"Streamer" Community App
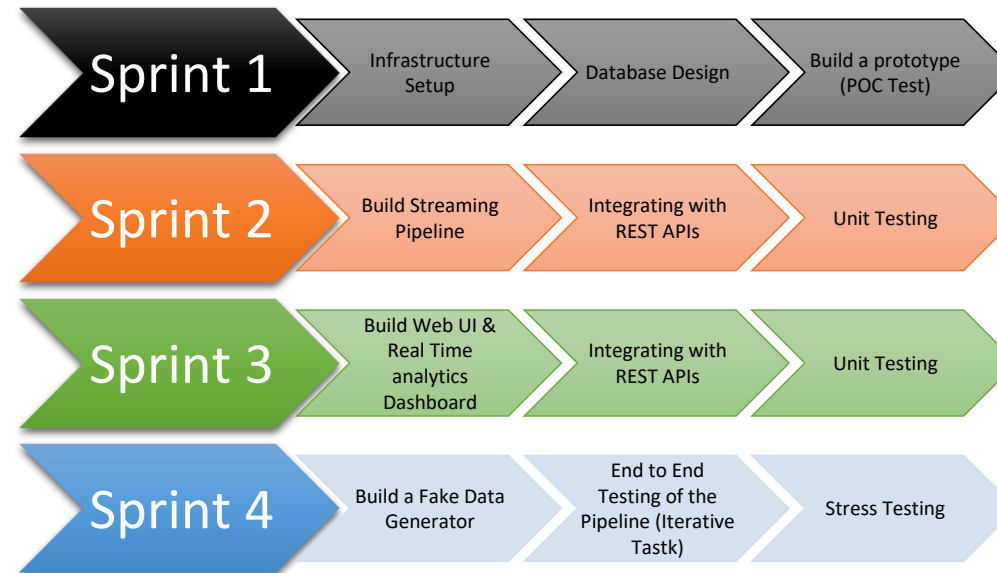
SPA, Assignment Part 1

Rishav Saigal
2019AB04045

Mohd Tauseef Ansari
2019AB04067

SDLC and Agile Methodology Used

| Sprint 1 | Infrastructure Setup | Database Design | Build a prototype (POC Test) |
Sprint 2	Build Streaming Pipeline	Integrating with REST APIs	Unit Testing
Sprint 3	Build Web UI & Real Time analytics Dashboard	Integrating with REST APIs	Unit Testing
Sprint 4	Build a Fake Data Generator	End to End Testing of the Pipeline (Iterative Tastk)	Stress Testing

SDLC method : As we know the application is related to social media, it must pass through a lot of testing and staging to get the perfect and most reliable software to the users. The best SDLC method to use is the Agile Method which launches an application in successive iterations and the design and code get cleaner and cohesive as the iterations proceed forward. Using this method, one can get a working prototype or MVP directly in a short time and then check for bugs and feasibility.

