

Ruim een kwart eeuw SAS Institute

Organisatiebrede analyses uit gelaagde BI-architectuur

Freek Kamst

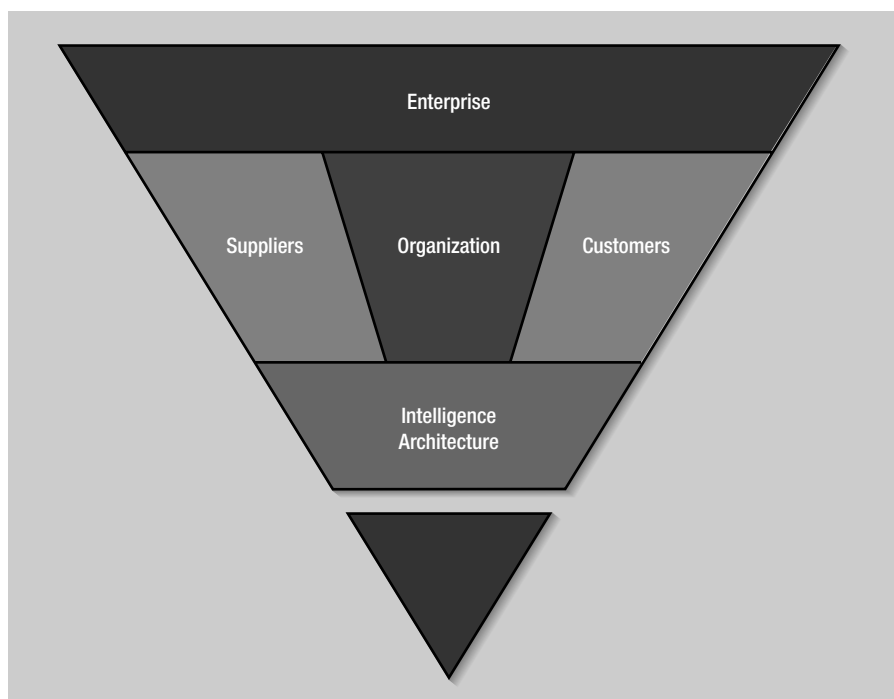
Toen Jim Goodnight in 1976, samen met drie medestudenten van de Universiteit van North Carolina, SAS Institute oprichtte, kon hij niet vermoeden dat zijn initiatief anno 2002 zou uitgroeien tot een wereldwijde onderneming met 8500 medewerkers. Toen al richtte SAS Institute zich op toepassingen die analyses konden uitvoeren op de enorme gegevensverzamelingen binnen mainframe-omgevingen. Vandaag de dag is SAS Institute de grootste niet-beursgenoteerde software-onderneming ter wereld en actief op het gebied van Enterprise Business Intelligence Suites.

Enterprise Business Intelligence Suites is de verzamelnaam van toepassingen die de basis vormen voor een organisatiebrede Business Intelligence-architectuur. Deze architectuur ondersteunt het gehele proces van gegevens tot informatie. Hierbij worden vanuit de operationele bronsystemen de gegevens verzameld en gestructureerd in een datawarehouse-omgeving opgeslagen. Vervolgens wordt vanuit deze gestructureerde omgeving met allerlei analytische toepassingen informatie geproduceerd ter ondersteuning van de tactische en strategische bedrijfsprocessen. Een trend van de laatste jaren is dat de leveranciers van Business Intelligence- en ETL-tools hun productontwikkeling

richten op de Enterprise Business Intelligence Suite-architectuur. Ook leveranciers van ERP (Enterprise Resource Management) en CRM (Customer Relationship Management) positioneren hun toepassingen in een dergelijke architectuur. SAS Institute heeft sinds 2001 ook een Enterprise Business Intelligence Suite-architectuur ontwikkeld, waarbij de SAS-toepassingen functioneel in een omgekeerde piramide worden weergegeven. Deze piramide geeft tevens weer hoe de informatiebehoefte zich ontwikkelt naarmate een organisatie groeit.

HET OPERATIONELE FUNDAMENT

Initieel begint iedere organisatie met de traditionele operationele systemen. Deze systemen zijn uiteraard noodzakelijk om de bedrijfsvoering te waarborgen, maar meestal niet geschikt om informatie te leveren ten aanzien van de tactische en strategische processen binnen een organisatie. Dit terwijl deze systemen in groten getale waardevolle gegevens bevatten waarmee het management beslissingen kan ondersteunen. De operationele systemen zijn in de structuur zichtbaar als de drie onderste grijze lagen van de piramide en worden hier het operationele fundament genoemd. De drie niveaus binnen het operationele fundament zijn onafhankelijk van elkaar. Het bovenste grijze niveau vertegenwoordigt de operationele applicaties die gegevens genereren vanuit de onderliggende database laag. Deze



FIGUUR 1: DE SAS INTELLIGENCE LAYERS.

