RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

“Extreme-Scale” Distributed Computing

Scientific Computing Environments 2025

Shantenu Jha, Matteo Turilli,

Mark Santcroos, Andre Merzky

Research in Advanced DIstributed Cyberinfrastructure
and Applications Laboratory (RADICAL)

http://radical.rutgers.edu

Distributed Computing in 2013

- We are at extreme scale today with respect to 2001!
  - Promise of scale? or Illusion of scale?
- Distributed Computing Infrastructure (DCI)
  - In the past 12 years we have seen the emergence of the first sustainable production distributed computing infrastructure (DCI)
  - We are still learning how to architect large-scale systems
    - Macroscopic vs microscopic theory of distributed systems!
- Distributed Computing Applications (DCA)
  - Many new types of applications have emerged
    - More than first-generation, i.e., distributed HPC and BoT HTC
  - Many local solutions, lack of end-to-solutions
    - Scaling remains difficult for *individual* scientists
    - $O(10^{-2})$ can do O(100) tasks each of O(10GB) over O(10) nodes

Distributed Computing in 2013

- The space of possible DCA is large (and rich), but developing DCA remains a hard undertaking
  - More than just submitting jobs; coordination of different components
  - Inconsistent and incomplete tools for deployment and execution
- DCI software environment is complex and fragile
  - Middleware functionality and semantics
  - Heterogeneous software and system access layers
  - Difficult to integrate services and extend tools
- There are missing abstractions
  - Conceptual abstractions that enable reasoning
    - .. distributed performance, decomposition/aggregation (application and system), trade-offs etc.,
    - Currently difficult to estimate and mostly irreproducible results
  - Implementation abstractions that enable effective engineering

Distributed Computing practice for Large-scale Science doi 10.1002/cpe.2897

Distributed Computing in 2013

- What is the primary issue(s) of current DCI?
  - How to deliver "well-defined" *capabilities* that go beyond underlying technologies, tools or infrastructure to implement/provide them?
- Fundamental conceptual gap in providing well-defined capabilities
  - Lack of reasoning and ability to estimate/calibrate performance
  - Two levels of conceptual abstractions to enable reasoning:
    I. Models that enable functional comparison for *individual components*, e.g., P* for Pilot-systems 10.1109/eScience.2012.6404423
    II. Models that enable reasoning at multiple, integrated levels to provide performance estimation and predictability
      - When and how to distribute? What and where?
      - A Linpack for distributed systems/applications?

Distributed Computing in 2025

- Fundamental Question
  - Will DC-2025 be qualitatively similar to DC-2013?
    - Same req and challenges for DC-2013, but increased scale?
  - If qualitatively new or different, how?
- How will DCI en route evolve?
  - Functionally new components, but not drastically different from the current ones (barring unpredictable breakthrough(s))?
  - Compute, data and network units will scale along predictable lines
  - Mostly smooth transition as scaled-up, but implementation and geographical heterogeneities will become increasingly significant
- What will DCI-2025 look like?
  - Loose coupling (DoE) or tight-coupling (XSEDE)?
    - Neither. Very different. Need a different language altogether.
  - Collective properties of units will be different.

Extreme Scale DC: ATLAS/HEP

- Observation:
  - ".. Distributed computing will persist " for integrated HPC + HTC
    Richard Mount (SLAC), c.f. http://goo.gl/pJzIjH
- Requirement:
  - ATLAS in >2018 needs:
    - Non-monolithic extreme-scale and integrated HPC + HTC
- Challenges:
  - Mostly economic, but also how to manage workload decomposition
  - Development and deployment of future supercomputing applications
    - Role for flexible execution strategies
- Question:
  - ".. Are systems of the complexity of ATLAS Distributed Computing sustainable long-term?"

Extreme Scale DC: Square Kilometre Array (SKA)

- Observation:
  - Integrating leadership-class resource (IBM Machine) with many compute resources for extreme-scale real-time data-analysis
- Challenges:
  - Centralized exascale computing and networking infrastructure.
  - Compute and data intensive, world-wide analysis.
- Requirements:
  - Antennas distributed over 5K Km, equivalent to a dish with a collecting area of a square kilometer.
  - Continuous coverage from 70 MHz to 10 GHz.
  - Computing infrastructure: 10 PF - 1EF processing power; 10 - 100 PB/h; 300 - 1500 PB storage.
  - Computing technology: Optical cross connects; Phase-change memory; Chip stacking?

