



工商数据中心架构创新之路

付宏伟

工商总局信息中心

2016年10月

目录

- ① 对架构的基本理解
- ② 工商数据中心的发展和现状
- ③ 工商数据中心改造的思路
- ④ 未来展望

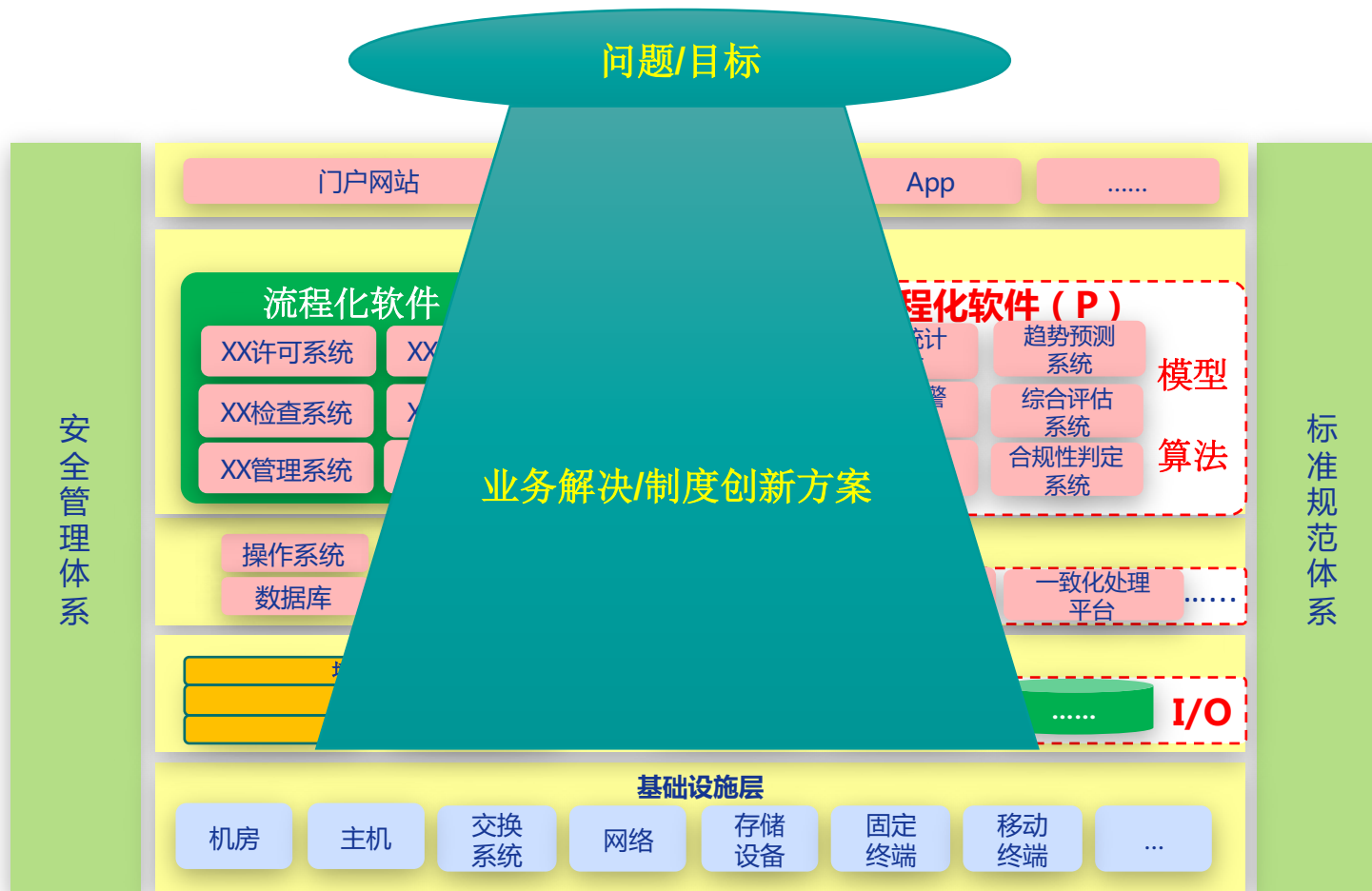


1

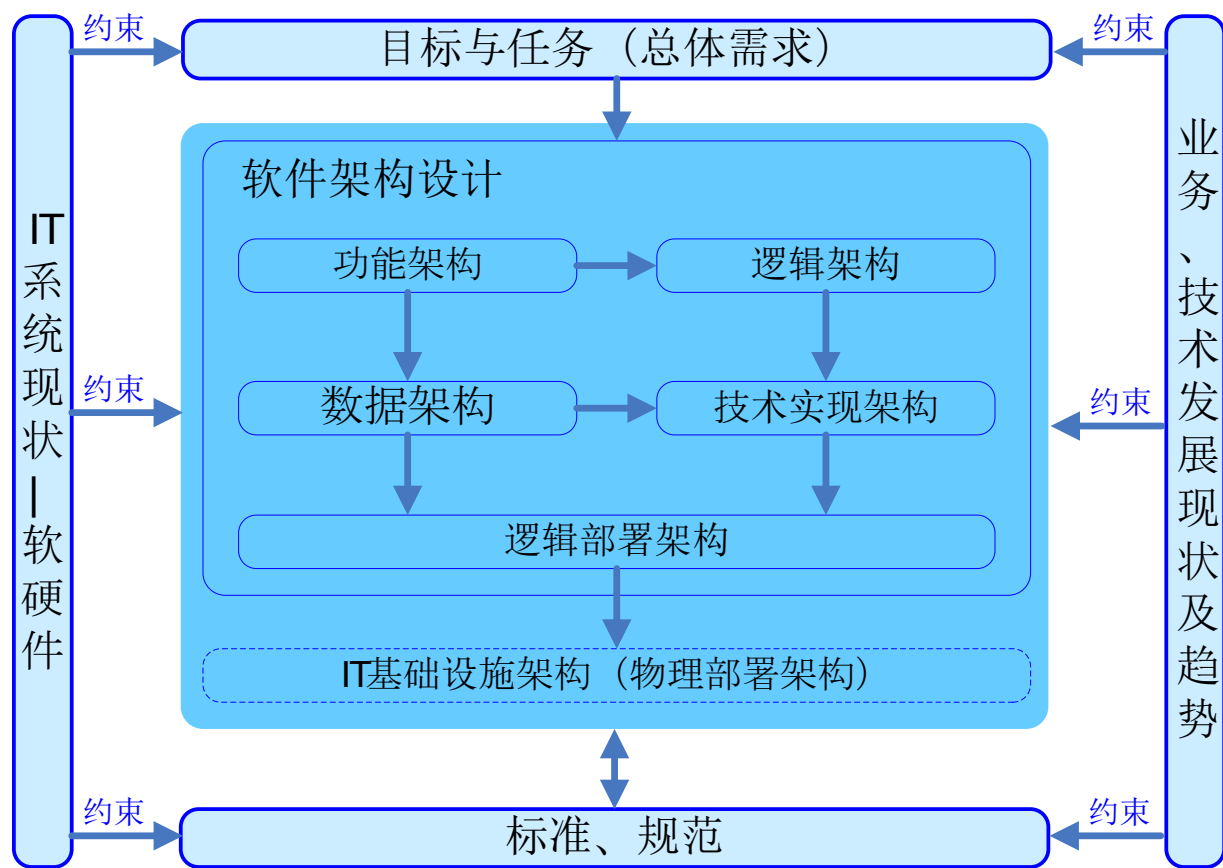
对架构的基本理解

架构服务于战略目标

提高部门职能履行水平和政务效能



架构设计方法

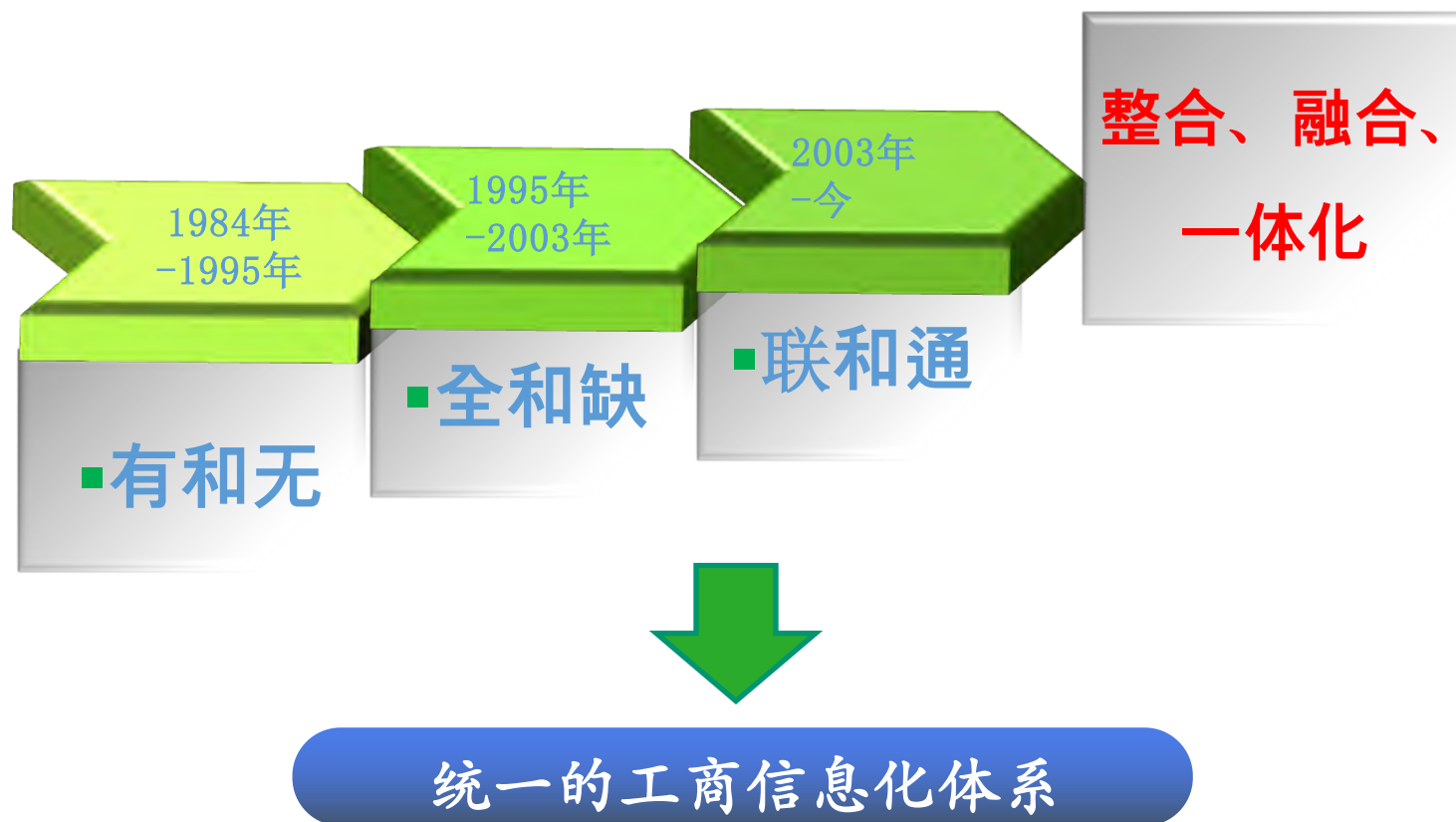




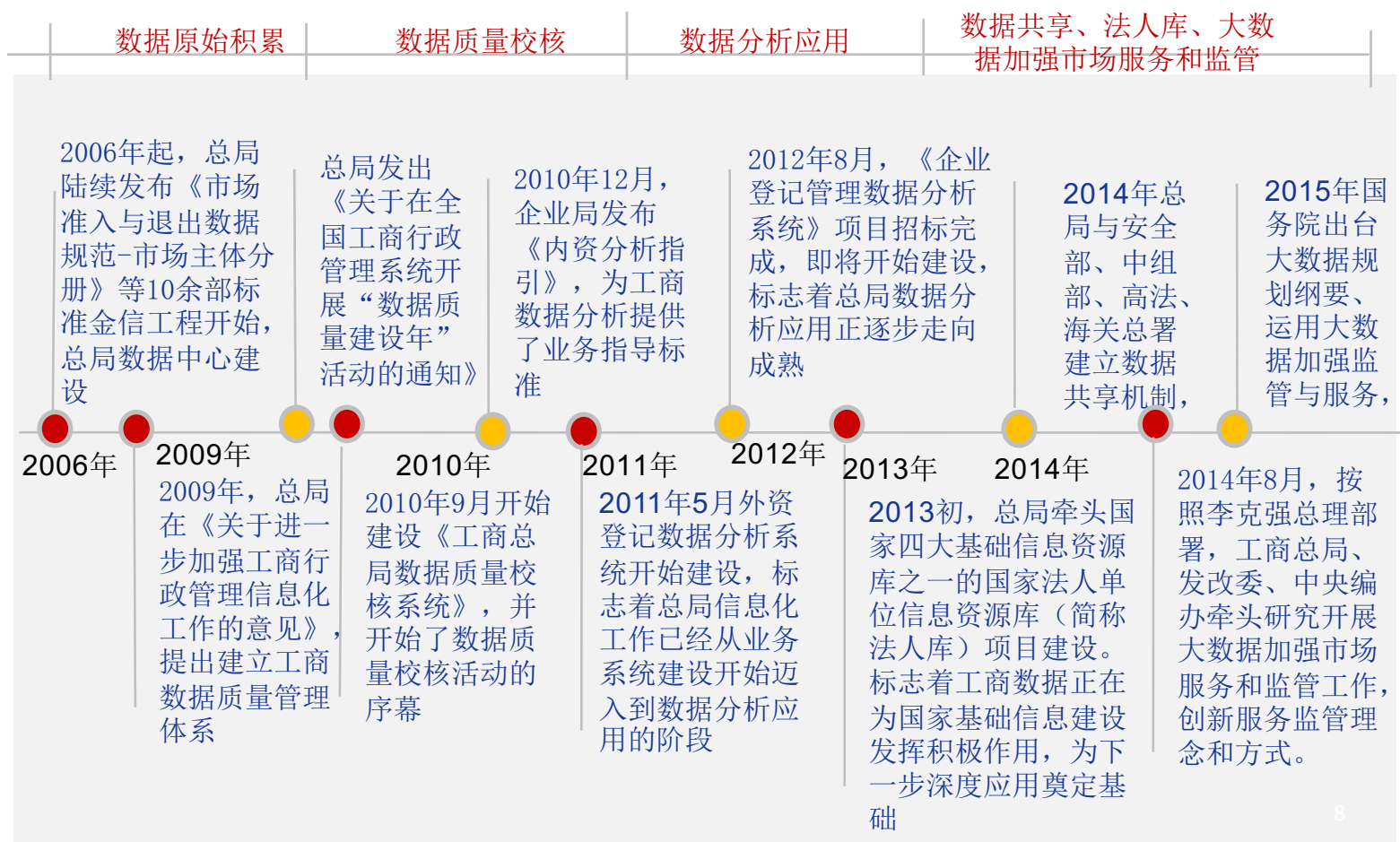
2

工商数据中心的发展和现状

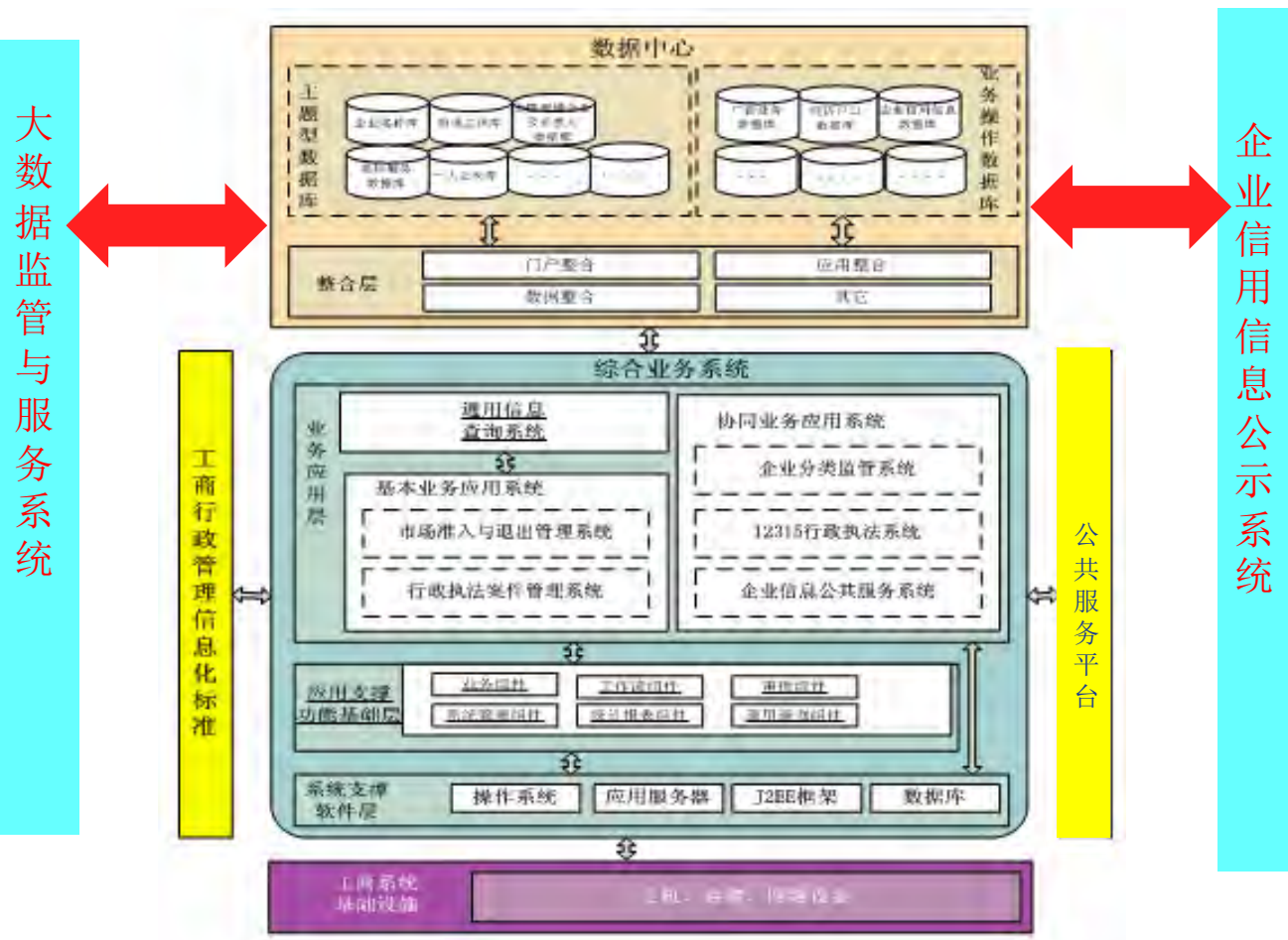
工商信息化发展概况



总局数据中心建设经过了十年的积累

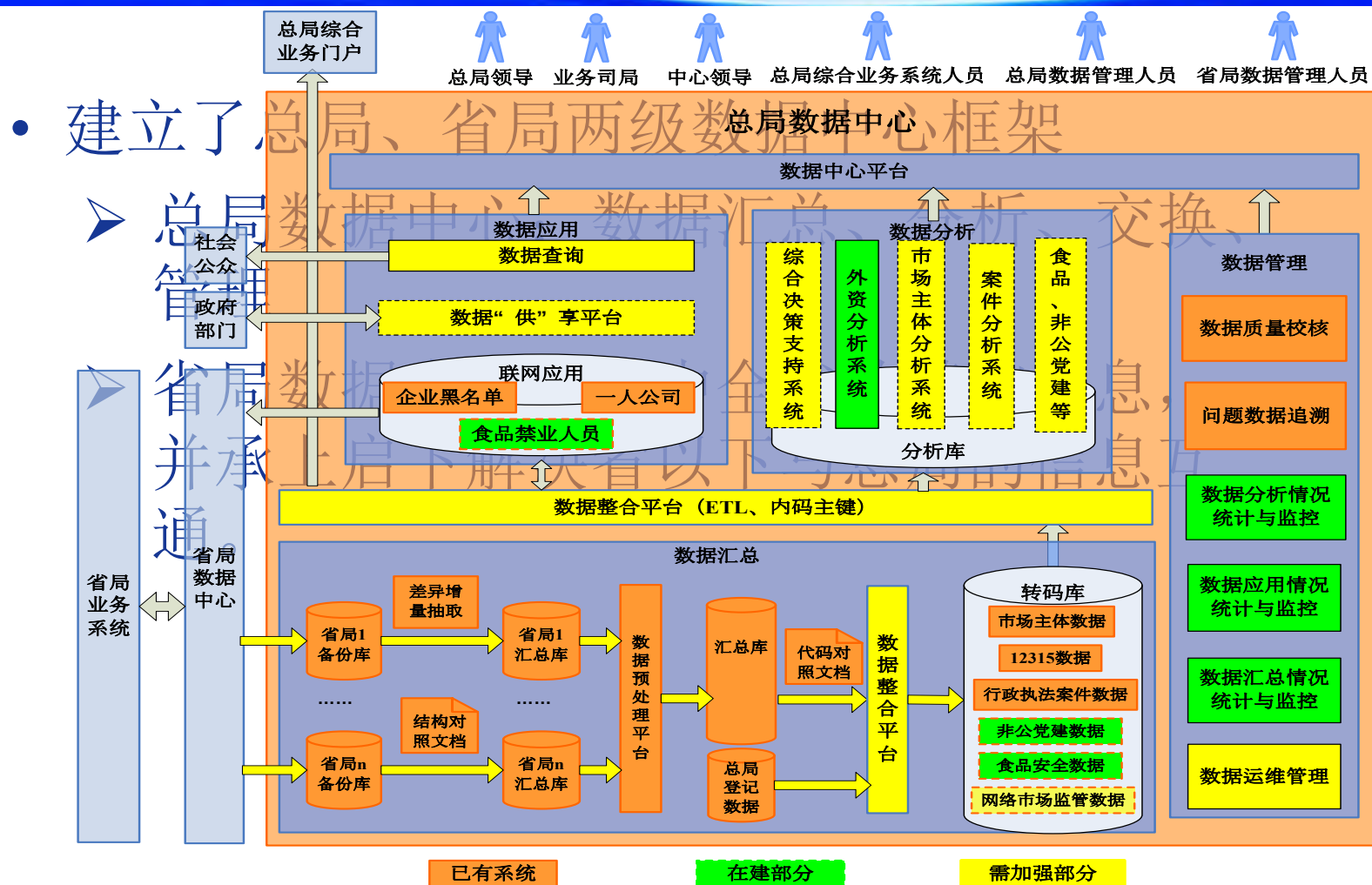


数据中心是工商信息化重点建设内容之一



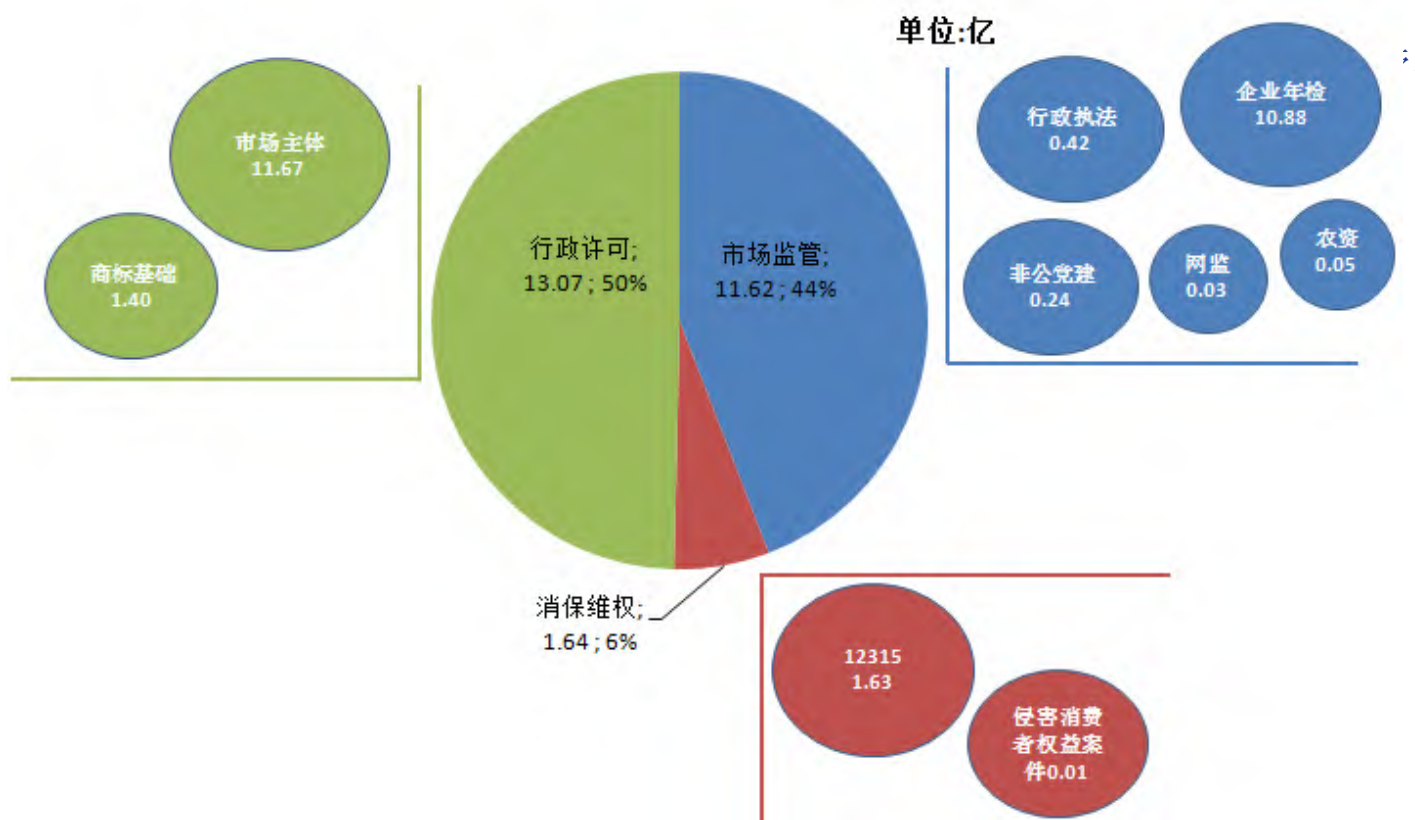
两级数据中心 框架

数据中心建设取得了显著的成效（1）



数据中心建设取得了显著的成效（2）

数据资源数据量情况



数据中心建设取得了显著的成效（3）

- 形成了数据质量管理为核心的数据保障体系
