Where Scalability Meets Efficiency: Using MATLAB and Cloud Computing to Run Our Internal Risk Model
MathWorks Finance Conference 2022
D. Rubio – 11 Oct 2022
## Introduction

<table>
<thead>
<tr>
<th>Background</th>
</tr>
</thead>
</table>
| • Swiss Re is one of the world's leading providers of reinsurance, insurance, and other forms of insurance-based risk transfer.  
• With headquarters in Switzerland, it uses a FINMA-approved internal capital adequacy model to comply with Swiss Solvency Test and Solvency II requirements.  
• The model is also used for business steering purposes and its usage has grown significantly over the past couple of years. |

<table>
<thead>
<tr>
<th>Status Quo</th>
</tr>
</thead>
</table>
| • The model is part of a complex technology ecosystem operated in a private cloud since 2019.  
• The current set-up is cost efficient and works well, improvements are regularly being introduced.  
• But some challenges remain which prevent the future ambitions to be achieved with the current platform. |

<table>
<thead>
<tr>
<th>Outlook</th>
</tr>
</thead>
</table>
| • After a successful exploration phase, we are kicking off a transformation program to migrate our risk models to Azure.  
• This transition will provide tangible benefits and fully aligns with Swiss Re’s public cloud strategy. |

<table>
<thead>
<tr>
<th>Objective of today</th>
</tr>
</thead>
</table>
| • Outline our cloud transformation journey.  
• Discuss the operational and financial benefits of the current platform, and how we are addressing the challenges.  
• Discuss the transition to public cloud and how we are partnering with Mathworks to customize the reference architecture for Azure. |
Swiss Re's capital position remained very strong in H1 2022, with the Group Swiss Solvency Test (SST) ratio above the 200–250% target range.

Capital Adequacy Ratio

Core Risks in internal model

PROPERTY & CASUALTY
- Natural catastrophe
- Man-made
- Costing & reserving
- Claims inflation

LIFE & HEALTH
- Lethal pandemic
- Mortality trend
- Longevity

FINANCIAL MARKET
- Credit spread
- Equity
- Foreign exchange
- FM inflation
- Interest rate
- Real estate

CREDIT
- Default risk
- Migration risk

Source: Swiss Re Risk Modelling
The private cloud has been a great learning experience, and has prepared us well for the next phase of the cloud journey.

**On-Premise 2017**
- Physical infrastructure
- Pay per plan
- Centralized service
- Limited product offering
- Managed OS

**Private Cloud 2019**
- Virtual infrastructure
- Pay per use
- Self-service
- Limited product offering
- Managed OS – IaaS OK

**Public Cloud 2023**
- Virtual infrastructure
- Pay per use + reservations
- Self-service
- Broad product offering
- Unmanaged OS – IaaS NOT OK
The internal model is deployed as the core of a complex technology ecosystem. A self-managed cluster provides the significant computing power required to perform large scale numerical calculations.
Our Cluster Manager scales programmatically the resources available and provides cost efficiency.

How it looks:

- Head Node
- Parallel Server
- Cluster Manager
- Worker Node X
- Worker Node Y
- Worker Node Z

How it works:

- Create blocker job
- Check cluster status (VMs, MJS, Worker)
- Check job queue
- **Start VM** / Install MJS / Start Workers
- Execute job
- Stop Workers / Uninstall MJS / **Stop VM**
Significant savings can be achieved thanks to on-demand scaling and pay-per-use. Average usage reduced despite substantial increase in demand.
Overcommitment in a virtualized set-up may lead to performance volatility at worker level, particularly for CPU-intensive applications, leading to longer calculation times.
Transforming a few highly-dependent jobs into many independent jobs can be a game changer, particularly in a cloud context.

### Cloud-optimized parallelization scheme

**Before:**
Fixed pool of 40 workers processes 40 equal, massive jobs.

**After:**
Scalable pool of workers asynchronously processes ~3000 small jobs.

### Benefits

<table>
<thead>
<tr>
<th>Benefit</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalability</td>
<td>Pool of workers can be enlarged (faster, more expensive) or made smaller (slower, cheaper) depending on business need.</td>
</tr>
<tr>
<td>Modularity</td>
<td>Individual jobs (=models) are disentangled. This increases flexibility for future model development.</td>
</tr>
<tr>
<td>Resilience</td>
<td>A single job failing no longer leads to a failure of the whole calculation. Instead, the failing jobs can be individually retried.</td>
</tr>
<tr>
<td>Performance</td>
<td>Performance is much less sensitive to a single low performing worker.</td>
</tr>
</tbody>
</table>

### Resilience improvement

**Old parallelization:** An error in any part of the model crashed the entire calculation.
**New parallelization:** Failed jobs are retried.

