

High Performance Computing in Clouds: Challenges and Solutions

Lúcia Drummond
(lucia@ic.uff.br)
Universidade Federal Fluminense

About me

Current Position

- **Full Professor at the Fluminense Federal University (UFF),
Institute of Computing**
- **Researcher at CNPq (Brazilian Research Agency)**

Research Interests

- Parallel and Distributed Computing
- High Performance Computing (HPC)
- Cloud Computing



Fluminense Federal University Niterói, Rio de Janeiro



Institute of Computing - UFF



Institute of Computing - UFF

<https://www.ic.uff.br>.

Our Graduate Program on Computer Science is among the 7 excellence programs in Brazil in this area

<https://www.ic.uff.br/pos-graduacao>.

Researches in several areas, such as Algorithms, Artificial Intelligence, Databases, HCI, Operations Research, Quantum Computing, Software Engineering, Virtual Reality, Cloud Computing, High Performance Computing, etc.

Outline

- Preliminary concepts
- Why move HPC applications to clouds?
- Some Challenges
- A Case Study: SIM@CLOUD
- Conclusions and Future work

Cloud Computing

National Institute of Standards and Technology (NIST) 2011: Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Emergence of Cloud Computing for HPC

- **From Grid to Cloud:**

Grid Computing: Early attempt to connect distant PCs and clusters via the internet to be used as an hpc platform

Cloud Computing: Born from underutilized business datacenters and virtualization.

- **Technical Synergy:**

Like supercomputers, cloud datacenters use clusters of general-purpose processors and high-speed networks.

Emergence of Cloud Computing for HPC

- **Scale Advantage:**

The total volume of cores in a cloud datacenter can be significantly larger than traditional supercomputers.

Cloud Computing and Supercomputing

Reinventing High Performance Computing: Challenges and Opportunities, Daniel Reed, Dennis Gannon, Jack Dongarra
March 8, 2022

“Change is now in the wind... The major cloud vendors have invested in global networks of massive scale systems that dwarf today's HPC systems. Driven by the computing demands of AI, these cloud systems are increasingly built using custom semiconductors, reducing the financial leverage of traditional computing vendors. These cloud systems are now breaking barriers in game playing and computer vision, reshaping how we think about the nature of scientific computation. Building the next generation of leading edge HPC systems will require rethinking many fundamentals and historical approaches ...”

Why Migrate HPC to the Cloud?

- No need to buy or maintain physical hardware/cooling.
- Eliminates the need for specialized hardware maintenance staff.
- Immediate access with no "queue wait times."

Challenges

- **Financial Cost**
- **Optimization:** Find the specific infrastructure that "fits" for each application.
- **Elasticity:** Adapting legacy apps to scale up and down dynamically
- **Security**

Cloud Computing and High Performance Computing

Several scenarios:

- **HPC in the Cloud:** application executed in the cloud
- **HPC plus Cloud:** the Cloud is used together with other HPC resources
- **HPC as a Service:** HPC treated as a service by the Cloud provider

A case study : SIM@CLOUD

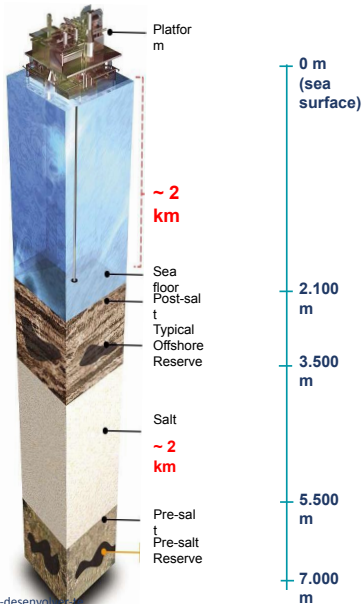
cooperation between UFF and Petrobras, published in Alan L. Nunes, Daniel B. Sodr , Cristina Boeres, Jos  Viterbo, L cia M. A. Drummond, Vinod E. F. Rebello, Luan Teylo, Felipe Albuquerque Portella, Paulo J. B. Estrela, Renzo Q. Malini:

