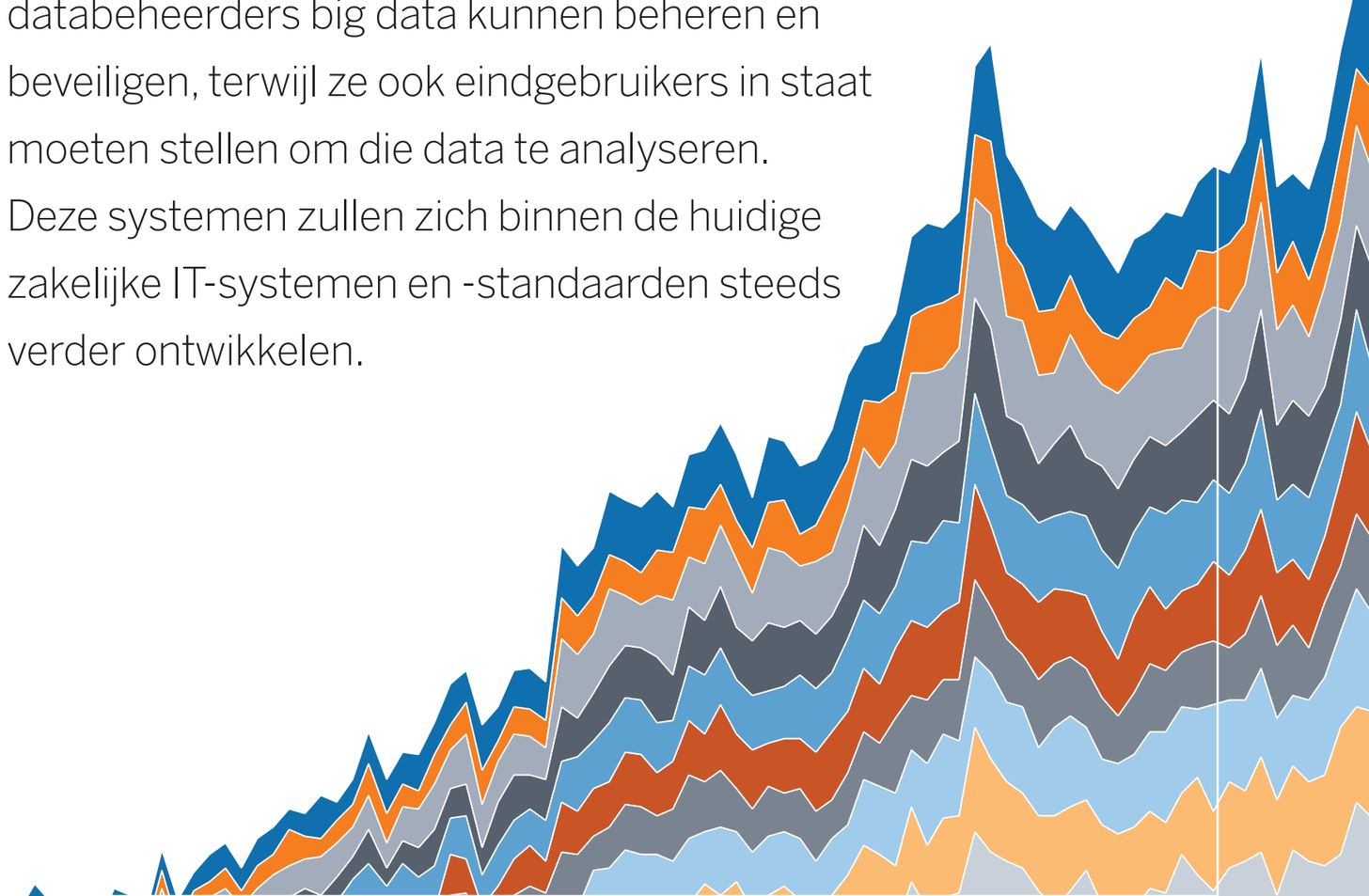
TOP TIEN

Big Data-TRENDS VOOR 2017

Bij Tableau kijken we jaarlijks naar wat er speelt in de branche. Onze bevindingen vormen de basis van onze lijst met de belangrijkste big data-trends voor het komende jaar. Dit zijn onze voorspellingen voor 2017.

