MySQL trends: patches, forks, en third party vendors

Substantiële waardetoevoeging

Roland Bouman

In DB/M 7 zijn in het artikel over de stand van zaken en de toekomst van MySQL de eigenschappen besproken die MySQL tot de meest populaire database voor webtoepassingen hebben helpen maken. Een van de meer prominente boodschappen is dat het relatief gemakkelijk is om met MySQL een replicatiecluster te bouwen, waardoor op basis van goedkope commodity hardware zowel de throughput als de beschikbaarheid van (web)applicaties flexibel kan worden geregeld (het scale-out principe).

In het voorgaande artikel [1] werd ook al even aangestipt dat het minder gemakkelijk is om de prestaties van een enkelvoudige MySQL server te verbeteren door het inzetten van meer krachtige hardware (scale-up). De laatste jaren is er een aantal onafhankelijke partijen opgestaan die zich juist zijn gaan richten om via scale-up de prestaties van enkelvoudige MySQL instances te verbeteren. In dit artikel wordt een aantal van deze vendors besproken met hun producten en de technieken die zij gebruiken om MySQL voor een bepaald doeleinde te perfectioneren.

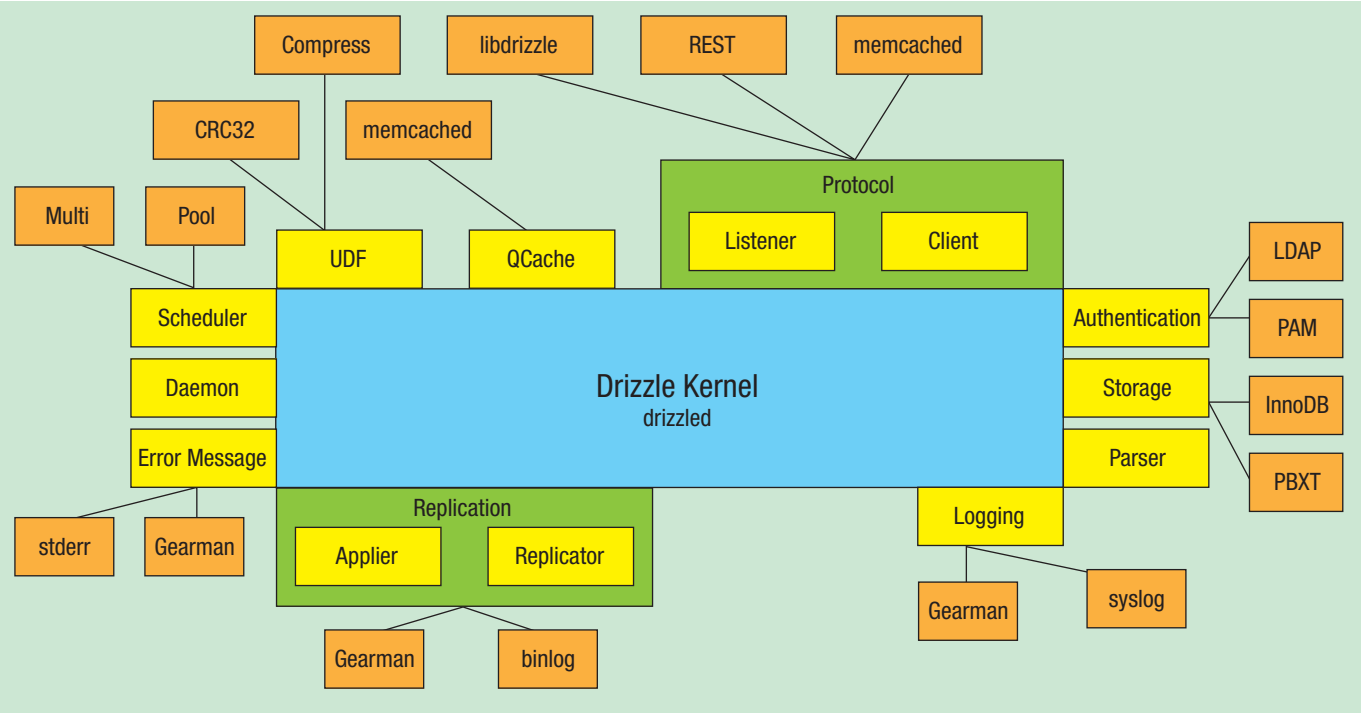
Waarom toch scale-up?

