



Netezza verovert de wereld

DWH performance in top gear

Hans Lamboo

Sinds enig tijd wordt de databasewereld geconfronteerd met de nieuwe technologie van zogenaamde datawarehouse appliances. Feitelijk zijn deze appliances database, server en storage in één eenvoudig te bedienen 'black box' met verbluffende prestaties, vrijwel oneindige schaalbaarheid en nauwelijks management. En dat alles tegen een vriendelijke prijs. Een van de succesvolle leveranciers is Netezza. "Just give me power and an IP-address."

Er komen vanuit verschillende kanten datawarehouse appliances op de markt. Naast Netezza noemen we DATAlegro, HP's Neoview en eigenlijk hoort ook oudgediende Teradata tot deze groep. Een markt van supergeoptimaliseerde en gecustomiseerde database servers, zonder verdere tooling. Een nieuwe trend dus.

De klant vertelt

In Londen werd voor de tweede keer een Europese Summit georganiseerd door Netezza (spreek uit nuh-tieh-zah). Opvallend waren de glunderende gezichten en de enthousiaste verhalen, en dan niet alleen die van de sales en de marketing manager. Nee, het waren juist de vele user story's van gebruikers uit diverse Europese landen die getuigden van Netezza's succes. Verhalen over performance, ook met de meest ingewikkelde query's, die soms 500 keer sneller was dan in het pre-Netezza tijdperk. Een Terabyte extra? Volgens een Engelse gebruiker een kwestie van een of meer Netezza Performance Servers (NPS) erbij zetten en inpluggen. Twee Britse telecombedrijven en één uit Slowakije.

De eenvoudige interface is bedoeld voor querying en connections

En uit Nederland? Ja, ook uit Nederland: Flora Holland, de grootste bloemenveiling van Nederland, die dit jaar nog zal fuseren met de Aalsmeerse veiling. Ook een Nederlandse Netezza-gebruiker maar niet aanwezig op de summit, is Center Parcs. De Netezza Performance Server ziet er sober uit, strak vormgegeven zonder poespas. Aan de voorzijde bevindt zich de eenvoudige interface, aan de achterzijde de connectoren en de power-aansluiting. That's it. In de presentaties noemt men perfor-

mances die wel 10 tot 100 keer beter zijn dan bijvoorbeeld Oracle, een veelgehoorde competitor. Amazon.com, een van de grote Amerikaanse gebruikers, had voorheen op hun Oracle DWH 20 uur nodig om Identify & Confirm patronen te ontdekken. Met de NPS duurt deze query slechts twee uur, zo melden zij. Nog een testimonial: Ahold produceert met de NPS 133 procent meer analyserapporten voor 20 procent meer eindgebruikers en heeft daarvoor bijna vier uur nodig; de oude configuratie met CDW/Oracle deed daar 72 uur over. Het Amerikaanse Catalina doet data mining projecten voor grote retailers. Sinds men Red Brick heeft vervangen door de NPS doet men 'gewoon' mining query's op een datawarehouse van 100 Terabyte, zo verhalen zij. Indrukwekkend.

Performance in top gear

Wat is het geheim van Netezza? De door CEO en oprichter Jit Saxona ontwikkelde technologie maakt gebruik van een aantal standaardcomponenten die op gepatenteerde wijze zijn samen-

Enige specificaties Netezza

Kleinste eenheid NPS 10025: 28 SPU's, 3 TB.

Grootste eenheid NPS 10800: 896 SPU's, 100 TB.

OS: RedHat Linux.

API's: SQL, ODBC, JDBC.

SQL Standards: SQL-92 compliant.

High-speed load/unload met ETL- en EAI-tools tot 500 GB per uur.

Backup & restore: Interoperable met Legato, Tivoli en Veritas tot 2 TB per uur.

DB Portability: IBM DB2, Informix, MS SQL Server, MySQL, Oracle, Red Brick, Sybase IQ, Teradata.

gevoegd in een SPU, een Snippet Processing Unit. De unit ziet eruit als een soort slede, die aan de achterzijde is voorzien van een disk van 400 GB, opgedeeld in drie partities: 100 GB system software, 150 GB primaire dataopslag en 150 GB voor een kopie van een willekeurige andere datapartitie. Vrijwel tegen de disk aan zit een Field Programmable Gate Array (FPGA), de processor en de overige circuits. Wat met name opvalt is de toepassing van

Stekker in het stopcontact, netwerk aansluiten en het werkt

de FPGA, immers een technologie die gebruikt wordt in auto's, DVD-spelers, video-camera's en andere plaatsen waar streaming data voorkomen – tot nog toe niet in PC's of servers. Door verschillende SPU's samen te voegen – bij de NPS is de kleinste eenheid 28 – kan Massive Parallel Processing (MPP) worden toegepast.

