BlueWhale蓝鲸

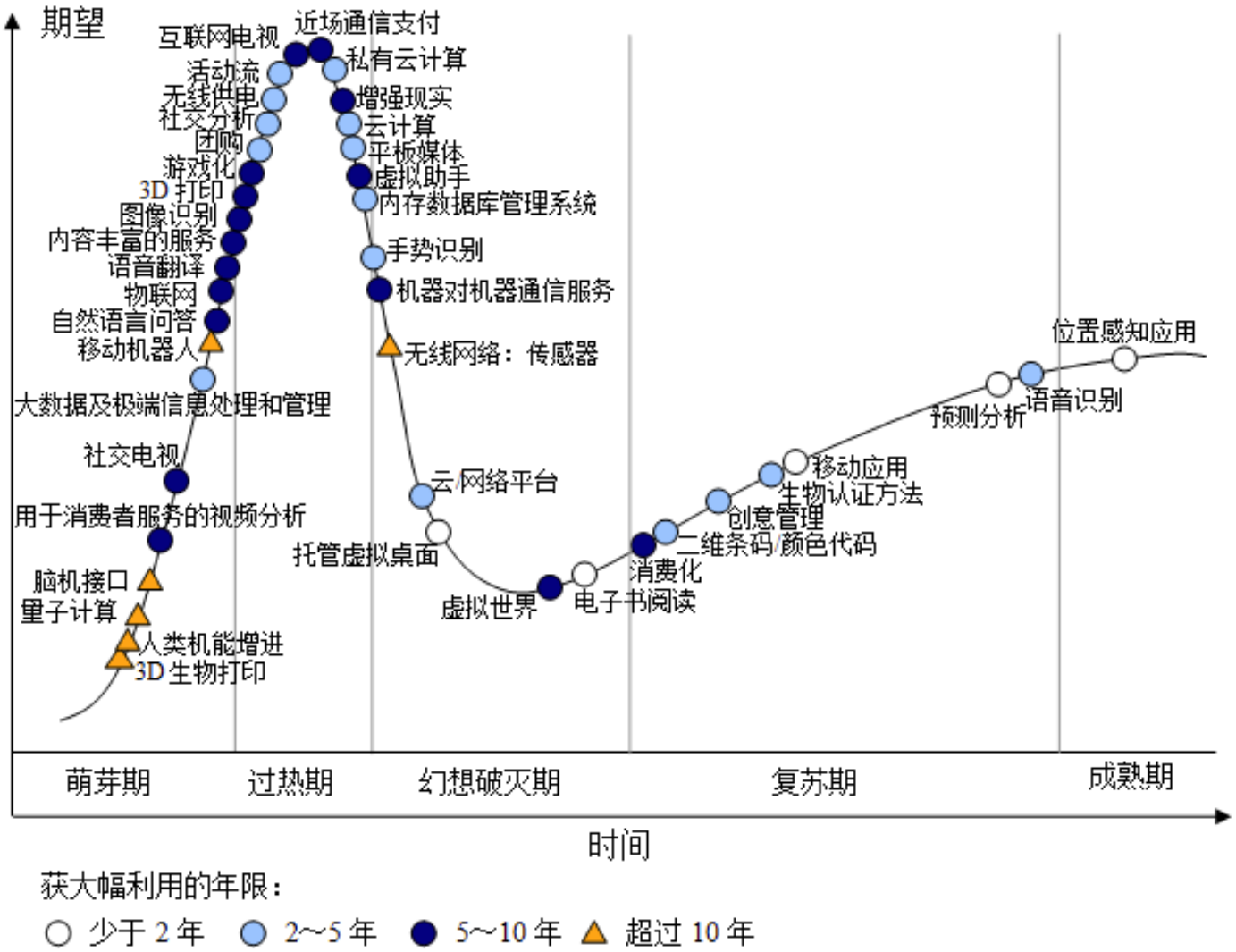
面向云应用环境的存储系统技术展望

许鲁

xulu@bwstor.com.cn

天津中科蓝鲸信息技术有限公司

Gartner 2011新兴技术成熟度曲线



版权所有@2010-2011天津中科蓝鲸信息技术有限公司

云应用环境的特点

- 规模大
  - 大量的计算节点
  - 大量的存储节点
  - 大量的应用种类
  - 大量的最终用户
- 应用深
  - 纵深的应用层次
  - 衍生的新型应用
- 要求高
  - 低成本的要求高
  - 可靠性的要求高
  - 易管理的要求高
  - 低功耗的要求高
  - 适用性的要求高