 - 集中开展了2010-2012“数据质量建设年”活动，印发了《工商行政管理数据管理办法》
 - 开发了数据质量评价校核管理系统，建立数据质量监控、问题发现、反馈、整改、跟踪、通报日常工作机制
 - 市场主体登记数据平均数据完整率、准确率、规范性分别达到97.81%、98.10%、97.27%，其他业务数据达到完整率95%、准确率98%。

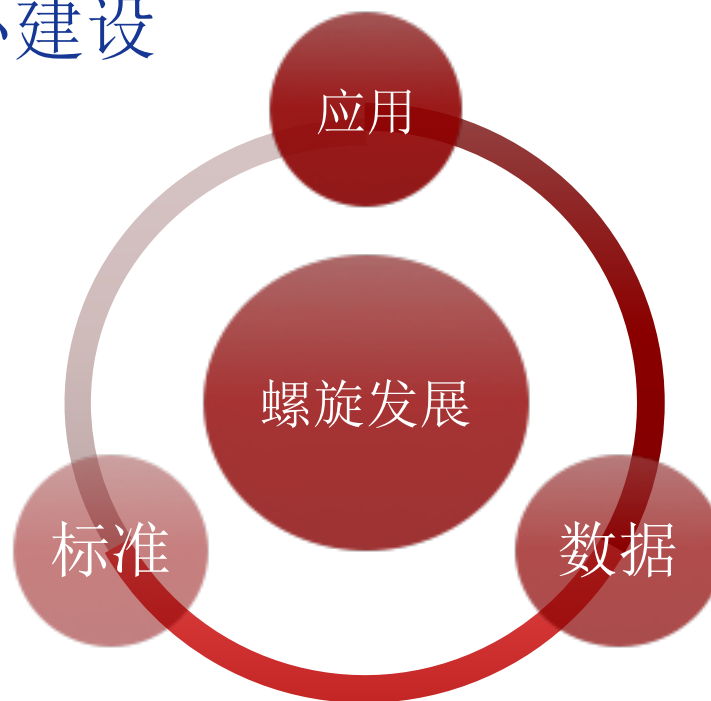
数据中心建设取得了显著的成效（4）

- 提升了工商监管、服务能力

- 联网应用：黑牌企业、自然人一人公司、失信被执行人，“一处失信，处处受限”
- 数据共享：住建部、人社部、民政部、国家统计局、国家税务总局、审计署等多个政府部门提供共享信息；服务地方经济社会发展，为北京金融行业数据分析、上海创新对口支援方式、河南中原经济区、京津冀一体化建设提供重要数据支撑。
- 数据分析：建设了企业（外资）登记管理数据分析系统，形成了《企业发展与宏观经济关系分析》和《十七大以来全国内资企业发展分析》、《全国外商投资企业数据分析报告》、《全国内资企业生存时间分析报告》等30余份报告，引起了广泛关注。

数据中心建设基本经验总结（1）

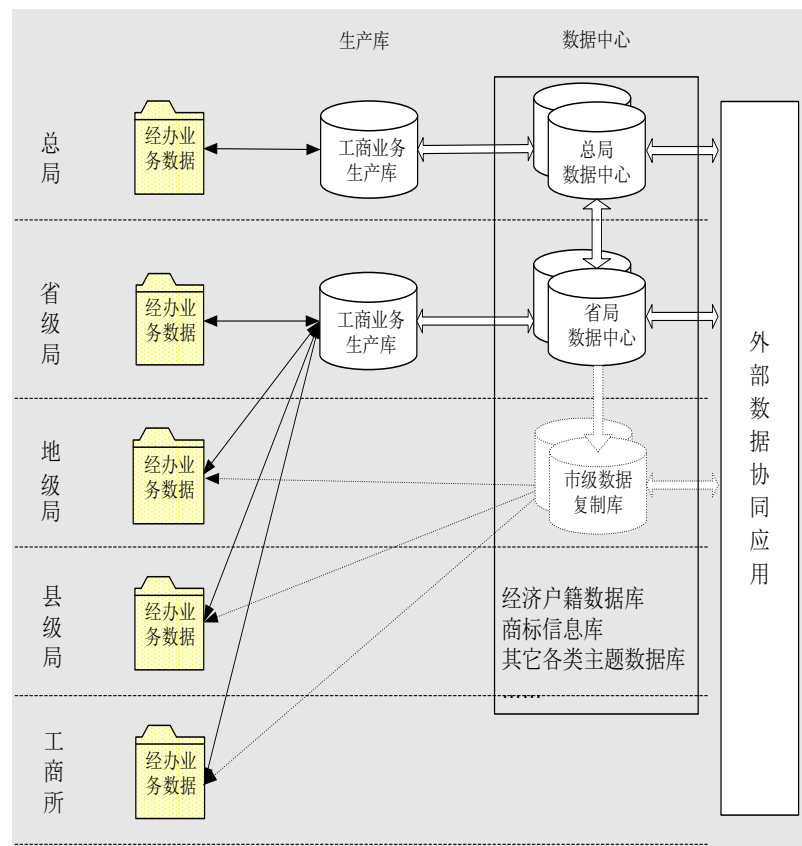
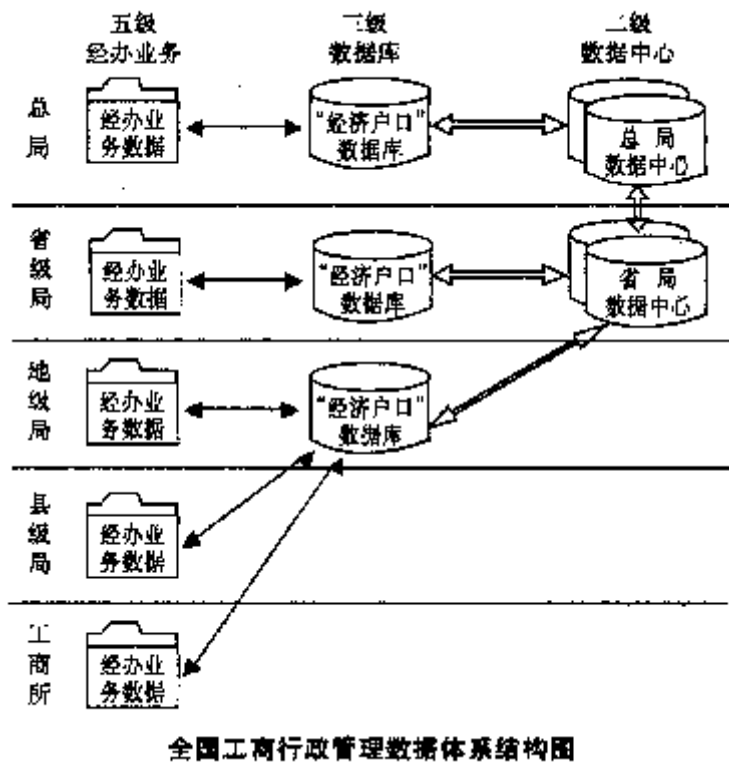
- 必须协调推进标准建设、业务系统建设和数据中心建设



