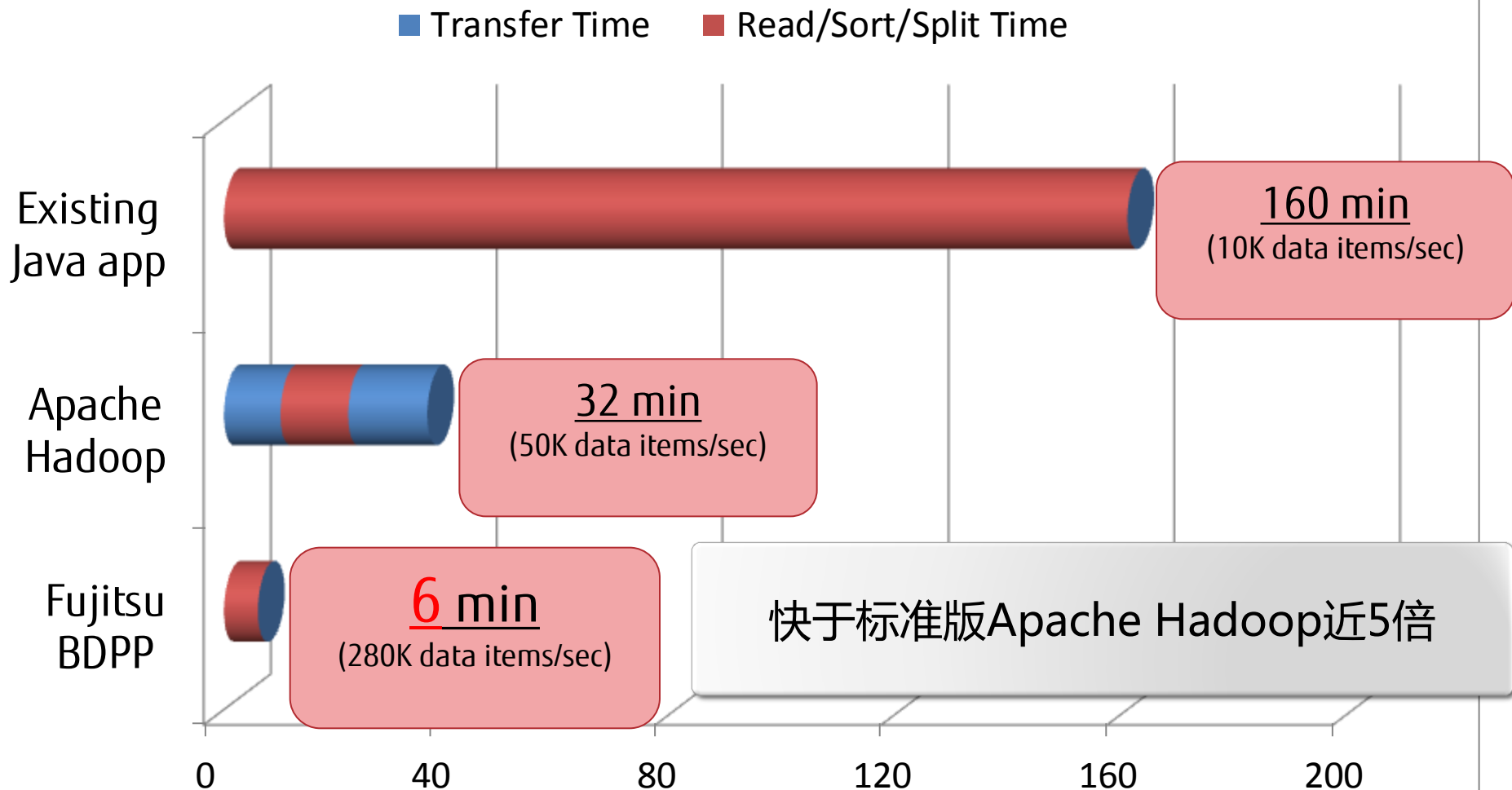
[富士通分布式文件系统]

