

Fakta om Datafordeleren

Fællesoffentlig datadistribution



Styrelsen for
Dataforsyning og
Effektivisering

21. november, 2018



Fakta om Datafordeleren

Dette dokument giver tekniske fakta om den Datafordeleren.

Dokumentet er et supplement til den konceptuelle beskrivelse i dokumentet "**Teknikken bag Datafordeleren**" og "**Modulopbygning af Datafordeleren**" samt de tekniske slides i præsentation fra Informationsmøde.

Områder som dækkes:

- Overordnede principper for distribution af data
- Generelt om den tekniske arkitektur
- Udviklingsplatform
- Servertyper, rolle og platform
- Oversigt over miljøer
- Sikkerhed
- Genskabelse af miljøer
- Dokumentation af AS-IS
- Skønnet kapacitet
- Referenceliste

Overordnede principper for distribution af data (Ref 8)

Med idriftsættelsen af den fællesoffentlige Datafordeler er der etableret en fællesoffentlig infrastruktur til distribution af Grunddata, med følgende overordnede principper:

- Alle Grunddata distribueres via Datafordeleren og Datafordeleren kan også anvendes til at distribuere andre relevante data end Grunddata.
- Opdatering af Grunddata sker fortsat gennem Registermyndighedens grænseflader. Registermyndigheden skal til Datafordeleren levere og vedligeholde en opdateret kopi af Grunddata med en frekvens, der er tilstrækkelig i forhold til brugernes behov for aktualitet.
- Eksisterende dataansvar er uændret, og Registermyndighederne sikrer udviklingen af specifikationer af Tjenester til udstilling på Datafordeleren. Tværgående databeskrivelser udarbejdes i samarbejde af de relevante Registermyndigheder.
- Datafordeleren distribuerer data via Online-opslag, hændelser og fil-distribution til både offentlige og private brugere.
- Datafordeleren giver adgang til Grunddata via standardiserede aftaler.
- Hvis en offentlig eller privat bruger i medfør af lovgivning, frikøbsaftaler eller betalingsaftaler har ret til at anvende data i Datafordeleren, så kan denne bruger som udgangspunkt hente disse data uden yderligere omkostninger.
- Datafordeleren opbevarer og distribuerer Grunddata i henhold til gældende lov

Generelt om den tekniske arkitektur Ref (5)

Arkitektur - Principper for datadistribution

- Systemet bruger åbne og offentlige standarder (Eks: XML, OCS, SAML, JSON, Geostandarder)
- Systemet består af løst koblede komponenter
- Trediepart “non-open source libraries and programs” der er brugt skal være “encapsulated in standardized interfaces” sådan at det er muligt at udskifte systemet og dets delkomponenter
- Systemet skal anvende moderne og velprøvede teknologier
- System-API'erne skal være fleksible og kunne modtage og sende data i forskellige formater
- Alle logs i hvert miljø tidssynkroniseres og samles i LogPoint

Arkitektur - Principper for datadistribution

- Systemet er løst koblet til dets datakilder, så datakilderne kan være nede for vedligeholdelse uden at forstyrre systemets ydeevne.
- Systemet skal være tilgængeligt for en bred vifte af brugere. Bedst praksis for brugervenlighed og handicapadgang skal følges.
- Systemet indeholder følsomme oplysninger. Systemet har et højt sikkerhedsniveau.
- Fejlhåndtering skal være robust og selvforklarende.
- Systemets ydeevne skal overvåges

Udviklingsplatform (ref (1), (5))

Primære udviklingsplatform er Microsoft Visual Studio 2013. Kode ligger i Team Foundation Server (TFS)

- C# er det primære udviklingssprog og følger "C# Coding Conventions (C# Programming Guide) fra Microsoft"
- Primære udviklingsplatform er Microsoft Visual Studio 2013
- Kode ligger i Team Foundation Server (TFS)
- Biblioteker fra 3 part er benyttet. Bla. kode og pakker fra NuGet, JavaScript biblioteker og andre pakker der kan være omfattet af licens inkl open source.
- Løsningen er udviklet i på .NET platformen og bruger komponenter fra 3. part, bla: MapServer, MapCache, GeoNetwork, FME og Snowflake Software.