21-Jun-2022 21:02:14: Calculating transactions in run-off...
Error using Calculate/setXlTransaction (line 1436)
Out of memory. Type "help memory" for your options.
21-Jun-2022 21:02:17: Failed job id 715. Retrying...

2022-06-21 21:24:00.279 INFO Workflow technically completed
2022-06-21 21:24:00.282 INFO Semantic status: COMPLETED

### Performance improvement

**Old parallelization:** Slower workers delay the entire calculation.
**New parallelization:** Faster workers no longer wait.

09-Sep-2022 14:28:36: Running jobs are:
CriticalIllness for slice 1 (elapsed: 54.62s, JobID:149)
CriticalIllness for slice 2 (elapsed: 54.61s, JobID:150)
CriticalIllness for slice 3 (elapsed: 48.33s, JobID:151)
CriticalIllness for slice 4 (elapsed: 38.01s, JobID:152)
CriticalIllness for slice 5 (elapsed: 35.03s, JobID:153)
CriticalIllness for slice 6 (elapsed: 35.73s, JobID:154)
CriticalIllness for slice 7 (elapsed: 35.86s, JobID:155)
CriticalIllness for slice 8 (elapsed: 23.91s, JobID:156)
CriticalIllness for slice 9 (elapsed: 18.55s, JobID:157)
CriticalIllness for slice 10 (elapsed: 12.42s, JobID:158)
CriticalIllness for slice 11 (elapsed: 2.30s, JobID:159)
CriticalIllness for slice 12 (elapsed: 0.66s, JobID:160)
LifeProtection for slice 13 (elapsed: 813.77s, JobID:1833)

09-Sep-2022 14:29:16: Jobs outstanding: 716

---

**Source:** Swiss Re Risk Modelling

---

D. Rubio | 11 Oct 2022 | MathWorks Finance Conference 2022
We have used a self-managed cluster as a simple way to enter Azure, and used the insights we gained to subsequently adopt the reference architecture.
Supported by Microsoft, we have used a self-managed cluster to optimize the infrastructure choices in Azure for our particular use case.

**Computing**
- High-performance computing (1:1): H-, HB- and HC-Series
- Memory optimized (2:1): E-Series (also non-hyperthreaded)

**Storage**
- File share standard: hot, cold and transaction optimized
- File share premium
- Blob storage

**Network**
- Accelerated networking ON/OFF

**OS**
- Windows Server 2016, 2019 and 2022

**Matlab version**
- 2019b, 2021b and 2022b

**With the elephants**

- **AZURE E32 NHT**: 0.81
- **AZURE E32**: 0.92
- **AZURE F336**: 1.09
- **AZURE HC44**: 1.08
- **AZURE H16**: 1.26
- **SWISSCOM**: 1.00

**With the mice**

- **AZURE E32 NHT**: 0.73
- **SWISSCOM**: 1.00
We can’t use the Azure Marketplace product out-of-the-box due to internal security restrictions, so we collaborated with Mathworks to customize the deployment process.

Original deployment steps:
- Use Mathwork’s own image with a given OS and Matlab version
- Deploy infrastructure components
- Apply Azure configuration (e.g. network)
- Run post-deployment scripts to install and configure Parallel Server

Mitigation actions:
- Migrate deployment template from ARM to Terraform (optional)
- Replace default VNET with own private VNET
- Remove public endpoints
- Amend post-deployment scripts so that Parallel Server avoids public endpoints
Combining a scale set with a smart scaling agent is another game changer, as it unlocks unprecedented computing capacity in a highly-efficient manner.

### Manual Scaling

- Manual scale
  - Depending on your needs

### Programmatic Scaling

#### Head Node

- Cluster Manager
- Parallel Server

#### Cluster

- VM Scale Set

### Custom Scaling

- Standard autoscale
- Predictive autoscale
- Scheduled autoscale

### Scaling use case: load testing the File Share

- 1Gbps: 1.01
- 2Gbps: 1.00
- 3Gbps: 1.00
- 4Gbps: 1.02
- 5Gbps: 1.04
- 6Gbps: 1.02
We worked with Mathworks to understand the ingredients of the Azure Marketplace image, and we created our own to have full flexibility.
An updated custom image removes the need for “online” OS updates, and results in increased availability and reduced operational support.
Infrastructure reservations represent an additional way to optimize the cost of computing resources, but predicting the right amount of reserved instances is not an easy task.

Cluster Size = 18 VMs
VM Cost = 1’000 USD
Discount = 29%
Any questions? Thank you!

Contact us
Daniel Rubio
Head Integrated Risk IT
Daniel_rubio@swissre.com

Acknowledgments
Danny Rasch, Salvi Amato, Francesco Atzeni, Tobias Frischknecht, Benjamin Troxler, Lionel Trebuchon
Edu Benet Cerda, Yannis Ben Ouaghrem, Alison Eele, Karen Tanner
Darko Mocelj, Daniele Rispoli, Shayne Kropp

Follow us