

VALSTYBĖS DUOMENŲ VALDYSENOS INFORMACINĖS SISTEMOS STANDARTINĖS DUOMENŲ DALIJIMOSI JUNGTIES APRAŠYMAS

1. Duomenų dalijimosi jungties bendravimas su Lietuvos statistikos departamento Duomenų valdymo platformos posistemiū (*Palantir Foundry*):

1.1. Duomenų dalijimosi jungtis veikia *Linux* virtualioje mašinoje Institucijos tinkle.

1.2. Komunikacija tarp jungties ir duomenų šaltinių (duomenų bazių, failų talpyklų ir pan.) vyksta Institucijos tinkle.

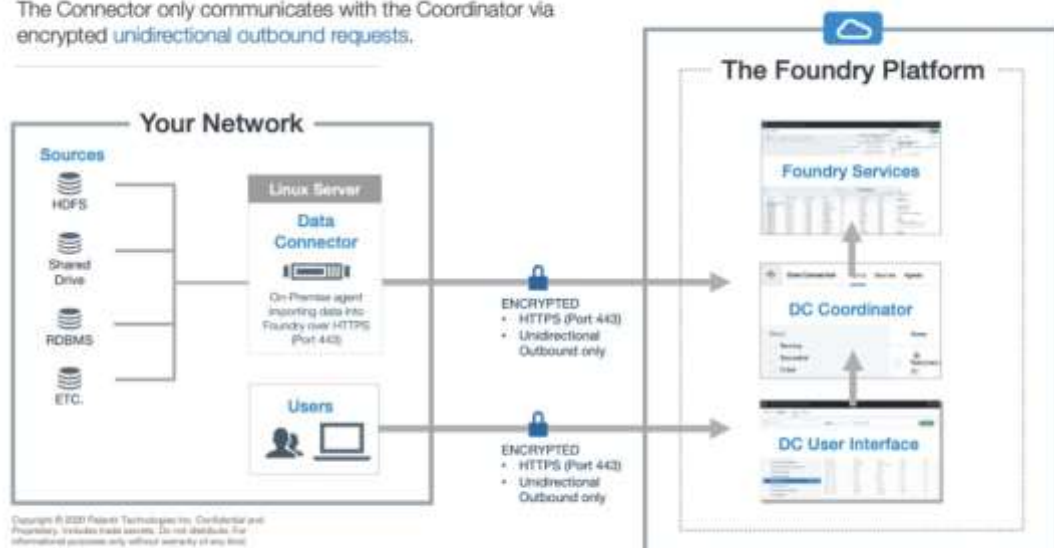
1.3. Jungtis turi turėti tiesioginį ryšį su duomenų šaltinio serveriu.

1.4. Komunikacija tarp jungties ir VDV IS vyksta šifruotomis išeinančiomis HTTPS užklausomis per 443 prievadą, t. y. jungtis inicijuoja šifruotą ryšį su Duomenų valdymo platforma (toliau – Platforma). Pati Platforma negali inicijuoti ryšio su jungtimi.

1.5. Tam, kad jungtis galėtų sėkmingai prisijungti prie Platformos, ugniasienės turi praleisti ryšį į `vdv.stat.gov.lt` (IP: 52.214165.165 ir 54.229.28.121) per TCP/443 prievadą.

Easily upload data from your network to your Foundry

The Connector only communicates with the Coordinator via encrypted unidirectional outbound requests.



2. Institucijos serverio techniniai reikalavimai. Duomenų dalijimosi jungtis gali veikti fiziniame arba virtualiame serveryje Institucijos tinkle. Siekiant užtikrinti greitaveiką, esant dideliems perduodamų duomenų kiekiams, rekomenduojama palaikyti šiuos reikalavimus:

2.1. CPU – 4 branduoliai;

2.2. Operatyvioji atmintis – 16 GB;

2.3. Disko vieta – 100 GB.

3. Šie reikalavimai itin priklauso nuo duomenų kiekio ir šaltinio tipo.

4. Operacinės sistemos konfigūracija:

4.1. Institucijos serveris, kuriame bus diegiama duomenų dalijimosi jungtis, turi naudoti *Linux* operacinę sistemą. Rekomenduojama distribucija – *CentOS 7*.

