Functionele specificaties

Omgevingsloket online

Point-in-time data recovery

Augustus 2018

Versie 2.14.0

Inhoudsopgave

1 Inleiding

1.1 Identificatie

Dit document is het functioneel ontwerp voor feature B06 betreffende de inrichting van Point-in-time data recovery (P-i-T recovery) in de database van het Omgevingsloket.

Dit document is een nadere uitwerking van een applicatief onderdeel van het beheerconcept zoals vastgelegd in … (dit document dient nog tot stand te komen).

1.2 Doel van dit document

Het doel van dit document is om de functionaliteit van Omgevingsloket online op zodanige wijze te beschrijven dat het configureren, modelleren en testen van Omgevingsloket online in Be Informed mogelijk wordt.

1.3 Scope en uitgangspunten

1.3.1 Scope

De scope van dit document beperkt zich tot nieuwe functionaliteit die in release 2.3 zal worden gebouwd binnen de Be Informed applicatie. Deze functionaliteit vormt onderdeel van het geheel aan maatregelen om het beheer op het systeem Omgevingsloket mogelijk te maken. Het betreft hier een aantal aanpassingen in de verschillende webapplicaties en aanpassing in de configuratie van de database.

Aanvullend zullen een aantal procesmatige maatregelen moeten worden genomen. Dit wordt aangegeven in hoofdstuk 3 (Checklist) en valt verder buiten de scope van dit ontwerp.

P-i-T recovery gaat over het herstellen van data wanneer deze corrupt is geraakt. Dit kan bijvoorbeeld door het uitvoeren van verkeerde SQL-statements op de database. Uitdrukkelijk vallen maatregelen voor uitvallen van een database niet onder dit onderwerp. Voor het uitvallen van een database worden andere (infrastructurele) maatregelen genomen, bijvoorbeeld door een fail-over database.

Ook blijft herstel mogelijk met het terug zetten van een back-up. De back-up voorziening wordt als aanvullend op P-i-T recovery gezien.

1.3.2 Uitgangspunten

De volgende uitgangspunten zijn gehanteerd bij het opstellen van dit ontwerp:

- De fysieke infrastructuur van het Omgevingsloket is ingericht en beschikbaar voor ten minste release 2.2. Hierin zijn opgenomen LOG4J en MySQL 5.0.

1.4 Leeswijzer

Dit document is als volgt opgebouwd.

Hoofdstuk 1 is een algemeen hoofdstuk waarin gegevens over dit document zijn opgenomen. Hoofdstuk 2 geeft een inleiding op het onderwerp P-i-T recovery. Het hoofdstuk geeft een voorbeeld van de praktijk van P-i-T recovery en geeft enkele aandachtspunten voor het verdere ontwerp. Hoofdstuk 3 geeft een checklist. Op basis van deze checklist kan een administratief proces worden ingericht in geval de Point-in-Time maatregel moet worden genomen. Ook geeft het een inkadering van de te realiseren maatregelen. Deze komen in de volgende twee hoofdstukken aan bod. Hoofdstuk 4 geeft aan wat er gerealiseerd moet worden aan logging. Hoofdstuk 5 geeft richtlijnen voor de configuratie van P-i-T recovery in de database.

2 Point-in-Time data recovery

> Point-in-time recovery in the context of computers is a system whereby a set of data or a particular setting can be restored or recovered from a time in the past. An example of this is Windows XP's feature of being able to restore operating system settings from a past date (before data corruption occurred, for example), or PostgreSQL's feature of being able to view a database table and its data as it was at a particular date in the past. Also, Time Machine for Mac OS X is an example of Point-in-time recovery.
> *(Wikipedia)*

Point-in-time data recovery (P-i-T recovery) gaat over het herstel van de database in geval dat er een verstorende productiefout in de database van Omgevingsloket is opgetreden. P-i-T recovery gaat over het herstel van corrupte data in de Omgevingsloket database. Het heeft geen betrekking op het herstellen van geüploade documenten of op het herstel gegevens in de backoffice systemen van de behandelende Bevoegd Gezagen. Op dit moment is hier de voorziening dat de back-up kan worden terug gezet. Wanneer data in een database corrupt (foutief) is geraakt, dan biedt een back-up voorziening onvoldoende mogelijkheden. Een back-up voorziening gaat om het fysiek terug zetten van data, terwijl P-i-T recovery gaat over herstel van de inhoud van data. In een aantal gevallen is dit niet genoeg. In deze situaties is P-i-T recovery een verfijndere maatregel die aan de applicatiebeheerder ter beschikking wordt gesteld. Met P-i-T recovery is hij in staat om de schade te beperken door niet een tabel als geheel terug te zetten, maar bepaalde opdrachten op de database ongedaan te maken.

