

# 以标准推动开源技术在行业的落地

公司：中国信息通信研究院

职位：大数据技术主管

演讲者：姜春宇



# 我们是谁? Who are we?



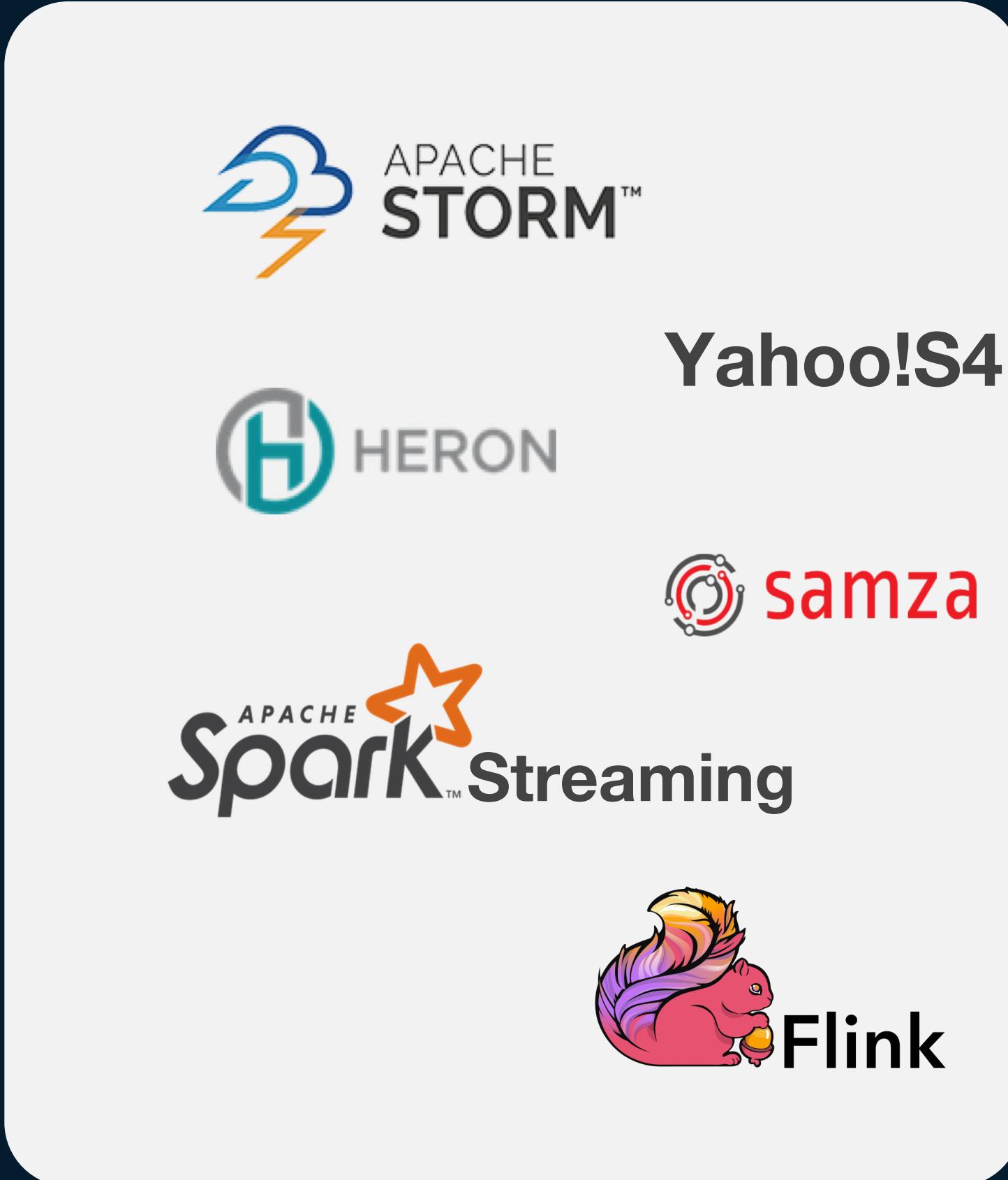
**大数据技术标准推进委员会** TC601 成立于2017年11月，是中国通信标准化协会 CCSA 下设的专业技术委员会，宗旨是凝聚产业链各个环节，识别和解决大数据发展面临的重大问题，以标准推进工作为纽带，推动大数据与实体经济深度融合。