Een belangrijke impuls in de richting van scale-up zijn de huidige trends in hardware-ontwikkeling. Scale-out was lange tijd een goede oplossing om met goedkope commodity hardware toch een robuuste database architectuur te bouwen. Maar de commodity hardware van vandaag de dag ziet er toch wel wat anders uit dan die van een jaar of vijf geleden. Zo is er al geruime tijd de trend gaande dat kloksnelheden van microprocessoren niet of nauwelijks verbeteren. In plaats daarvan krijgen computers wel steeds meer microprocessoren tot hun beschikking, hetzij door meerdere cores op een chip te plaatsen, hetzij door de computer met meerdere chips uit te rusten. Dit betekent voor veel software toepassingen een uitdaging, omdat deze rekenkracht eigenlijk alleen effectief benut kan worden door het vergroten van het aantal parallelle operaties. Dat het nodig is om de beschikbare capaciteit te benutten, staat ook als een paal boven water: een niet benutte computer neemt wel nog steeds evenveel ruimte in, verbruikt (bijna) evenveel stroom, en produceert (bijna) evenveel hitte, en daarmee hebben we gelijk de belangrijkste componenten in de kosten van een datacentrum genoemd. Nu is het juist voor database software moeilijk om alle microprocessoren te benutten: veel tijd gaat verloren aan het ophalen en opslaan van data, en in die tijd kan er binnen hetzelfde proces of dezelfde thread niets anders gedaan worden dan wachten. Dit is dus niet specifiek voor MySQL, het geldt eigenlijk voor alle traditionele database-architecturen.

Een andere impuls wordt gevormd door de vrij recente ontwikkelingen op het gebied van storage technologie, en dan met name de flash-technologie en solid state drive (SSD) technologie. Traditionele magneetschijven zijn in de loop der jaren vooral in capaciteit toegenomen, maar niet uitgesproken sneller geworden. Op het eerste gezicht lijkt dit een goede zaak, want omdat er steeds meer behoefte bestaat om zaken te bewaren nemen de datavolumes immers toe. Maar juist voor databasetoepassingen zijn schijven met veel capaciteit minder optimaal. Een grotere capaciteit per schijf betekent immers dat er voor random access operaties meer en langer gezocht moet worden en dit draagt dus alleen nog maar meer bij aan wachttijden, en dus minder benutting van de beschikbare processoren. Solid state drives brengen hierin verandering. Deze drives zijn op dit moment nog vrij duur en bieden minder opslagcapaciteit dan de traditionele magneetschijven. Maar daar staat tegenover dat bij SSD's de wachttijd voor random access leesacties zo'n 10 tot 100 maal lager is dan voor magneetschijven. Behalve dat zijn SSD's in potentie ook zuiniger wat betreft energieverbuik, en produceren ze minder warmte.

Patches en verschillende MySQL versies

Zoals is vermeld zijn er grenzen aan het vermogen van MySQL om op te schalen. Nu zijn er vanuit Google [2], maar ook via enkele consultancyfirma's broncode-aanpassingen (zogenaamde 'patches') beschikbaar om een aantal van deze problemen te verhelpen. Deze patches zijn evenals MySQL zelf ook open source. Hoewel de patches al enige tijd geleden aan de MySQL ontwikkelaars zijn aangeboden, is het proces om deze verbeteringen



Afbeelding 1: Drizzle Kernel.

te accepteren en te incorporeren tot nu toe erg langzaam geweest, hetgeen weer meer kritiek heeft geoogst.
Sommige firma's brengen daarom hun eigen versie van de MySQL server uit. Ze nemen daarbij de broncode van de standaard MySQL server, en voegen daaraan hun eigen patches toe, en soms ook die van buiten het bedrijf. Voorbeelden hiervan zijn OpenQuery, Proven Scaling en Percona. Deze bedrijven zijn eigenlijk geen softwareleveranciers in de enge zin: in de eerste instantie betreft het hier consultancybedrijven die proberen om heel direct op de eisen van hun klanten aan te sluiten door een beter schaalbare MySQL versie uit te brengen.