Dit ontwerp wordt om twee redenen opgesteld: ten eerste is P-i-T recovery relatief eenvoudig in te regelen op de database. Het betreft hier een standaard voorziening die op een eenvoudige wijze geactiveerd kan worden. Ten tweede biedt dit ontwerp een aanknopingspunt om een checklist op te stellen van acties die moeten worden ondernomen wanneer er een productiefout in de database optreedt. Deze checklist is ook van belang als P-i-T recovery niet wordt doorgevoerd. Deze checklist is ook van belang als alleen de standaard back-up beschikbaar is en voor alle handmatige acties die verder nodig zijn.

2.1 Scenario: Hoe werkt P-i-T recovery in de praktijk?

In dit scenario wordt de situatie van een applicatiebeheerder beschreven na dat de P-i-T recovery op basis van dit functioneel ontwerp is gerealiseerd. Op basis van dit scenario wordt de aanvankelijke scope verder afgebakend en worden de concrete maatregelen die in dit ontwerp worden geadresseerd, benoemd.

Bij de functioneel applicatiebeheerder komt het signaal binnen dat bepaalde gegevens in de database niet juist zijn. Hij herinnert zich dat de dag hiervoor een databasebeheerder onderhoud aan het plegen was. De beheerder kijkt in de database en ziet tot zijn schrik dat er een tabel ontbreekt. Een herstelactie is nodig. Op basis van het transactielog van P-i-T recovery, is het mogelijk om te achterhalen waar en wanneer het fout is gegaan. Om te kijken wie er moet worden geïnformeerd over de herstelwerkzaamheden, i.c. de personen en instanties waarvan mogelijkerwijs gegevens verloren zijn gegaan, kunnen worden achterhaald met behulp van het recovery trail logbestand. Dit bestand bevat onder anderen de gegevens van de dossiers die in het afgelopen uur zijn gewijzigd. Voor deze dossiers wordt de checklist afgelopen. Als onderdeel van deze checklist is het synchroniseren van deze dossiers opgenomen. In deze synchronisatieslag worden de geüploade bestanden die geen referentie meer hebben in de database, van de fileserver verwijderd. Indien nodig wordt de aanvrager geïnformeerd over het opgetreden probleem. Aan de aanvrager wordt gevraagd om zijn aanvraag na te lopen op inconsistenties. Binnen het uur is de database weer hersteld en zijn de ontbrekende tabellen weer in oude staat terug gezet. De gebruikers waarvan het vermoeden bestaat dat ze een aanvraag hebben ingediend in de periode dat de database inconsistent was, zijn geïnformeerd.

Uit bovenstaand scenario blijkt het voordeel van P-i-T recovery. Wanneer P-i-T recovery niet is ingeregeld, dan kan men in geval van calamiteiten alleen terug vallen op de dagelijkse back-up. Met P-i-T recovery is het mogelijk om gemaakte fouten beter te herstellen, waardoor het probleem van gegevensverlies aanzienlijk minder groot is.

2.2 Aandachtspunten

Uit het scenario blijken ook een aantal andere zaken, waarbij men met de inrichting van deze functie rekening moet houden. Hierbij worden eerst de onderdelen genoemd die in dit ontwerp verder worden uitgewerkt:

- P-i-T recovery moet worden ingesteld. P-i-T recovery moet als één van de maatregelen worden gezien, waarmee het herstel van een database mogelijk wordt gemaakt. Andere maatregelen zijn de back-up voorziening en de fail-over inrichting van de omgevingen. Met P-i-T recovery kan terug worden gegaan tot elk moment in de tijd. Variërend van een seconde tot dagen.

- Wanneer P-i-T recovery wordt ingesteld, bestaat de noodzaak om een recovery trail logbestand in te richten, op basis waarvan kan worden achterhaald welke vergunningaanvragen en andere belangrijke databasetransacties hebben plaats gevonden.

Naast de technische realisatie van P-i-T recovery, moeten ook enkele procesmatige zaken worden geregeld. In dit ontwerp wordt volstaan met het aangeven van deze zaken. De daadwerkelijke realisatie hiervan valt buiten de scope van dit ontwerp en in dit FO worden nog geen voorstellen gedaan voor automatisering, een dergelijke functionaliteit zou in een latere versie toegevoegd kunnen worden:

- Verschillende betrokkenen moeten worden geïnformeerd:
    - Vergunningaanvragers, zodat ze hun ingediende vergunningaanvragen kunnen beoordelen op volledigheid na herstel.
    - Bevoegd Gezagen, zodat zij adequate maatregelen kunnen nemen om hun zaaksysteem weer gesynchroniseerd te krijgen met de herstelde database.

