

Deze website gebruikt cookies om de beste gebruikerservaring mogelijk te maken. [Meer informatie](#)
Akkoord

BI-Platform.

Door: Sjoerd Janssen

DW&BI Summit 2019: Prepare for success



Deel dit bericht

Op 27 en 28 maart vond alweer de zesde editie van de Data Warehousing en Business Intelligence Summit plaats. Naast dat er is ingegaan op de nieuwste ontwikkelingen op het gebied van databases, data architectuur en data science, is er vooral veel aandacht besteed aan het hoe. Hoe bereik je succes: Hoe pak je het aan, hoe richt je het in en hoe organiseer je het?

Hoe pak je het aan?

Lawrence Corr, voormalig associate bij Ralph Kimball, legt tijdens zijn sessie op het Summit zijn aanpak voor requirementsmodellering op het gebied van datawarehousing en Business Intelligence uit. Hij gaat daarbij eerst in op de manco's van het gebruik van entity relationship diagrammen:

- *Ze helpen niet in een heldere communicatie omdat de business ze vaak niet begrijpt;*
- *Ze vertellen geen verhaal;*
- *Ze spreken niet tot de verbeelding.*

Zijn methode, Data Modelstorming, heeft deze nadelen niet. Het belangrijkste onderdeel binnen de Modelstorming aanpak bestaat uit het per gebeurtenis (te herkennen aan een werkwoord) verzamelen van de antwoorden op de volgende vragen: Wanneer, waar, hoe, hoeveel, waarom, wie en wat. Deze methode kun je bovendien heel goed in interactieve workshops, met behulp van

simpele post-its, toepassen. Het vormt daarmee een goede aanpak voor het verzamelen van requirements vanuit een businessperspectief.

In zijn twee sessies tijdens de Summit, gaat de zeer ervaren data scientist Keith McCormick in op het ontwikkelen, maar vooral ook toepassen van data science modellen in de praktijk. Hij benadrukt daarbij dat je voor een succesvolle toepassing van data science het niet als een ontwikkelproject, maar als een proces moet aanvliegen. En daarbij zijn een goede voorbereiding en uitvoering van de implementatie in de organisatie wellicht nog belangrijker dan de accuratesse van het model. Hij besteedt dan ook een complete sessie aan de voorbereidingen die je moet treffen voor een succesvolle adoptie. In die sessie geeft hij aansprekende voorbeelden van de weerstand die hij in de praktijk is tegengekomen. Zo geeft hij een voorbeeld van een vervoersbedrijf waarbij de monteurs het ontwikkelen van een model voor preventief onderhoud vooral beschouwden als het in twijfel trekken van hun deskundigheid. Ze zagen het model als een middel voor het senior management om met de vingers te kunnen wijzen. In een ander voorbeeld van een model voor voorspelling van productverkoop, wordt het model na verloop van tijd niet meer gebruikt omdat er geen vertrouwen in het model is. Dat blijkt echter niet te liggen aan de accuratesse van het model, maar wel aan een juiste toepassing van het model. Er was onvoldoende aandacht besteed aan het betrekken en informeren van de eindgebruikers.

Kent Graziano legt in zijn eerste sessie op de Summit uit hoe je de twaalf principes achter het Agile Manifesto kunt toepassen in een datawarehousing- en Business Intelligence-project. Kent, auteur en gerenommeerd spreker op het gebied van data modeling, data architecture en datawarehousing, ziet daarbij Data Vault modellering als een middel om agile te kunnen werken. Dit omdat een Data Vault model iteratief kan worden opgebouwd en uitgebreid zonder de noodzaak tot herbouw en het opnieuw laden van de historische data. Andere technieken die kunnen ondersteunen in het agile werken in een Business Intelligence- en datawarehouse-omgeving zijn codegeneratoren en het virtualiseren van je rapportagelaag, zodat je deze ook iteratief kunt aanpassen zonder impact op de dataopslag. Denkt Kent dat het daarmee realistisch is om in sprints van twee weken te werken? Misschien wel. Maar ten eerste is het doel niet om binnen twee weken iets te leveren, maar in lijn met de agile principes, in ieder geval binnen enkele weken. Ten tweede gaat Kent creatief om met wie er in het project als klant wordt gezien. Wellicht is dat in het geval van een DWH-aanpassing niet de business, maar een BI-ontwikkelaar die deze aanpassing nodig heeft om een rapport te kunnen ontwikkelen.