A Framework for Executing Long Simulation Jobs Cheaply in the Cloud. IC2E 2024: 233-244



Petrobras

- Brazilian Energy Company
- Oil and Gas Exploration and Production is still its core business
- Reference in ultra-deepwater exploration
- Why HPC at Petrobras?
 - Two main disciplines/workloads:
 - Geophysics – Seismic Processing
 - Reservoir Engineering – Reservoir Simulation
 - Latin American No.1 in TOP500 and Green500 lists
 - Multi-million dollar public cloud user



(1) <https://www.agencia Petrobras.com.br/pt/inovacao/petrobras-monta-supercomputador-para-desenvolvimento-de-tecnologias-18-01-2023/>

(2) Petrobras Strategic Plan 2023-2027

Motivation

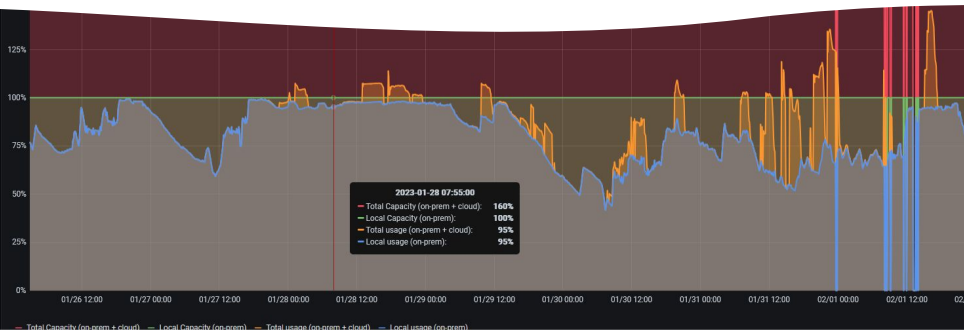
- Simulation is a key analysis tool for industry and requires substantial amounts of computing power
- These executions can lead to unpredictable computational demands, meaning that
 - on-premises infrastructures may quickly become ill-equipped to meet their needs
- Cloud platforms appear to offer *limitless computing resources and storage space*
 - *The question is “By leveraging cloud resources, can industry run their simulations with enhanced efficiency, more cost-effectively, and with less effort?”*

Petrobras

- Investments in **on-premise** infrastructure often **lag behind** capacity demands

- Cloud bursting:
 - Handle peaks in demand
 - Minimizes job wait times

- Challenges:
 - Variety of VM **types and regions**
 - different pricing options:
 - **On-Demand, Spot, Reserved Instances, Saving Plans, and Capacity Blocks**

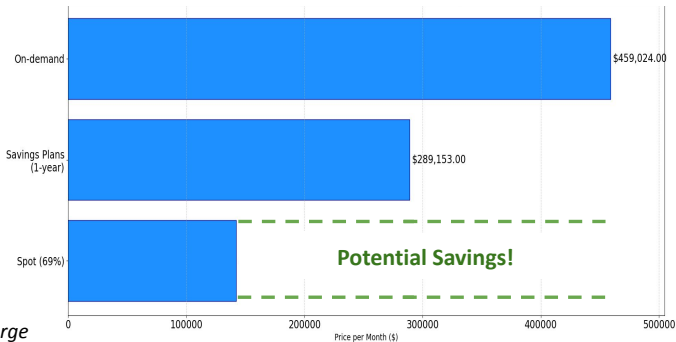


Challenges

- Cloud computing is too expensive
- Cloud bursting is available under different pricing options
On-Demand, Spot, Reserved Instances, Saving Plans, and Capacity Blocks
- Different from an on-premises cluster, users must consider the **cost** in addition to the execution time of their job

Spot VMs

- Potentially the cheapest market option
- Always cheaper than their On-Demand equivalents
- But can be terminated by the provider at any time