TC601, established in November 2017, is a professional technical committee under CCSA of China Communications Standardization Association. Its purpose is to identify and solve the major problems in big data area, and promote the deep integration of big data and real economy with standard promotion as a link.

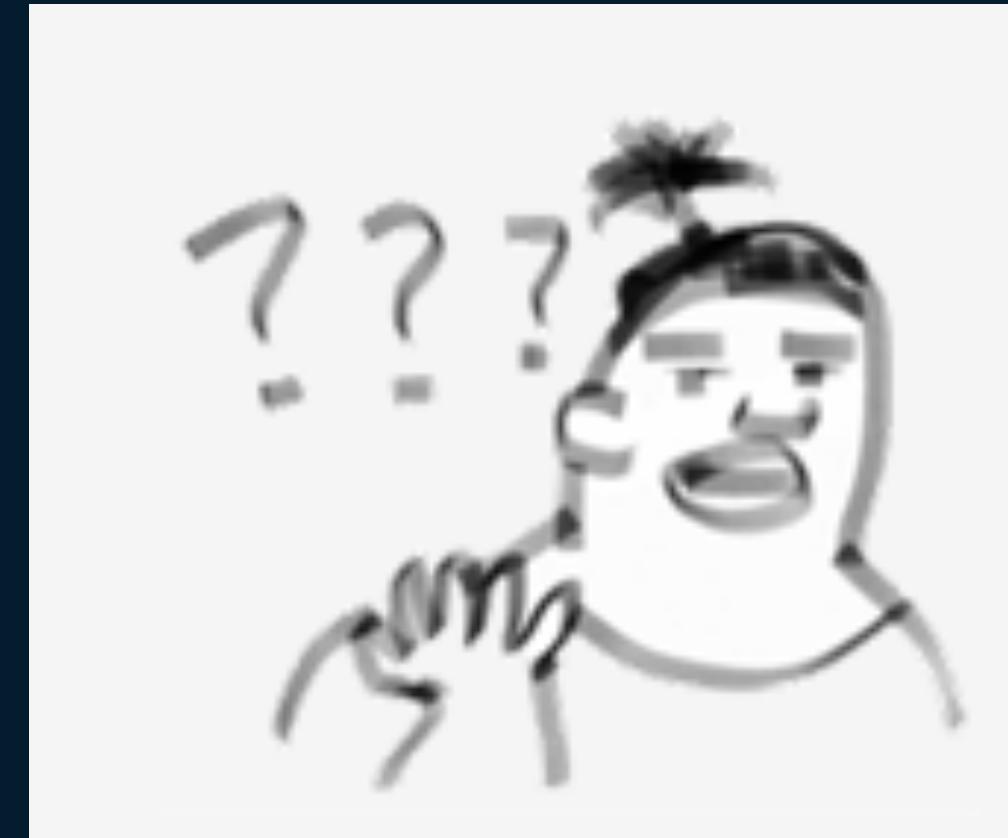
泛  
标  
准

- 1 宣传国家标准化法律、法规和方针政策，促进主管部门与会员之间的交流与沟通；
- 2 开展大数据领域技术和产品标准研究，组织会员开展标准符合性试验与评估验证工作；
- 3 开展数据开放、数据流通、数据资产管理等研究，开展相关咨询和培训服务；
- 4 搭建政、产、学、研、用共同参与的交流平台，促进大数据与行业应用深度融合；
- 5 搭建国内外大数据技术和产业的交流合作，积极参与国际标准化工作。

