01

第1章 大数据关键技术

授课教师：[李阳源]

15138617705@163.com

主页：https://sites.google.com/view/dalguri



目录

CONTENT

大数据时代

大数据概念

大数据的影响

大数据关键技术

大数据计算模式

代表性大数据技术



1.5 大数据计算模式

1.5 大数据计算模式



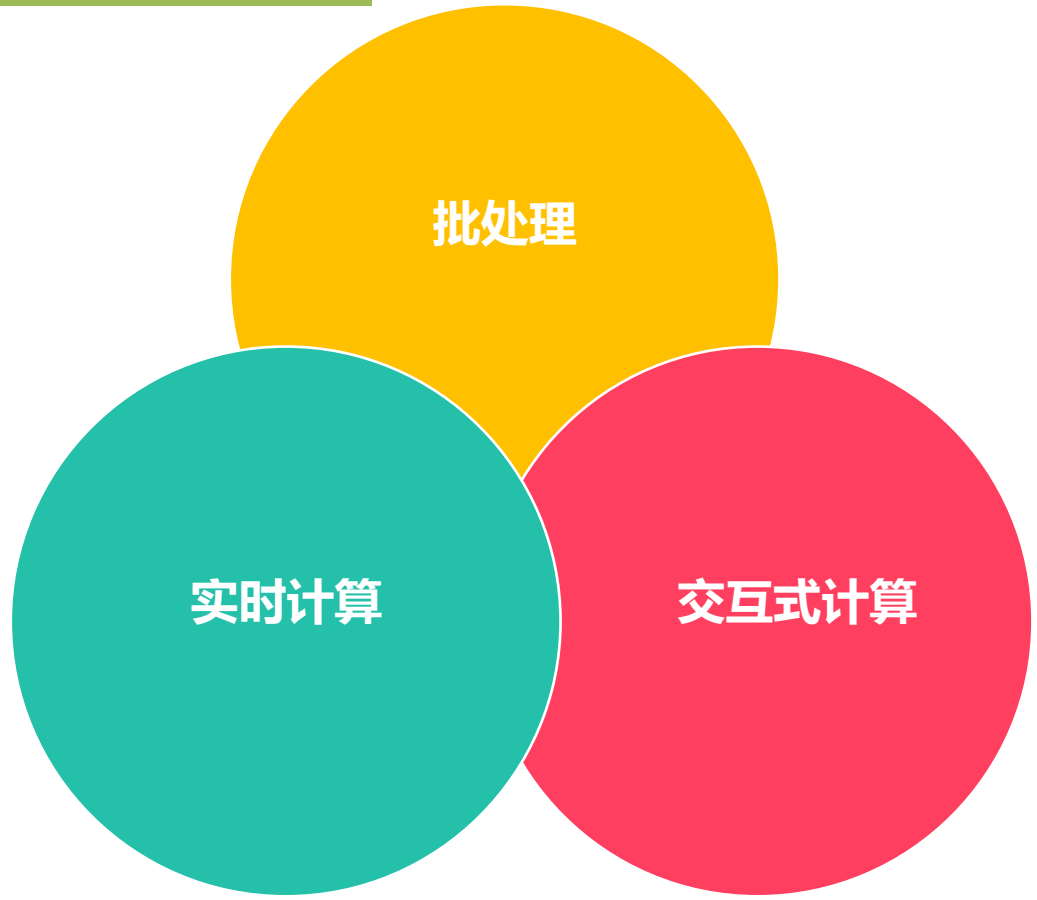
不同的计算模式需要使用不同的产品



正如“不同的锁需要使用不同的钥匙”企业中不同的应用场景属于不同的计算模式，需要使用不同的大数据技术

1.5 大数据计算模式

不同的计算模式需要使用不同的产品



1.5 大数据计算模式



典型的计算模式

01 批处理计算

02 流计算

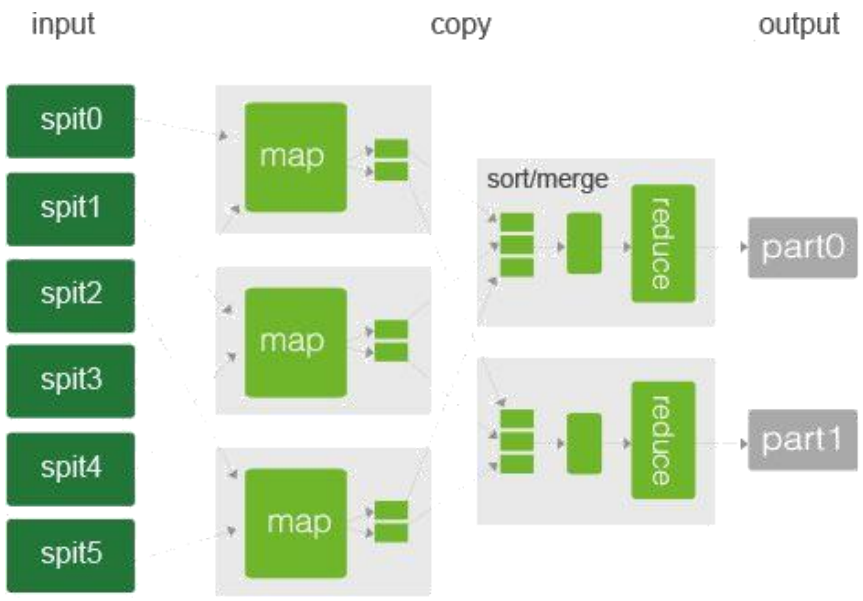
03 图计算

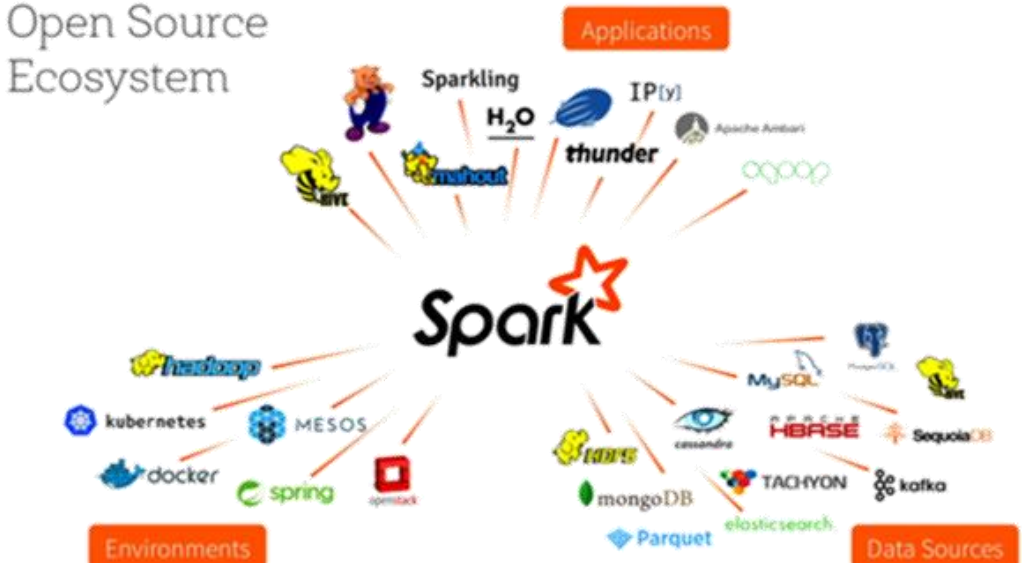