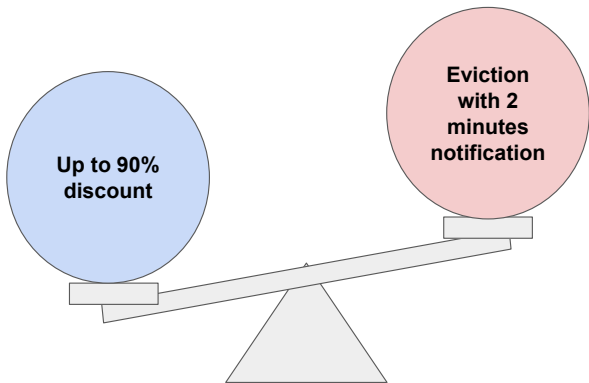
100 X C5.24xlarge

- Computed Optimized
- 96 vCPUs & 192 GiB Memory

Source: <https://calculator.aws> (region sa-east-1)

Spot VMs

- Cheaper than On-Demand
- Offer the same computational power
- Can be terminated by the provider at any time

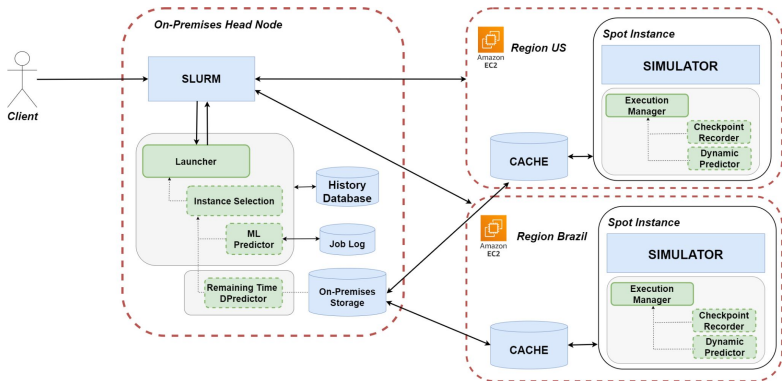


Sim@Cloud

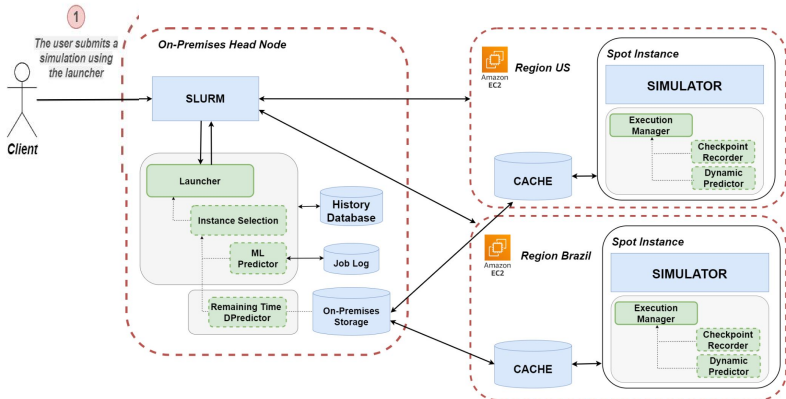
Main goals

- Execution of large-scale industrial and scientific simulations across heterogeneous cloud environments
- **Reduce monetary costs of executing simulations**
- Handle the entire life cycle of simulations running in the cloud
 - responsible for selecting the most appropriate and available VM instance type across different cloud regions and markets (**On-Demand** and **Spot**).

