

Dataconversie met Oracle Spatial

Realworld klantendag

Jeffrey Janssen



19 september 2013

Voorstellen



- ① Jeffrey Janssen
- ② Computer Science & Engineering (TU/e)
- ③ Realworld Systems
- ④ Datamigraties
 - Alliander
 - Stedin

Agenda



1 Architectuur

- Inleiding
- Ontwerp migratie

2 Rapportage

- Ontwerp logging
- Rapportages

3 Performance

- Do's en don'ts
- Voorbeeld

4 Ervaringen

- Ervaringen
- Vergelijking met FME

Datamigratie



- 1 Overzetten van gegevens van een of meerdere bronsystemen naar een doelsysteem
- 2 Combineren en/of splitsen van brongegevens
- 3 Bewerkingen op alfanumerieke gegevens
 - Eenheden, schrijfwijzen, standaardwaarden
 - Mapping
- 4 Bewerkingen op ruimtelijke gegevens
- 5 Aantonen van volledigheid en kwaliteit

Mapping



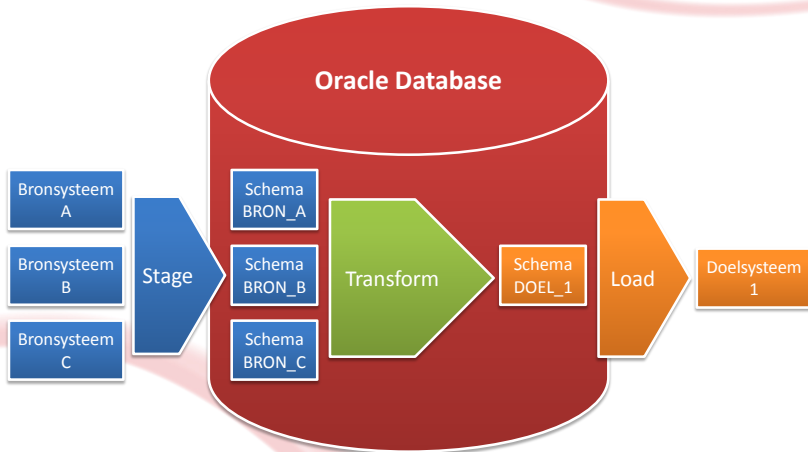
- ① Mapping van attribuutwaarden (types, keuzelijsten)
- ② Excelbestanden in te vullen door gebruikers
- ③ Vaste prefixen voor invoer- en uitvoerkolommen
- ④ Package met functiegenerator
 - Per uitvoerkolom één mappingfunctie
 - Invoerkolommen zijn de parameters
- ⑤ Mappingfuncties
 - Opzoeken aan de hand van bronwaarden
 - Melding en standaardwaarde als bronwaarde niet bestaat

Uitgangspunten



- 1 Onderhoudbaar
- 2 Flexibel
- 3 Herhaalbaar
- 4 Geautomatiseerd
- 5 Goede performance

Migratieschema



Ontwerp migratie



① Stage

- Inlezen brondata in Oracle
- “Read only”

② Transform

- Converteren brondata naar datamodel doelsysteem
- Logging en rapportages

③ Load

- Geconverteerde data inlezen in doelsysteem
- Logging en rapportages

Logging



- ① Tellingen op bron- en doeltabellen
- ② Traceability van records
- ③ Controle op volledigheid
- ④ Registratie
 - Relaties
 - Uitval
 - Warning
- ⑤ Rapportages
 - Balans
 - Overzicht
 - Warning/uitvallijst

Balans



- ① Telling op brontabel, relaties en uitval
- ② Geeft transformatie weer
- ③ Opgebouwd uit combinaties van bron- en doeltabellen
- ④ Verantwoording van *alle* bronrecords

LS Kasten		
829	[INSTAP_TRANSFORM]i_lsp_aansluitkast_m	533 [EGIS]e_kast
748	[INSTAP_TRANSFORM]i_lsp_verdeelkast_m	818 Errors:[INSTAP_TRANSFORM]i_lsp_aansluitkast_m ⇒ [EGIS]e_kast
8,442	[TKV] dtk_ruimten	226 Errors:[INSTAP_TRANSFORM]i_lsp_verdeelkast_m ⇒ [EGIS]e_kast
10,019	o	8,442 Errors:[TKV] dtk_ruimten ⇒ [EGIS]e_kast
		10,019
LS/OV Moffen		
3,760	[INSTAP_TRANSFORM]i_lsp_aftaklas_m	18,380 [EGIS]e_kv04_ov_mof
3,942	[INSTAP_TRANSFORM]i_lsp_eindlas_m	
102	[INSTAP_TRANSFORM]i_lsp_looddop_m	
6	[INSTAP_TRANSFORM]i_lsp_schutmof_m	
10,570	[INSTAP_TRANSFORM]i_lsp_verbindingslas_m	
18,380	o	18,380

