

DE 10 HETASTE
Big data-
TRENDERNA FÖR 2017

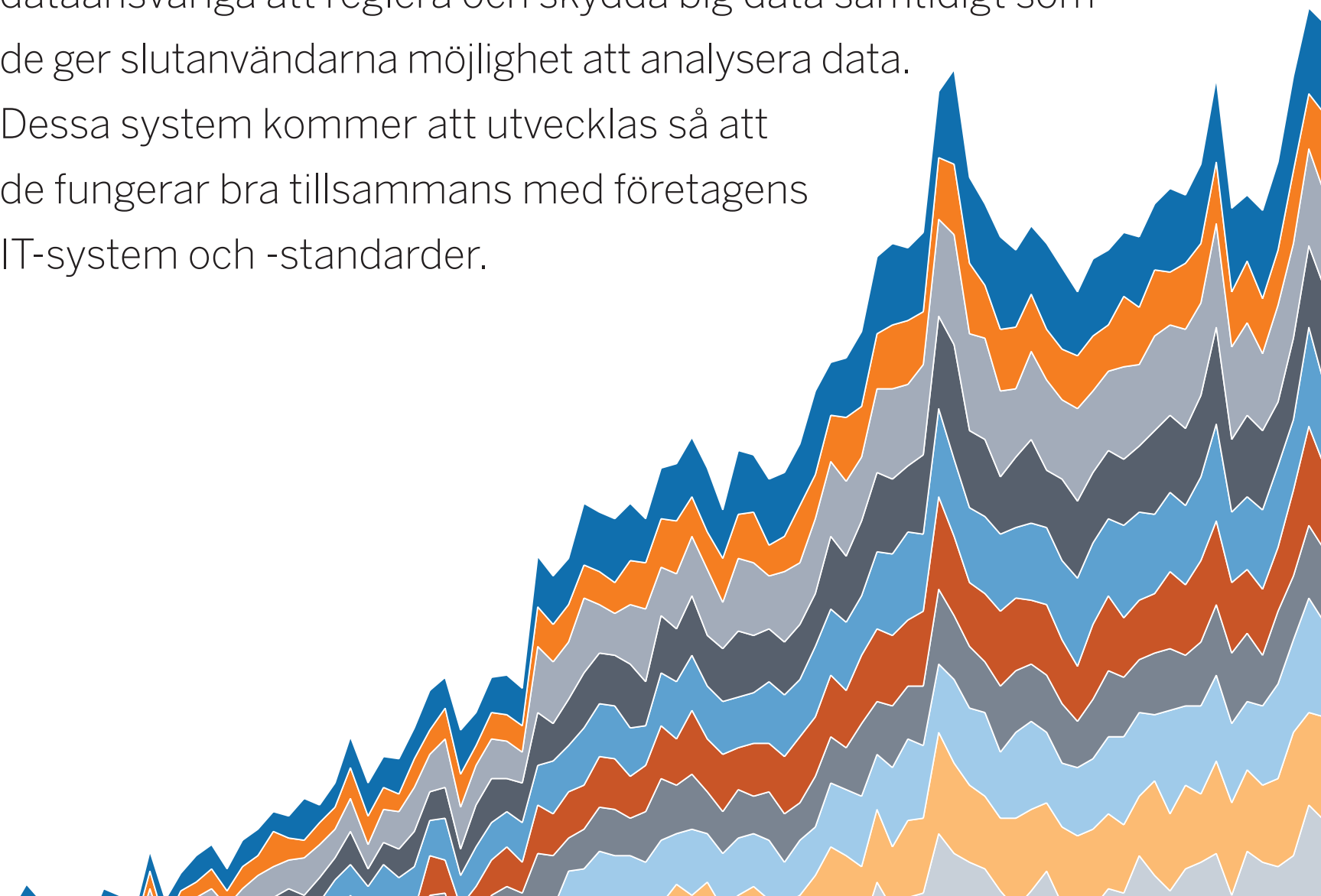




Varje år startar vi på Tableau ett samtal om vad som händer i branschen. Diskussionen leder fram till vår lista över de främsta big data-trenderna för det kommande året. Här är våra prognoser för 2017.