标准规范应用、应用**产生数据**、**数据促进应用**

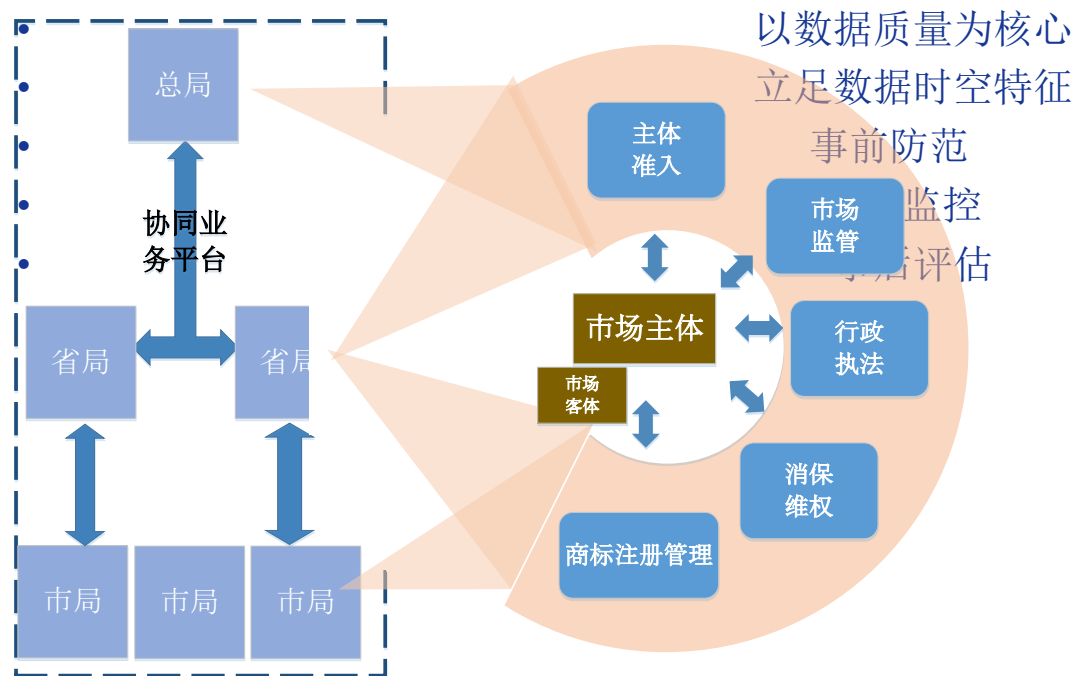
数据中心建设基本经验总结（2）

- 必须坚持走物理集中的道路



数据中心建设基本经验总结（3）

- 必须建立工商特色数据治理体系



数据治理体系与“多业务、多层级、多系统”紧密关联

数据中心建设基本经验总结（4）

- 必须大力推进数据应用

- 数据必须能够支撑领导决策分析，尤其是与处理好数据与统计的关系，否则数据分析将受限
- 数据必须能够支撑监管与服务，否则数据的价值就不能更好的发挥
- 只有数据应用的引领才能更好的推进数据中心建设

数据中心建设的问题和难题

- 数据中心建设具有金字塔特征，牵涉面广，不单纯是一个技术问题，是一个管理问题，目前缺乏约束力
- 数据中心建设符合木桶原理，牵涉到31省，往往短板制约整体效果
- 目前两级数据中心数据管控能力不足、运行支撑能力不足，总局省局工作耦合度高，数据上报不及时、不完整、不稳定、不一致等问题
- 统计报表与数据中心生成数据不一致，制约了数据分析工作的开展
- 08年建设至今基础设施没有升级过，性能受限

小结





3

工商数据中心改造的思路

国家大数据战略要求

国务院办公厅关于促进电子政务协调发展的指导意见（国办发〔2014〕66号）

- 进一步加快人口、法人单位、空间地理、宏观经济等国家基础信息资源库共建共享；
- 推进信息资源共享共用和数据开放利用

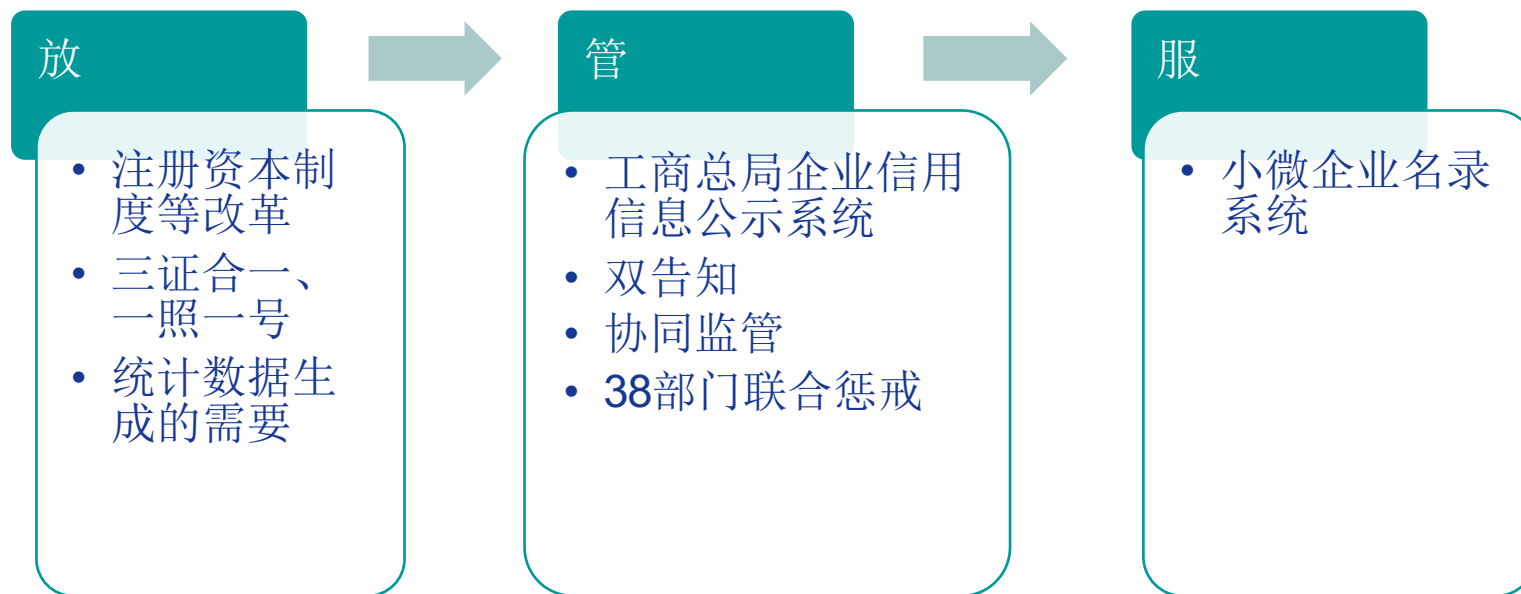
国务院办公厅关于运用大数据加强对市场主体服务和监管的若干意见（国办发〔2015〕51号）

- 加强和规范政府数据采集。
- 建立政府信息资源管理体系
- 高效采集、有效整合、充分运用政府数据和社会数据

国务院关于印发促进大数据发展行动纲要的通知（国发〔2015〕50号）

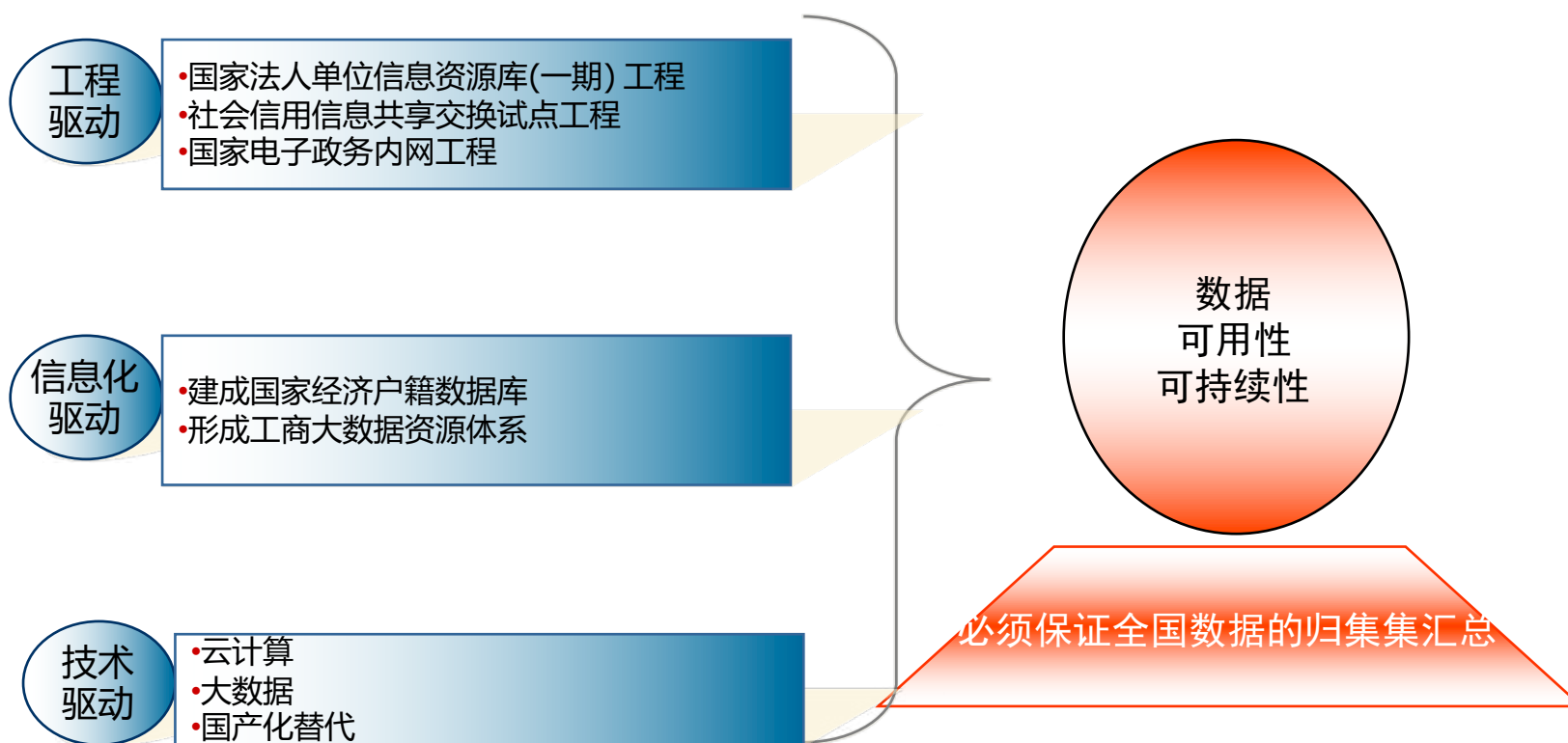
- 加快政府数据开放共享，推动资源整合，提升治理能力
- 政府数据资源共享开放工程
- 国家大数据资源统筹发展工程
- 政府治理大数据工程

国家深化商事制度改革要求



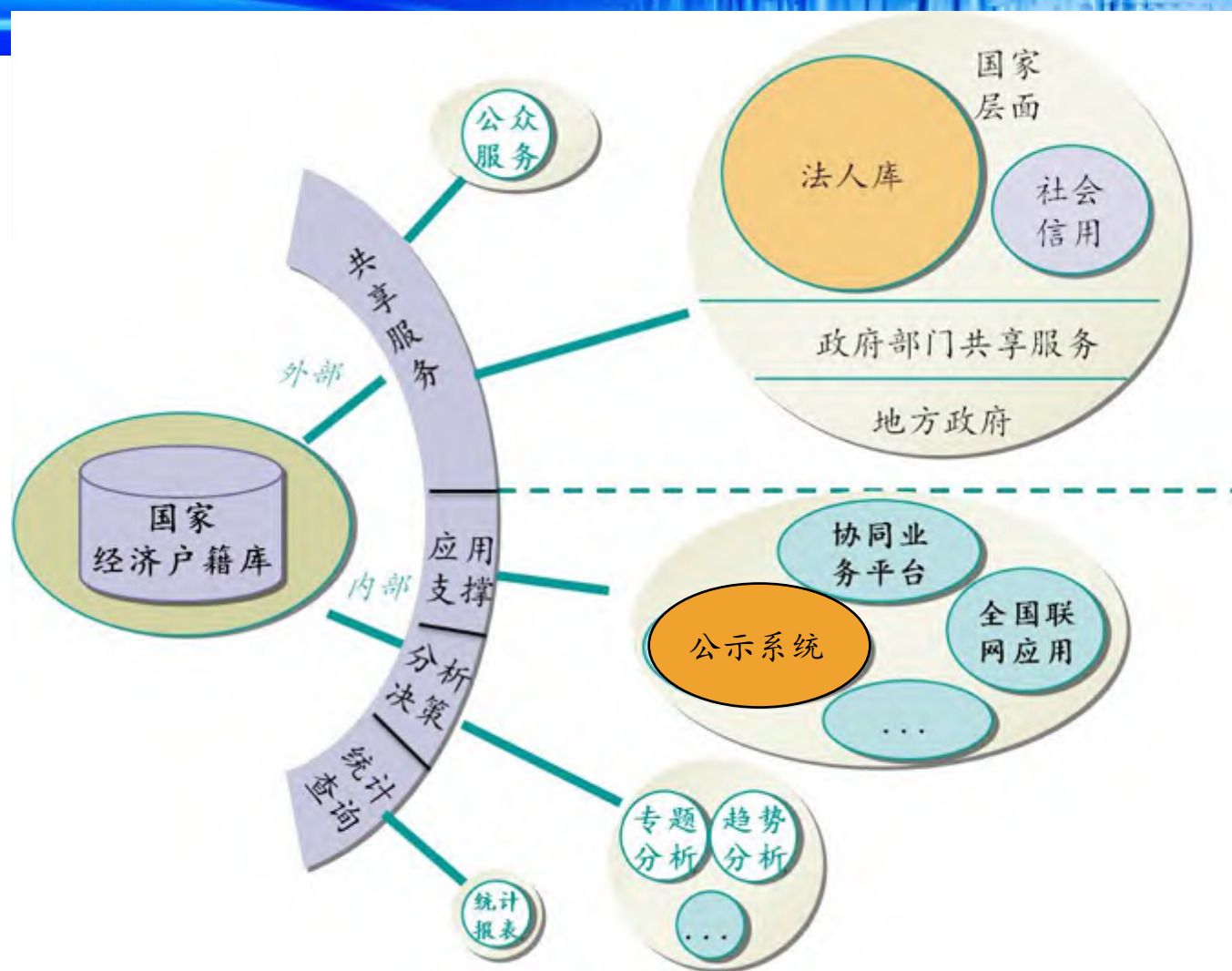
工商部门正面临着深化商事制度改革、加强事中事后监管、适应体制调整的新形势。

信息化建设要求



小结

16/11/2



数据中心升级改造原则

- ◆应用导向、突出服务
- ◆统筹规划、分步实施
- ◆保护投资、深化完善
- ◆技术前瞻、实用为本
- ◆着眼未来、注重运维

数据中心升级改造总体思路

- 从分析型主题库建设向数据仓库、生产系统转变
- 从满足工商应用向满足国家、部门、社会应用转变
- 从单一数据质量管理向构建数据治理体系转变
- 从传统IOE架构向分布式、云计算模式转变

