



Mitä on datatyö

Mimmit Koodaa -webinaari
Sanna Uusimäki

5.5.2023



Ketä te kuuntelette?

Tarina on miten sen kertoo..

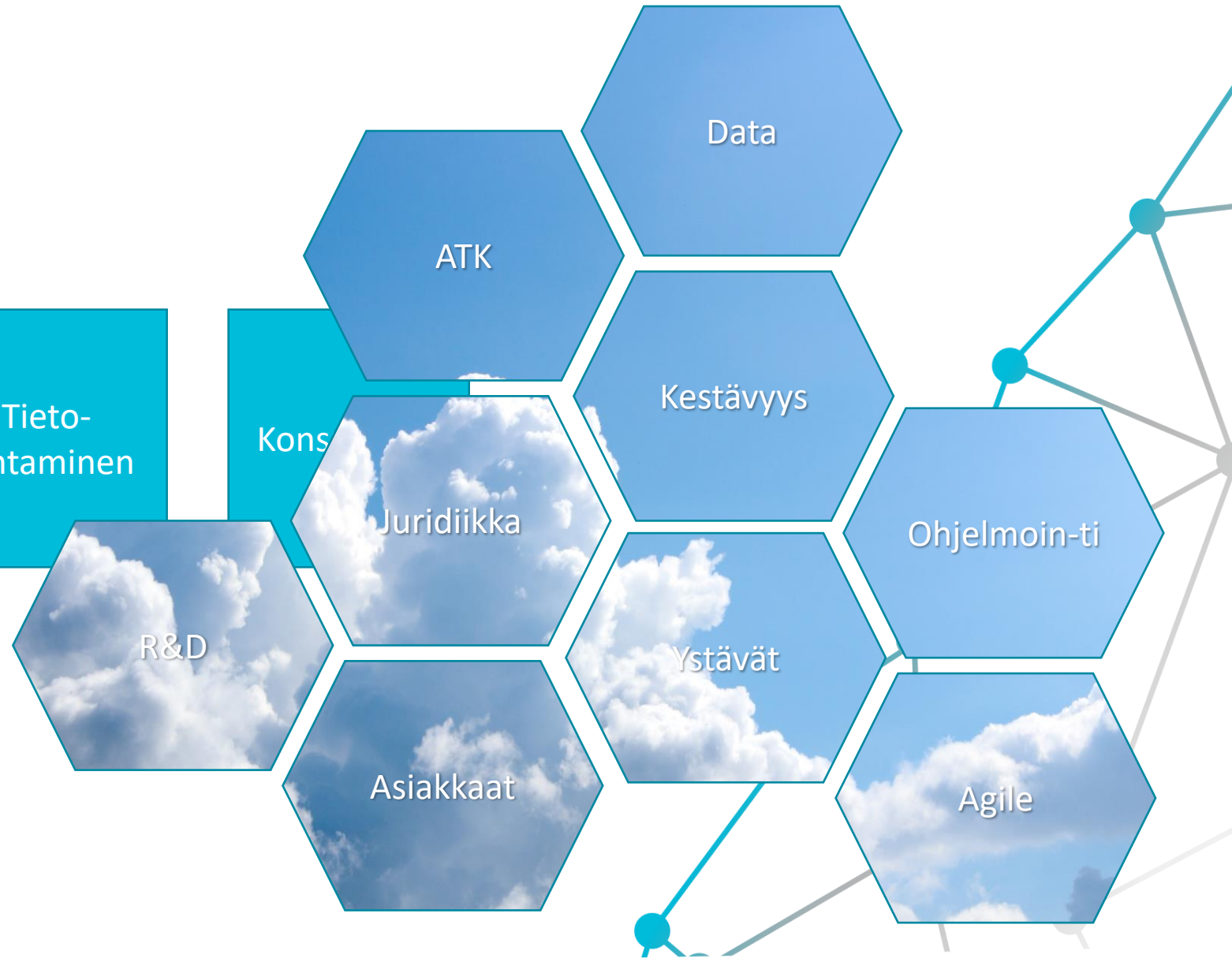
Lukio

Ohjelmisto-
tekniikka

Tieto-
johtaminen

Kons

Mutta on totta, että ensimmäinen
työtehtäväni datan parissa oli jo
kandityöni 16 vuotta sitten.





