

# 存储系统评测的 规范、工具与方法

CCF信息存储技术专委会·存储测试工作组：

肖侬 武永卫 周可 陈志广 张武生 陈俭喜 董勇 张晓 曾令仿 任峰

# 目录

---

- 一：存储基础设施重要性
- 二：典型应用场景存储能力需求
- 三：存储评测工具与方法
- 四：测试案例

# 目录

---

一：存储基础设施重要性

二：存储典型应用场景能力需求

三：存储评测规范、工具与方法

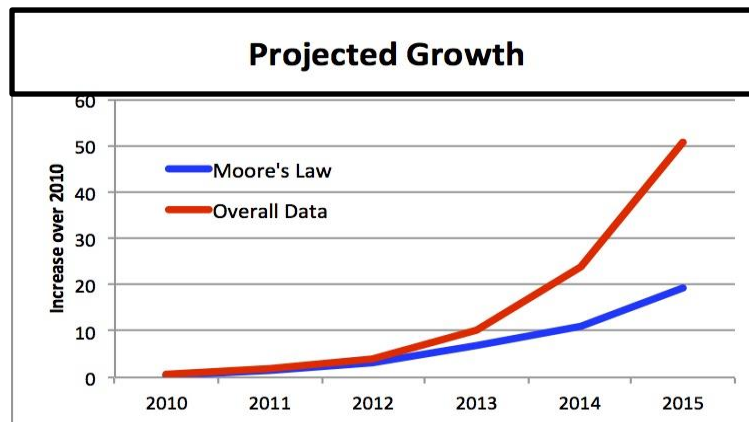
四：测试案例

# “存、算、网”是构建信息系统的3个关键环节，数据存储能力是底座

数据已成为继土地、劳动力、资本、技术之后的第五大生产要素  
称为继理论、实验、计算之后的第四维

- ◆ 数据是**财富**，是**战略资源**，数据驱动创新发展
- ◆ 存储系统承载数据，是网络空间安全的**核心基础设施**

- ◆ 存储、计算和网络是信息系统三大部件
- ◆ 信息系统已进入以数据/存储为中心的时代



数据增长速度超过摩尔定律

- 国务院《促进大数据发展行动纲要》将**大数据上升为国家战略**
- 存储成本**超过50%**
  - 公共信息服务设施
  - 工业领域(地球物理...)

