

Thema 2.2 Software Engineering - Network & Security Engineering : IT Infrastructures

Casusbeschrijving UNWDMI

Achtergrond

Dat het wereldklimaat in de laatste tientallen jaren aanzienlijke veranderingen ondergaat, is genoegzaam bekend. Over de oorzaken wordt weliswaar onder meteorologen nog gediscussieerd, maar de feiten dienen zich onmiskenbaar aan. Gevolgen van de klimaatverandering zijn extremere weerscondities in de vorm van superstormen, overstromingen en perioden van langdurige droogte. Deze condities gaan vaak gepaard met menselijk leed. De Verenigde Naties als organisatie richt zich sinds haar oprichting in 1949 op het voorkomen van menselijk lijden, in eerste instantie door de wereldvrede te waarborgen, maar later ook door ontwikkelingshulp, onderwijs en het behoud van cultuurgood te stimuleren. Vanuit haar primaire doelstelling heeft de VN besloten zich sterk te maken om de gevolgen van klimaatverandering tegen te gaan. Enerzijds biedt de VN noodhulp aan getroffen gebieden, anderzijds wil de VN een hoofdrol spelen in de wereldwijde informatievoorziening over de weerstoestand, zodat lokale weerdiensten snel weeralarmen kunnen afgeven, als gevolg waarvan de negatieve gevolgen van extreme weerscondities voor de lokale bevolking tegengegaan kunnen worden. Ten behoeve van de deze informatievoorziening is onlangs het United Nations Weather Data Management Institute (UNWDMI) opgericht.

Het United Nations Weather Data Management Institute

Het UNWDMI heeft zich tot doel gesteld gegevens over de wereldwijde toestand van het weer te verzamelen, op te slaan en beschikbaar te stellen aan belangstellenden, zoals lokale weerdiensten, commerciële weeradviesbureaus en overige partijen die belang hebben bij weerinformatie in de breedste zin van het woord. De organisatie heeft zijn hoofdkantoor in Groningen. Deze vestigingsplaats mag opzienbarend heten, maar is het gevolg van effectief lobbywerk door de Nederlandse regering. Bovendien biedt Groningen door een uitstekende Internet-infrastructuur een belangrijke voorwaarde voor een organisatie van een dergelijke statuur.

Het UNWDMI is sinds een paar jaren operationeel. Er is eerst een organisatie in de steigers gezet, een bedrijfsgebouw en commitment van alle leden van de VN om van dit instituut een succes te maken. Het gevolg hiervan is dat alle VN-leden hebben toegezegd dat hun lokale weerstations hun gegevens voortaan doorseinen naar het UNWDMI in Groningen en dat hun weerdiensten hun informatie voortaan in Groningen betrekken. Het eigendom en beheer van de lokale weerstations blijft in handen van de verschillende weerdiensten.

De organisatie van het instituut is tamelijk overzichtelijk. Gelet op de kritiek vanuit vooral de USA op de soms geldverslindende VN-instellingen is gekozen voor een eenvoudige opzet van de organisatie. Bovendien wordt het instituut geacht zijn kosten te dekken door inkomsten uit de dienstverlening. Het UNWDMI wordt geleid door een managementteam, bestaande uit de algemeen directeur en vier afdelingshoofden.

Het instituut kent de afdelingen Data Acquisition, ICT Services, Application Development & Maintenance en Service Management.

- De afdeling Data Acquisition onderhoudt contacten met de eigenaren van lokale weerstations en is verantwoordelijk voor een adequate stroom van weergegevens vanuit deze weerstations.
- De afdeling ICT Services richt zich op het beheer van de technische infrastructuur om de dienstverlening aan de klanten van UNWDMI vorm te geven.
- De afdeling Application Development & Maintenance is verantwoordelijk voor de ontwikkeling en het onderhoud aan de applicaties die nodig zijn om weergegevens op te slaan en beschikbaar te stellen aan de klanten.
- De afdeling Service Management onderhoudt contacten met de klanten van het UNWDMI, stelt in overleg met klanten specificaties op van de klantvraag en sluit contracten af met klanten.

Deze afdelingen worden terzijde gestaan door een stafbureau waar standaardzaken als financieel management en personeelsmanagement ten uitvoer worden gebracht. Ook treedt het stafbureau op als projectbureau. In totaal zijn er slechts 90 mensen werkzaam op het instituut. In het onderstaande overzicht valt te lezen hoe de verdeling van de medewerkers over de afdelingen is.