- De P-i-T recovery moet zodanig zijn ingericht dat deze aansluit bij de gangbare reactietijd op een calamiteit in de infrastructuur tot het moment dat een tijdelijke voorziening kan worden getroffen (bijv. het instellen van de “sorry”-pagina).

- Omdat het herstel van data centraal is geregeld, moet ook de coördinatie van de acties naar aanleiding van het databaseherstel centraal worden geregeld.

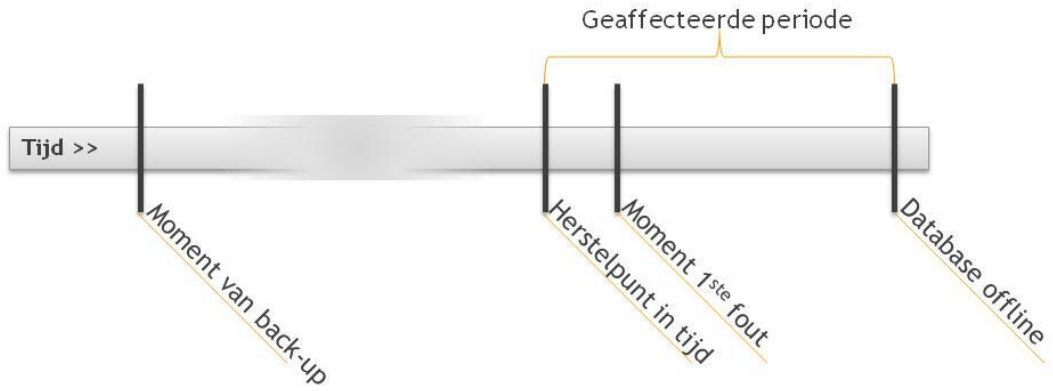
3 Checklist

Wanneer een situatie zich voordoet waarin gebruik wordt gemaakt van P-i-T recovery of wanneer de back-up van de database terug wordt gezet, dan moeten de volgende acties worden uitgevoerd:

1. Het Omgevingsloket moet tijdelijk offline worden gezet.
   Alle webapplicaties moeten offline worden gezet omdat deze allemaal gebruik maken van de zelfde database. Na de herstelwerkzaamheden, moeten de applicaties weer online worden gezet.

2. De betrokkenen moeten worden geïnformeerd.
   De betrokkenen zijn:
   a. alle vergunningaanvragers die in de geaffecteerde periode een vergunningaanvraag hebben ingediend of in behandeling hebben;
   b. alle bevoegd gezagen die betrokken zijn bij deze vergunningaanvragen; en
   c. alle betrokkenen die in de geaffecteerde periode aan een vergunningaanvraag zijn gekoppeld.

   De geaffecteerde periode betreft de tijd tussen het moment waar de database naar terug wordt gezet en het moment dat Omgevingsloket offline wordt gezet. De betrokkenen zijn te achterhalen op basis van een analyse van het betreffende recovery trail logbestand waarin deze zijn opgenomen. Dit is een handmatige actie.



3. De vergunningaanvragers moeten worden geïnformeerd over het feit dat er gegevens verloren zijn gegaan als gevolg van een calamiteit. Zij zullen de aanvraag die is ingediend opnieuw moeten indienen en bestanden die zijn geüpload opnieuw aanbieden. De bestanden gaan mogelijkerwijs niet meer te koppelen omdat de metagegevens van een bestand in de database staan. Als deze metagegevens verloren zijn gegaan, dan moet het bijbehorende bestand ook worden verwijderd.

4. De bevoegd gezagen en andere betrokkenen moeten worden geïnformeerd zodat ze de behandelsystemen die aansluiting hebben op het Omgevingsloket kunnen synchroniseren met de terug gedraaide vergunningaanvragen.
   Hiervoor moeten zij de beschikking krijgen over gegevens van de vergunningaanvragen die in de geaffecteerde periode zijn ingediend. Deze gegevens worden geregistreerd in het recovery trail logbestand. Daarnaast moet het bevoegd gezag geïnformeerd worden over de geaffecteerde periode. Dit

is een handmatige actie.

5. Het bevoegd gezag moet handelingen verrichten om een aantal zaken weer gesynchroniseerd te krijgen. De volgende zaken moeten door het bevoegd gezag zelf worden geregeld:
   a. Berichten die in de backoffice systemen zijn verwerkt, moeten ongedaan worden gemaakt.
   b. Documenten die in de periode zijn geüpload en vergunningaanvragen in de vorm van PDF-documenten moeten worden verwijderd.
   c. Indien nodig moeten Organisatiebeheer en Gebruikersbeheer worden geverifieerd op verlies van gegevens.
   d. Autorisatiegegevens moeten worden geverifieerd op volledigheid.

