VIDA 分布式视频处理框架产品白皮书

文档名称	Vida V1.0 分布式媒体处理框架产品白皮书
发布者	成都索贝数码科技股份有限公司
日期	2018-9-28
版本	Version 1.0

sobey
www.sobey.com

目　录

第一章 概述 2

第 1 节 背景 2

第 2 节 目标 2

第二章 VIDA 分布式视频处理框架 3

第三章 VIDA Grid 存储系统 4

第 1 节 系统架构 4

第 2 节 产品特性 4

第四章 VIDA MapReduce 计算系统 7

第 1 节 分布式并行视频处理耗时构成 7

第 2 节 整体处理效率计算 8

附录：测试记录 9

第 1 节 测试配置 9

第 2 节 操作步骤 10

第 3 节 测试数据 10

第一章 概述

第1节 背景

随着时代的发展，对视频大数据的存储、分析和挖掘的需求已摆在媒体行业的面前。视频数据的数据量大、传输负载重、计算耗时等特点，使得传统的单节点、多任务的处理办法难以满足高效快速处理的需求。特别是针对视频的单一处理任务，解决的办法有限。

对于视频的单一处理任务，处理思路一般集中在两点

- 并行化计算过程
- 提高文件系统的读写效率

以上两种方式都可以在一定程度上提升视频处理倍数。但若是不改变以往视频处理框架和文件存储方式，对于 100Mbps~1500Mbps 量级的广电 HD/UHD 视频素材，处理速度达到一定水平后就会触及瓶颈。

原因在于节点的增多也使得分布式处理过程中**分片、传输、拼接过程消耗增大**，消耗将抵消多节点计算提升的效率，也就是说无法简单地通过增加系统中节点来继续提升处理效率，如何突破这一瓶颈成为了媒体行业很有研究意义的课题。

第2节 目标

索贝公司在上述背景下提出了 VIDA 分布式视频处理框架，主要目的是突破单一视频处理任务瓶颈，充分发挥分布式结构优势，**实现视频计算平台的性能随节点“准线性”扩展**，最终实现数十倍乃至上百倍的高码率处理效率；另一方面该框架也提供更加适合视频存储、具有更高可靠性和扩展性的存储能力。

第二章 VIDA 分布式视频处理框架

VIDA 分布式视频处理框架由多个超融合架构组成，框架在逻辑上分为 **VIDA MapReduce 计算系统**和 **VIDA Grid 存储系统**，如图 1 所示。