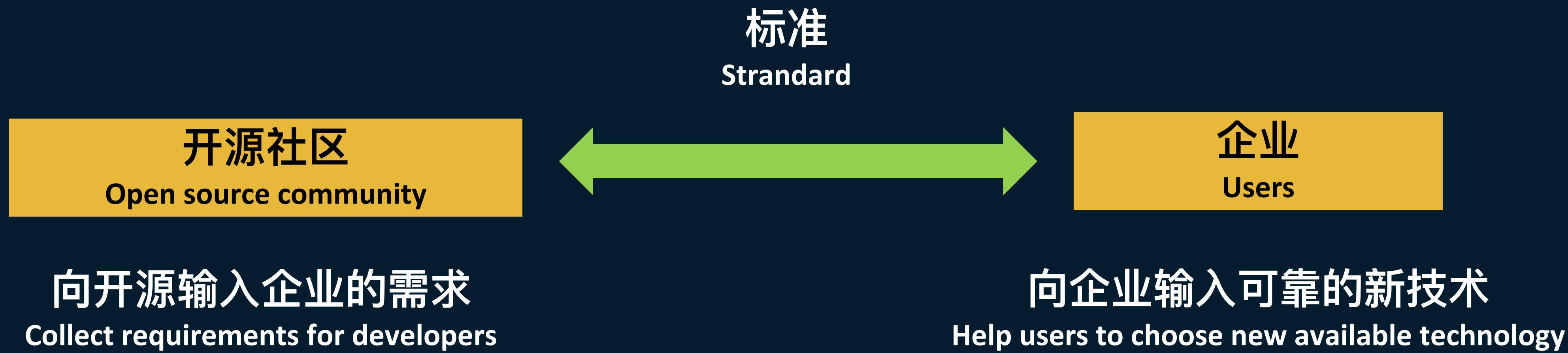
# 为什么需要标准? Why standard is important?



技术怎么选? 用户很头大  
So many technologies? It's  
really confusing for users.



# 我们在做什么? Link users and developers



# 流计算的标准 Stream processing standard



所以我们在2018年，联合8家企业做了第一个分布式流处理平台的标准。

So in 2018, we co-work with eight companies to make the first standard for distributed stream processing platform.



# 流计算的标准 Standard detail



标准定义了分布式流处理平台相关的名词术语，规定了一个优秀的产品应具备的7大类技术要求和51项具体技术指标。

The standard defines terminology related to distributed stream processing platform, specifies seven categories of technical requirements and 51 specific technical indicators for an excellent product.

基本功能 Function	分布式流处理平台应具备的处理实时数据的基本能力 <b>Ability to process real time data</b>
兼容能力 Compatibility	分布式流处理平台应具备的与主流硬件、软件及大数据生态的兼容性 <b>Compatible with mainstream hardware, software and big data ecosystem</b>
管理能力 Management	分布式流处理平台应具备的对用户友好的管理能力 <b>Operational and Maintenance Management Ability</b>
容错能力 Fault tolerance	分布式流处理平台在硬件或软件异常情况下仍能保证数据与服务稳定运行的能力 <b>Ensuring stable operation of data and services in case of hardware or software breakdown</b>
扩展性 Scalability	分布式流处理平台在线扩缩容的能力 <b>On-line scaling capability</b>
安全性 Security	分布式流处理平台应具备的安全保障能力 <b>Safety and security capability</b>
性能 Performance	分布式流处理平台应考察的性能指标项

## 大数据标准化体系

### 工具类标准

- 分布式批处理平台
- 分布式分析型数据库
- 数据集成工具
- 数据管理平台
- 数据挖掘平台
- 商业智能（BI）分析平台
- 分布式数据库
- 时序数据库
- 分布式流处理平台
- .....

