

第一章作业

1.

- 大数据管理的数据特征：可以用4V来刻画，就是大容量（*volume*），多类型（*variety*），快变化（*velocity*）和低质量（*veracity*）。
- 大数据管理的系统特征：
 - 从封闭世界到开放世界：从结构化数据到非结构化数据，要支持不同的数据模型；数据无模式或难以实现确定；数据操作是用户自定义的、更加复杂的操作；负载的不确定性更大
 - 从精确数据到不精确数据：大数据系统从单一的量的管理到量质融合管理的转变
 - 从数据管理到知识管理：数据很多是自描述的，因此如何获取并管理知识是关键问题之一
- 大数据管理的应用特征：
 - 以对象为中心进行数据组织，实现数据汇聚
 - 以机器学习为主要应用类型
 - 以社会发展第四范式：“数据驱动”，为解决问题新模式

2.

- 典型组成特征：
 - 多引擎系统架构
 - 支持不同数据模型的引擎将会在大数据管理系统中并存
 - 从逻辑上实现多引擎的融合
 - 对外提供统一的数据存取访问接口
 - 混合负载系统架构
 - 大数据上存在多种不同的计算模式
 - 需要支持混合负载，在同一份数据上支持不同的计算模式
 - 分布式系统架构
 - 大数据系统一定是分布式的系统，具有高可扩展性
 - 一般认为，大数据系统是架构在普通服务器上的，不是高端的服务器，因此，需要分布式容错机制、高可伸缩的体系架构
- 典型技术的分层架构：分为以下五层
 1. 查询引擎：如sql等
 2. 计算引擎：如hadoop等
 3. 存储管理：根据存储的形式不同，有图，文档，键值对，关系等存储方式

- 4. 分布式缓冲区管理
- 5. 分布式文件系统
 - 此外，还有原生应用、负责资源管理与调度的应用

3.

- (1) 两个方向：
 - 模型扩展：面向对象数据库与XML数据库
 - 专用系统：one-size-does-not-fit-all
- (2) GFS,MapReduce,Bigtable
- (3) 提出了SpannerDB,F1,其中F1具备无阻塞schema变更技术