Hoe richt je het in?

Dagvoorzitter en spreker Rick van der Lans constateert tijdens de openingssessie op de Summit dat technologie steeds meer invloed heeft op de architectuur, vanwege het feit dat er in database-land steeds meer producten zijn die een architectuurkeuze afdwingen. Zo zijn er steeds meer gespecialiseerde producten met beperkte use cases zoals NoSQL databases. Daarnaast zijn er producten met unieke interne architecturen zoals analytical databases. Hij stelt zichzelf daarom de

vraag: Zouden we een data-architectuur moeten definiëren om vervolgens de best passende producten te selecteren, of zouden we ons in de data-architectuur meer moeten laten leiden door de (nieuwe) mogelijkheden die bepaalde producten bieden? Daarbij is hij, in tegenstelling tot wat hij in de schoolboeken geleerd heeft, steeds meer geneigd tot het laatste. Waar we dan wel veel behoefte aan hebben zijn architectuurontwerpprincipes die helpen de juiste keuze te kunnen maken. Voorbeelden die hij daarvan aanhaalt: Data-opslag en -toegang moeten technologie-agnostisch zijn, data processing moet zoveel mogelijk bij de data zelf plaats vinden en de metadata moet gecentraliseerd worden opgeslagen en beheerd.

Ook Kent Graziano geeft een sessie over data architectuur. Hij behandelt daarin eerst de oorspronkelijke datawarehouse-architecturen van Bill Inmon en Ralph Kimball. Vervolgens bespreekt hij ook nieuwere varianten van Bill Inmon (DW 2.0tm) en Daniel Lindstedt (Data Vault). In de praktijk komt hij echter vele hybride vormen tegen en zijn er ondanks deze architecturen nog steeds veel obstakels die organisaties verhinderen data in acties om te zetten. Hij erkent de volgende uitdagingen:

- *In de praktijk is het vaak nog te omslachtig om nieuwe databronnen toe te voegen;*
- *De laadtijden zijn te lang;*
- *De architecturen ondersteunen ad hoc analyses en data science onvoldoende;*
- *Er wordt te veel tijd besteed aan handmatige administratie;*
- *Reporting performance wordt nog te vaak beïnvloed door laadactiviteiten op het systeem.*

Het is volgens hem dan ook tijd voor een nieuwe datawarehouse-architectuur, die deze tekortkomingen adresseert. Een datawarehouse zal pas een succes zijn als:

- *Het tijdig kan worden geladen;*
- *Het onafhankelijk van het type data snel te bevragen is door zowel data scientists als business gebruikers;*
- *Het een lage total cost of ownership met zich meebrengt.*

Als oplossing voor deze uitdagingen beschrijft Kent tijdens zijn sessie een cloudgebaseerde datawarehouse-architectuur waar computing en storage uit elkaar zijn gehaald en welke ook kan omgaan met semigestructureerde data. Daarbij moet dan wel worden opgemerkt dat Kent tegenwoordig werkzaam is als Chief Technology Evangelist van Snowflake Computing, een bedrijf dat een datawarehouse-oplossing volgens deze architectuur aanbiedt.

Hoe organiseer je het?

Nigel Turner, al meer dan 25 jaar werkzaam in het information management gebied en vicevoorzitter van de Data Management Association van het Verenigd Koninkrijk, gaat in zijn eerste sessie tijdens de Summit in op de organisatie van data governance en het essentiële belang van een datastrategie. Een datastrategie is een midden- tot langetermijnplan voor de verbetering, het management en de exploitatie van data binnen de organisatie, en de manier waarop dit kan worden

gerealiseerd. En in een tijd waarin veel organisaties datagedreven willen opereren, wordt deze strategie steeds belangrijker. Data governance begint dan ook met het vaststellen van deze strategie, voordat er wordt nagedacht over zaken als de organisatie, processen, data management en meetwaarden, cultuur en communicatie en als laatste de tools en technologie. De verantwoordelijkheden kunnen daarbij worden ingericht op basis van de processen, systemen, datadomeinen of organisatieonderdelen.

In zijn sessie besteedt Nigel veel aandacht aan communicatie. Iedereen die data produceert of met data werkt moet zich bewust zijn van de datastrategie en datakwaliteit. Daarbij moeten we dan ook het jargon en de wollige taal, die in de praktijk veel voorkomen, vermijden.