Sanna Uusimäki

Data Solution Architect – Research and Development Lead



Sanna is a data solution architect at Cloud1. She has worked as data professional since 2009 in the roles that have been both hands on and leading roles. She has done report implementation, source system analysis, data modelling, data ware housing. She is now responsible of leading data development and research internally at Cloud1.

Email

Sanna.uusimaki@cloud1.fi

Tel.

+358(0)505405730

[Sanna Uusimäki | LinkedIn](#)

Sanna's main professional strengths are deep understanding of data and as a talented conversationalists a quick adaptation to the new client environments. She turn thoughts into action and she able to see how all the pieces and resources can be arranged for maximum productivity. She has a talent to see how different people can work together productively.

“Helping the clients to leverage their data is what drives me. Data platform architecture, data modelling, continuous development are my core competencies as a data architect. I am known to be able to translate business needs into efficient development backlog” she summarizes.

Core capabilities and recent achievements

Capability to understand and translate customer needs into business and technology solutions, providing consultation on the customer digital journey and digital transformation, and development and implementation of quality business digital solutions.

Cloud solution requirements analysis, solution design, conducting workshops and delivering projects.

DataOps and Scrum master practical knowhow that is supported by Scaled Agile Framework for enterprises certification.

Assignment history includes

Client Lead, Healthcare company (2021 – present)

Continuous Development of Cloud Data Platform, and on-prem Datawarehouse. Azure Data Services, Microsoft SSIS, Atlassian tools.

Data Solution Architect, Data Project Manager (11/2019 – present)

Cloud Data Platform development for clients in Chemistry, Healthcare, Banking with multivendor teams and team size varying from 4 – 8

Research and development team lead at Cloud1 (08/2022 – present)

Data Delivery Lead, Delivery and Governance at Cloud1 (04/2021-06/2022)

Data and Analytics consultant, Accenture (2011-2019)

Data Consultancy, Nordics Data Visualization Capability lead and multiple client projects. In 8 years at Accenture she took part in data related projects with data modeling, data analysis, reporting, to data warehousing. She lead a data warehouse continuous development team of 14 developers. She has done a SAFe Certificate (Scaled Agile Framework for enterprises).

Industry experience includes

- Healthcare
- Banking
- Insurance
- Retail
- Distribution Automation

Education and main certifications

Master's degree in Tech, Tampere University of Technology
Thesis “Implementing Metrics for Managing Software R&D Projects”



Esiintyjien LinkedIn



Sanna Uusimäki

Research and Development Lead for Platform
Solutions at Cloud1 Oy



Aleksi Saarinen

Data Engineer at Cloud1





Minä väitän

Tavoitteena on, että tämän jälkeen on helpompi etsiä tietoa.



Hypetermit vaihtuvat, mutta perusasiat pysyvät

Data	Tietokanta	Tietovarasto	Data-alusta
Data ei itsessään ole mitään. Se tarvitsee otsikot ollakseen informaatiota, ja vasta päätökset tuovat arvoa.	Rakenteellinen kokoelma dataa, jonka tallentamiseen on useita eri työkaluja. Tietokanta on myös juridinen termi.	Tietovaraston on mallinnettu datan varastointipaikka, jossa on tyypillisesti historiointi käytössä.	Datan käsittelyyn tarkoitettu kokoelma tiedon varastoinnin, muokkaamisen, jakelun ja hallinnan työvälineitä ja käytäntöjä (Data Governance).
Raportointi	Analytiikka	Koneoppiminen	Tekoäly
Datan ja information keräämistä ja muokkaamista päätöksenteon tueksi. Datan visualisointia.	Datan analysointia uuden tiedon luomiseksi. Sisältää myös datan visualisointia.	Algoritmeja, jotka on kirjoitettu korjaamaan itseään rakenteellisen datan perusteella. Tyypillisemmin jonkin työkalun osa.	Tällä hetkellä erittäin nopeasti kehittyviä malleja tekstin, puheen, kuvan, videon tuottamiseen. Voi olla itsenäinen työkalu..tai robotti.



