DryadLINQ

Presented By: Mert Coskun

Written by: Yuan Yu, Micheal Isard, et al.
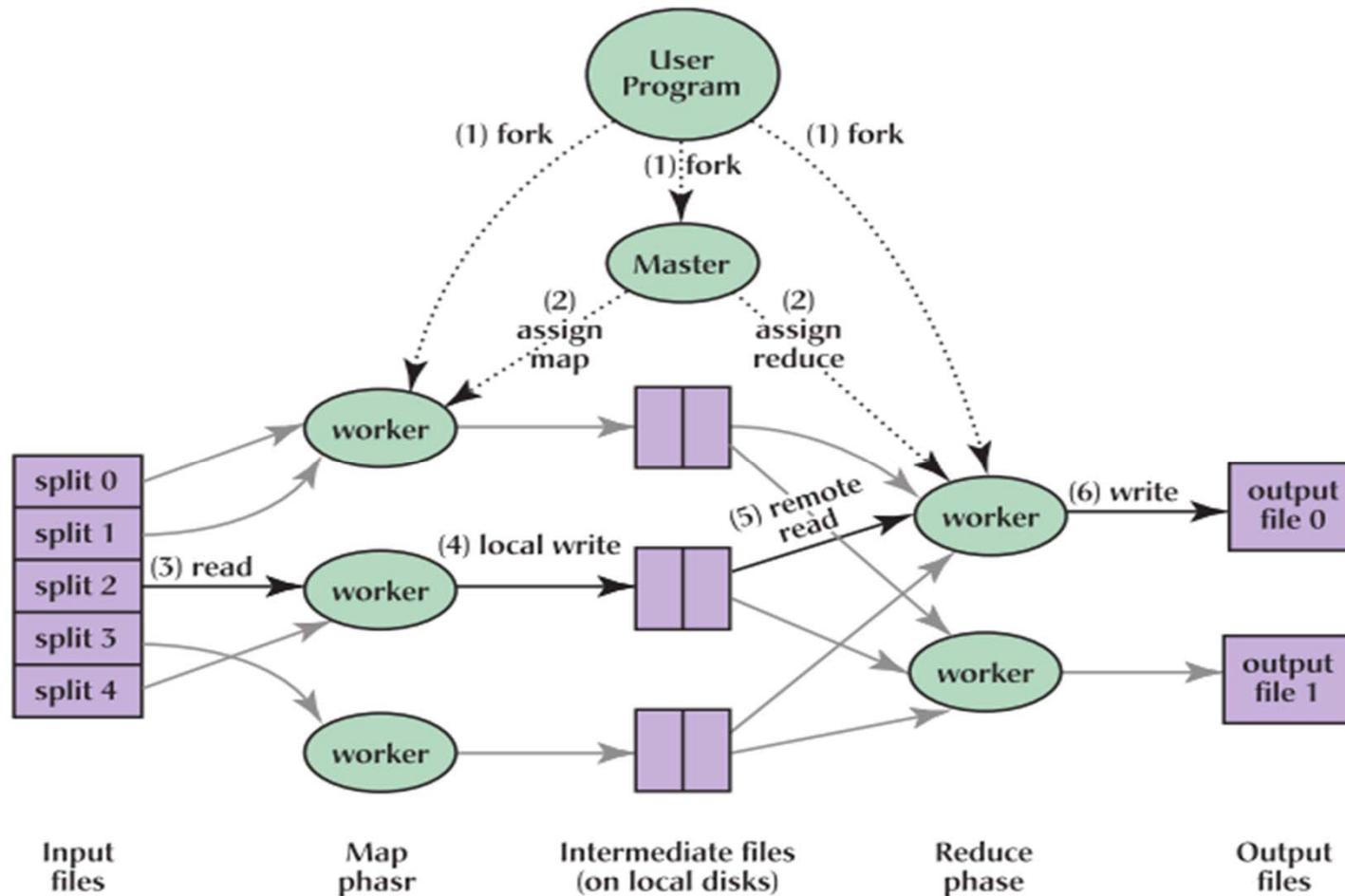
Problem

