**Praktični zadatak**

Zadatak je da se implementira strukturirani repozitorijum koji može poslužiti kao jedinstveni izvor podataka za analizu i izvještavanje u sljedećem kontekstu.

*Kompanija je instalirala skup uređaja za monitoring vremenskih prilika i kvaliteta vazduha. Uređaji su instalirani na različitim lokacijama od strane različitih klijenata. Svaka lokacija je unutar oblasti određenog tipa i to: industrial, commercial, residential. Uređaji koji su jednom instalirani ne mogu da mijenjaju lokaciju. Moguće je da uređaji budu izbrisani iz sistema.*

*Dešava se da uređaj izgubi konekciju i tada server ne prima podatke. Uređaji mogu slati podatke duplikate.*

*Poseban mikroservis jednom dnevno eksportuje podatke o uređajima i snima ih kao CSV datoteku u određenu lokaciju odakle ih preuzima vaš sistem.*

*Poseban mikroservis očitava podatke sa uređaja i snima ih kao JSON datoteku u određenu lokaciju odakle ih preuzima vaš sistem. Novi direktorijum se kreira za svaki datum i to sa nazivom oblika received=yyyy-mm-dd.*

Potrebno je osmisliti i implementirati sljedeće Spark pipeline-ove koji se sastoje od moguće više koraka i to:

1. Device information pipeline. Podaci o uređajima u centralnom repozitorijumu se update-uju na način što se formira lista zapisa koji odgovaraju jednom uređaju. Prvi zapis ima polje PARENT=NULL, a kod svakog sljedećeg zapisa ažurira se ovo polje tako da sadrži ID zapisa koji se odnosi na podatke koji se mijenjaju.
2. Device data pipeline. Potrebno je omogućiti istorijski prikaz podataka. Podaci koji su primljeni sa više od 12 sati zakašnjenja upisuju se u posebnu tabelu centralnog repozitorijuma. Podaci koji imaju jednake vrijednosti za kolone *timestamp* i *device* su duplikati. U centralni repozitorijum upisuje se samo posljednje primljeni podatak duplikat, dok se ostali upisuju u posebnu tabelu.
3. Device performance pipeline koji za svaki uređaj prikazuje broj poruka i broj poruka duplikata na dnevnom nivou za određeni period vremena.
4. Report pipeline koji prikazuje prosječne dnevne/mjesečne vrijednosti parametara za svaku lokaciju.

Svaki dataset u centralnom repozitorijumu treba da bude snimljen kao delta tabela (delta.io). Sistem treba da učitava potrebne parametre iz config fajlova.

Zadatak se radi samostano i brani usmeno u janurskom ispitom roku. Obavezni dio rješenja je i tehnička dokumentacija (nekoliko stranica teksta gdje objašnjavate detalje implementacije).