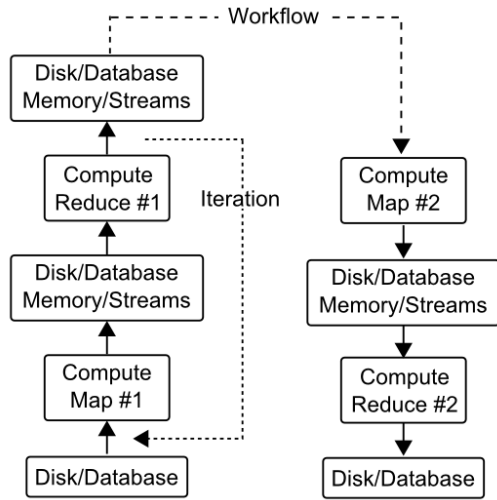
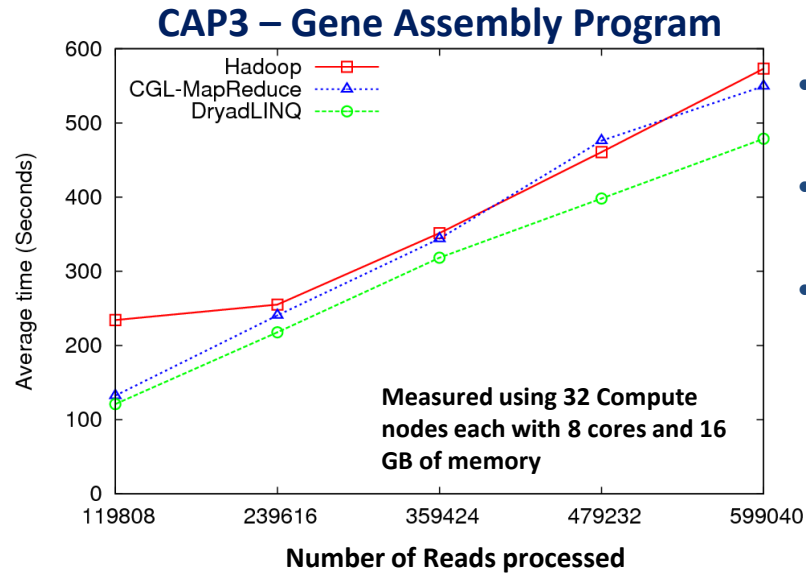


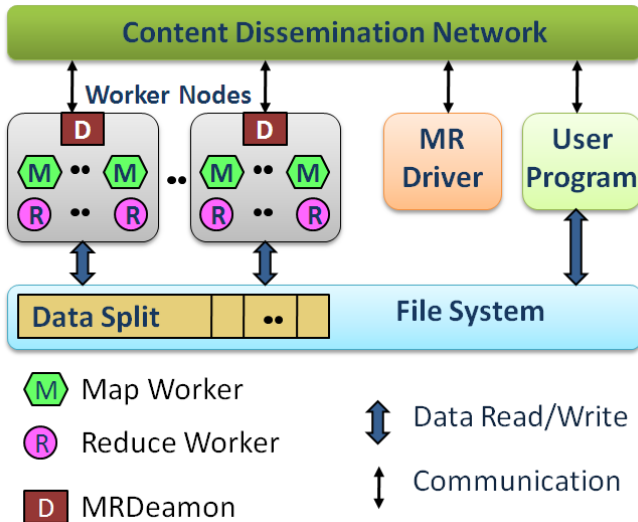
# Architecture and Performance of Runtime Environments for Data Intensive Scalable Computing



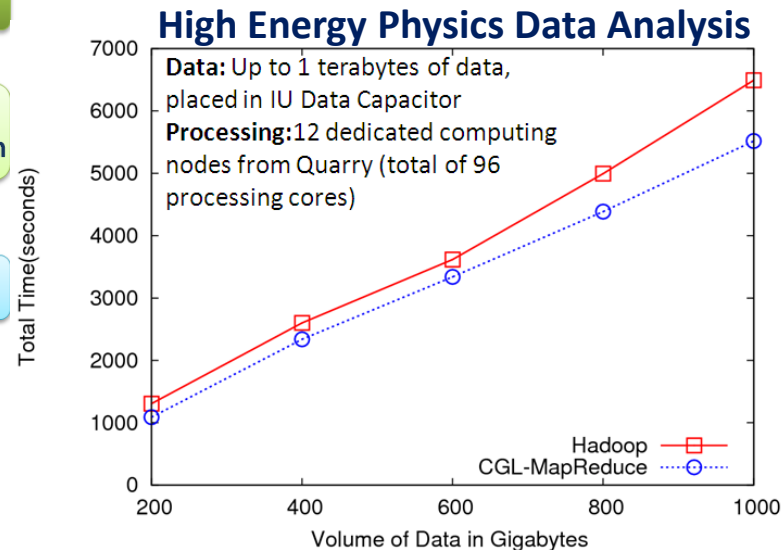
**Data/compute intensive applications implemented as MapReduce “filters”**



- Compute intensive application
- Embarrassingly parallel operation
- All runtimes performs equally well