富士通大数据处理系统



应用系统直接存取待处理数据

■ 1亿数据的整理,分割处理



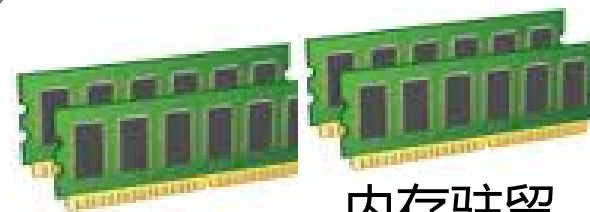
解决I/O瓶颈



缓存



缓存



内存驻留



基于硬盘
数据处理



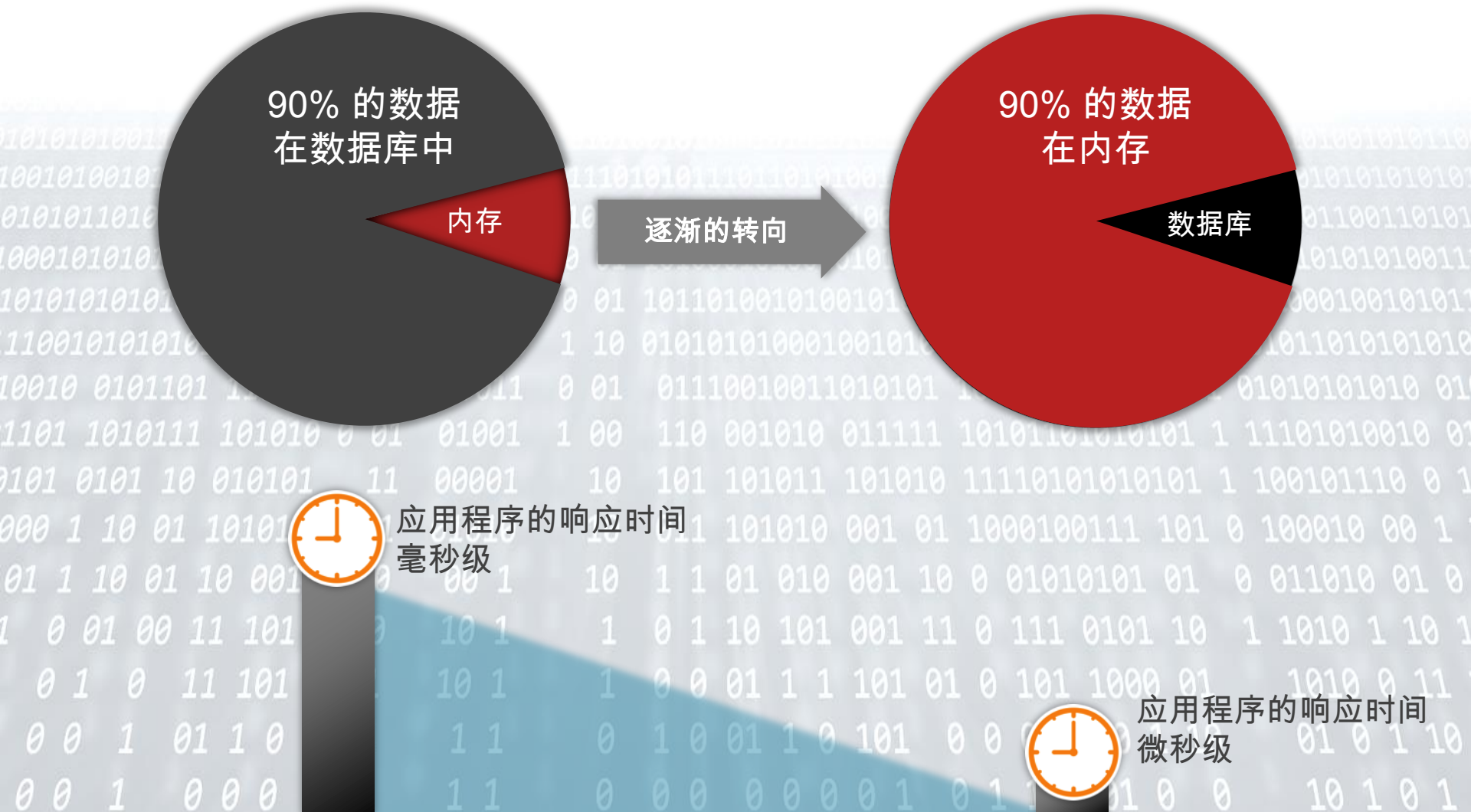
带有缓存
数据处理



基于内存
数据处理

演进

存储介质决定了数据处理的性能



内存的价格使得大数据在内存中处理成为可能

内存处理

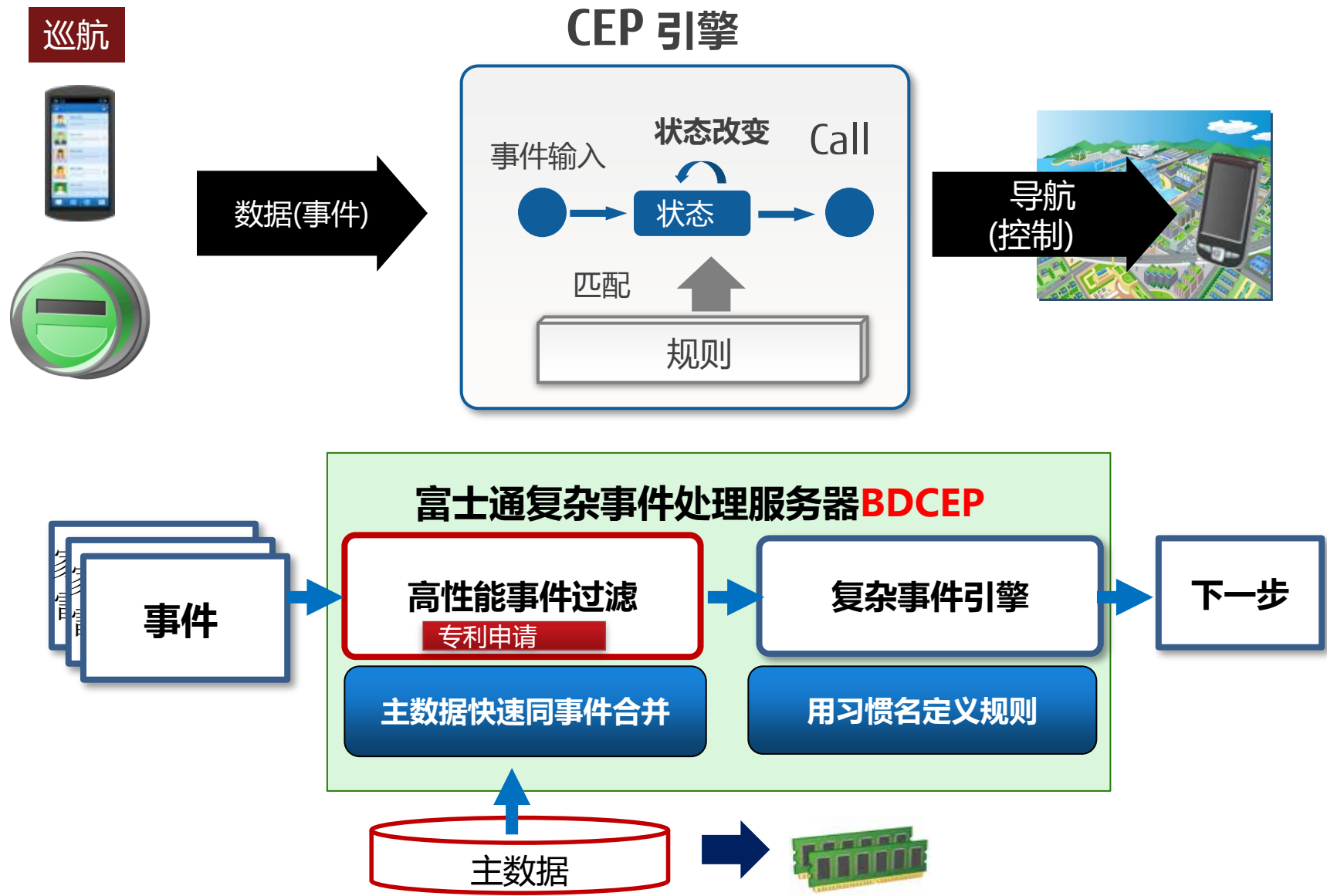
充分利用内存价格的下降

大数据

从数据中挖掘有价值的信息