De 10 hetaste big data-trenderna 2017

2016 var en milstolpe för big data då fler företag lagrade, bearbetade och utvann värde från data i alla former och storlekar. 2017 kommer system som stödjer stora volymer av både strukturerad och ostrukturerad data att fortsätta stiga. Marknaden kommer att kräva plattformar som hjälper dataansvariga att reglera och skydda big data samtidigt som de ger slutanvändarna möjlighet att analysera data. Dessa system kommer att utvecklas så att de fungerar bra tillsammans med företagets IT-system och -standarder.



BIG DATA

1

Big data blir snabbt och lättillgängligt: Alternativ utvecklas för att göra Hadoop snabbare

Visst kan du utföra maskininlärning och genomföra attitydanalyser med Hadoop, men den första frågan som många ställer sig är: Hur snabb är den interaktiva SQL:en? SQL är trots allt kanalen för användare som vill använda Hadoop-data för snabbare, mer repeterbara KPI-dashboards samt explorativ analys.

Det här behovet av snabbhet har gjort att man börjat använda snabbare databaser som [Exasol](#) och [MemSQL](#), Hadoop-baserad lagring som [Kudu](#) och tekniker som möjliggör snabbare frågor. Användning av SQL-on-Hadoop-motorer ([Apache Impala](#), [Hive LLAP](#), [Presto](#), [Phoenix](#), och [Drill](#)) samt OLAP-on-Hadoop-tekniker ([AtScale](#), [Jethro Data](#) och [Kyvos Insights](#)). Dessa fråge-acceleratorer suddar ut gränserna mellan traditionell lagring och big data.

LÄS MER: [AtScale BI on Hadoop benchmark Q4 2016](#)

Big data är inte längre bara Hadoop: Specialbyggda verktyg för Hadoop har blivit föråldrade

Under tidigare år har vi sett flera tekniker öka med big data-vågen för att uppfylla behovet av analys på Hadoop. Men företag med komplexa, heterogena miljöer vill inte längre införa en siloåtkomstpunkt för BI och för en enda datakälla (Hadoop). Svaren på deras frågor begravs i en mängd källor som sträcker sig från registersystem och molnlager till strukturerad och ostrukturerad data från både Hadoop- och icke-Hadoop-källor. (För övrigt börjar även relationsdatabaser att bli redo för big data. SQL Server 2016 la t.ex. nyligen till stöd för JSON.)

2017 kommer kunder att kräva analyser på all data. Plattformer som är data- och källoberoende kommer att frodas medan de som är **specialbyggda för Hadoop** och inte kan distribueras för användningsfall kommer att falla i glömska. **Slutet för Platfora** är ett tidigt tecken på denna trend.

LÄS MER: [Inte så sunt förnuft: Big data-lager](#)



Företag utnyttjar datasjöar redan från start för att öka värdet

En datasjö är som en konstgjord reservoar. Först bygger du en fördämning i ena änden (ett kluster), sedan fyller du den med vatten (data). När du är färdig med sjön börjar du använda vattnet (data) för olika ändamål som att producera el, dricka och återskapa (prediktiv analys, ML, datasäkerhet osv.).

Fram till och med nu har påfyllningen av sjön varit ett mål i sig. 2017 kommer det att förändras när affärsmotiveringen för Hadoop skärps. Företagen kommer att kräva repeterbar och smidig användning av sjön för snabbare svar. De kommer att noga överväga affärsresultat innan de investerar i personal, data och infrastruktur. Detta kommer att främja ett starkare samarbete mellan [verksamheten och IT](#). Självserviceplattformar kommer att få djupare erkännande som ett verktyg för att utnyttja big data-resurser .

