2017 可信云大会
可信云标准新一代
客户满意是未来
Cloud Computing Summit 2017.7.25-26

大数据产品能力评测
赋能企业大数据能力建设

中国信息通信研究院
数据中心联盟 大数据技术与产品工作组组长
2017年7月25日

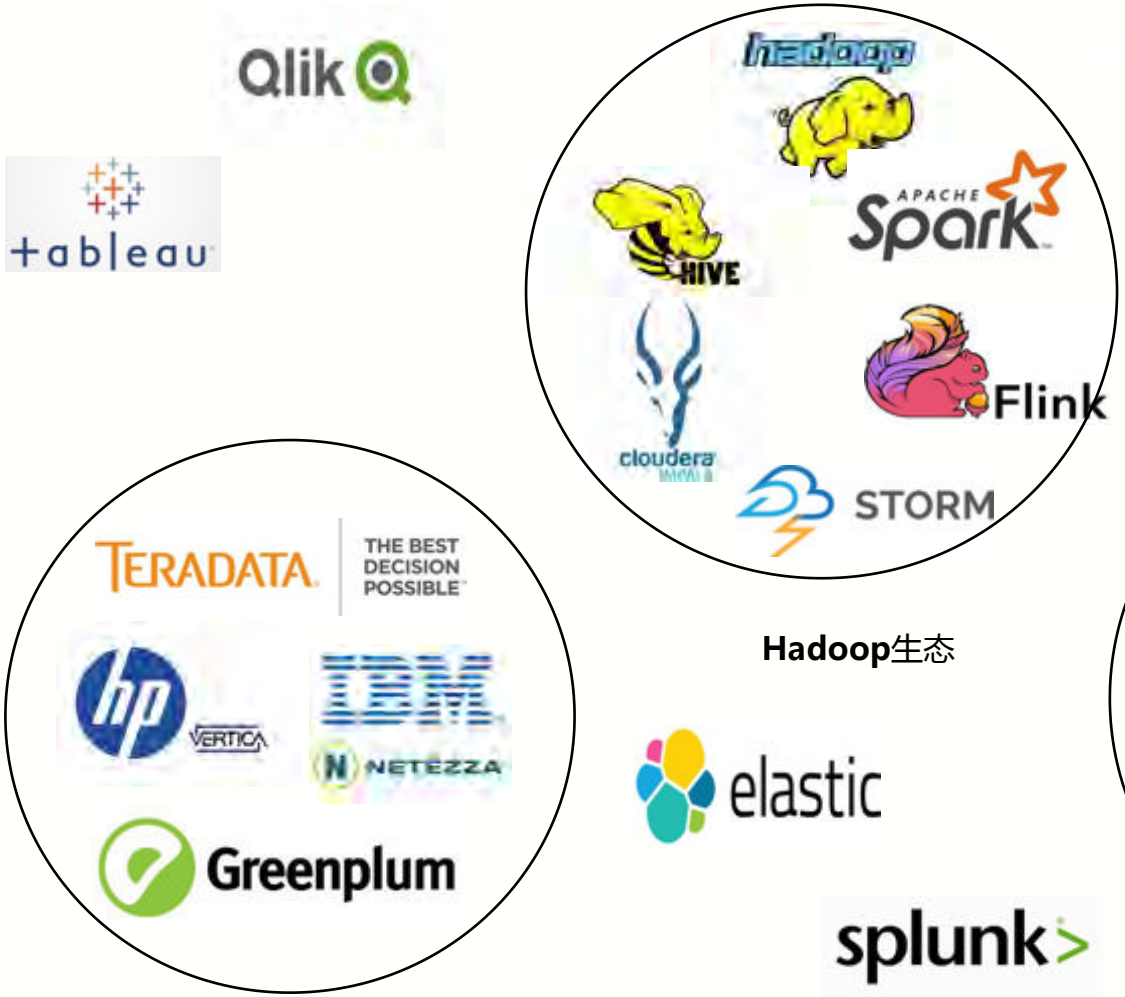
提纲

1 背景

2 大数据产品能力评测实践

3 下一步计划

大数据时代三架马车



Hadoop生态

MPP数据库

NoSQL数据库

大数据时代技术架构的特点

- 以开源技术为主导
  - Hadoop生态（HDFS、Hive、HBase等）
  - NoSQL数据库生态（MongoDB、Cassandra等）
- 多种技术和架构并存
  - 传统数据仓库
  - MPP架构
  - Hadoop架构（多种SQL on Hadoop解决方案）
  - Spark、Flink、Storm等计算引擎
  - Elastic Search、Splunk等技术