De top 10 big data-trends voor 2017

2016 was een heel belangrijk jaar voor big data, aangezien steeds meer organisaties data in allerlei indelingen en bestandsgroottes zijn gaan opslaan en verwerken en de hieruit voortvloeiende bevindingen ook succesvol hebben toegepast. In 2017 zullen systemen die ondersteuning bieden voor grote hoeveelheden gestructureerde en ongestructureerde data, nog vaker worden geïmplementeerd. De markt heeft platformen nodig waarop databeheerders big data kunnen beheren en beveiligen, terwijl ze ook eindgebruikers in staat moeten stellen om die data te analyseren. Deze systemen zullen zich binnen de huidige zakelijke IT-systemen en -standaarden steeds verder ontwikkelen.



BIG DATA

1

Big data wordt snel en toegankelijk: er komen meer opties voor een snellere uitvoering van Hadoop

Natuurlijk kunt u machine learning toepassen en sentimentanalyses uitvoeren op Hadoop, maar de eerste vraag die mensen vaak stellen is: hoe snel is de interactieve SQL? SQL is uiteindelijk toch hét kanaal voor zakelijke gebruikers die Hadoop-data willen gebruiken voor snellere, herhaalbare KPI-dashboards en verkennende analyses.

Deze behoefte aan snelheid heeft een impuls gegeven aan het gebruik van snellere databases zoals Exasol en MemSQL, aan het gebruik van Hadoop-gebaseerde opslag zoals Kudu en aan het gebruik van technologieën die snelle query's mogelijk maken. Dankzij het gebruik van SQL-on-Hadoop-engines (Apache Impala, Hive LLAP, Presto, Phoenix en Drill) en OLAP-on-Hadoop-technologieën (AtScale, Jethro Data en Kyvos Insights) zorgen deze queryversnellers ervoor dat de grenzen tussen traditionele warehouses en de wereld van big data nog verder vervagen.

MEER INFORMATIE: AtScale BI on Hadoop benchmark Q4 2016 (AtScale BI op Hadoop, benchmark Q4 2016)

2

Big data is niet meer synoniem aan Hadoop: speciaal gemaakte Hadoop-tools worden overbodig

De afgelopen jaren zagen we tijdens de big data-golf verschillende technologieën opkomen om aan de behoefte aan analytische functies in Hadoop te voldoen. Bedrijven met complexe, heterogene omgevingen willen echter niet langer werken met afzonderlijke BI-toegangspunten per gegevensbron (zoals Hadoop).
De antwoorden op hun vragen zitten verstopt in vele databronnen, variërend van 'systems of record' tot cloud warehouses en van gestructureerde tot ongestructureerde data van zowel Hadoop- als niet-Hadoop-gebaseerde opslagsystemen (zelfs relationele databases zijn steeds meer voorbereid op big data. Zo is aan SQL Server 2016 onlangs JSON-ondersteuning toegevoegd).

In 2017 willen klanten allerlei soorten data kunnen analyseren. Data- en brononafhankelijke platformen zullen snel groeien, terwijl platformen die specifiek voor Hadoop zijn gemaakt en niet overal kunnen worden geïmplementeerd, zullen verdwijnen. De overname van Platfora is een eerste indicator van deze trend.

MEER INFORMATIE: Uncommon sense: The big data warehouse (Uncommon Sense: het big data-warehouse)

Organisaties maken direct optimaal gebruik van data lakes voor meer rendement

Een data lake is als een waterreservoir. Eerst maakt u een damwand (cluster) en vervolgens vult u het reservoir met water (data). Nadat het reservoir is gevuld, gebruikt u het water (de data) voor allerlei verschillende dingen, bijvoorbeeld om elektriciteit te genereren, als drinkwater of voor recreatie (voorspellende analyse, ML, cyberbeveiliging etc.).

Tot nu toe was het vullen van het reservoir een doel op zich. In 2017 gaat dat veranderen, omdat de zakelijke verantwoording voor Hadoop steeds strikter wordt. Organisaties zullen een herhaalbare en flexibele toepassing van het data lake vereisen om snellere informatievoorziening te realiseren. Organisaties zullen al hun bedrijfsresultaten zorgvuldig willen bekijken voordat ze investeren in personeel, data of infrastructuur. Op deze manier wordt een krachtigere samenwerking tussen bedrijf en IT gestimuleerd. Ook worden selfserviceplatformen steeds meer erkend als tool voor het optimaal benutten van big data-middelen.

MEER INFORMATIE: Maximizing data value with a data lake (De waarde van data maximaliseren met een data lake)

4

Architecturen worden volwassen en stappen af van uniforme structuren

Hadoop is niet meer slechts een batchverwerkingsplatform voor gebruikssituaties binnen de data science. Inmiddels is het een veelzijdige engine voor ad-hocanalyses geworden. Hadoop wordt zelfs gebruikt voor operationele rapportage over dagelijkse workloads; het soort rapportage dat vanouds wordt verwerkt door datawarehouses.