SQL on ikuista

- On lukemattomia tapoja tuottaa, käsitellä ja siirtää dataa
- Structured query language (SQL), joka on kehitetty datan tallentamiseen ja muokkaamiseen relaationaalisessa tietokannassa.
- SQL:n looginen rakenne on ja pysyy, se on käytännössä joukko-oppia. Jokainen data-ammattilainen on joskus googlannut “left outer join”



Iso osa datasta on roskaa

- Joten taitavin kierrättäjä voittaa!
- Datan ymmärtäminen luomisen yhteydessä auttaa sen hyödyntämisessä myöhemmissä vaiheissa.
- Datan elinkaaren voi ajatella virtana, jossa sen korjaaminen on järkevintä heti, kun se syntyy.
- Datan käsittelyssä vaaditaan viestintä
- Mikä tahansa organisaatio hyötyy, kun ymmärtää omaa dataansa hiukan. Sitä kutsutaan hyväksi datakulttuuriksi.
- Filter input, escape output – vain luotettavasta datasta on hyötyä.
- Ymmärrä, miten data voidaan yhdistää muuhun dataan. Tarvitaan avain!





Pilvi on itseasiassa jonkun toisen tietokone

- Oletko tänään käyttänyt pilveä?
- Valtavat määrät dataa päätyy pilveen, eli jonkun toisen tietokoneelle, tiedätkö kenen?
- Pilvi on hiukan sama kuin keskitetyt tietokoneet olivat
 - Mahdollisuus skaalaukseen, ja nopeaan implementointiin
 - Uhat sinällään samat kuin ennenkin





Tietosuoja, tietoturva ja muu datatyön säätely

- Tietosuoja ja tietoturva mainitaan usein yhdessä, mutta ovat kaksi eri asiaa.
- Tietosuojalla tarkoitetaan henkilöön liittyvän datan käytön säätelyä.
- Tietoturva on datan suojaamista väärinkäytöksiltä ja tuhoutumiselta.
- Tietoturva on aina käytettävyyden ja turvallisuuden kompromissi.

- Vaatii erittäin hyvää sääntöjen tuntemista, että niitä kannattaa venyttää.
- Datatyössä on hyvin usein henkilörekisterin käsittelijänä, joten aiheen hyvä tuntemus ja jatkuva päivittäminen on pakollista





Eksplisiittisyys on kultaa eli datatyö vaatii tarkkuutta



Datan parissa täytyy usein miettiä, miten ratkaisu elää vuosia eteenpäin. Datatyön olennainen osa on datan tehokas käsittely ja fiksu varastointi.

Jos ei mitään muuta ota tästä esityksestä mukaan, niin ota se, että

VANHA on eksplisiittinen ja UUSI on aina tulkinnanvarainen.

Uudelleennimeä aina VANHA.