04 查询分析计算

1.5 大数据计算模式

批处理计算

MapReduce是批处理计算模式的典型代表

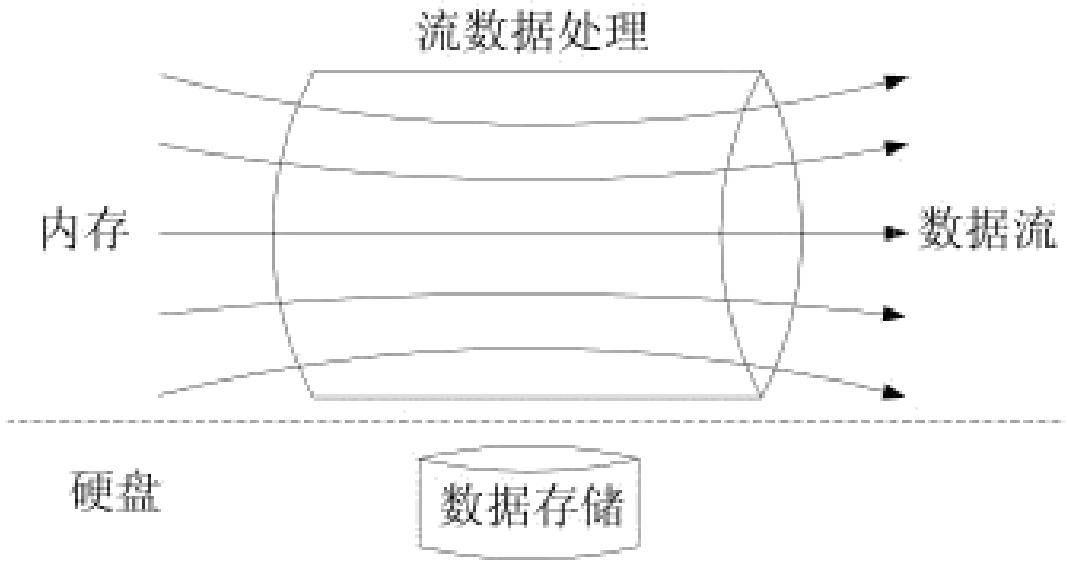




1.5 大数据计算模式

流计算

流数据需要实时处理，给出实时响应，否则分析结果就会失去商业价值

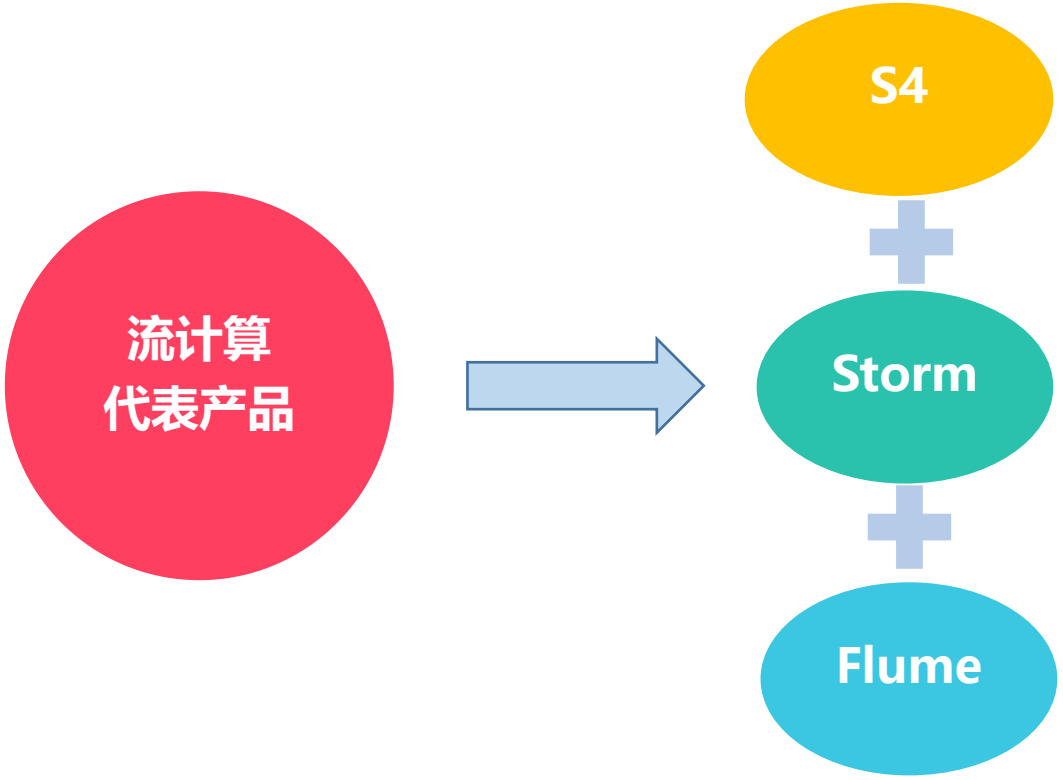


1.5 大数据计算模式



流计算

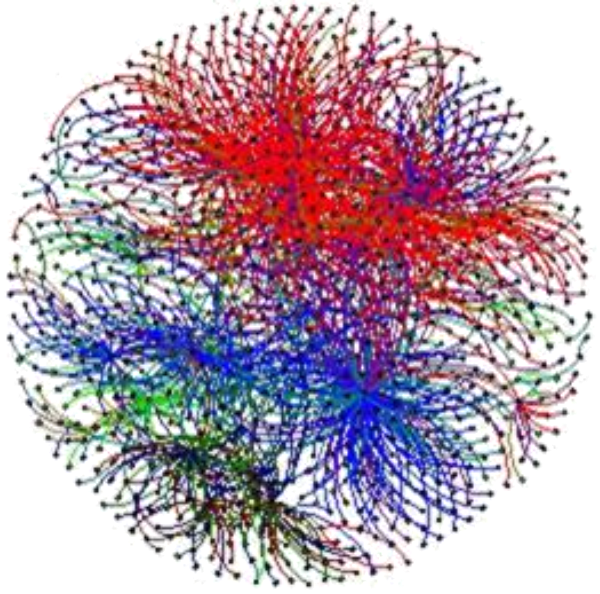
流数据需要实时处理，给出实时响应，否则分析结果就会失去商业价值

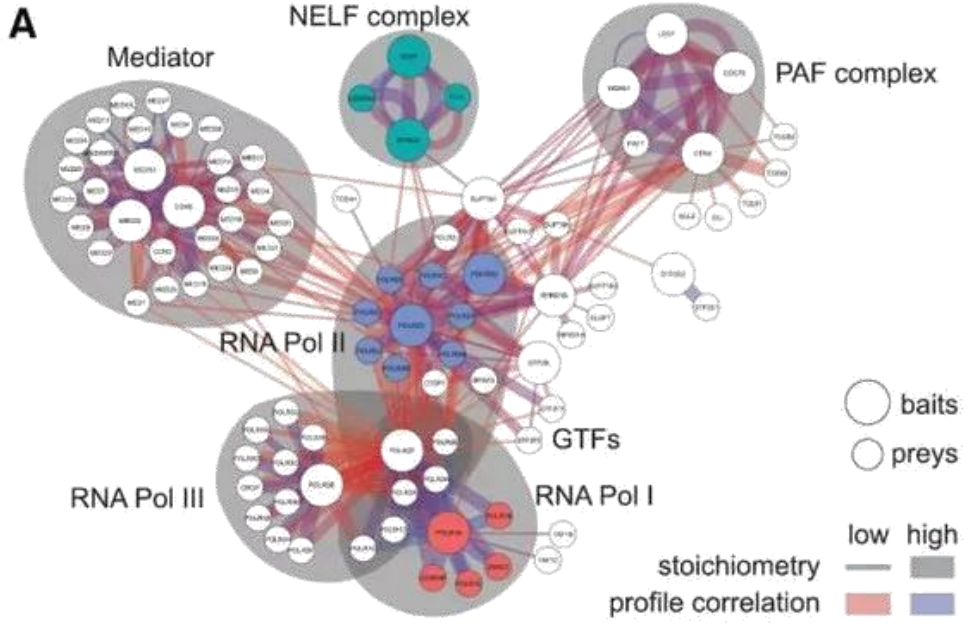


1.5 大数据计算模式

图计算

图计算代表软件：Google Pregel





1.5 大数据计算模式

图计算



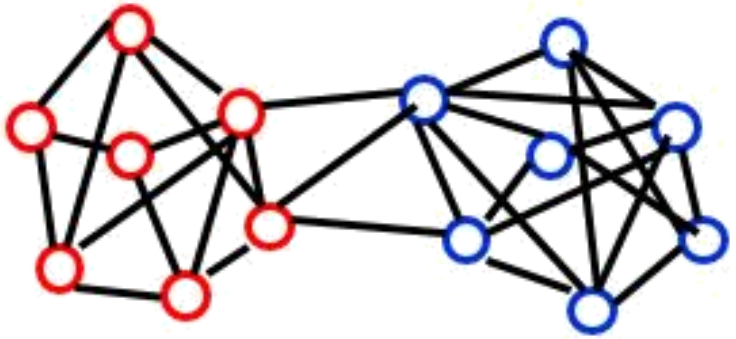
地理信息系统