Overzicht



- ① Telling op relaties en doeltabel (voor en na)
- ② Controle of alle records ook daadwerkelijk zijn toegevoegd
- ③ Opgebouwd uit lijst doeltabellen

Target table	Parent	Inserted	Count	Expected	Diff
[EGIS]_e_aardkabel	○	275	275	275	○
[EGIS]_e_afdekplaat	○	1,442	1,442	1,442	○
[EGIS]_e_afdekplaat_inhoud	○	3,827	3,827	3,827	○
[EGIS]_e_anode	○	1	1	1	○
[EGIS]_e_anode_e_aarddraad	○	1	1	1	○
[EGIS]_e_bescherming	○	3,497	3,497	3,497	○
[EGIS]_e_bescherming_comp	○	9,758	9,758	9,758	○
[EGIS]_e_bescherming_comp_inhoud	○	8,356	8,356	8,356	○
[EGIS]_e_beveiliging_asri_hs	○	276	276	276	○
[EGIS]_e_beveiliging_hs	○	1,718	1,718	1,718	○
[EGIS]_e_beveiliging_trafo_hs	○	10	10	10	○
[EGIS]_e_diepte	○	296	296	296	○
[EGIS]_e_doorvoeren_isolatoren	○	76	76	76	○
[EGIS]_e_drukluchtinstallatie_hs	○	25	25	25	○

Warning/uitvallijst



- ① Overzicht van warning- en uitvalmeldingen
- ② Gegroepeerd per reden, bron- en doeltabel

Uitval Transform

Source table	Target table	Code	Description	Count
[INSTAP_TRANSFORM] i_hsp_doorsnede_m	[EGIS]_e_dwarsprofiel	NIEDP1_U02	Record is een vlak; alleen de lijnen worden gemigreerd	1,752
[INSTAP_TRANSFORM] i_msp_doorsnede_m	[EGIS]_e_dwarsprofiel	NIEDPM_U01	Geometrie moet uit minstens 5 coördinaten bestaan om een dwarsprofiel te genereren	34
[INSTAP_TRANSFORM] i_msp_doorsnede_m	[EGIS]_e_dwarsprofiel	NIEDPM_U02	Doorsnede heeft geen bakje op een uiteinde	49
[INSTAP_TRANSFORM] i_lsp_doorsnede_m	[EGIS]_e_dwarsprofiel	NIEDPM_U03	Richting van de doorsnede kan niet worden bepaald	1
[TKV]_dtk_ruimten	[EGIS]_e_kast	TIEKAM_A01	Samengevoegd	522
[TKV]_dtk_ruimten	[EGIS]_e_kast	TIEKAM_A02	Samengevoegd	11
[INSTAP_TRANSFORM] i_lsp_verdeelkast_m	[EGIS]_e_kast	TIEKAM_U01	Verdeelkast wel in Instap, niet in TKV	1
[TKV]_dtk_ruimten	[EGIS]_e_kast	TIEKAM_U02	Verdeel/aansluitkast niet in Instap, wel in TKV	6,964
[INSTAP_TRANSFORM] i_lsp_verdeelkast_m	[EGIS]_e_kast	TIEKAM_U03	Verdeelkast heeft geen kastnummer in Instap	225
[TKV]_dtk_ruimten	[EGIS]_e_kast	TIEKAM_U04	Record is geen kast in TKV	945
[INSTAP_TRANSFORM] i_lsp_aansluitkast_m	[EGIS]_e_kast	TIEKAM_U13	Aansluitkast heeft geen kastnummer in Instap	818
[TKV]_dtk_rekken	[EGIS]_e_kvo4_verdeelrek	TIEKRM_U01	Kast is uitgevallen; rek valt ook uit	7,762
[TKV]_dtk_groepen	[EGIS]_e_kvo4_veld	TIEVDM_U01	Rek is uitgevallen; veld valt ook uit	56,837

① Tussentabellen maken

- Langzame query met veel joins is vaak sneller als deze opgesplitst wordt
- Apart filteren op geometrie en attributen soms noodzakelijk vanwege Spatial Index

② Bij ingewikkelde bewerkingen op geometrie werken met coördinaten of vertices (`MDSYS.VERTEX_[SET_]TYPE`)

- `SDO_UTIL.GETVERTICES()`

Don'ts



- 1 **UPDATE** statements met join op grote tabellen
 - Behalve bij join op primary key; (**UPDATE SELECT**)
 - Extra tussentabel maken voor **UPDATE**
- 2 Spatial Join tussen twee tabellen met veel records vermijden indien mogelijk