数据中心升级改造总体思路



- 源数据丢失
- 增量异常
- 未正常上报
- 不上报部分业务数据
- 数据与文档不同步
- 源表锁定

数据源监控
数据同步机制

- 清洗转换周期长
- 源表锁定
- 上报与抽取冲突

省局负责标准化
总局制定规格
省局上报监控系统

每日对账
并行计算
基础升级

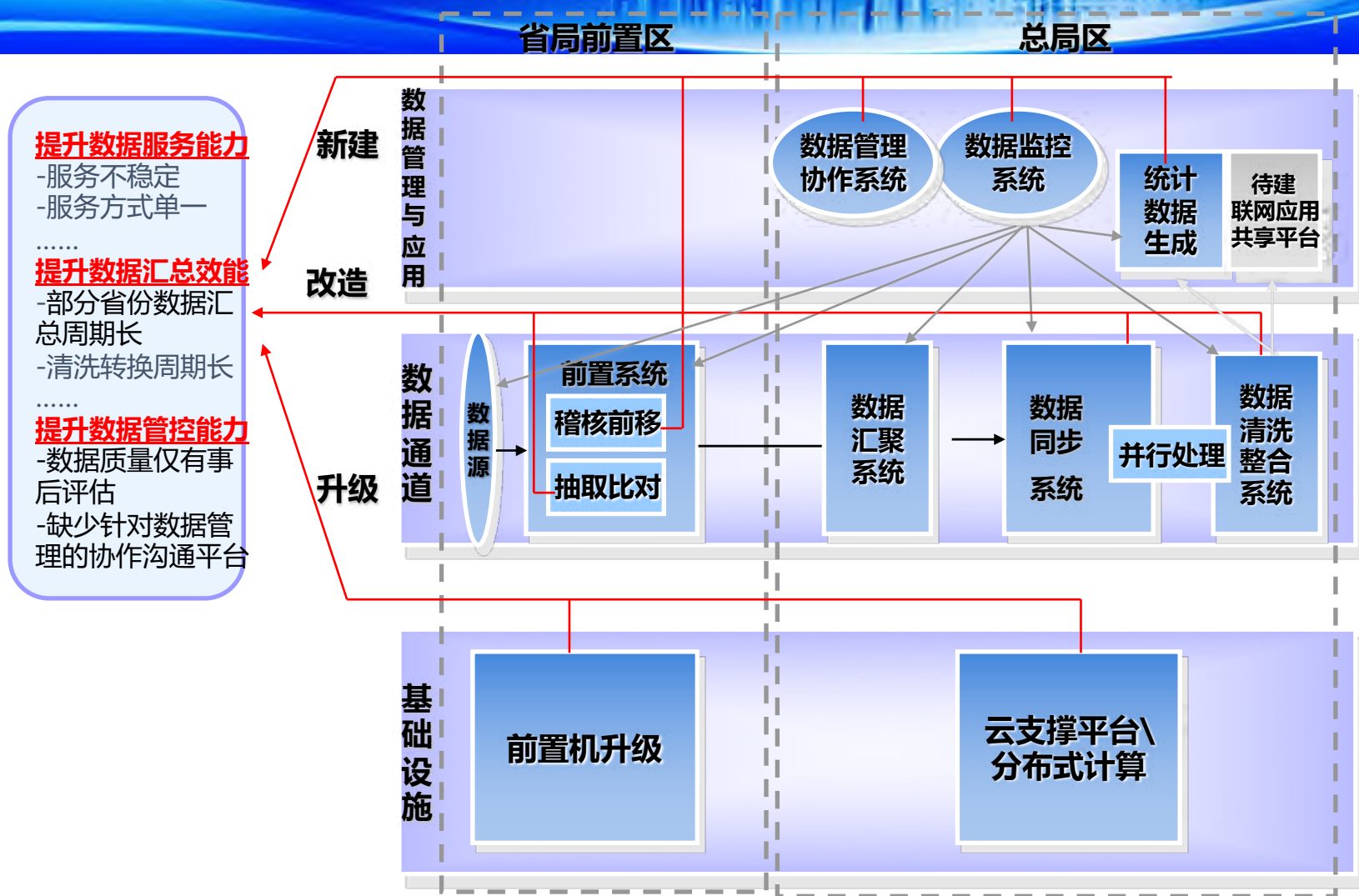
- 没有数据监控手段
- 缺少针对数据管理的协作沟通平台

监控平台
管理协作平台

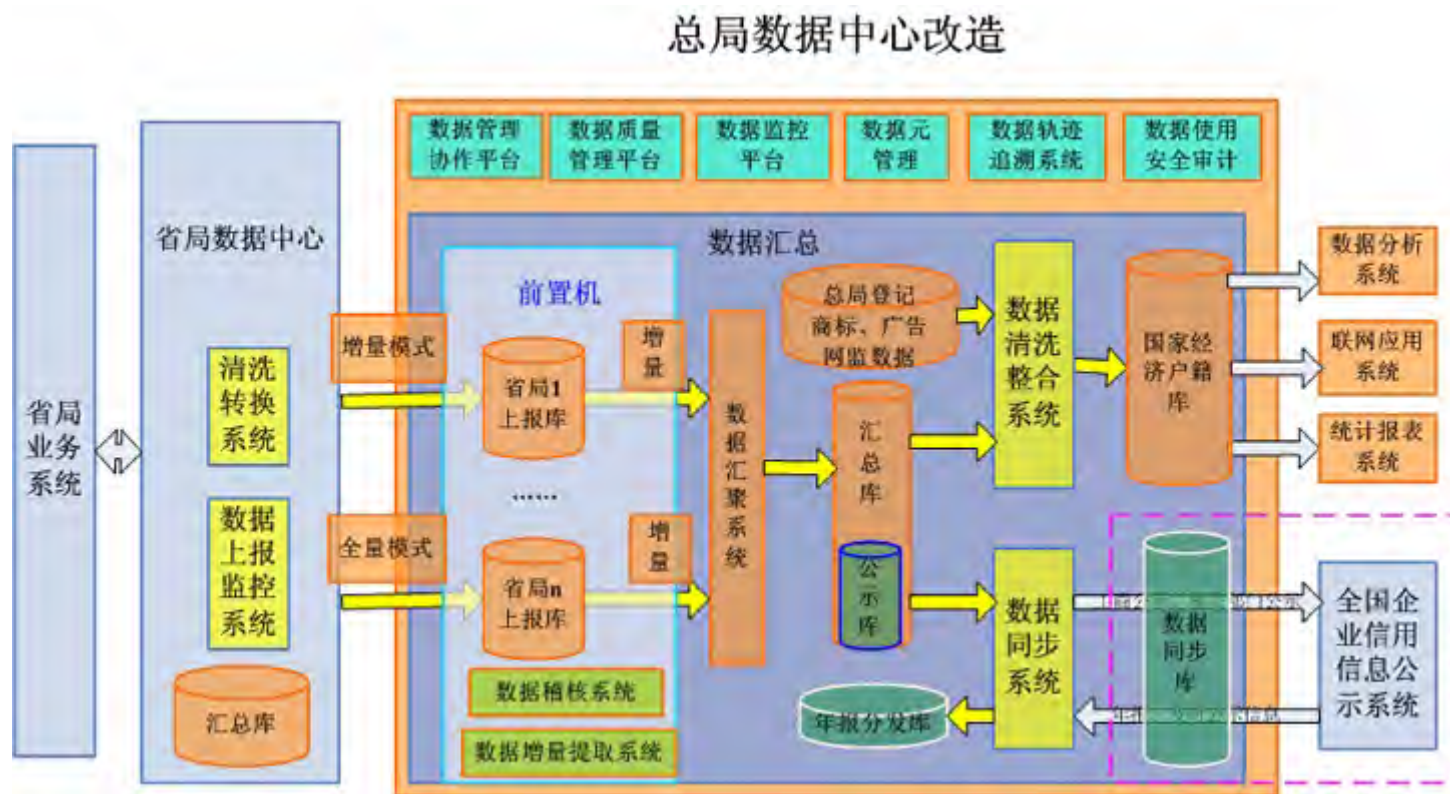
技术路线

- ◆ 分布式并行计算
 - ◆ MPP架构
 - ◆ 数据处理流水线
- ◆ 云服务的关键技术
 - ◆ 资源虚拟化
 - ◆ 数据处理虚拟化
 - ◆ 数据主题化
- ◆ 数据治理关键技术
 - ◆ 元数据、标准、质量、模型管理
 - ◆ 采集、存储、数据仓库

数据中心系统构成视图



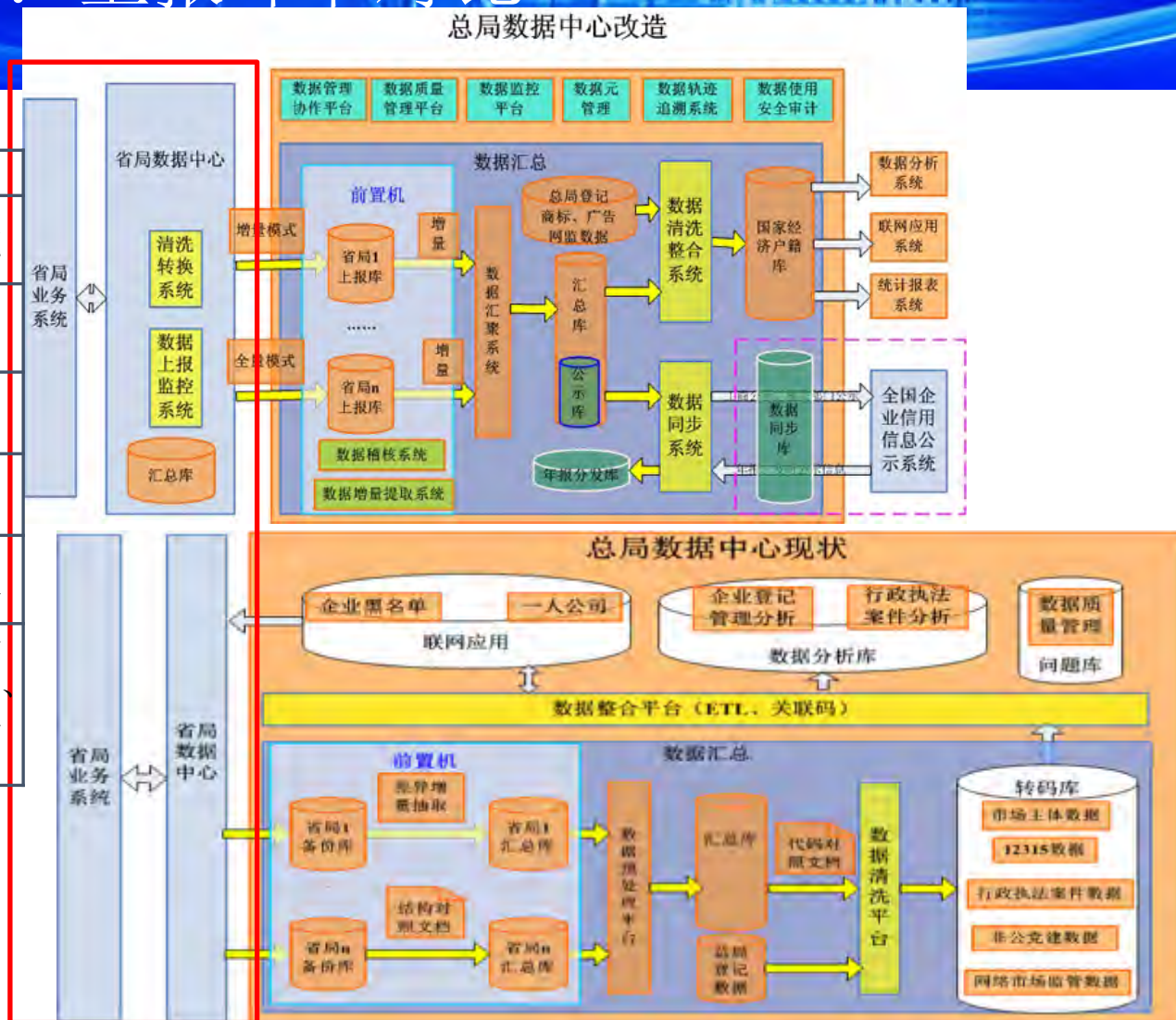
数据中心构成视图



整体优化：分布式数据库、分布式处理、并行计算、云计算、基础设施升级
可扩展优化：考虑节点纵向可扩展、业务横向可扩展

关键环节0：上报环节对比

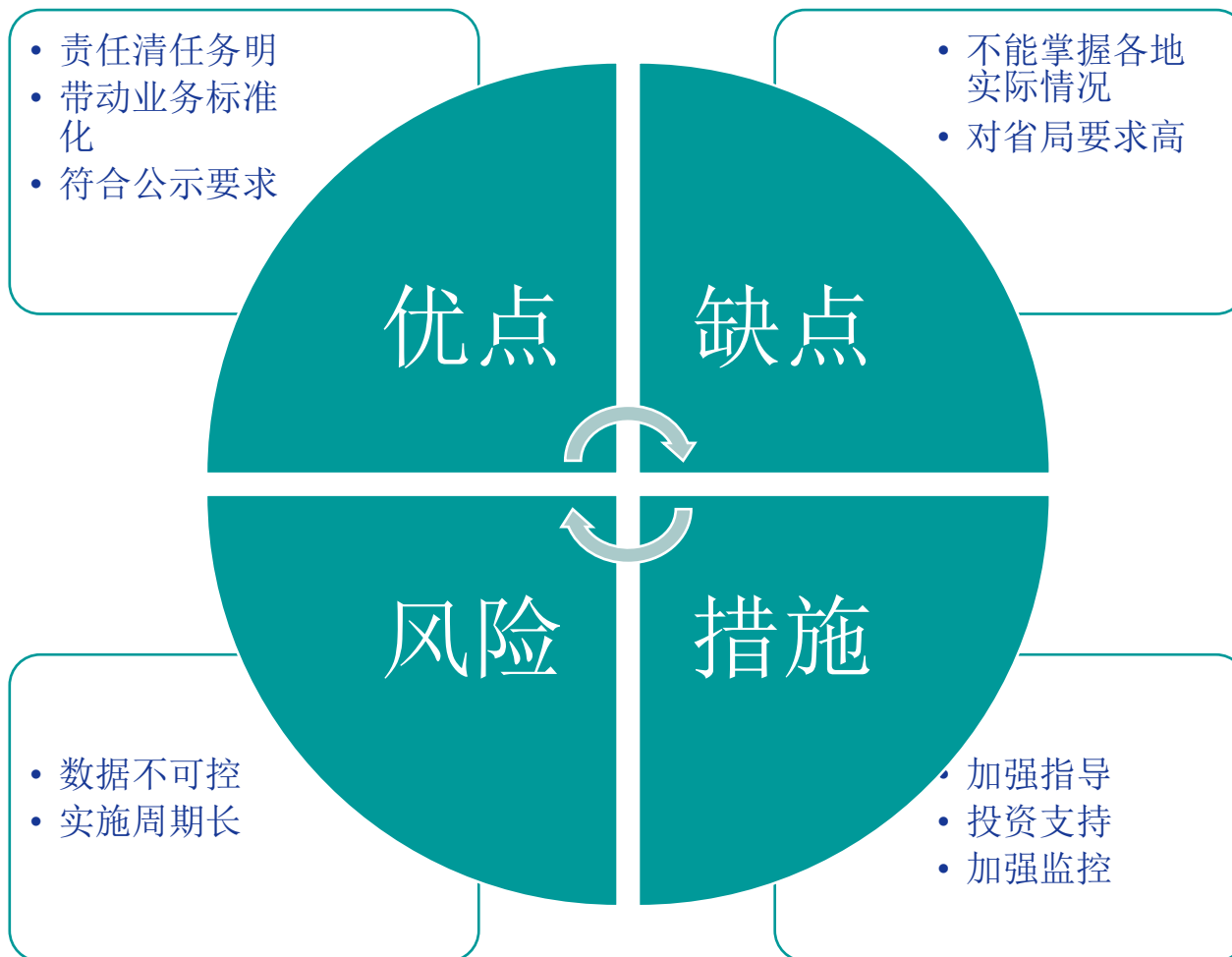
对比	现状	改造
数据格式	生产库格式	总局格式
数据库类型	传统	分布式
上报模式	全量	增量
上报管理	无监控	有监控
上报周期	每天、每周	4h、每天
清洗转换	无监控	有、公示部分转义、统计部分聚合



对省级要求：强化省局数据预处理

- 统一到总局标准：31省独立建设、建设年份跨度较大、符合标准程度不一，源数据异构
- 统一时点更新：大集中与两级部署并存，全省大集中22个，计划单列市、副省级单独建设
- 统一数据库类型：有的省存在Oracle、db2、sybase、sql server、infomix其中多种数据库
- 统一汇集到经济户籍库：一般是登记办案、12315、农资、网监分别独立建库