In zijn sessie over datakwaliteit in relatie tot Business Intelligence en datawarehousing besteedt hij veel aandacht aan masterdata management. Masterdata betreft immers data die veel gedeeld wordt binnen en buiten de organisatie en daarom veel aandacht verdient. En dat datakwaliteit van sterke invloed is op het succes van datawarehousing- en BI-projecten staat als een paal boven water, maar dat betekent niet dat we problemen met de datakwaliteit dan ook daar moeten oplossen. Het is immers beter om te voorkomen dan om te genezen. Daarvoor is het dan weer van belang om een duidelijke data governance ingericht te hebben.

Als prikkelende afsluiter van de Summit pleitte Rick van der Lans voor het organiseren en inrichten van een Data Marketplace. Het gebruik van data is immers tegenwoordig zo veranderd en divers dat we onze data delivery-keten anders moeten organiseren. Wanneer we alleen precies op maat gemaakte producten voor de dataconsumenten produceren, doen we immers heel wat aannames die weleens onjuist kunnen zijn. Aannames die weleens onjuist zouden kunnen zijn:

- IT heeft alle data onder beheer;
- Gebruikers hebben geen behoeften om zelf data sets toe te voegen;
- Gebruikers weten precies wat ze willen (of zoals Henry Ford ooit zei: “If I had asked people what they wanted they would have said faster horses”);
- De gebruikers begrijpen de Business Intelligence tools voldoende om met self-service eigen inzichten te verkrijgen;
- Gebruikers kunnen zelf hun eigen rapporten ontwikkelen.

Kortom, wellicht is de tijd rijp voor confectie in plaats van op maat gemaakt. Wellicht moeten we niet meer alleen afgaan op wat er precies gevraagd wordt, maar dataproducten ontwikkelen waarvan we denken dat ze gebruikt gaan worden. Dan zouden we vervolgens op basis van het daadwerkelijke gebruik kunnen bijsturen. Dat vraagt wel een andere manier van organiseren en inrichten. Vergelijk het met een productaanbieder. We hebben dan research, development, marketing en sales nodig en de ontwikkelde producten moeten ook goed te vinden zijn.

Nederlandse inbreng

Naast bovengenoemde gerenommeerde internationale sprekers waren er tijdens de Summit ook

boeiende sessies van enkele Nederlandse sprekers. Zo ging Martin Kersten, befaamd wetenschapper op het gebied van database-architecturen en verbonden aan het Centrum voor Wiskunde en Informatica, diep in op wat er de laatste tien jaar nu echt is veranderd in de databasemarkt. Lex Pierik, senior Business Intelligence consultant, behandelde in zijn sessie met sprekende voorbeelden hoe data-driven storytelling kan worden toegepast om nieuw verkregen inzichten te communiceren. Welke architecturen geschikt zijn voor het verwerken en distribueren van grote hoeveelheden streaming data werd behandeld in een sessie van Rutger Rienks, thought leader data strategy bij KPN.

Samen met deze sessies geeft de goedbezochte Summit een mooi overzicht van de huidige stand van zaken en nieuwste ontwikkelingen op het gebied van datawarehousing, Business Intelligence, data science en datastrategie. Maar het meest waardevol zijn nog wel de inzichten die het de deelnemers biedt om succesvol te kunnen zijn in dit vakgebied. Hoe pak je het aan, hoe richt je het in en hoe organiseer je het? Prepare for success.

Tags:

[Datawarehousing](#), [Business Intelligence](#), [Analytics](#), [Data Science](#), [Data](#), [Masterdata](#), [Metadata](#), [Datakwaliteit](#), [Data Governance](#), [Self-service](#)

Company:

[DWBI Summit](#), [Adept Events](#)

Sjoerd Janssen

Senior Consultant Insights & Analytics bij Philips en lid redactieadviesraad BI-Platform.

[Alle blogs van deze auteur](#)

Gerelateerde Berichten

-

[04/04 - Cloudera kondigt twee Edge Data Management-oplossingen aan](#)

-

[04/04 - Dataiku kondigt release van Dataiku Free en Lite Edition aan](#)

-

[04/04 - SAS en Siemens sluiten partnerovereenkomst](#)