Probleemstelling



- 1 Tabel met veel punt-geometrieën (± 2 miljoen)
- 2 Zoek welke objecten dichter dan 1 cm bij elkaar liggen
 - (Deze daarna verplaatsen zodat er geen punten > 1 cm van elkaar af liggen)
- 3 `SDO_WITHIN_DISTANCE()` - Doorlooptijd 2:19 uur

Probleemstelling



- 1 Tabel met veel punt-geometrieën (\pm 2 miljoen)
- 2 Zoek welke objecten dichter dan 1 cm bij elkaar liggen
 - (Deze daarna verplaatsen zodat er geen punten $>$ 1 cm van elkaar af liggen)
- 3 `SDO_WITHIN_DISTANCE()` – Doorlooptijd 2:19 uur
- 4 `SDO_NN()` – Doorlooptijd 1:15 uur

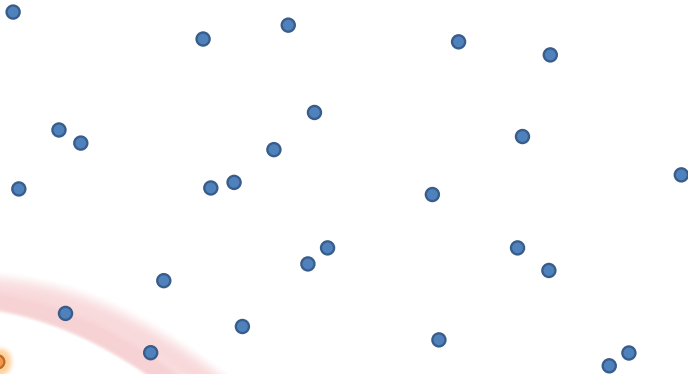
Alternatieve oplossing



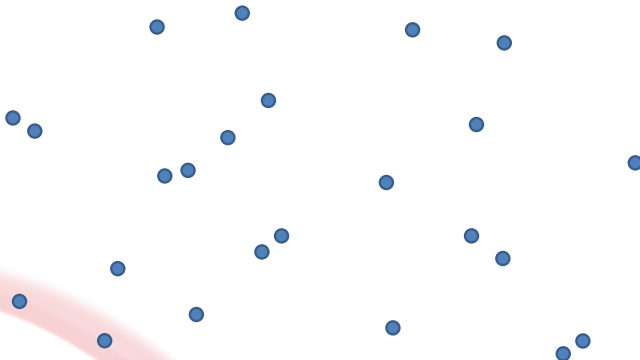
“Sliding Window” methode:

- 1 Haal alle objecten op, inclusief coördinaat
- 2 Sorteer op x en itereer
- 3 Vergelijk het “huidige” object met de objecten die (qua x-coördinaat) er maximaal 1 cm vanaf liggen
- 4 Bereken de exacte afstand