Backend Technology

- Rammeværk for entiteter som ORM
- Unity er valgt som IoC container
- Active Directory er brugt som User Storage, og AD FS vil blive brugt til at aktivere systemet til at bruge en fødereret sikkerhedsopsætning.
- MSSQL Server er brugt til at gemme alle systemdata, mens PostgreSQL vil blive brugt til lagring af alle data, der er tilgængelige til distribution (Basic Data).

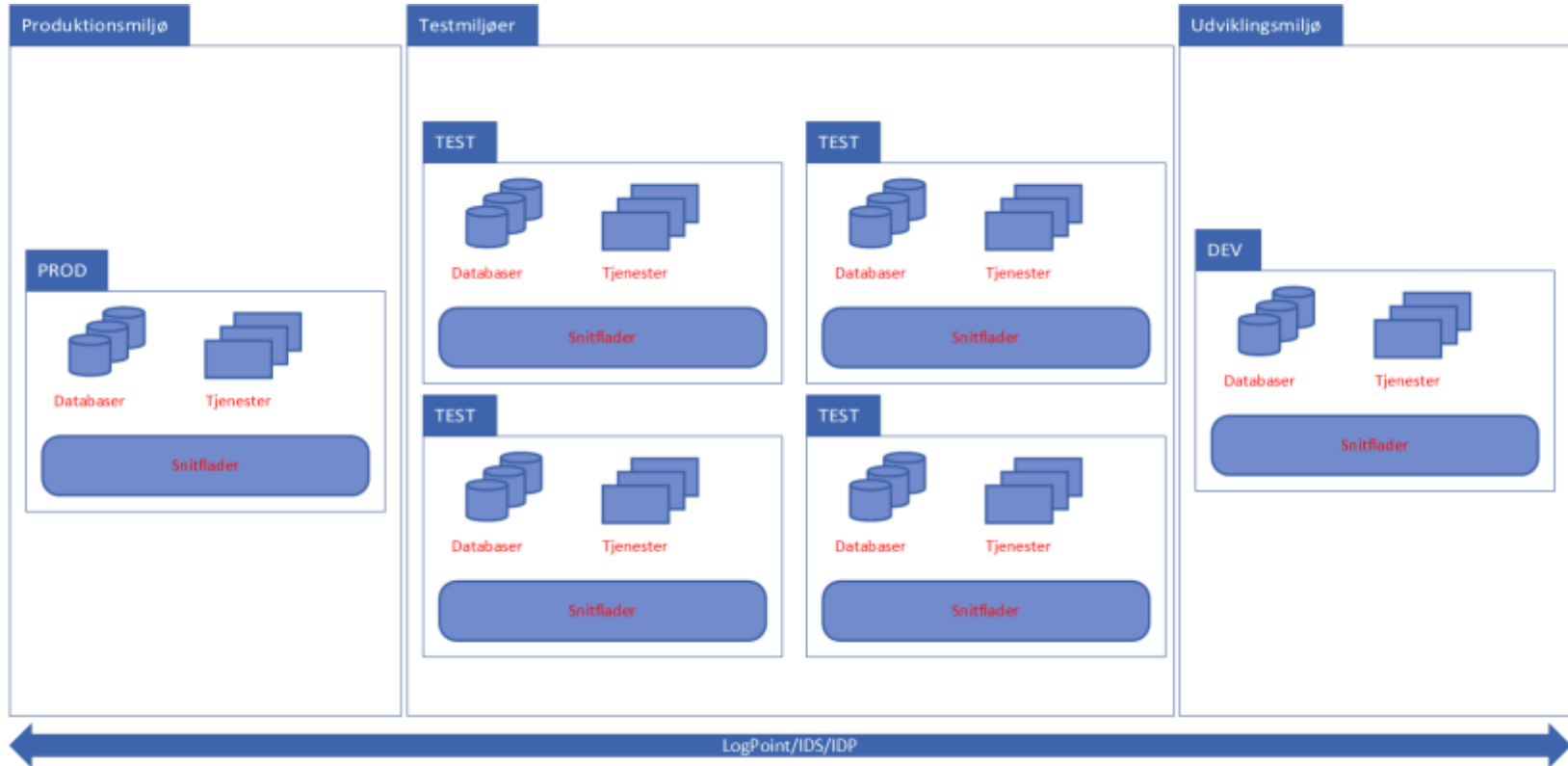
Frontend Technology

- AngularJS
- JQuery
- Bootstrap
- ASP.NET MVC til hosting af front-end applikationen