数据重要性日益凸显

## ■ 受控、可信的存储设施是数字经济高质量发展的基石

- ◆ 大量数据保存在非可信存储系统中成为严重安全隐患
- ◆ 可信的存储系统，也是**信息安全的重要保障**



数据保存在不受控设备中风险多样

# 目录

---

一：存储基础设施重要性

**二：典型应用场景存储能力要求**

三：存储评测规范、工具与方法

四：测试案例

# 新需求：新联接、新应用带来数据爆炸式增长，对存储能力提出新需求



根据预测，到2030年，全球每年新增数据量将超过1000ZB量级

## 存的更多

- 医疗：小鼠大脑仅13个神经元就高达**1TB**数据量
- 能源勘探：采集从二维->三维，数据量**增加10倍**

## 存的更久

- 银行票据：至少保存**15年**以上，通常都不删除
- 医疗：基因数据保存**30+年**

## 24 存的更可靠

- 金融：实时在线风控**7\*24小时**保障资金安全
- 煤矿作业：**7\*24小时**安全生产，监控不能中断

## 存的更安全

- 金融：交易数据、影像数据、第三方数据等
- 安平：治安、出入境、经侦、刑侦等

1000亿

联接数

5G, 8K, VR

30x

流量

TB级 >>> PB级

传统数据

视频&图片数据

企业IT

# 金融行业：生产交易存储场景，可靠性须达到6个9，确保业务持续性



手机网民 **96.4%**  
渠道业务压力浪涌式

网上交易、移动支付取代  
线下业务，支持上亿在线  
用户秒杀、节日大促



新建数据中心 **重构容灾**  
双活/多活架构转型

数据中心灾备建设，应对  
自然灾害、断网断电导致  
的业务中断、数据丢失

国际结算



交易系统



银证转账



**全渠道  
交易系统**

信贷业务



移动营销



移动银行



海量数据 **精准** 分析  
反欺诈从 **事后** 到 **实时**

• **性能**：网上交易要求**十万级IOPS**，**0.5ms时延**

① **OLTP模型**

• **可靠性**：可靠性**6个9**，**7x24**提供服务；

① **硬件冗余** ② **数据修复** ③ **软件故障检测**

• **数据安全**：数据中心**灾备双活**纳入金融行业标准

① **AA双活** ② **数据加密**

• **能效**：**金融机构**数据中心运营成本**电费占68%**

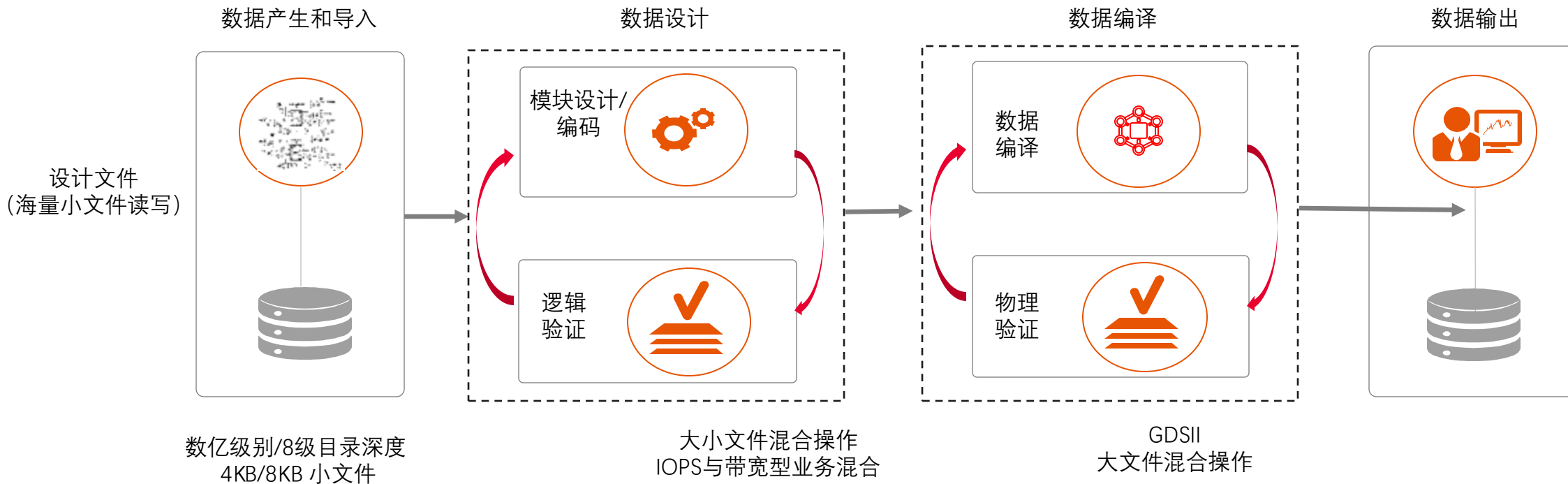
① **性能能效比** ② **重删压缩**



# 高性能计算：存储需低时延、高带宽，并支持大规模并发IO、高频元数据操作

计算全流程耗时几十小时，存储性能影响计算周期

任意环节失败将导致高性能计算失败，存储可靠性影响效率



**性能：**数万核同步工作带来高并发IO，IO时延低于8~10ms

- ① 高吞吐带宽模型
- ② 高IOPS随机读写模型

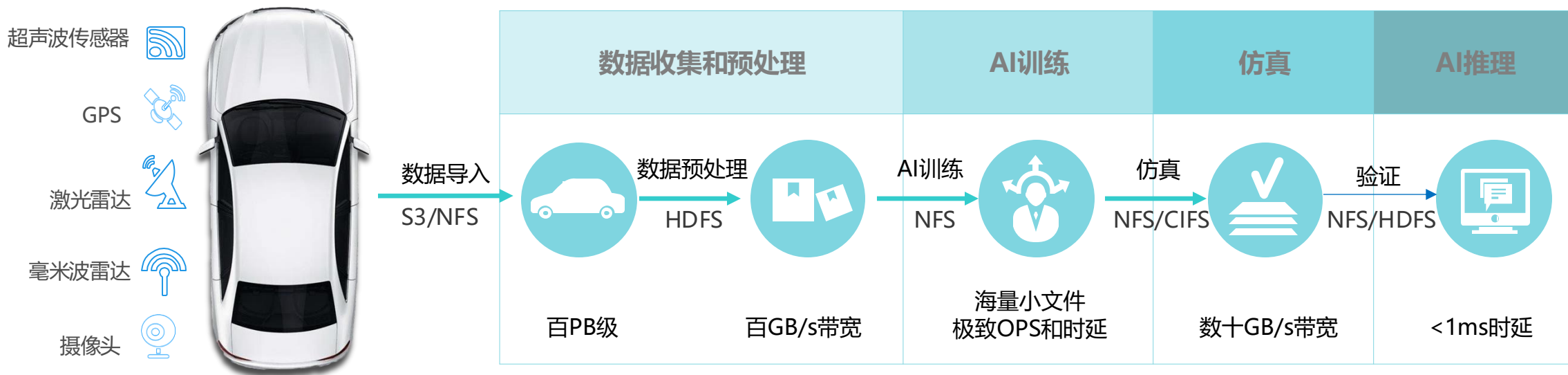
**可靠性：**存储系统故障对上层计算软件不感知

- ① 数据重构
- ② 亚健康检测
- ③ 数据修复

**容量：**每天生产10T数据，单套存储系统**PB级容量需求**

- ① 容量密度
- ② 大比例EC

# 自动驾驶：混合负载，多协议融合、支持多语义、并实现性能无损



## L3演进到L4

50倍路测里程，50倍数据增长

**性能：**同时要求高带宽300GB/s和

高OPS**混合负载**，超低时延

- ① 高吞吐带宽模型
- ② 小对象TPS模型
- ③ 大对象带宽模型
- ④ 高IOPS随机读写模型

模型

**兼容性：**S3、NFS、CIFS、

HDFS**多协议接口**兼容、互通

- ① 文件协议
- ② 大数据协议
- ③ 对象协议
- ④ 跨协议互通

**容量：**每车每天64TB数据，

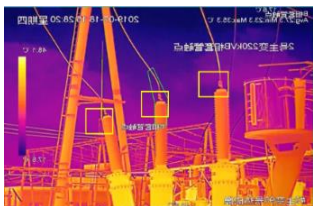
要求单集群**百PB级**

- ① 容量密度



# 边缘计算：海量数据在边缘产生，要求大容量、低功耗

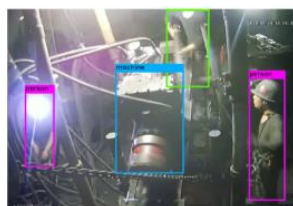
5G+IoT+AI使能电力、矿山等行业边缘数据存储/处理应用加速



5G视频巡检，  
温度采集



井下巡检，减  
少不安全行为



瓦斯抽采安全  
AI故障预测

## 边缘海量数据处理

多样数据统一接入，汇聚管理  
海量数据存储，转发

## 边缘AI推理分析

边缘就近实时分析  
业务就近闭环

# 75%

数据将来自边缘和终端

# 50%

数据需要在边缘处理

Source: IDC Report 2021Q1

企业分支

## 1 容量

- 弹性EC
- 重删压缩

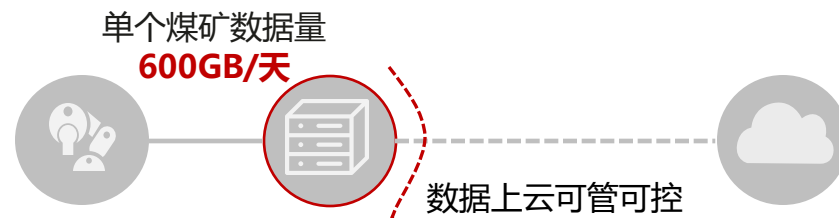
## 2 能效

- 容量能效比
- 性能能效比

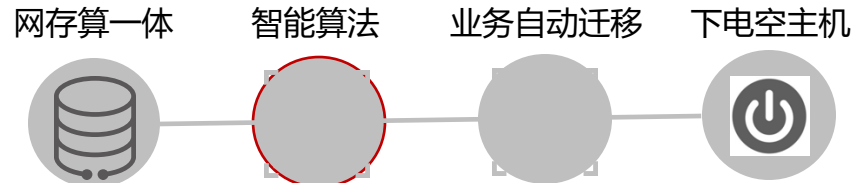
## 3 性能

- OLTP模型
- OLAP模型

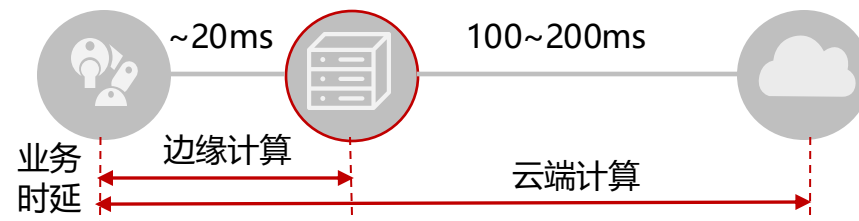
海量数据本地计算和存储，上云可管可控



智能运维，软硬协同，节能降耗



存储时延1ms以内，缩短端到端处理时延



边缘数据存储与处理需要新基础设施满足近端处理诉求



# AI大模型：大模型开发应用面临三大挑战，对存储能力提升诉求强烈

## ① 数据归集和预处理时间长

数据源分散，归集困难  
数据频繁搬迁，TB预处理需10天

算力占比10%

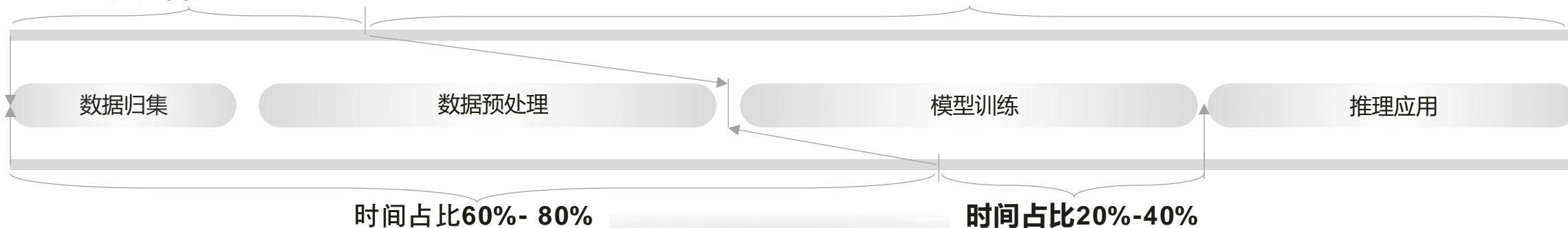
## ② 模型训/推算力消耗大

小文件加载慢，GPU等待时间长  
Checkpoint恢复时间长，GPU利用率低

算力占比90%

## ③ AI大集群业务稳定挑战大

单点故障导致整集群不可用  
训练全部中断，计算资源被浪费



### 专业存储诉求

智

数据智能流动

数据**快速归集**、**高效清洗**、**高效流转**能力，  
缩短大模型数据准备时长 (**20亿文件30+天**)