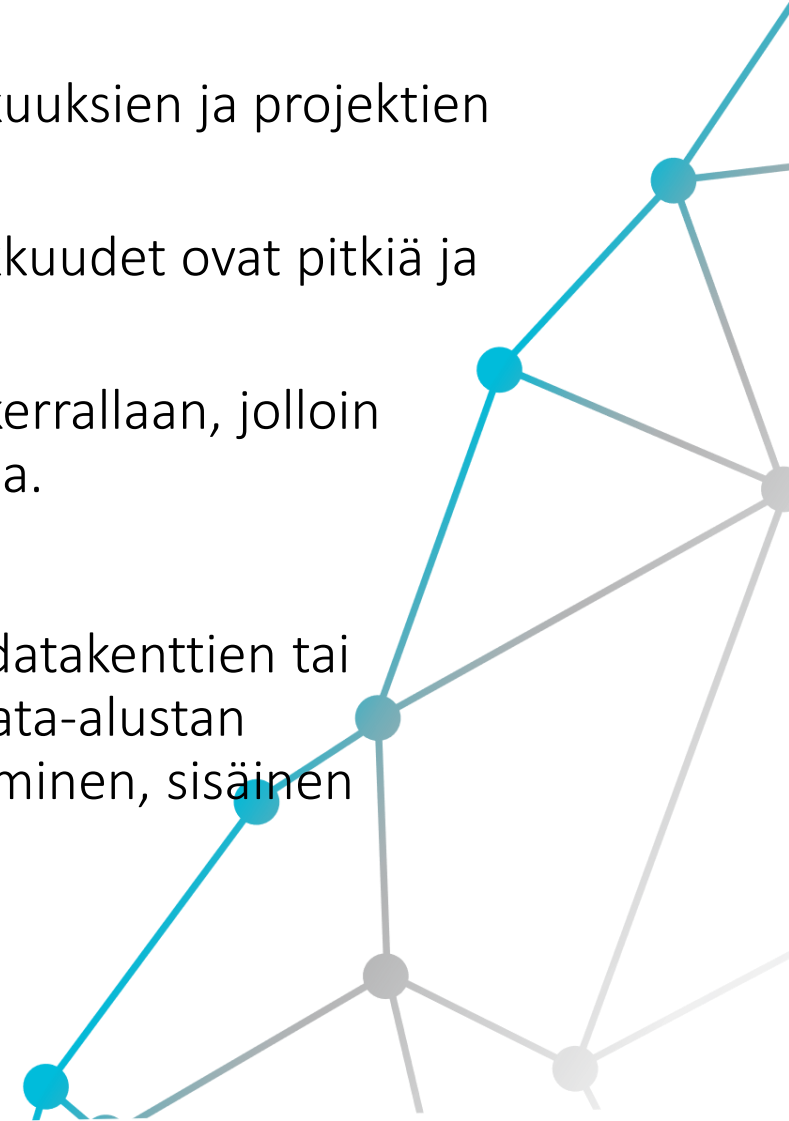


Missä nyt olen töissä



Tyypillinen työviikko Cloud1:lla

- Datatyötä tehdään meillä tiimeissä. On erilaisia tiimejä asiakkuuksien ja projektien mukaan.
- Tyypillisesti työskennellään tutun asiakkaan kanssa, sillä asiakkuudet ovat pitkiä ja asiakkaat tulevat tutuiksi.
- Tavallisesti kukin työskentelee noin kahdessa asiakkuudessa kerrallaan, jolloin datatyölle tyypillinen intensiteetin vaihtelu saadaan mitigoitua.
- Viikkopalavereissa sovitaan, mitä kukakin tekee seuraavaksi.
- Kehitystehtäviä ovat esim. tuotannon ylläpitotoimet, uusien datakenttien tai kokonaisuuksien tuominen data-alustalle, kokonaan uuden data-alustan pystyttämiseen liittyvät tehtävät, datan hallintamallin kehittäminen, sisäinen tuotekehitys ja niin edelleen.

