

Data Centers

数据中心是数万甚至数十万规模的机器集群

设计时需要考虑的问题有

- 成本
- 能耗
- 可靠性（一般通过冗余实现）
- 网络 IO
- 交互性作业和批处理作业的平衡
- 并行性
- 可扩展性
-

Cloud Computing

云计算基于数据中心的支持

Multi-Tenant Isolation：将不同用户隔离，互不影响

包括对计算资源的分配，以及用户应用的隔离

Load Balancing：请求的模式通常不是平滑的（自相似性），需要 load balancing

解决方法：background job，或是把不同的用户集合起来

能耗也是一个需要考虑的问题，有 10% 以上的成本用于能源的消耗

硬件也会出错，如何在这种情况下提供高可用性也是需要考虑的问题

Programming Models

map reduce

spark

以及分布式文件存储系统，如谷歌的 GFS，hadoop 的 HDFS

通过提供冗余的数据来保证数据可靠

Consistency in Distributed Systems

CAP 定理

- Consistency: Every read receives the most recent write or an error
- Availability: Every request receives a (non-error) response – without guarantee that it contains the most recent write
- Partition tolerance: The system continues to operate despite an arbitrary number of messages being dropped (or delayed) by the network between nodes

分布式系统不可能满足超过两个上述要求

以及如何在分布式系统中节点结果不一致时达成共识