快

极致读写性能

并行**xxTB/s**大带宽和**x亿高OPS**，  
提升训练效率，让**GPU少等待**

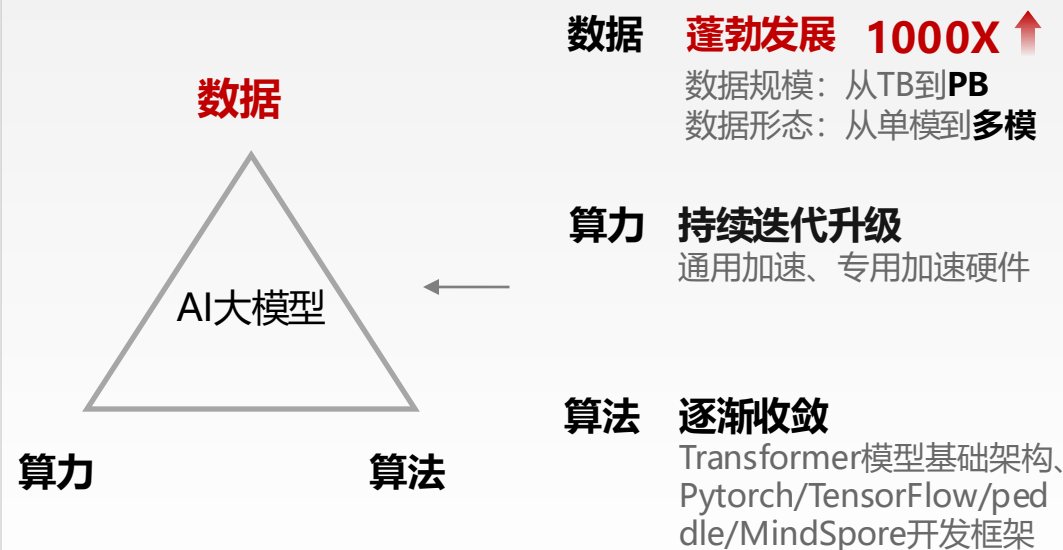
稳

极致高安全可靠

**故障域隔离**，**业务永远在线**  
端到端**自主可控**，**分钟级断点续训**

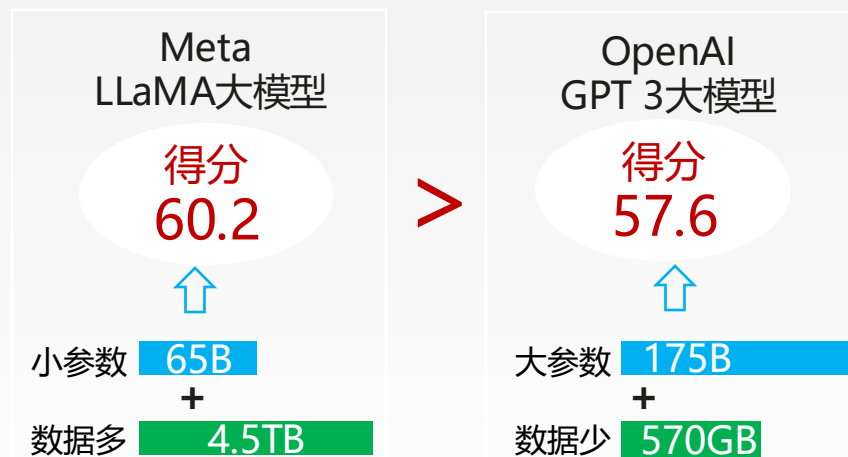
# 数据决定AI智能的高度：缺数据，难AI

## 数据是AI大模型的核心要素



## 数据规模和质量决定AI智能的高度

优质数据越多，模型效果越好



Source: Meta AI 官方论文《LLaMA: Open and Efficient Foundation Language Models》，大模型在OBQA测试的得分

作为数据载体，数据存储成为AI大模型的关键基础设施

# 安全：存储是数据安全防御体系中的最后一道防线，不可或缺

## 全球数据安全事件频发

- 21年4月，某厂商遭勒索共计，索要赎金5000万美元
- 21年3月，欧洲OVH机房着火，360万个网站瘫痪
- 21年12月，日本某大学超算系统数据误删，3400个文件丢失
- 22年6月，限制共享漏洞，国内系统面临攻击风险
- 22年12月，某车漆用户数据泄露，赎金200万美元

## 全球加快数据安全立法



### 中国：

- 《网络安全法》
- 《数据安全法》
- 《关键信息基础设施安全保护条例》

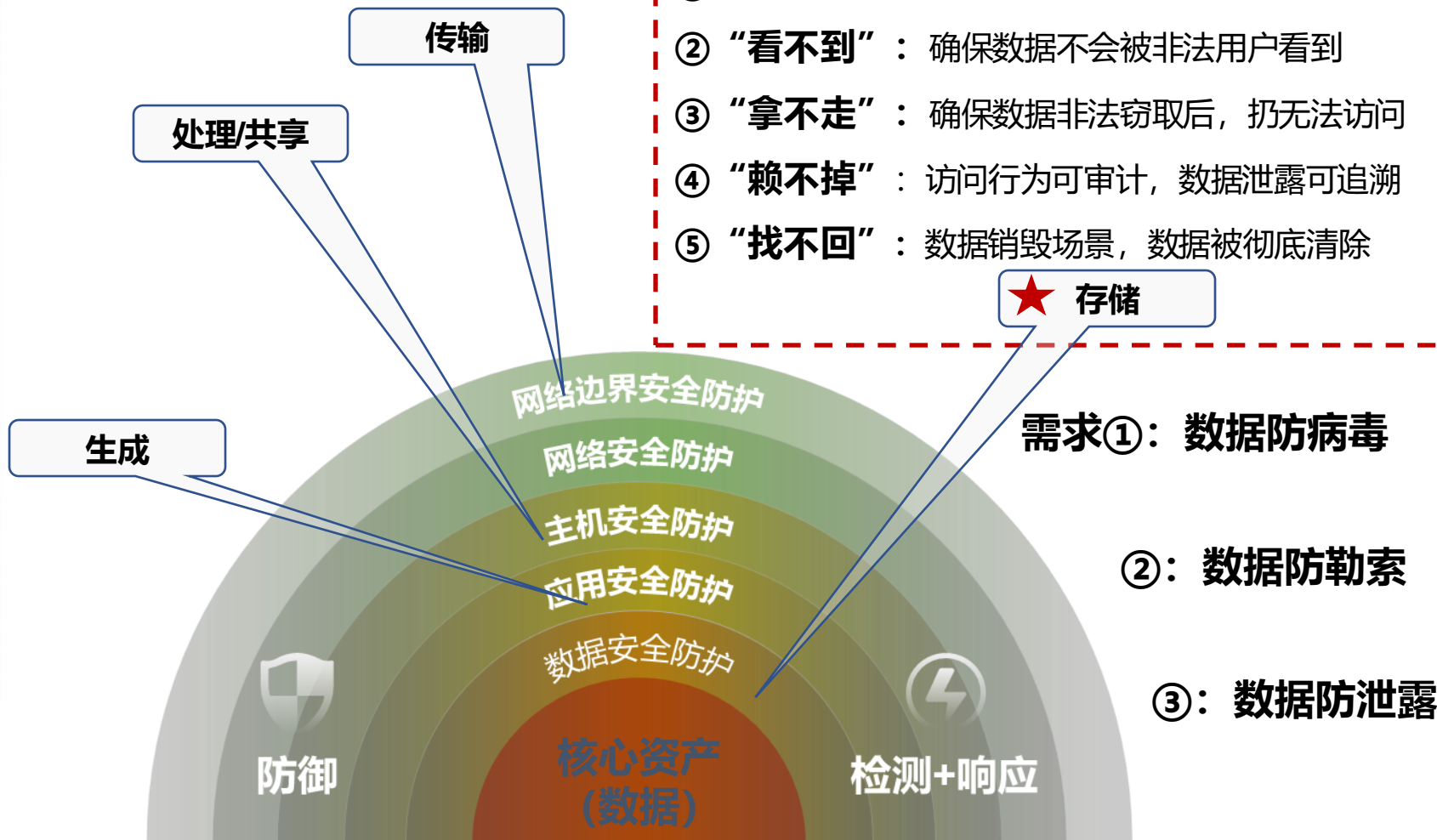


### 海外：

- 《欧盟通用数据保护条例》
- 《英国数据保护法》

存储的数据保护能力在数据安全纵深防御体系中不可替代

- ① “进不去”：确保非法用户无法访问数据
- ② “看不到”：确保数据不会被非法用户看到
- ③ “拿不走”：确保数据非法窃取后，仍无法访问
- ④ “赖不掉”：访问行为可审计，数据泄露可追溯
- ⑤ “找不回”：数据销毁场景，数据被彻底清除