- Data Sizes vs. CPU performance

- Decreasing Acceptable Execution Time

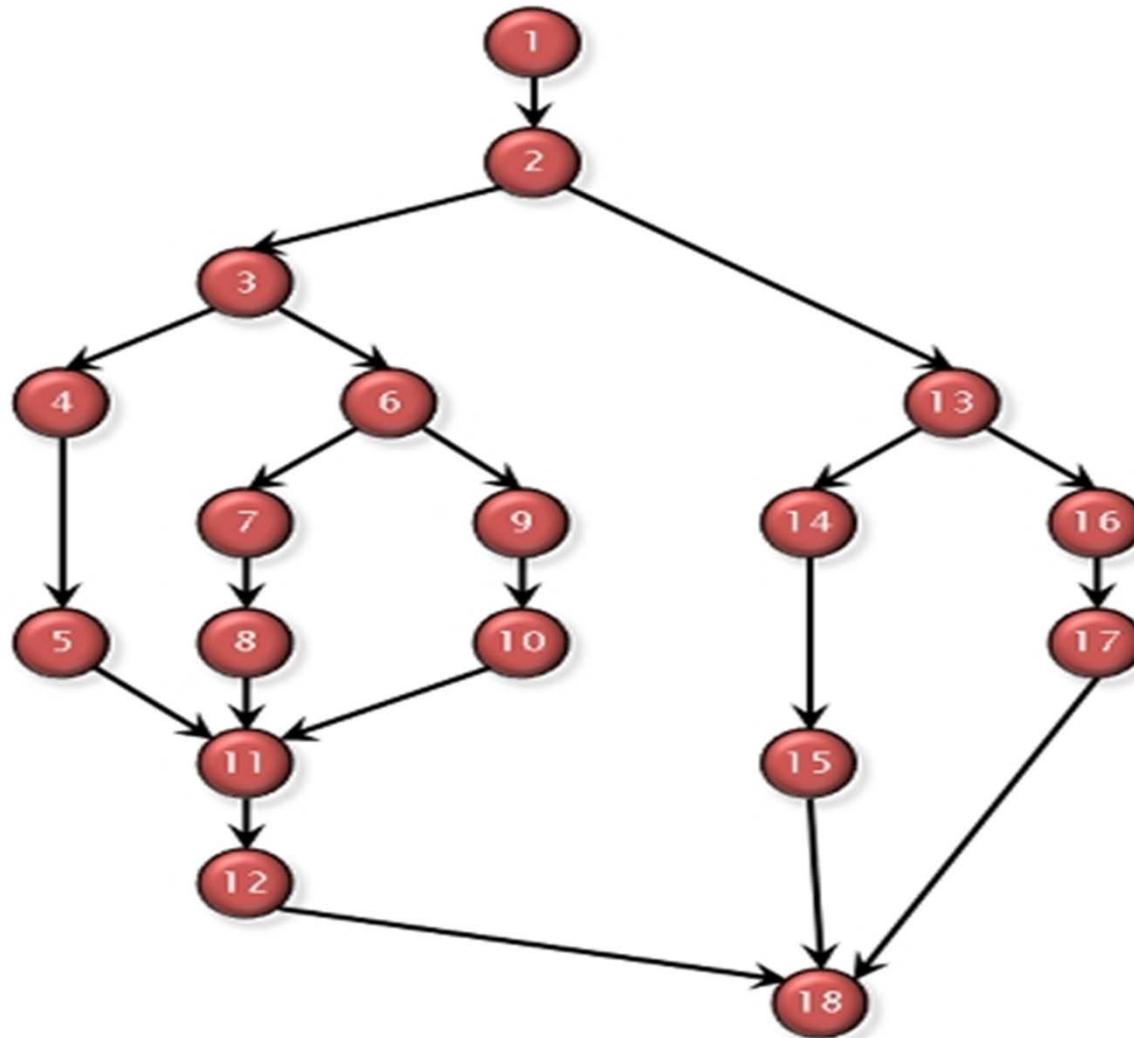- **How to program for clusters?**

- Another Motivation?

Simple Survey

	Google	Yahoo / Facebook	Microsoft
Language	Sawzall (2005)	Pig (2006) Hive (2008)	**DryadLinq (2008)**
Execution	MapReduce (2004)	Hadoop	Dryad (2007)
Storage	BigTable (2004) GFS (2003)	Hbase HDFS	Cosmos Azure SQLServer