In-Memory 举例1：大数据预处理 Interstage **BDCEP**



利用故障预测提高设备运行效率

- 检测异常现象和已知的出错样本
- 当预测到出错将会出现，系统提出检修请求

机器标识

扭矩

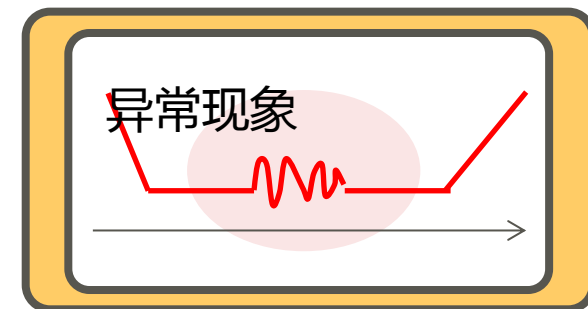
转速

温度

实时分析

✓ 检测出错

异常现象



仓库数据

生产线数据

安全标准

易损零件的检测



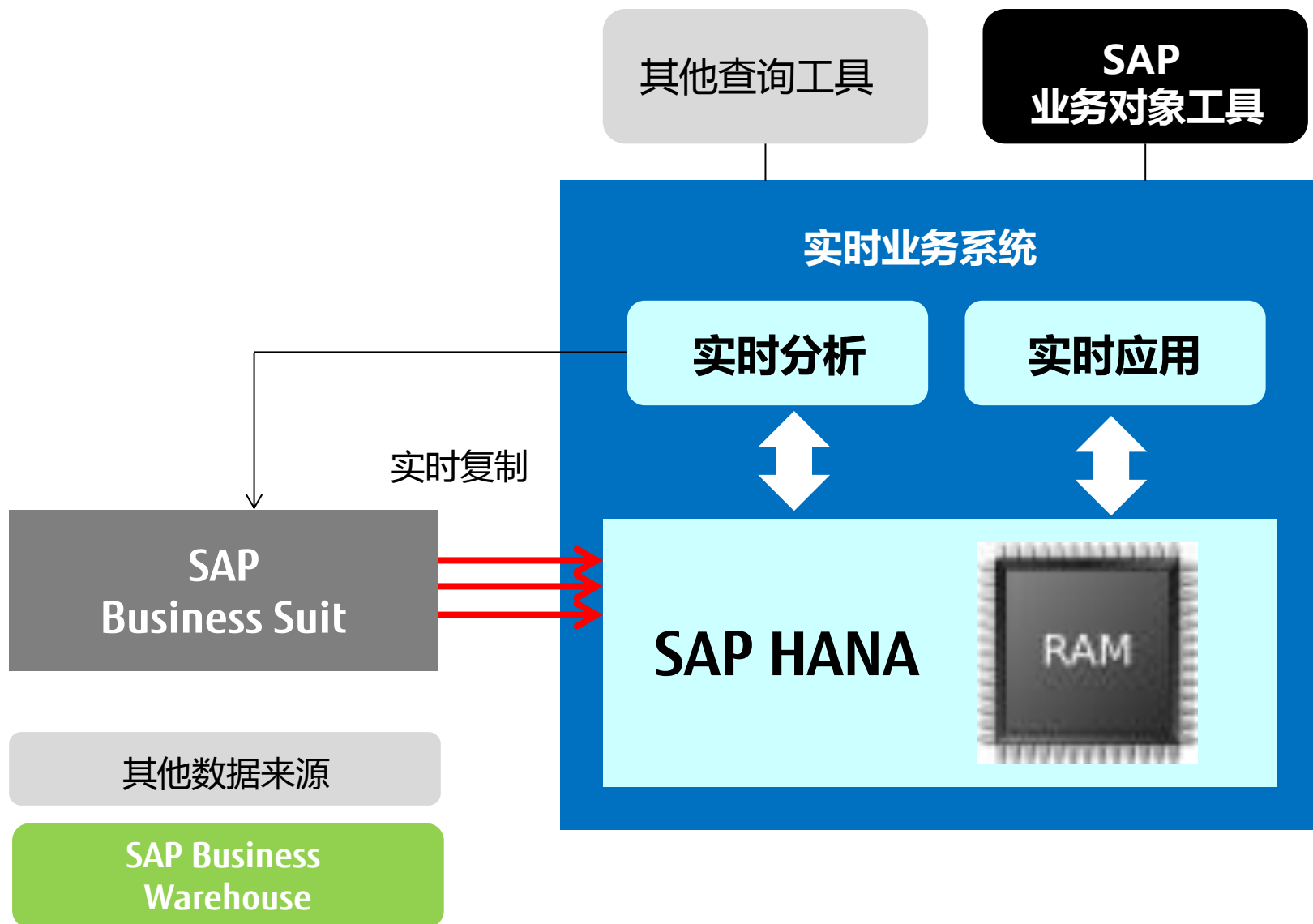
向工程师提供报告

机器故障预测



对生产线发出警报

In-Memory 举例2：SAP HANA



富士通 SAP HANA Solution 解决方案

单节点

