Architectural Lessons: Look Back In Order To Move Forward

G. Djorgovski\(^1\), S. Catlett\(^2\), D. Crichton\(^3\), T. Huang\(^4\), B. Hughes\(^5\), E. Lai\(^6\), A. Mahabal\(^7\), D. Pineo\(^8\), T. Plante\(^9\)

1. California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, U.S.A.
2. Eastman 14, 101 N Columbus Street, Suite 205, Alexandria, VA 22314, U.S.A.
3. Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, U.S.A.

Abstract

True elegance of scalable and adaptable architectures is not about incorporating the latest and greatest technologies. Its elegance is measured by its ability to scale and adapt as its operating environment evolves over time. Architecture is the link that bridges people, processes, policies, interlinks, and technologies. Architectural development begins by obtaining an understanding of which technologies interface to the problem domain. It follows by the development of a domain architecture that manages information and knowledge flows that everyone can use [E. Alexander, 1979]. Architecture is the true artful. Like all masterpieces, the values and strength of architectures are measured not by the volume of publications, it is measured by its ability to evolve. An architect must look back in order to move forward. This talk discusses some of the prior works including onboard data analysis as well as data systems management and federated Nodes that interact dynamically through a Peer-To-Peer (P2P) paradigm. It is a dynamically distributed architecture where each Node consists of common software stack with standardized interfaces. This conceptual architecture serves as a high level system blueprint for the definition, construction, and deployment of both existing and new components to ensure that they can be unified and integrated into an evolutionary national infrastructure for EarthCube.

Capabilities which drives the conceptual architecture:

1. New architectural approaches across the entire science mission geoscientific domain and data lifecycle. These will increase the science yield by scaling and improving the integration of each of the various components of the lifecycle.
2. Increasing capability within EarthCube to support data reduction as part of managing bandwidth capabilities.
3. Integrated analytics as part of the data pipeline lifecycle.
4. Ability to quickly construct new analytic centers that can unify archives, computing capabilities, and software services to bring heterogeneous data sets as well as models/simulations together for analysis.
5. Systems that can react to increased velocity of data across the lifecycle with new technology approaches for data triage and capture.
6. Interoperability with other agencies including capture of data from their instruments and integration with their ground systems, archives, and analytic capabilities.
7. Quantitative uncertainty management within scientific inferences for the application of algorithms across the data lifecycle as data volumes increase.

Best Practices

- Common Software Stack
- Common Data Model that integrate with software
- Standard Interfaces
- Service-Oriented Architecture (SOA)
- Decoupling storage, compute, and data management
- Federated search
- Analytic services
- Visualization

Architectural Principles

- Scalability
- Community Driven
- Open Science
- Interoperability
- Sustainability
- Distributed

These activities were carried out at the Jet Propulsion Laboratory. California Institute of Technology, under a contract with the National Aeronautics and Space Administration. © 2015 California Institute of Technology. Government sponsorship acknowledged.