"De FPGA is een van de sleutels tot ons geheim", beaamt CTO Justin Lindsey. "Samen met de MPP zorgt dat voor geweldige performance. Alle data worden gelijktijdig en gelijkmatig verdeeld

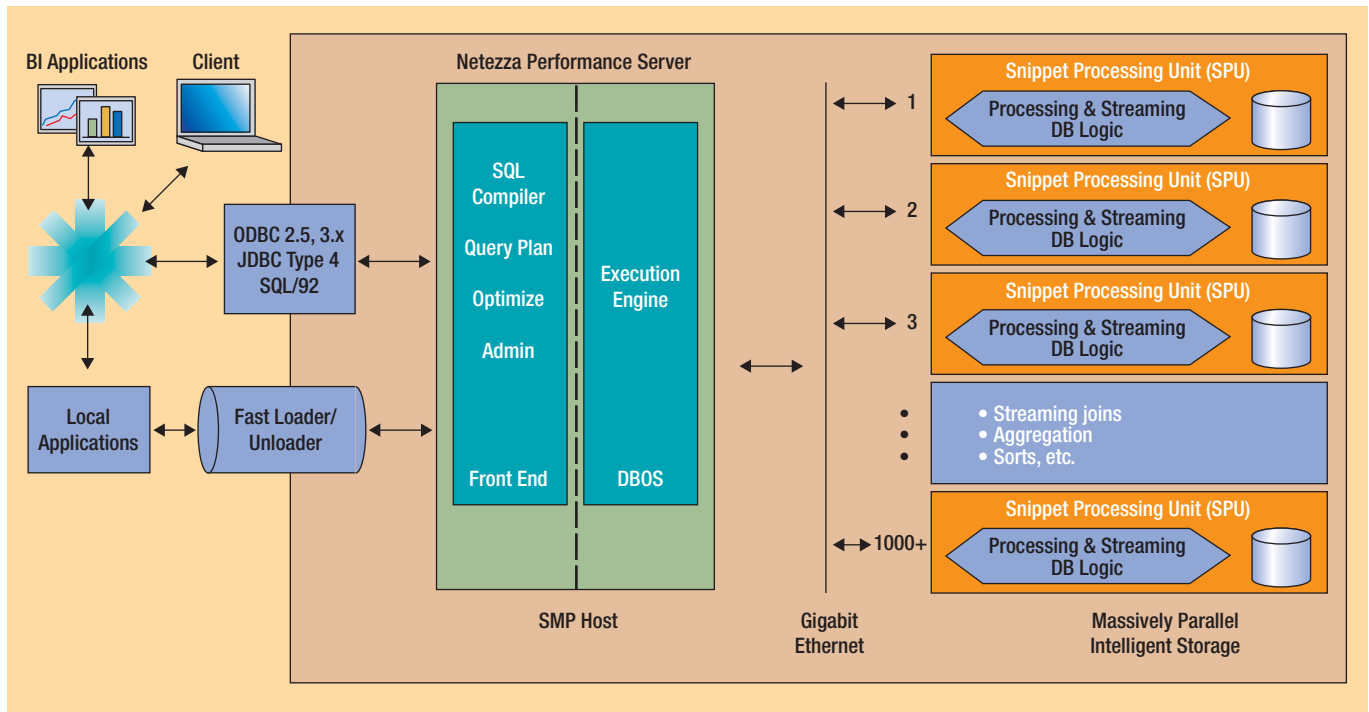
over de disks in de relationele database, gebaseerd op Ingres. Deze is van nature paralliseerbaar; daardoor kan een gebruiker op 50 records werken, en een ander tegelijkertijd op 50 andere records. Staan de data eenmaal waar ze moeten staan, dan worden ze gefilterd door de FPGA. Drives zijn per definitie traag. Dat vangen we op door data snel 'op' de disk te filteren, waardoor er veel minder data naar de CPU's gaan – precies de bottleneck in DWH-omgevingen. Zo verliezen we dus helemaal geen tijd." Lindsey vervolgt: "In de FPGA hebben we allerlei database-logica ondergebracht. Bij elke query parametriseren we bepaalde waarden en zenden die naar de FPGA. Een SQL query komt binnen, wordt geoptimized en omgezet in een werkplan waarin een verzameling snippets zit. Dit gebeurt bij elke query opnieuw. Je zou kunnen zeggen dat we de hardware voor elke query optimaliseren."