# 典型业务场景对存储能力的需求汇总

典型应用场景	性能						可靠性						数据保护		数据安全			兼容性				容量		能效		
	OLTP模型	OLAP模型	高吞吐带宽模型	高IOPS随机读写模型	小对象TPS模型	大对象带宽模型	软件检测	数据重构	数据修复	亚健康检测	硬件冗余	掉电管理	故障域管理	快照克隆	远程复制、同步复制	双活	防病毒	防勒索	数据加密	文件协议	块协议	大数据协议	协议互通	重删压缩	容量密度	存力能效比
金融交易类	●						●	●	●	●	●	●	●	●	●	●	●	●		●			●			●
科学计算			●	●			●	●		●	●	●	●				●	●		●				●		●
自动驾驶			●	●	●	●	●	●	●	●	●						●	●		●		●	●	●		●
边缘计算	●	●					●	●		●		●							●		●		●	●	●	●

# 目录

---

- 一：存储基础设施重要性
- 二：存储典型应用场景能力需求
- 三：存储评测规范、工具与方法**
- 四：测试案例



# 当前存储评测多以考核性能为主，其评测维度和评价方法尚未充分涵盖实际业务场景对存储系统真实能力、特性需求：**需要建立一套综合评测体系**

评测名称	IO500	SPC	TPC	SPECstorage Solution
负载	HPC的5类应用场景： <input type="checkbox"/> IOEasy、 <input type="checkbox"/> IOHard、 <input type="checkbox"/> MDEasy、 <input type="checkbox"/> MDHard、 <input type="checkbox"/> Find	<input type="checkbox"/> SPC-1企业关键业务应用场景下的负载模型，覆盖结构化数据为主的数据库、电子邮件等在线事务处理（OLTP）应用； <input type="checkbox"/> SPC-2基准测试主要针对顺序I/O应用环境。主要覆盖大文件读写、非线性编辑和大表查询，测试带宽负载场景。	<input type="checkbox"/> 在线交易：TPC-C、TPC-E <input type="checkbox"/> 决策支持：TPC-H、TPC-DS、TPC-DI <input type="checkbox"/> 虚拟化：TPCx-V、TPCx-HCI <input type="checkbox"/> 大数据：TPCx-HS、TPCx-BB <input type="checkbox"/> IOT：TPCx-IoT <input type="checkbox"/> AI：TPCX-AI <input type="checkbox"/> 通用：TPC-Energy、TPC-Pricing	5类业务场景： <input type="checkbox"/> SWBUILD、 <input type="checkbox"/> VDA、 <input type="checkbox"/> EDA_BLENDED、 <input type="checkbox"/> AI_IMAGE、 <input type="checkbox"/> GENOMICS
指标	<input type="checkbox"/> 带宽 <input type="checkbox"/> 元数据性能	<input type="checkbox"/> IOPS	<input type="checkbox"/> 流量指标(Throughput) <input type="checkbox"/> 性价比(Price/Performance)	<input type="checkbox"/> 可以提供的最大可持续吞吐量
测试工具	<input type="checkbox"/> io500（开源）	不公开	TPC提供开源工具	<input type="checkbox"/> SPEC SERT Suite
测试方式	自测并提交结果，免费	送测，评测机构负责，收费	自测并提交结果，服务费	自测并提交结果，收费

- 可靠性
- 可信性
- 可维护性
- 扩展性
- 兼容性
- 能效

- 工况性能(R/W混合比，文件大小混合比，目录深度混合比，)
  - 元数据
  - 带宽
  - 延迟
  - 持续性

# CCF信息存储技术专委会 存储评测

标题 页面 结果

- 1 简介
  - 1.1 背景
  - 1.2 目标与挑战
  - 1.3 一般性规定
    - 1.3.1 被测对象
    - 1.3.2 测试相关实体
    - 1.3.3 测试组织
    - 1.3.4 测试工具
- 2 评测维度
  - 2.1 性能
  - 2.2 可靠性和可用性
  - 2.3 数据保护
  - 2.4 生态协议
  - 2.5 扩展性
  - 2.6 能效
  - 2.7 可管理性
- 3 典型场景
  - 3.1 AI场景
  - 3.2 高性能计算场景
  - 3.3 关系数据库场景
  - 3.4 大数据
  - 3.5 云计算
- 4 评测设计
  - 4.1 测试用例数据集生成
  - 4.2 基于真实应用运行场景的负载模拟
  - 4.3 性能测试与能力测试
  - 4.4 测试用例
- 5 测试指引
  - 5.1 测试准备
  - 5.2 测试执行
  - 5.3 状态监控与收集
  - 5.4 结果收集与分析

## CCF 信息存储技术专委会 存储测试规范说明

### 卷 0•测试规范基础

序号	修订日期	修订内容	文件版本	修订人
1				
2				
3				
4				
5				

标题 页面 结果

- 1 AI业务对存储的要求
  - 1.1 大语言模型业务特征
  - 1.2 AI业务的通用要求
- 2 AI存储测评指标
  - 2.1 性能
    - 2.1.1 AI训练模型规模测试
    - 2.1.2 多AI训练任务并发性能测试
  - 2.2 可靠性
    - 2.2.1 冗余可靠性
    - 2.2.2 故障切换
    - 2.2.3 数据不丢失
  - 2.3 能效
  - 2.4 扩展性
- 2.5 数据保护
  - 2.5.1 快照
  - 2.5.2 防勒索
- 3 测试操作指南
  - 3.1 准备
    - 3.1.1 被测系统安装
    - 3.1.2 测试工具安装
    - 3.1.3 配置被测系统
    - 3.1.4 配置测试工具
  - 3.2 运行
  - 3.3 结果收集与验证
  - 3.4 结果提交

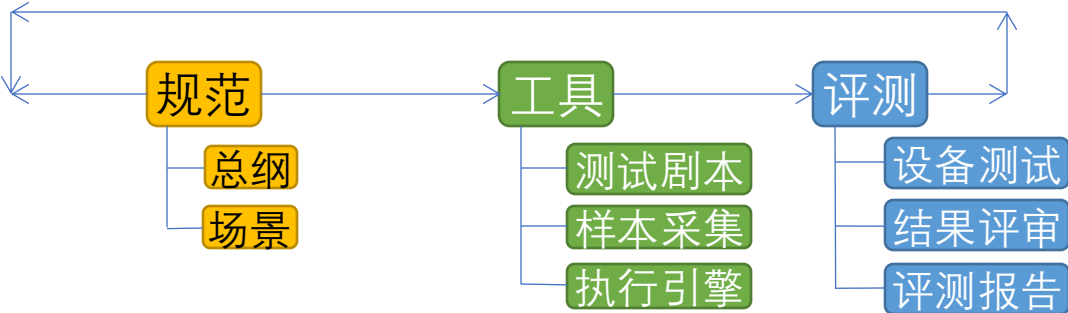
## CCF 信息存储技术专委会 存储测试规范说明

### 卷 1•人工智能场景

序号	修订日期	修订内容	文件版本	修订人
1				
2				
3				
4				
5				

- 卷2•高性能计算场景(并行文件系统, NetCDF, HDF5, ...)
- 卷3•关系数据库场景(sqlDB)
- 卷4•大数据场景(nosqlDB, graphDB, HDFS, ...)
- 卷5•云计算场景(虚拟化, 超融合, 存算一体, 存算分离, ...)

