iiHadoop, 101-summary

Anoop

April 03, 2019

an extension of Hadoop framework
it improves upon the i^2 MapReduce paper,
How? = Dynamic Scheduling of tasks and async ability

iterative computations on changing data
only recompute small incremental data

D + delta(D) as input to MapReduce in the incremental run
Reuse of intermediate results

Similarity:

incremental computations on the small fraction of data that is affected by changes

i^2 MapReduce

i2MapReduce uses a static scheduling policy
requires equal number of map and reduce tasks to allow asynchronous execution

iiHadoop

i2MapReduce uses a dynamic scheduling policy
iterations are asynchronously executed

scheduler creates the map and reduce tasks that perform
incremental computations on these changing data
Dynamic => kill them when they are completed

PageRank