Database Schema
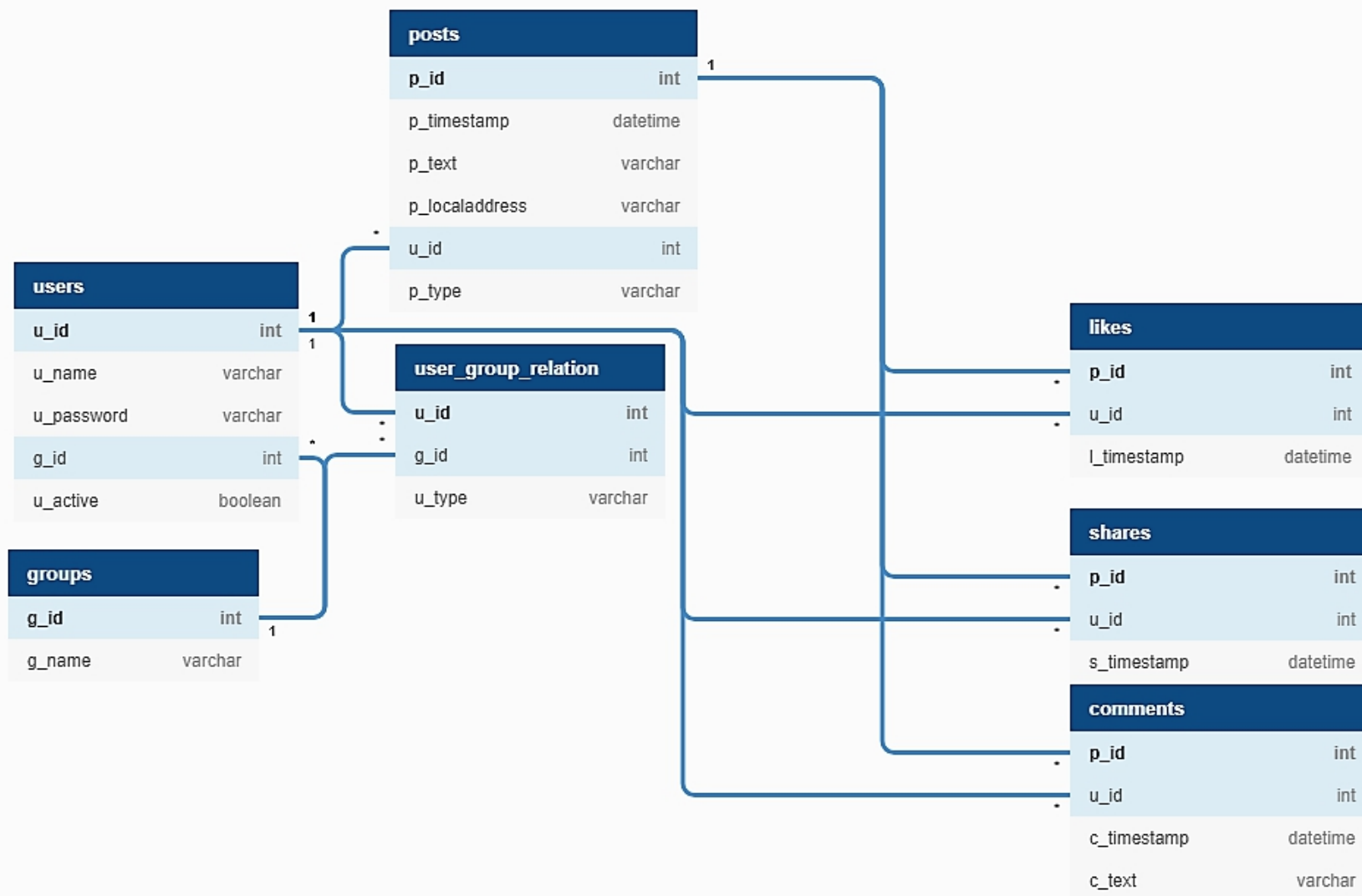
Stack of Technology Used

Frontend

Backend

FastAPI
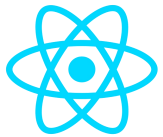
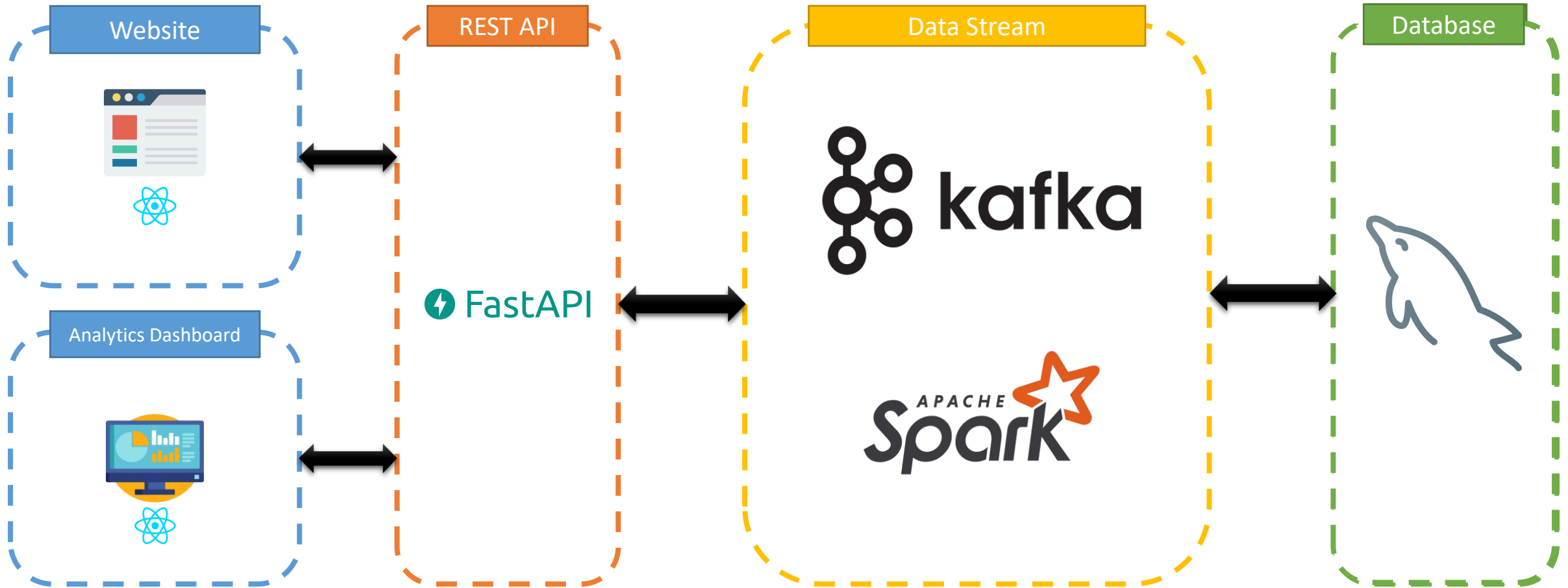**Data Streamer & Real Time Analytics**