In 2017 zullen organisaties op deze behoefte aan hybride inspelen door zich binnen de architectuur te richten op situatiespecifieke ontwerpen. Voordat bedrijven zich toeleggen op een datastrategie, wordt er een groot aantal factoren nader onderzocht. Denk hierbij aan gebruikersrollen, vragen, volumes, toegangsfrequentie, datasnelheid en de mate van samenvoeging. Deze moderne architecturen worden afgestemd op de behoeften. Ze combineren de beste selfservicetools voor datavoorbereiding, Hadoop Core en analytische platformen voor eindgebruikers op dusdanige wijze dat deze tools opnieuw kunnen worden geconfigureerd zodra er sprake is van nieuwe behoeften. De flexibiliteit van deze architecturen zal uiteindelijk bepalen voor welke technologieën wordt gekozen.

MEER INFORMATIE: The cold/warm/hot framework and how it applies to your Hadoop strategy (Het aanpasbare framework en wat het betekent voor uw Hadoop-strategie)

5

Investeringen in big data worden vooral bepaald door variatie, niet door volume of snelheid

Gartner beschrijft big data met drie V's: informatiemiddelen met 'high-volume', 'high-velocity' en 'high-variety' (hoog volume, hoge snelheid en grote variatie). Hoewel alle drie de V's aan een opmars bezig zijn, lijkt variatie de grootste drijfveer voor investeringen in big data te worden. Dit blijkt uit een recent onderzoek van New Vantage Partners. Deze trend zet door omdat bedrijven steeds meer bronnen willen integreren en zich willen richten op de langetermijneffecten van big data. Er komen steeds meer verschillende data-indelingen bij, van JSON zonder schema tot geneste typen in andere databases (relationeel en NoSQL), tot niet-standaard data (Avro, Parquet, XML), waardoor men niet meer om het gebruik van connectors heen kan. In 2017 worden analytische platformen beoordeeld op de mogelijkheid om voor al deze verschillende bronnen rechtstreekse connectiviteit te bieden.

MEER INFORMATIE: Variety, not volume, is driving big data initiatives (Variatie en niet volume is de drijvende kracht achter big data-initiatieven)

6

Met Spark en machine learning wordt big data een belangrijke speler

Het big data-platform Apache Spark, ooit onderdeel van het Hadoop-ecosysteem, wint bij grote ondernemingen aan populariteit. Uit een onderzoek dat is uitgevoerd onder data-architecten, IT-managers en BI-analisten, blijkt dat bijna 70% van hen liever Spark dan MapReduce gebruikt. Dit is mogelijk toe te schrijven aan het feit dat MapReduce is gericht op batchverwerking en zich niet leent voor interactieve toepassingen of streamverwerking in real time.

Dankzij deze big-compute-on-big-data-capaciteiten zijn platformen met rekenintensieve functies voor machine learning, kunstmatige intelligentie en grafische algoritmen populair geworden. Vooral Microsoft Azure ML is in trek vanwege de gebruiksvriendelijkheid voor beginners en de eenvoudige integratie met bestaande Microsoft-platformen. Als ML beschikbaar wordt voor het grote publiek, zal dat leiden tot meer modellen en toepassingen (en de petabytes aan data die dit met zich meebrengt). Naarmate computers leren en systemen slimmer worden, zijn de ogen steeds meer gericht op providers van selfservicesoftware en hoe zij deze data toegankelijk kunnen maken voor de eindgebruiker.

MEER INFORMATIE: Why you should use Spark for machine learning (Waarom u machine learning Spark zou moeten gebruiken)