6. De database moet worden terug gezet tot het eerstvolgende punt waarvan bekend is dat de database correct en volledig is geconstateerd.

3.1 Benodigde voorzieningen

In deze paragraaf wordt aangegeven welke voorzieningen nodig zijn om bovenstaande acties uit te kunnen voeren. De volgende voorzieningen zijn nodig:

- Een onderhoudspagina die kan worden getoond tijdens de herstelwerkzaamheden.
  De standaard onderhoudspagina kan hiervoor worden gebruikt. Hier hoeven dus geen extra voorzieningen voor te worden gerealiseerd.

- Een informatiepagina waarin de situatie wordt beschreven, zodat een ieder zich kan informeren.
  Over deze pagina moet nog verder een beslissing worden genomen. Omdat een dergelijke pagina relatief eenvoudig is te realiseren en met name een redactioneel karakter heeft, valt deze pagina buiten de scope van dit ontwerp.

- Overzicht van de betrokkenen.
  Dit overzicht is samen te stellen op basis van het recovery trail logbestand. Dit is een voorziening die moet worden gerealiseerd. Het niveau van logging moet zodanig zijn dat snel een overzicht is samen te stellen van alle betrokkenen, bij voorkeur gegroepeerd per bevoegd gezag. Met dit overzicht kan een (lokale) beheerder de volgende acties uitvoeren:
  - Synchroniseren van de backoffice systemen.
  - Synchroniseren van de bestanden op de fileserver.

  Deze acties dienen vooralsnog handmatig te gebeuren. Voor volgende releases van OLO kan hierbij aan een geautomatiseerde oplossing worden gedacht. Dit valt voor nu verder buiten scope.

- Configuratie van de database.
  De database moet zodanig worden geconfigureerd dat P-i-T recovery mogelijk is.

4 Recovery trail logging

Recovery trail logging maakt gebruik van de logginginfrastructuur zoals beschreven in het ontwerp “Foutdetectie, -analyse en herstel op applicatieniveau”. De configuratie van de logginginfrastructuur is zodanig ingesteld, dat de transactie logregels in een apart recovery trail logbestand worden gezet. Door aan te sluiten op de logging infrastructuur kan dit logbestand op elke willekeurige locatie (bijv. een logging server) worden weg geschreven.

Op basis van dit logbestand moet het op een eenvoudige manier mogelijk zijn om de verschillende betrokkenen te achterhalen. Dit bestand moet met de nodige zorgvuldigheid worden behandeld omdat hier privacy gevoelige informatie in is vastgelegd.

De volgende informatie wordt in de vorm van logregels in dit logbestand vastgelegd:

[1] Het indienen van een aanvraag.
Wanneer een aanvraag wordt ingediend dan wordt dit vastgelegd in een of meerdere logregels. De informatie die wordt vastgelegd bevat ten minste de volgende gegevens:
   a. De identificatie van de aanvraag (id en naam/titel);
   b. De identificatie van de aanvrager (naam, e-mailadres);
   c. De identificatie van het bevoegd gezag (naam, e-mailadres).

[2] Het versturen van een bericht.
Elk bericht dat tijdens behandeling van een aanvraag wordt verstuurd, geeft een mogelijke wijziging in de backoffice systemen aan. Hier moet derhalve op gereageerd worden. De informatie die wordt vastgelegd in logregels is de volgende:
   a. De identificatie van de aanvraag (id en naam/titel);
   b. De identificatie van het bericht (type bericht);
   c. De identificatie van het bevoegd gezag (naam, e-mailadres).

[3] Het koppelen van een betrokkene aan een aanvraag.
Wanneer een betrokkene (bijv. een adviseur) aan een aanvraag wordt gekoppeld, dan wordt deze informatie vastgelegd als het gaat om een betrokkene die geen deel uit maakt van het Bevoegd gezag waarbij de vergunningaanvraag in behandeling is. De volgende informatie wordt vastgelegd:
   a. De identificatie van de aanvraag (id en naam/titel);
   b. De identificatie van de betrokkene (naam, e-mailadres);
   c. De identificatie van het bevoegd gezag (naam, e-mailadres).

