ISSUE 37 | OCTOBER 2010

SOCIETY OF ACTUARIES Technology Section

CompAct

ELECTRONIC NEWSLETTER

CONTENTS

QUICK LINKS

ACTUARIES LOOKING TO THE SCAI FOR ANSWERS

by Van Beach, FSA, MAAA

When running stochastic projections, the actuaries' need for computing power is insatiable. A multi-tier Scalable Cloud Actuarial Infrastructure (SCAI) may hold the solution.

Imagine the following scenario: Results are needed immediately, but the in-house compute cluster (i.e., the grid of servers used to parallelize computations) is already fully utilized and there is a queue of jobs that will keep the resources occupied for days. A job could get bumped, or priorities reordered, but even with all 200 computing resources (cores) in the cluster available, running another 1,000 scenarios will require more than 10 hours to complete.

The 10-hour run time is certainly a dramatic improvement over running on a single computer (which was the situation before the compute cluster was installed), but is still far from ideal.

The drivers are well-known–greater reliance on actuarial models for risk management and reporting, more complex models, stochastic methodologies, etc. The same situation seems to play out every quarter-end–everyone needs the cluster at the same time. There is no indication the problem will go away.

A better solution is needed.

Instead of 200 cores available 365 days a year, a more effective distribution might be 2,000 cores 36.5 days a year. At certain times,

Council

Links of Interest

Fiction Contest

Howard Callif, Editor

SOA Staff
Meg Weber, Staff Partner

Sue Martz,
Section Specialist

Sam Phillips, Staff Editor

20,000 cores for 3.65 days would be the optimal configuration. If computing capacity were marginal cost, these would all be equivalent. In addition to marginal cost, the ideal solution would also provide infinite, on demand, high performance resources. While still nirvana, several Scalable Cloud Actuarial Infrastructure (SCAI) options make this vision more than just a daydream.

Clouds and SCAI

Before proceeding further, some additional definitions are required. While there are many definitions of a "cloud," I prefer an inclusive version, which suggests that a cloud is a computing resource–server, storage, software, etc.–that is available via a network, often an Internet connection. Typically, the cloud operates on an on demand, pay-per-use business model and provides the illusion of being an infinite resource. The cloud can be internal or external (i.e., provided by a third party).

I'm defining a "SCAI" (pronounced like "sky") to be a subset of the resources available in the cloud. First, a SCAI is an infrastructure (servers, storage, network, etc.), rather than a software application. The software can be initiated from the desktop, but is executed in the SCAI. Second, the SCAI can be used to parallelize and distribute actuarial projections. Third, the SCAI resources are effective in supporting actuarial calculations (not all infrastructures can handle the unique performance requirements of actuarial models).

In summary, a cloud infrastructure that can be leveraged for parallelizing actuarial calculations is a SCAI. By this definition, an in-house cluster of computers is a SCAI: a bank of servers, centrally available via a network, used to distribute calculations for actuarial models.

I'm going to discuss four computing options that I believe will play key roles in providing the flexibility and capacity to meet actuarial demands as part of a SCAI:

1. Multi-core computers
2. In-house compute clusters
3. Third-party clusters
4. Cloud platforms

Multi-core Computers

The power packed into a desktop computer continues to spiral

upward, largely due to additional processors and cores. These create the potential for a self-contained "cluster" where actuarial projections can be parallelized across the cores. No sharing or queuing is required when the cluster is your desktop, and it is often much easier and more efficient to run a projection locally rather than trying to schedule time on a shared cluster of computers.

If budgets don't allow each user to have a multi-core computer, a common practice is to have a bank of shared "work" machines that are available via Windows Remote Desktop, essentially creating a local cloud of work machines. With this arrangement, users log into the more powerful work machines when projections need to be run. If the data reside on a local network, moving the job execution from your personal computer to a shared work computer should be a snap.

In-house Compute Clusters

In recent years, the prevalence of in-house compute clusters to support actuarial calculations has continued to increase, providing a computational "backbone" for actuarial projection work. The key innovation that triggered this expansion was the introduction of Microsoft HPC: a cost-effective, centralized operating system that turns a bank of servers into a scalable resource that can be made broadly available.

The introduction of Microsoft HPC effectively removed the cost hurdle for implementing industrial-strength distribution, scheduling, and cluster management software. While hardware costs and IT support are still issues that must be addressed, many companies have been able to implement small clusters to greatly alleviate computing bottlenecks for their actuaries. For most companies, an in-house cluster–a fixed set of dedicated resources, maintained internally, and managed via Microsoft HPC–is their primary source of actuarial computing capacity.

Third-party Clusters

The concept of providing a shared compute cluster that many companies can leverage during periods of peak demand has a lot of intuitive appeal–until met with the practicalities of actuarial computing requirements. The biggest challenges to this business model have been the unique performance profiles of actuarial projections and the correlation of peak demand across all potential customers; everyone needs the extra capacity during the same periods during the year, so there is no way to fill in "valleys" with

another company's "peaks."

However, there are examples of vendors who are proving to be quite successful in providing their infrastructure as a service offering to insurers. An example is R Systems NA, Inc., in Champaign, Ill., which highlights several key traits required to be a viable alternative in this market.

