GAIA: Strengthening the Reliability of Datacenter Computing with Fast and Scalable Distributed Consensus

Abstract:

To deal with the rapidly increasing volume of data, more and more software applications run within a datacenter containing numerous computers. To harness the massive computing resources, applications are deployed with two major types of infrastructures: schedulers and virtual machines. These infrastructures have brought many benefits, including improving resource utilization, balancing loads, and saving energy. Unfortunately, as an application runs on more computers, minor computer failures will occur more likely and can turn down the entire application, causing severe disasters such as the 2015 NYSE trading halts and recent Facebook outages. Existing infrastructures lack support to ensure high-availability for applications.

This GAIA project takes a holistic methodology to greatly improve application availability with three objectives. First, we will create a fast, scalable distributed consensus protocol for general applications. Distributed consensus provides strong fault-tolerance: it replicates the same application on different computers and always enforces same inputs on these computers, as long as a majority of computers are alive. An open challenge is that traditional consensus protocols are too slow mainly because their messages go through OS kernels and software TCP/IP layers. We tackle this challenge by creating a fast consensus protocol with an ultra-fast kernel bypassing technique called Remote Direct Memory Access (RDMA). Our preliminary results published in [SOSP '15] show that our protocol can support diverse, unmodified applications, and our latest protocol was 20X~31X faster than traditional protocols even when running on 35X more computers.

Second, we will construct the first fault-tolerant scheduler by integrating our protocol with popular datacenter schedulers. To seamlessly achieve resource allocation and application replication, we propose a new replication-aware resource allocation workflow for our scheduler. Preliminary results published in [APSys ’16] show that this scheduler can efficiently support Memcached, a widely used key-value store.

Third, we will make our protocol and virtual machines (VM) form a mutual-beneficial eco-system. This eco-system not only leverages the VM hypervisor layer to automatically replicate applications, but it also introduces a new VM live migration approach for computer load balance. To migrate application execution states to remote computers, prior live migration approaches incur substantial application down time and resource consumption on local computer. Our new approach need only migrate a consensus leadership, which consumes almost zero time and resource.

We believe that our expertise on building reliable distributed systems and our preliminary results will help us achieve all the three objectives. By greatly strengthening the availability of

datacenter applications, this project will benefit almost all computer users and software vendors, including many financial platforms in HK. This project will also advance various datacenter techniques (e.g., migration) and attract researchers to build more reliable infrastructures.

Long term impact:

To deal with the rapidly increasing volume of data, more and more software applications run within a datacenter containing numerous computers. Many applications are mission-critical and naturally demand high reliability and performance, including financial platforms, social network platforms, military services, and medical services. To harness the massive computing resources, applications are deployed with two major types of infrastructures: schedulers and virtual machines. These infrastructures have brought many benefits, including improving resource utilization, balancing loads, and saving energy.

Unfortunately, as an application runs on more computers, minor computer failures will occur more likely and can turn down the entire application, if the failure computer runs a critical computation. For instance, recently, both New York Stock Exchange (NYSE) and Nasdaq have experienced outages of their whole site or delays of IPO events due to minor machine errors. In addition, social network applications such as Facebook tend to be online 24-7, but minor computer failures have turned down the whole Facebook site for several times in the last few years. All these outage events have led to huge money lost.

A key problem of these disasters is that datacenter infrastructures lack high availability support for general applications. As a result, although some applications have built ad-hoc replication approaches to improve their own availability, most other applications still suffer.

This proposed GAIA project takes a holistic methodology to tackle this problem. It first plans to build a fast, scalable replication protocol, it then integrates this protocol into two major infrastructures, potentially benefiting all applications. We envision significant impacts for this project in different terms.

In the near term, Falcon (Objective 1) can largely improve both the scale and performance of many replication systems. For instance, a notable key-value store system called Scatter deploys 8~12 replicas in each Paxos group, and now it can deploy hundreds of replicas in each group and achieves much better performance. Overall, a fast, scalable, and general service, Falcon may significantly promote the deployments of Paxos and improve both the performance and fault-tolerance of various systems in datacenters.

In the intermediate term, by realizing Objective 2 and 3, we will greatly strengthen the availability of datacenter applications and benefit almost all computer users and software vendors, including many financial platforms in HK. Because HK has lots of financial platforms which naturally desire high availability in their operational hours, this GAIA project can bring practical benefits to these platforms and avoid horrible outages such as the 2015 NYSE trading halts.

In the long term, we anticipate that this project will advance various datacenter techniques (e.g., live migration) and attract researchers to build more reliable infrastructures. As datacenter emerges to be a "giant computer", and a future datacenter OS for such a computer has gradually come up. Therefore, consensus protocols, schedulers, and virtual machines will become essential datacenter OS components, and the outcomes of this project will eventually be adopted in a future datacenter OS.