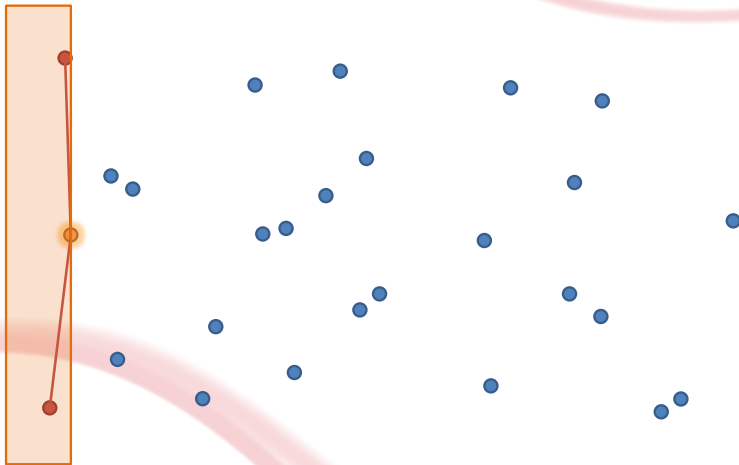
Sliding Window



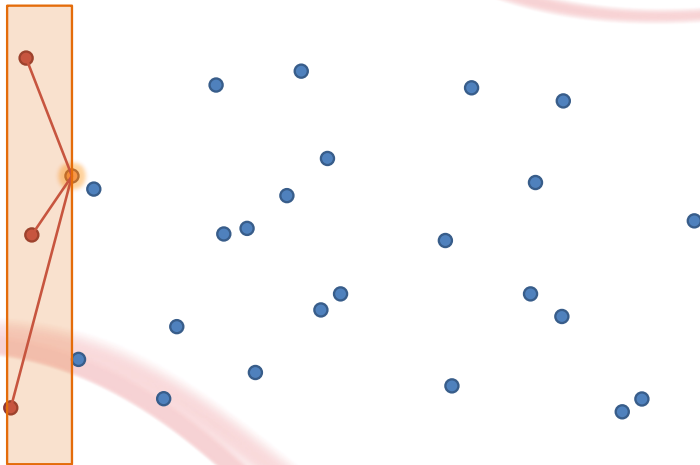
Sliding Window



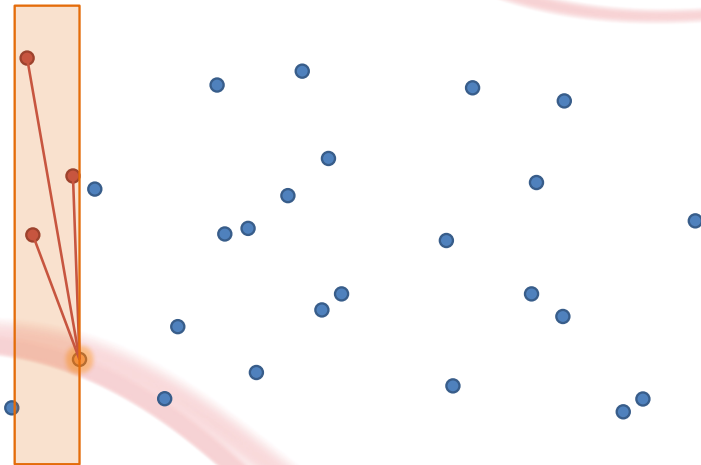
Sliding Window



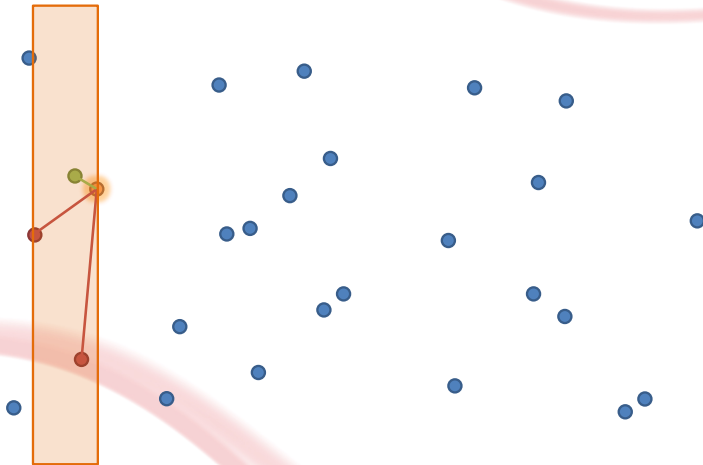
Sliding Window



Sliding Window



Sliding Window



Resultaat



- 1 Alleen berekenen van de exacte afstand tussen een (klein) aantal kandidaten
- 2 Snel: 12 minuten doorlooptijd
- 3 Verplaatsen en zoeken van nieuwe plek is mogelijk in dezelfde loop
- 4 Geheugenintensief omdat de hele set in het geheugen geladen moet worden
- 5 Script ook bruikbaar voor andere problemen:
 - Bepalen welke kabels aan elkaar gekoppeld zijn (gebruik begin- en eindcoördinaat)
 - Puntengroepen bepalen en de middelste bepalen

Uitgangspunten



- ① Onderhoudbaar
- ② Flexibel
- ③ Herhaalbaar
- ④ Geautomatiseerd
- ⑤ Goede performance

Algemeen



- ① **Betere performance in Oracle**
 - Data hoeft niet ingelezen en weggeschreven te worden (over het netwerk)
- ② **Complexiteit en onderhoudbaarheid ongeveer gelijk**
 - In Oracle directere koppeling tussen complexiteit van script en operaties
- ③ **In Oracle zijn meerdere geometriekolommen in een tabel mogelijk**

Operaties

- 1 Joins lastig in FME; FeatureMerger is vaak traag
- 2 Ruimtelijke relaties beter te leggen in Oracle
- 3 Opbouw netwerktopologie en tracing beter in Oracle
- 4 Mappingfuncties (vooral op meerdere kolommen) makkelijker in Oracle
- 5 Uitval in FME komt vaak al uit de transformer, niet apart registreren zoals in Oracle
- 6 `SDO_GEOM.SDO_BUFFER()` is beperkt in Oracle (alleen ronde hoeken)

Debugging



- ① Data in FME is “vluchtig”, tussentabellen in Oracle blijven bestaan
 - Bij debuggen extra logging toevoegen in FME
- ② Individueel uitvoeren van losse SQL statements in plaats van complete script
- ③ FME heeft een viewer, handig om data te bekijken bij het debuggen

Debugging



- ① Data in FME is “vluchtig”, tussentabellen in Oracle blijven bestaan
 - Bij debuggen extra logging toevoegen in FME
- ② Individueel uitvoeren van losse SQL statements in plaats van complete script
- ③ FME heeft een viewer, handig om data te bekijken bij het debuggen
 - Viewer gebruiken in combinatie met Oracle

Einde presentatie

Vragen?