

# Overhaul EMCO's data integrity

## Internship Projects/Mentors

Title	Overhaul EMCO's data integrity
Status	CANDIDATE SELECTED
Difficulty	MEDIUM

## Description

The Edge Multi-Cluster Orchestrator (EMCO) is a software framework for intent-based deployment of cloud-native applications to Kubernetes clusters, spanning enterprise data centers, multiple cloud service providers and numerous edge locations. It is architected to be flexible, modular and highly scalable. It is aimed at various verticals, including telecommunication service providers.

The primary task of this internship project is to enhance EMCO in order to achieve a higher degree of data integrity and consistent/predictable operation.

Since EMCO is a software framework architected as multiple microservices that exchange data, the integrity of such data (and the preservation of accurate, up-to-date references, as well as predictable lifecycle transitions) is of the utmost importance.

This work expands on existing efforts done within the project to maintain references between resources (and control which operations are possible at which time), with the ideal end-goal of making it impossible to get EMCO in a non-recoverable state.

## Additional Information

Related issue IDs on GitLab (may not represent the full scope):

- <https://gitlab.com/groups/project-emco/-/epics/4>
- <https://gitlab.com/project-emco/core/emco-base/-/issues/151>
- <https://gitlab.com/project-emco/core/emco-base/-/issues/181>
- <https://gitlab.com/project-emco/core/emco-base/-/issues/98>

## Learning Objectives

- Understand the value of presenting and brainstorming ideas towards common goals
- Experience and learn best practices in version control, open source and the review process, using git and GitLab
- Achieving data integrity and deterministic state machines in a distributed system

## Expected Outcome

The very first task (<1 week) will be to [run the upstream examples](#) and attempt to break the internal state of EMCO by modifying them, making API calls out of order, or providing incorrect parameters.

The student is expected to deliver multiple contributions, as a series of patches, throughout the internship, to resolve or improve the current referential integrity issues observed between microservices, or under certain edge cases (such as when dependent services respond unpredictably).

Depending on the difficulty of improving basic data integrity (reproducible by the API examples), additional integrity-related tasks may be pursued:

- implementing or enhancing rollback/update operations
- clarifying intent vs. deployment-level states in code and in documentation
- if these tasks still prove to not be sufficient, the next step would be working on implementing support for an open-source NoSQL database backend for EMCO, which also relates to data integrity.

Since EMCO is still in its early days, the expected deliverables may also change throughout the internship, but the general umbrella of work items will remain with the intent of addressing data integrity and consistency issues.

Documentation for all implemented features is also a part of the expected outcome. Both user and developer-facing documentation.

The following epic will be used to track the progress:

- <https://gitlab.com/groups/project-emco/-/epics/4>

# Relation to LF Networking

EMCO

## Education Level

The student is required to be enrolled in an undergraduate program (BS or MS).

## Skills

The following skills are desired in the student candidate:

- Linux and the common shells/shell scripting languages.
- Go lang (or similar programming languages as long as comfortable with them).
- Experience with databases and data stores.
- Good understanding of boolean logic and mathematical logical systems.
- Good mental model of software interfaces and RESTful APIs.
- Mindset: self-driven and highly-motivated to troubleshoot, solve problems and learn.

## Future plans

With a solid data integrity as the backbone of EMCO, the project will be able to move faster when it comes to implementing new features, reducing the time it takes to guarantee that the new feature won't break existing data integrity/consistency/lifecycle considerations, since there will be a clearer contract between user intent and the actual state of EMCO.

In the future, other database backends and event streaming technologies may be plugged into EMCO, reusing data integrity considerations and code.

## Preferred Hours and Length of Internship

The preferred length of internship is Part-time (20 hours a week for 24 weeks starting in early-summer and ending in late-fall).

Alternatively, we are also open to Full-time (40 hours a week for 12 weeks starting in early-summer and ending in late-summer)

Working hours are flexible but the student should try to be reachable and responsive throughout 9am-4pm (pacific time) since that's when the team is most active.

## Mentor(s) Names and Contact Info

[Click here to apply](#)

Please read all instructions before applying. Include Resume, proof of school enrollment, and participation permission from school/employer

Igor DC | [contact me](#) | Intel Corporation

The student will be provided with guidance from the mentor above and regular meetings will be set-up.