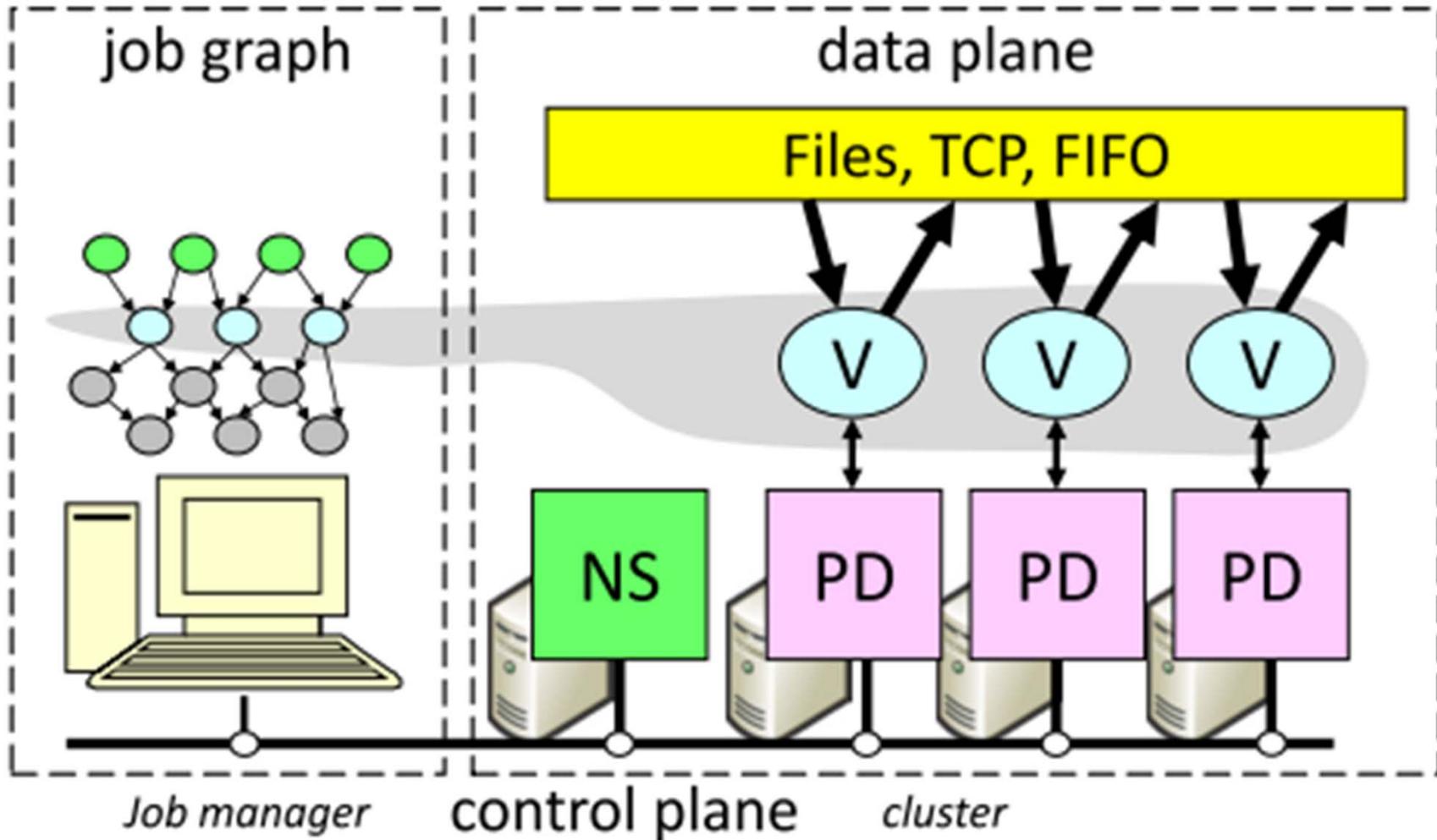# MapReduce

Dryad

DryadLinq

- Aim: Simplifying language

- Focus: Hybrid

- Solution: Dryad + Linq

