

Antti Jaatinen

Työkoneiden datankeruun käytännön ongelmia

Tiivistelmä: Työkoneista anturein mitattua dataa hyödyntävät koneoppimis- ja analytiikkasovellukset kohtaavat usein vaikeuksia datan laadun takia. Tässä esitelmässä kerron kokemuksia tästä ongelmasta tiedonkeruun näkökulmasta ja esitän joitain ehdotuksia tilanteen parantamiseksi.

Esitelmä perustuu henkilökohtaisiin kokemuksiin aiheesta Metso Outotecin automaatio- ja digitalisaatiotyössä. Esitän myös joitain parannusehdotuksia suunnittelijoille.

Avainsanat: datan laatu, asioiden internet, reunalaskenta, etävalvonta

Antti Jaatinen: Valmet Automation, email: antti.jaatinen@valmet.com

1 Johdanto

Tietojärjestelmästä saatu tapahtumadata on kovin erilaista verrattuna ”todellisesta maailmasta” mittaamalla saatuun dataan nähden. Jälkimmäinen voi olla monella tapaa värjättyä tai virheellistä. Ilmiö on tunnettu mittaus- ja säätötekniikan alalla jo vuosikausia, mutta ei välttämättä ole kaikille datankeruun ja analytiikan parissa toimiville tuttua. Lisäksi massatuotetussa koneautomaatiossa joudutaan tekemään kompromisseja verrattuna teollisiin automaatiojärjestelmiin, saati sitten puhtaisiin mittausjärjestelmiin. Mittausjärjestelmän rakentaminen ja ylläpito ovat aina haastavia tehtäviä, ja työkoneiden ympäristössä se saattaa jopa olla mahdotonta anturien altistuessa vaikeille ympäristöolosuhteille.

Metso Outotecin digitalisaatiotyössä tuotiin Metrics-alustaan useita uusia laitetyppejä. Näille tarvittiin uusia konnektiviteettiratkaisuja. Tässä työssä samat haasteet tulivat eteen uudestaan ja uudestaan, joten sisäisiä dokumentteja konnektiviteetin vaatimuksista. Suurimmat huolenaiheet olivat tietoturva ja datan laatu.

Data-alkemian harjoittajilla saattaa olla ylioptimistisia odotuksia datan laadun (lähinnä tarkkuuden ja

tiheyden) suhteen. Tällaiset asiat korjaantuvat helpoiten ekskursiolla lähelle dataa tuottavaa laitetta, sillä todellisuus datan alkuperästä hämärtyy helposti pelkkiä numeroita tai käyriä katsoessa.

2 Datan keruun ongelmalähteitä työkoneissa

Anturista aloittaessa, sen ominaisuudet harvoin välittyvät ns. pilveen asti. Anturin luotettava toiminta-alue, herkkyys ja tarkkuus ovat kyllä jossain tiedossa, mutta niiden vaikutusta dataan tuskin mietitään. Eikä ole harvinaista, että anturin tyyppi vaihtuu (halvempaan) sarjatuotannon aikana ilman, että muutoksesta tiedotetaan myös datan käyttäjille.

Anturin ominaisuuksien lisäksi mittausjärjestelmän (tässä tapauksessa yleensä koneautomaation) käyttöönotossa on virheen mahdollisuuksia. Näitä ovat anturin kalibrointi ja skaalaus. Miten varmistetaan niiden oikeellisuus? Tallentuvatko näiden parametrien muutokset niin, että ennen ja jälkeen uudelleenkalibrointia saadut mittaustulokset ovat vertailukelpoisia?

Automaatiojärjestelmien kaapeissa olevat luurangot paljastuvat armottomasti, kun ne kytketään tiedonkeruujärjestelmään. Näitä ovat esimerkiksi mittausalueen ylä- ja alarajojen käyttö erikoistarkoituksiin, kuten vaikkapa mittausalueen alarajan käyttäminen vikasignaalina. Tällöin ei tietenkään voida tietää, onko joku arvo todella nolla vai onko anturiin tullut vika.

