devoteam

Få et indblik i data operations på cloud

Data operations i skyen

Simon Stubben &
Rune Bendix Wittchen
04.05.2021

Innovative technology consulting for business.

DevoTalks

En webinar serie af digitale indspark

VELKOMMEN

DIGITALE
indspark

FORSKELLIGE
emner

SKIFTENDE
oplægsholdere

20 + 10
minutter

Lidt praktik inden start

Har du spørgsmål
kan du stille dem i chat-funktionen, og de vil blive besvaret efter 'talk'en'

Vil du gerne gense slides
bliver denne DevoTalk optaget og gjort tilgængelig på sitet devotalks.dk

Generer "video-vinduet"
af dagens taler i fht. at se slides, kan du selv drag'n'drop'e vinduet til et passende sted

Agenda

01 Dataprojekter, skitsering af baggrunde

02 Modenhed i dataprojekterne

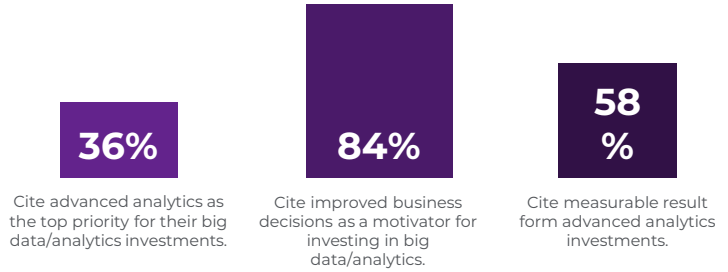
03 Hvad er DataOps, og hvordan ledes projekter?

04 Data-opbevaring og -behandling i skyen

05 DataOps værktøjer løser udfordringer

06 Hvordan måles DataOps succes?

Data-projekter: skitsering af baggrunde



Source: New Vantage Big Data Executive Survey 2018

Management:
Ledelsesredskaber til styring af aktiviteter og ressourcer på baggrund af data.

Business:
Forretningsudfordringer som kan løses med viden

Challenges:
Dem kender i bedst selv

Data:
Katalog over tilgængelig data i organisationen

Diverse kilder:
Data i form af rapporter eller filer i organisationen.

Systemkilder:
Data samlet i en database

Employees:
Ressourcer med interesse i at "dykke ned i data"