One size doesn’ t fit all

大数据技术带来的挑战

- 大数据技术厂商
  - 技术跟进和选型难
  - 难以实现差异化发展
  - 竞争无序，没有基本的门槛，劣币驱逐良币
  - 底层平台离用户太远，无法引起用户重视
- 用户
  - 技术和产品选型困难
  - 技术体系复杂、人才储备不足
  - 使用和运维的门槛提升

标准化和评测的意义

供应商

- 保证厂商之间有序竞争
- 建立行业门槛，提升供应商服务能力

共性的评估体系和标准

将复杂的产品转化为容易理解的指标

解决供应商和用户之间巨大的信息鸿沟

用户

方便用户选型

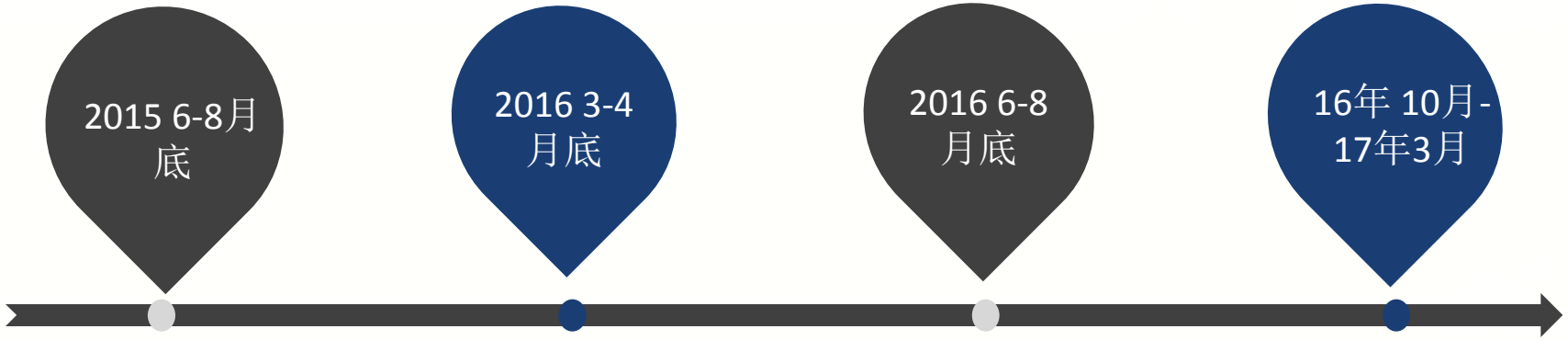
提纲

1 背景

2 大数据产品能力评测实践

3 下一步计划

大数据产品能力评测发展历程



第一批评测

华为、移动、星环、中兴、Ucloud5家企业通过了大数据基准测试

第二批评测

腾讯云、新华三、星环、明略数据、东方金信、博易智软6家企业通过了Hadoop基础能力评测

第三批评测

2家厂商通过了Hadoop基础能力评测，5家厂商参加了性能专项评测

第四批评测

9家厂商参加本次Hadoop基础能力评测，5家MPP数据库厂商参加了MPP数据库基础能力评测

已经有24家的25个产品通过评测，其中包括21款Hadoop产品，5款MPP数据库产品

第五批测试正在进行中，目前有10家企业参与性能评测，有10家参与基础能力评测

国内首个大数据产品评测体系

评测体系

基础能力

指标导向

性能

场景导向

Hadoop平台

MPP数据库

NoSQL数据库

Hadoop基础能力测试2.0

可用性	运维管理	兼容性	功能	安全	多租户	易用性	扩展性
Namenode主节点失效恢复	自动化部署	ODBC兼容性	数据导入	认证	租户管理	**工作流创建**	集群动态扩展
Datenode节点失效恢复	资源监控	JDBC兼容性	SQL任务能力	授权	资源管理	**工作流管理**	集群动态收缩
HMaster节点失效恢复	作业监控	SQL支持度	NoSQL数据库	加密	资源隔离	**工作流监控**	
RegionServer节点失效恢复	集群操作	传统数据库同步	机器学习	审计	**资源监控**		
ResouceManager节点失效恢复	故障管理	跨不同数据库表关联操作	流处理能力				
Hive Server失效恢复	日志管理	异构硬件兼容性					
HDFS备份恢复	配置管理	操作系统兼容性					
HBase备份恢复	权限管理						
双集群互备	无宕机升级						
运维管理节点失效及恢复							