### 数据服务类

- 通用准则
- 征信产品服务
- 金融风控产品
- 数据流通中心

## 我们的其他标准实践

新立项时序数据库、分布式事务数据库、流计算、知识图谱和多方安全计算等10个标准，另外尝试了面向用户行为数据分析解决方案的标准，目前已经形成技术产品类标准32项。

Ten new standards, such as time series database, distributed transaction database, stream computing, knowledge map and multi-party security computing, have been established. At present, 32 technical product standards have been published .

# 如何验证标准

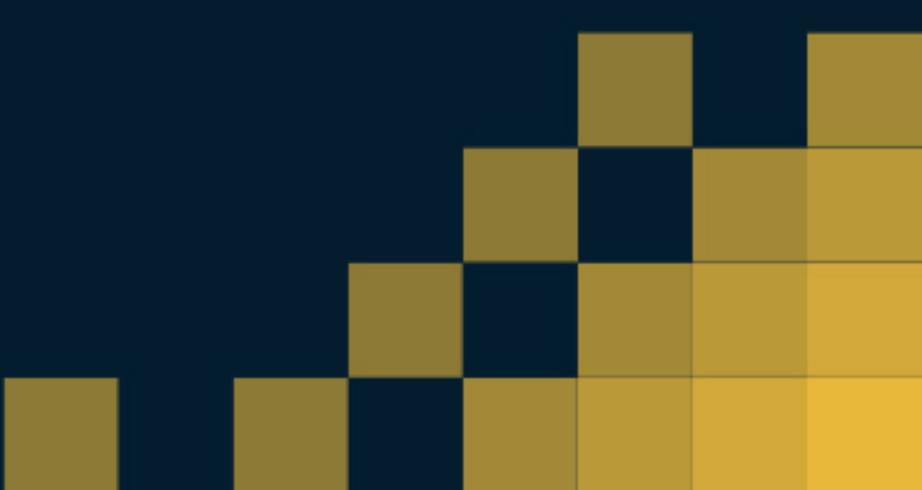


经过4年的发展，DCA大数据产品能力测评，已经成为国内权威的评测体系，成为厂商产品研发和用户采购选型的风向标。

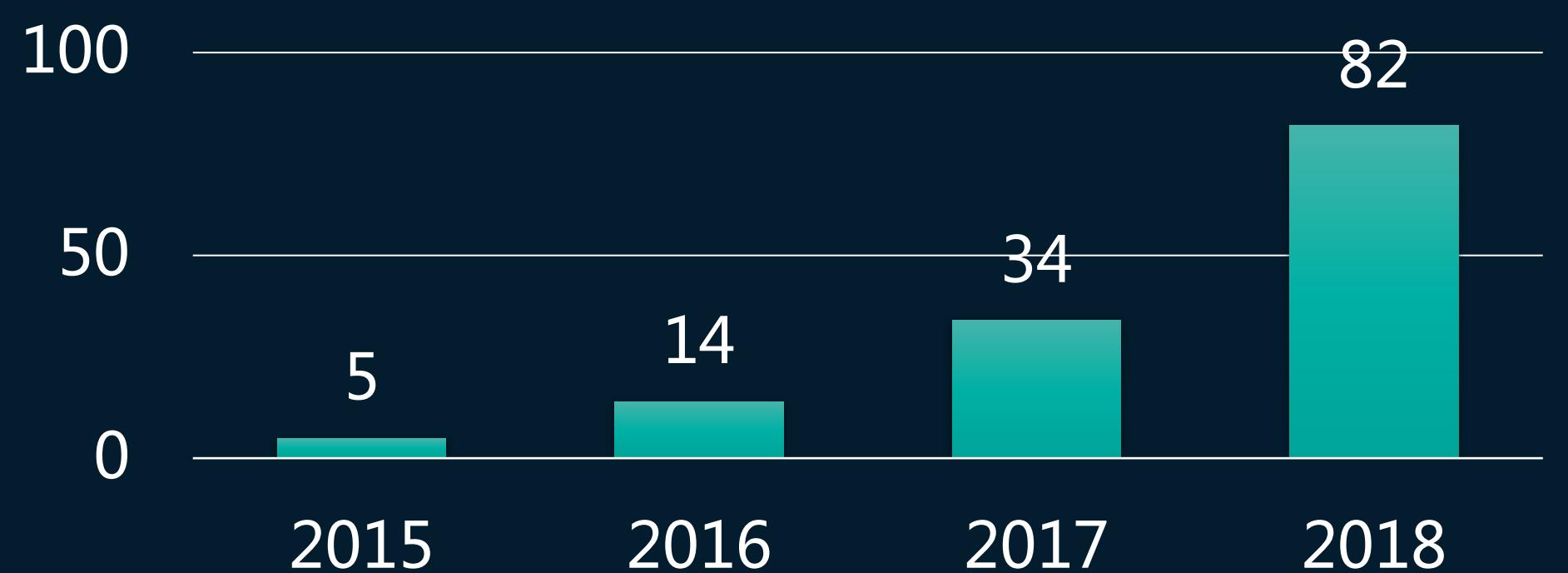
**After four years of development, DCA big data product capability evaluation has become an authoritative evaluation system in China.**