- 向上扩展 -

SAP HANA
数据库

SAP HANA 服务器
PRIMERGY RX 600
or
PRIMERGY RX900

支持内部存储系统

- 1 x 服务器
- 内存 128GB to 1TB
- 不具备高可用性和灾备配置
- 适合中小企业非关键业务

多节点

- 向外扩展 -

SAP HANA
数据库

SAP HANA
服务器 1
PRIMERGY RX 600
512 GB RAM

SAP HANA
服务器 2
PRIMERGY RX 600
512 GB RAM

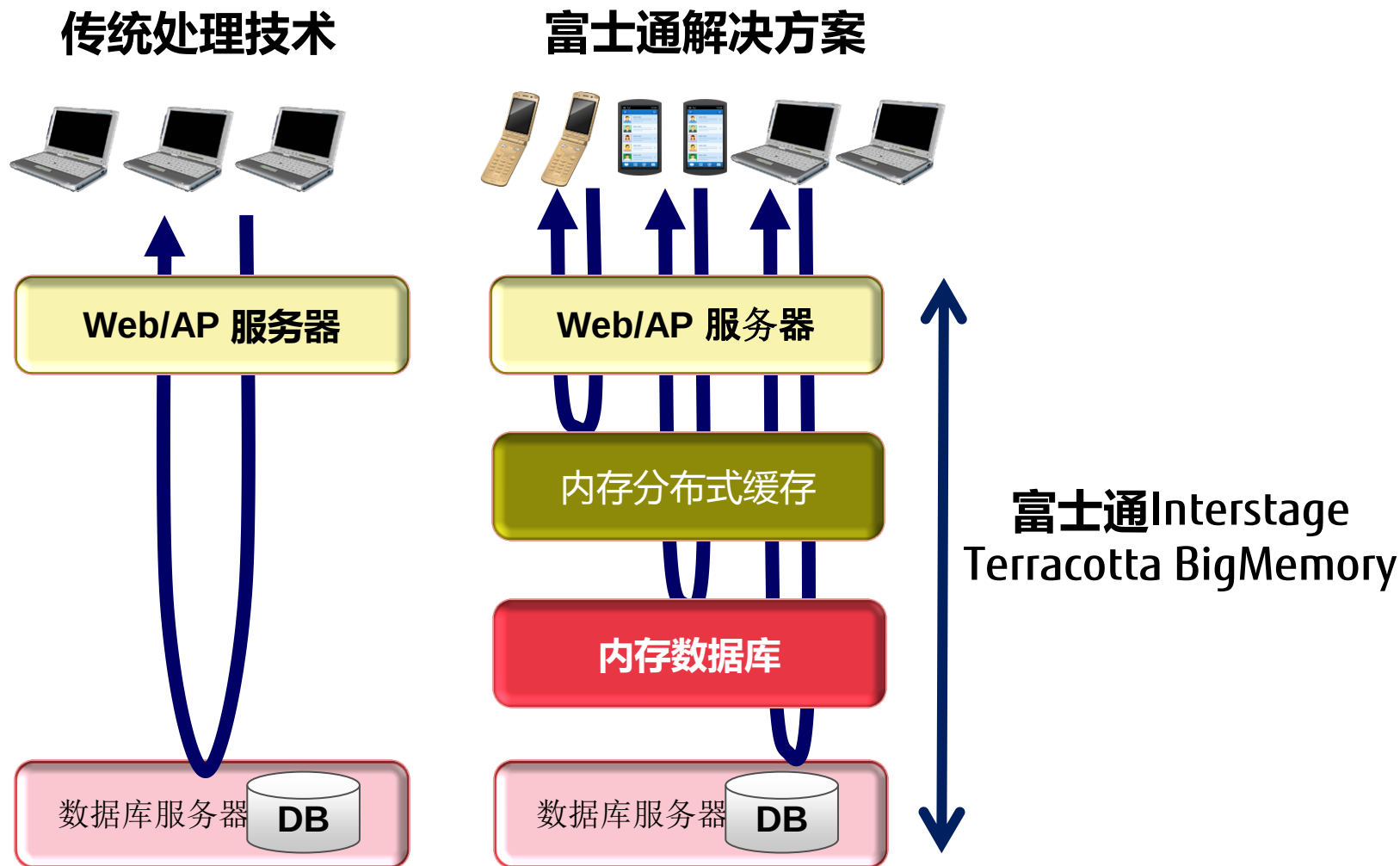
SAP HANA
服务器 N
PRIMERGY RX 600
512 GB RAM

