

# DistCoz: Tell Me What to Optimize in My Distributed Application

Steffen Maass\*, Mohan Kumar\*, Prof. Taesoo Kim

Georgia Institute of Technology

\* Student

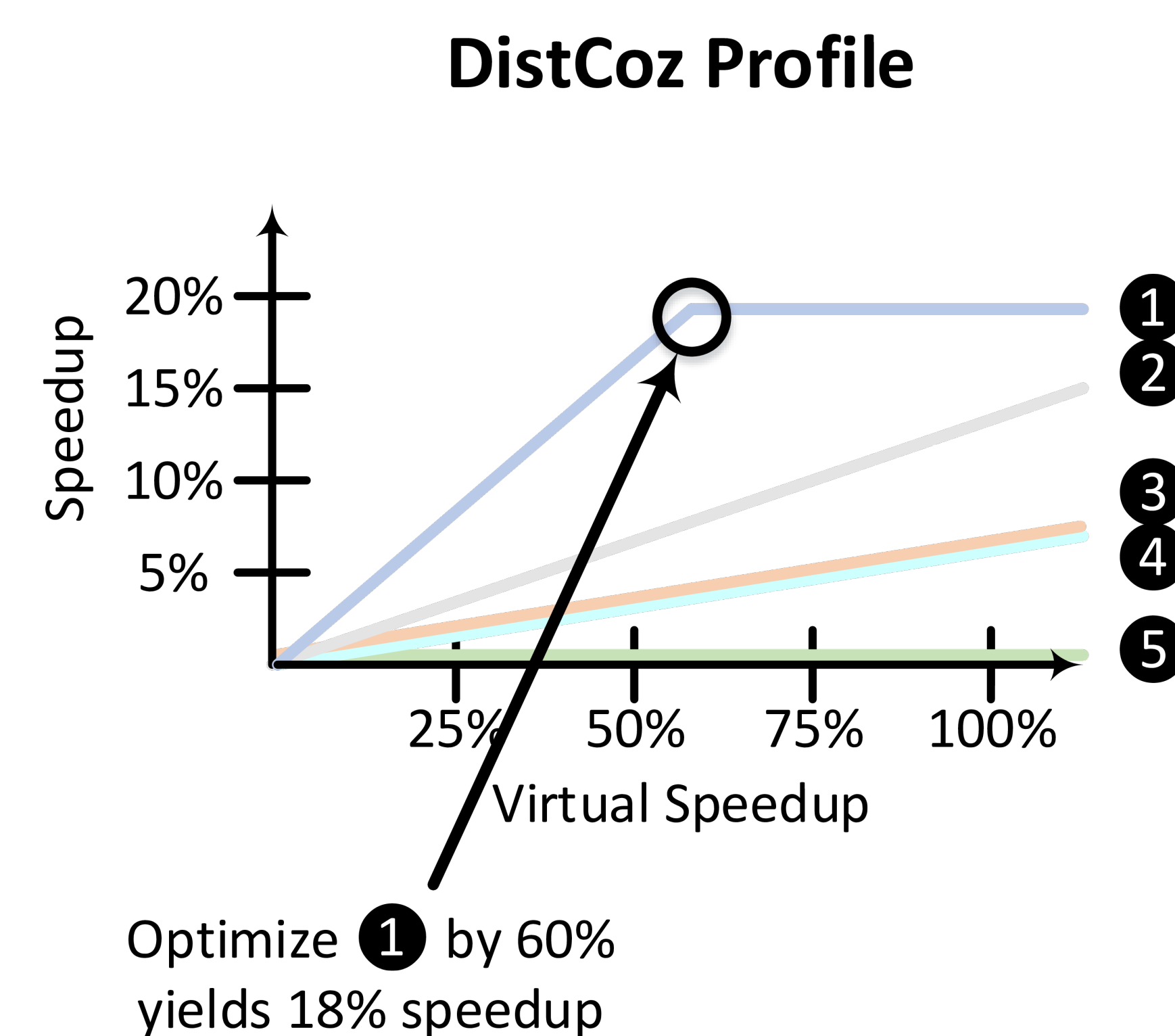
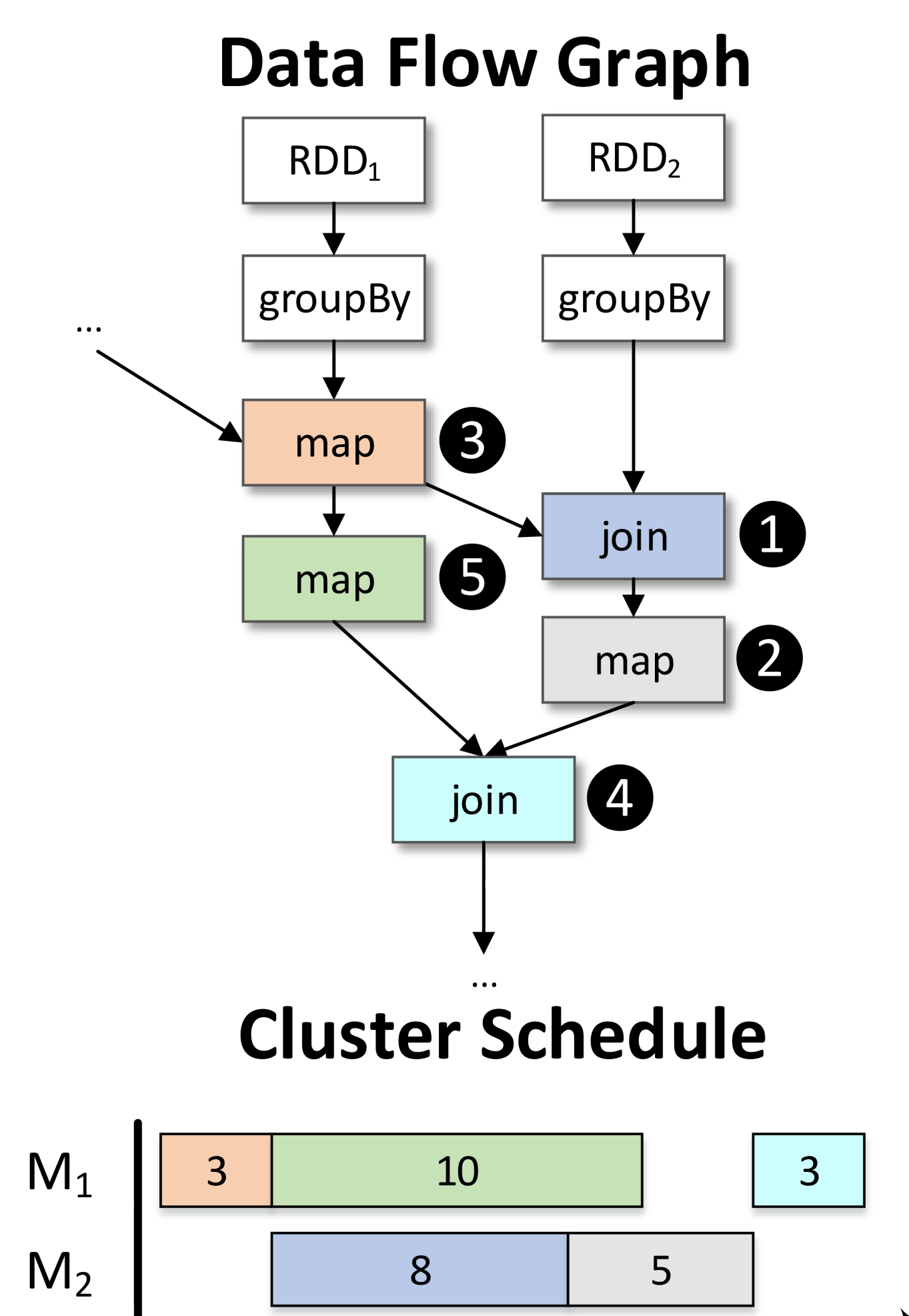
- Optimizing a distributed application:
  - How to find **optimization candidates**?
    - ⇒ Current techniques:
      - CPU-intensive subtasks (e.g. found via *perf*)
      - Critical path analysis
  - How to evaluate **potential improvement**?
    - ⇒ Implement, test, repeat
    - ⇒ Wastes developers' time!

- Use *virtual speedup*
  - ⇒ Simulate speedup by *slowing down* other components
- Extend to distributed setting
  - ⇒ Apply *virtual speedup* to RPCs

- Output: speedup vs. virtual speedup
- Allows direct correlation between **local speedup** and **global speedup**

**Code**

```
AlsApplication.scala:  
val als = new ALS()  
  .setMaxIter(50)  
  .setRegParam(0.01)  
  .setUserCol("userId")  
  .setItemCol("movieId")  
  .setRatingCol("rating")  
val model = als.train(trainingData)  
...  
ALS.scala:  
...  
val RDD1 = loadFromHdfs(matrixA)  
val RDD2 = loadFromHdfs(matrixB)  
val input1 = RDD1.groupBy(..).map(..) ③  
val input2 = RDD2.groupBy(..)  
  
val output1 = input1.map(..) ⑤  
val intermediate2 = input2.join(input1) ②  
val output2 = intermediate2.map(..) ①  
  
val result = output1.join(output2) ④  
...
```



- Reduce search space?
- Run on every request?
- Apply to investigation of latency issues?