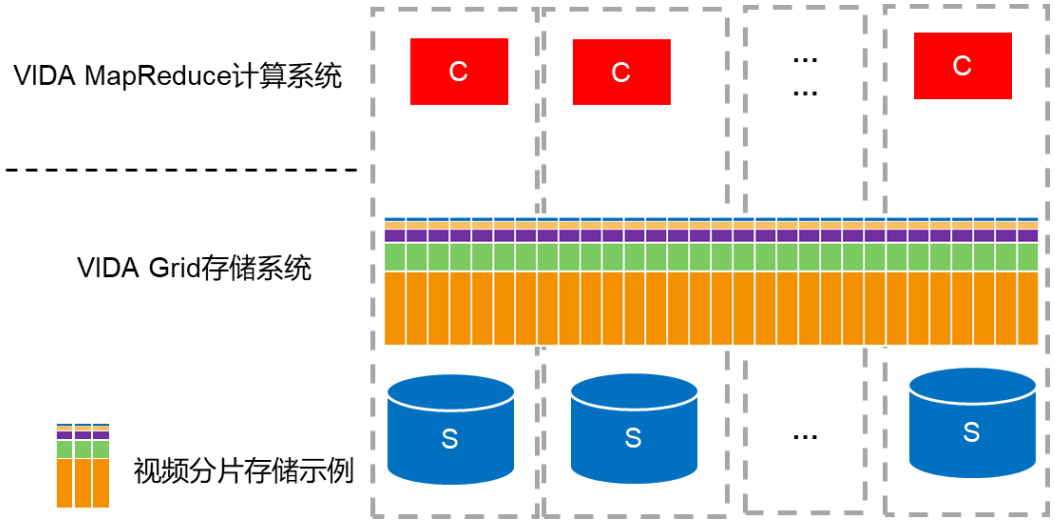


图 1 架构示意图

- **VIDA Gid 存储系统：**

由视频对象和存储节点构成，一方面视频对象经过时域上的划分后，各时段内容和不同码率的视音频文件存储在不同节点中，以快速地不同场景下提供对应码流内容。

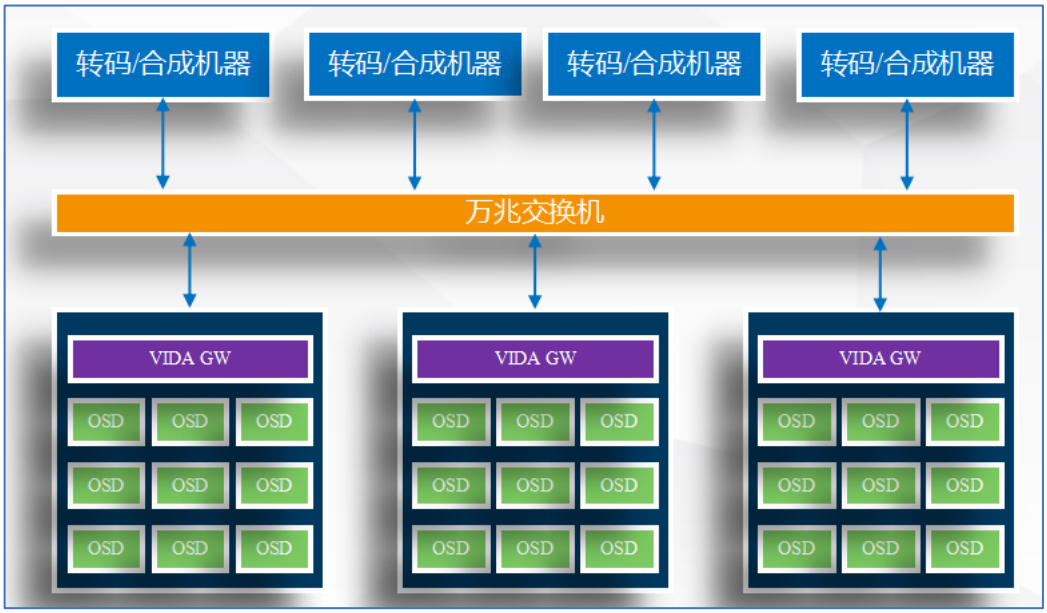
- **VIDA MapReduce 计算系统：**

将多个融合架构中的计算核心汇聚起来并进行有效管理，用于分布式视频计算。

第三章 VIDA Grid 存储系统

VIDA Grid 分布式视频对象存储系统，是一个专门针对视频打造的，具有高性能、高可用、高扩展特点的对象存储系统。VIDA Grid 针对视频特性综合了视频文件格式、对象存储、高性能并行计算等多方面的技术。VIDA Grid 作为业内首款面向视频处理的分布式对象存储系统，既具有对象存储的优势，也拥有索贝视频技术的加成，使存储理解视频文件，将显著提高视频制作效率，可广泛应用于广电媒资、高性能计算、数据中心、互联网运营和大型企业等各行业领域，助力超高清时代发展。

第1节　系统架构



VIDA Grid 基于 OSD 管理所有存储在系统中的用户数据。每个节点包含若干对象存储设备 OSD，用户的数据都存在 OSD 上。用户数据通过 VIDA GW 网关程序和客户端进行交互，VIDA GW 支持按标准 S3 协议进行数据读写，并在此基础之上实现数据的追加写、修改和闪拼等操作。

第2节　产品特性

高性能

VIDA Grid 通过业务集群和存储集群提供基础的能力支撑，具备高并发、高吞吐的服务特点。

- 扁平化存储结构，不会因碎片化小文件多，目录结构复杂而导致性能出现瓶颈。
- 分布式部署，对对象进行切片存储，具备极高的读写吞吐能力。
- 3 节点读带宽可达到 3300MB/s，写带宽可达到 1800MB/s，性能随节点数增加而提高。