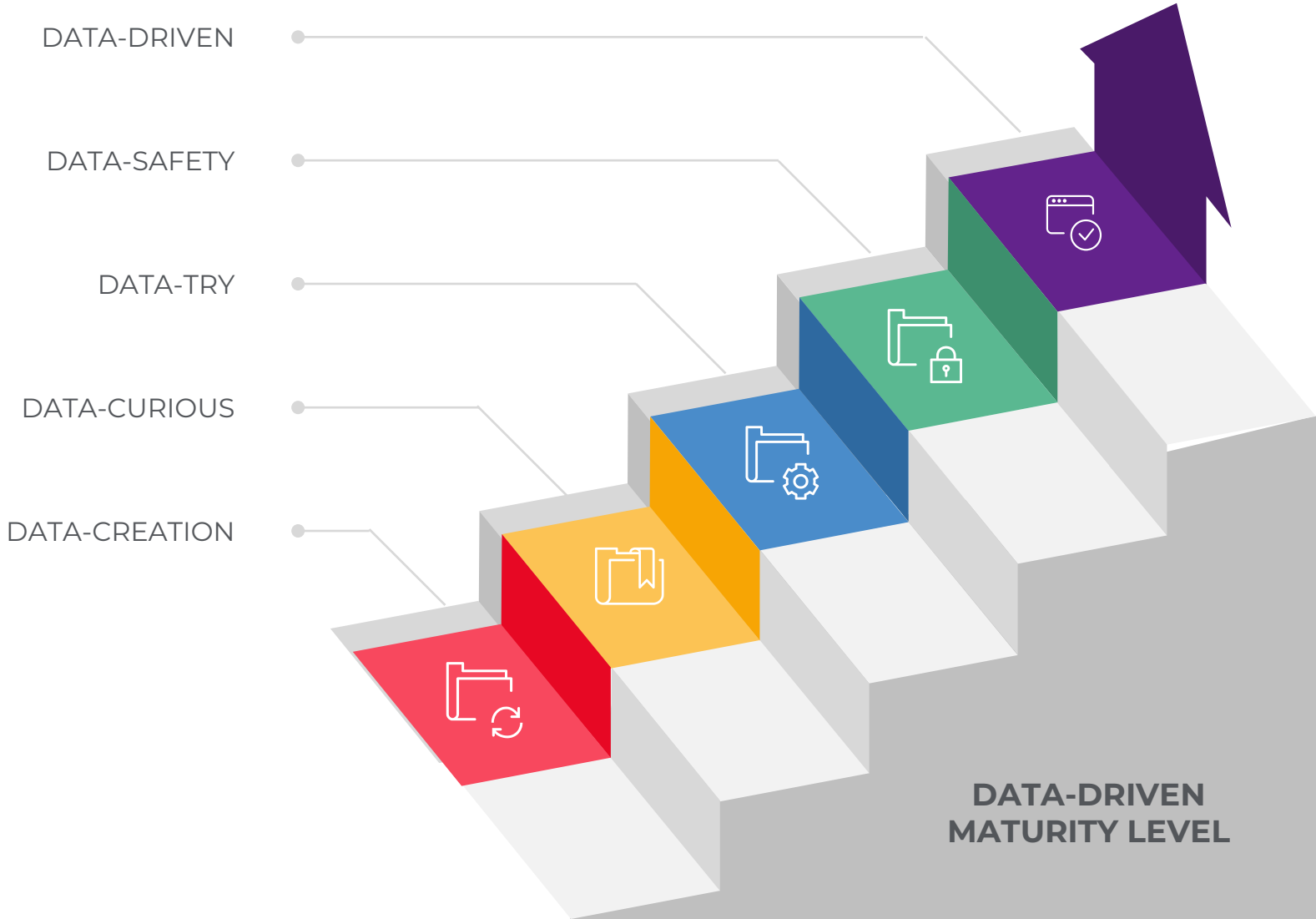
Fresh:
Nyligt rekrutterede med andre kompetencer

Experienced:
Nuværende ansatte med historisk erfaring

Dataprojekter kræver modenhed, opbakning og DataOps!

Et projekts datamodenhed kan testes med DataOps redskaber

Skrøbeligheder i eksplorative dataprojekter må accepteres. Og man må erkende at det kan være en svær opgave at lykkes.



Hvad er DataOps/Data Operations?

DataOps er et rammeværk for projektstyring og automatisering af opgaveløsning på dataområdet. 01

Definerer et udgangspunkt for at opstarte dataprojekter, og “evangelisere best practices”. 02

Formålet er at udvikle en konsistent, valideret og sikker opsætning for bearbejdning af data. 03

18 DataOps principper hos dataopsmanifesto.org

DataOps strategi, proces og ledelse

1. Opstil en DataOps strategi som understøtter en value proposition:
   En **nedbrydning af silotænkning og manuelle processer** for databehandling.

2. Lad DataOps understøtte processer som udvikler forretningen:
   DataOps skal sikre nye **samarbejder mellem interessenter** i forretningen.

3. Ledelse skal støtte at data bliver en del af forretningens drivende kraft:
   FTE'er med datafokus skal have **arbejdsgange støttet af DataOps**

Andre forventninger til DataOps projekter

01

Hele projektplanen og udviklingsteamet behøver ikke at være defineret fra start.