First, they are able to provide flexible scheduling to address peak demand periods. While this has been a stumbling block for others, R Systems provides their services to multiple verticals, thus allowing insurance peaks and valleys to be offset by other industries' uncorrelated usage patterns. Second, R Systems' hardware is current, turns over often and is tailored for high performance computing. Third, they have gained an understanding of the unique performance characteristics of actuarial models and are quite skilled in tuning their environment for high performance. Finally, their costs create a solid value proposition, as compared with in-house computing resources.

Using a third-party comes with risks–counterparty, availability, security, etc.–but the potential to augment an in-house cluster with additional capacity for peak periods has tremendous value for many insurers.

Cloud Platforms

Taking the infrastructure as a service concept one step further, a cloud platform (a platform as a service) provides the hardware, storage, network, availability, support and more, all in a pay-as-you-use package. For actuarial computing needs, however, not all platforms are created equal. The definition of a cloud infrastructure suggests the illusion of being infinite. The reality is that not all clouds can maintain this illusion– especially as the resource demands increase.

When considering the computational demands of actuarial projections–thousands of scenarios per projection, multiple projections per user, several users per company across potentially hundreds of companies, it is clear that only the largest cloud platforms will be viable options. Microsoft Azure is one cloud platform that has potential to be part of the actuarial computing capacity solution. It is still in its relative infancy, but Microsoft's commitment to providing this service and the trajectory of resources being brought online suggests Azure will be a viable for the long term.

Further, the potential to provide (the illusion of) infinite, on demand, marginal cost resources is clearly a game changer when trying to meet actuarial computing demands.

Putting It All Together

The supposition might be that a company should choose one of the above options for their SCAI. What I would suggest, however, is that the SCAI should consist of all four: multi-core work computers, an in-house cluster, third-party cluster(s), and a cloud platform. Each provides a different value proposition and meets a different need. Simplistically, the usage might be as follows:

- Day-to-day model development and testing is done on shared work computers.
- The in-house cluster is available for on demand ad hoc model runs, larger tests, benchmarks, etc., during off-peak periods. During peak periods, the in-house cluster is used for the mission-critical runs and generally as the first tier of capacity
- The third-party cluster might be arranged to meet 80 to 90 percent of peak demand and scheduled only during peak periods. The third-party cluster provides overflow capacity needed above the steady-state levels met by the in-house cluster. Given there are often better marginal rates with advance notice, this need should be planned as much as possible.
- The cloud platform is then used for unscheduled or exceptionally heavy periods of capacity demands. It provides the final tier of capacity when other sources have been exhausted.

Going back to our original challenge:

- 200 cores available 365 days a year could be met with an in-house cluster,
- 2,000 cores 36.5 days a year could be met with a third-party cluster, and
- 20,000 cores for 3.65 days could be provided by a cloud platform.

In the case of the cloud platform, the resources are marginal cost, (provide the illusion of being) infinite, and on-demand. If high-performance resources are required, a third-party cluster can provide these. Yes, this is more than just a daydream. A SCAI that consists

of all four options provides flexibility and capacity to meet essentially any flavor of computing demand.

The Role of the Application

Each of the above–multi-core computers, in-house clusters, third-party clusters and cloud platforms–provide powerful options for meeting the actuarial capacity challenge. However, not all options will necessarily be available (or effective) for a given actuarial projection software application. There are at least two levels to consider: integration and optimization.

First, an application needs to provide an integration layer to make the computing option available. For example, there needs to be an integration layer between the application and Microsoft HPC for the Microsoft compute cluster to be available as a compute option. Third-party cluster providers might be able to use multiple integration options, but the integration layer is still required. Cloud platforms again require specific development to be able to be utilized as part of a SCAI.

It is a different question to ask whether the application has been optimized for the given computing option. For example, an application might be able to take advantage of all cores in a multi-core computer with the use of Microsoft HPC software. However, this introduces a complexity and overhead that would not be required if the application could automatically detect and utilize the available cores. Working with a cloud platform like Azure also requires a tremendous amount of effort and expertise on the part of the application vendor to provision and utilize the cloud resources optimally.

In short, the application needs to be integrated and optimized for the given computing option. Each application will have a different level of integration, optimization, and value for the given computing option. So the optimal SCAI configuration will be application specific, in addition to company specific.

Smart Modeling

Lest we forget amid all this infrastructure discussion that modeling techniques also play a role in addressing the capacity challenge. Cluster modeling and scenario reduction techniques are two examples of methodologies that are very effective in reducing computation time while maintaining a high level of accuracy.

Summary

The need for actuarial computing capacity is not going to recede. To meet regulatory and risk management requirements, companies will need to find solutions to this capacity challenge. There are multiple options–multi-core computers, in-house clusters, third-party clusters, and cloud platforms–that can be leveraged, assuming your actuarial projection software can support the option. The question is no longer "if" additional capacity is needed: The question is "how" the capacity can most effectively be provided. When considering this question, actuaries should look to the SCAI for answers.

Van Beach, FSA, MAAA, is the MG-ALFA product manager at Milliman and can be reached at van.beach@milliman.com.

Actuaries
Risk is Opportunity.®

475 North Martingale Road, Suite 600 Schaumburg, Illinois 60173
Phone: 847.706.3500 Fax: 847.706.3599 www.soa.org