<i>Afdeling</i>	<i>Aantal medewerkers</i>
Directie	1
Data Acquisition	10
ICT Services	45
Application Development & Maintenance	12
Service Management	10
Stafbureau	12

Klanten en diensten

De belangrijkste dienst die het UNWDMI levert is het beschikbaar stellen van informatie over de wereldwijde toestand van het weer. De weergegevens zijn afkomstig van lokale weerstations, die eigendom zijn van lokale weerdiensten. Deze gegevens worden door UNWDMI via datacommunicatieverbindingen verzameld en, via een getrapte structuur, in een centrale omgeving opgeslagen. Door middel van een (web)applicaties kunnen de klanten een selectie uit deze gegevens opvragen. De ontvangstinfrastructuur en de klantapplicaties worden ontwikkeld en onderhouden door het UNWDMI. De klant geeft in het servicecontract met het instituut op welke informatie hij uit de database wil onttrekken en welke kwaliteitsnormen daarbij moeten gelden. Het contract bevat eveneens een tariefstelling, die door commerciële klanten rechtstreeks aan het UNWDMI voldaan moet worden en die voor landelijke weerdiensten verrekend wordt in het contributietarief voor de VN. Looptijd van een servicecontract is doorgaans één jaar.

Klanten zijn naast alle landelijke weerdiensten commerciële weeradviesbureaus, onderzoeksinstituten en andere partijen die belang hebben bij weerinformatie. In

totaal verwacht UNWDMI ongeveer duizend servicecontracten te kunnen afsluiten. De jaaromzet bedraagt € 8 miljoen.

In de afgelopen jaren is er een toenemende behoefte aan “micro”-metingen; micrometingen betekent zoveel mogelijk meetpunten per land om heel nauwkeurig het weer in beeld te kunnen brengen. Daar waar in het begin sprake was van 5000 weerstations, is dit aantal inmiddels gegroeid naar 100.000. Er is ambitie om door te groeien naar 1 tot 5 miljoen weerstations, om dus zo lokale weersverschijnselen en – verschillen zo goed mogelijk in kaart te brengen.

Aangezien de naam UNWDMI wat lastig bekt is het ondertussen gemeengoed geworden om als alternatief UMI te hanteren.

Verzameling, opslag en beschikbaarstelling van gegevens

De weergegevens zijn afkomstig van een zeer uitgebreid netwerk van in totaal 100.000+ weerstations, die wereldwijd zijn opgesteld in een grid. Het aantal is volop in ontwikkeling en neemt snel toe. Deze stations sturen 24 uur per dag en 7 dagen per week elke seconde een aantal gegevens naar Groningen zoals:

- Identificatiecode van het station
- lokale datum en tijd
- weergegevens: temperatuur, dauwpunt, luchtdruk, zichtbaarheid, neerslag, sneeuwdiepte, bewolking, windrichting, windsnelheid, gebeurtenissen.

De identificatiecode is voor elk van de weerstations uniek. De gegevens van één enkele meting worden opgestuurd in de vorm van een XML-bestand.

Bij het meten en verzenden van de meetwaarden gaat er nog wel eens wat mis. Zeker in afgelegen gebieden op de wereld kunnen er storingen in de weerstations optreden, die niet zo snel verholpen kunnen worden. Als gevolg daarvan kunnen meetwaarden soms irreëel zijn of zelfs ontbreken. Daarom vindt er in de systemen in Groningen een controleslag plaats voordat de meetgegevens opgeslagen worden in de centrale database. Het systeem gaat daarbij als volgt te werk.

- Indien één of meer meetwaarden ontbreken, worden ze door het systeem berekend door middel van extrapolatie van de dertig voorafgaande metingen. Dit komt ongeveer in 1% van alle gevallen voor.
- Een meetwaarde voor de temperatuur wordt als irreëel beschouwd indien ze 20% of meer groter is of kleiner is dan wat men kan verwachten op basis van extrapolatie van de dertig voorafgaande temperatuurmetingen. In dat geval wordt de geëxtrapoleerde waarde $\pm 20\%$ voor de temperatuur opgeslagen. Voor de andere meetwaarden wordt deze handelswijze niet toegepast.

De hier beschreven handelswijze is volkomen geaccepteerd in de wereld van de meteorologie. Alle instellingen die weergegevens opslaan passen dit systeem toe en alle gebruikers van weerinformatie weten dat dit systeem toegepast wordt en nemen daar genoegen mee. Dat geldt niet alleen voor landelijke weerdiensten, maar ook voor commerciële weeradviesbureaus en onderzoeksinstituten.

Na de correctie worden de gegevens opgeslagen in een centrale databaseomgeving. Het logisch databasemodel ziet er als volgt uit:

station (stn, name, latitude, longitude, country, timezone)
latitude en longitude bepalen de positie van het weerstation
timezone is een aanduiding van de tijdzone waar het weerstation zich bevindt
in het geval een weerstation zich op zee bevindt, is country gelijk aan NULL

timezone (timezone_id, gmt_offset, dst)
UT is Universal Time en is gelijk aan de Nederlandse tijd minus 1 uur

measurement Deze kun en dien je zelf af te leiden uit de te ontvangen gegevens.