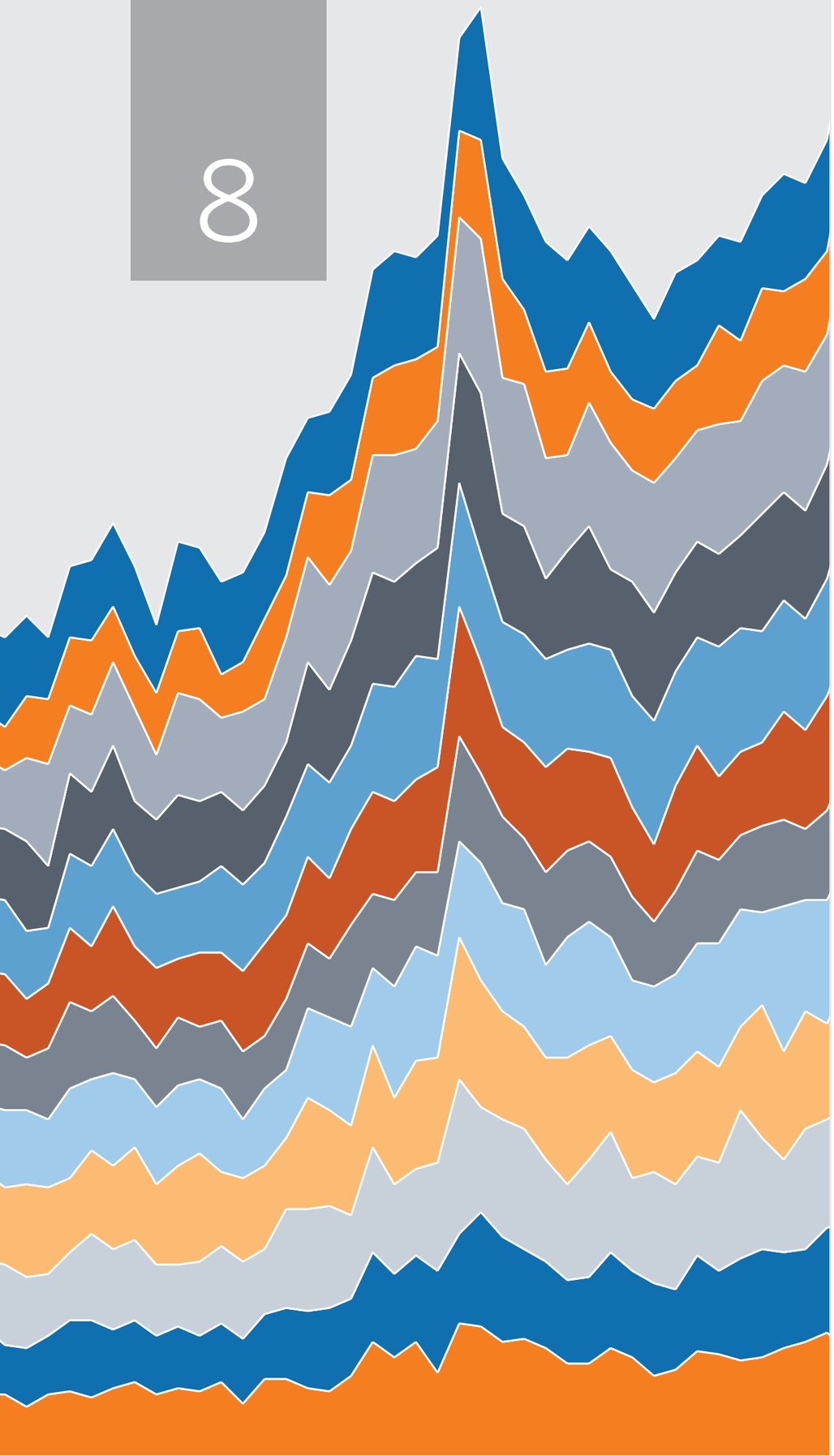
7

Het samenkomen van IoT, cloud en big data biedt nieuwe mogelijkheden voor selfserviceanalyse

Het lijkt erop dat alles in 2017 iets zal hebben dat informatie naar de thuisbasis zendt. IoT genereert enorme hoeveelheden gestructureerde en ongestructureerde data en deze data worden steeds meer geïmplementeerd in cloudservices. De data zijn vaak heterogeen en worden toegepast op verscheidene relationele en niet-relationele systemen, van Hadoop-clusters tot NoSQL-databases. Terwijl het proces waarmee data worden vastgelegd dankzij allerlei innovaties op het gebied van opslagservices en beheerde services steeds sneller is geworden, vormen de toegang tot en het inzicht in de data zelf nog steeds flinke uitdagingen. Daarom groeit de vraag naar analytische tools die moeiteloos verbinding kunnen maken met veel verschillende in de cloud gehoste databronnen en die deze bronnen ook kunnen combineren. Deze tools stellen bedrijven in staat om allerlei soorten en op verschillende locaties opgeslagen data te verkennen en visualiseren, en zo de verborgen mogelijkheden van hun IoT-investering te benutten.