1.5 大数据计算模式

图计算



社交网络数据
就是图结构数据



1.5 大数据计算模式

查询分析计算





1.5 大数据计算模式



查询分析计算

代表产品	Google Dremel	Hive	Cassandra

1.5 大数据计算模式



大数据计算模式	解决问题	代表产品
批处理计算	针对大规模数据的批量处理	MapReduce、Spark等
流计算	针对流数据的实时计算	Storm、S4、Flume、Streams、Puma、DStream、Super Mario、银河流数据处理平台等
图计算	针对大规模图结构数据的处理	Pregel、GraphX、Giraph、PowerGraph、Hama、GoldenOrb等
查询分析计算	大规模数据的存储管理和查询分析	Dremel、Hive、Cassandra、Impala等

1.5 大数据计算模式理解测验（选择题）



第1题

大数据计算模式的选择主要取决于什么？

A. 数据的存储方式

B. 数据的清洗手段

C. 应用场景与处理需求

D. 数据采集的频率

1.5 大数据计算模式理解测验（选择题）



第2题

以下哪一项不是常见的大数据计算模式？

A. 批处理计算

B. 流计算

C. 图像识别计算

D. 查询分析计算

1.5 大数据计算模式理解测验（选择题）



第3题

MapReduce 最适合用于哪种计算模式？

A. 图计算

B. 实时数据处理

C. 批处理计算

D. 查询交互式分析

1.5 大数据计算模式理解测验（选择题）



第4题
为什么流计算要求系统具备实时处理能力？
A. 因为数据是静态的
B. 因为计算任务很复杂
C. 因为商业价值依赖于快速响应
D. 因为只能使用单机计算

1.5 大数据计算模式理解测验（选择题）



第5题

以下哪一个不属于流计算的代表工具？

A. Storm

B. Flume

C. S4

D. Hive

1.5 大数据计算模式理解测验（选择题）



第6题
图计算主要用于哪类数据结构的处理？
A. 表格型数据
B. 时序型数据
C. 图结构数据（节点与边）
D. 文本日志

1.5 大数据计算模式理解测验（选择题）



第7题
下列哪一个场景最适合采用图计算？
A. 批量生成销售报表
B. 分析网页之间的链接结构
C. 实时监控传感器温度变化
D. 查询指定用户的购买历史

1.5 大数据计算模式理解测验（选择题）



第8题
查询分析计算的主要目标是？
A. 实现数据压缩
B. 提升存储安全性
C. 支持大规模数据的交互式查询
D. 自动识别图像内容

1.5 大数据计算模式理解测验（选择题）



第9题

以下哪一项不属于查询分析计算的代表系统？

A. Hive

B. Dremel

C. Cassandra

D. Pregel

1.5 大数据计算模式理解测验（选择题）



第10题
下列说法正确的是：
A. 流计算不能处理实时数据
B. 批处理计算适用于每秒产生大量数据的情况
C. 图计算无法用于社交网络分析
D. 查询分析计算强调快速响应和交互性



Thank you