大数据产品基础能力认证包括八大项：功能、运维、多租户、可用性、安全、兼容性、易用性、扩展性，总共44项测试用例

MPP数据库基础能力测试

功能	运维	安全	扩展性	可用性	兼容性
数据类型	安装部署	身份认证	动态扩展测试	关闭进程	访问接口兼容性
操作符	资源监控	加密	快速扩展测试	网络故障	CPU兼容性
函数	服务管理	审计	扩容不中断业务	整机宕机	大数据体系兼容性
DML操作	会话管理	权限	缩容		
表连接查询	作业管理	备份			
子查询	故障管理	集群灾备			
表空间	锁管理				
临时表	动态诊断事件				
索引	缓存管理				
事务支持	用户管理				
自定义函数	节点组管理				
存储过程	存储分配管理				
查询工具	分布及分区管理				
导入导出工具	资源负载管理				
系统表/视图支持	升级				
外部表					
集群间Dblink					

MPP数据库基础能力认证包括六大项：功能、运维、安全、扩展性、可用性、兼容性总共48项测试用例

Hadoop性能测试方法2.0用例分布

SQL任务	NoSQL任务	机器学习
I/O密集型任务	数据并发导入	Kmeans 无监督聚类
CPU密集型	95%的读，5%的写	SVM
报表任务	50%的读和50%的写	
分析型任务		
交互式查询		

Hadoop平台性能专项认证包括**SQL任务、NoSQL任务、机器学习和批处理**四类任务，总共15个测试用例

- SQL测试覆盖30TB数据规模
- NoSQL测试有20亿条数据的读写

通过企业名单

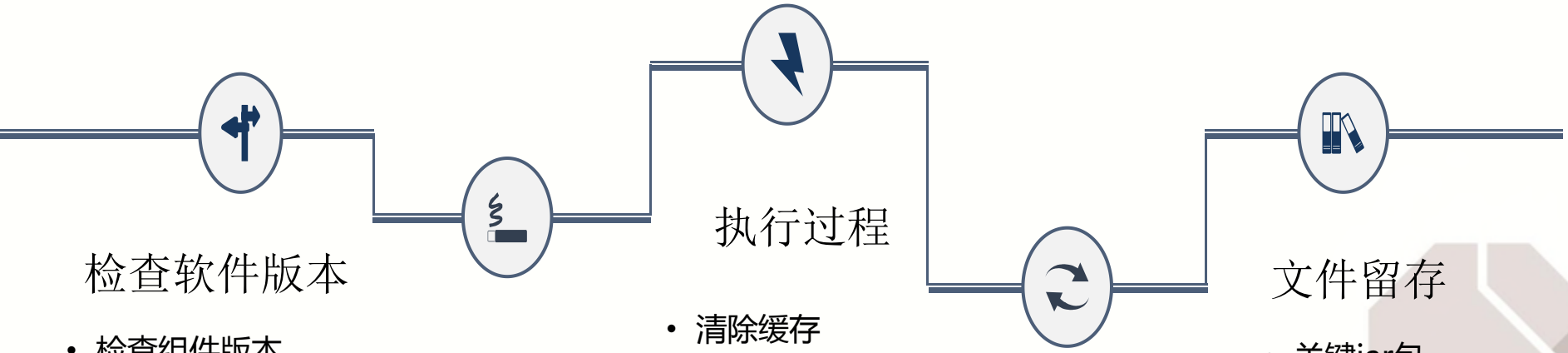
公司全称	大数据产品基础能力评测		大数据产品性能评测
	Hadoop平台基础能力评测	MPP数据库基础能力评测	Hadoop平台性能
华为技术有限公司			第一批（2015年）
星环信息科技（上海）有限公司	第二批 （2016年）		第一批（2015年）、第三批（2016年）
中兴通讯技术有限公司	第四批（2017年）		第一批（2015年）
中移（苏州）软件技术有限公司		第四批（2017年）	第一批（2015年）
上海优刻得信息有限公司			第一批（2015年）
新华三集团	第二批（2016年）		第三批（2016年）
腾讯云计算（北京）有限责任公司	第二批（2016年）		第三批（2016年）
北京东方金信科技有限公司	第二批（2016年）		第三批（2016年）
北京明略软件系统有限公司	第二批（2016年）		
博易智软（北京）技术股份有限公司	第二批（2016年）		
北京百分点信息科技有限公司	第三批（2016年）		第三批（2016年）
北京国双科技有限公司	第三批（2016年）		
深圳前海信息技术有限公司	第四批（2017年）		
中国电子科技集团公司第二十八研究所	第四批（2017年）		
航天信息股份有限公司	第四批（2017年）		
联想（北京）有限公司	第四批（2017年）		
成都四方伟业软件股份有限公司	第四批 （2017年）		
杭州泰一指尚科技有限公司	第四批 （2017年）		
北京中联润通信息技术有限公司	第四批 （2017年）		
中国华戎控股有限公司	第四批（2017年）		
天津南大通用数据技术股份有限公司		第四批（2017年）	
天津神舟通用技术有限公司		第四批（2017年）	
贵州易鲸捷信息技术有限公司		第四批（2017年）	
北京酷克数据科技有限公司		第四批 （2017年）	