LÄS MER: [Maximera datavärdet med en datasjö](#)

4

Lösningar som är mogna att avvisa ramverk med samma storlek för alla

Hadoop är inte längre bara en gruppbearbetningsplattform för dataanalytiska användningsfall. Den har blivit ett mångsidig motor för ad hoc-analyser. Den används även för operativ rapportering om daglig arbetsbelastning av det slag som traditionellt hanteras av datalager.

Under 2017 kommer företag att svara på dessa hybridbehov genom att eftersträva lösningar som är specifikt utformade för användningsfallen. De kommer att undersöka en rad olika faktorer inklusive användaridentiteter, frågor, volymer, åtkomstfrekvens, datahastighet och aggregeringsnivå innan de bestämmer sig för en datastrategi. Dessa moderna referenslösningar kommer att vara behovsstyrda. De kommer att kombinera de bästa självserviceverktygen för dataprep, Hadoop Core och analysplattformar för slutanvändare på ett sätt som kan konfigureras utifrån hur dessa behov utvecklas. Vilken teknik man väljer kommer i slutändan att bero på flexibiliteten hos dessa lösningar.

LÄS MER: [The cold/warm/hot framework and how it applies to your Hadoop strategy](#)



5

Variation, inte volym eller hastighet, driver investeringar i big data

Gartner definierar big data som informationstillgångar med hög volym, hög hastighet, hög variation (kallas de tre V:na på engelska; volume, variety, velocity). Medan alla tre V:n växer är variation den enskilt största drivkraften för investeringar i big data. Det syns i resultaten från en **färsk undersökning** av nya Vantage Partners. Denna trend kommer att fortsätta växa eftersom företag försöker integrera flera källor och fokusera på "**den långa svansen**" av big data. Från schemafria JSON-databaser och kapslade typer i andra databaser (relationsdatabaser och NoSQL) till icke-platt data (Avro, Parkett, XML) – dataformaten ökar och anslutningar blir avgörande. 2017 kommer analysplattformar att utvärderas baserat på deras förmåga att ge direkta anslutningar till dessa olika källor.

LÄS MER: [Variety, not volume, is driving big data initiatives](#)

Spark och maskininlärning gör big data tydligt

[Apache Spark](#) var en gång en del av Hadoops ekosystem och är nu på väg att bli den big data-plattform som företag föredrar. I en [undersökning](#) med dataarkitekter, IT-chefer och BI-analytiker svarade nästan 70 % att de föredrog Spark framför det befintliga MapReduce, som är batch-orienterat och inte lämpar sig för interaktiva applikationer eller bearbetning av realtidsströmmar.

Dessa förmågor till big-compute-on-big-data har förbättrat plattformar med beräkningsintensiv maskininlärning, AI och grafalgoritmer. Framförallt Microsoft Azure ML har vuxit tack vare sin nybörjarvänlighet och enkla integration med befintliga Microsoft-plattformar. Att öppna upp ML till massorna kommer att leda till att det skapas fler modeller och tillämpningar som genererar flera petabyte av data. I takt med att maskiner lär sig och system blir smartare kommer alla att hålla ögonen på de leverantörer som tar fram mjukvara för självservice för att se hur de gör denna data lättillgänglig för slutanvändaren.

LÄS MER: [Why you should use Spark for machine learning](#)

7

Mötet mellan IoT, moln och big data skapar nya möjligheter för självserviceanalyser

Det verkar som om allt år 2017 kommer att ha en sensor som skickar information tillbaka till moderskeppet. IoT genererar stora volymer av strukturerad och ostrukturerad data och en allt större andel av denna [data används i molntjänster](#). Uppgifterna är ofta heterogena och finns i flera relations- och ickerelationssystem – från Hadoop-kluster till NoSQL-databaser. Medan innovationer inom lagring och hanterade tjänster har påskyndat insamlingsprocessen utgör åtkomst och förståelse av data fortfarande en viktig sista utmaning. Det gör att kraven växer på analytiska verktyg som sömlöst ansluter till en mängd olika molnbaserade datakällor och kombinerar dem. Sådana verktyg gör det möjligt för företag att utforska och visualisera alla typer av data, oavsett var de lagras. Detta hjälper dem att upptäcka dolda möjligheter som finns i IoT-investeringarna.