> Sommige firma's brengen hun eigen versie van de MySQL server uit

Het laatstgenoemde Percona gaat nog een stapje verder, en heeft zelfs zijn eigen lijn van de InnoDB engine ge-'forked' en onder de naam XtraDB uitgebracht [3]. Recentelijk heeft dit bedrijf de 'hot backup' tool Xtradb Backup uitgebracht onder een open source licentie [4]. XtraDB Backup is ook compatible is met de InnoDB engine waarvan XtraDB is afgeleid. Daarmee gaat Percona nadrukkelijk de concurrentie aan met Oracle, die behalve eigenaar van de open source InnoDB engine ook eigenaar is van InnoDB hot backup [5]. Hoewel de InnoDB storage engine beschikbaar is onder een open source licentie, is InnoDB hot backup een proprietary product, dat alleen onder de voorwaarden van een commerciële softwarelicentie verkrijgbaar is.
Ook het bedrijf Monty Program ab is in deze categorie een aparte vermelding waard. Dit bedrijf is zeer recent opgericht door Michael Widenius ('Monty'), de oprichter en hoofdontwikkelaar van het originele MySQL product. Monty Program ab levert een aangepaste MySQL versie genaamd MariaDB, met daarin een transactionele opvolger van de MyISAM engine genaamd Maria [6]. Maar MariaDB is meer dan MySQL+Maria: het bedrijf heeft zichzelf zeer nadrukkelijk tot doel gesteld om leverancier te worden van de beste open source MySQL versie. Door veel mensen in de MySQL community wordt Monty Program ab gezien als een partij die kan garanderen dat MySQL in actieve ontwikkeling zal blijven, ook na de op handen zijnde overname door Oracle.

Forks

Door sommigen worden de net besproken aangepaste MySQL versies aangemerkt als *forks*. Toch is dit niet helemaal juist: bij een echte fork wordt de bestaande codebase als uitgangspunt genomen voor verdere ontwikkeling, echter zonder de intentie om *backwards compatible* te blijven met het originele product. Omdat de ontwikkelingen in de loop der tijd tussen het origineel en de fork dusdanig uiteen gaan lopen, wordt het gaandeweg onmogelijk om verbeteringen die voor één van de branches worden ontwikkeld, terug te voeren in de andere branch. Dit is vooralsnog niet aan de orde bij de zojuist genoemde alternatieve MySQL versies: er wordt juist geprobeerd om de patches aan te passen aan de MySQL mainline, en tot dusver wordt deze aangevoerd door Sun Microsystems.

Op dit moment is er één echte MySQL fork in ontwikkeling: Drizzle [7]. Het initiatief om Drizzle uit MySQL te forken komt vanuit Sun Microsystems, en de vaste kern van Drizzle ontwikkelaars wordt ook gevormd door werknemers van Sun. Drizzle is gebaseerd op de recente MySQL 6.0 codebase, maar daaraan zijn direct vergaande versimpelingen en wijzigingen aangebracht ten einde tot een meer schaalbare en modulaire database server te komen. Drizzle is qua datatypes en feature set dan ook niet compatibel met MySQL.