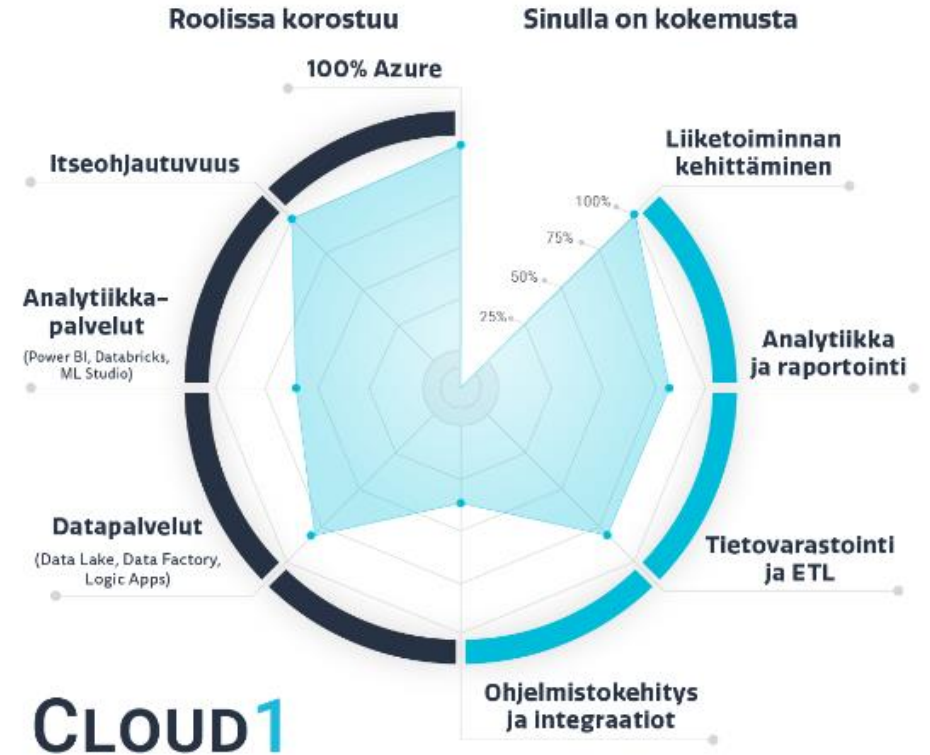




Datatyöläinen suunnittelee ja ratkoo

- Datan parissa on erilaisia toimenkuvia ja tyypillisesti termit vaihtuvat melko nopeasti
 - Väitän
 - Analytiikka ja Raportointi
 - ML/AI/Robotiikka
 - Data-alusta / tietovarastointi
 - Integraatioalusta / Service layer / etc.
 - Sovelluskehitys
 - Data scientist / data analyst
 - Data engineer
 - Data architect
 - Data solution architect
 - Integration architect
 - Mutta myös palvelupäällikkö, asiakasvastaava, projektipäällikkö, scrum master, product owner...

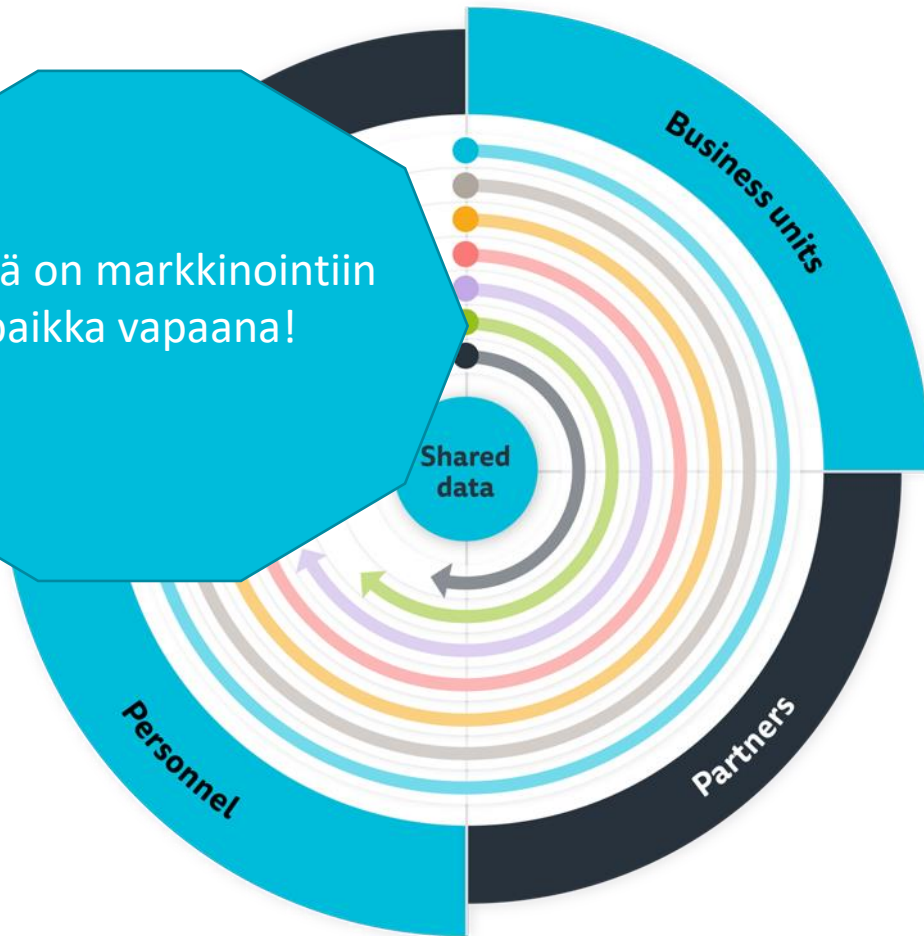
Azure Data Architect





Cloud1 – Architects of digitalization

Meillä on markkinointiin paikka vapaana!