- Novelty: General purpose language into distributed program

- In a Sentence: PLinq extended to D.S.
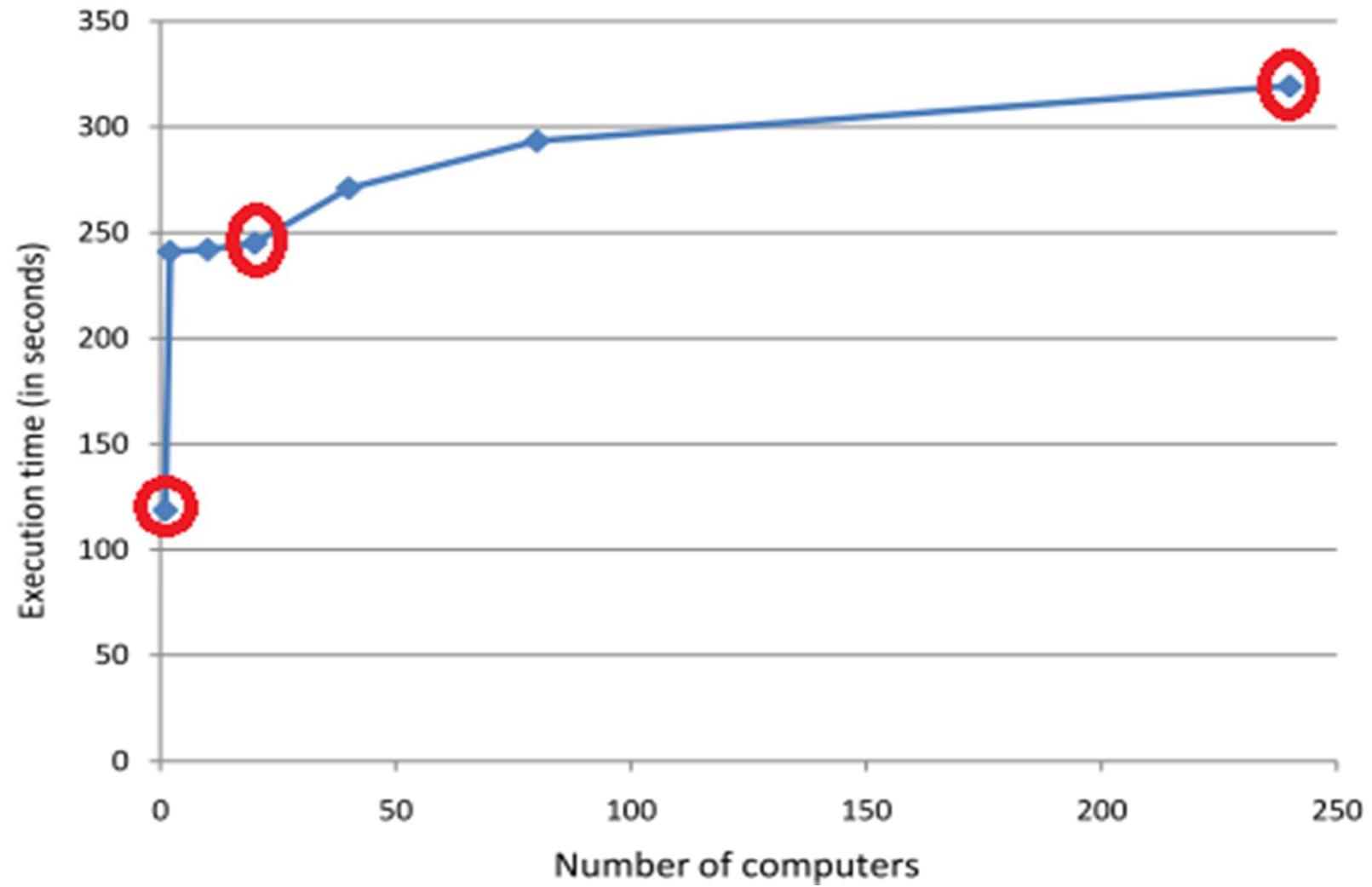
Under the Hood