Een belangrijk verschil met MySQL is dat Drizzle in opzet veel opener probeert te zijn dan MySQL. Zo worden allerlei discussies over te ontwikkelen functionaliteit openlijk gevoerd via publiekelijk toegankelijke mailinglijsten, en is het vrij eenvoudig om als niet-Sun werknemer toch deel te nemen aan het ontwikkelproces. Door uit te gaan van een minimalistische kern (microkernel) en meer gespecialiseerde functionaliteit bewust te isoleren in plug-ins wordt het eenvoudiger om aan het product wijzigingen door te voeren en patches bij te dragen. Was MySQL al uitgerust met de *pluggable* storage engine architectuur, in Drizzle wordt dit idee nog weer verder doorgevoerd om behalve storage engines ook zaken als logging, netwerkprotocol en zelfs de parser pluggable te maken. Dit is te zien in afbeelding 1 [8]. In de afbeelding is de eigenlijke Drizzle server in het midden afgebeeld (blauw). Deze is uitgerust met een groot aantal publieke interfaces (geel). Voor bijna al deze interfaces zijn er reeds implementaties in de vorm van plugins beschikbaar (oranje).

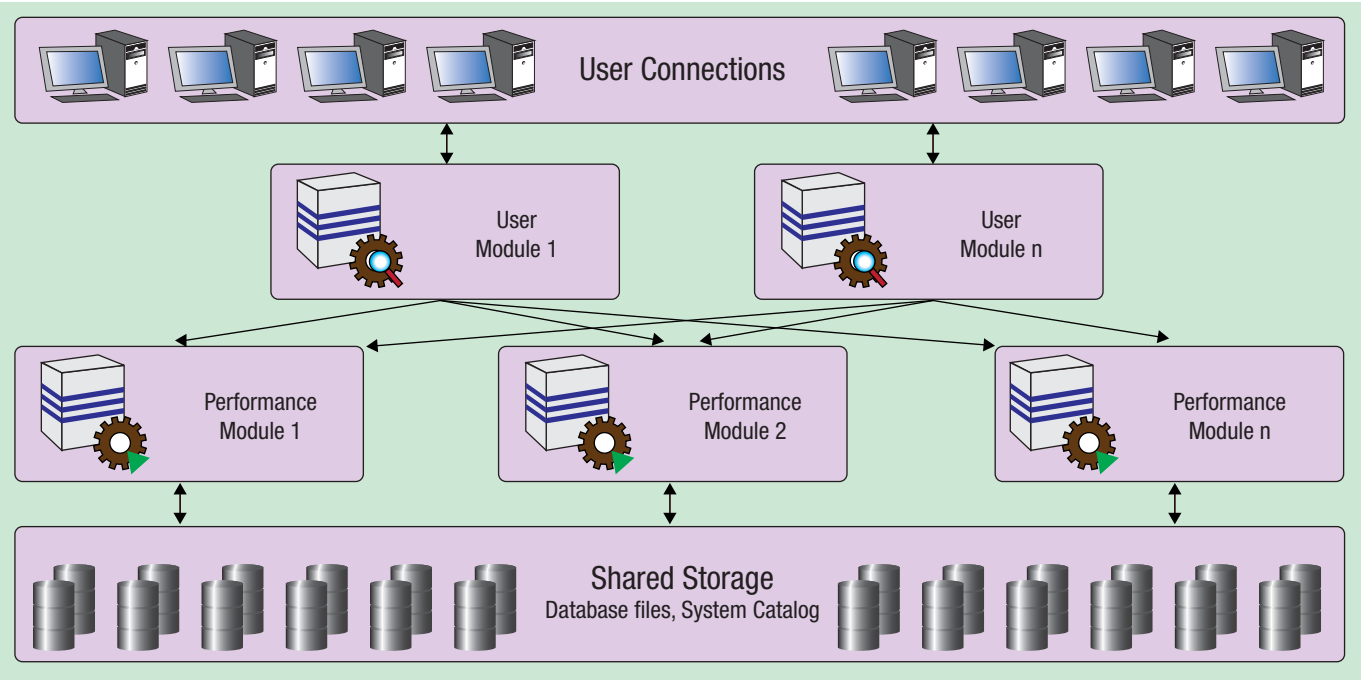
Third party MySQL vendors

De laatste jaren is er steeds meer interesse voor het MySQL serverproduct te bespeuren vanuit fabrikanten van special-purpose databases. Door hun product in te bouwen in de MySQL server (soms in combinatie met gespecialiseerde hardware), maken zij hun eigen product toegankelijk voor MySQL applicaties, en voorzien ze het daarmee van een SQL front-end zonder zelf te hoeven investeren in het bouwen en onderhouden van een server platform. Daarnaast levert dit connectiviteit en compatibiliteit op met MySQL applicaties, omdat zaken als het netwerkprotocol, en in mindere mate de ondersteunde datatypes en het SQL dialect gehandhaafd blijven.

Third party Storage Engines

Sommige vendors leveren een puur softwarematige oplossing. Voorbeelden hiervan zijn Infobright en Calpont.
Infobright is een analytische database [9]. Infobright heeft een eigen MySQL storage engine die gebouwd is volgens het *column store* principe: data zijn per kolom opgeslagen in plaats van per rij, en worden bovendien ook nog eens sterk gecomprimeerd. De storage engine bevat ook een metadatalaag (de zogenaamde 'Knowledge Grid') en een aangepaste optimizer, zodat de juiste data snel kunnen worden gezocht zonder altijd te hoeven decomprimeren. De knowledge grid houdt ook bepaalde statistieken bij, waardoor aggregaten (zoals COUNT() en SUM()) supersnel kunnen worden berekend. Daarmee is dit product vooral gericht op datawarehousing en Business Intelligence applicaties.

Infobright levert zowel een community als een enterprise editie. De community editie is 100 procent open source (op basis van een GPL licentie). Echter voor dit product kan van Infobright geen ondersteuning verkregen worden, en daarnaast bevat het minder features of minder geavanceerde features dan de enterprise editie. Voorbeelden van dergelijke features zijn de bulk-