Cloud1 Oy was founded in 2012 and is privately owned in Finland.

Our foundation builds on: visionary, humanity, fairness, professionalism and expertise.

Cloud1 provides with consulting services for building data capabilities on Azure.

Data Platforms, AI capabilities, Business Driven Reporting, Integrations, Governance guidances and Ops (support) -services

Cloud1 has a strong collaboration with Microsoft as a trusted partner.

Cloud1 has earned 6 Gold Competencies.

60+

Personnel
(80+ with
Devisioona)

500+

Azure
Projects

50+

Customers
(TOP200
Finland)

10 Milj.

Revenue
estimate
in 2023



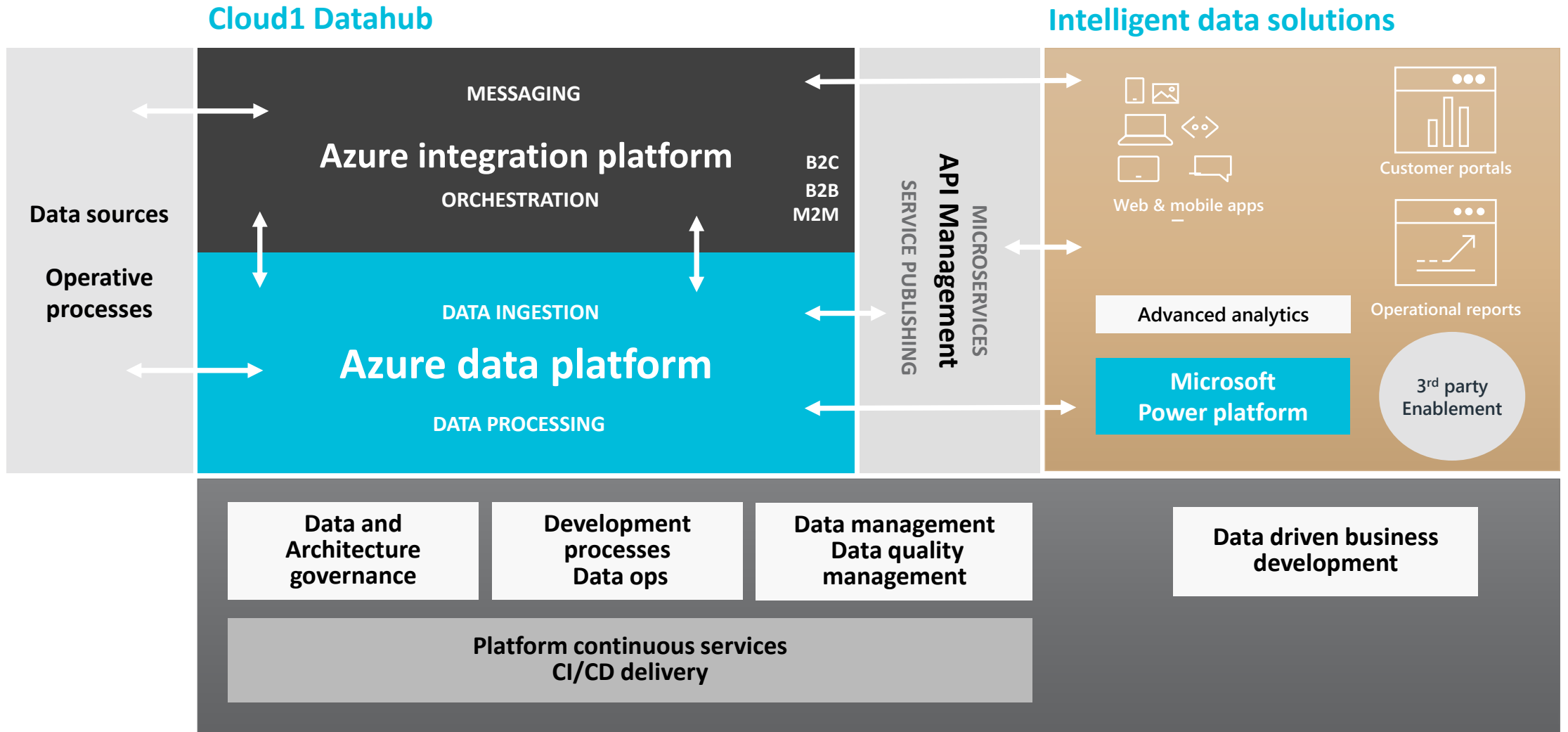
Datahub

Datahub is a term that Cloud1 uses of the concept of platform that stores, processes, manages, and shares data from various sources.

It is for making it easier to access, analyze, and utilize data for various applications. Datahub is for facilitating data discovery, data storing, data access, and data collaboration.

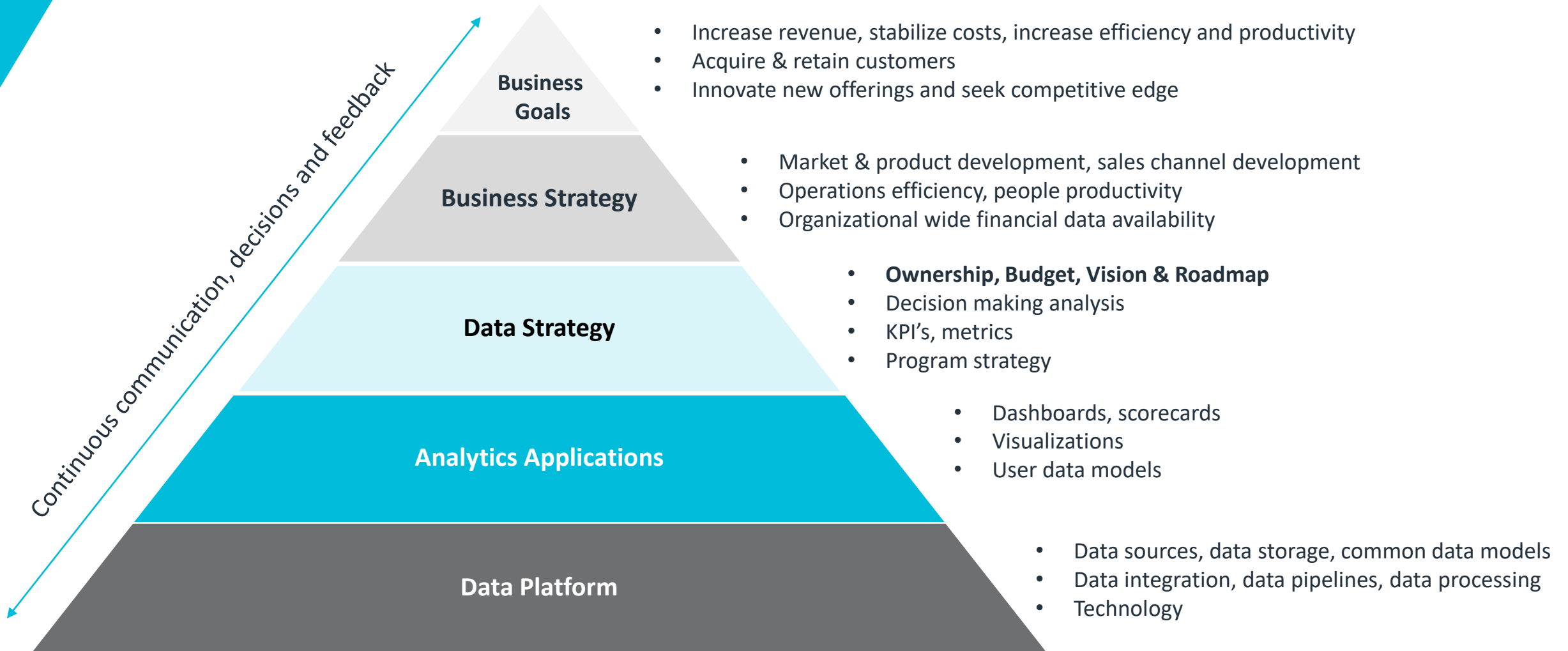


Cloud1 Datahub concept





Data Culture & Customer Business Value





Miten data-alalle pääsee?