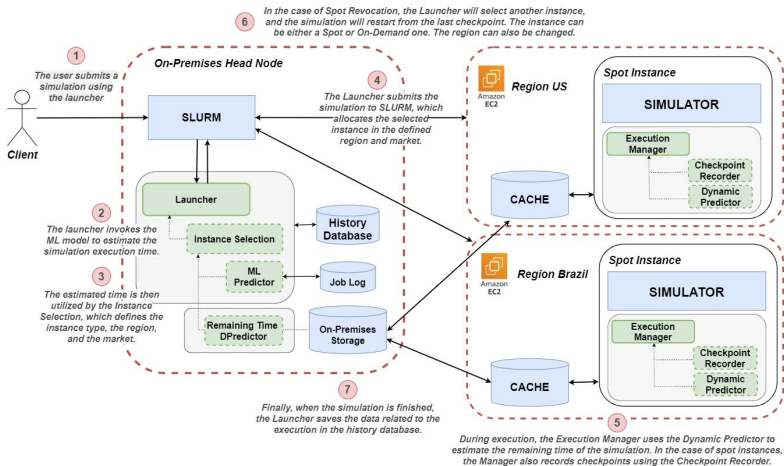
Sim@Cloud Architecture



Sim@Cloud Architecture



Sim@Cloud Architecture



Launcher Instance Selection

Algorithm 1: Makespan prediction, performance, price and I/O overhead based VM Selection

Input: $tsize, restarted, VMREF$

```
1: {*** Step 1: Makespan Estimation ***}  
2: if NOT  $restarted$  then  
3:    $makespan \leftarrow ML\text{-PREDICTOR}()$   
4: else  
5:    $makespan \leftarrow \frac{PF(VMREF)}{PF(vm_{prev})} \times DYNAMIC\text{-PREDICTOR}()$   
6: end if  
7: {*** Step 2: VM Selection ***}  
8:  $vm_{best} \leftarrow VMREF$   
9:  $COST_{best} \leftarrow price(vm_{best}) \times makespan$   
10: for  $vm_i \in VMSET$  do  
11:    $R \leftarrow Region(vm_i)$   
12:    $MAKESPAN' \leftarrow PF(vm_i) \times makespan + (\frac{tsize}{BW_R})$   
13:   if  $vm_i$  is Spot then  
14:      $n\_ckp \leftarrow \lceil PF(vm_i) \times makespan / INTERVAL \rceil - 1$   
15:      $time\_ckp \leftarrow n\_ckp \times (CP\text{-LAT} \times CP\text{-OHD}_R)$   
16:      $MAKESPAN' \leftarrow time\_ckp + MAKESPAN'$   
17:   end if  
18:   if  $COST_{best} > (price(vm_i) \times MAKESPAN')$  then  
19:      $vm_{best} \leftarrow vm_i$   
20:      $COST_{best} \leftarrow price(vm_i) \times MAKESPAN'$   
21:   end if  
22: end for  
23: return  $vm_{best}$ 
```

Get estimated runtime (use ML for new job, dynamic for a previously interrupted one)

For reference, calculate the cost of using the baseline On-demand instance

For each feasible VM, estimate the total execution time considering data transfer delays to the region

For each spot instance, compute the additional checkpoint overhead

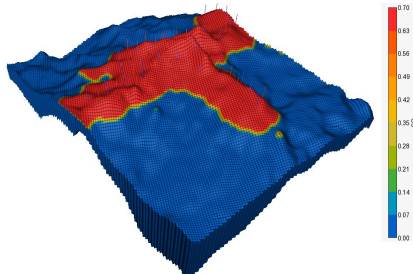
Select the VM with the lowest monetary cost

Evaluation Workload and Environment

- Pre-Salt Semisynthetic Model
 - Black-oil simulator CMG IMEX
 - 50-years simulation based on a real workload
 - Short version (20-years simulation)
- On-premise Storage
 - Rio de Janeiro
 - Data sovereignty requirement

Model	Makespan
Pre-Salt	~ 9h
Short Pre-Salt	~ 3h

Some of the predefined parameters	
VM_{REF}	c5a.24xlarge
VM_{SET}	c5.24xlarge, c5a.24xlarge, c6a.24xlarge, c6i.32xlarge
$INTERVAL$	1800 seconds