Servertyper, Roller og platform (Vejledende oversigt) (ref (2))

Servere er i primært virtuelle Antal af servere og instanser er vejledende	1 produktionsmiljø ca 34 server instanser Antallet af miljøer og servere er usikkert	4 testmiljøer ca 110 server instanser Antallet af miljøer og servere kan svinge meget	1 udviklingsmiljø ca 20 server instanser Antallet af miljøer og servere kan svinge meget
Servertyper (x) giver ca. antal instanser med denne servertype	Domain Controller (1) Active Directory Federation (1) Applikation Server (9) Database Server (6)	Samme opsætning som produktion. Dog med lidt færre instanser og servere. De enkelte testmiljøer er forskelligt konfigureret.	Varierer i forhold til udviklingsopgaven
Roller (Applikation) (x) angiver ca antal instanser med denne rolle.	Federated Security Area Active Directory Federation Batch host (3) FME, FTP file upload application Snowflake Gopublisher (2) Claims Database PostgreSQL Master (3) PostgreSQL Slave (2) FME Engine Filehost (2) Gateway (2) GeoNetworks MapServer/Cache Monitor agent SALT Master KeyValueStore CouchBase Administration Website FME Web server ADFS Proxy MapCache Self Service Website Switchboard (2) Update chanel Web Application Proxy	Samme opsætning som produktion. De enkelte testmiljøer er forskelligt konfigureret.	Varierer i forhold til udviklingsopgaven
Platform (OS) (x) angiver ca antal instanser på denne platform Windows Servere 2012 R2 er på nyeste patch	Domain Controller 2012 R2 (Scale Out) Windows Server 2012 R2 - Domain Member Windows Server 2012 R2 - Workgroup (4) Ubuntu 14.04 (2) Microsoft SQL Server 2014 - Workgroup Ubuntu 14.04 (14) Windows Server 2012 R2 - Workgroup (5) Windows Server 2012 R2 with IIS 8.5 - Workgroup (5)	Samme opsætning som produktion. De enkelte testmiljøer er forskelligt konfigureret.	Varierer i forhold til udviklingsopgaven

Oversigt over miljøer (Ref (8))



Sikkerhed (ref (6+7))

Sikkerhedszone	Dataindhold	Autentifikation/autorisation teknologi	Autorisationstype
S0	<ul style="list-style-type: none">- Offentligt tilgængelige data og tjenester til levering af disse data.- Selvbetjeningsportalen	<ul style="list-style-type: none">- Brugernavn/password (FTP+HTTP/HTTPS)- NemID (OCES) FOCES/VOCES/MOCES- NemLog-in (SAML2)- AD FS	<ul style="list-style-type: none">- Anonym- Kendt bruger- Begrænset
S3	<ul style="list-style-type: none">- Administrativ data- Administrativ portal	<ul style="list-style-type: none">- NemLog-in (SAML2)- NemID (OCES) FOCES/VOCES- SSH2 (SFTP)- Registreret IP adresse- AD FS	<ul style="list-style-type: none">- Begrænset (Rollebaseret)
S5	<ul style="list-style-type: none">- Følsomme data og tjenester til udstilling af disse- Komponenter til modtagelse af data fra dataleverandørerne	<ul style="list-style-type: none">- NemLog-in (SAML2)- NemID (OCES) FOCES/VOCES- SSH2 (SFTP)- Registreret IP adresse- AD FS	<ul style="list-style-type: none">- Begrænset (Rollebaseret)

- Sikkerhedszone opdelingen er dækket af “no read up, no write down”, mellem niveauerne.
- Kommunikation mellem sikkerhedszoner foregår kun igennem køer der kun kan flytte data til en højere sikkerhedszone, fx S0->S3/S5.