- Vaatii kykyä oppia. Myös työssä jo pitkään olleet opiskelevat jatkuvasti.
- Peruspalikat
 - Millaisia erilaisia data-ammattilaisia on
 - SQL:n osaaminen perusteita pitemmälle on aina hyödyllistä
 - Python tai joku muu ohjelmointikieli on erittäin hyödyllinen
- Opiskeluun on paljon mahdollisuuksia
 - [Home - MOOC.fi](#) (Mimmit tietää)
 - [Exam DP-203: Data Engineering on Microsoft Azure - Certifications | Microsoft Learn](#)
 - Useilla konsultointitaloilla on lisäksi erilaisia trainee-ohjelmia
 - Cloud1 Datatyöpaja!
- Jos osaaminen on jo niin tästä linkistä:
[Oletko dataosaaja? Hae mukaan huippuammattilaisten joukkoon! | Cloud1](#)

ng / Browse / Get started with data engineering on Azure / Introduction to data engineering on Azure /

< Previous

Unit 5 of 6 ▾

Next >

Knowledge check

✓ 200 XP

3 minutes

Choose the best response for each of the questions, then select Check your answers.

1. Data in a relational database table is...

Structured

✓ That's correct. Relational databases contain structured data in tables.

Semi-structured

Unstructured

2. In a data lake, data is stored in?

Relational tables

Files

✓ That's correct. Data in a data lake is stored in files.

A single JSON document

3. Which of the following Azure services provides capabilities for running data pipelines AND managing analytical data in a data lake or relational data warehouse?

Azure Stream Analytics

Azure Synapse Analytics

✓ That's correct. Azure Synapse Analytics includes functionality for pipelines, data lakes, and relational data warehouses.

Azure Databricks

Next unit: Summary

Continue >



Cloud1 kutsuu Mimmit Datatyöpajaan!

Meillä on ilo kutsua Mimmit meille Vuorikadulle Helsinkiin datatyöpajaan!

14.6. keskiviikkona klo 12 alkaen

15 osallistujaa

Paikalla on data-ammattilaisia erilaisilla taustoilla

Ilmoittautuminen

Mimmit Koodaa –sivuston kautta.





Tietoa meistä | Cloud1

Mikä meille on tärkeää?

Toteutamme vaativat integraatio-, arkkitehtuuri- ja datahankkeet nopeasti ja tehokkaasti Microsoft Azure -teknologian ja huippuosaamisen avulla.

Inhimillisuus

Kunnioitamme ja arvostamme toisiamme. Huomioimme ihmisten yksilölliset tarpeet ja panostamme hyvinvointiin. Lisäarvo syntyy ihmisten välisessä vuorovaikutuksessa.

Ammattimaisuus

Panostamme työn laatuun. Ehdotamme vain ratkaisuja, joiden takana voimme seistä. Teemme toteutuksia, joiden lopputuloksesta voi olla ylpeä.

Reiluus

Tarjoamme työyhteisömme jäsenille tasa-arvoiset mahdollisuudet, ja kannustamme osallistumaan päätöksentekoon. Jokaisella on mahdollisuus kehittää omaa osaamistaan ja muotoilla työnkuvaa vahvuuksiensa sekä kiinnostuksenkohteidensa mukaan.

Asiantuntijuus

Suunnittelemme ja toteutamme ratkaisut loppukäyttäjää ajatellen. Tarjoamme räätälöityjä ratkaisuja, jotka vastaavat asiakkaidemme ainutlaatuisiin tarpeisiin.

Edelläkävijyys

Haastamme totut konventiot, uskallamme ajatella ja tehdä toisin. Kehitämme rohkeasti toimintaamme, näytämme suuntaa ja luomme uutta.