预处理要求：省局负责数据标准化



两级数据中心分工

16/11/2

✕ 总局数据中心

- ✕ 国家经济户籍库
- ✕ 制定数据汇总标准
- ✕ 数据稽核反馈
- ✕ 数据轨迹追溯
- ✕ 年报数据分发
- ✕ 统一监控

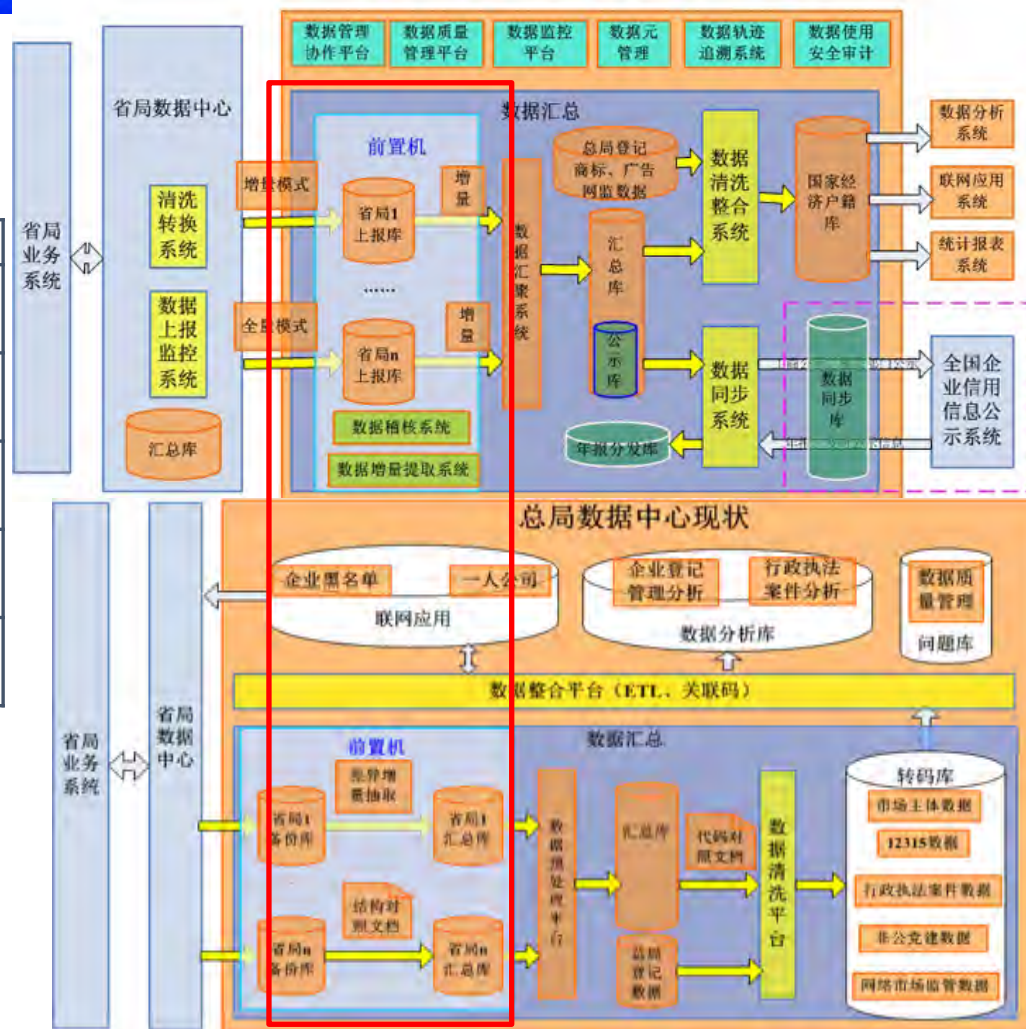
✕ 省局数据中心

- ✕ 省级经济户籍库
- ✕ 数据标准化
- ✕ 数据上报监控
- ✕ 数据质量把关
- ✕ 年报数据下载
- ✕ 沟通协作

关键环节1：采集环节对比

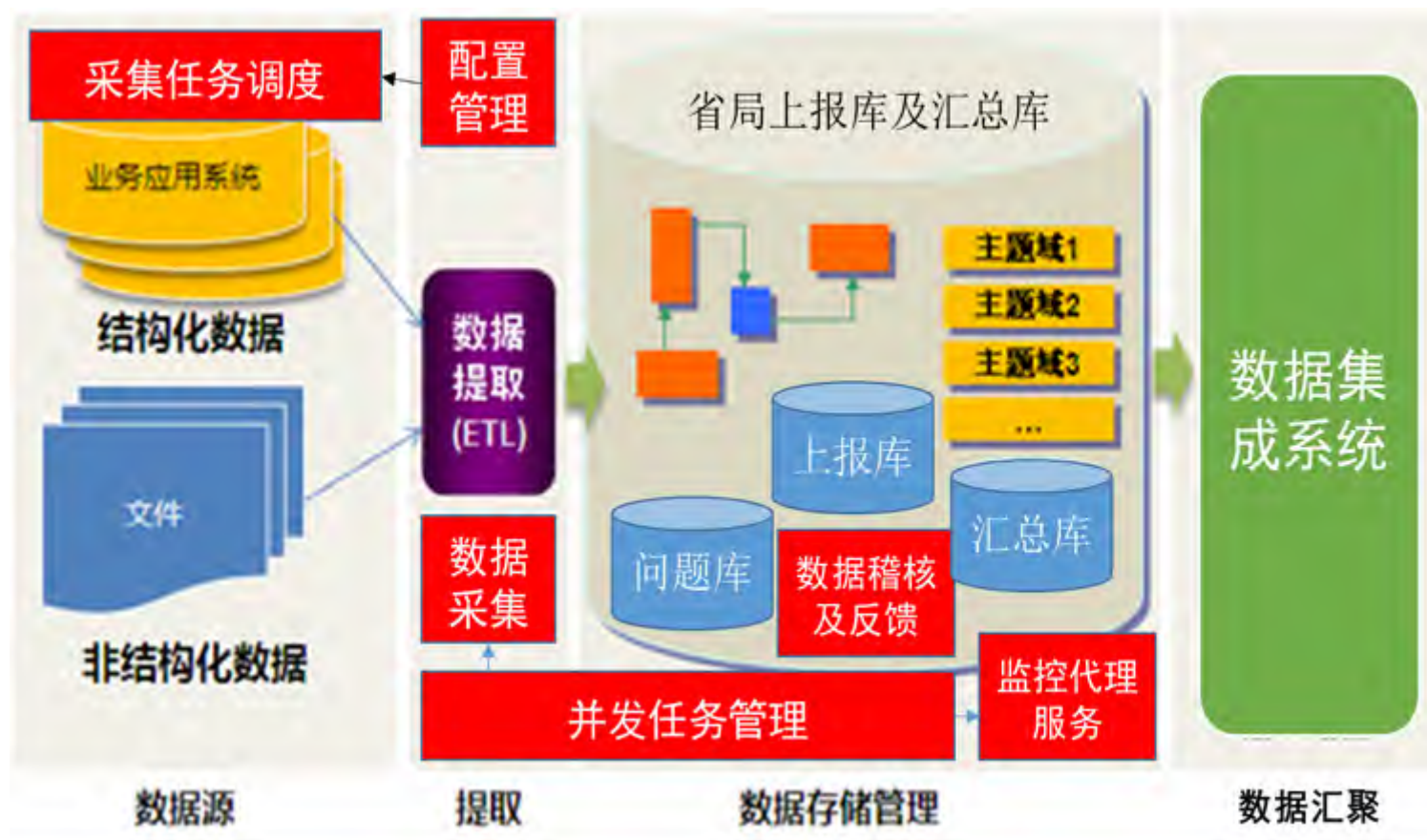
对比	现状	改造
采集方式	异构转换	同构复制
数据库类型	传统	分布式
稽核模式	少量	全面
稽核反馈	无	新增
优先级采集	无	公示优先

总局数据中心改造



前置系统：“大前置”思路

16/11/2



前置机配备

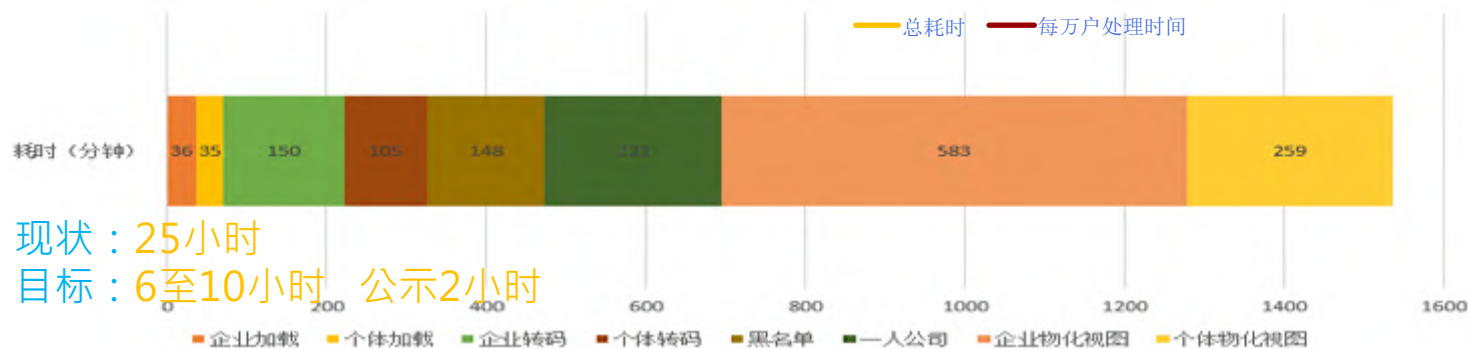
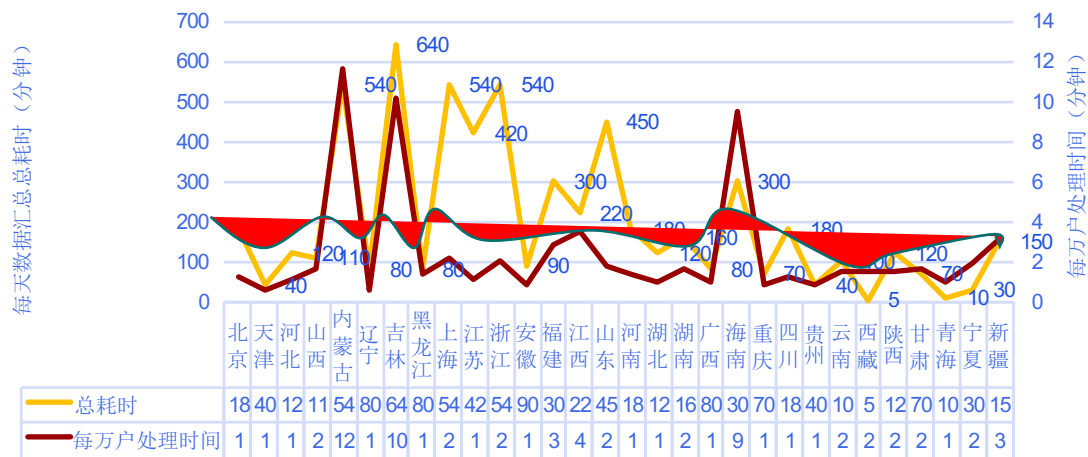
16/11/2

广东全量：3天

数据迁移：1620

数据稽核：300

数据抽取：1440



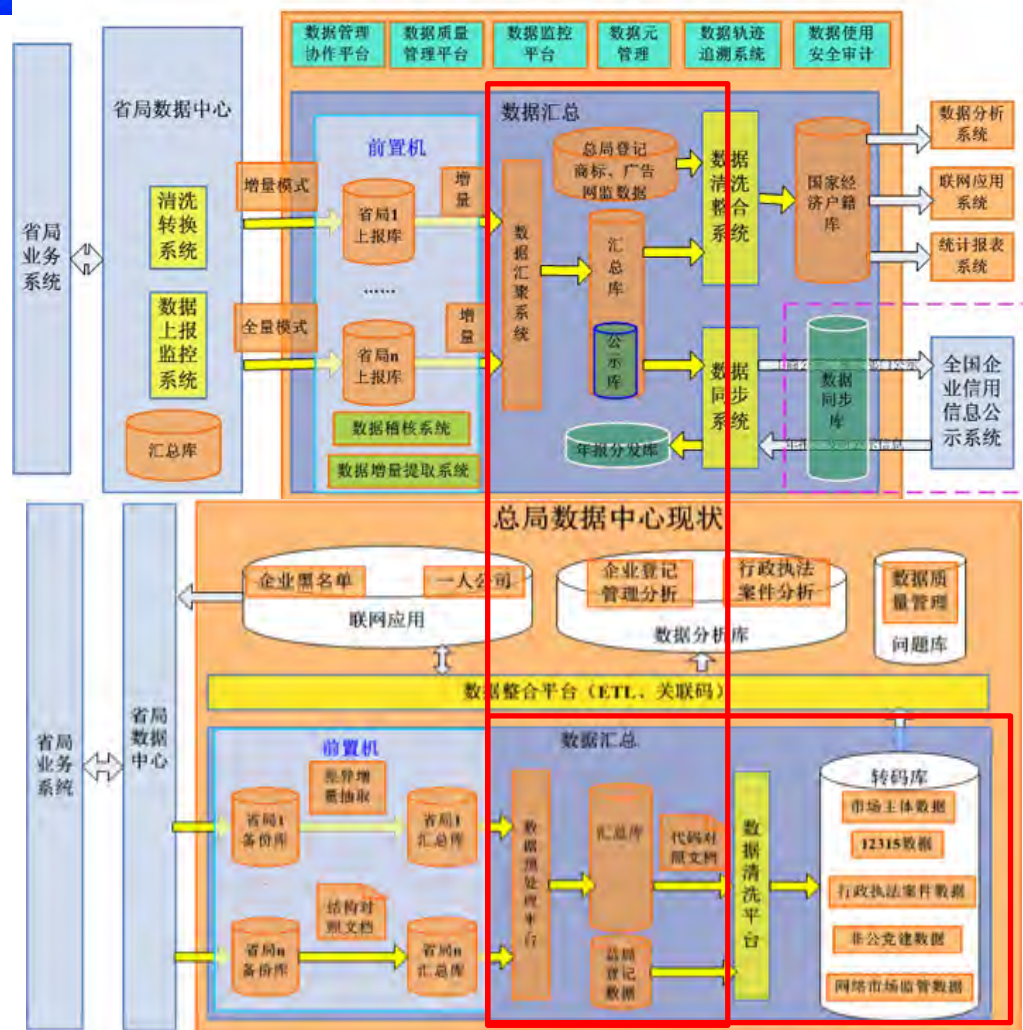
现状：25小时

目标：6至10小时 公示2小时

关键环节2：全国汇总环节对比

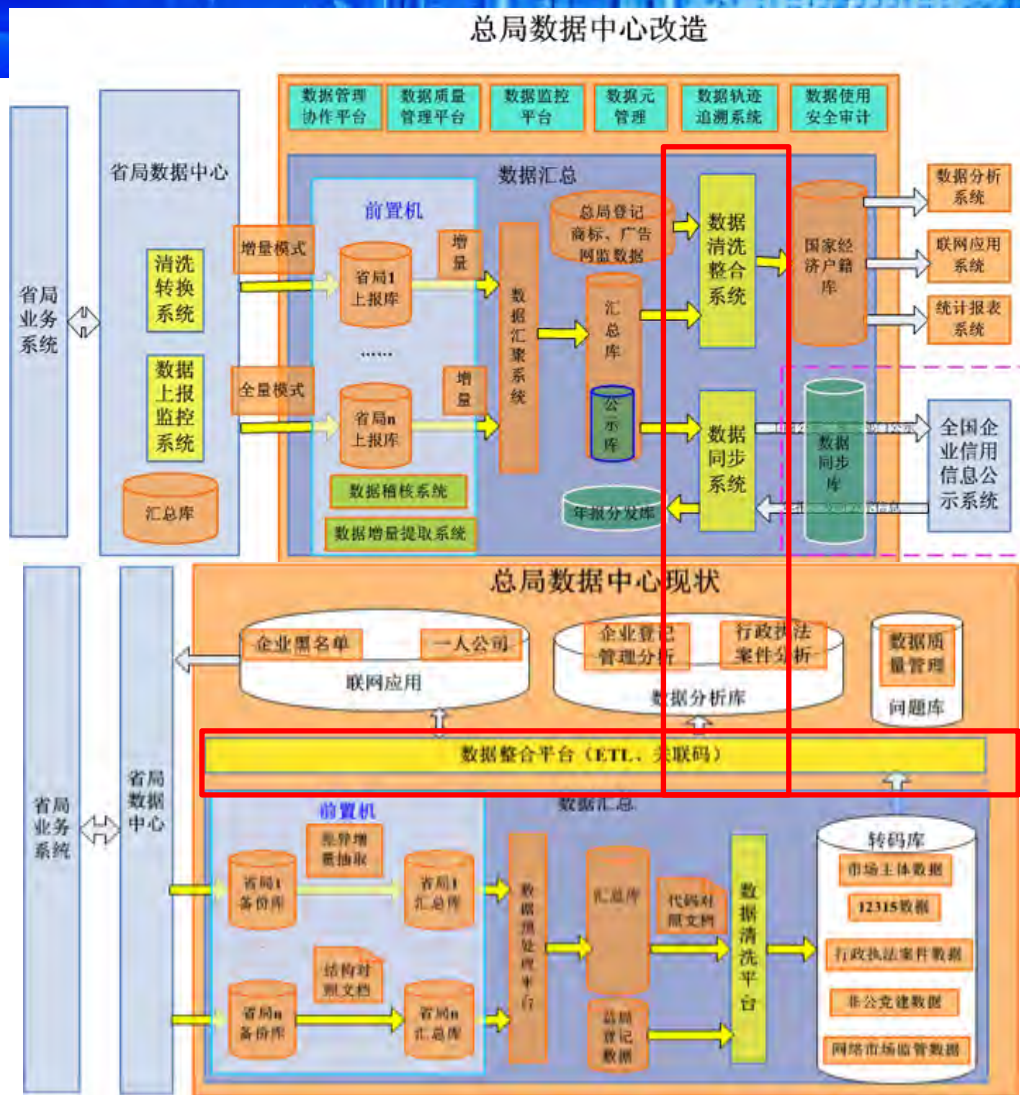
对比	现状	改造
处理方式	主体并行、省份并行	增加主题内并行
数据库类型	传统	分布式
稽核模式	人工检查	自动监测各环节
清洗转换	代码转换	无

总局数据中心改造



关键环节3：数据服务环节对比

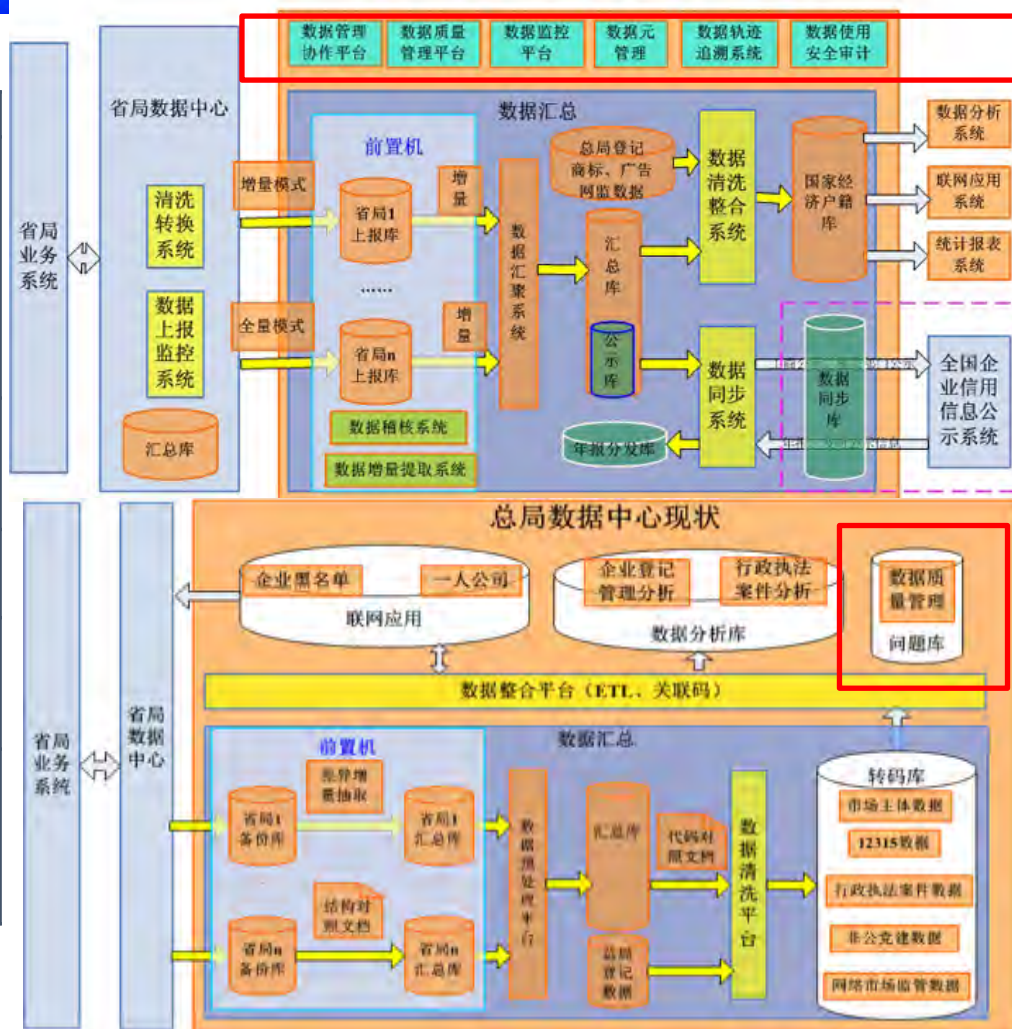
对比	现状	改造
处理方式	人工	自动同步
数据格式	文件	数据库、接口
同步机制	多线程	并行处理
清洗转换	大量	减少
地方共享	黑名单	年报等公示信息