APACHE Spark   kafka

Database

MySQL

Social Media platforms receives a lot of load and data on the servers, so it is most important to choose the suitable tech stack. Let's point the best technology that one can use to create the perfect application.

High Level Architecture

Streaming Data-flow Architecture

Web UI

Data Generator

API

⚡ FastAPI

Data Stream

kafka
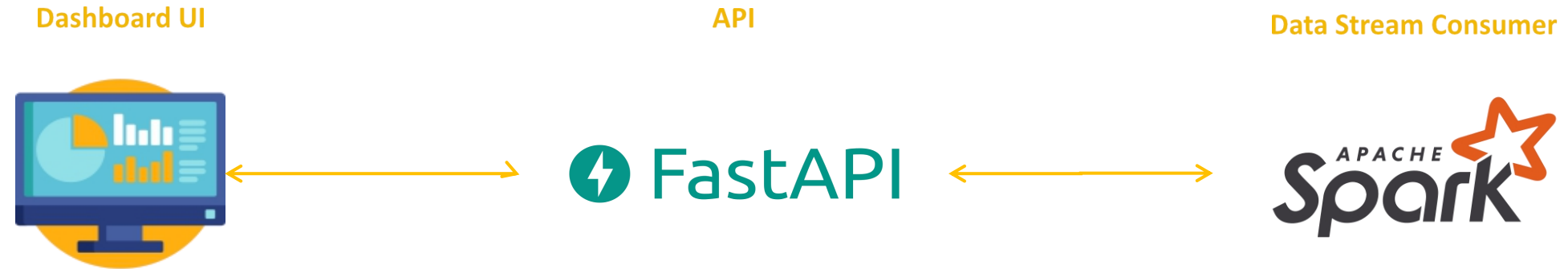
SPARK

MySQL Database

File Storage
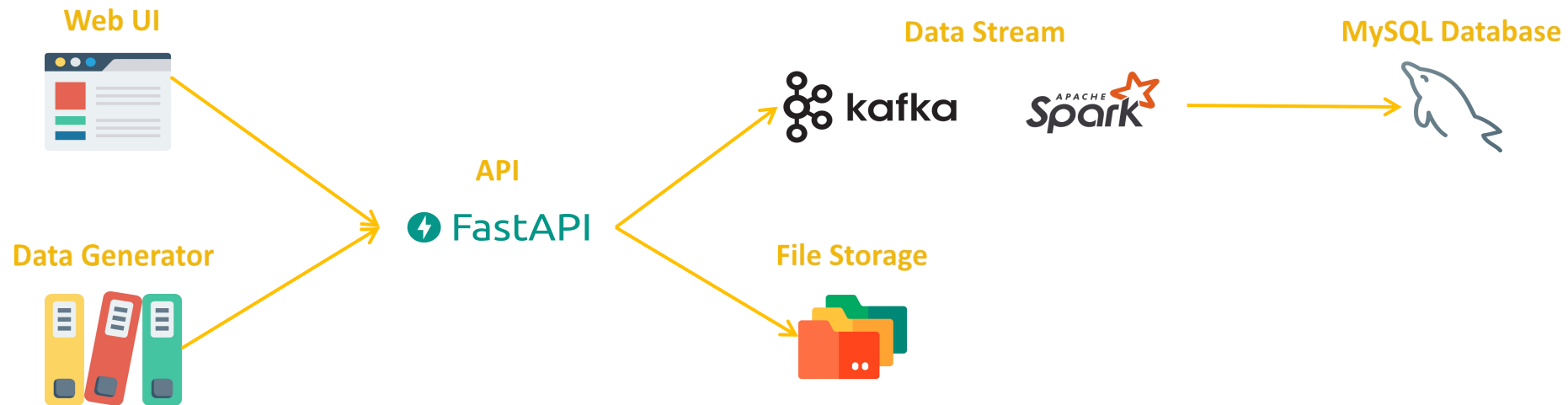
Web UI is user application from where data will be ingested through REST APIs , likewise through fake data generator and that will also be happening through REST API. In Backend REST APIs request and these data items will be passed to Data Stream i.e kafka and pyspark will handle real time data and will do some pre-processing and Analytics. After the processing of the data it will be ingested to MySQL Database