### • 工作流程

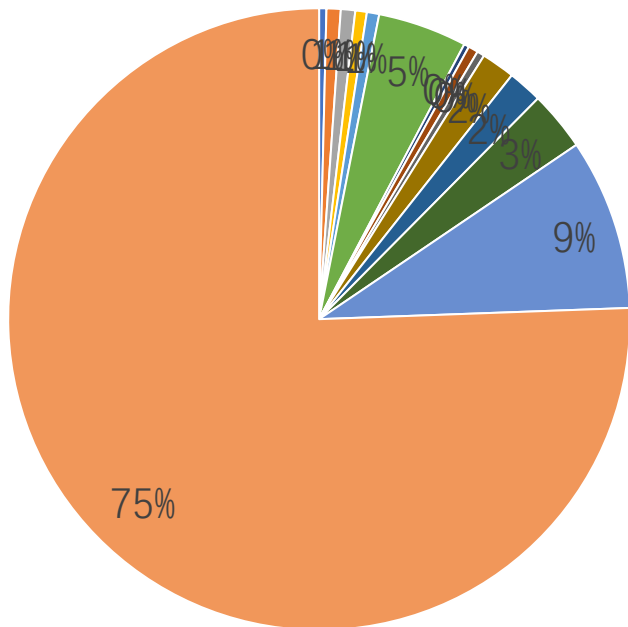


# 面向真实业务场景的IO建模

大小区间	public-文件大小MB	group-文件大小MB	home-文件大小MB	合计-文件大小MB	public-数量个	group-数量个	home-数量个	合计-数量个
小于100k	4627224	2864053	10543818	18035095	202210209	131169907	362309721	695689837
100k~256k	7608821	6771143	20726080	35106044	50434431	44714895	162195750	257345076
256k~1M	15962689	8736380	10853870	35552939	34614155	25943476	22762403	83320034
1M~2M	17729290	3576444	7123459	28429193	12717878	3056706	5072732	20847316
2M~4M	21548591	388523	8520276	30457390	7561101	142111	3146274	10849486
4M~8M	8931606	524174	210877696	220333476	1651490	96311	46620298	48368099
8M~16M	5145700	539456	7052420	12737576	462363	51134	623091	1136588
16M~32M	15355044	489220	8405759	24250023	730973	20386	413288	1164647
32M~64M	2857115	3577955	12114685	18549755	65648	72034	272932	410614
64M~128M	54719507	3479538	24667014	82866059	792423	43975	309471	1145869
128M~256M	11788761	3979406	69231895	85000062	53105	20565	368266	441936
256M~512M	47343780	11586257	88086593	147016630	121456	33950	233732	389138
512M~1G	69320024	66040762	288040427	423401213	93273	96133	485010	674416
1G以上	439872979	243533944	2911469577	3594876500	57798	34605	704286	796689
				4756611955				1122579745

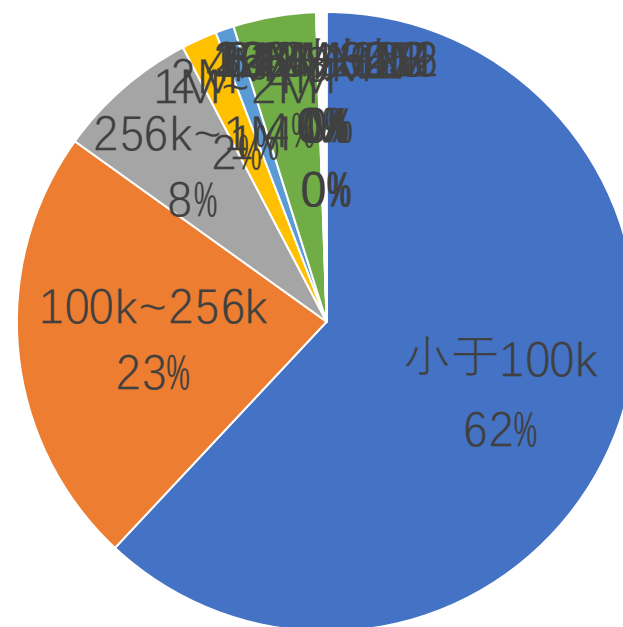
# 面向真实业务场景的IO建模

### 文件大小分布



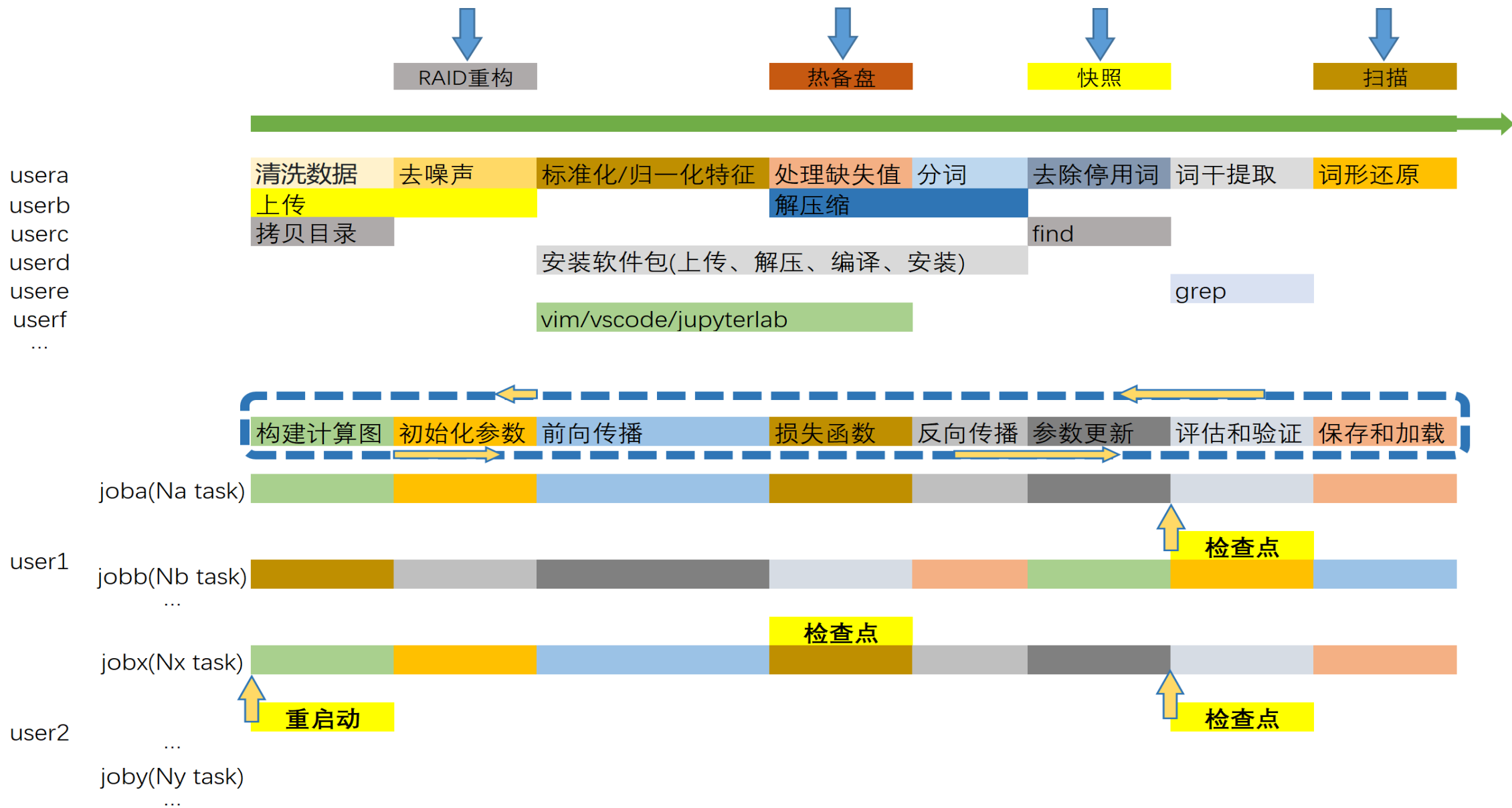
- 小于100k
- 100k~256k
- 256k~1M
- 1M~2M
- 2M~4M
- 4M~8M
- 8M~16M
- 16M~32M
- 32M~64M
- 64M~128M
- 128M~256M
- 256M~512M
- 512M~1G
- 1G以上

