

# 一年黑沟 成长之路

# SACC 第十届中国系统架构师大会







# 百度云Spark优化及大规模应用

张建伟

# 大数据平台架构

运营管理	- 数据分析与治理(Data Analytic and Government) -		
系统管理	开发、分析和挖掘 DataLab	报表可视化 DataViz	数据管理和服务 DataGov
服务管理	- 离线数据处理(Data Factory) -		- 在线数据服务(Data Warehouse) -
安全管理	<b>数据计算</b> 工作流调度 大数据计算 BMR(Hadoop/Spark) 流式计算 BSC(Spark Streaming)		分析数据库 Palo
资源管理			40 ± ₩.40 ±
监控报警	数据存储 文件存储 HDFS 对象存储 Bos		搜索数据库 Elasticsearch
滚动升级	- 数据采集传输(Data Collect and Transfer) -		
健康预检	结构和非结构数据采		式数据传输 Kafka

### 发展历程



# 大数据计算核心引擎 Spark



#### Spark 性能优化 - OAP (Optimized Analytics Package)

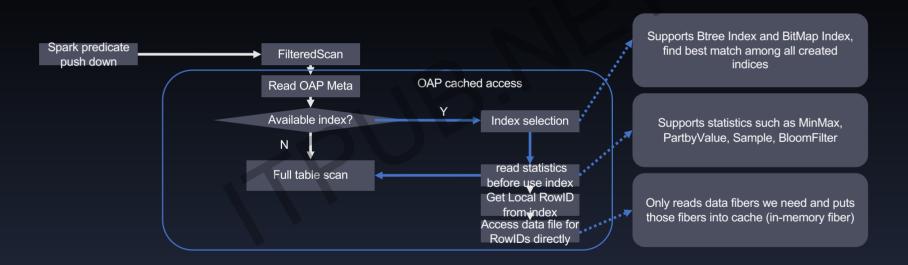


#### **OAP Files and Fibers**

Index meta Column (Fiber) #1 Column (Fiber) #2 statistics Column (Fiber) #N RowGroup #1 Index data structure RowGroup #2 (Index Fiber) Index meta statistics RowGroup #N OAP index Index data structure OAP data (Index Fiber) files One Index file for every data file OAP index files

OAP meta file

#### OAP Internals - index



Select xxx from xxx where age > 29 and department in (Baidu, Al-Lab)

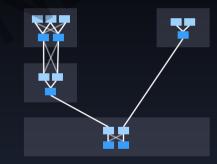
#### Spark 性能优化 — AE ( Adaptive Execution )

Reduce 数量该如何设置?

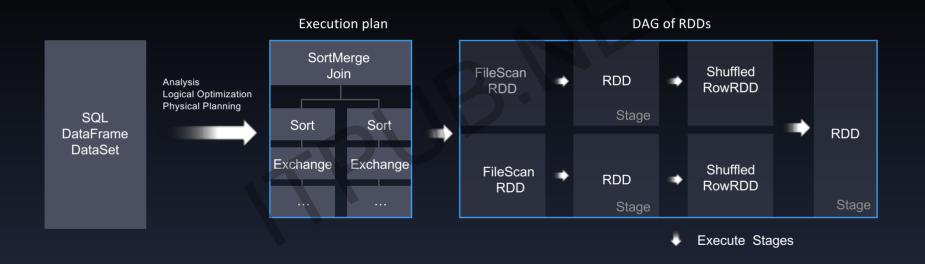
太小: Spill, OOM

太大:调度开销,更多IO,输出更多小文件

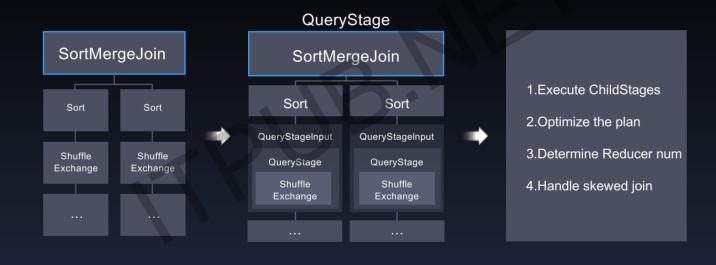
同样的 Reduce 个数不适合每个 Stage



#### Spark 性能优化 – AE – Normal Execution Mode



### Spark 性能优化 – AE - Adaptive Execution Mode



#### Spark 性能优化 – AE – auto setting the reduce number

- Target size per reducer = 64 MB.
- Min-Max shuffle partition number = 1 to 5



### Spark 性能优化 – AE – Optimize Join Strategy at Runtime

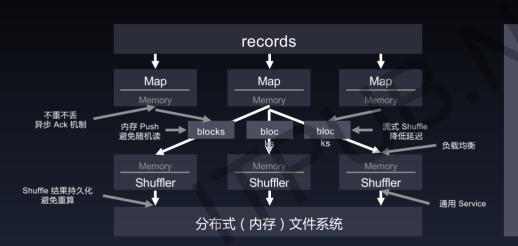


#### Spark功能优化 – CP (continuous processing) – 与社区联合开发

(DataFrame/Datasets/SQL) input Structured Streaming Incremental Planner streams Structured Streaming processing modes: Micro-batch Processing Output Micro-batch and Continuous Processing (default) streams input Continuous Processing (low-latency,optional) tables . Offset log state

Streaming Query

#### Spark 稳定性优化 – 独立 Shuffle 服务



#### 收益

- 1. 流式shuffle:减少旧 shuffle map merge、reduce pull 时间消耗
- 2. Shuffler 内存聚合:聚合度高,减少IO
- 3. 中间数据持久化:避免重算(对大 dag 作业尤为重要)

# 基于 Spark 的可信数据计算平台



在安全可信的环境下,探索数据价值

可信计算环境 (AC)

Sgx 加固的 spar 计算标规 数仓之间的 多方计算(A∩B)

Sql 版本的联合 多方计算服务 隐私数据查询 分析(A->B)

差分隐私解决数据 查询隐私泄漏

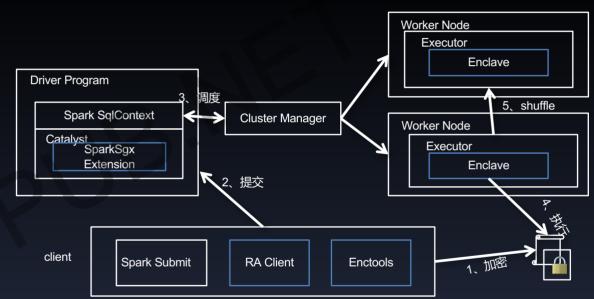
### SparkSgx整体架构

#### • 集群

- SparkSgx Extension
  Catalyst的扩展转换执行计划 cluster
- EnclaveSpark算子实现

#### • 客户端

- RA Client 远程认证客户端
- Enctools加密工具



# THANK YOU