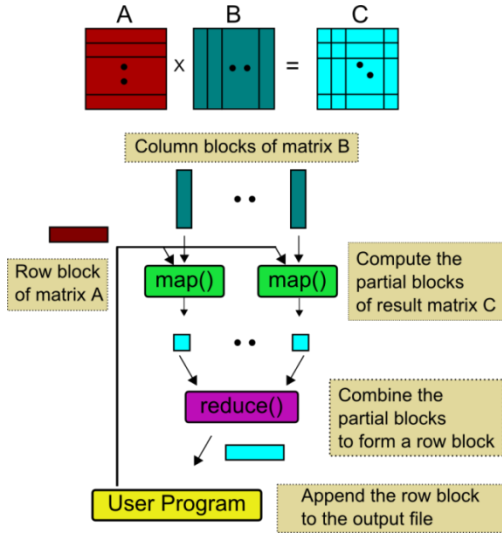


**Architecture of CGL-MapReduce**

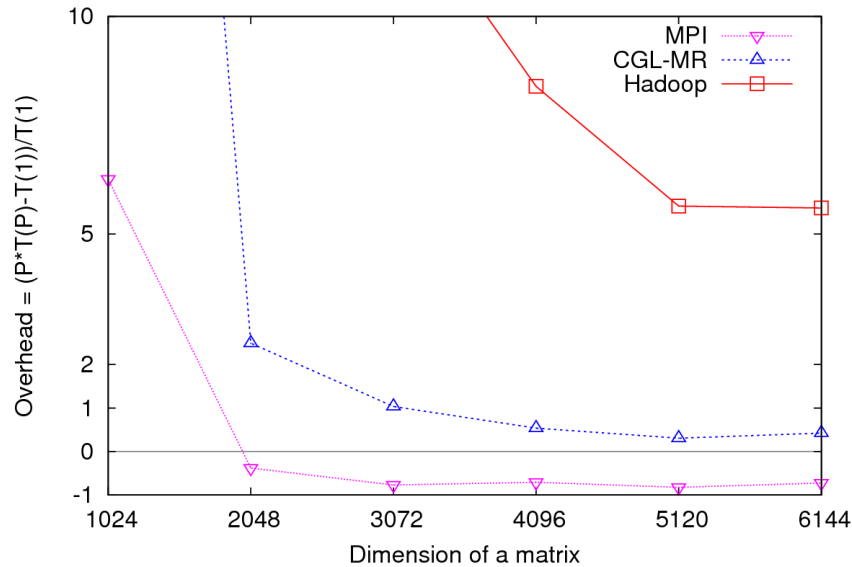


- Data intensive application
- MapReduce style parallel operation
- Both runtimes perform comparably well

# Iterative MapReduce- Kmeans Clustering and Matrix Multiplication

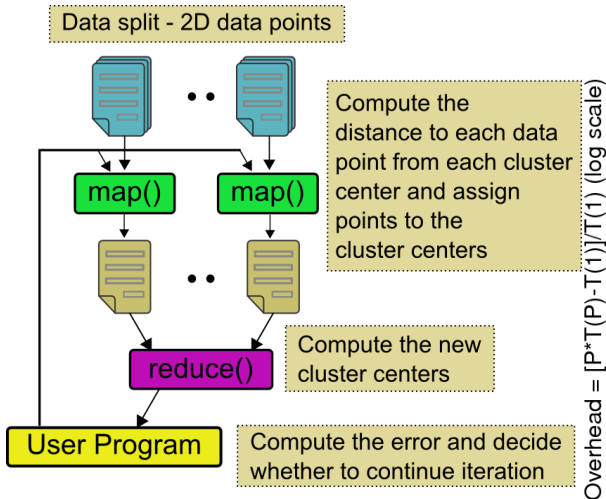


Overhead of parallel runtimes – Matrix Multiplication

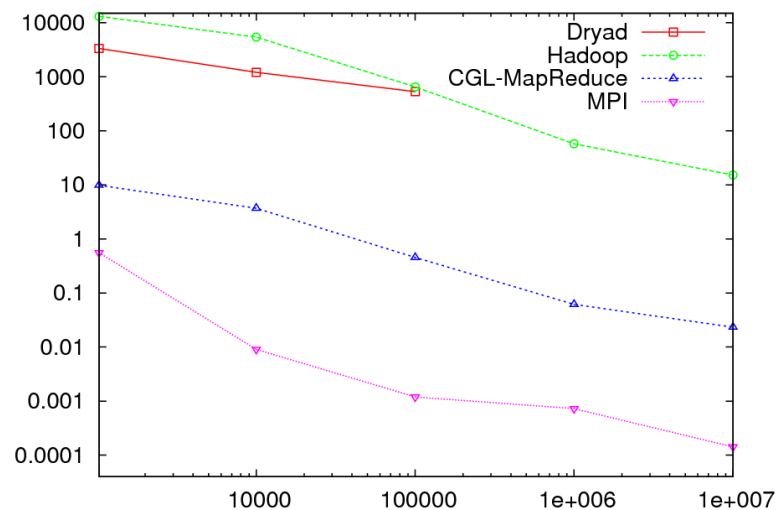


- Compute intensive application  $O(n^3)$
- Higher data transfer requirements  $O(n^2)$
- CGL-MapReduce shows minimal overheads next to MPI