关键环节4：数据管理环节对比

对比	现状	改造
质量管理	终端评价为主, 业务问题与技术问题纠缠	突出前端, 全程管理, 及时反馈, 业务问题为主
追溯管理	关联码	数据仓库切片、全程数据库
安全管理	基础安全	应用安全
可视化	少、不直观	全面监控、易用、便于协作
用户	数据处为主	扩展到业务司局、地方工商局

总局数据中心改造



数据管理平台：提升管控能力



“人员”协作管理、“数据”监控管理

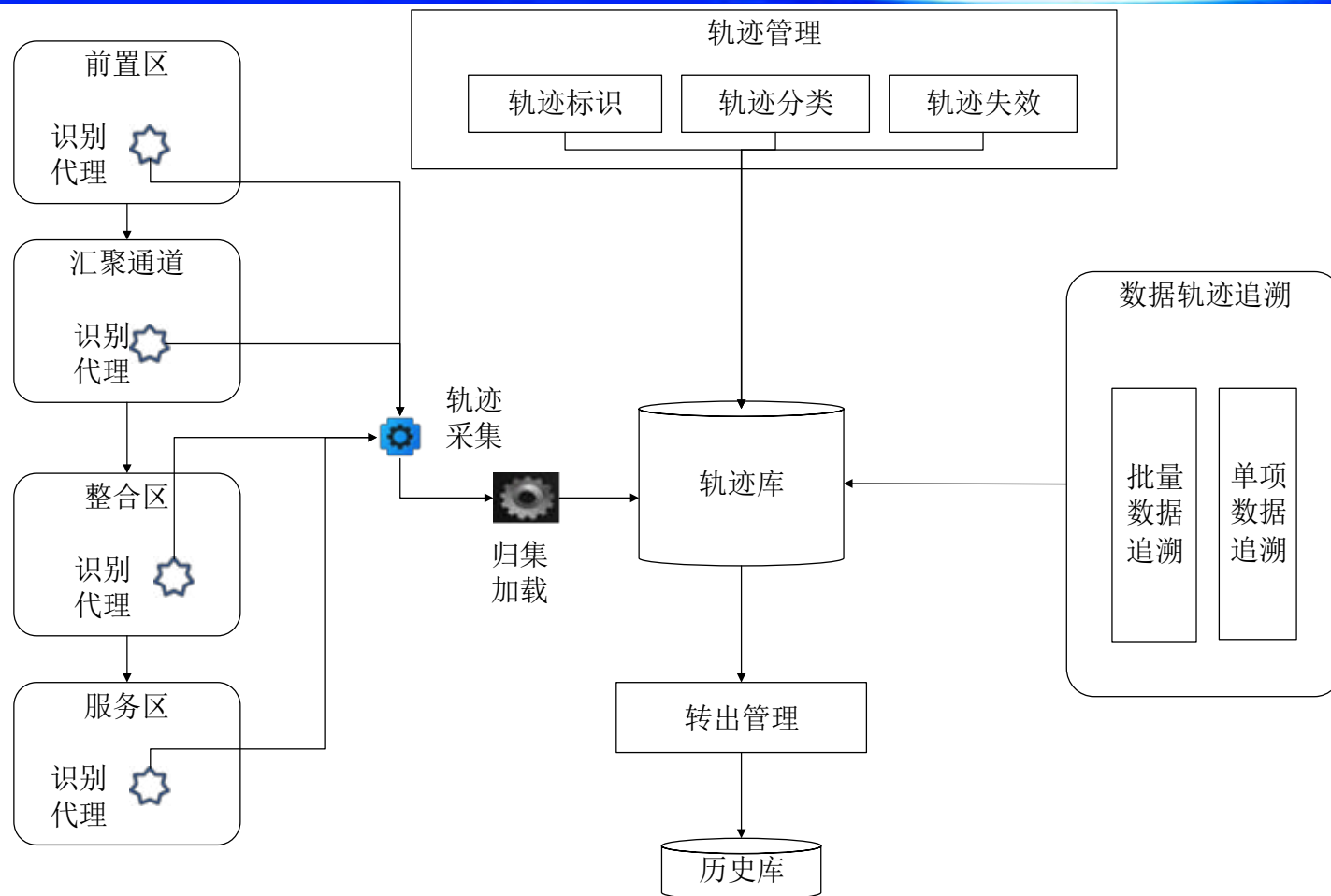
数据管理平台：统一监控系统

16/11/2



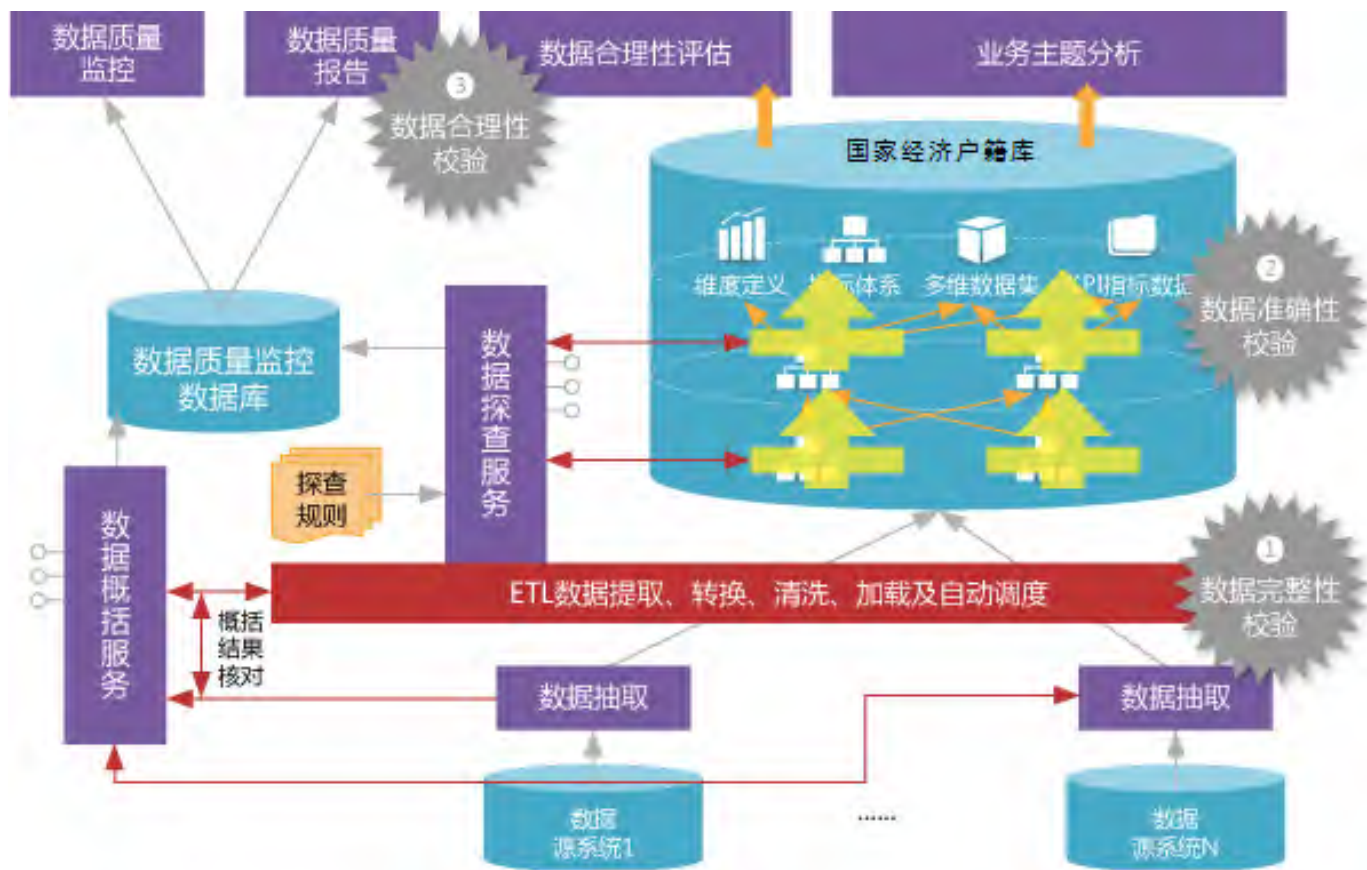
数据管理平台：数据轨迹追溯系统

12/11/2

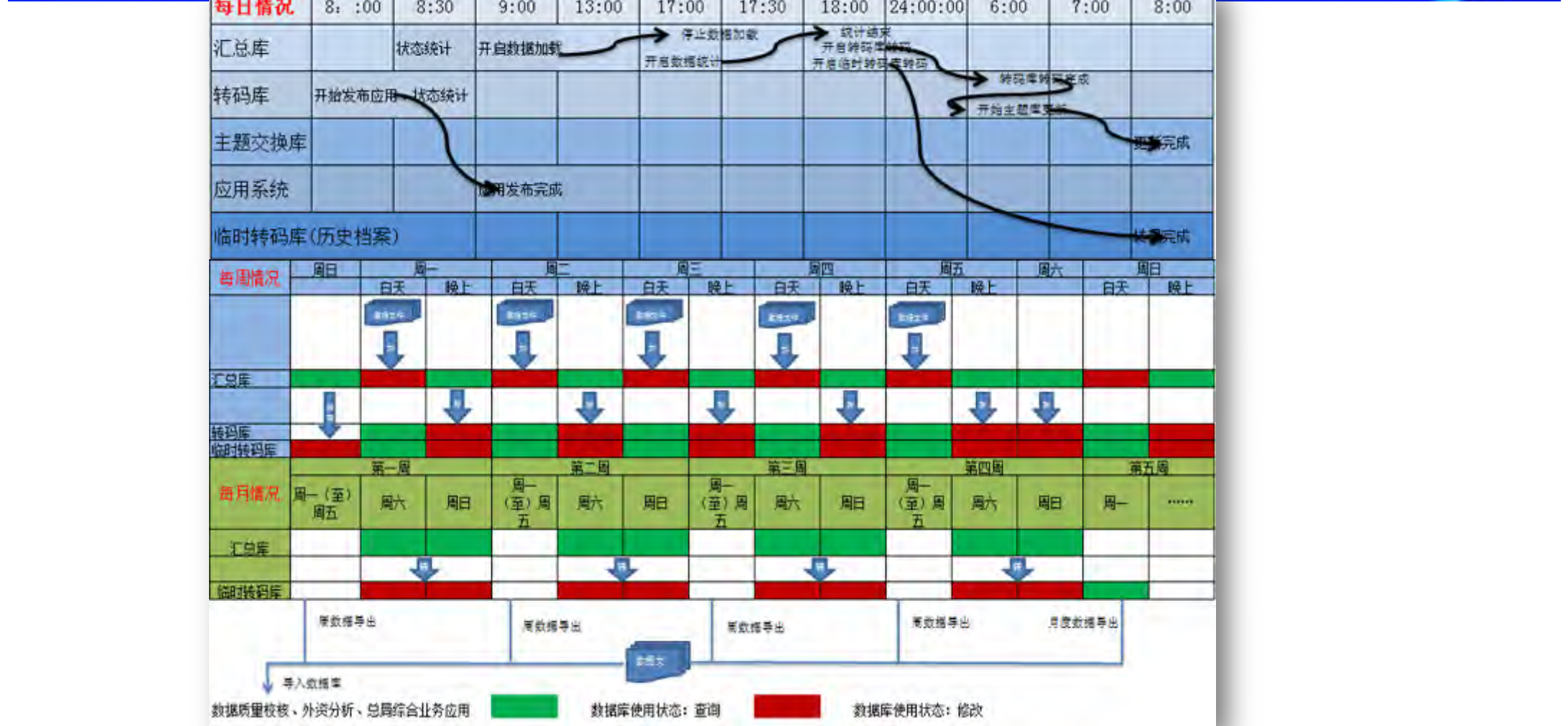


数据管理平台：数据质量管理

16/11/2



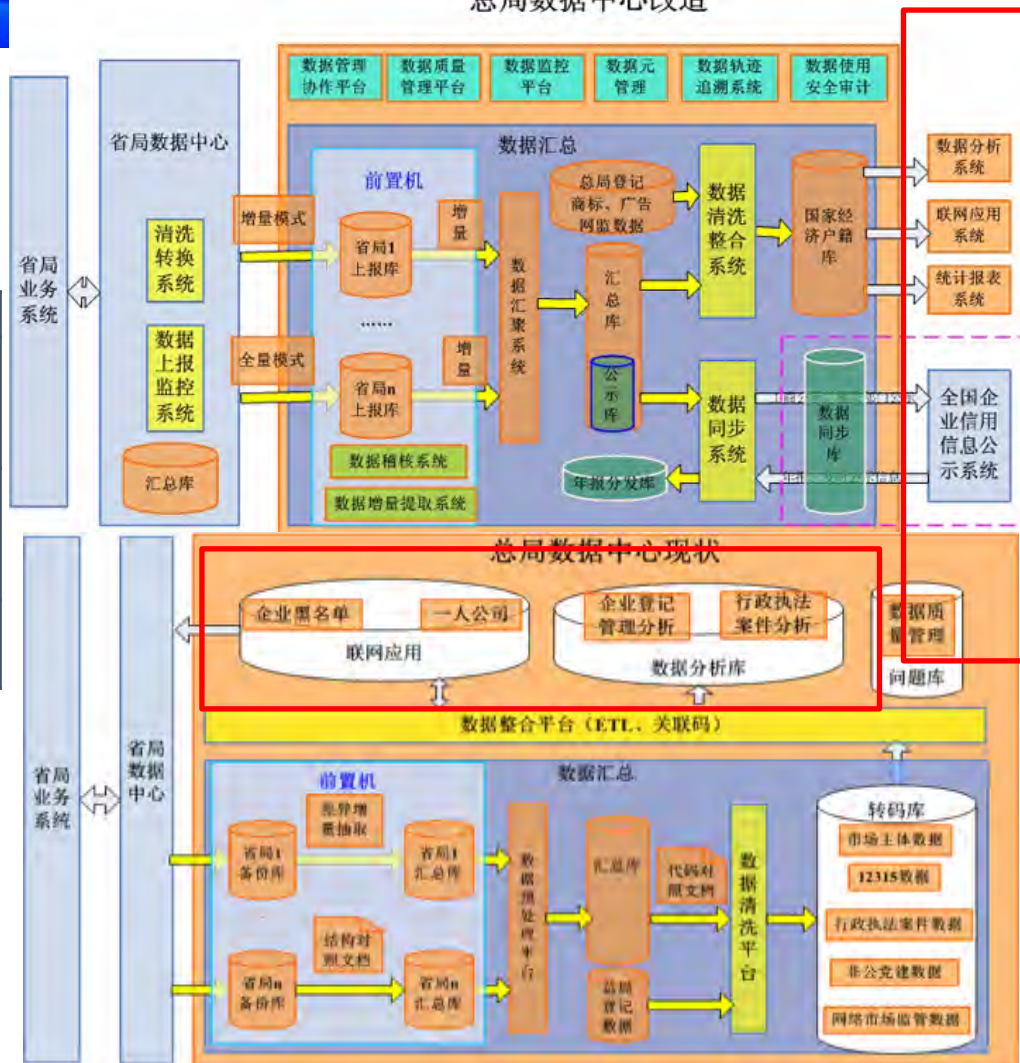
数据管理平台：提升服务能力



关键环节5：数据应用环节对比

对比	现状	改造
统计	时效低、不一致性	快速、结果一致
公示	无	及时、完整、原始一致
专项查询统计	临时性居多	系统化

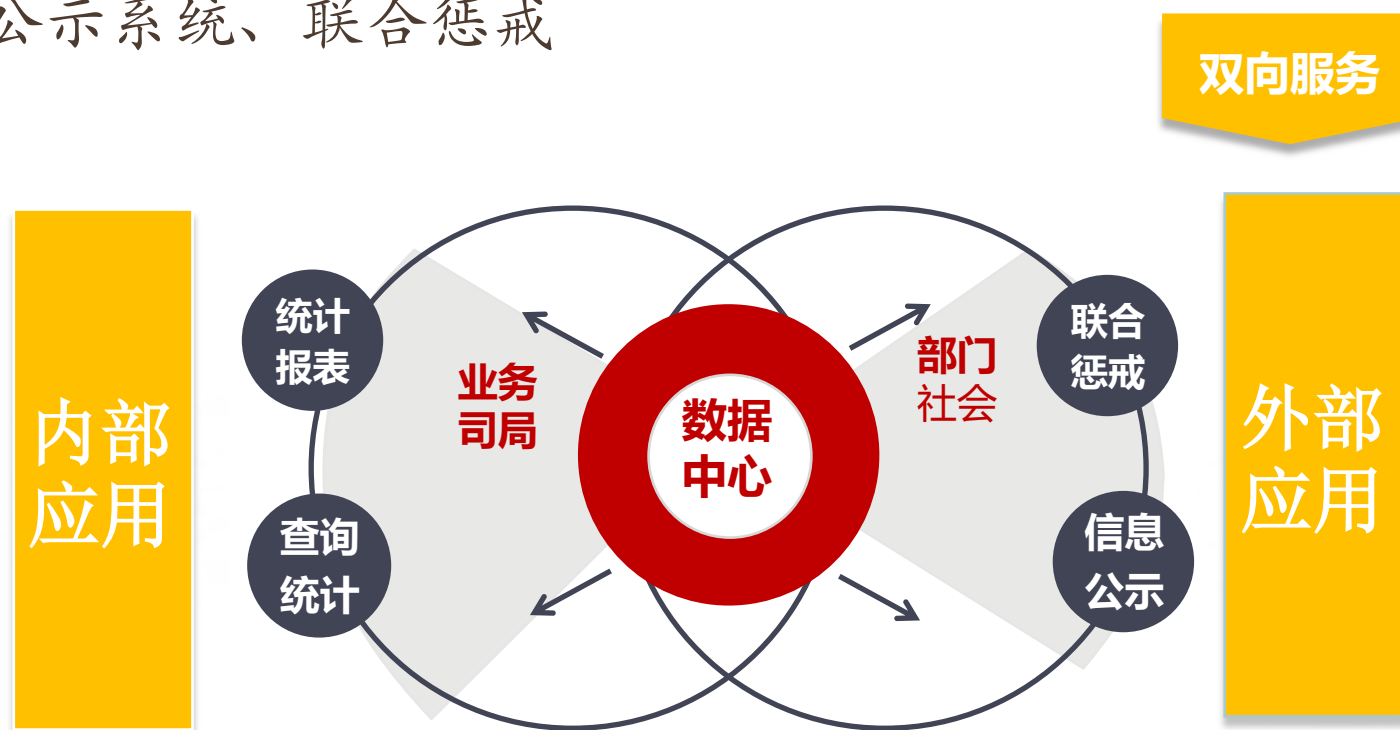
总局数据中心改造



建设两类核心应用系统

