Towards resilience in federated model training using the FEDn framework

Domain: Data Engineering
Project supervisor: Salman Toor
Email: salman.toor@it.uu.se

Introduction

Federated machine learning has opened new avenues for privacy-preserving data analysis. It is an emerging area of research where most of the current efforts focus on the algorithmic details and communication overhead required to train accurate models. Despite the phenomenal progress in the field, we still lack production-grade federated machine learning frameworks that adhere to fundamental properties such as scalability, resilience, security and performance in geographically distributed settings. To fill this gap, we have designed and developed the FEDn framework. FEDn is an open-source framework dedicated to address federated machine learning challenges at scale. The foundation of the FEDn architecture is based on the map-reduce paradigm, a well-known scalable distributed systems design. The overall architecture consists of three tiers. The first tier consists of geographically distributed clients. The second tier is based on combiners - an intermediate tier responsible for load balancing and robustness. The third tier consists of a single reducer component - responsible for building global models. The FEDn architecture adheres to a number of characteristics important to have a highly scalable geographically federated model training environment. However, there are areas that need further improvements. One of them is the third tier of the framework's architecture. The Reducer component in the architecture lacks a high level of resilience. This can be achieved by having different replication strategies. This project will be an opportunity to understand the requirements of a real-world framework, hands-on experience in architecting and implementing a small-scale solution and design different experiments to systematically evaluate the proposed solution.

Task

Within the scope of the project, the task will be to first understand the federated model training process, deploy FEDn framework, and design a component that enhances the real-time resilience in the third tier of the FEDn architecture.

Required expertise

1. Good understanding of Python programming
2. Basic understanding of Docker containers
3. Basic understanding of distributed computing environment

Relevant courses

- Applied Cloud Computing
- Data Engineering 1
- Data Engineering II

Important links:

- FEDn article: https://arxiv.org/abs/2103.00148
- Git repository: https://github.com/scaleoutsystems/fedn