

Архитектуре система великих скупова података

КТ1 – Спецификација пројекта

Марко Кубурић

Мастер – НРС

Јануар 2026

Мотивација

- Климатске промене и загађење ваздуха представљају глобални и локални изазов.
- Пораст температуре утиче на:
 - учесталост топлотних таласа,
 - атмосферску стагнацију,
 - акумулацију загађујућих материја.
- Комбинација историјских и актуелних података омогућава:
 - анализу дугорочних трендова,
 - разумевање краткорочних екстремних догађаја.

Циљ пројекта

- Пројектовање и имплементација архитектуре за обраду великих скупова података.
- Интеграција:
 - пакетне (batch) обраде климатских података,
 - стриминг (real-time) обраде података о загађењу.
- Анализа односа климатских услова и квалитета ваздуха.

Домен проблема

- Домен: климатологија и мониторинг квалитета ваздуха.
- Кључне варијабле:
 - температура, падавине, ветар, соларна радијација,
 - концентрације загађујућих материја (PM2.5, PM10, O₃).
- Просторни аспект: глобално, са фокусом на Европу и Београд.
- Временски аспект: дугорочни трендови и реално време.

Batch скуп података 1

Историјске температуре по градовима

- Извор: Berkeley Earth – Climate Change: Earth Surface Temperature.
- Линк: <https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>
- Фајл: GlobalLandTemperaturesByCity.csv.

Каррактеристике Batch скупа 1

- Формат: CSV; величина: ~509 MB; више од 8 милиона редова.
- Временски опсег: 1743–2013.
- Грануларност: месечни подаци по граду.
- Филтрирано на европске градове ($\sim 34^{\circ}\text{N}$ – 72°N , 25°W – 45°E).
- Колоне: датум, просечна температура, несигурност, град, држава, географске координате.

Улога Batch скупа 1

- Анализа дугорочних температурних трендова.
- Идентификација климатских аномалија и сезоналности.
- Пружање историјског контекста климатских промена.

Batch скуп података 2

ERA5-Land реанализа

- Извор: Copernicus Climate Data Store (ECMWF ERA5-Land).
- Тип: месечни агрегати климатских варијабли.
- Формат: NetCDF (.nc).

Каррактеристике Batch скупа 2

- Временски опсег: 1950–2024; величина: \sim 2.5 GB.
- Просторна покривеност: цела Европа (72°N – 34°N , 25°W – 45°E).
- Просторна резолуција: \sim 9 km (\sim 266.000 тачака).
- 12 варијабли:
 - температура на 2 m, температура површине,
 - укупне падавине, испарање,
 - соларна радијација (SSRD, SSR, STR),
 - компоненте ветра (u10, v10), притисак,
 - влажност земљишта, снежни покривач.

Улога Batch скупа 2

- Физички заснован опис климатских услова целе Европе.
- Анализа по латитудним зонама, планинским регионима и градовима.
- Крос-валидација са Berkeley Earth мерењима на 50+ градова.

Streaming скуп података

Квалитет ваздуха у реалном времену

- Извор: AQICN – World Air Quality Index API.
- Линк: <https://aqicn.org/api/>
- Формат: JSON; реално време.

Каррактеристике Streaming скупа

- Подаци по градовима и мерним станицама.
- Учесталост: минутна / часовна.
- Показатељи: AQI, PM2.5, PM10, O₃, NO₂, SO₂.

Повезаност Batch и Streaming података

- Batch: климатски контекст – 75 година ERA5 + 270 година Berkeley Earth.
- Streaming: тренутно стање квалитета ваздуха у Београду (AQICN API).
- Stream-batch join: повезивање преко локације и месечне климатологије.
- Резултати: Elasticsearch + Kibana (стриминг), MongoDB + Metabase (batch).

- Климатолог: анализа дугорочних климатских промена на нивоу Европе – градијент загревања, снежни покривач, суши, енергетски биланс.
- Еколошки аналитичар: праћење загађења и утицаја климе у реалном времену – стагнација ветра, ефекат испирања, сезонске аномалије AQI.

Питања за пакетну обраду (1/3)

- Q1: Како се стопа загревања разликује по латитудним зонама Европе? (Арктичка амплификација)
- Q2: Колика је сагласност ERA5 реанализе и Berkeley Earth станичних мерења за 50+ градова?
- Q3: Како се снежни покривач мења по деценијама – Алпи vs Скандинавија vs Карпати?
- Q4: Колико је суша учесалија у Медитерану у односу на континенталну Европу? (Композитни индекс суше)

Питања за пакетну обраду (2/3)

- Q5: Да ли се расподела падавина по сезонама помера током деценија?
- Q6: Где се у Европи налазе зоне највећег потенцијала ветроенергије? (Atlas ветра)
- Q7: Да ли постоје фазе затамњења (dimming) и посветљивања (brightening) соларне радијације?

Питања за пакетну обраду (3/3)

- Q8: Који европски градови имају најизраженији ефекат урбаног топлотног острва (UHI)?
- Q9: Да ли се учсталост климатских екстрема убрзава? (Топлотни таласи и хладни таласи)
- Q10: Како се површински енергетски биланс (Bowen ratio, испаравање) мења кроз деценије?

Питања за стриминг обраду

- S1: Каква је корелација AQI индекса са ERA5 климатологијом за одговарајући месец? (stream-batch join)
- S2: Да ли стагнација ветра (<2 m/s) корелира са високим AQI? (Детекција алармних стања)
- S3: Да ли пад притиска и висока влажност сигнализирају "washout-- пад AQI након кише?
- S4: Колико је тренутна температура аномална у односу на Berkeley Earth историју 1743–2013?
- S5: Rolling z-score детекција аномалија AQI у прозору од 2 сата.