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

存储的重要性

- 数据的重要性
- 存储的重要性
  - 存储是数据的载体

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

存储系统在云应用环境中的重要性

- 低成本的影响
  - 成本主要来源之一
- 可靠性的影响
  - 数据是重要的资产
  - 服务可靠性的基础
- 易管理的影响
  - 系统管理主要难点
- 低功耗的影响
  - 功耗主要来源之一
- 适用性的影响
  - 主要限制因素之一

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

环境发展趋势

- 应用规模化
  - 服务器与存储的分与合
    - “分”：机会
    - “合”：发展
- 部件大众化
  - CPU、内存、各种器件
  - 控制器、电源、机箱
- 存储软件化
  - 在以通用部件为基础的硬件环境中，解决存储系统的可靠性、高性能、可扩展等问题

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

主要内容

- 局域存储系统——关注性能、扩展性、可靠性、可用性、低成本问题
  - 存储设备集群
  - 元数据服务集群
  - 文件服务集群（NAS集群）
- 局域数据管理——关注此基础上的多维视图、资源利用、可靠性问题
  - 通用索引机制
  - 自动分级（包括生命周期）、去冗
  - 备份、副本
- 广域存储系统——关注多中心间的数据共享、分布、传输、安全问题
  - 统一文件视图
  - 按需数据分布
  - 按效数据传输
- 广域数据管理——关注此基础上数据布局、副本控制、传播跟踪问题
  - 按照策略要求将数据在中心间位置布局
  - 按照策略要求将数据在中心间副本布局
  - 按照策略要求将数据在中心间复制跟踪

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

数据存储系统挑战

- 满足成本要求
- 数据量迅速膨胀
  - 海量存储空间
  - 容量的按需扩展
- 多种访问模式且量大
  - 高并发高带宽低时延
  - 性能的按需扩展
- 业务连续性要求高
  - 高可靠，容错能力强
  - 故障恢复对系统影响低
  - 系统无缝扩展
- 自动化系统管理

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

存储系统的特点

- 高性价比
- 高可靠性
- 高适用性
- 高可控性
- 高可动性
- 可伸缩性
- 可管理性
- 可集成性

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高性价比的存储系统

- 高性能、高利用率、低成本
  - COTS
- 基本点

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 磁盘级
- 设备级
- 集群级
- 文件级

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 磁盘级
  - LSE/ME
    - Scrubbing
    - IDR
    - Vertical Parity

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 设备级
  - RAID
    - 磁盘局部错误的处理与恢复
    - 重构过程的粒度化和渐进式重构过程
    - 重构过程的并行化
      - ✓ 大量磁盘资源的利用
  - N+M
    - 纠错码（Erasure codes）实现的一种实际抽象
    - N是数据块的数量，M是校验块的数量
    - Reed Solomon码使用多项式 计算，需要更强的计算能力

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 集群级
  - 设备间
    - Mirror
    - RAID（N+M）
  - 多路径支持
  - SAN级访问接口

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 文件级
  - ObjectRAID
    - N+M
  - 文件服务的高可用
  - 文件访问接口

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 元数据服务集群
  - 问题
    - 大规模数据处理集群支持及可靠性
  - 挑战
    - 动态能力
    - 服务开销
    - 复杂程度
  - 技术
    - 元数据服务粒度
    - 元数据服务分布
    - 元数据服务聚合
    - 元数据服务均衡
    - 元数据服务接管

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可靠的存储系统

- 文件服务集群
  - 问题
    - 大众化
  - 挑战
    - 文件服务粒度
    - 文件服务聚合
    - 文件服务均衡
    - 文件服务接管