4.2. Operacinėje sistemoje reikia:

4.2.1. sukurti vartotoją ir grupę. Ir vartotojo, ir grupės pavadinimas gali būti „palantir“;

4.2.2. suteikti vartotojui teises /opt kataloge.

Pavyzdys: `chown -R palantir:palantir /opt`.

5. Duomenų talpyklų tipai: įdiegta duomenų dalijimosi jungtis gali prisijungti ir pasiekti duomenis iš įvairių duomenų talpyklų. 8 punkte yra išvardytos dažniausiai pasitaikančios ir naudojamos talpyklos. Prireikus prijungti jungtį prie talpyklos, kuri nėra pateikta žemiau, prašome susisiekti su VDV IS administratoriais el. paštu vdv-prisijungimas@stat.gov.lt.

6. Duomenų bazės. Duomenų dalijimosi jungtis gali prisijungti prie įvairių duomenų bazių: *Oracle, PostgreSQL, MariaDB, MySQL, Microsoft SQL Server*, kitų (pagal poreikį).

7. Tam, kad jungtis galėtų prisijungti, reikalingas duomenų bazės vartotojas, turintis skaitymo (angl. *read only*) teises į sutartyje nurodytus duomenų rinkinius. Jungtis jungsis prie duomenų bazės iš serverio, kuris yra įdiegtas Institucijos vidiniame tinkle. Vidinis jungties serverio IP adresas turėtų būti praleistas ugniasienėje ir duomenų bazės konfigūracijoje.

8. Rinkmenų talpyklos. Rinkmenų ir katalogų pavidalu saugomi duomenys dažnai yra pasiekiami per tokius serverius kaip SFTP/FTP, HDFS.

9. Norint pasiekti šiuos duomenis, rinkmenų talpyklos serveryje turi būti sukurtas vartotojas, kuris turėtų skaitymo (angl. *read*) teises kataloge, kuriame yra rinkmenos su reikalingais duomenimis. Kaip ir duomenų bazės atveju, vidinis jungties serverio IP adresas turėtų būti praleistas ugniasienėje ir rinkmenų talpyklos konfigūracijoje.

10. Talpykla vidiniame jungties serveryje:

10.1. yra galimybė duomenis paimti iš serverio, kuriame yra įdiegta jungtis. Tokiu atveju jungtis nesijungtų prie Institucijos duomenų talpyklų. Susijungimas vyktų ne aplikacijos, o infrastruktūros lygmeniu;

10.2. vidinis serveris, kuriame yra įdiegta jungtis, gali prijungti (angl. *mount*) bendrinamą (angl. *shared*) failų sistemą, naudodamasis NFS ar SMB/CIFS protokolais. Tokiu atveju jungtis matytų rinkmenas lokaliai ir galėtų jas pasiekti;

10.3. taip pat į jungties serverį duomenys gali būti persiunčiami FTP/SFTP ar kitais protokolais ar būdais į katalogą, iš kurio jungtis lokaliai galėtų pasiekti duomenis. Tokiu atveju pati jungtis nesijungtų prie duomenų šaltinių, tačiau Institucija turi užtikrinti, kad duomenys atsinaujintų reguliariai, be trikdžių ir tokiu pačiu formatu bei schema.

11. Duomenų sinchronizavimas:

11.1. Dažnumas. Jungtis ima duomenis reguliariai ir nustatytu dažnumu. Dažnumas gali būti konfigūruojamas minučių, valandų, dienų, mėnesių tikslumu. Poreikiui esant, galimas ir sudėtingesnis konfigūravimas.

11.2. Metodai. Galimi šie duomenų paėmimo būdai:

11.2.1. momentinis (angl. *snapshots*) – paimami visi duomenys, kurie yra pasiekiami paėmimo momentu;

11.2.2. duomenų papildymas (angl. *append*) – paimami tik nauji duomenys, kurie buvo atnaujinti po paskutinio paėmimo;

11.2.3. duomenų atnaujinimas (angl. *update*) – paimami tik nauji duomenys. Jei atnaujinto failo pavadinimas sutampa su anksčiau paimto failo pavadinimu, senasis failas yra pakeičiamas nauju. Šis būdas dažniausiai naudojamas su rinkmenų talpyklos šaltiniais.