测试实验室环境

锐捷RG-S6220-48XS4QXS 万兆交换机

组件	配置	台数
CPU	2*英特尔至强 E5-2620 v3 2.4GHz,15M 缓存	32
内存	4*16GB RDIMM, 2133 MT/s	31
	8*16GB RDIMM, 2133 MT/s	1
硬盘	10*1.2TB 10K RPM SAS 6Gbps 2.5英寸 热插拔硬盘	22
	10*1.2TB 10K RPM SAS 12Gbps 2.5英寸 热插拔硬盘	10
网卡	单口万兆网卡	32
交换机	锐捷RG-S6220-48XS4QXS 万兆交换机	1

审核方法与步骤



检查软件版本

- 检查组件版本
- 是否使用测试工具
- 组件列表

数据检查

- 数据大小
- 对于表检查行数、列数
- 数据内容
- 建表语句
- 副本数
- 执行脚本

执行过程

- 清除缓存
- 任务正常执行
- 集群的资源使用情况

结果检查

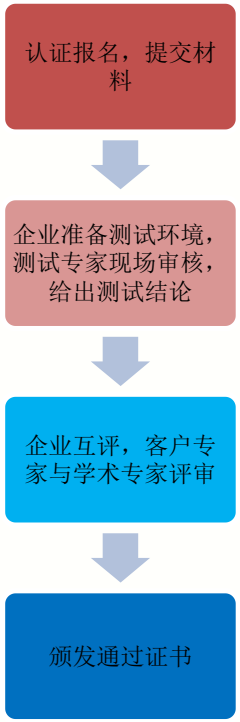
- 记录测试时间
- 检查结果是否执行正确

文件留存

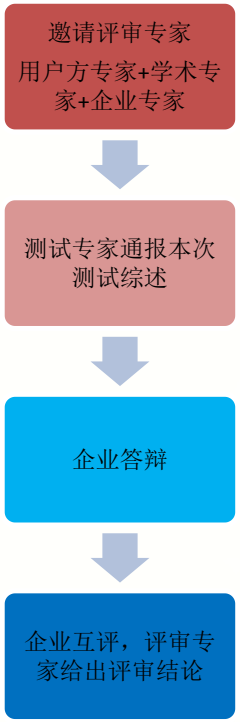
- 关键jar包
- 执行脚本
- 执行日志

自愿参加、公开公正的认证流程

认证流程：



评审流程：



大数据产品能力评测总结

- 权威性：数据中心联盟联合国内20多家大数据企业一起制定大数据产品评测标准，得到
- 全面性： 从功能完备性和性能两个角度来衡量大数据产品
- 严谨性：完善的现场测试流程，企业互评和专家评审流程
- 高认可度：24家25个产品通过评测，企业覆盖互联网和IT领域各类巨头，评测证书作为加分项多次写入大数据平台招标书
- 领先性
  - 性能是在统一平台（32台集群）、统一测试数据、统一测试工具、统一测试周期、统一测试规则下进行的
  - 业界领先的测试集群规模和集群配置（32台集群规模）
  - 10项以上用例使用TB级别的数据规模

衡量大数据产品的维度

- **功能完备性**
  - 功能、运维、可用性、安全、扩展、兼容性、多租户、易用性等8大维度
- **性能**
  - 产品本身的易部署性稳定性、易运维性、性能
  - 考察参测团队综合使用大数据平台的能力
    - 环境部署与集群规划
    - 测试工具的使用
    - 多任务调优能力
    - 时间进度安排
    - 集群的故障处理与运行维护
- **稳定性**