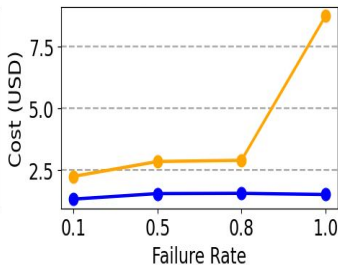
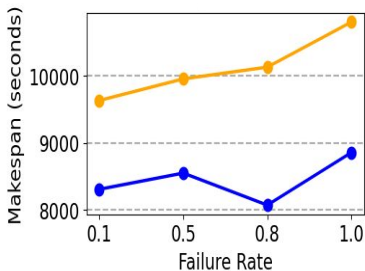


Evaluation Spot market price variability

- Each execution is composed of 4 simulation experiments, each subjected to different revocation (failure) rates
 - the Amazon EC2 Metadata Mock tool was used to replicate the two-minute spot revocation alert
- Observed that Spot instance prices can change significantly from time to time
- Same set of experiments executed on two different weeks
 - results to be presented are an average of three separate executions

Evaluation Spot market price variability

- Execution 1 (typical AWS spot market demand)
 - motivated by their prices at the time, explored several newer generation instances (c6a and c6i)
 - case $\lambda = 0.8$ used a spot c6i at unusually low price
- Execution 2 (unusually high spot market prices)
 - predominant use of older generation instances (c5 and c5a), higher average number of revocations
 - case $\lambda = 1.0$, the elevated average cost was caused by one execution suffering five revocations (max permitted), and terminating on a on-demand c6i inst.
- Variability in prices and availability underscores the need for an instance selection strategy

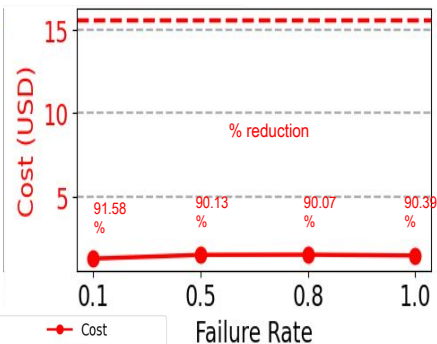
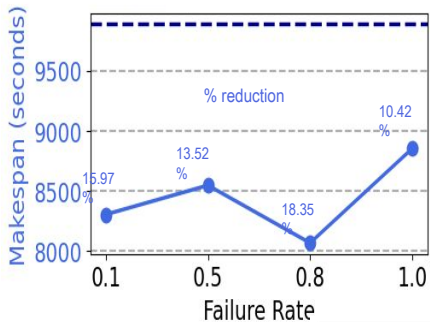


Execution1 Execution2

Evaluation

Reducing cost with Spot instances

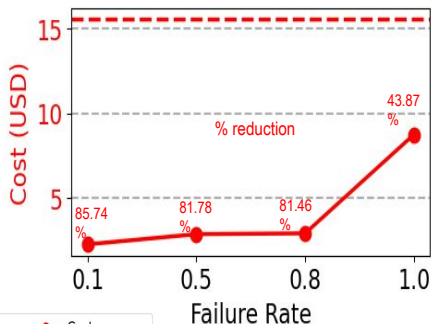
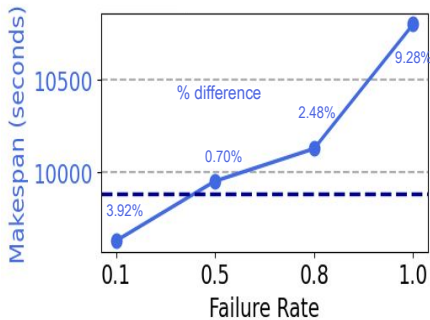
- Short Pre-salt Execution 1 (under typical AWS spot market demand)
 - In relation to the user's expectation (time of 9,883s and cost of \$15.55 on the reference VM)
 - performances are at least 10% better with costs 90% lower, on average



