

GraphQL-pohjainen data-alusta IoT-sensoriverkolle

Kuvaus kehitetystä data-alustasta

1. Yleiskuvaus

Hankkeessa rakennettiin GraphQL-rajapinta, joka tarjoaa yhdenmukaisen ja turvallisen pääsyn IoT-sensoriverkkojen mittausdataan opetuskäyttöön.

Alustaa on kehitetty Laatusuoritus etäopiskeluun-hankkeessa, jossa alustalla tuetaan virtuaalista koulutusta ja etälaboratorioita tarjoamalla opetuksen käyttöön todellisista ympäristöistä kerättyä reaaliaikaista mittausdataa. Kehitystyössä on huomioitu etäkoulutuskäytön mukanaan tuomat vaatimukset ja erityistarpeet.

Alusta tukee laajaa kirjoa mitattavia suureita useilta eri aloilta:

- **Ympäristön valvonta:** lämpötila, kosteus, valoisuus, värinä, melu, nesteen pinta, veden laatu, sähkönkulutus, paine, jännite ja yleishavainnot
- **Ilmanlaatu:** hiukkaspitoisuus, pöly ja kaasumittaukset (mm. CO , NO , SO , CO, O)
- **Sää:** tuulen nopeus ja suunta, sademäärä ja sateen tyyppi METAR-koodeineen

Jokainen mittaus sisältää metatiedot: anturin, mittauspisteen, projektin, aikaleiman, yksikön ja tiedon siitä, onko kyseessä raaka vai käsitelty arvo.

2. Datavirta alustalla

Data kulkee alustalla seuraavanlaisesti:

Sensoriverkko. Anturisolmut mittaavat ympäristöstään suureita (esim. lämpötila, kosteus, hiukkaspitoisuus) ja lähettävät mittaukset verkon kautta eteenpäin.

Yhdyskäytävä (GW). Verkon yhdyskäytävä vastaanottaa solmujen lähettämät mittaukset ja välittää ne IoT-alustalle palvelimelle.

Palvelin ja tietokannat. Saapuva data jaetaan kahden tietokannan kesken:

- *Aikasarjatietokanta* tallentaa varsinaiset mittausarvot aikaleimoineen ja mahdollistaa tehokkaan haun mille tahansa ajanjaksolle.
- *Metatietokanta* tallentaa kuvaukset projekteista, antureista, mittauspisteistä, verkoista ja käyttöoikeuksista. Metatiedot pysyvät ennallaan myös silloin, kun anturi vaihdetaan esimerkiksi rikkoontumisen vuoksi.

GraphQL-rajapinta. Rajapintapalvelu yhdistää nämä kaksi lähdettä: asiakkaan kyselyyn palautuu mittausdata, joka on automaattisesti rikastettu asiakuanuuluvilla metatiedoilla. Asiakassovelluksen ei tarvitse tietää, missä tai missä muodossa data on tallennettu.

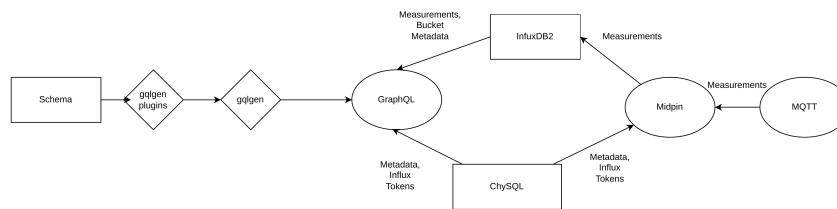


Figure 1: Kaavio datan virtaamisesta

3. GraphQL-rajapinnan toteutus

Rajapinta toteutettiin Go-kielessä skeemalähtöisesti (schema-first): datakuvaus määritellään ensin GraphQL-skeemana, ja palvelinkoodi sekä tyyppimäärittelyt generoidaan skeemasta työkalujen avulla. Skeema on jaettu loogisiin osiin mitausten aihealueen mukaan (sää, ilmanlaatu, ympäristö), mikä helpottaa uusien mittaustyyppien lisäämistä ilman muutoksia olemassa oleviin osiin.

Keskeisiä ratkaisuja:

- **Moniverkkokyselyt.** Yksi kysely voi hakea dataa useasta sensoriverkosta rinnakkaisesti, jolloin koko anturiverkon tilanne saadaan yhdellä pyynnöllä.
- **Joustava ajallinen haku.** Kyselyissä voi käyttää sekä absoluuttisia että suhteellisia aikavälejä (esim. viimeinen tunti).
- **Kyselyjen rajaaminen.** Palvelin rajoittaa kyselyjen monimutkaisuutta ja välimuistittaa toistuvat kyselyt suorituskyvyn varmistamiseksi.