### 文件个数

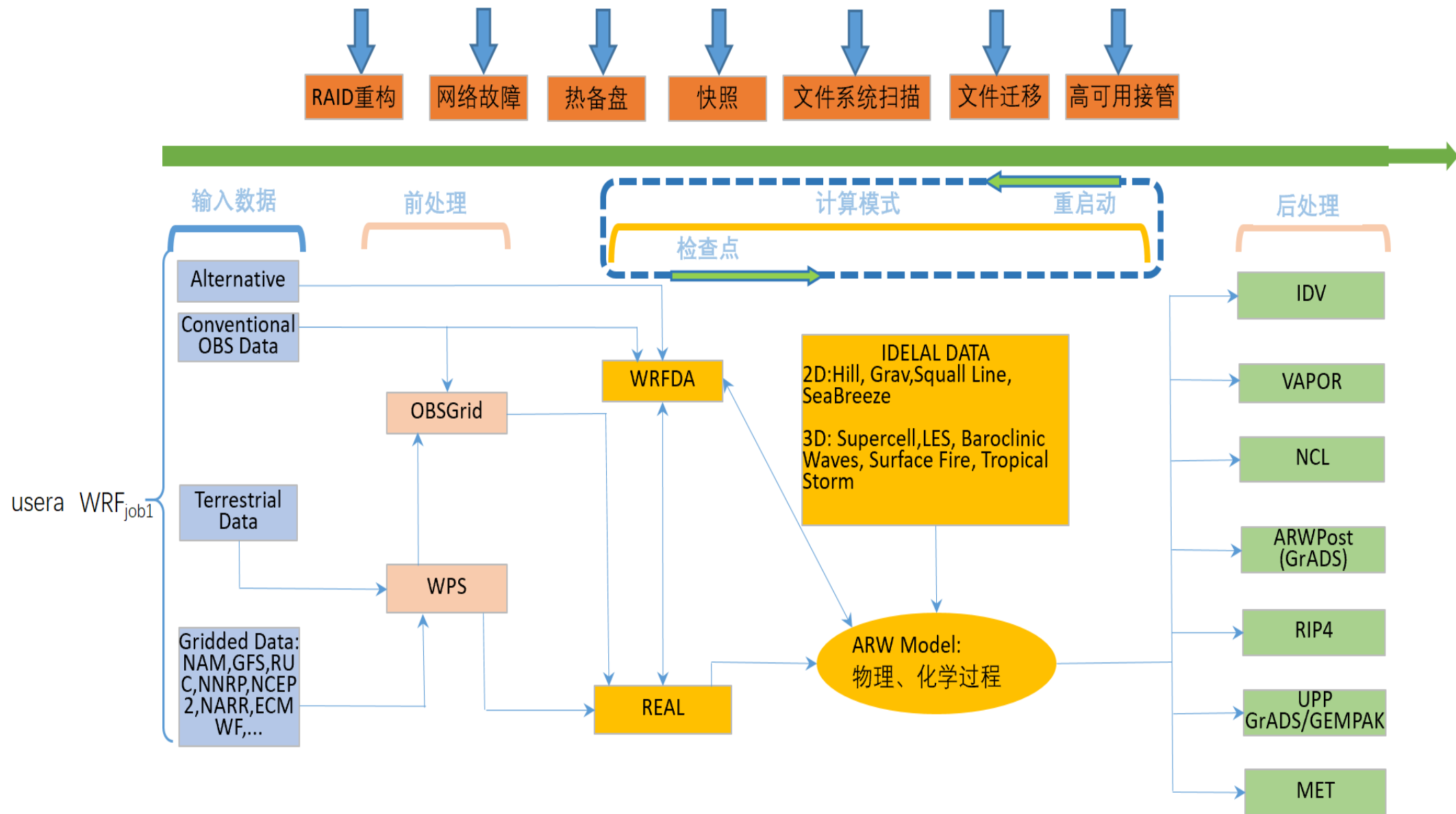


- 小于100k
- 100k~256k
- 256k~1M
- 1M~2M
- 2M~4M
- 4M~8M
- 8M~16M
- 16M~32M
- 32M~64M
- 64M~128M
- 128M~256M
- 256M~512M
- 512M~1G
- 1G以上

# 面向真实业务场景的IO建模



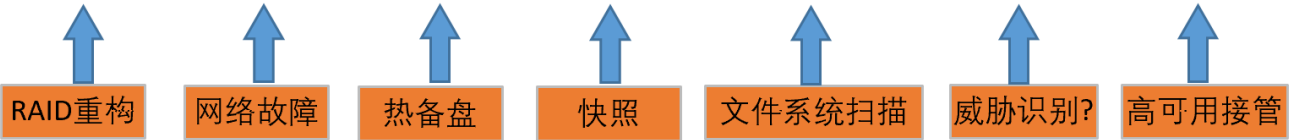
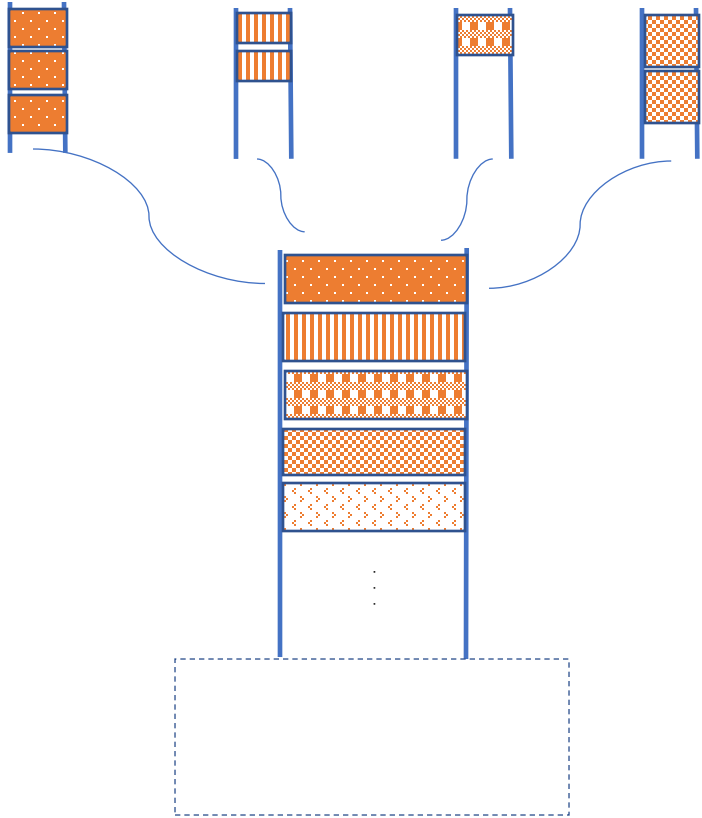
# 面向真实业务场景的IO建模