Dat klinkt dus allemaal als pure hardware-technologie, maar dat wil Lindsey graag nuanceren. "De intelligentie van de FPGA is ingebouwd op aanwijzingen van onze zeer ervaren database-experts. Daarin schuilt zeker een belangrijk stuk van het geheim: niet alleen de toepassing van de FPGA, maar ook de wijze waarop we dat doen." De processor en de FPGA zitten bijna tegen de disk aan. De data bevinden zich dus direct naast de processor en fysiek in hetzelfde circuit, hetgeen een extra performance-winst geeft.



Performance Management Strategy	Data warehouse development	Reporting and analysis
<ul style="list-style-type: none">- Design strategic management model- Metrics definition- Performance Management Roadmap	<ul style="list-style-type: none">- Architecture design- ETL design- Data Quality Assurance- ETL Development	<ul style="list-style-type: none">- Reporting- Dashboard development- Data mining and analysis
		

www.quintica.nl



Afbeelding 1: Netezza Performance Server.

Eenvoudig

De kleinste NPS bevat 28 SPU's (inclusief twee hot-spares) en is goed voor ruim 3 TB aan data. Zoals gesteld kan de grootste NPS 100 TB aan, met 896 SPU's. Zo eenvoudig de interface en zo strak de buitenzijde zijn, inwendig is de NPS dus technologisch zeer gecompliceerd. In zijn presentatie vergelijkt Lindsey de NPS met het gebruik van een koelkast. "Stekker in het stopcontact, netwerk aansluiten en het werkt. Op de meeste plaatsen waar we NPS installeren, heeft de machine alleen een IP-adres nodig. Vaak is de NPS binnen een paar uur al helemaal gebruiksklaar, soms voor het IP-adres bekend is. Het motto is dus: give me power and an IP-address."

De eenvoudige interface is bedoeld voor querying en connections. Niet voor tuning en andere DBA-zaken. Die heeft de NPS niet nodig. Daarmee wordt de DBA een hoop werk uit handen genomen. "Dat klopt. De DBA krijgt dus meer tijd voor belangrijke dingen, zoals praten met en vooral luisteren naar de business, en het efficiënt bouwen van applicaties. Meer tijd voor andere projecten", zegt Lindsey. "De meeste DBA's die ik ken managen allemaal meer dan één systeem. Daar hebben ze nu dus meer tijd voor."

Conclusie

De grootste uitdaging die Netezza heeft is het overtuigen van prospects van de verbluffende prestaties, schaalbaarheid en onderhoudsvrijheid van de NPS. En technologie-gebaseerde bedrijven hebben over het algemeen weinig kaas gegeten van marketing. In het geval van Netezza zijn het vooral de zeer

tevreden gebruikers die het evangelisatiewerk voor Netezza doen: zij zijn graag bereid een presentatie te geven over hun ervaringen met de NPS. Er bestaat zelfs al een community op het web. Inmiddels heeft Netezza ook in ons land een groeiend aantal enthousiaste gebruikers en zelfs een officiële partner, Inergy uit Woerden. Inergy's Peter Dieleman is niet alleen onder de indruk van de buitengewone performance en de eenvoud waarmee de NPS ingericht en onderhouden kan worden. Hij gelooft zozeer in het concept dat hij het aandurft om klanten de NPS te leveren met een aanpak die hij 'fixed price and fixed data' noemt. Netezza sloot inmiddels partnerships met BI-leveranciers Actuate, Business Objects, Cognos, Hyperion, Information Builders, MicroStrategy, SAS en SPSS; op het gebied van data-integratie met Ab Initio, Ascential, Business Objects, Informatica en Sunopsis. Andere bedrijven zullen volgen.

Bovendien heeft men inmiddels aangekondigd de gang naar de beurs te maken. Is er dan helemaal niets aan te merken op Netezza? Weinig. Veel gebruikers zien in een datawarehouse appliance een geschenk uit de hemel. Maar het is misschien goed erbij te bedenken dat het geen artikel is dat vanuit de schappen geleverd wordt. Elke bestelling wordt op maat gemaakt en ingeregeld. Het succes zou hierbij roet in het eten kunnen gooien. Hopelijk blijft de performance van Netezza – dat overigens 'resultaten' betekent in het Urdu – als leverancier net zo goed als die van hun datawarehouse appliance.

Hans Lambou is hoofdredacteur van Database Magazine.