—●— Makespan —●— Cost
- - - Baseline Makespan - - - Baseline Cost

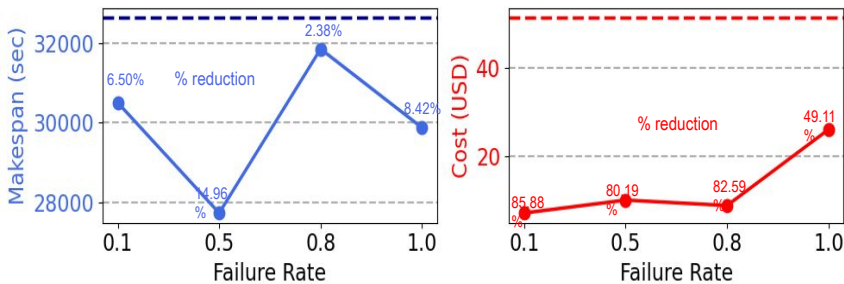
Evaluation Reducing cost with Spot instances

- Short Pre-salt Execution 2 (when spot market prices were unusually high)
 - In relation to the user's expectation (reference on-demand VM)
 - slower executions (up to 10%) may occur but costs are still lower (> 43%), on average



Evaluation Reducing cost with Spot instances

- Pre-salt Execution - long (~9h) simulation
 - executions frequently reach the revocation limit and complete on an on-demand instance
 - but still obtain average **reductions** in both the **makespan** and **monetary cost**



Evaluation Reducing cost with Spot instances

For $\lambda = 1.0$, the execution faced 5 revocations, but execution time similar to the case of $\lambda = 0.1$

- The C6a.24xlarge spot instance was the preferred choice in both executions
 - during the execution period, it offered the best cost-performance
- Reasons the execution uses a variety of instance types:
 - quarantining mechanism avoids reselection of the same instance type after revocation
 - opt to sacrifice the best cost-performance option for one with a smaller likelihood of being revoked

λ	Instantiation	Selected VM	Region	Market	Exec. Time (sec)	ERT_{ref} (sec)	ML Predictor	Cost
0.1	t_0	c6a.24xlarge	SA-East-1	Spot	31,456		36,000 (s)	US\$ 6.93
1.0	t_0	c6a.24xlarge	SA-East-1	Spot	3,718	32,627	36,000 (s)	
	t_1	c5.24xlarge	SA-East-1	Spot	10,921	27,429		
	t_2	c6a.24xlarge	SA-East-1	Spot	263	-		
	t_3	c6i.32xlarge	SA-East-1	Spot	3,718	20,506		
	t_4	c6a.24xlarge	SA-East-1	Spot	263	-		
	t_5	c6i.32xlarge	US-East-1	On-Demand	11,864	-		
	Total					30,747		

λ	Instantiation	Selected VM	Region	Market	Exec. Time (sec)	ERT_{ref} (sec)	ML Predictor	Cost
0.1	t_0	c6a.24xlarge	SA-East-1	Spot	31,456		36,000 (s)	US\$ 6.93
1.0	t_0	c6a.24xlarge	SA-East-1	Spot	3,718	32,627	36,000 (s)	
	t_1	c5.24xlarge	SA-East-1	Spot	10,921	27,429		
	t_2	c6a.24xlarge	SA-East-1	Spot	263	-		
	t_3	c6i.32xlarge	SA-East-1	Spot	3,718	20,506		
	t_4	c6a.24xlarge	SA-East-1	Spot	263	-		
	t_5	c6i.32xlarge	US-East-1	On-Demand	11,864	-		
	Total					30,747		

Conclusions and Future Work

- Sim@Cloud was able to reduce the monetary cost of executions, even in cases involving several Spot revocations
- The framework's decisions are transparent to the user and require no intervention, which makes adopting cloud environments for HPC workloads easier.
- Carry out further evaluations with respect to:
 - different instance types/classes, other geographical regions and cloud provider
 - the performance variability of cloud instances
 - dynamic job rescheduling/migration to meet SLA objectives, for example
- The framework is cloud provider agnostic ...
 - ... but adjustments and experiments are still pending