Afbeelding 2: De InfiniDB architectuur van Calpont.

loader, die weliswaar aanwezig is in de community editie, maar een veel minder goede performance biedt dan de in de enterprise editie aanwezige bulk loader. Een ander voorbeeld is de syntax: de enterprise editie heeft een min of meer complete SQL-implementatie, maar in de community editie zijn INSERT, DELETE en UPDATE statements niet beschikbaar. Met dit gemis is de bulkloader de enige mogelijkheid om data in het systeem te krijgen, en kunnen data alleen verwijderd worden door te partitioneren en een partitie te DROP'pen. Al met al lijkt Infobright's community editie meer op een 'trial' versie dan een volwaardig software product. Ondanks dat de enterprise editie alleen commercieel verkrijgbaar is, is zij wel bedoeld als een low-cost oplossing voor het bouwen van multiterabyte datawarehouses.

> Drizzle is qua datatypes en feature set niet compatibel met MySQL

Calpont is eveneens een analytische database [10]. Net als Infobright is ook dit een op datawarehousing toegesneden column-oriented storage engine. In tegenstelling tot Infobright voorziet Calpont's architectuur ook in grootschalig parallelisme (MPP, massive parallel processing) om analytische query's snel op te kunnen lossen. Dit wordt gerealiseerd via een gedistribueerde database-architectuur, welke door Calpont met de term InfiniDB wordt aangeduid.
De InfiniDB architectuur kent drie verschillende soorten componenten: user modules, performance modules, en storage (zie afbeelding 2). Een user module is een aangepaste MySQL server die de binnenkomende SQL query's opbreekt in kleinere eenheden die onafhankelijk van elkaar kunnen worden uitgevoerd. De user module zet de stukken van de opgeknipte SQL query vervolgens door naar één of meerdere performance modules. De performance modules zijn verantwoordelijk voor het daadwerkelijk uitvoeren van de query's en het ophalen van de data. De performance modules halen daartoe data op uit de eigenlijke storage eenheden, maar werken tegelijkertijd ook als een cache, zodat vaak gevraagde data direct uit het geheugen geserveerd kunnen worden. Na het ontvangen van de partiële resultaten van de performance modules construeren de user modules daaruit het uiteindelijke queryresultaat, om het vervolgens naar de client terug te sturen. Een belangrijk punt is dat in een kenmerkende InfiniDB architectuur zowel de user modules als de performance modules meervoudig aanwezig zijn, en in principe als fysiek separate computers (nodes) draaien. Schaalbaarheid kan worden verkregen door meer nodes met hetzij performance modules, hetzij user modules toe te voegen. Calpont is commercieel verkrijgbaar en adverteert evanals Infobright een low-cost alternatief te zijn voor het bouwen van volwaardige datawarehousingtoepassingen.