高可靠

VIDA Grid 通过提供对象数据多份冗余或者纠删码和保证多份对象的数据一致性来提供对象数据的高可靠性。拥有负载均衡机制。系统中任何一个组件（包括硬盘、节点）发生故障，系统将自动检测、隔离该故障；如果出现磁盘或节点失效，系统自动通过其他节点在系统其余的可用空间中快速恢复受故障影响的数据。

- 数据高可靠，多重权限控制策略，保障数据安全。
- 强大容错性，数据多重备份，自我恢复机制，节点故障无数据丢失。
- 服务高可用，通过负载均衡设计，可用性达到 99.99%。

高扩展

VIDA Grid 所有业务、存储节点采用分布式集群方式工作，各功能节点、集群都可以独立扩容，整个扩容过程对用户完全透明。

- 具有极佳的在线扩展能力。在不中断业务的情况下，实现存储空间的大规模在线扩展，存储空间可以从几十 TB 线性扩展到几十 PB 甚至更多。
- 采用分布式架构，各控制节点之间完全独立，当数据的读写性能或网络带宽不足时，可简单的在集群中增加控制节点来实现性能的线性增长。

S3 标准协议

VIDA Grid 支持 S3 标准协议。

- 支持 AWS version2 和 version4 两种签名方式
- 可以使用 S3 协议进行桶管理、对象管理、权限控制
- 支持 S3 browser 等第三方工具上传下载

数据追加和修改

VIDA Grid 可对已经封口的文件追加写入数据。

VIDA Grid 可对已经封口的文件任意位置的任意数据进行修改。

文件闪拼

分片转码/合成后的文件需要拼接成完整文件，VIDA Grid 可将若干分片文件瞬间合并为目标文件。

- 支持多种 4K 文件格式拼接
  - ✓ MXF
  - ✓ MP4
  - ✓ MOV
- 大数据量文件秒级别拼接
  - ✓ 以对象引用方式拼接，极少存储空间占用
  - ✓ 支持大量分片文件瞬间拼接为目标文件

断点续传

当上传意外终止，用户再次上传该文件时，可以从中断处继续上传，减少重复上传时间。

在下载或上传时，如果碰到网络故障，可以从已经上传或下载的部分开始继续上传下载未完成的部分，不必从头开始上传下载。

第四章 VIDA MapReduce 计算系统

VIDA 拥有针对视频特性而设计的一套先进的技术架构，它综合了高性能并行计算、视频文件格式、对象存储等多方面的技术。打破传统顺序视频处理方式，结合大数据批处理机理和 CPU+GPU 分片并行转码/合成操作，线性提高视频的转码/合成处理效率，运用对象引用技术，使分片文件瞬间形成目标文件，称之为 VIDA MapReduce。该项计算系统的设计目标，是实现视频运算性能随设备数量的扩展“准线性”扩展。

第1节 分布式并行视频处理耗时构成

如果对一个视频内容进行分布式并行处理，耗时包括任务调度分配耗时、执行器任务启停耗时、视频处理耗时、文件拼接耗时。

- **任务调度分配耗时**

这个过程指一个任务被调度管理器分配给执行器，由于它仅仅是管理数据的运算，相对简单，耗时是一个几乎固定常量,耗时一般在 1 秒以内，记为 t1。

- **执行器任务启停耗时**

这个过程一般包括执行器的视频文件头部读取和分析、数据结构的构建、解码器动态加载、完成后的文件封口等，耗时一般在 5 秒以内，记为 t2。

- **视频处理耗时**

这个过程是视频处理最主要耗时，消耗 CPU、GPU、磁盘高负荷运转，并以最大化利用硬件运算能力为目标，记为 t3。

- **文件拼接耗时**

这个过程是指各个执行器执行完的结果，需要拼接成整体文件，这个过程由一个单一执行器，需要将所有结果文件从磁盘中读出，拼接成一个文件重新写入，记为 t4，对于高码率的视频文件，该过程使用传统拼接方法耗时有时候甚至不亚于视频处理耗时。