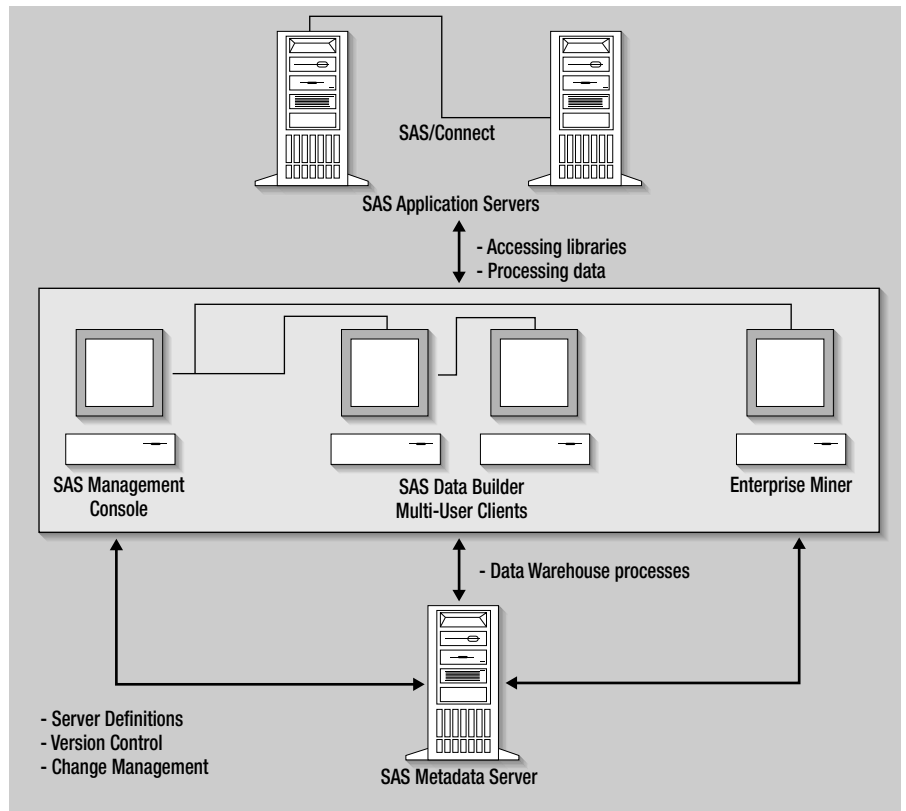
databases zijn operationeel op de infrastructuur die het onderste puntje van de piramide vertegenwoordigt. De infrastructuur bevat alles wat nodig is om een operationele omgeving draaiende te houden, inclusief hardware, systeemsoftware, mensen en overige middelen.

DE INTELLIGENCE ARCHITECTUUR

De Information Delivery Architectuur (zie figuur 2) vormt de technologisch basis voor alle oplossingen in de Intelligence Architectuur-laag die boven op het operationele fundament is gebouwd. Deze architectuur omvat de toepassingen voor het extraheren, transformeren, opslaan en produceren van informatie uit de operationele systemen. Deze laag vertegenwoordigt de basisfunctionaliteit van SAS Institute zoals deze 25 jaar geleden vorm heeft gekregen. De ETL-component wordt hier vertegenwoordigd door het product *SAS Warehouse Administrator* en bevat een eerste generatie ETL-tool. Eerste generatie ETL-tools kenmerken zich door het genereren van programmatuur die op de bron- en doelsystemen worden uitgevoerd om zo een gegevensextractie tot stand te brengen. Dit in tegenstelling tot tweede generatie ETL-tools die gebruik maken van een

DataStep is een uiterst krachtige taal

aparte server en de extractie via een gegevensstroom laten uitvoeren. *SAS Warehouse Administrator* genereert een specifieke eigen taal *DataStep*, maar kan ook C-, Cobol- of SQL-programmatuur aanmaken. *DataStep* is een uiterst krachtige taal waarmee complexe extracties en transformaties kunnen worden uitgevoerd. Hierbij kan de datawarehouse-architectuur naar eigen inzicht worden ontworpen met een onbeperkt aantal datawarehouses en datamarts. De onderliggende database is in principe een SAS Institute-eigen struc-