End-User Dataflow

Web UI

API

MySQL Database
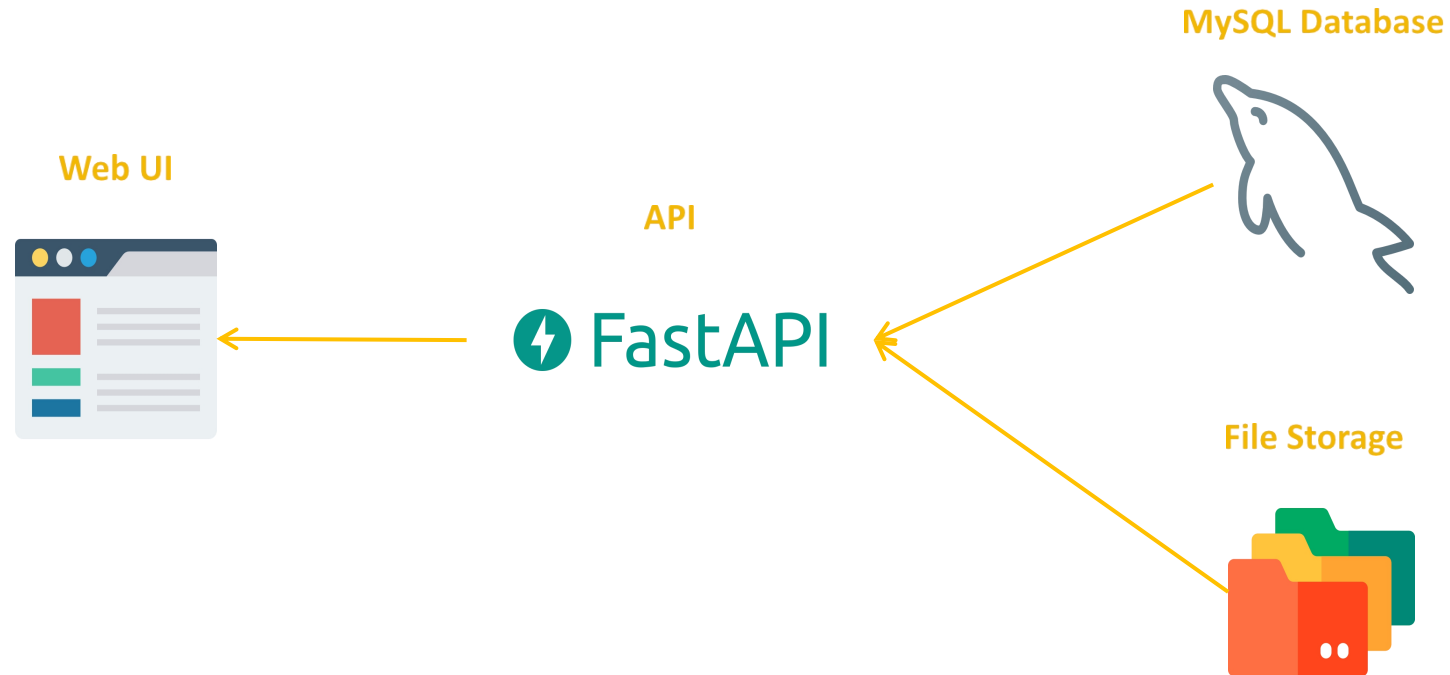
File Storage

This is a simple workflow for viewing data from database, as we don't require any Analytics at user end. We are simply fetching the request through REST API from MySQL Database and File Storage and viewing at Web UI.

Streaming Data Analytics Framework

Dashboard UI

API

Data Stream Consumer

Here comes the interesting part where we are planning to implement basic ML algorithm. In this we are using Data Stream Consumer (Spark) for Analytics and the response through REST API. And from REST API we will be visualizing at Dashboard UI. This process is a to-and-fro.