Datatyypit osataan ottaa huomioon varsin hyvin, mutta datan kvantisoitumista ei niinkään paljon. Tämä siis tarkoittaa sitä, kuinka isoa mittausarvon muutosta yhden bitin muutos tarkoittaa. Anturin liian iso toiminta-alue verrattuna mitattavaan alueeseen voi aiheuttaa saman ongelman, ja yhdistetty vaikutus on tuhoisa.

Väylät eivät yleensä ottaen vääristä dataa, mutta voivat aiheuttaa toisenlaisia ongelmia, jotka aiheutuvat liian tiheän datankeruun tavoittelusta. Ensiksi ei ole varmaa, että luettavan väylän tila on atominen, eli että kaikki luetut datapisteet kuvaavat samaa ajanhetkeä.

Järjestelmä päivittää niitä usein syklisesti ja vastaavasti tiedonkeruu kerää niitä toisella syklillä. Toisekseen, jos väylä perustuu kyselyihin, ne kuormittavat koneautomaatiota ja voivat hidastaa sen toimintaa ja pahentaa ensiksi mainittua ongelmaa.

3 Edge-laskennalle aiheutuvia vaatimuksia

Käytän tässä nimitystä ”edge-laite” koneautomaatioon kytketystä tiedonkeruu- ja käsittelylaitteesta. Vastaavasta laitteesta käytetään usein nimitystä ”gateway”, mutta koska termillä on tarkka merkitys tietoliikennetekniikassa, haluan käyttää eri termiä. Tiedonkeruujärjestelmä on samoin liian suppea termi tällaiselle laitteelle. Edge-laite pystyy kompensoimaan edellä mainittuja ongelmia. Automaation ja edgen kuuluisi olla työpari.

Käytännössä automaatiojärjestelmän päivitykset ovat vaikea asia, jota ei välttämättä voi tehdä etänä, kun taas edge-laite on nykyaikaisempi, jatkuvasti verkkoon yhteydessä oleva laite. Tämä johtaa siihen, että edgeen kehitetään erilaisia virityksiä kompensoimaan koneautomaation ominaisuuksia. Näin menetellessä aiheutuu kuitenkin painetta edgen versioiden ja varianttien hallinnalle. Muistetaanko koneautomaatiota päivittäessä vaihtaa edgenkin konfiguraatio? Tämä ei todennäköisesti ratkaise ongelmaa, vaan siirtää sen edgen konfiguraatioiden versionhallintaan, koska tarvittava edgen softaversio tulee riippuvaiseksi automaation versiosta.

Tietoliikenne ja datan tallennus ei ole ilmaista, joten edge-laitteissa tavallisesti on hystereesiä datan lähetyspuolella, eli datapistettä ei lähetetä, ellei se ole muuttunut yli tietyn rajan. Kustannukset yhden tavun lähettämiseksi eivät ole suuret, mutta kumulatiivinen kustannus datan ja koneiden määrän kasvaessa alkaa jo vaikuttaa. Datan säilyttäminen verkossa ei myöskään ole ilmaista, ja siirto edullisemmasta pitkäaikaisäilöstä laskentaa varten työmuistiin on lisäkustannus.

Tietoturva on syytä muistaa, että koneautomaatiota ei luultavasti ole suunniteltu kytkettäväksi verkkoon. Tietoturva on luultavasti heikolla tasolla. Kannattaa teettää auditointi ja olla tietoinen ongelmista. On myös hyvä ymmärtää, että automaatiojärjestelmä, edge-laite ja tiedonsiirtoratkaisu käyttävät kaikki samaa TCP/IP protokollaa, ja hyökkäys etäyhteyden kautta on mahdollinen, vaikkakin epätodennäköinen.

4 Ratkaisuja

Ratkaisuja näihin ongelmiin toki löytyy. Idealistisesti ajatellen koneautomaation pitäisi tukea nykyaikaista

tiedonkeruuta ja olla tietoturvallisesti kytkettävissä edge-laitteen kautta verkkoon. Käytännössä tämä kasvattaa projektien laajuutta liian paljon, mutta projektit tulisi jakaa uusien ja vanhojen järjestelmien osalta niin, että näiden kehitystä voidaan vetää erikseen. Tiedonkeruun ja automaation tulee toimia yhteen, ja projekteissa on varauduttava luurankojen paljastumiseen, eikä ongelmia saa piilotella. Tämä on yrityksen toimintakulttuurin mittari.