**基础能力**  
指标导向

**性能**  
场景导向



# 测评观察 Test observation

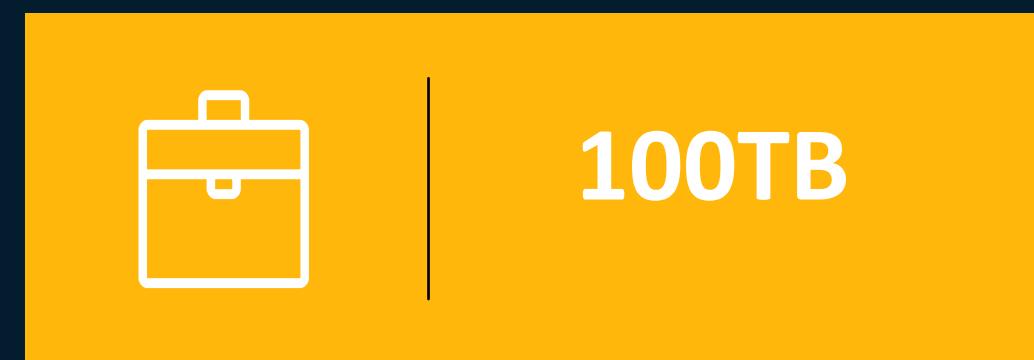


134个

目前已经有83产品通过评测，完成测试数量134个。

**At present, 83 products have passed the evaluation and 134 tests have been completed.**

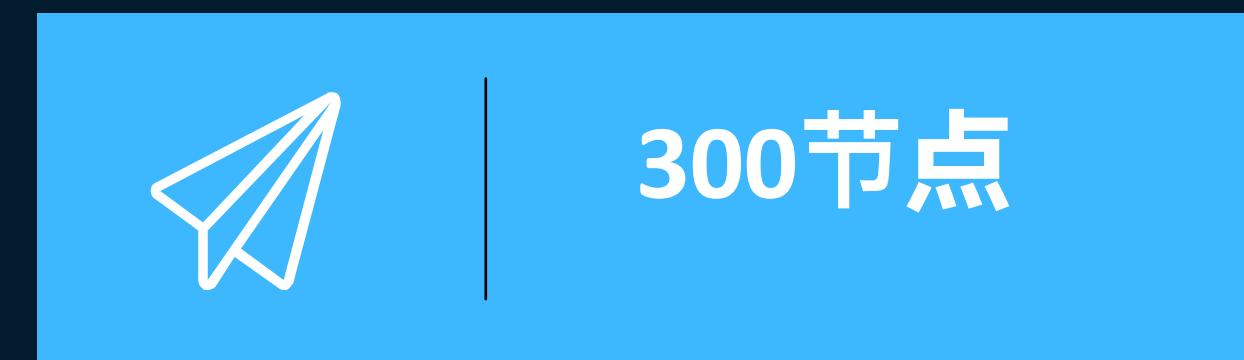
# 测评观察 Test observation



全球领先数据规模  
性能最大测试数据100TB

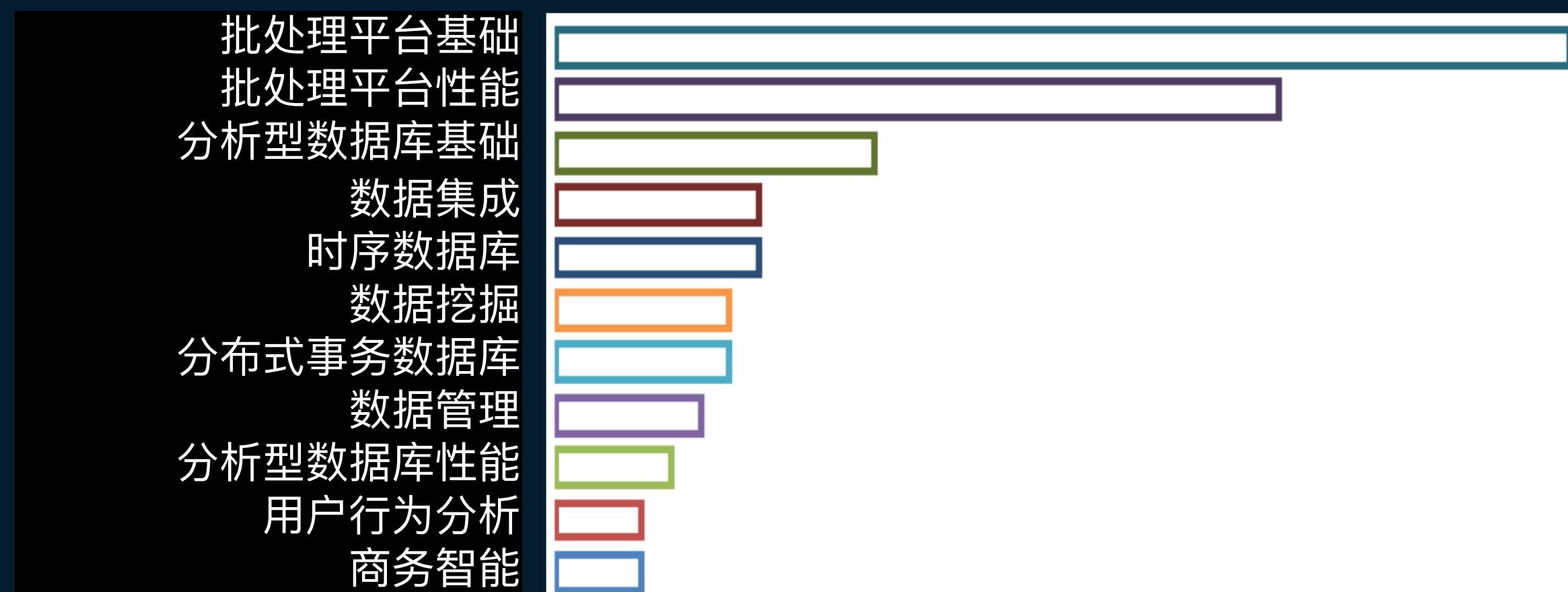


最大基础能力集群规模  
阿里完成10000节点批  
处理平台基础能力评测



最大的性能测试规模  
华为通过了300节点物理机性能评测  
阿里通过300个云服务器性能评测  
新华三通过了200节点物理机性能评测

## 最喜欢的评测项目



# 后续合作 Future Collaboration



- 与Flink社区合作发布流处理平台实践白皮书
- Publish Steam Processing Platform White Paper with Flink Community
- 开展面向企业用户的技术培训
- Organize Technical Tutorial for Enterprise Users
- 进行分布式流处理平台性能标准的制定
- Making Benchmark for Performance
- 启动流处理平台第一轮测评
- Start the test for Stream Processing Platform



THANKS

