



# APACHE TEZ

Directed Acyclic Graph Framework

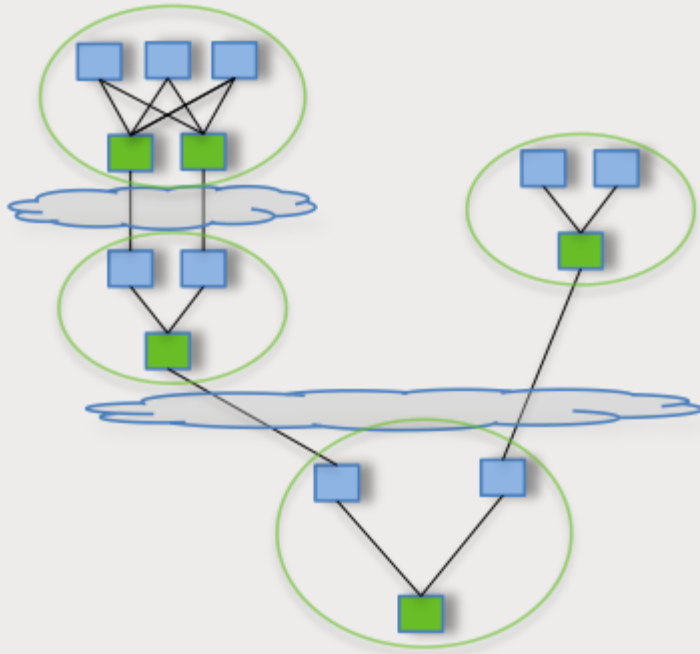


# What is Tez?

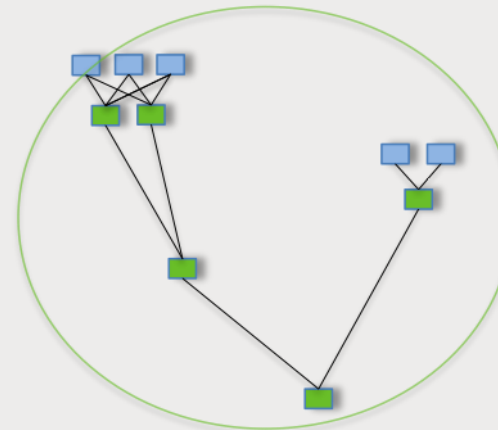


- Another bit of infrastructure you can just use
  - *Makes your Hive, Pig, or MapReduce jobs faster!*
  - *It's an application framework clients can code against as a replacement for MapReduce*
- Constructs Directed Acyclic Graphs (DAGs) for more efficient processing of distributed jobs
  - *Relies on a more holistic view of your job; eliminates unnecessary steps and dependencies.*
- Optimizes physical data flow and resource usage

# Directed Acyclic Graphs



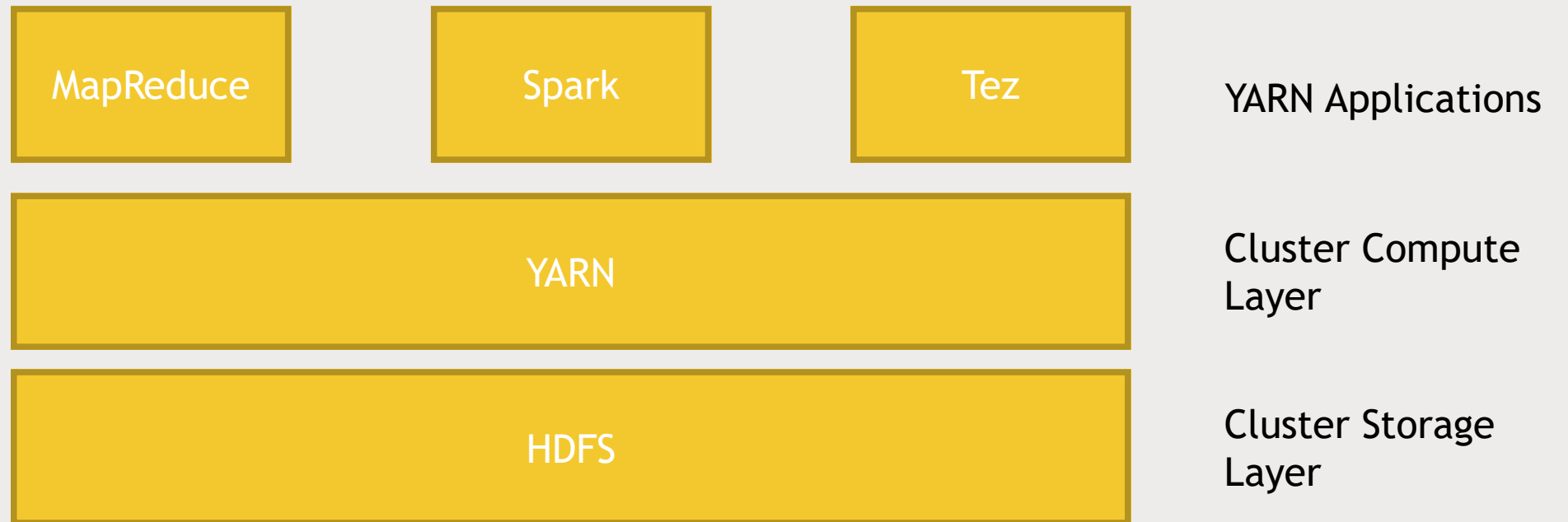
Pig/Hive - MR



Pig/Hive - Tez

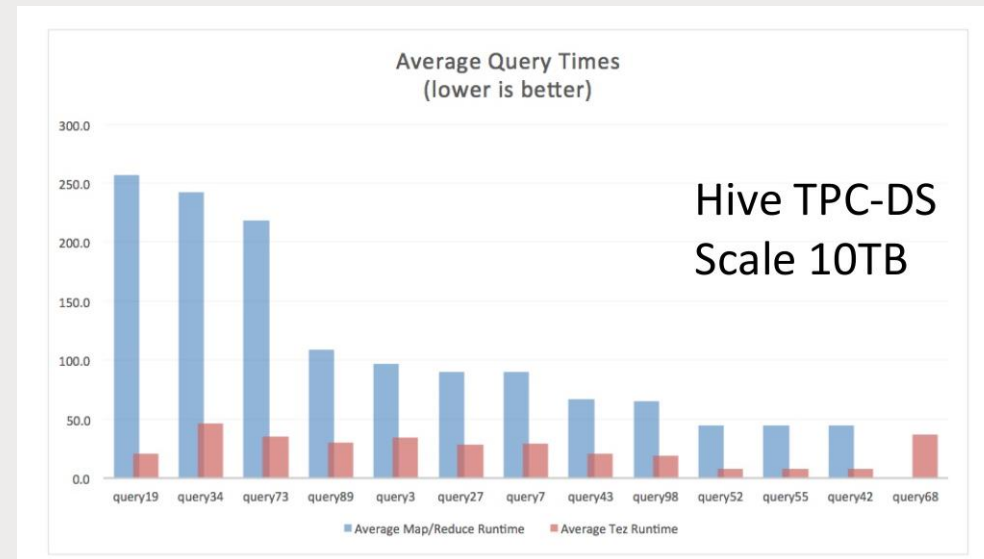
optimize steps,  
run some of them together/at once

# Where Tez sits



# Just tell Hive or Pig to use it.

- It probably does by default anyhow.
- It really is a lot faster!



# Let's try it out

- Compare performance of a Hive query using Tez vs. MapReduce

