

# Measurement & analysis of distributed workflows

- Objectives: Performance objectives range from throughput to response-time/latency
- Shared resources:
  - Insight depends on system wide state (resource sharing for storage, network)
  - In storage & network, congestion can be very difficult to characterize and control
- Separability of bottleneck causes can vary:
  - Sometimes bottlenecks are interconnected and combine network, storage, memory, and processor.
- Many programming environments and OS environments (layers of virtualization)
  - Complex frameworks with compiled to dynamic interpretation
  - Layers of system virtualization (VM, containers, serverless) hide performance
  - Requires very different measurement, attribution, diagnostics, modeling techniques
  - each cloud coordination and platform framework is different
- Measurement and attribution challenges:
  - Storage paradigms: HDF (logical data) vs. Parallel semantics vs. POSIX layer (blocking, async, etc). Network: GridFTP vs. TCP (etc).
  - Networking paradigms: unreliable network (ethernet), reliable/high-performance (e.g., InfiniBand), large-scale "RDMA" (CXL)