02

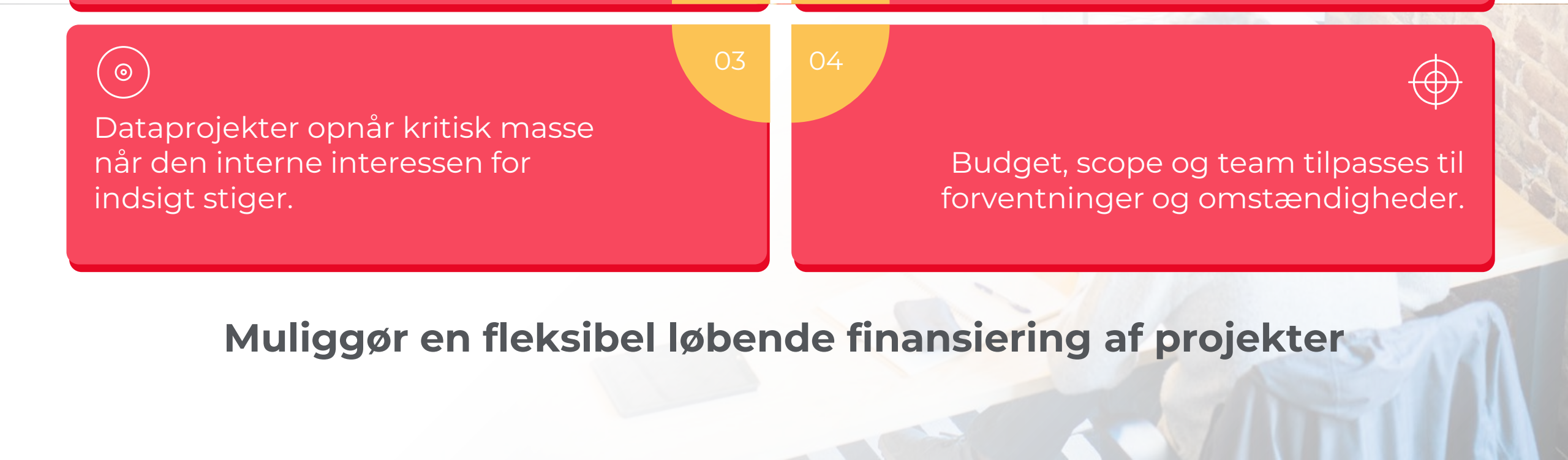
Elementer udvikles og evalueres løbende, men distribueres først ved tilfredshed med løsning.

03

Dataprojekter opnår kritisk masse når den interne interessen for indsigt stiger.

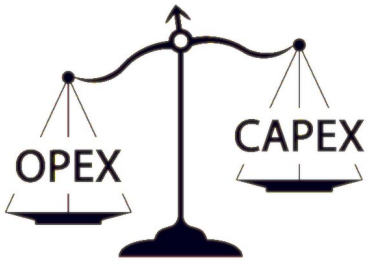
04

Budget, scope og team tilpasses til forventninger og omstændigheder.



Muliggør en fleksibel løbende finansiering af projekter

Dataopbevaring i skyen - nyt og gammelt



1
Fra velkendte relationelle on-prem databaser

Til cloud håndterede og optimerede databaser -- af samme type.

2
Eller videre til cloud-only instanser (DBaaS)
Bliv forberedt til avanceret DataOps.

BigQuery

Cloud SQL

SQL

Amazon DynamoDB

3
Afregning pr. forbrug vs. Investeringsudgifter

Dataopbevaring der imødekommer dine retningslinier

Compliance

- Lever op til bl.a.
  - ISO 27001, SOC 1-3, CIS, GxP
  - GDPR, FSA (Finanstilsynet)
- Kontraktmæssig garanti for at data ikke forlader EU

Sikkerhed

- Datacentre er fysisk sikrede, standard policies, adgangsstyring og netværks-segmentering
- Indbyggede sikkerhedsredskaber, der bl.a. kan opdage trusler

Fortrolighed

- Fuld kontrol over adgang til data - hvem, hvordan, hvornår, hvorfra
- Al data er per standard krypteret - hvormed udbyderne ikke kan få adgang

Databehandling i skyen - principperne er de samme

Ændringen ved at flytte databehandling til skyen, er en omstilling fra lokalt datamiljø til et som ligner - men udnytter skyens services.

