Data Ingestion Pipeline

Project Deliverable

Course B1AI

Class:

Author: François LANGE

Date: 2026/06/19

Table of Contents

The current version is #ident "@(#)$Format:LocalFoodAI_lanfr144:generate_docs.py:Francois Lange:lanfr144@school.lu:2026/06/16 21:48:22:Francois Lange:lanfr144@school.lu:2026/06/16 21:48:22:2a8ed056889f3b796f9266feda591b12b72f3b96:HEAD -> main, origin/main:$"

Data Ingestion Pipeline

Overview

The application utilizes `data_sync.sh` to update the OpenFoodFacts dataset.

Online Mode

Run `bash data_sync.sh --online`. The script will download the latest CSV directly from the official servers and trigger the ingestion pipeline.

Offline Mode

Drop a `en.openfoodfacts.org.products.csv` file into the `/data` folder and run `bash data_sync.sh`. The script detects the file and triggers the Docker ingestion container.

References

- **OpenFoodFacts Dataset & API Catalog**: Detailed food ingredients database. (https://world.openfoodfacts.org/)
- **Ollama Local LLM Inference Engine**: Lightweight instruction-following llama3.2 runtimes. (https://ollama.com/)
- **Zabbix Enterprise Telemetry and Monitoring**: System health and performance logging. (https://www.zabbix.com/)

Index