LÄS MER: [Tableau om att lösa IoT:s sista utmaning](#)

Självservice för dataförberedelser blir mainstream eftersom slutanvändare börjar forma big data

Att göra Hadoop-data åtkomlig för företagsanvändare är en av de största utmaningarna i vår tid. Ökningen av plattformar för självserviceanalyser har förbättrat denna resa. Men användarna vill förbereda sin data för analys snabbare och enklare vilket är särskilt viktigt för de som arbetar med flera datatyper och -format.

Rörliga självserviceverktyg för dataprep tillåter inte bara att Hadoop-data förbereds vid källan utan gör också data tillgänglig som ögonblicksbilder för snabbare och enklare utforskning. Vi har fått många innovationer inom området från företag som är fokuserade på slutanvändarens dataprep för big data såsom [Alteryx](#), [Trifacta](#) och [Paxata](#). Dessa verktyg sänker ribban för att delta för [sena Hadoop-anammare](#) och [eftersläntrare](#) och kommer att fortsätta att få dragkraft under 2017.

LÄS MER: [Varför självserviceprep är avgörande för big data](#)

Big data växer upp: Hadoop bidrar till företagens standarder

Vi ser en växande trend där Hadoop bli en central del av IT-landskapet för företag. Och under 2017 kommer vi att se fler investeringar i de säkerhets- och regleringskomponenter som omger affärssystem. Apache Sentry tillhandahåller ett system för att upprätthålla detaljerad, rollbaserad åtkomst till data och metadata som lagras på ett Hadoop-kluster. [Apache Atlas](#) skapades som en del av dataregleringsinitiativet och ger företag möjlighet att tillämpa konsekvent dataklassificering över ekosystemet för data. [Apache Ranger](#) ger centraliserad säkerhetsadministration för Hadoop.

Kunderna börjar förvänta sig dessa typer av funktioner från företagens RDBMS-plattformar. Dessa funktioner sätts främst i nya big data-tekniker och eliminerar därmed ytterligare en barriär mot företagens införande.

LÄS MER: [The phases of Hadoop maturity: Where exactly is it going?](#)

10

Kataloger för metadata hjälper människor att hitta big data som är värd att analysera

Under en lång tid kastade företag bort data eftersom de hade för mycket att bearbeta. Med Hadoop kan de behandla mycket data, men data är oftast inte organiserad på ett sätt så att de hittas.

Kataloger för metadata kan hjälpa användare att upptäcka och förstå relevant data som är värd att analysera med hjälp av självserviceverktyg. Denna lucka fylls av företag som [Alation](#) och [Waterline](#) som använder maskininlärning för att automatisera arbetet med att hitta data i Hadoop. De katalogiserar filer med taggar, upptäcker förhållanden mellan datatillgångar och ger även sökförslag via sökbara användargränssnitt. Det hjälper både datakonsumenter och -förvaltare att minska den tid det tar att lita på, hitta och söka i data på ett korrekt sätt. 2017 kommer vi att se ökad medvetenhet och större efterfrågan för självserviceinhämtning, vilket kommer att växa som en naturlig förlängning av självserviceanalyser.

LÄS MER: [Datakataloger som ett strategiskt krav för datasjöar](#)



Om Tableau

Det är enklare än du tror att integrera datavisualisering i dina detaljhandelsprogram och -processer.

Tableau Software hjälper människor att se och förstå information, oavsett hur mycket det är eller hur många system den lagras i. Anslut, blanda, visualisera och dela data-dashboards snabbt och sömlöst från datorn till iPad-enheten. Skapa och publicera kontrollpaneler med automatiska datauppdateringar och dela dem med kollegor, partner eller kunder – inga programmeringskunskaper krävs. Börja använda den kostnadsfria testversionen idag.

[TABLEAU.COM/TRIAL](https://tableau.com/trial)