需要配置外部存储器

- 有多台服务器组成的SAP HANA 系统
- 增加计算节点和存储(最大可扩展至8TB)
- 已包括高可用性配置,灾备为可选
- 适合中大企业关键业务

共享存储
NetApp
FAS 3240





In-Memory 在通用 JAVA+DB 企业应用中的应用

■ 性能

- 将关键业务数据大部分驻存内存,业务应用系统的性能得以大幅度提高。

■ 规模

- 可管理TB级内存数据及每秒几10万的数据交易量

■ 管理

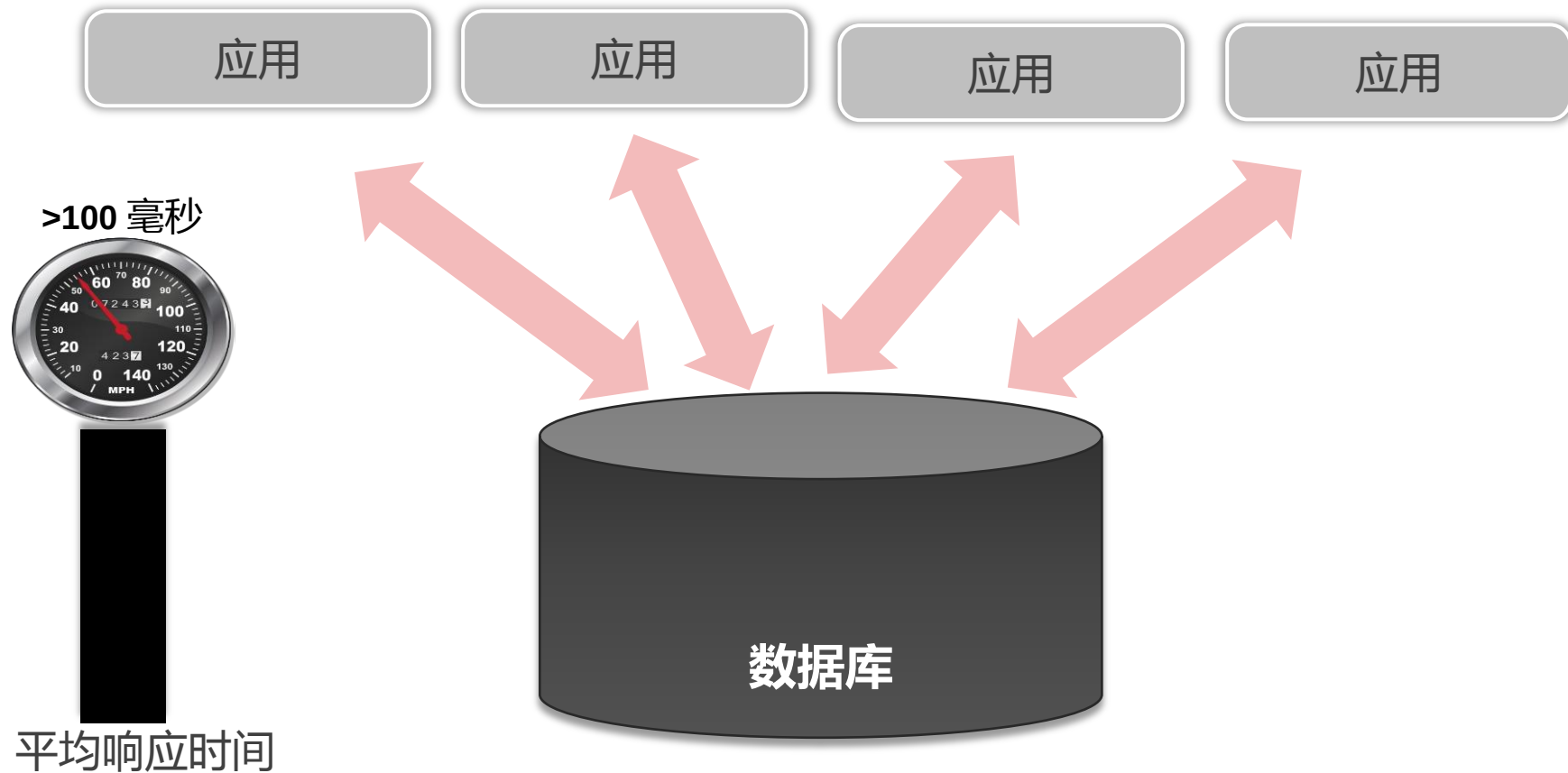
- 管理,监控系统 and 数据以实现系统的可靠和高可用性

■ 简易

- 快速部署, 简化配置, 容易存取和无缝升级



数据库的 I/O 瓶颈制约了应用系统的成本,性能和系统可扩展性



响应速度



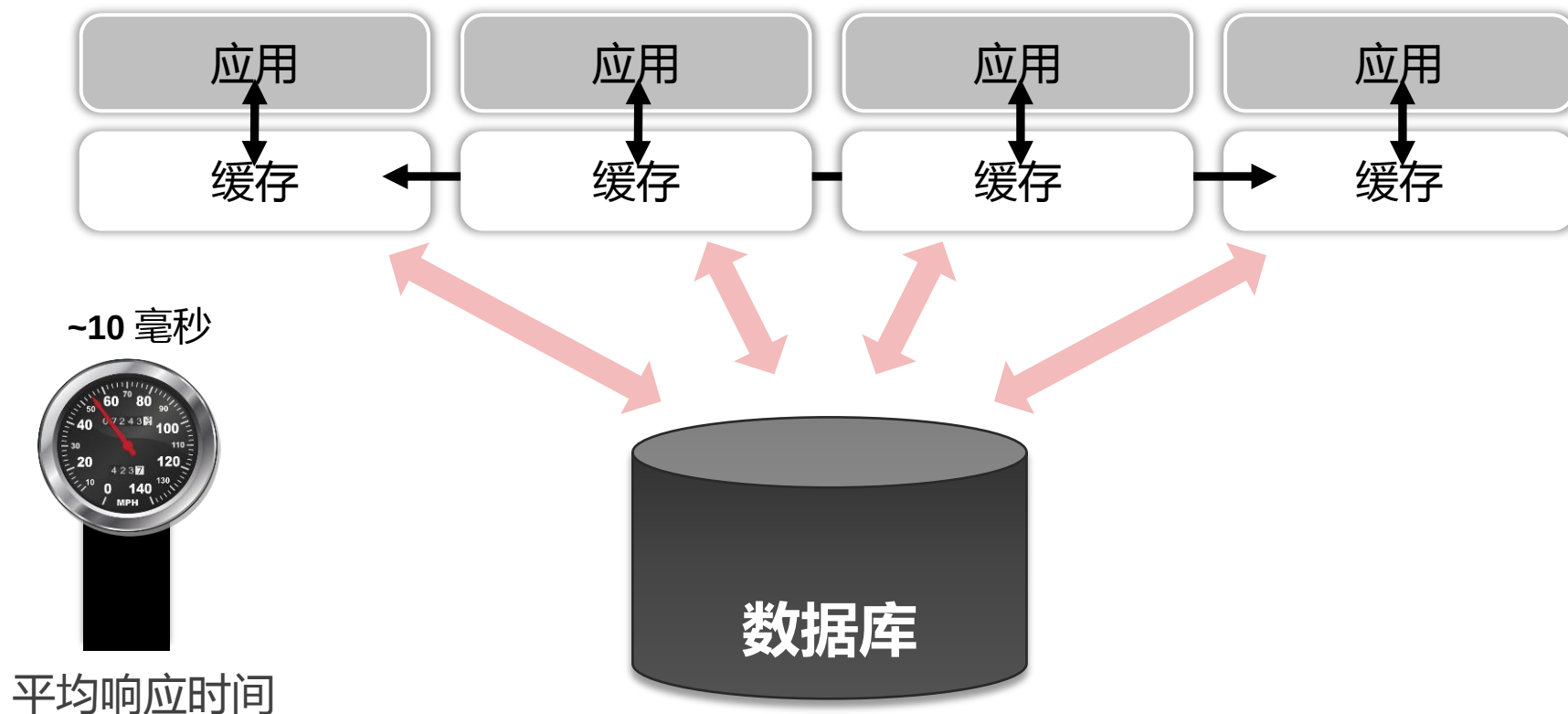
系统成本



可扩展性



缓存的引进大大改善了数据库的 I/O 压力



响应速度



系统成本