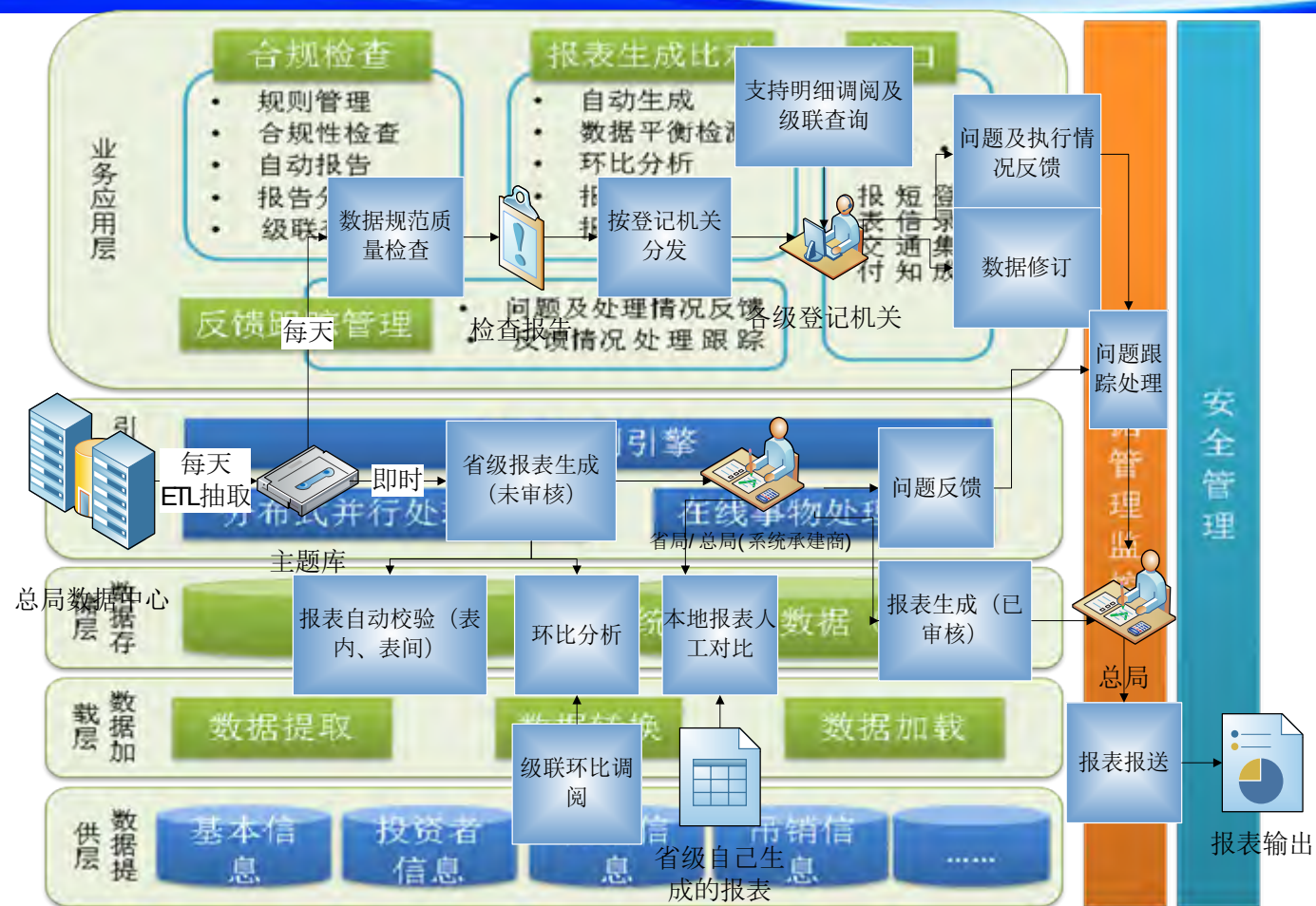
16/11/2

- 建设统计数据生成系统
- 建设业务司局专项查询统计系统
- 支撑公示系统、联合惩戒



统计数据生成系统

16/11/2



业务司局专项查询统计系统

16/11/2



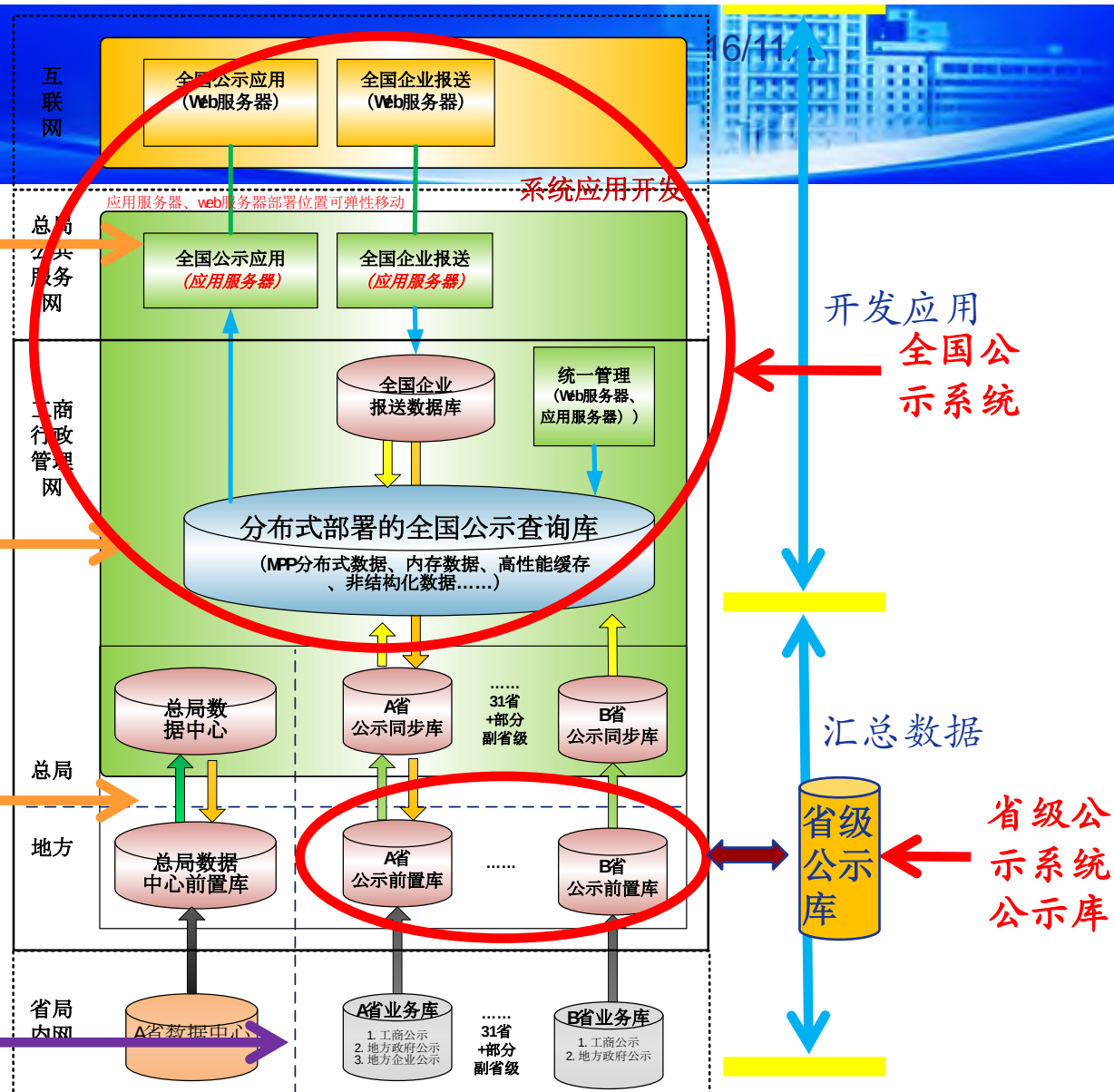
支撑公示

公示应用

分布式数据库以及云支撑平台

公示数据汇总

地方局改造工作

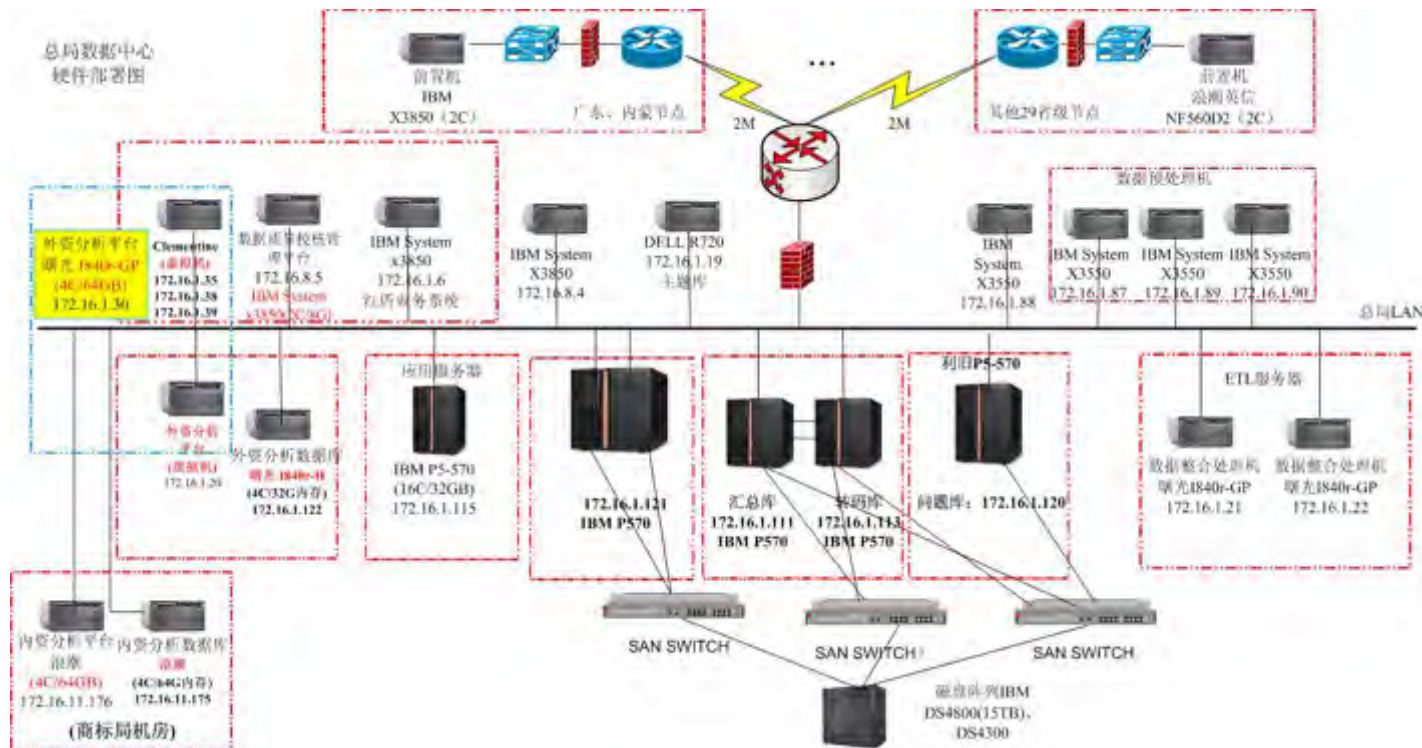


升级改造内容：一云一库两平台三系统



一云：替换“IOE”

- IOE中的I为主：小型机、DS4800、DB2、x3850、Datastage、SPSS、congnos、WebSphere、MQ
- 辅助O：oracle、OWB

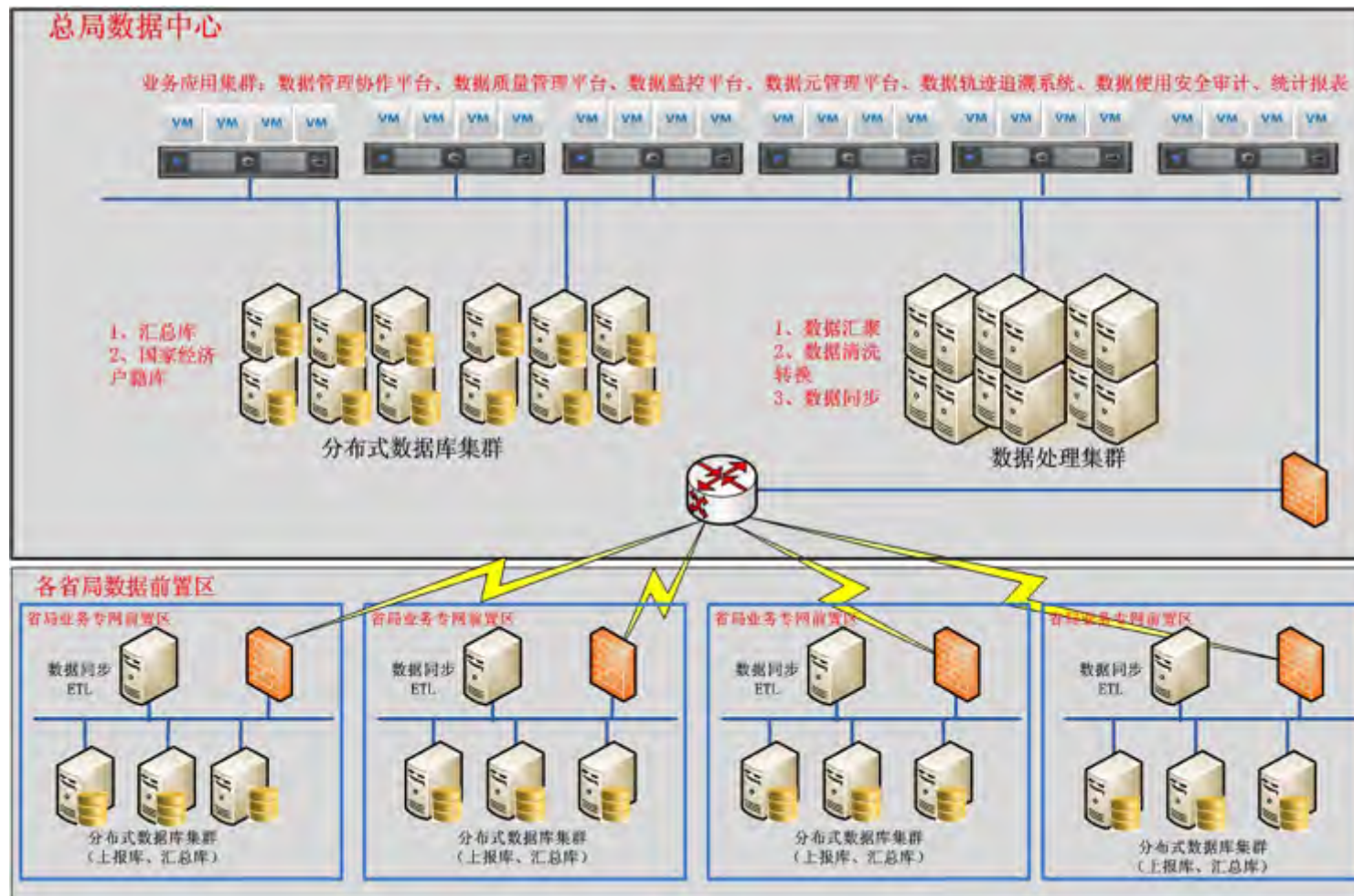


一云：PaaS层升级

16/11/2

- MPP数据库：share-nothing分布式体系结构、支持sql89 sql92 sql99、与Oracle等数据库兼容
- 应用中间件：要支持国产MPP数据库、集群版
- 数据同步工具：支持日志、时间戳等多种方式
- ETL工具：支持多数据源、集群版
- 此外总局将尝试使用hadoop、内存计算等技术进行快速统计分析