Genskabelse af miljøer ref (3)

I forbindelse med migrering af infrastrukturen skal følgende komponenter formodentlig genskabes hos en ny leverandør:

- Installation og konfiguration af netværks komponenter, herunder switches, routers, firewalls og load-balancere
- Opsætning og konfiguration af VLAN's, specielt ift. sikkerhedsopdeling i zoner
- Installation og konfiguration af IPS/IDS/DDoS og lign sikkerhedsudstyr
- Installation og konfiguration af overvågningsudstyr til monitorering og logning
- Installation, konfiguration og provisionering af fysiske servere, server-farme, virtuelle servere
- Installation og konfiguration af storage systemer (SAN, LUNs, lokale diske mv)
- Installation og konfiguration af backup systemer (diske, cartridges, robotter mv)
- Opsætning og konfiguration af cloud baserede servere og services.

Ovenstående liste er vejledende og der kan derudover være komponenter hos den nuværende Leverandør, som skal reetableres i en ny leverandørs miljø.

Dokumentation af AS-IS (Ref(4))

I forbindelse med overtagelsen af systemet overdrages eksisterende dokumentation. Kvaliteten vil følge kontrakten med den nuværende leverandør. Her står bla:

Uddrag af tekst fra kontrakt med den nuværende leverandør

Leverandørens dokumentation skal udarbejdes og løbende vedligeholdes i overensstemmelse med god skik i it-branchen, således at dokumentationen understøtter Kundens behov for et driftssikkert, skalerbart og overdrageligt System.

Skønnet kapacitet (ref(10))

Oversigt nedenfor viser forecast over anvendelse af Datafordeleren fra det oprindelige udbudsmateriale. Ibrugtagning af Datafordeleren har af flere årsager ikke fulgt den forventede takt. Der er dog intet der pt. indikerer, at niveau og stigningstakt for anvendelse skulle afvige væsentligt fra det oprindelige forecast. Dog er der en forventning om, at transaktionsmængden bliver lidt mindre, idet brug af filudtræk er mere efterspurgt. Datamængde og –trafik forventes dog at ligge på det forecast'ede niveau. Det skal endvidere bemærkes, at varighed af paralleldrifft ift. CPR og CVR er øget i forhold til det oprindelig forventede.

De store Geodata registre forventes idriftsat og ibrugtaget i 2019.

I forhold til forecast i tabellen nedenfor er det derfor forventningen, at anvendelse af Datafordeleren vil nå et omfang fra midt 2019 til midt 2020 som modsvarer et niveau mellem 2015/2016 og 2016/2017 nedenfor.

Skønnet kapacitet over 8 år	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019	2019/2020	2020/2021	2021/2022
Antal forespørgsler/måned	0	216.967.579	364.167.033	491.411.086	614.263.857	767.829.821	959.787.276	1.199.734.096
Registerdata	0	27.795.078	29.322.033	30.723.585	38.404.482	48.005.602	60.007.003	75.008.754
Geodata (simple)	0	75.669.000	133.938.000	184.275.000	230.343.750	287.929.688	359.912.109	449.890.137
Geodata (almindelige)	0	82.290.038	145.657.575	200.399.063	250.498.828	313.123.535	391.404.419	489.255.524
Geodata (komplekse)	0	31.213.463	55.249.425	76.013.438	95.016.797	118.770.996	148.463.745	185.579.681
Datamængder (GB)	0	31.035	39.843	42.251	52.814	66.017	82.522	103.152
Datatrafik (GB/måned)	0	16.796	33.021	46.087	57.609	72.011	90.013	112.517

Referenceliste

Ref (1): Developers guideline v 1.4

Ref (2): DFD_Server_instance_list v 1.1

Ref (3): Internal document

Ref (4): Bilag 5, Dokumentation v 4

Ref (5): Architecture Description v 2.4

Ref (6): Brugervejledning - Administrationsportal v 2.6

Ref (7): Datafordeler System documentation v 3.0

Ref (8): Underbilag 3B, Leverandørens løsningsbeskrivelse v. 5.6

Ref (9): Bilag 03 - Leverancebeskrivelse_v3.1

Ref (10): Bilag 12A – Vederlagsberegningsskema_v. 3.1

Ovenstående er de primære kilder. Teksten i denne præsentation kan være justeret og tilrettet.