Migrating Legacy Applications into the Cloud
Supporting Heterogeneous Cloud Platforms
Safe Cloud Updates

Submission to "The Science of Cloud Computing" NSF PI Meeting

Name: Iulian Neamtiu, Assistant Professor.
Affiliation: Department of Computer Science and Engineering, University of California, Riverside.
Email: neamtiu@cs.ucr.edu
Topics:
4. Programming Models for the Cloud
7. Cloud Debugging, Certification, and Update

1 Summary of Current Research Activity

Our current research is focusing on elastification (making legacy applications scalable in the Cloud), and developing infrastructure for Cloud-wide dynamic updates.

Program elastification. Monolithic server programs are gradually migrating into the Cloud. While Cloud platforms are elastic, today's server programs are not written to run well on elastic platforms, i.e., utilize a wide-range of available resources. We are currently working on a tool called Elastin that transforms inelastic programs—programs written with a fixed set resources in mind—into elastic applications that run on elastic (Cloud) platforms by adapting, at runtime, to changes in the available resources. We believe this approach can be a stepping stone towards migrating legacy applications to the Cloud.

Support for dynamic updates to Cloud applications. Cloud service providers need to provide continuous service while accommodating the fast evolution pace (up to several updates per week) characteristic of Cloud applications. To address the tension between frequent updates and high-availability, we are working on an approach to updating Cloud applications via safe, on-the-fly updates. Expanding on our past work on dynamic updates to applications and database schemas, we are currently working on dynamic updates to Cloud Compute and Storage services. Our current focus is on allowing developers to construct Cloud applications in a natural way, while at the same time providing compiler and runtime infrastructure that (1) permits dynamic updates to applications, and (2) ensures update safety, i.e., following a dynamic update, applications continue to operate as expected, at the new version.

2 Future Research Problems

We believe that migrating legacy applications into the Cloud, making Cloud applications scalable and fault-tolerant, adapting Cloud applications to new hardware platforms, and ensuring safe (dynamic) Cloud application updates, present significant future research challenges and opportunities.

2.1 Self-adaptive Cloud Applications from Legacy Applications

Companies have huge amounts of legacy code that must be migrated to the Cloud, and rewriting this code from scratch for the Cloud is not always feasible. There are three main problems with running legacy applications in the Cloud: scalability, fault-tolerance, and administrative burden. Ideally, applications would adapt to changes in available resources, would use isolation and migration to achieve fault-tolerance, and would automate the deployment and management. Current Cloud infrastructure support for these features is limited, e.g., scalability is achieved by starting and stopping application instances, but for that applications have to be designed from scratch around a variable number of instances. We believe that automating scalability and fault-tolerance are particularly important research goals. Our initial results (Section 1) show that we can construct scalable, elastic applications from inelastic legacy applications relatively easily. Infrastructure for achieving application scalability and fault-tolerance, i.e., language, compiler and safety analyses, as well as tools that monitor application performance and make adaptation decisions constitute a promising area of research.

2.2 Supporting Heterogeneous Cloud Platforms

To tailor resources to the needs of individual applications, Cloud service providers are making new hardware instances available, e.g., Amazon has begun offering "Cluster GPU" instances, and Intel advocates Atom-based Green Cloud computing. We believe that Cloud applications should be able to take full advantage of these new resources and other possible future hardware platforms such as CUDA or FPGAs. For that, applications need the support of compiler, runtime and VM systems. For example, an application could be compiled to x86 binary, OpenCL and HDLs. When available (and advantageous), the applications should be able to transition among these heterogeneous platforms. The infrastructure for this (compiler and programming language support), as well as translation and migration tools constitute a promising research area.

2.3 Safe Updates to Cloud Applications

To stay competitive, online service providers must release very frequent updates; anecdotal evidence suggest several updates ("pushes") a week at Facebook and Google. There are two main problems with frequent updates: (1) insufficient testing leads to update errors and buggy new versions, and (2) frequent updates mean frequent service unavailability, or frequent service degradation.

We believe these two problems can be tackled with formal techniques for guaranteeing Cloud update safety, and with on-the-fly Cloud application updates. There are, however, important research challenges that need be addressed: first formally modeling software updates and ensuring end-to-end update consistency guarantees (from the client to Compute Services to Storage services); second, to allow Cloud application update on-the-fly, new dynamic software update mechanisms need to be designed and implemented.