一云：IaaS层升级



一库：制定国家经济户籍库数据汇总标准

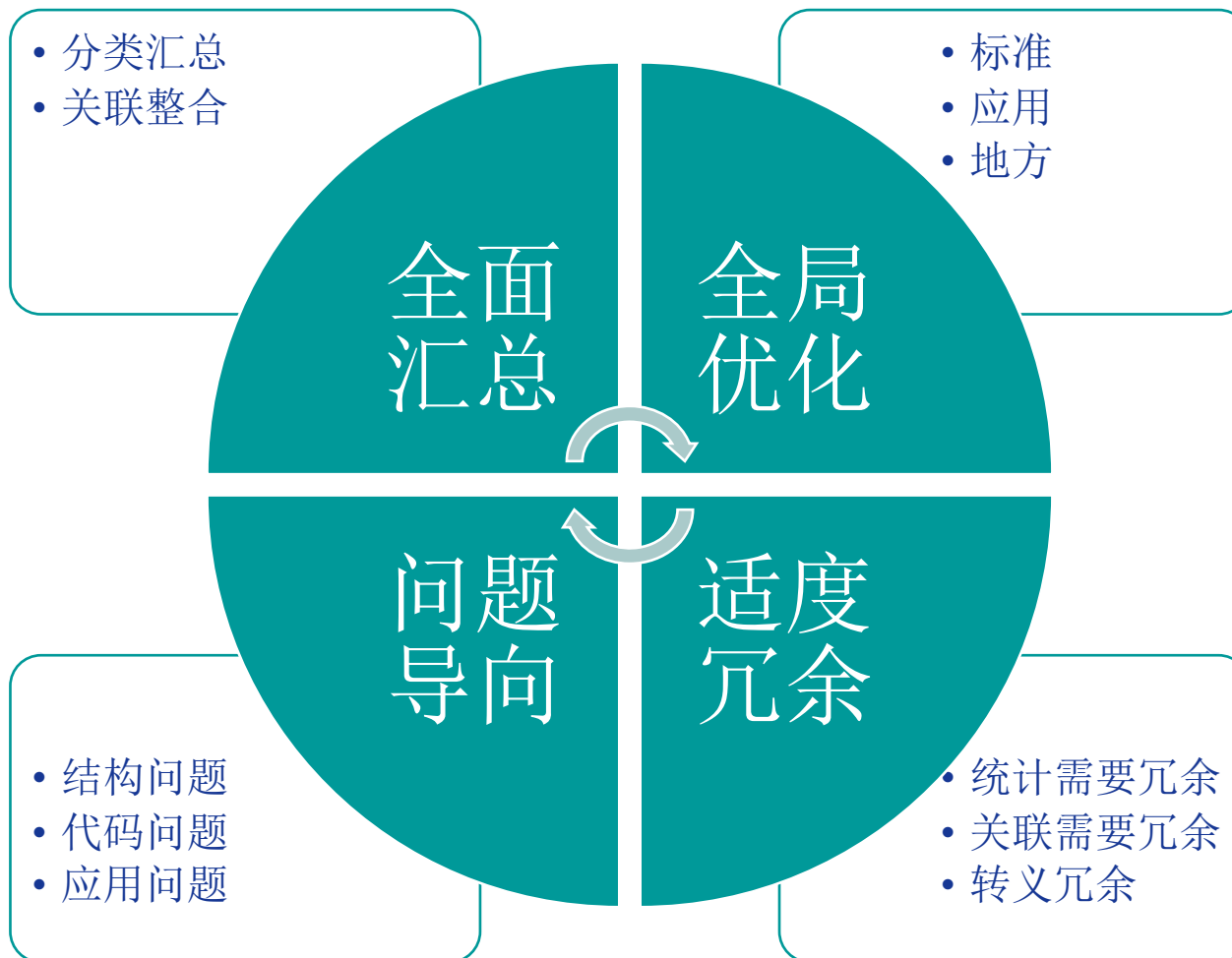
●●● 梳理数据标准规范
●●● 梳理文书格式

●●● 梳理统计报表
●●● 梳理业务需求

●●● 梳理数据中心
●●● 梳理地方业务系统

国家经济
户籍库数
据汇总标
准

一库：数据汇总标准设计原则



□ 总局

□ 一云

□ 一库

□ 两平台

□ 三系统

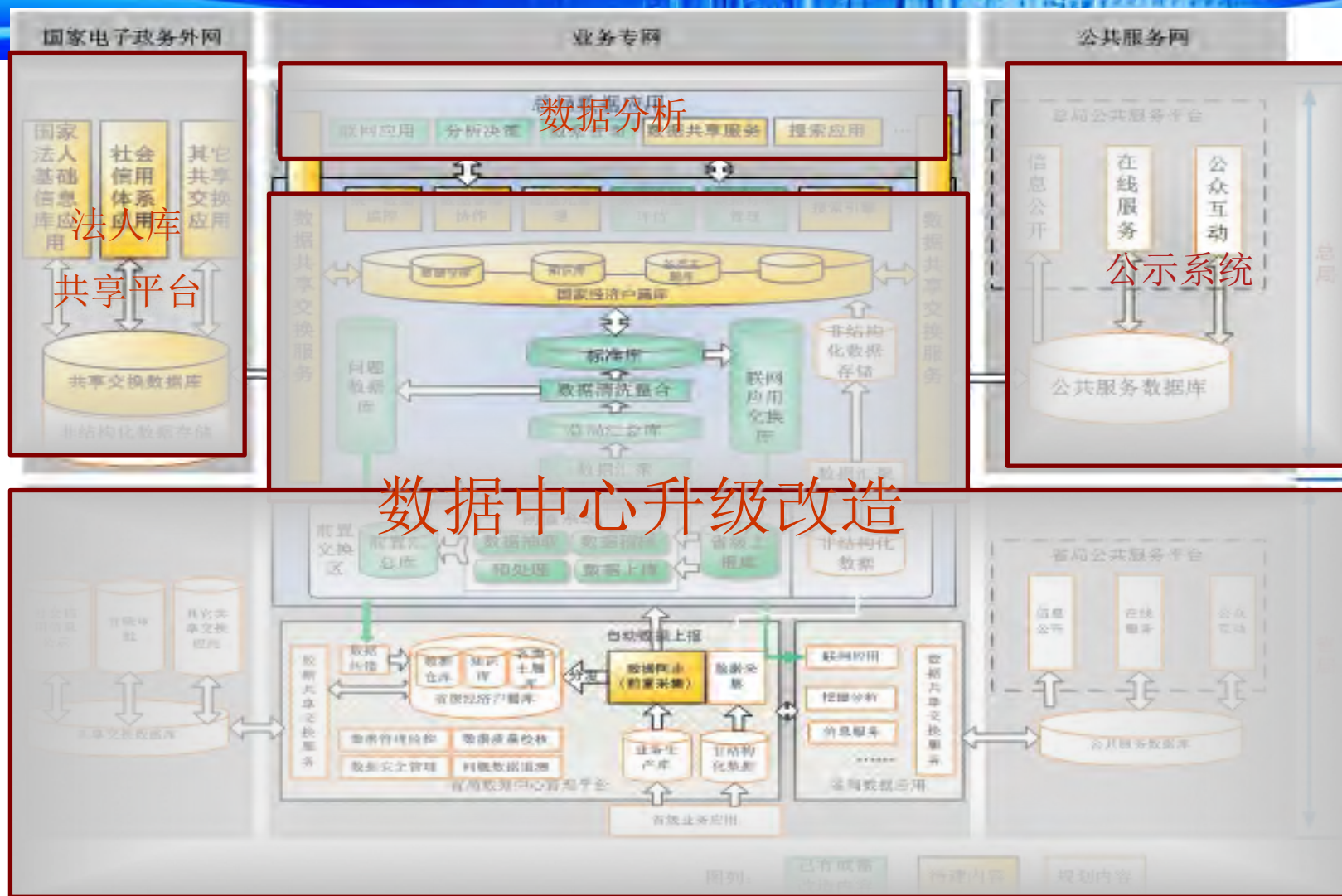
□ 省局

□ 数据中心

□ 业务系统



后续建设内容：法人库、共享平台、分析平台

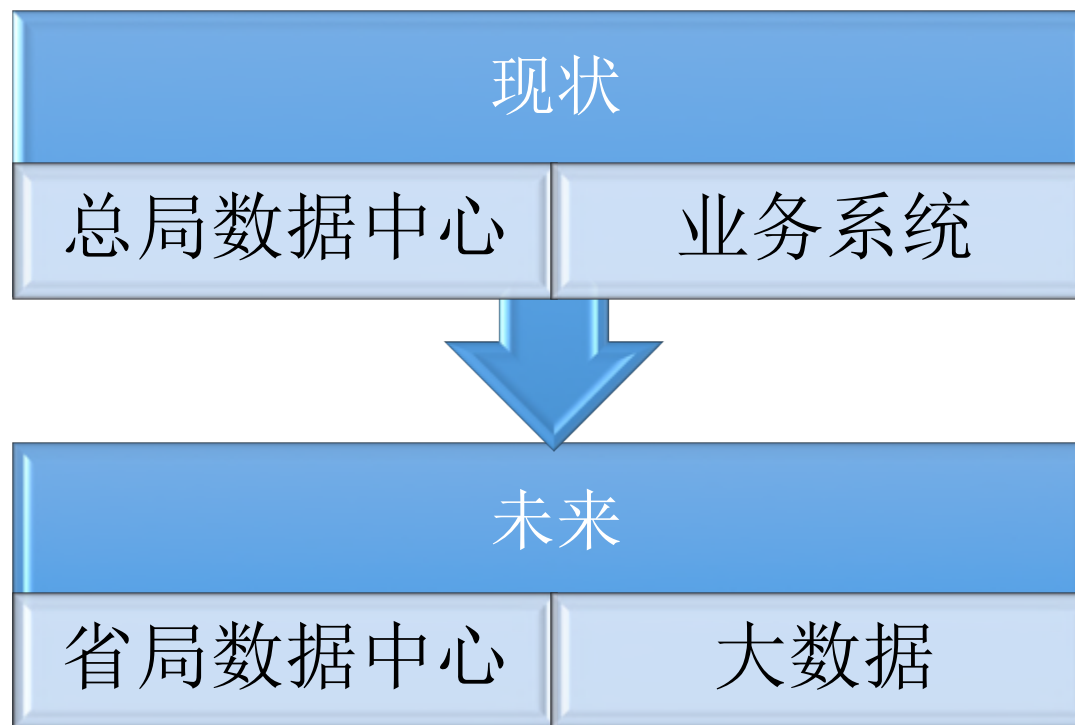




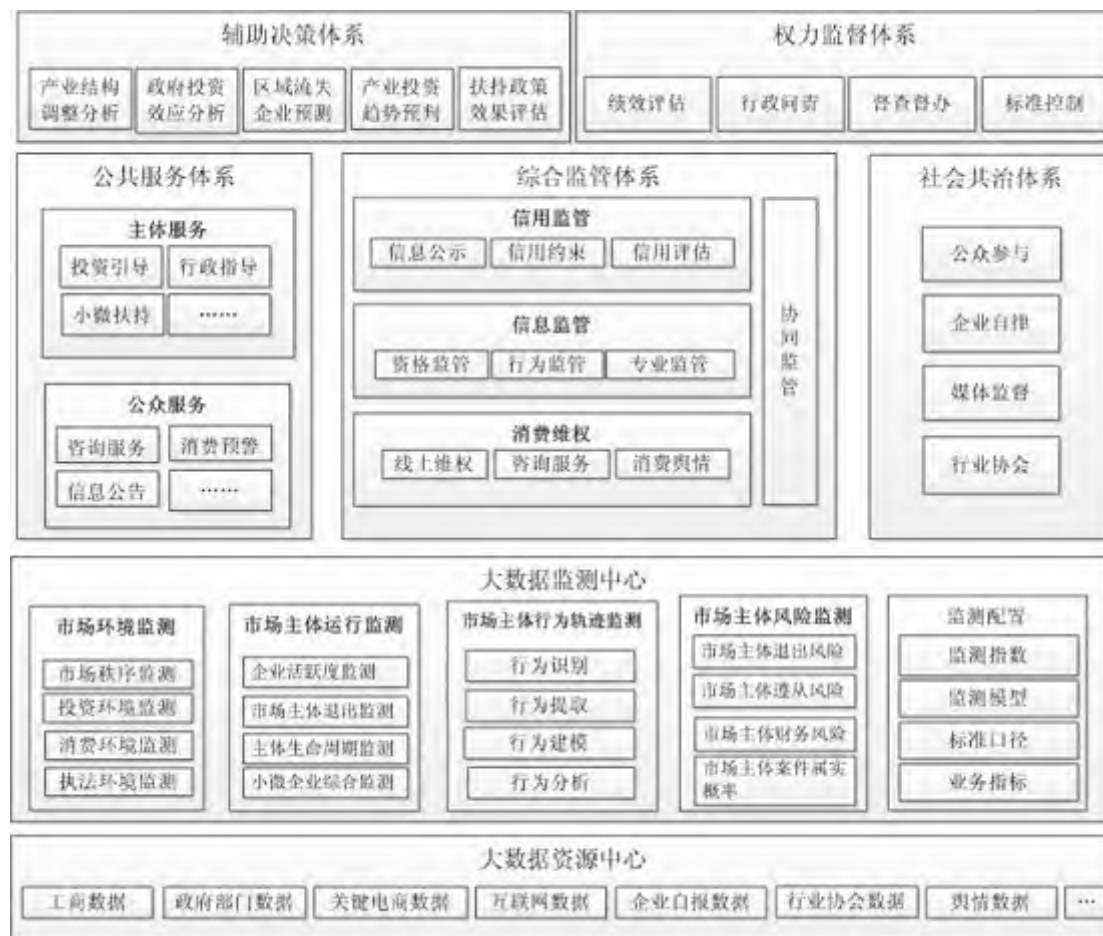
4

未来展望

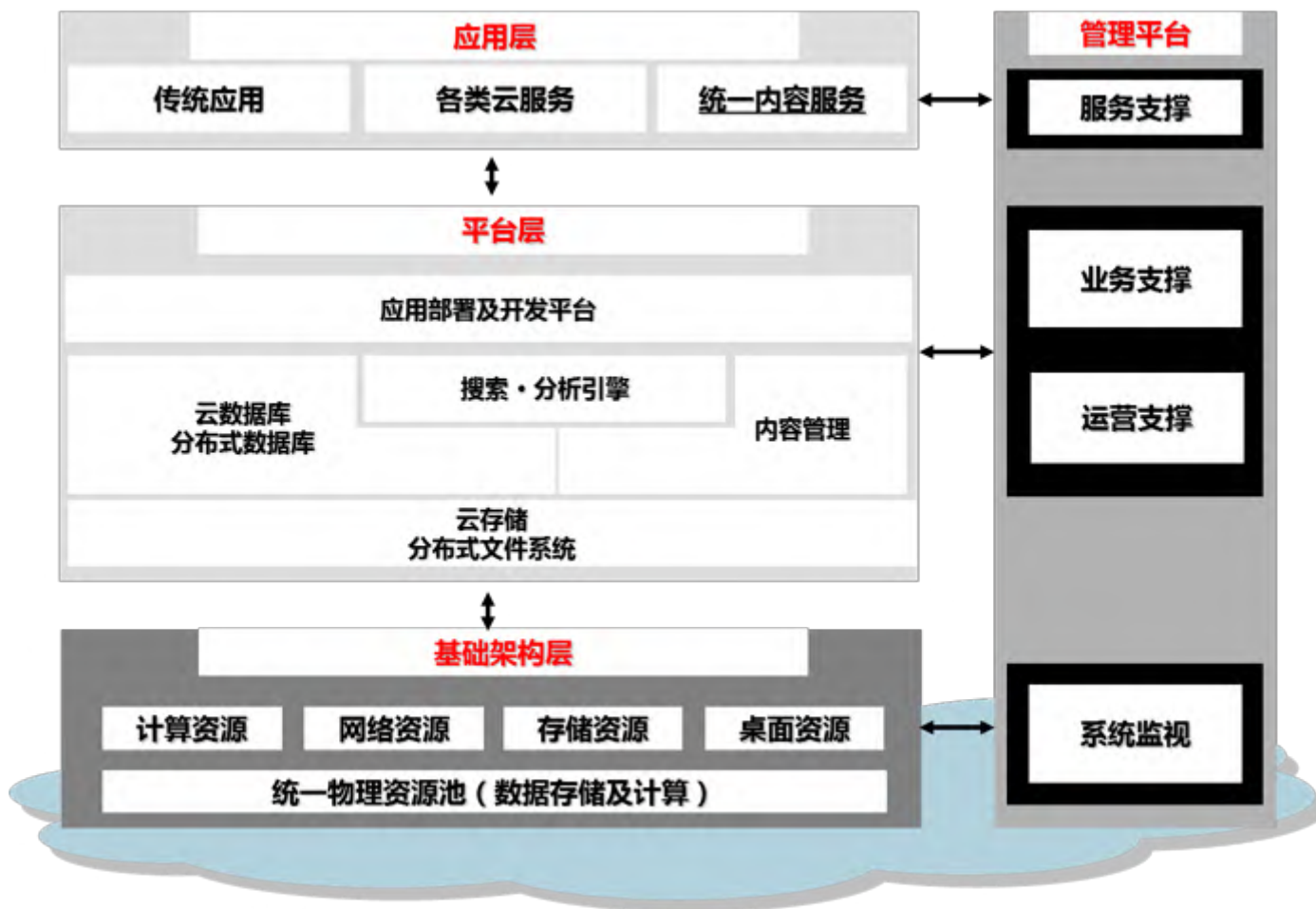
数据中心建设新阶段



大数据时代来临---业务架构革新



大数据时代来临----技术架构革新



小结：用大数据精准监管



融合工商数据、部门数据、社会数据，转变监管手段，提升监管能力

谢谢!

