为什么需要并行计算?

1. 提高计算机性能的主要手段 1. 2. 3.
2.
3.
4.

弗林分类 1. 2. 3. 4.

并行类型分类 1. 2. 3. ( )

存储访问结构分类 1. 2. 3.

系统类型分类 1. 2. 3. 4.

并行程序设计模型/方法分类 1. 2. 3.

并行计算的主要技术问题

1.
2.
3.
4.
5.
6.
7.
8.
9.

Amdahl 定律揭示了

MapReduce 定义：MapReduce 是的 1.；2.，3.

对付大数据并行处理：

上升到抽象模型：

上升到构架：

MapReduce 提供的主要功能：、、、、

MapReduce 主要设计思想与特点：

1.
2.
3.
4.
5.
6. ( )

MapReduce 的工作过程

1.
2.
3.
4.
5.
6.
7.
8.

9. _____

MapReduce 失效处理与性能优化方面的主要措施

1. 失效处理 1. 2. 3.
2. 性能优化三点：1. 2. 3.

Google GFS 基本设计原则

1. _____
2. _____
3. _____

Google GFS Master 的主要作用：1. 2. 3.

BigTable 设计动机和目标：

1. _____
2. _____
3. _____

Combiner 设计目的与作用：_____

Partitioner 设计目的与作用：_____

InputFormat：_____

InputSplit：_____

RecordReader：_____

NameNode：_____

DataNode：_____

JobTracker：_____

TaskTracker：_____

HDFS 基本特征

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____

HDFS 基本文件访问过程

1. _____
2. _____
3. _____

HDFS 可靠性设计 1. 2. 3. 4. 5. 6. 7. 8.

HBase 设计目标：_____

HBase 功能特点：1. 2. 3. 4. 5. _____

Hive 组成模块：1. 2. 3. 4. 5.

什么是 Distributed Cache？_____

全局参数 Configuration 的目的？_____

什么时候实现 Writable 接口？_____

什么时候实现 WritableComparable 接口？_____

为什么有 Spark？

1. _____
2. _____
3. _____

Spark 的技术特点

1. _____
2. _____
3. _____

RDD：_____

Application：_____

Job：_____

Stage：_____

Task：_____

Spark 的执行过程

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____

构建 RDD 世系关系的优势：1._____ 2._____

RDD 包含_____、_____、_____、_____