Extreme Scale DC: Human Brain Project

- Observation:
  - New application types and scenarios are necessary to create and simulate multi-scale brain models
- Requirements
  - In situ analysis of multi-Petabyte datasets.
  - System software and middleware supporting interactive computational steering and visualisation support.
- Challenges:
  - Integration of hierarchical storage-class memory in software for Big Data analytics.
  - Platform independence through the provision of high-level APIs and user-transparent programming paradigms
  - Virtualisation of the entire system including communication
- Question:
  - Using “current” technologies complemented by brain-inspired communication and computing sub-systems?

DC-2025: Foundational Requirements

- Support a broad range of DCA requirements
  - e.g. Large-scale simulations, big-data repositories, real-time computing, scientific experiments at global scale
  - Novel application classes: Adaptive Applications
- Balanced DCI and support for scaling along all dimensions
  - Scaling-up, Scaling-out, Scaling-across
- Separate capability from technology used to provide functionality
  - Capability: Well-defined and aggregated functionality, without regard to how, or the specific technology/approached used
    - e.g., Num. of tasks, throughput, probabilistic bounds on time-to-completion, performance of resources, data (volumes/ transfer/storage ability)

DC-2025: Foundational Challenges

- Distributed Computing Infrastructure:
  - Federate diversified set of resources at multiple levels
    - E.g. how/when to federate leadership machines with other less powerful machines?
  - Manage complexity and heterogeneity of infrastructure
    - Flexible deployment and execution
  - Providing capabilities
    - What functional units, and how to compose functionality?
    - Designing a federated system that scales along 1 dimension is relatively easy compared to scaling along >1 dimensions
- Distributed Computing Applications:
  - When and how to distribute? What and where to distribute?
  - Manage transition from static to adaptive applications

RADICAL Research Agenda

- Need to federate systems to provide well-defined capabilities from heterogeneous dynamic components with varying levels of control
- How to provide well-defined capabilities?
  I. Well-defined capability amidst heterogeneous, dynamic resources requires *flexible federation* of resources and services
  II. Reasoning about performance
    - Can design for randomness but not for unpredictable behaviour
  - Combination of reasoning (across possible configurations) and flexible federation points to a role for next-generation middleware
- How will applications utilize systems?
  - For given capability appropriate execution strategy is determined
  - Interoperability: DCI level? DCA level Interoperability?

RADICAL Research Agenda: Next-Generation Middleware

- Design Objective and Role of Next Generation Middleware
  - Provide well-defined capabilities
  - *Next Generation Middleware will be defined to be that which we can add to existing middleware layer(s) to provide systems based upon well-defined capability rather than a technology, or a specific execution strategy (say HTC or HPC), or a specific usage mode!*
- Federation via middleware:
  - Test Case: How would we federate XSEDE and OSG?
    - New complementary and non-destructive models of federation required
    - Adaptive execution strategy and flexible federation
    - Can't remove complexity, can only manage it, belief that it is best done with such middleware that supports interoperablity

Summary

- DC-2025 will look somewhat like DC-2013
  - Applications will scale-up along predictable lines
  - Individual DCI components will scale-up along predictable lines
- DC-2025 may look like DC-2013
  - Greater divergence between community vs individual applications?
  - How will community CI (ATLAS, LSST/SKA, *EONs) be federated with national-scale DCI (XSEDE, OSG, leadership-class machines)?
- DC-2025 will not look like DC-2013
  - Scale of heterogeneity, degrees-of-freedom will need addressing
    - "Just do it" wont work: will need a more reasoned approach
    - Complexity of treating individual resources will be too great
      - Collective "Properties and Design" Principles
  - We posit: Fresh perspective on reasoning and federation of resources and thus providing well-defined capabilities

Acknowledgement

- AIMES: Integrated Middleware Framework for Extreme Collaborative Science, Office of Advanced Scientific Computing and Research, Department of Energy ER26115/DE- SC0008591
  - Also Daniel Katz and Jon Weissman
- NSF CAREER Award, Division of Advanced Cyberinfrastructure (ACI), OCI-1253644
- RADICAL Members