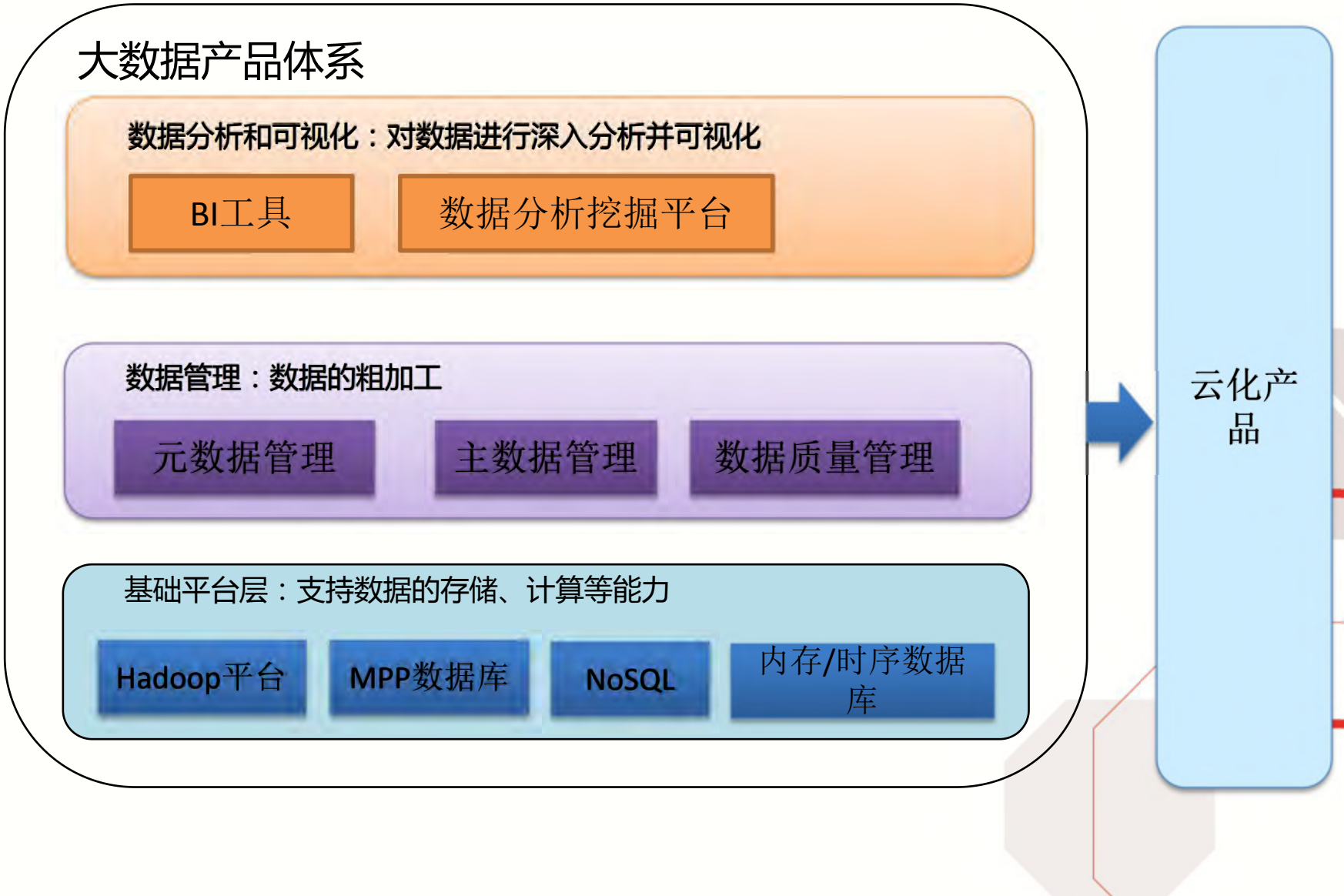
提纲

1 背景

2 大数据产品能力评测实践

3 下一步计划

未来评测体系规划


大数据产品体系
数据分析和可视化：对数据进行深入分析并可视化
BI工具
数据分析挖掘平台
数据管理：数据的粗加工
元数据管理
主数据管理
数据质量管理
基础平台层：支持数据的存储、计算等能力
Hadoop平台
MPP数据库
NoSQL
内存/时序数据库
云化产品

请各位专家批评指正

谢谢！

中国信息通信研究院
标准所移动与大数据研究部