Published Work

MScheduler: Leveraging Spot Instances for High-Perf the Cloud

IEEE CloudCom 2023

Prediction of Reservoir Simulation Jobs Times Using a Real-World SLURM Log

XXIV Simpósio em Sistemas Computacionais de Alto Desempenho 2023: 49-60

A Framework for Executing Long Simulation Jobs Cheaply in the Cloud

12th IEEE International Conference on Cloud Engineering 2024

High Performance Computing in Clouds: Moving HPC Applications to a Scalable and Cost-Effective Environments

Springer, 2023

Two-Step Estimation Strategy for Predicting Petroleum Reservoir Simulation Jobs Runtime on an HPC Cluster

Wiley, CCPE, 2025

Prediction of Reservoir Simulation Jobs Times

Using a Real-World SLURM Log

Alan L. Nunes¹, Felipe A. Portella^{1*}, Paulo J. B. Estrela¹, Renzo Q. Malini¹, Bruno Lopes², Arthur Bittencourt², Gabriel B. Leite², Gabriela Coutinho², Lúcia Maria de Assumpção Drummond²

MScheduler: Leveraging Spot Instances for High-Performance Reservoir Simulation in the Cloud

Felipe A. Portella^{1*}, Paulo J. B. Estrela¹, Renzo Q. Malini¹, Luan Teylo¹, Josep L. Bernat², Lúcia M. A. Drummond¹

¹Petroleo Brasileiro SA
²Raxosiana Supercomputing
³Instituto de Física
⁴Universidade Federal do Rio de Janeiro

{felipeportella, paulo.estrela, renzo}@petrobras.com.br

Abstract—Petroleum reservoir simulation uses computer models to predict fluid flow in porous media, aiding in oil production. Engineers create numerous simulation models, often requiring significant computational resources, which are not always available within the on-premise infrastructure. Commercial public cloud platforms offer many advantages, such as virtually unlimited scalability and pay-as-you-go pricing. This paper introduces MSCHEDULER, a scheduler framework for reservoir simulations at Petrobras, a Brazilian energy company. It efficiently executes jobs in the public cloud using Virtual Machines (VMs) to reduce costs and job completion times with VM termination. Contributions include a novel methodology for reservoir simulation checkpoint-based scheduling, and an analysis of the strategy cost reduction from Petrosim.

Index Terms—Cloud Computing, Spot Instances, Reservoir Simulation

1. INTRODUCTION

Petroleum reservoir simulation is a subset of reservoir engineering that employs sophisticated computer models to predict fluid flow (typically oil, water, and gas) through porous media [1]. Engineers can make decisions throughout the simulation and effectively forecast future oil production. The process involves designing models of oil fields and performing computationally demanding calculations to make extraction decisions. The reservoir is the business if they do it. Reservoirs. Usually, to refine a technique known as primary objective of RM is model parameters such that of the reservoir replicates production period. This leads to reduced uncertainty. It is computationally demanding because of the same reservoir

Edson Borin · Lúcia Maria A. Drummond · Jean-Luc Gaudiot · Alba Melo · Maicon Melo Alves · Philippe Olivier Alexandre Navaux · Editors

High Performance Computing in Clouds

Moving HPC Applications to a Scalable and Cost-Effective Environment



Concurrency and Computation: Practice and Experience

WILEY

SPECIAL ISSUE PAPER

Two-Step Estimation Strategy for Predicting Petroleum Reservoir Simulation Jobs Runtime on an HPC Cluster

Alan L. Nunes¹, Bernardo Galvão¹, Bruno Lopes¹, Felipe A. Portella¹, José Vitorino¹, Lúcia M. A. Drummond², Luciano Andrade², Miguel de Lencastre², Paulo J. B. Estrela¹, Renzo Q. Malini¹

¹Instituto de Computação, Universidade Federal Fluminense (UFF), Niterói, Brazil | ²Petroleo Brasileiro S.A. (PETROBRAS), Rio de Janeiro, Brazil

Correspondence: Alan L. Nunes (alan@uff.br)

Received: 28 June 2024 | Revised: 25 November 2024 | Accepted: 11 January 2025

© 2025 IEEE. Personal use of this article is permitted, but republication and redistribution in any form is prohibited.