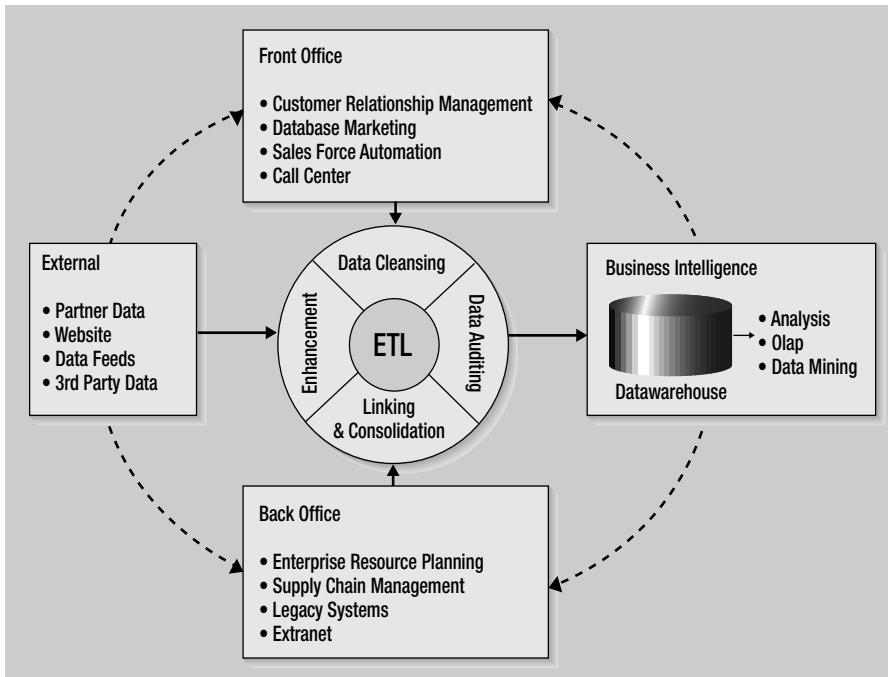
FIGUUR 2: DE INFORMATION DELIVERY ARCHITECTUUR.

tuur, die zo is opgezet dat analyses en rapportages snel en efficiënt kunnen worden geproduceerd. Inmiddels behoort een Oracle- of DB2-database ook tot de mogelijkheden. Om alle gegenereerde programmatuur tijdig en sequentieel te kunnen laten draaien, beschikt *SAS Warehouse Administrator* over een geavanceerde scheduler, waarmee het mogelijk is om over verschillende systemen toepassingen en scripts te activeren die afhankelijk zijn van elkaar. Sinds de laatste versie van *SAS Warehouse Administrator* worden de metagegevens in deze laag op een aparte server beheerd. Deze Metadata Server beheert in een repository alle metagegevens van de gehele Intelligence Layer-structuur. De voordelen hiervan zijn vooral in het wijzigingsbeheer zichtbaar. Bijvoorbeeld wanneer een wijziging in een rapportage plaatsvindt, zal het gehele traject naar de bron in kaart kunnen worden gebracht en worden de extractieprocedures opnieuw gegenereerd. Andersom wanneer er een wijziging plaatsvindt op één van de bronsystemen kan dezelfde procedure richting rapportageomgeving worden uitgevoerd.

De Information Delivery Architectuur is gebaseerd op een centrale ETL-omgeving waarbinnen het *SAS Management Console* voor het beheer van de extractie- en transformatieprocessen zorgdraagt. De *SAS Data Builder Multi-User Clients* zijn de interfaces voor het onderhoud en de bouw van de extractie- en transformatieprocessen. De *Enterprise Miner* verwerkt ten slotte de opdrachten uit de front-end SQL-queries op het datawarehouse en zorgt ervoor dat deze optimaal worden uitgevoerd.

METAGEGEVENS

Zoals beschreven worden alle metagegevens centraal beheerd op een Metadata Server. De rapportage- en analysefunctionaliteit op deze server zorgen ervoor dat er snel en adequaat inzicht kan worden gegeven in mogelijke wijzigingen op de gehele Enterprise Business Intelligence structuur. Het metagegevensmodel is volgens het standaard CWM (Common Warehouse Model) opgezet en de gegevens zijn uitwisselbaar met andere toepas-



FIGUUR 3: DE CENTRALE ETL-OMGEVING.

singen die voldoen aan de specificaties van CWI (Common Warehouse Interface). Hiermee heeft SAS Institute op het gebied van haar productontwikkeling ook de stap gezet om zich opener naar de markt toe te positioneren. Door de gesloten architectuur van SAS was het voorheen complex om niet-SAS-producten aan de SAS-structuur te koppelen, maar door de metagege-

Een van de hoofdredenen waarom CRM-projecten mislukken, is het ontbreken van kwaliteit van de gegevens.

vens op een aparte server beschikbaar te stellen, is het nu ook voor andere producten mogelijk om hier gebruik van te maken. Alle lagen boven de Intelligence Layer maken gebruik van de metagegevens die in deze laag worden opgeslagen.