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高适用的存储系统

- 问题
  - 多态环境
  - 众多应用
- 技术
  - 面向应用边界的资源管理
  - 面向应用模式优化的资源管理
    - 支持应用定制的资源管理

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可控的存储系统

- 问题
  - 资源共享
    - 资源共享的充分性
    - 资源共享的低廉性
    - 资源共享的可控性
- 技术
  - 面向共享资源的可控管理
- 应用
  - 质量保证

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

高可动的存储系统

- 问题
  - 数据是水，存储是桶
  - 水在桶间一定可流动
    - 可流动的程度决定着系统的灵活度
      - ✓ 资源共享的低廉性
- 挑战
  - 动态状态信息较小
  - 静态状态信息很大
- 技术
  - 在线数据“瞬间”迁移
    - 在线数据，在线应用，数据迁移
    - 瞬间I/O处理栈的迁移（动态状态）
    - 渐进式有效数据迁移（静态状态）
- 应用
  - 负载均衡，系统伸缩，系统节能，系统管理

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

可伸缩的存储系统

- 问题
  - 需求的动态性
    - 容量需求
    - 性能需求
- 技术
  - 文件系统按照“卷粒度”的映射/去映射支持
  - 文件系统按照“卷粒度”的数据迁移支持
  - 存储设备按照“卷粒度”的资源管理

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

可管理的存储系统

- 问题
  - 存储管理与应用的静态式思维
  - 存储规划
  - 需求的动态性
    - 容量需求
    - 性能需求
- 技术
  - 存储资源管理的粒度化
  - 存储资源管理的动态化

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

可集成的存储系统

- 问题
  - 存储系统是一个/一堆黑箱子
    - Box Mover
  - 存储系统的封闭式思维
- 技术
  - 存储系统自身的可集成使能技术
    - 垂直流水线

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

局域数据管理

- 数据属性与分类的关联关系
  - 数据属性是文件内在的
    - 文件系统的原生支持
      - ✓ <Key，Value>
  - 数据分类是面向应用的
- 数据多视图支持
  - 面向数据管理的虚拟文件系统
    - 可跨越多底层文件系统
    - 支持原生属性排序索引
    - 支持应用定义数据分类
    - 支持各种数据分类视图
    - 支持各种分类策略驱动

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

广域存储系统

- 广域范围的NAS集群
  - 每个文件都拥有一个唯一的属主站点
    - 属主站点拥有该文件的整体数据
  - 两级虚拟文件系统
    - 第一级：多个站点文件存储系统的Union FS
    - 第二级：对于每个文件，提供多个站点文件缓存系统的Union FS
  - 数据服务调度策略
    - 根据文件数据访问和分布情况，决定数据的服务站点
    - 根据文件数据访问和分布情况，决定数据的缓存站点
- 站点文件存储系统
  - 属于本地站点的文件
- 站点文件缓存系统
  - 属于其它站点文件或片断的本地缓存
    - 与FS Cache的关系？
- 站点间数据传输系统
  - 采用P2P传输技术，提供多点并发的数据传输

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

广域存储系统

- 统一文件视图
  - 所有外部访问和内部访问都是一致视图
- 按需数据分布
  - 根据访问需求所有文件或片断可被任何站点缓存
- 按效数据传输
  - 根据访问需求所有文件或片断可被任何数据拥有站点提供

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

广域数据管理

- 面向广域数据管理的虚拟文件系统
  - 按照策略要求将数据在中心间位置布局
  - 按照策略要求将数据在中心间副本布局
  - 按照策略要求将数据在中心间复制跟踪

版权所有@2010-2011天津中科蓝鲸信息技术有限公司

BlueWhale蓝鲸

愿我们携手共创奇迹！

天津中科蓝鲸信息技术有限公司
TianJin ZhongKe BlueWhale Information Technology Co.,LTD

地址：天津市华苑产业区海泰发展6道6号海泰绿色产业基地F座5门201

电话：022-85689592/93（天津）
010-59851155 （北京）
0571-88843820 （浙江）
027-59847184 （湖北）

网址：www.bwstor.com.cn