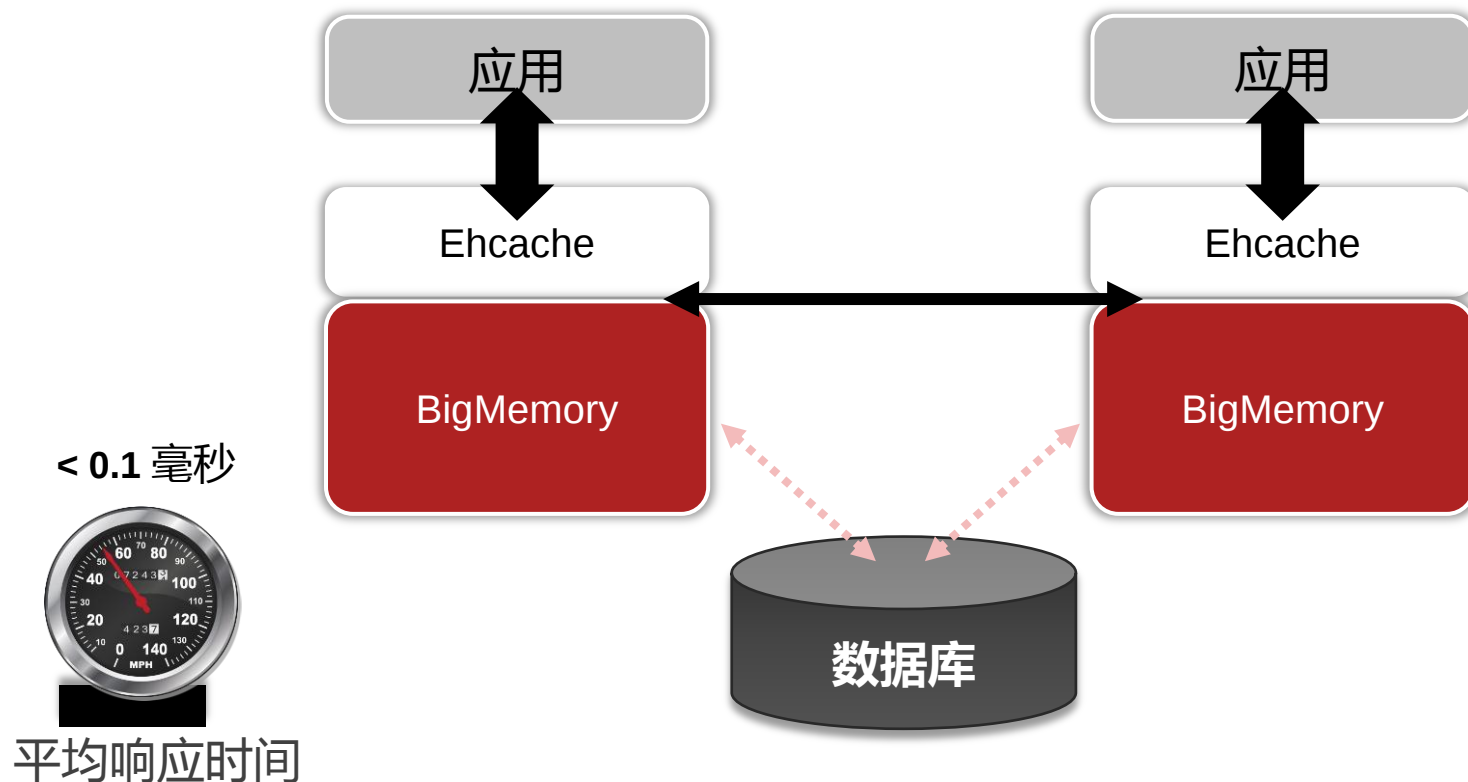


可扩展性



富士通 Interstage BigMemory解决方案

结合Ehcache，BigMemory进一步提高了 JAVA+DB 系统的性能

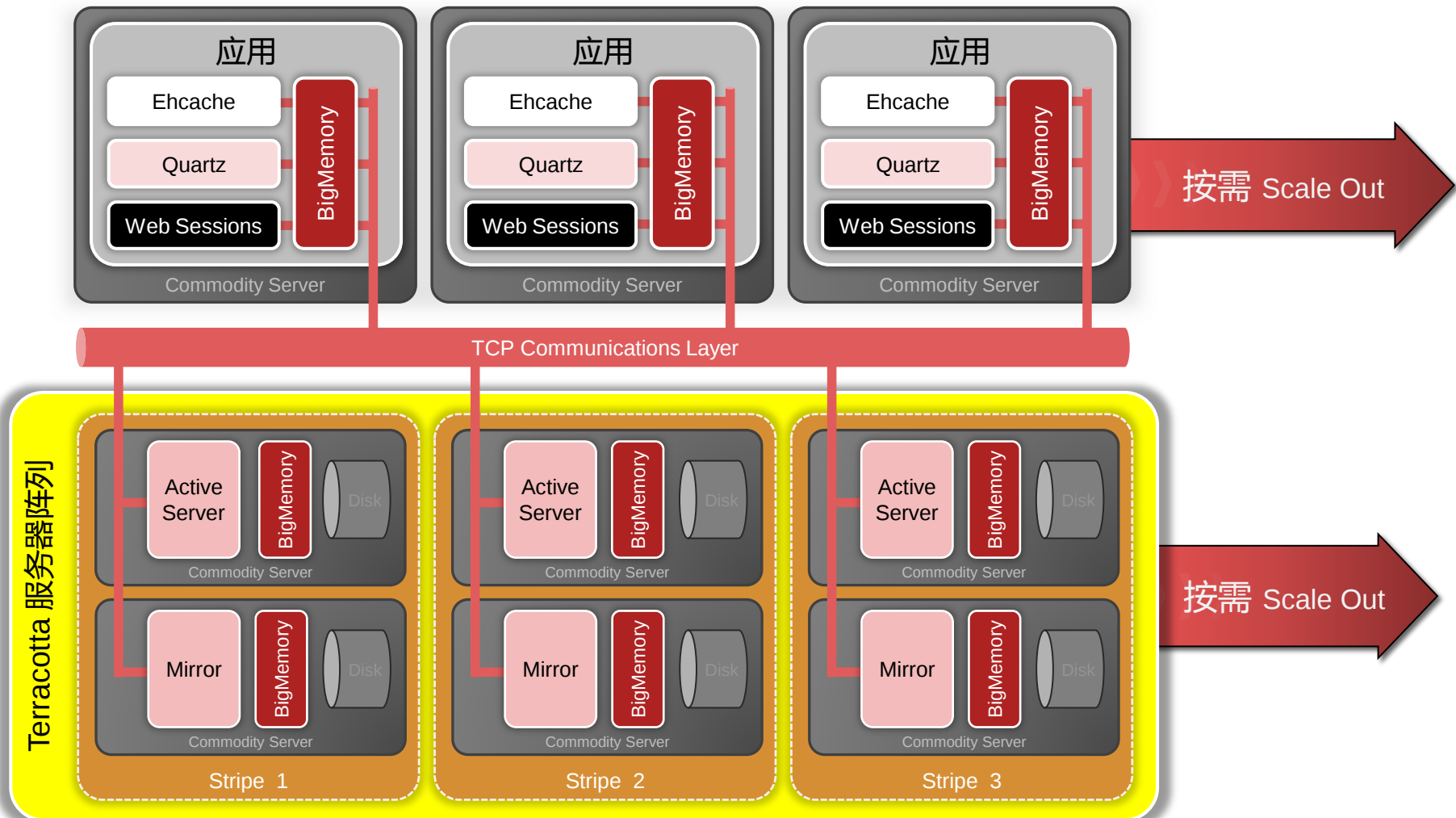


响应时间   

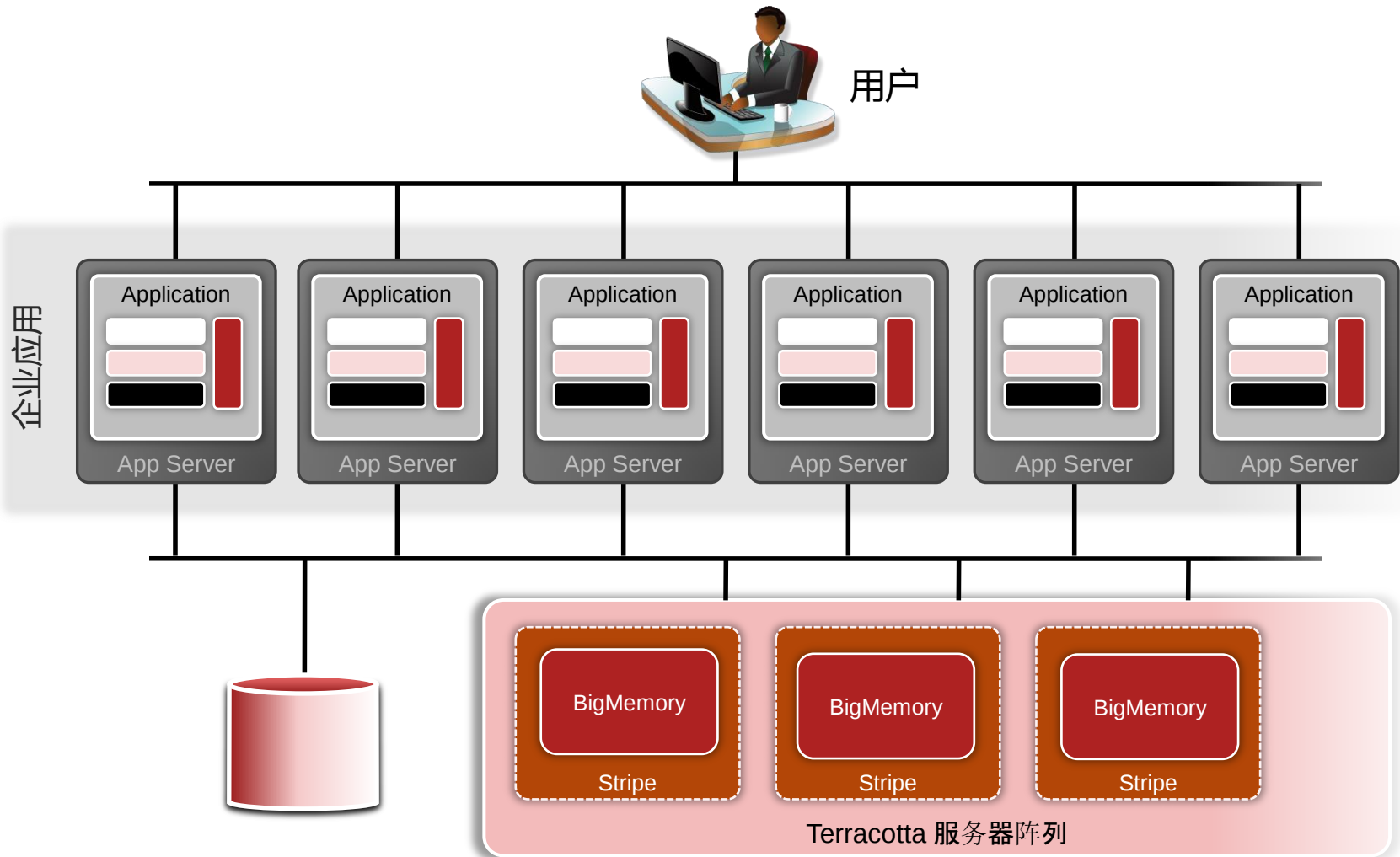
系统成本   

可扩展性   

Terracotta BigMemory服务器阵列平台