SAS Institute is ervan doordrongen dat gegevenskwaliteit essentieel is voor een juiste verslaglegging. Wanneer de basisgegevens voor een analytische toepassing niet betrouwbaar zijn, dan heeft een analyse of rapportage geen toegevoegde waar-

de. Een van de hoofdredenen waarom CRM-projecten mislukken, is het ontbreken van kwaliteit van de gegevens. Immers, wanneer de gebruikers niet de juiste informatie uit een toepassing verkrijgen, zal men de applicatie spoedig niet meer gebruiken. Op het gebied van gegevens-cleansing en -auditing heeft SAS Institute twee jaar geleden het bedrijf DataFlux gekocht, die op dat moment al veel ervaring had met oplossingen op het gebied van gegevenskwaliteit binnen SAS omgevingen. Door de toepassingen in het ETL-gedeelte te integreren, is er extra functionaliteit ontstaan die de kwaliteit van de gegevens kan waarborgen. Het ETL-gedeelte binnen de Intelligence Architectuur speelt een centrale rol bij het beschikbaar stellen van informatie aan alle Business Intelligence-toepassingen en daarom is deze laag dan ook essentieel voor het succes van een Enterprise Business Intelligence-omgeving.

DE ORGANISATORISCHE INTELLIGENCE-ARCHITECTUUR

Deze laag bestaat uit een aantal specifieke toepassingen die gebruik maken van de informatie uit Information Architectuur-

laag. Alhoewel de toepassingen geïntegreerd zijn met de functionaliteit van de onderliggende laag, kunnen deze ook op andere databases of datawarehouses worden aangesloten. Bijvoorbeeld een datawarehouse binnen een Informatica-omgeving kan ook gebruik maken van de analytische toepassingen van SAS Institute. De Organisatorische Intelligence-laag bestaat uit verschillende modules die onafhankelijk van elkaar op zogenaamde SAS Applicatie Servers operationeel zijn. Voorbeelden van analytische toepassingen waar SAS Institute momenteel succesvol mee is, zijn:

- Supplier Relationship Management
- Financial Management Solutions
- IT Management Solutions
- Human Capital Management
- Customer Relationship Management

De meeste van de hiervoor beschreven toepassingen bestaan uit een standaard gegevensmodel en bijbehorende extractiemodules inclusief rapportages en analyses. De toepassingen maken gebruik van de basisgegevens uit het datawarehouse en slaan hun tussenresultaten op in databases op de SAS Applicatie Servers. Deze gegevens kunnen in latere analyses weer gebruikt worden.

DE STRATEGIC PERFORMANCE MANAGEMENT-LAAG

De bovenste laag van de piramide zorgt ervoor dat een organisatie de mogelijkheid krijgt om prestatiegericht te sturen door inzicht te geven in klanten, de organisatie en de leveranciers. De strategische doelen, prestatiemetingen en verbeteracties worden vastgelegd in zogenaamde scorecards en de resultaten hiervan kunnen worden gepubliceerd via een intranet, zodat medewerkers stuurinformatie kunnen analyseren, communiceren en nieuwe meetwaarden kunnen invoeren. De Strategic Performance Management-laag is pas te realiseren wanneer de onderste lagen zijn ingevuld. Hiermee kunnen we tevens concluderen dat naarmate de Business Intelligence-toepassingen binnen

een organisatie zich ontwikkelen ook de informatiebehoefte evenredig toeneemt.

TOT SLOT

De kerncompetentie van SAS Institute is gebaseerd op een reeds lange beproefde technologie. Echter, het nadeel van oude technologieën is dat deze niet zijn ontwikkeld conform de open standaarden. SAS Institute heeft dit probleem inmiddels onderkend en conformeert zich steeds meer aan open standaarden zoals XML, CWM en CWI. De integratie met niet-SAS-producten wordt hierdoor eenvoudiger en daarom kunnen in het verleden gedane investeringen niet als verloren worden beschouwd. Het investeren in een

SAS-omgeving blijft echter nog steeds voorbehouden aan grote organisaties, maar nu Siebel, SAP en Oracle zich naast Microsoft ook gaan richten op middelgrote en kleine ondernemingen, zal ook SAS

Oude technologieën zijn niet ontwikkeld conform open standaarden

Institute hier binnenkort op moeten reageren. In de Verenigde Staten heeft men al succes met een ASP (Application Service Provider) oplossing voor de farmaceutische- en detailindustrie. Hiermee kunnen

organisaties via het internet gebruik maken van analytische applicaties tegen relatief kleine investeringen. Door een financieel sterke positie en veel te investeren in de ontwikkeling van haar producten, blijft SAS Institute een innoverende onderneming waar de komende jaren nog veel van verwacht zal worden. ●

Freek Kamst (Freek.Kamst@newcom.nl) is consultant bij Newcom Information Systems.