MySQL Appliances

Andere vendors combineren een door hen aangepaste MySQL versie met gespecialiseerde hardware tot een appliance – een kant en klare server die geoptimaliseerd is voor een bepaald doeleinde. Hier volgt een aantal voorbeelden:
Kickfire; opnieuw een analytisch product [11]. Kickfire levert een oplossing volgens wat zij zelf noemen de 'Kickfire Equation': Column store + Compression + SQL Chip = performance. Met de term 'SQL Chip' wordt bedoeld: gespecialiseerde hardware waarmee het mogelijk wordt een stroom van data parallel te verwerken; en daarmee ook een vorm van MPP. Kickfire is alleen op commerciële basis verkrijgbaar als een appliance, en opnieuw is het de insteek van de vendor om een low-cost datawarehousing platform te bieden.
Schooner MySQL Appliance; deze appliance combineert een voor schaalbaarheid aangepaste MySQL versie met snel SSD geheugen voor storage [12]. Eigenlijk is dit product vooral bedoeld als een scale-up alternatief voor traditionele MySQL (web)applicaties.
Virident Greencloud server for MySQL; net als de Schooner appliance is dit product vooral bedoeld als scale-up alternatief voor webapplicaties [13].

Deze vendors maken dankbaar gebruik van MySQL's pluggable storage engine architectuur. Maar meestal is dat nog niet voldoende om echt alle voordelen van hun oplossing te genieten. Om goed om te kunnen gaan met een column store is het bijvoorbeeld nodig om de optimizer aan te passen. Om effectief gebruik te maken van zoiets als Infobright's knowledge grid moet ook de executor aangepast worden. Deze vendors voegen dus op het softwarevlak substantieel waarde toe aan het gewone MySQL product.

Literatuur

1. *DB/M 2009/7: MySQL: De stand van zaken.*
2. *Site: google-mysql-tools – http://code.google.com/p/google-mysql-tools/*
3. *Site: Xtradb – www.mysqlperformanceblog.com/2008/12/16/announcing-percona-xtradb-storage-engine-a-drop-in-replacement-for-standard-innodb/*
4. *Site: Xtradb Backup – https://launchpad.net/percona-xtrabackup*
5. *Site: Innodb hot backup - www.innodb.com/products/hot-backup/order/*
6. *Site: Maria FAQ – http://askmonty.org/wiki/index.php/Maria*
7. *Site: Drizzle – https://launchpad.net/Drizzle*
8. *Site: Drizzle modularity – http://oddments.org/?p=56*
9. *Site: Infobright technology –www.infobright.com/Products/Technology/*
10. *Site: Calpont – http://www.calpont.com/*
11. *Site: Kickfire – http://www.kickfire.com/*
12. *Site: Schooner Appliance for MySQL – www1.schoonerinfotech.com/products/mysql-appliance.html*
13. *Site: Virident Greencloud server for MySQL – www.virident.com/bhive/c/4/3130/Virident GreenCloud_MySQL_DataSheet_June15.pdf*

Roland Bouman is webapplicatie- en BI-ontwikkelaar voor Strukton Rail en auteur.

Met dank aan Eric Day (Sun Microsystems) en Robin Schumacher (Calpont).