Edge-laitteen vaatimuslista on jo pitkä, mutta sen tulisi kyetä parantamaan koneautomaatiosta tulevaa dataa ja suorittamaan laskentaa sille. Missään tapauksessa tuotantojärjestelmästä ns. raakaa dataa ei tulisi lähettää pilveen sellaisenaan.

Edge-laitteen tulee kyetä puskuroimaan ja käsittelemään dataa. Datan siivous ja prosessointi tulisi tehdä niin lähellä laitetta kuin mahdollista, ”jalostusasteen” kasvaessa pilveä kohti mentäessä.

Edge-laitteen tietoturva ei ole tässä esitelmässä juurikaan sivuttu, mutta aihe on laaja, ja kannattaa tarkkaan miettiä omia kykyjä pitää laitekanta päivitettyinä ja turvallisena.

4.1 Ratkaisuja automaatiojärjestelmän suunnasta

Mittaustekniikan perusteiden kertaus. Anturien oikea valinta, skaalaus ja kalibrointi.

Tietojen käsittely insinööriyksiköissä, ei koskaan jännitteinä, virtoina tms.

Virhetilanteet on voitava viestittää muuta kanavaa pitkin kuin mittausarvot.

Automaation rajapintojen määrittelyt on otettava osaksi muutostenhallintaprosesseja ja tuoterakenteita. Ne on vakioitava niin hyvin kuin mahdollista.

Oleta, että ”konnektiviteetti” todella tarkoittaa automaation liittämistä verkkoon ja altistamista sen vaaroille.

4.2 Ratkaisuja datan suunnasta

Jokainen pilveen lähetettävä datapiste olisi oltava perusteltavissa.

Pohdi, mitä ilmiöitä haluat mitata, ja perusta tiedonkeruun tiheys ilmiön nopeuteen.

Lähetä vain muuttunut data, ja suunnittele järjestelmäsi niin, että tästä ei aiheudu laskennoille

ongelmia.

Älä yritä kompensoida automaation ongelmia edgen konfiguroinneilla. Ongelma siirtyy versioiden ja riippuvuuksien hallintaan, kun automaatio aikanaan korjataan.

Datan siivous, käsittely, vikatilanteiden tunnistus ja muu valmistelu Edgessä on silti suositeltavaa.

4.3 Ratkaisuja organisaation suunnasta

Konnektiviteettiprojektit eli laitteen liittäminen osaksi jotain palvelua eivät ole yksinkertaisia projekteja. Ne vaativat usean eri organisaation osallistumista kokonaisen ratkaisun saavuttamiseksi. Jokaisella osallistuvalla ryhmällä saattaa olla motiivi osoptimoida omaa tekemistään, joten jonkun on katsottava, että kokonaisuus toimii. Hänelle on annettava myös resurssit ja valta tähän.

Koneautomaation kohdistuu jo nyt suuria vaatimuksia ja ne ovat monimutkaisia reaaliaikaisia ja turvallisuuteen liittyviä järjestelmiä. Niiden laitteisto on myös rajoittuneempaa ympäristön takia.

Järjestelmien liittäminen toisiinsa paljastaa haasteet armottomasti. Tähän ei pidä suhtautua liian vakavasti eikä aloittaa syyttelyä. Organisaation kypsyyden merkki on tällaisesta haasteesta selviäminen.

Liitettävyyys on kertaluokkaa helpompi tehdä uusiin laitteisiin kuin jälkiasentaa (retrofit). Vältä jälkiasennuksia, jos mahdollista, tai erota ne erillisiksi projekteiksi.

Älä kuuntele "data scientisteja" liikaa. Mikään työkoneesta saatavan datan laatu ei luultavasti ole tarpeeksi hyvää. Yritä löytää kompromissiratkaisu esimerkiksi keräämällä enemmän dataa tutkimuskäyttöön vain muutamasta koneesta eikä kaikista. Onnistuneet algoritmit viedään laajempaan käyttöön ja toteutetaan Edge-laskennassa.