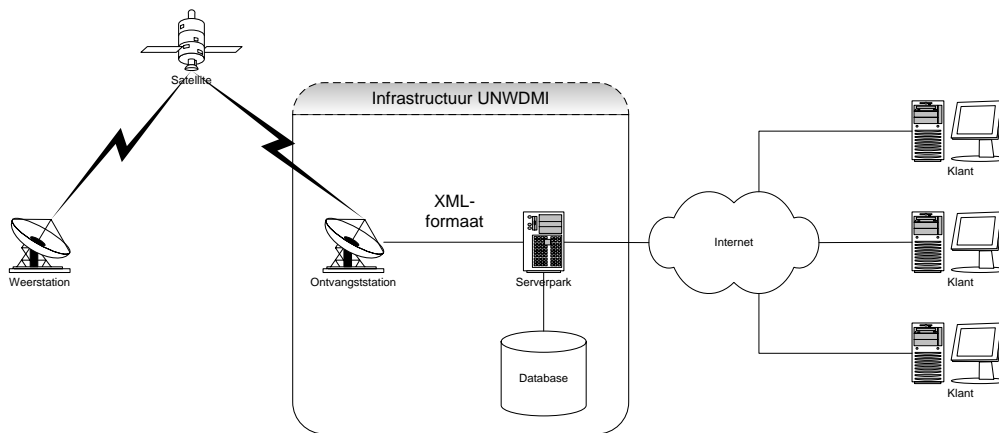
We merken op dat als gevolg van de schaalgrootte van de database het niet noodzakelijk is dat de tabellen elk als één bestand op de schijfeenheden geïmplementeerd zijn. Hoe dat moet, is nog niet uitgewerkt.

Op basis van een servicecontract kan een klant informatie ontsluiten uit deze database. Het contract beschrijft onder andere:

- met welke frequentie een klant informatie mag ophalen: bijvoorbeeld elke 24 uur, elk uur of altijd. Bij een beperkte frequentie wordt ook een tijdstip in het contract vastgelegd.
- welke selecties van gegevens een klant wenst. Hieraan zijn geen beperkingen gesteld. Voorbeelden zijn:
 - alle weergegevens van alle weerstations in een bepaald land gedurende de afgelopen 24 uur;
 - alle weergegevens op een bepaald tijdstip van alle weerstations op een bepaalde breedtegraad;
 - alle weergegevens van alle weerstations op elk vol uur in de afgelopen week.
- welke kwaliteits- of servicenormen er van toepassing zijn met betrekking tot de beschikbaarheid, informatie-integriteit en responstijd van de beschreven ICT-dienst.
- het tarief voor de dienst die in het contract beschreven is.

Technische architectuur

De technische infrastructuur, die is gestoeld op een maximale capaciteit van 100.000 weer stations, is een aantal jaar geleden ontworpen. Het systeem begint echter tegen limieten aan te lopen.



figuur 1: de technische architectuur

Probleemstelling

De UMI is nu een paar jaar operationeel. De limieten aan de schaalbaarheid van het ene centrale systeem komen op het gebied van CPU-kracht en ook opslagsnelheid in zicht. Opslag en verwerking in de cloud is om meerdere, plausibele, redenen door het bestuur van de UMI afgewezen. Er moet dus aan een andere, haalbare oplossing worden gewerkt. Hiervoor zijn meerdere alternatieven.

Een alternatief dat overwogen wordt is om de werkload van alle stations te verdelen over meerdere “lichtgewicht” systemen. Lichtgewicht betekent: beperkte rekenkracht, maar ook zeer beperkt stroom; groen dus.

Uiteraard wil het UMI eerst een proef doen om te zien of dit een haalbaar alternatief is. Bij de proef worden de weerstations in groepen van 8000 verdeeld. Elke lichtgewicht server moet dan alle verwerking van deze 8000 stations doen, en dus ook de opslag. Aangezien de lichtgewicht server zelf maar zeer beperkte opslag heeft moet dit gebeuren op een andere, centrale, opslagserver. Tijdens het werken aan de opdrachten krijg je hiervoor een (centrale) testserver toegewezen. Tussen het lichtgewicht en het centrale systeem is een netwerk aanwezig, zodat de gegevens tussen beide systemen verzonden kunnen worden.

Jullie hulp ingeroepen bij twee onderdelen:

- het ontwikkelen van de applicatie die in staat is om 8000 weerstations te ontvangen (volledige, realtime afhandeling) op de lichtgewicht server. Het is opgelegd dat dit in Java gebeurt.
- De inrichting van beide servers gecombineerd, de manier waarop de centrale opslag geregeld is en de manier waarop beide servers met elkaar communiceren (fileshares/rsync/etc) staat open. Voor de proef krijg je één virtuele server toegekend en één lichtgewicht server. Uiteraard dien je ook de juiste JVM op het lichtgewicht systeem in te richten.
- Data die verzonden wordt over het Internet dient versleuteld te zijn om te voorkomen dat derden de data gaan manipuleren of dat ze de data afvangen en geen licentiekosten betalen
- Zowel de lichtgewicht server als de opslag server dienen alleen maar de noodzakelijke processen te runnen die nodig zijn om de dienst te realiseren (dit met het ook op security). Er mag geen ongeoorloofde toegang tot de systemen mogelijk zijn.