Data dokument

Database

Data logik

Interface

Rå data	Transformeret data	Publiceret data

FTP server

Cloud Function

Cloud Composer

Importér til BigQuery

Automatiseret validering og transformering

Publisér data vha. Data Studio

DevOps vs DataOps - forskellige verdener

DevOps er IKKE DataOps

	DevOps	*DataOps*
Målgruppe	**Softwareudviklere & slutbrugere**	**Data skabere og databehandlere**
Kvantificeres pba.	**Software funktionalitet**	**Data output**
Kvalitetssikres pba.	**Kode review, automatiserede og funktionelle tests, monitorering**	**Data governance, automatiserede tests, monitorering**

Redskaber - et udpluk

Google Cloud

python™

CSV

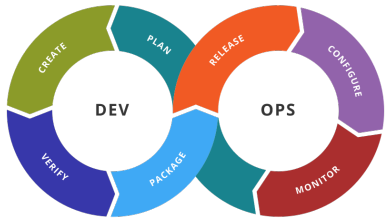
Cloud Functions

Cloud Composer

JIRA

great_expectations

Apache Airflow



Opsummering

Tre hovedbudskaber

- En enorm mængde ressourcer gør skyen til et fremragende sted til at bearbejde data.
- DataOps opstiller et rammeværktøj hvori denne type udfordringer imødegås.
- Få lavet en strategi og plan, der passer din organisation.

Spørgsmål

- Hvordan står jeres organisation overfor data projekter?
- Hvor moden mener I, at jeres organisation er ift. behandling af data?

Dataprojekter kræver modenhed, opbakning og DataOps!

Den industri-nære forskning på området, analyserer sig frem til følgende:

85% of big data projects fail

(Gartner, 2017)

87% of data science projects never make it to production

(VentureBeat, 2019)

“Through 2022, only 20% of analytic insights will deliver business outcomes”

(Gartner, 2019)

DataOps opstiller et rammeværk hvori denne type udfordringer imødegås.

DataOps principper

It's a team sport to embrace change, that requires daily interactions by self-organized teams. 01

Analytics is code and analytics is manufacturing of value. Value working analytics. 02

Orchestrate analytics to make it reproducible in disposable environments. 03

Quality is paramount, so monitor quality and performance to improve cycle times of upgrades. 04

Kondenseret fra de 18 DataOps principper hos dataopsmanifesto.org

Dataopbevaring i skyen - genkendelige databaser

Velkendte relationelle on-prem databaser

- MS SQL Server
- Oracle DB
- PostgreSQL
- MongoDB
- MySQL
- IBM Db2
- SAP HANA

Cloud-only instanser (DBaaS - ready for DataOps)

- Google BigTable (på Google/GCP)
- Azure CosmosDB (på Microsoft Azure)
- AWS Redshift (på Amazon AWS)
- Snowflake (på Azure/AWS/GCP)
- . . .

Cloud håndterede og optimerede databaser

- Azure SQL (managed SQL Server på Azure)
- Oracle Autonomous Database (managed Oracle DB)
- Amazon RDS for PostgreSQL (managed PostgreSQL på AWS)
- MongoDB Atlas
- Amazon Aurora (managed MySQL på AWS)
- IBM Db2 on Cloud (managed DB2 på IBM cloud)
- SAP HANA Cloud (managed SAP DB)

For more info: db-engines.com

Alternative slide 7

DataOps strategi, proces og ledelse

01

Opstil en DataOps strategi som understøtter en value proposition:

En **nedbrydning af silotænkning og manuelle processer** for databehandling.

02

Lad DataOps understøtte processer som udvikler forretningen:

DataOps skal sikre nye **samarbejder mellem interessenter** i forretningen.

03

Ledelse skal støtte at data bliver en del af forretningens drivende kraft:

FTE'er med datafokus skal have **arbejdsgange støttet af DataOps**