Optimizations

- Pipelining
- Redundancy elimination
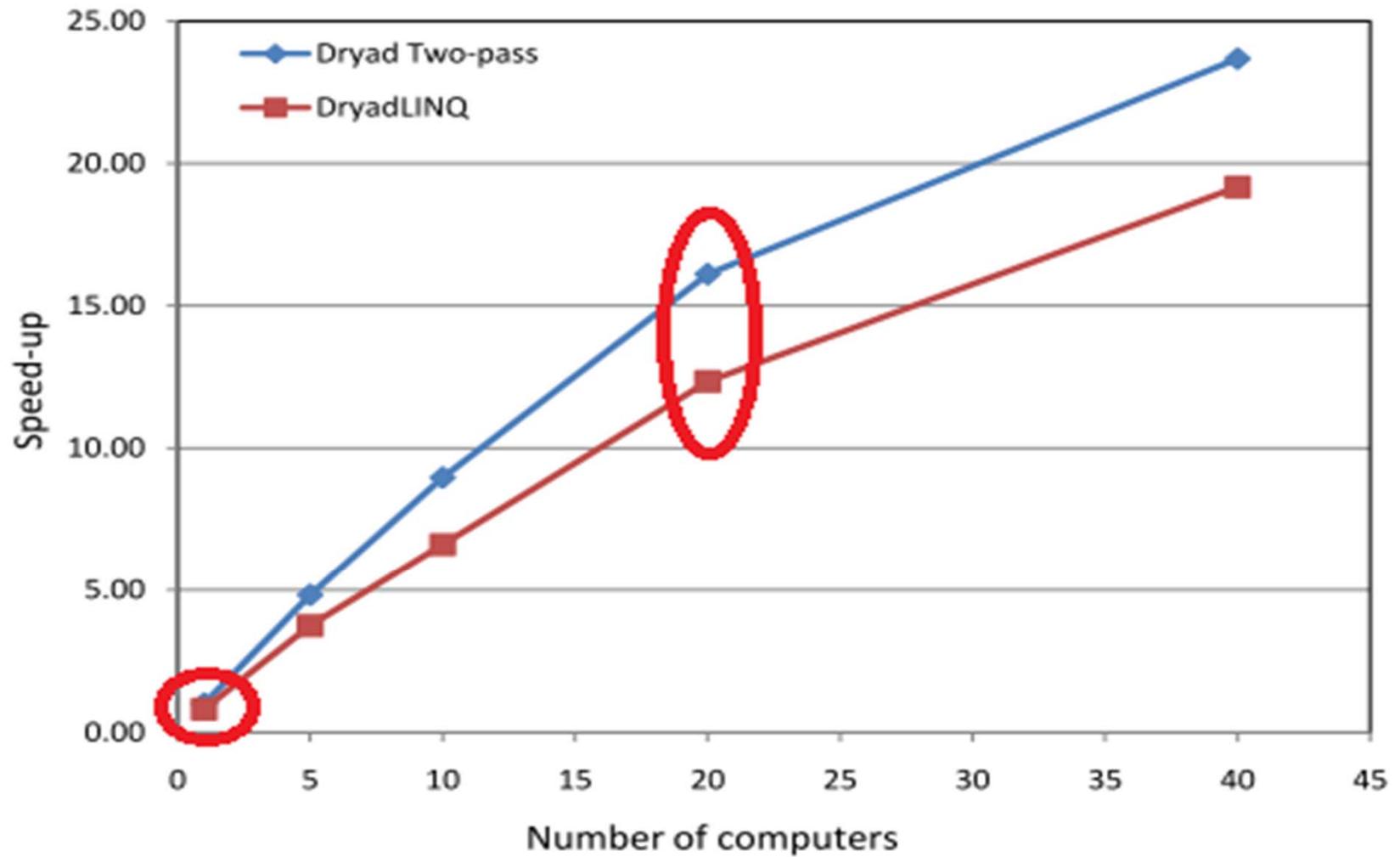- Eager Aggregation
- I/O reduction (vague)

- Dynamic Optimization

- OrderBy

Terasort

Scalability

The Good

- No new domain specific language

- Automatic Optimisations

- Plinq

- Easy Debugging!!!

The Questionable

- Good for small clusters?

- Apply?

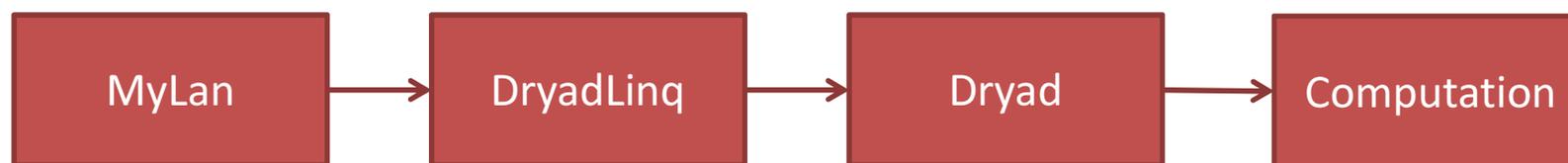- Success in Algorithms with Random Access

The Bad

- Database Look Ups

- .Net dependency

- Writable with Linq? Apply?

Why I chose MapReduce

- Better Tested

- Multiple Actions in Single Step (Average)
  - Apply ?

- Testing (Cluster)

Why I chose MapReduce

Thank You!

- Questions?