• 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**

```scala
// AhsApplication.scala:
val als = new ALS()
  .setMaxIter(50)
  .setRegParam(0.01)
  .setUserCol("user_id")
  .setItemCol("movie_id")
  .setRatingCol("rating")
val model = als.train(trainingData)
```

**Data Flow Graph**

**Cluster Schedule**

**DistCoz Profile**

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