

Using Automated Planning in data centers fault tolerance systems

Douglas Trajano

Master's Degree in Computer Science
Pontifical Catholic University of Rio Grande do Sul - PUCRS
School of Technology, Porto Alegre, Brazil
douglas.trajano@edu.pucrs.br

Abstract

Data centers moves the digital world and incidents can cause financial losses or impact users globally. Develop intelligent systems that can respond and act in this scenario is an area of study of artificial intelligence. With it in mind, we propose an automated planner to act or help engineers with possible failures and events in our data centers. Several possible scenarios will be mapped and an automated planner will be created to take actions in this use case.

Introduction

Nowadays, IT operations are crucial for business continuity, it generally includes redundant or backup components and infrastructure for power supply, data communication connections, environmental controls (e.g. air conditioning, fire suppression) and various security devices.

Data Centers are complex systems encompassing a wide variety of technologies that are constantly evolving. Data Centers house critical computing resources in controlled environments and under centralized management, which enables enterprises to operate around the clock or according to their business needs. (1)

When a failure occurs in one of data center's components, engineers need to perform some tasks. The first one is keep architecture running, so the backup hardware is triggered. Then, a troubleshooting is necessary and possible solutions are applied. In this part we have a lot of opportunities with an automated planner, because troubleshooting is a form of problem solving, and we can treating the issue as a search problem.

Automated planning is interested in domain-independent general approaches to planning. For solving a particular problem, a domain-independent planner takes as input the problem specifications and knowledge about its domain. (2)

The planner will be responsible for finding the actions that needs to be applied in the infrastructure. The goal is to reduce the impact of negative events and component failures. You must also maintain the workload of the appropriate components to avoid future problems. In the Section Technical approach I discuss more about how I will do that and

evaluate whether the results were positive or not. In addition, in Section Project Management I show the schedule that I will follow to execute this proposal.

Technical approach

1. Problem domain

Data centers are dynamic and sensitive to failures. Worker nodes are activated as needed, depending on the workload required for a given time. Safety equipment needs to work together to keep computers in optimal conditions. In my understanding, there are several objects involved in this problem.

- Computer
 - Worker node
 - Storage
 - Master node
- Network
 - Rack switch
 - Core switch
- Environmental control
 - Air conditioning
 - Fire suppression
- Energy
 - UPS (uninterruptible power supply)
 - Generator
 - Power Distribution

2. Planning research

First, a research on how we can model this specific problem in the planning domain is required. The initial idea is to solve the problem with heuristic functions. The planner goal is determined in the beginning of the algorithm execution. In this case, the expected result is a step-by-step well-described way to achieve the goal. For example, the objective is reduce overhead in *Cluster A*; the result will be a text with the instructions to route new deployments to other cluster (i.e. set *cluster A* unavailable to new deployments, move *n* pods to *cluster B*).

3. Results Evaluation

In this domain, we have several objects that need to be supervised, some negative events or failures can have

a workaround solution defined and have no impact for users, otherwise, critical actions like restart a server, replace by another one or open a ticket for repair, can be required.

Project Management

The project's objective is to validate this study with an MVP (minimum viable product). I will work on agile methodology, so each sprint will have a planning session to define the next tasks properly. The deadline for submitting papers is June 28, 2021. Below you can see a table with the expected sprints and the start and end dates.

The project's goal is validate this study with a MVP (minimum viable product). At the end is expected to have a paper to explain the research, a repository with the code. I will work on agile methodology, so each sprint will have a planning session to define the next tasks properly. The deadline for paper submission is June 28, 2021. Below we can see a table with the expected sprints and the start and end dates.

Sprint	Start	End
1	May 24, 2021	May 30, 2021
2	May 31, 2021	June 6, 2021
3	June 7, 2021	June 13, 2021
4	June 14, 2021	June 20, 2021
5	June 21, 2021	June 27, 2021

An estimate of tasks is provided below:

- **Sprint 1 and 2:** Develop domain and problems using PDDL.
- **Sprint 3 and 4:** Develop an automated planner and evaluate the results.
- **Sprint 5:** Revision and paper development.

Conclusion

In this work, we aim to investigate the possibility of using automated planning to help engineers in the crucial task of manage hardware in a data center. We believe that our proposed methods can open new possibilities to further research in how automated planning can be applied in IT infrastructure operations. We expect to achieve satisfying results and get a better understanding of the problem at hand.

References

- Arregoces, Mauricio, and Maurizio Portolani. Data center fundamentals. Cisco Press, 2003.
- Ghallab, Malik, et al. Automated planning : theory and practice. Boston, Elsevier Science, 2004.