GreenLight Technical Report

2009

Service-Oriented Architecture

Ingolf Krueger, Pavel Pevzner, Steve Briggs, Matthew Arrott: GreenLight SOE

Research Activities

We have been exploring the applicability of domain modeling techniques to manage the complexity of an integrated architecture for a software-intensive system-of-systems such as the GreenLight instrument. Modeling plays an important role in all requirements engineering activities, serving as a common interface to domain analysis, requirements elicitation, specification, assessment, documentation, and evolution. Initially, domain models are created to describe the existing system in which the software should be built, covering stakeholders, human actors that interact with the system, hardware devices, and the environment in which the system will operate. In addition to behavior, domain models define "the language" of the system by capturing domain entities in a structural way [Jackson 1993]. Then, deficiencies in existing systems (i.e., traditional datacenters) and objectives for the target system are more clearly identified. During requirements elicitation, alternative models for the target system are created, which may define different boundaries between the target system and its environment. Models can help in defining the questions for stakeholders and surfacing hidden requirements. Ultimately, the requirements have to be mapped to the precise specification of the system and the mapping should be kept up to date during the evolution of requirements or the architecture.

Our efforts have focused on identifying the set of stakeholders for the Instrument, capturing a preliminary set of requirements regarding data management such as “which are the green parameters of interest?”, “what green parameters can be measured?”, “what accuracy is necessary for measurements?”, etc. Building such understanding regarding the data collection process is crucial for the later development of green experiments which aim to identify energy consumption in relation to the utilization of the various resources made available by the GreenLight instrument. By means of a sequence of working meetings with other PIs and their team members, we identified two classes of system-related data producers: the communication infrastructure and the underlying physical computing infrastructure; and several classes of environment data producers (e.g., cooling fans, cold water intake, temperature sensors, etc.). There are more classes of data consumers with requirements as diverse as reading CPU temperature at 10ms intervals, to optimizing cooling fans air flow at minute granularity, to measuring PDU phases per rack and building live depictions of the power load of the Instrument.

As part of the ongoing requirements elicitation process the resulting specifications are checked for errors such as incompleteness, contradictions, ambiguities, inadequacies in respect to the real needs – which all can have negative effects on the system development costs and the quality of the resulting product. The choice of modeling notations is often a tradeoff between readability and powerful reasoning techniques: natural language is very flexible, useful for communicating requirements, but cannot capture relationships and is often an expression of subjective reasoning [Zave 1995, Zave 1997]; applied/semi-formal models (e.g., entity-relationship diagrams, UML diagrams, structured analysis) typically have a graphical representation which is very useful when communicating with stakeholders and often offers simulation and animation capabilities; and formal notations (e.g., KAOS [Dardenne 1991, Dardenne 1993], SCR [Heninger 1978, Heninger 1980], process algebra, Promela / SPIN [Holzmann 1997]) capture precise semantics, which supports rich verification techniques.

We used UML-style diagrams to capture critical elements regarding the power utilization of the Instrument and its constituents. Through an agile development process, we continuously incorporate the requirements and derived models into an integrated architecture for a

cyberinfrastructure foundation for the Instrument. The intention is to build a loosely coupled service-oriented architecture (SOA) to incorporate the hardware and software constituents of the Instrument and enable dynamic resource management in response to power related policies. For this purpose, we are working on resource models for the GreenLight resources (e.g., CPUs, virtual machines, physical nodes, storage elements, etc.). The models are then used to define classes of resources that may share or have dependencies in their usage policies and scheduling algorithms (e.g., scheduling of virtual machines on the actual physical machines and between physical machines).

We have also been working on models and implementation for a Proteomics research platform to enable bioinformatics scientists to define specialized mass-spectrometry analysis workflows and execute them on the GreenLight Instrument. These models are relevant for inferring relationships between particular proteomics tools, resource utilization models, and energy utilization in response to their execution. The results drive an upcoming design for a computational mass spectrometry cyberinfrastructure by optimizing computing architectures in relation with the proteomics experiments intended to be executed (e.g., using general purpose CPUs vs GPGPUs with specialized adaptations of the proteomics algorithms), the communication bandwidth required for transporting the data between the processing nodes, and centralized or distributed storage infrastructure (e.g., exploiting the data locality principles that apply to data intensive algorithms).

The result of our activities, in the long-run, will be a loosely coupled service-oriented architecture (SOA) to integrate multiple hardware and software components of the instrument, allowing specific policies to be enforced by pluggable entities enabling on-demand monitoring and control of the GreenLight Instrument. In effect, the SOA will become a principled way for accessing the resources offered by the instrument under energy and other quality of service policies.

For more information please contact Ingolf Krueger, ikrueger@ucsd.edu