富士通 BigMemory 在云计算环境中的应用

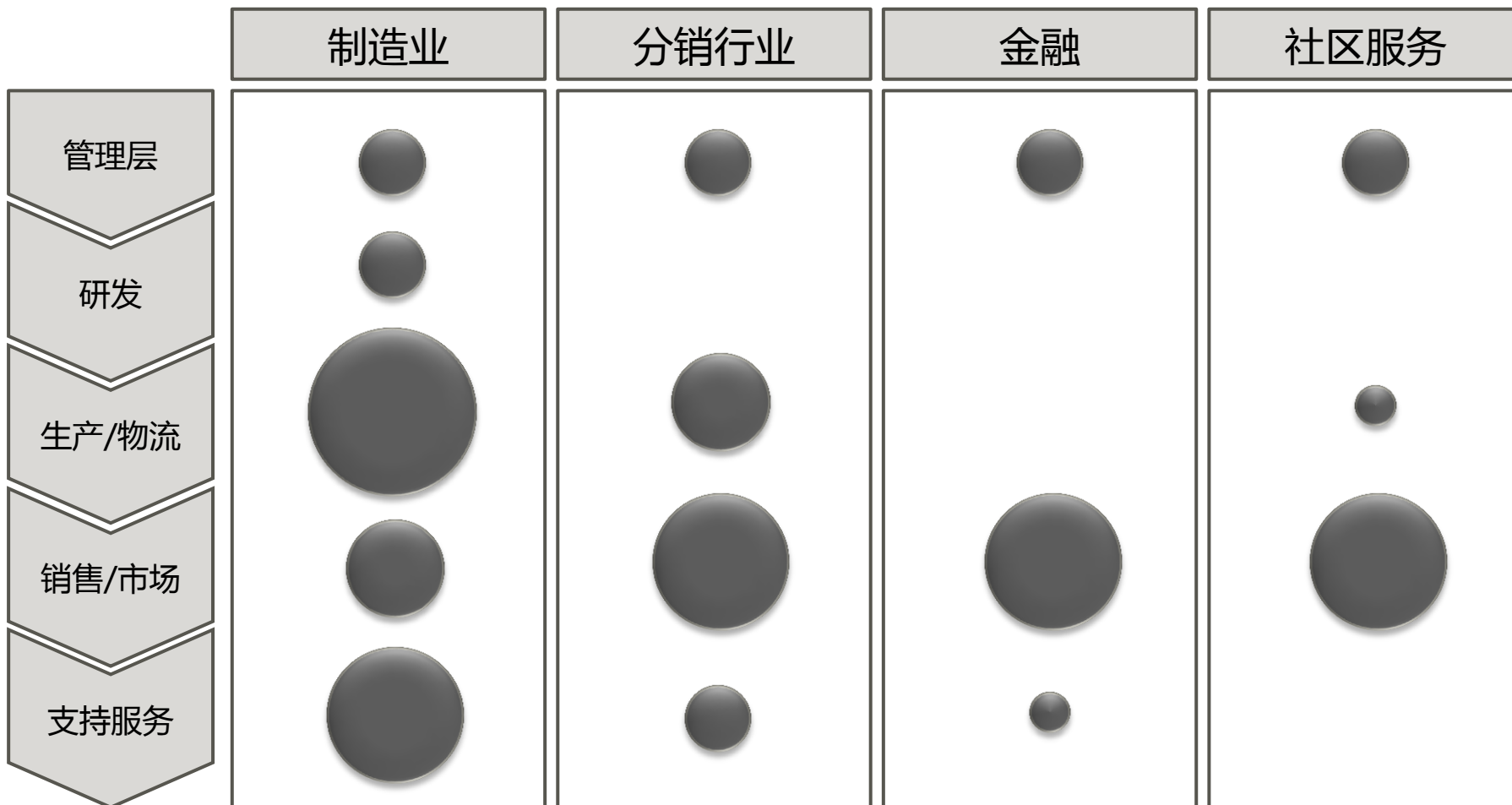


富士通提供全面大数据实时处理技术



2014 年大数据的市场预测

● 市场大小

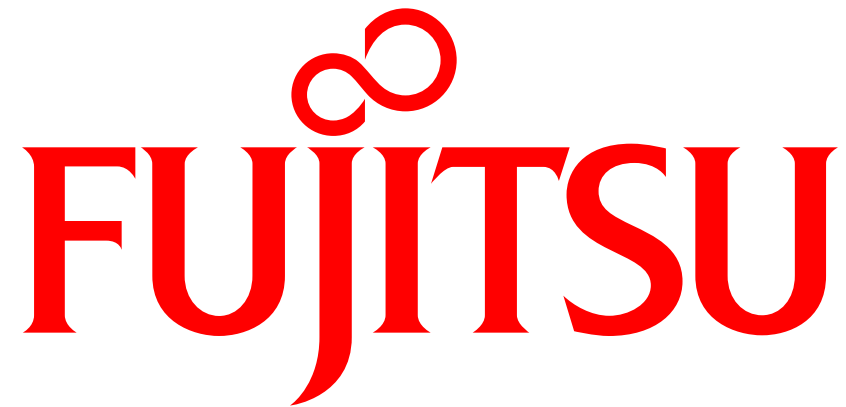




融合,创新开拓大数据应用

创建大数据企业处理平台

深挖大数据,为企业业务
提供有价值的信息



shaping tomorrow with you