MEER INFORMATIE: Tableau on solving IoT's last-mile challenge (Tableau over de laatste uitdagingen op het gebied van IoT)

8



Selfservice-datavoorbereiding wordt mainstream naarmate eindgebruikers meer invloed hebben op big data

Hadoop-data toegankelijk maken voor zakelijke gebruikers is een van de grootste uitdagingen van dit moment en de opkomst van analytische selfserviceplatformen heeft hier een positieve invloed op gehad. Maar zakelijke gebruikers willen graag nog minder tijd besteden aan het voorbereiden van data voor analyses en het proces eenvoudiger maken. Dit is vooral belangrijk als men te maken heeft met een variatie aan datatypen en -indelingen.

Met flexibele selfservice-datavoorbereidingstools kunnen Hadoop-data niet alleen bij de bron al worden voorbereid, maar kunnen er ook snapshots van de data worden ingezet om onderzoek sneller en eenvoudiger te maken. We hebben hier al heel veel innovaties besproken van bedrijven die zich bezighouden met de voorbereiding van big data voor eindgebruikers, zoals Alteryx, Trifacta en Paxata. Dankzij deze tools is het voor mensen die nu pas kennismaken met Hadoop gemakkelijker om in te stappen en daarom zullen ze in 2017 steeds populairder worden.

MEER INFORMATIE: Why self-service prep is a killer app for big data (Waarom een app voor eigen voorbereiding (selfservice) ideaal is voor big data)

9

Big data wordt volwassen: Hadoop draagt bij aan zakelijke standaarden

We zien dat Hadoop steeds meer een essentieel onderdeel wordt van het zakelijke IT-landschap. In 2017 gaat er nog meer geïnvesteerd worden in de beveiligings- en beheercomponenten van zakelijke systemen. Apache Sentry voorziet in een systeem voor verfijnde, rolgebaseerde autorisatie van data en metadata opgeslagen in een Hadoop-cluster. Apache Atlas is ontwikkeld als onderdeel van het databeheerinitiatief en geeft bedrijven de mogelijkheid binnen het volledige data-ecosysteem consistente dataclassificatie toe te passen. Apache Ranger voorziet in centraal beveiligingsbeheer voor Hadoop.

Tegenwoordig verwachten klanten dit soort functionaliteit van hun zakelijke RDBMS-platformen. Deze capaciteiten spelen een belangrijke rol in opkomende big data-technologieën, waardoor nog een drempel voor gebruik binnen het bedrijfsleven wordt weggenomen.

MEER INFORMATIE: The phases of Hadoop maturity: Where exactly is it going? (De ontwikkeling van Hadoop: waar gaat het heen?)

10

De opkomst van catalogussen met metadata helpt mensen bij het vinden van analysewaardige big data

Bedrijven hebben lange tijd data genegeerd of weggegooid omdat er te veel was om te kunnen verwerken. Met Hadoop kunnen bedrijven enorm veel data verwerken. Het nadeel is dat de data vaak niet zijn geordend op een manier die het doorzoeken ervan bevordert.

Catalogussen met metadata maken het gebruikers gemakkelijker om te zoeken naar en inzicht te krijgen in relevante data die geschikt zijn voor analyse met selfservicetools. Bedrijven als Alation en Waterline spelen hierop in. Deze bedrijven gebruiken machine learning om het zoeken naar data in Hadoop te automatiseren. Bestanden worden ingedeeld met behulp van tags, er worden relaties tussen datamiddelen vastgesteld en er worden middels doorzoekbare UI's zelfs querysuggesties gedaan. Zowel dataconsumenten als databeheerders zijn hiermee minder tijd kwijt aan het vinden en onderzoeken van betrouwbare data. In 2017 wordt men zich nog bewuster van selfserviceonderzoek en zal de vraag hiernaar toenemen (als een natuurlijke uitbreiding van selfserviceanalyses).

MEER INFORMATIE: Data catalogs as a strategic requirement for data lakes (Datacatalogussen als een strategische vereiste voor data lakes)

Over Tableau

De integratie van datavisualisatie in uw retailprogramma's en -processen is eenvoudiger dan u denkt.

Tableau Software helpt u om data inzichtelijk en begrijpelijk te maken, ongeacht hoeveel het is of in hoeveel systemen het is opgeslagen. Datadashboards kunnen snel worden gekoppeld, gecombineerd, gevisualiseerd en gedeeld. Naadloos van pc tot iPad. Maak en publiceer dashboards met geautomatiseerde data-updates en deel deze met collega's, partners of klanten. Programmeerervaring is niet nodig. Probeer het vandaag nog gratis uit.

TABLEAU.COM/TRIAL