## Iterative MapReduce algorithm for Matrix Multiplication



Overhead of parallel runtimes – Kmeans Clustering



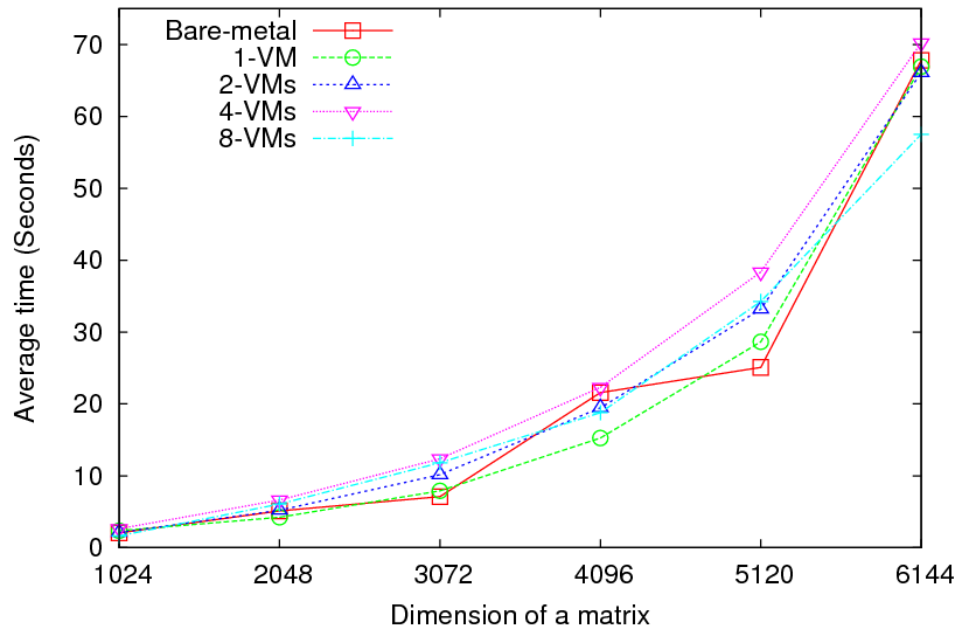
- $O(n)$  calculations in each iteration
- Small data transfer requirements  $O(1)$
- With large data sets, CGL-MapReduce shows negligible overheads
- Extremely higher overheads in Hadoop and Dryad

## Kmeans Clustering implemented as an iterative MapReduce application

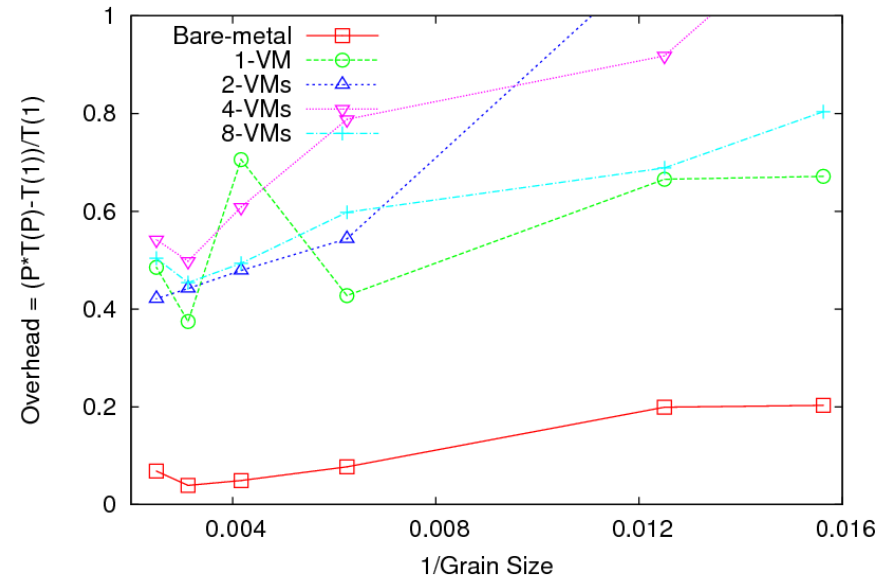
# High Performance Parallel Computing on Cloud

- Performance of MPI on virtualized resources
  - Evaluated using a dedicated private cloud infrastructure
  - Exactly the same hardware and software configurations in bare-metal and virtual nodes
  - Applications with different communication: computation ratios
  - Different virtual machine (VM) allocation strategies {1-VM per node to 8-VMs per node}

Performance of Matrix multiplication under different VM configurations



Overhead under different VM configurations for Concurrent Wave Equation Solver



- **$O(n^2)$  communication** ( $n$  = dimension of a matrix)
- **More susceptible to bandwidth than latency**
- **Minimal overheads under virtualized resources**

- **$O(1)$  communication** (Smaller messages)
- **More susceptible to latency**
- **Higher overheads under virtualized resources**