第2节　整体处理效率计算

综上所述，分布式视频处理时长公式为

$$t = t1 + t2 + t3 + t4$$

假设 k 为单台机器视频处理速度（固定），待处理视频文件长度为 d（单位：秒）；n 为执行器数量；则 t3 可表示为：

$$t3 = \frac{d}{kn}$$

整体处理效率 K 表示为：

$$K = \frac{d}{t1 + t2 + {}^{d}/_{kn} + t4}$$

如果$\frac{d}{kn} \gg t1 + t2 + t4$, 有$K \approx kn$，即实现准线性增长。

基于目前的处理能力，在大多数视频处理任务中，$k \subset (0.5,2)$, VIDA MapReduce 技术下可实现 t1<1s，t2<3s，**借助于 VIDA Grid 的闪拼能力可实现 t4<2s**，即$t1 + t2 + t4 < 6s$，因此当$\frac{d}{kn} > 60s$时，可认为准线性，一般的，**对一个 m 分钟素材，当执行器数量$n < m$时，可认为处理性能随 n 准线性增长。**测试结果证明了我们的结论。

附录：测试记录

第1节　测试配置

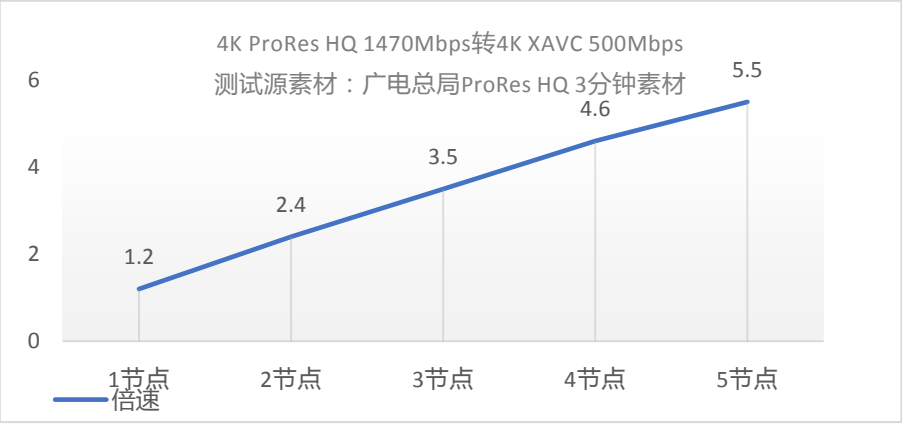
序号	设备名称	配置说明	型号
1	分布式媒体对像存储	Intel Xeon Silver 4114 2.2G, 10C/20T×2	DELL PowerEdge R740XD
		内存：16GB RDIMM×8, 2666MT/s, Dual Rank	
		系统硬盘：960GB SSD×2	
		数据硬盘：4TB 7.2K RPM NLSAS 12Gbps 512n 3.5in Hot-plug Hard Drive×12	
		网卡：Intel X520 DP 10Gb DA/SFP+ Server Adapter×2	
		其它：冗余电源	
		VIDA GRID 分布式媒体对像存储软件	
2	分布式渲染集群节点	CPU：INTEL Xeon 6132 2.6GHz（十四核）×2	HP Z8G4
		内存：8GB DDR4-2666 ECC 内存 × 24（共196GB）	
		系统盘： 480 GB SATA Enterprise SSD 固态硬盘×1	
		内置数据盘：1TB 7200 rpm SATA 硬盘 * 1	
		显卡：AMD Radeon Pro WX9100 16GB 专业图形显卡 ×1	
		其他：集成声卡、集成网卡、鼠标键盘	
		操作系统：Windows 10 Pro for Worksatations 64bit	
		万兆网卡 intel x520DP	
		电源：1125 瓦	
		VIDA MapReduce 分布式渲染集群软件	
3	分布式转码集群节点	CPU：INTEL Xeon 6132 2.6GHz（十四核）×2	HP Z8G4
		内存：8GB DDR4-2666 ECC 内存 × 12（共96GB）	
		系统盘： 480 GB SATA Enterprise SSD 固态硬盘×1	
		内置数据盘：1TB 7200 rpm SATA 硬盘 * 1	
		显卡：NVIDA P2000 ×2	
		其他：集成声卡、集成网卡、鼠标键盘	
		操作系统：Windows 10 Pro for Worksatations 64bit	
		万兆网卡 intel x520DP	
		电源：1125 瓦	
		VIDA MapReduce 分布式转码集群软件	