4. Datan haku

Dataa haetaan GraphQL-kyselyillä. Seuraava esimerkki hakee tunnin mittausdatan verkosta, mukaan lukien metatiedot:

```

query GetMeasurements {
  measurements(networkName: ["verkon nimi"], start: "-1h", end: "now()") {
    metadata {
      project {
        name
      }
      sensor {
        id
        name
      }
      poi {
        id
        name
      }
      raw
    }
    data {
      timestamp
      data {
        ... on TemperatureMeasurement {
          value
          unit
        }
      }
    }
  }
}

```

Kyselyitä voidaan suodattaa verkon, mittauspisteen, anturin, klusterin ja ajanjaksolla. Lisäksi saatavilla ovat yhteenvetokyselyt, joilla listataan käytettävissä olevat verkot, anturit, mittauspisteet sekä se, mistä ajanjaksoista dataa löytyy.

5. Käyttöoikeudet ja tietoturva

Pääsy dataan hallitaan käyttöavainten ja klusterien avulla:

- Jokainen käyttäjä tai järjestelmä saa käyttöavaimen, jolle on myönnetty luku- ja/tai kirjoitusoikeudet tietyihin klustereihin. Klusterit ryhmittelevät mittauspisteitä.
- Alusta valvoo oikeuksia automaattisesti: vaikka kyselyssä viitattaisiin rajattuun dataan, palautetaan vain siihen sallittu osuus.
- Kirjautuminen tapahtuu ulkoisella tunnistautumispalvelulla, jonka jälkeen pyynnöt tunnistetaan avaimella. Tuntematon käyttäjä ei näe rajapinnan kuvausta eikä dataa.
- Eri organisaatiot, osastot tai projektit, jotka jakavat samaa alustaa, näkevät vain oman datansa.

Näin sama alusta palvelee turvallisesti useampaa tahoa samanaikaisesti.

6. Käyttöönotto

Palvelu on konttitoimitettu (Docker/Podman), joten se on helppo ottaa käyttöön erilaisissa ympäristöissä. Rajapinnan mukana toimitetaan selainpohjainen kehittäjätyökalu (GraphQL playground), jolla kyselyjä voi kokeilla interaktiivisesti ja tutustua rajapinnan itsekuvaavaan dokumentaatioon.

7. Midpin – mittausdatan keruu ja tallennus

Alustaan kuuluu myös Midpin-middleware-palvelu (Go-kielellä toteutettu sovel-luspaketti), joka toteuttaa luvussa 2 kuvatun datavirran vastaanotto- ja tallennusosan.

Palvelu toimii seuraavasti:

- **MQTT-kuuntelu ja jäsenitys.** Midpin yhdistyy MQTT-brokeriin TLS-salatulla yhteydellä ja tilaa aiheet, jotka on määritetty metatietokantaan. Saapuvat viestit jäsentää aiheeseen liitetty ajuri (esim. LoRa, Home Assistant, RuuviTag, Sensative Strip, IOLINK). Jokainen ajuri on sidottu tiettyyn MQTT-aiheeseen, ja uusi laitetyyppi lisätään kirjoittamalla sille oma ajurinsa sekä liittämällä se aiheiden ja ajurien väliseen hakutaulukkoon. Aiheet päivitetään automaattisesti määrääjoin.
- **Datan rikastus ja tallennus.** Jäsentämisen jälkeen data etenee dekorointiketjussa: ensin haetaan metatietokannasta sensoriin ja mit-tauspisteeseen liittyvät lisätiedot, jonka jälkeen mittausarvot muotoillaan InfluxDB-kirjoituspisteiksi. Dekorointiketjun lopuksi tallennuslogiikka kirjoittaa mittausarvon InfluxDB:hen projektikohtaisesti bucketiin ja tallentaa anturin viimeisimmän arvon ja aikaleiman metatietokantaan. GraphQL-rajapinta hyödyntää näitä tietoja kyselyissään.

8. Yhteenveto

Hankkeessa toteutettu data-alusta yhdistää hajallaisten sensoriverkkojen mit-tausdatan ja metatiedot yhdeksi, turvalliseksi ja joustavasti kyseltäväksi kokon-aisuudeksi. Data kulkee sensoriverkosta yhdyskäytävän kautta palvelimelle ja tietokantoihin, joista GraphQL- rajapinta palvelee sitä asiakassovelluksia ja analytiikkaa varten. Skeemalähtöinen toteutus tekee uusien mittaustyyppien lisäämisestä suoraviivaista, ja klusteripohjainen oikeudenhallinta mahdollistaa alustan jakamisen usean organisaation kesken.

Data-alustaa voidaan hyödyntää myös virtuaalisessa koulutuksessa, jossa opiskelijat voivat harjoitella ja analysoida reaaliaikaista dataa eri toim-intaympäristöistä. Avoimen datan tarjoaminen koulutusorganisaatioille ja tutkimusryhmille laajentaa alustan hyötyjä projektiorganisaation ulkopuolelle.