De logregels die worden aangemaakt voldoen aan de beschrijving zoals aangegeven in het ontwerp “Foutdetectie…”. Op basis van dit ontwerp bestaat een logregel uit de volgende onderdelen:

- Datum en tijd van de gebeurtenis;
- Code (deze is nodig om de logregel identificerend te krijgen, zodat deze later kan worden gerouteerd naar het juiste logbestand)
- Omschrijving, waarin de gewenste informatie is opgenomen.

5 Instellen database

De database wordt zodanig geconfigureerd dat P-i-T recovery mogelijk is. Dit is standaard functionaliteit van MySQL.

De basisinstelling van P-i-T recovery is dat herstel tot elk moment tussen het moment van herstel en vier dagen terug in de tijd kan worden gegaan. Er wordt hier gekozen voor een periode van vier dagen omdat in deze periode in ieder geval geacht wordt een back-up te zijn gemaakt.

Bij de inrichting van de omgeving moet rekening worden gehouden met voldoende schijfruimte zodat het transactie logbestand kan worden opgeslagen. In het transactie logbestand zijn alle handelingen in de database gedetailleerd vastgelegd, zodat recovery mogelijk is.

5.1 Omvang van de logging

De omvang van het transactielogbestand is een minimaal een factor 2 en maximaal een factor 4 van de omvang van de database. Dit komt omdat de statements in het algemeen langer zijn dan de uiteindelijke data die in de tabellen komt.

Het onderstaande voorbeeld geeft hiervan een illustratie:

```
INSERT INTO LVOORGANISATI-
ON(ORGANISATIONID,NAME,CONTACTPERSOON,URL,EMAILADRES,VOOROVERLEGTOEGESTAAN,PUBLIC
ATIE,SYNCHRONISATIE,VERZENDMETHODE,ORGANISATIONTYPE,POSTADRES,BEZOEKADRES,TELEFOON
NUM-
MER,OMSCHRIJVING,TOELICHTING,FAXNUMMER,EMAIL_ALGEMEEN,VERANTWOORDELIJKE,CBS_NR,COM
MUNICATIEMETAANVRAGER,VOLLEDIGGEBRUIKLVO)

VALUES

('ppp','Provincie','CP van Drenthe', 'www.drenthe.nl', 'lvoprovincie@gmail.com', 'J', 'N', 'J', '1', 'provin-
cie', null, null, null, null, null, null, 'lvoprovincie@gmail.com', 'V van Drenthe', null, 'J', 'J');
```

Het gehele statement komt in de logfiles terecht, terwijl in de database alleen de daadwerkelijke waardes komen te staan. De totale benodigde opslag is afhankelijk van hoever er terug moet kunnen worden gegaan in de tijd. Als men één dag terug in de tijd wil gaan dan moet extra schijfruimte ter grootte van de dagelijks hoeveelheid toegevoegde gegevens maal een factor 3. Dus als op één dag 1Gb aan gegevens wordt toegevoegd, moet er minimaal 3Gb extra schijfruimte bij komen. Als men langer in de tijd terug wil, dan moet dit aantal worden vermenigvuldigd met het aantal dagen dat men terug wil (dus als je drie dagen terug wilt, moet er 3x3Gb = 9Gb extra schijfruimte bij komen).

Wanneer het logbestand wordt gecomprimeerd, dan wordt de omvang van het bestand 10 keer kleiner.

5.2 Richtlijnen

Verdere richtlijnen voor het instellen van P-i-T recovery zijn te vinden in de gebruikershandleiding van MySQL 5.0 (zie ook kader).

MySQL 5.0 reference manual

(http://dev.mysql.com/doc/refman/5.0/en/point-in-time-recovery.html)

Point-in-time recovery refers to recovery of data changes made since a given point in time. Typically, this type of recovery is performed after restoring a full backup that brings the server to its state as of the time the backup was made. (The full backup can be made in several ways, such as those listed in Section 6.2, “Database Backup Methods”.) Point-in-time recovery then brings the server up to date incrementally from the time of the full backup to a more recent time.

Point-in-time recovery is based on these principles:

The source of information for point-in-time recovery is the set of incremental backups represented by the binary log files generated subsequent to the full backup operation.

The mysqlbinlog utility converts the events in the binary log files from binary format to text so that they can be executed or viewed. mysqlbinlog has options for selecting sections of the binary log based on event times or position of events within the log.

Executing events from the binary log causes the data modifications they represent to be redone. This enables recovery of data changes for a given span of time. To execute events from the binary log, process mysqlbinlog output using the mysql client.

Viewing log contents can be useful when you need to determine event times or positions to select partial log contents prior to executing events.

Saving the output in a file is useful as a preliminary to executing the log contents with certain events removed, such as an accidental DROP DATABASE. You can delete from the file any statements not to be executed before executing its contents.