保存基准样例并验收资源变更前后的环境与关键输出

保存迁移基线，按输入、ID与事先定义的数值规则验收，再测扩容后的完整性能。

扩容或换机器后，先证明旧流程仍按约定得到正确结果，再比较速度。只检查软件能启动、文件大小接近或相关性很高，都可能漏掉样本错位与关键差异。本篇在变更前保存基准样例，用明确规则验收输入、环境和关键输出，区分允许差异与真正故障。

变更前准备能重跑的基线

保存代表输入与数据库指纹、参数、软件及包版本、工作流、随机设置、线程、运行命令和关键结果。样例应覆盖实际步骤及易出问题的数据，不只留一个帮助命令。列明哪些输出必须精确相等，哪些连续数值允许事先批准的误差，哪些随机算法需要重复运行与分布评价。

对象	本例约定	验收要点
输入与标签	字节指纹、样本/基因集合和顺序	防止换文件或错配ID
环境	依赖、系统、CPU/GPU和驱动	包列表相同不等于所有底层相同
输出	完整度、结构和关键数值	不只比较文件体积或相关性
性能	同方法完整时间与资源峰值	结果验收先于加速结论

先验输入，再按关键字段比较

在新环境先检查基线输入校验，并执行同一工作流。二进制容器格式或PG记录变化时，不宜只用整个输出文件哈希判断分析等价；选择有科学含义的验收字段。下面仅比较feature,value两列连续数值CSV，按ID对齐并拒绝重复、缺失和非有限值，容差由项目事先指定。

```
import csv,math,sys
def values(path):
    with open(path,newline='') as f:
        reader=csv.DictReader(f);assert reader.fieldnames==['feature','value']
        out={}
        for row in reader:
            key=row['feature'];assert key and key not in out
            value=float(row['value']);assert math.isfinite(value)
            out[key]=value
    assert out
    return out
old,new=values(sys.argv[1]),values(sys.argv[2])
rtol,atol=map(float,sys.argv[3:5])
assert all(math.isfinite(v) and v>=0 for v in [rtol,atol])
assert old.keys()==new.keys(),'feature set changed'
failed=[k for k in old if not math.isclose(old[k],new[k],rel_tol=rtol,abs_tol=atol)]
print('features',len(old),'failed',len(failed),
      'max_abs_error',max(abs(old[k]-new[k]) for k in old))
assert not failed,failed[:10]
```

例如保存为compare_values.py后，使用python3 compare_values.py old.csv new.csv 1e-6 1e-8；这两个容差仅是命令示例，不是通用标准。整数计数需要精确整数比较，不能先转float损失大整数精度；矩阵、VCF或结构预测则应另定义对应的标签、范围与质量规则。

对差异逐层定位

先排输入/参考/参数变化，再看版本、BLAS或底层库、线程、随机性、CPU指令与GPU实现。行顺序变化不必然是数值变了，但ID集合或样本映射改变必须解释。不能删掉不一致行、临时放宽容差或关闭必要步骤，让迁移测试看起来通过。

对随机或并行算法，固定种子也未必保证不同构建逐字节一致，应按该方法的可复现范围设计重复测试，保留分布、关键结论和失败案例。若参考或方法升级，应另列方法变更，不能把所有差异归到硬件扩容，也不能称为完全复现旧协议。

验收后再决定切换和资源配置

记录测试覆盖、输入规模、实际环境、检查项、结果与未解决差异。通过小样例后再验证代表完整任务及目标并发，测完整时间、作业RAM和临时盘；小样例低峰值不能线性保证全规模配置。旧环境和回退材料按约定保留，切换后继续核首批真实任务。

基线应随正式方法版本保存，而不是每次出问题后临时重建。明确谁批准验收、谁处理差异、何时切换及怎样回退，能够让后续加内存、换GPU或迁移服务器都有可检查依据，资源推荐也因此对应真实流程和容量需求。

参考资料

Python math.isclose官方容差定义 https://docs.python.org/3/library/math.html#math.isclose

Python CSV按字段读取 https://docs.python.org/3/library/csv.html

GNU SHA2输入指纹 https://www.gnu.org/software/coreutils/manual/html_node/sha2-utilities.html

conda-pack迁移的可复用条件 https://conda.github.io/conda-pack/