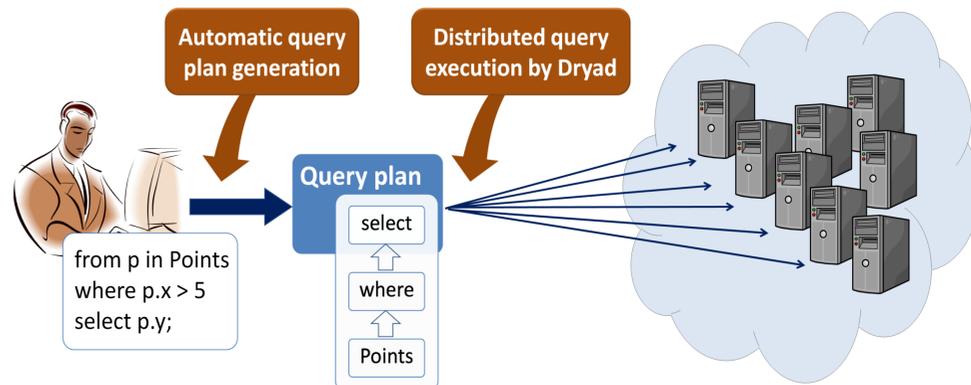
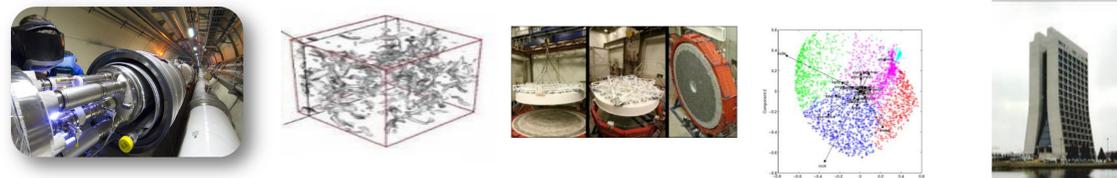


Computation and data intensive applications are increasingly prevalent, especially in research, as data acquisition and storage costs continue to drop towards zero.

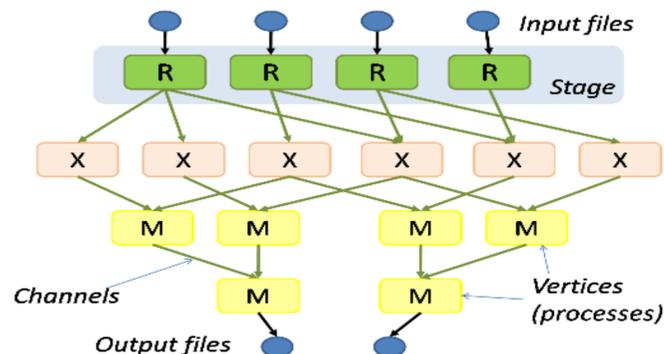
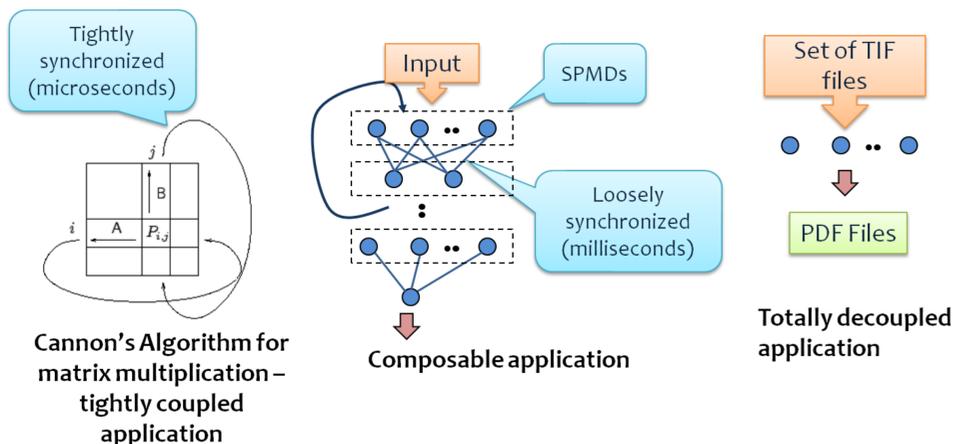
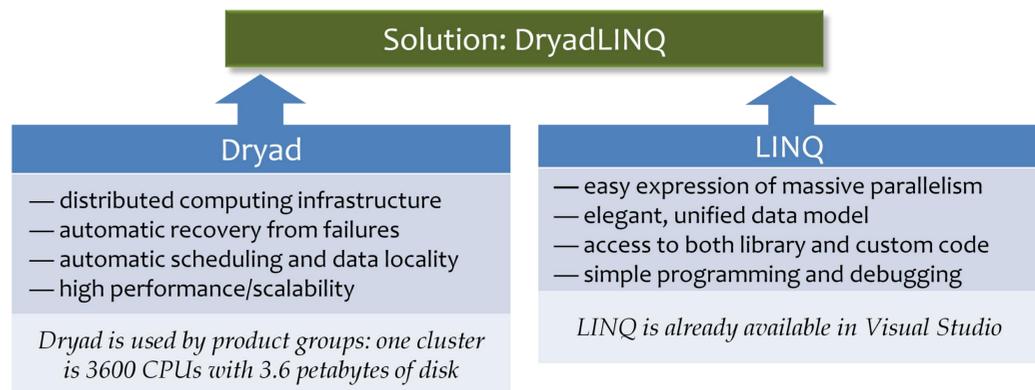


DryadLINQ: Making Large-Scale Distributed Computing Simple
Yuan Yu, Dennis Fetterly, Úlfar Erlingsson, Jon Currey, Michael Isard — MSR Silicon Valley

Composable Applications

- **How to support these large scale applications?**
 - Efficient parallel/concurrent algorithms and implementation techniques
- **Key observations**
 - Most of these applications are:
 - A Single Program Multiple Data (SPMD) program or a collection of SPMDs
 - Exhibits the **composable** property
 - Processing can be split into small sub computations
 - The partial-results of these computations are merged after some post-processing
 - Loosely synchronized (Can withstand communication latencies typically experienced over wide area networks)
 - Distinct from the closely coupled parallel applications and totally decoupled applications
 - With large volumes of data and higher computation requirements, even closely coupled parallel applications can withstand higher communication latencies?

Problem: Large-scale computing on PC clusters is difficult	
Lots of resources (3000 PCs at \$5000): — 24 x 10 ¹² cycles/second (12,000 cores) — 24 terabytes of DRAM — 4 PB of storage (12,000 disk arms)	But, very hard to utilize: — hard to program 12,000 concurrent threads — something breaks every day (often a disk) — deployment, management is a big headache



- The computation is structured as a directed graph
- A Dryad job is a *graph generator* which can synthesize any directed acyclic graph
- These graphs can even change during execution, in response to important events in the computation
- Dryad handles job creation and management, resource management, job monitoring and visualization, fault tolerance, re-execution, scheduling, and accounting

Composable class can be implemented in high-level programming models such as Dryad