# 面向真实业务场景的IO建模

gjob1	grep, 循环执行, 每次执行完后要清理内存中的缓存, 确保从物理文件中实际读取内容
gjob2	grep, 循环执行, 每次执行完后要清理内存中的缓存, 确保从物理文件中实际读取内容
...	
sjob1	stat, 循环执行, 每次执行完后要清理内存中的缓存, 确保从物理文件中实际读取内容
sjob2	stat, 循环执行, 每次执行完后要清理内存中的缓存, 确保从物理文件中实际读取内容
...	
fjob1	find, 循环执行, 每次执行完后要清理内存中的缓存, 确保从物理文件中实际读取内容
fjob2	find, 循环执行, 每次执行完后要清理内存中的缓存, 确保从物理文件中实际读取内容
...	
trjob1	tar, 压缩文件到存储目录, 循环执行, 每次执行完后清理内存中的缓存, 保留生成的压缩文件
trjob2	tar, 压缩文件到存储目录, 循环执行, 每次执行完后清理内存中的缓存, 保留生成的压缩文件
...	
twjob1	tar, 解压缩文件到存储目录, 循环执行, 每次执行完后清理内存中的缓存
twjob2	tar, 解压缩文件到存储目录, 循环执行, 每次执行完后清理内存中的缓存
...	
cpjob1	cp, 拷贝文件到存储目录, 循环执行, 每次执行完后清理内存中的缓存
cpjob2	cp, 拷贝文件到存储目录, 循环执行, 每次执行完后清理内存中的缓存
...	
mjob1	合并文件, 将分散到各个目录的文件按照一定的大小进行合并, 保存到新目录里, 循环执行
mjob2	合并文件, 将分散到各个目录的文件按照一定的大小进行合并, 保存到新目录里, 循环执行
...	
trainjob1	有x个任务在启动/重启动, 有y个任务在训练中, 每个任务的并发度是 $C_1, \dots, C_{x+y}$ , 每个任务的训练时间为 $T_1, \dots, T_{x+y}$
trainjob2	



# 自动化评测工具

**目标：**应满足任务设计和编排、任务调度和执行、负载生成、数据采集、数据分析和评估、报告输出等评测需求，同时支持测试流程自动化、结果输出标准化；各功能模块的数据及配置信息采用标准格式文件，支持开放定义；测评工具支持标准开放接口支持不同存储设备的接入及自动测试。

## 评测系统总体架构图



## 易部署、易使用、易维护

- ✓ 基于Python/java，实现测评场景构造、测试、评估。
- ✓ 方案能力组件解耦，灵活编排内外部工具。
- ✓ 测评用例以xml配置的形式承载，具备高可维护性。
- ✓ 外部接口可通过配置，实现不同存储特性拓展。

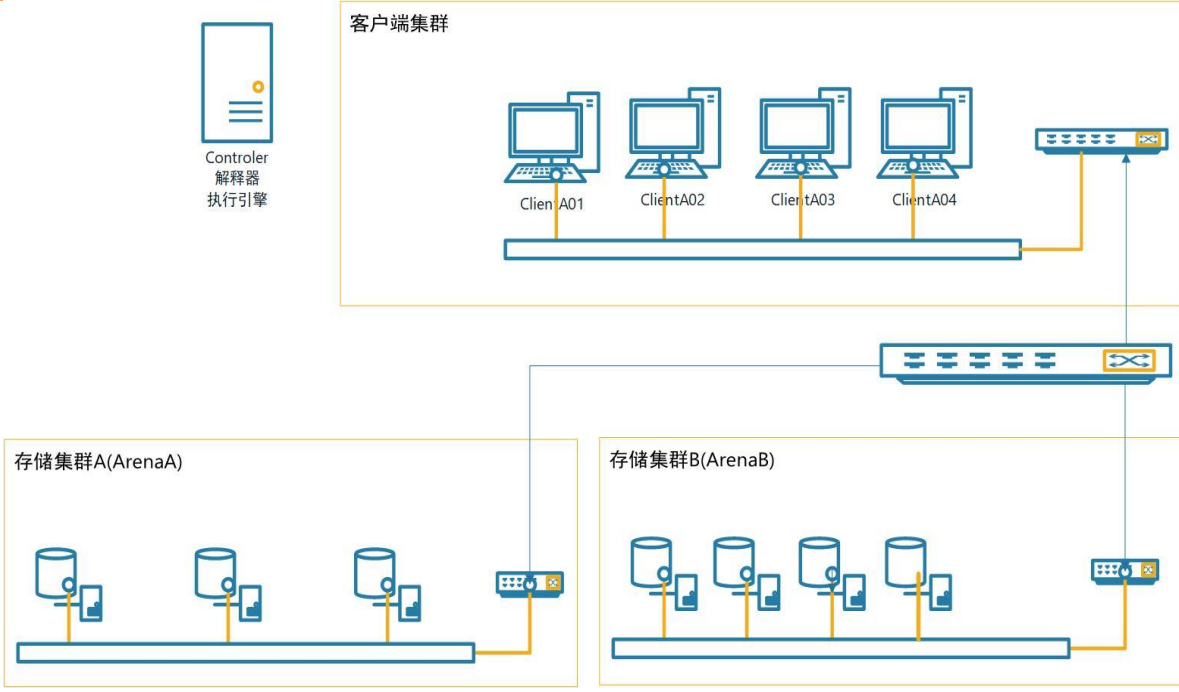


# 自动化评测工具



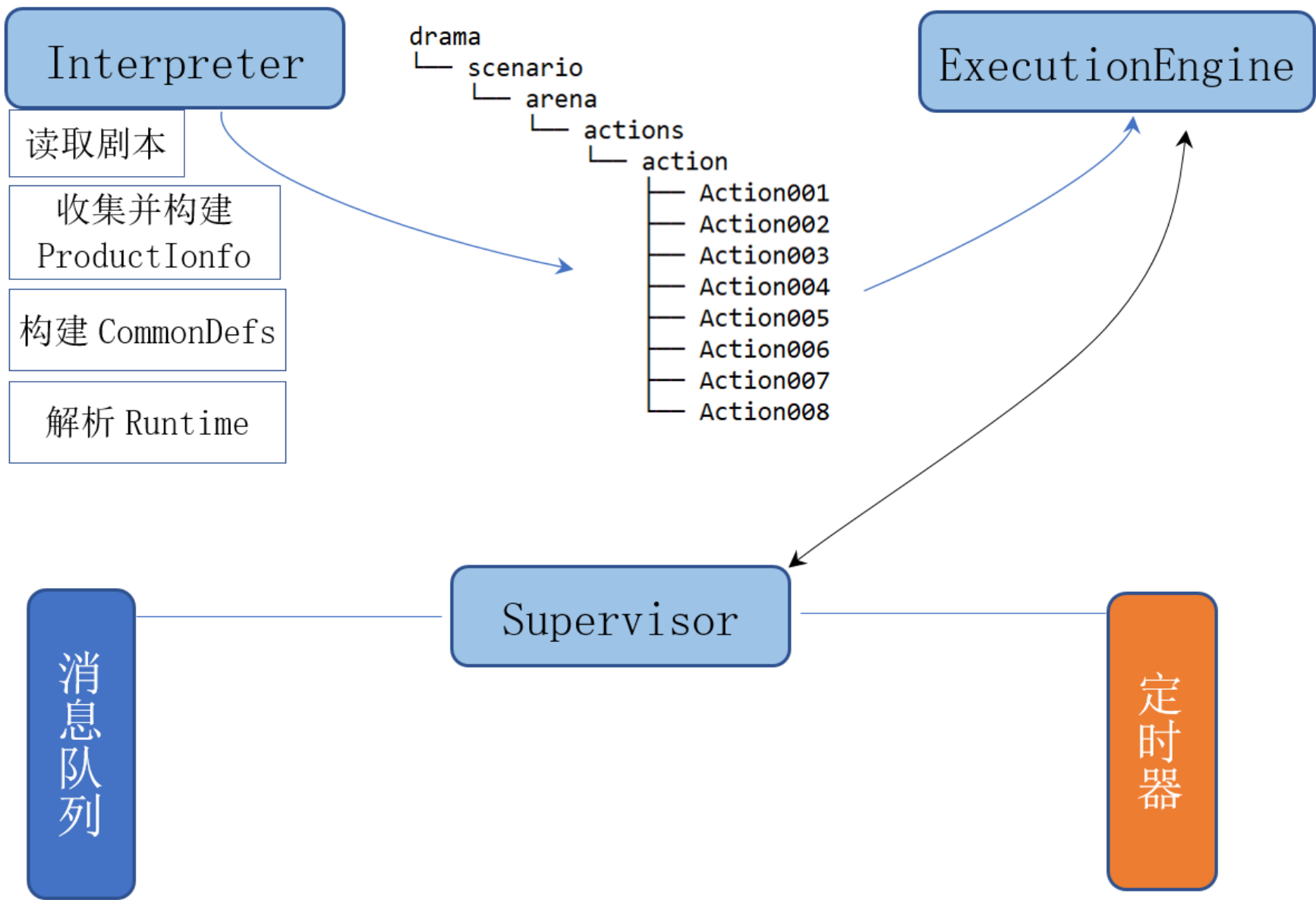
timeline

- clientnode 001
- clientnode 002
- clientnode 003
- clientnode 004
- .....