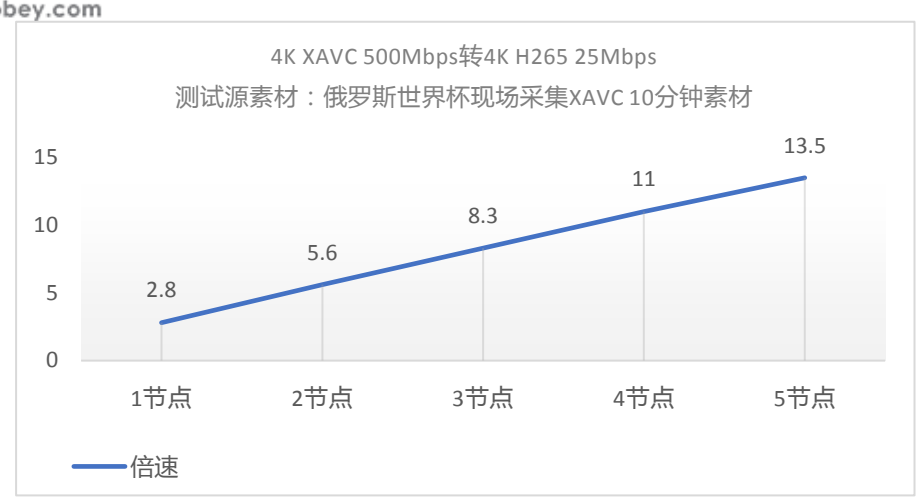
第2节　测试步骤

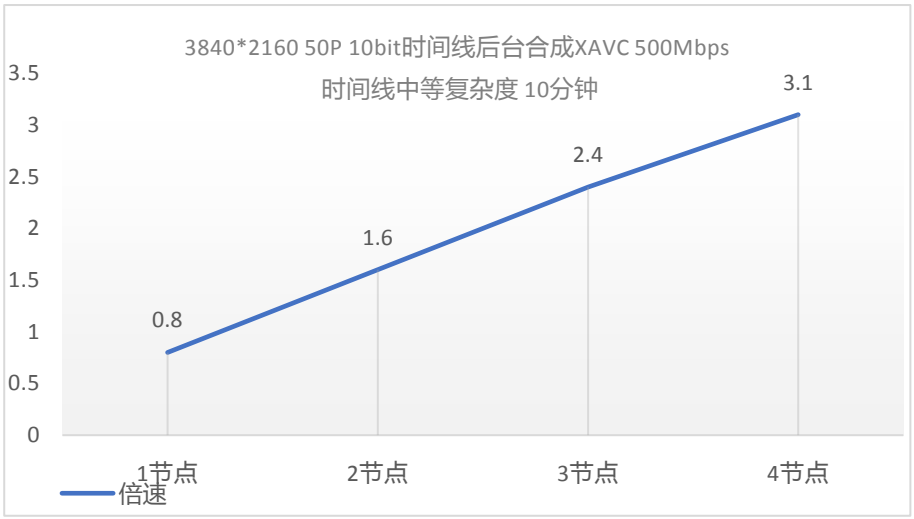
转码测试操作步骤	
1	添加转码策略：在 VIDA MPC web 页面，选择“策略管理”后，添加转码策略，指定视频转码格式
2	发起转码任务：在 VIDA MPC web 页面，选择“任务管理”→“添加任务”，输入任务名称，选择需要转码的视频和策略，点击“提交”
3	查看任务进度：在 VIDA MPC web 页面，选择“任务管理”，查看任务进度

合成测试操作步骤	
1	使用 MCH-4K 系统中 NOVA11 制作一条 3840*2160 50P 10bit 中等复杂度时间线
2	使用 MCH-4K 系统中 NOVA11 发起后台合成
3	查看任务进度：在 VIDA MPC web 页面，选择“任务管理”，查看任务进度

第3节　测试数据







根据以上测试数据得出结论：转码/合成速度随节点数增加呈线性增长趋势。

第四章第 2 节整体处理效率计算中，通过公式推理，推断出：转码/合成性能会随机器增加呈“准线性增长”趋势。

测试结果与公式推理结果一致。