Action	Action	Action	Action	Action	Action	Action	Action	Action	Action
Arena(Device 1)		Arena (Device 2)		Arena (Device 3)		Arena (Device 4)		Arena (Cluster A)	
Senario(HPC)				Senario(AI)...				Senario(HPL)...	
Drama(Storage)								Drama(Cluster)	

# 自动化评测工具

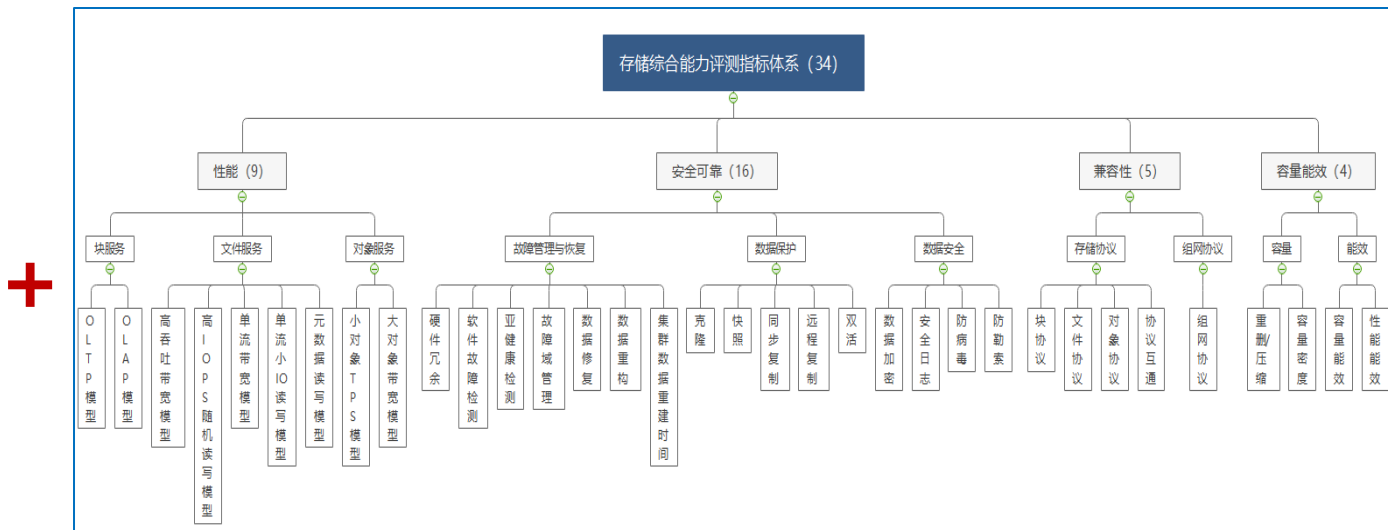


# 目标：构建科学实用的工具，形成贴合业务场景的存储评测规范、方法

从系统特性和实际应用场景IO特征、实际业务应用条件出发，充分考虑实际业务中各个指标的重要程度以及考虑指标类之间、同一指标类各指标项之间的关联程度，建立一个贴近业务、科学实用的评测规范：基础评价指标和方法、组合测试/场景测试评测标和方法，设计整体评测流程，为存储系统提供前瞻性、差异化、场景化的综合能力度量。



典型业务场景



评测指标体系

# 目录

---

一：存储基础设施重要性

二：存储典型应用场景能力需求

三：存储评测规范、工具与方法

**四：测试案例**

# 测试案例

## AI场景存储测评报告

### 总览

测试开始时间: 2024-09-20 15:17:53

测试结束时间: 2024-09-20 16:25:17

测试总时长: 1:07:23

用例总数: 10

用例总览:

用例	类型	是否必选	指标
AI训练模型规模测试	性能测试	是	带宽/容量 (MBPS/TB) : 0.59 OPS/容量 (OPS/TB) : 5.79 时延(ms): 2.17
多AI训练任务并发性能测试	性能测试	是	带宽/容量 (MBPS/TB) : 0.1 OPS/容量 (OPS/TB) : 0.92 时延(ms): 4.74
系统数据盘故障	可靠性测试	是	
存储接口卡故障	可靠性测试	是	
存储RAID/EC重构	可靠性测试	是	
控制器/节点故障	可靠性测试	是	
快照功能	数据安全测试	是	
防勒索检测	数据安全测试	是	
基础能效	能效测试	是	
节点在线扩展	扩展性测试	是	

### AI训练模型规模测试

开始时间: 2024-09-20 15:17:53

结束时间: 2024-09-20 16:06:53

执行时长: 0:48:59

执行结果: Pass

### 多任务混合负载-B

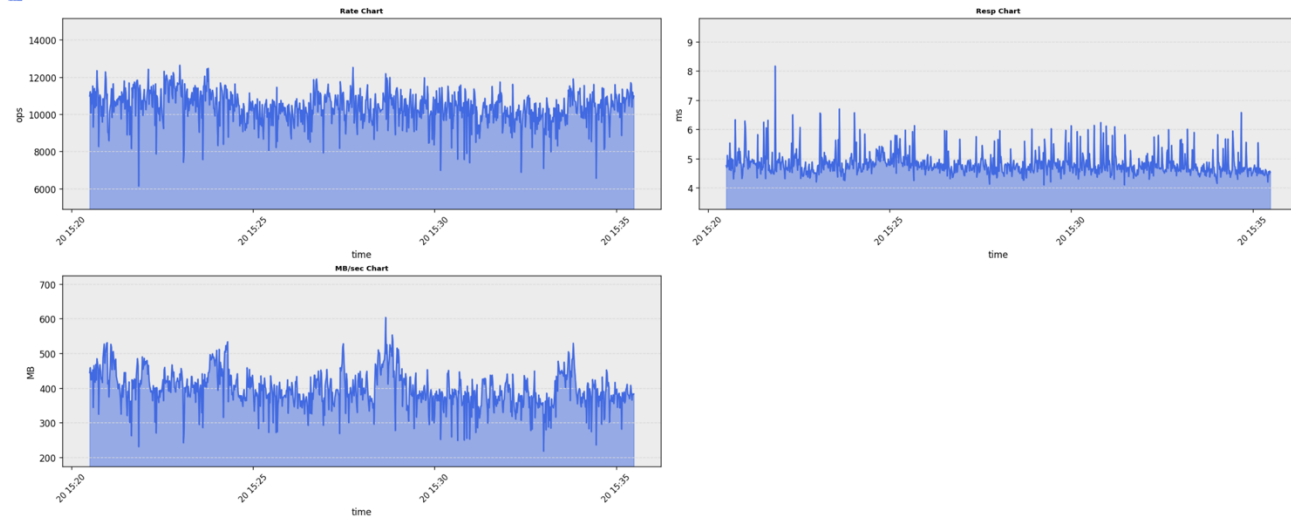
Rate	MB/sec
37914.1622	3838.6068

### AI模型: 元数据导入

Time	Rate	Resp	MB/sec	Getattr_rate	Getattr_resp	Write_rate	Write_resp	MB_write	Read_rate	Read_resp	MB_read
2024-09-20 15:35:29	10375.6733	4.796	396.2993	2057.2078	0.0099	3167.9633	14.186	249.631	5150.5022	0.932	146.6684

### 性能数据图

概览



谢谢