TOP 10 TRENDURI

Big Data

PENTRU 2017

tableau

În fiecare an, Tableau inițiază o discuție despre ceea ce se întâmplă în domeniu. Discuția definește principalele tendințe în big data pentru anul următor. Acestea sunt predicțiile noastre pentru 2017.

Top 10 trenduri Big Data pentru 2017

2016 a fost un an de referință pentru big data, datorită faptului că din ce în ce mai multe organizații au început să stocheze, să proceseze și să valorifice date de toate formele și dimensiunile. În 2017, sistemele care susțin volume mari de date atât structurate, cât și nestructurate vor continua să ia amploare. Piața va cere platforme care să-i ajute pe custozii de date să controleze și să asigure protecția big data, oferindu-le în același timp utilizatorilor finali posibilitatea de a analiza datele. Aceste sisteme vor ajunge la maturitate, funcționând bine în interiorul sistemelor și standardelor IT de tip enterprise.

BIG DATA

1

Datele big data devin rapide și abordabile: opțiunile se extind, grăbind proiectul Hadoop

Desigur, puteți realiza machine learning și sentiment analysis pe platforma Hadoop, dar prima întrebare care se pune de obicei este: cât de rapid este SQL-ul interactiv? Până la urmă, SQL face legătura cu utilizatorii business care vor să folosească datele Hadoop pentru tablouri de bord KPI mai rapide și mai repetabile precum și pentru cercetare analitică.

Această nevoie de viteză a alimentat adoptarea bazelor de date mai rapide, precum Exasol și MemSQL, a aplicațiilor de stocare bazate pe platforma Hadoop, precum Kudu, și a tehnologiilor care permit interogări mai rapide. Prin folosirea motoarelor SQL-on-Hadoop (Apache Impala, Hive LLAP, Presto, Phoenix și Drill) și a tehnologiilor OLAP-on-Hadoop (AtScale, Jethro Data și Kyvos Insights), aceste acceleratoare de interogări au ca efect eliminarea diferențelor dintre depozitele de date tradiționale și universul big data.

LECTURĂ SUPLIMENTARĂ: BI-ul realizat de AtScale pe Hadoop benchmark în trimestrul 4 din 2016 („AtScale BI on Hadoop benchmark Q4 2016”)

2

Big Data nu mai e doar Hadoop: instrumentele destinate Hadoop devin învechite

În anii trecuți, am văzut cum câteva tehnologii au luat amploare odată cu valul big data pentru a satisface necesitățile de analiză în Hadoop. Dar companiile cu medii complexe și eterogene nu mai vor să adopte un punct de acces BI izolat, destinat unei singure surse de date (Hadoop). Răspunsurile de care au nevoie se găsesc într-o mulțime de surse, ce variază de la sisteme de înregistrare, până la depozite de date din cloud și la date structurate și nestructurate, atât de pe platforme Hadoop, cât și din alte surse. (Chiar și bazele de date relaționale sunt acum pregătite pentru big data. De exemplu, SQL Server 2016 a implementat de curând compatibilitatea cu JSON.)

În 2017, clienții vor cere analize pentru toate datele. Platformele care suportă mai multe tipuri de date și de surse se vor dezvolta, iar cele care sunt create special pentru Hadoop și nu pot fi implementate în orice situație nu vor mai fi folosite. Renunțarea la Platfora este un indicator al acestei tendințe.

LECTURĂ SUPLIMENTARĂ: Sens neobișnuit: depozitul pentru big data („Uncommon sense: The big data warehouse")

3

Organizațiile folosesc data lakes încă de la început pentru a crește valoarea adăugată

Un data lake este ca un lac de acumulare. Mai întâi construiești un baraj (un cluster), apoi lași lacul să se umple cu apă (datele). Odată ce lacul este umplut, începi să folosești apa (datele) în diverse scopuri, precum generarea de electricitate, apă potabilă sau activități de recreere (analize predictive, ML, securitate cibernetică etc.).

Până acum, menținerea nivelului de apă din lac a fost un scop în sine. În 2017, lucrurile se vor schimba, deoarece așteptările de la platforma Hadoop vor crește. Organizațiile vor avea nevoie să folosească lacul în mod repetabil și rapid pentru a obține răspunsuri mai prompte. Înainte de a investi în personal, date și infrastructură, organizațiile vor lua în considerare consecințele pentru business. Acest fapt va duce la crearea unui parteneriat mai strâns între business și IT, iar platformele self-service vor fi din ce în ce mai apreciate drept instrumentele care valorifică activele big data.

LECTURĂ SUPLIMENTARĂ: Maximizarea valorii datelor cu un data lake („Maximizing data value with a data lake”)

4

Arhitecturile ajung la maturitate, începând să respingă frameworkurile universale

Hadoop nu mai este doar o platformă de procesare în bloc a datelor utilizate în scopuri științifice. A devenit un motor multifuncțional pentru analize ad-hoc. A fost folosită chiar și pentru rapoarte operaționale zilnice, tipul de rapoarte prelucrate în mod tradițional prin depozitele de date.

În 2017, organizațiile vor răspunde acestor nevoi hibride cerând designuri de arhitecturi destinate unor utilizări specifice. Înainte de a alege o anumită strategie privind datele, organizațiile vor analiza mai mulți factori, precum așteptările utilizatorilor finali, întrebările, volumele de lucru, frecvența de accesare, viteza datelor și nivelul de agregare. Aceste arhitecturi moderne se vor baza pe nevoi. Vor combina cele mai bune instrumente self-service de pregătire a datelor, Hadoop Core și platforme de analiză pentru utilizatorii finali în moduri care să poată fi reconfigurate în funcție de evoluția nevoilor. În cele din urmă, flexibilitatea acestor arhitecturi va afecta opțiunile tehnologice.

LECTURĂ SUPLIMENTARĂ: Framework-ul rece/cald/fierbinte și cum se aplică strategiei dvs. pentru Hadoop („The cold/warm/hot framework and how it applies to your Hadoop strategy”)

5

Varietatea datelor atrage investițiile în big data, nu volumul sau viteza

Gartner definește big data ca acele date care respectă „cei trei V”: volum mare, viteză mare și varietate mare. Cu toate că toți cei trei V sunt în creștere, varietatea devine treptat principalul factor care generează investiții mari în big-data, așa cum rezultă dintr-un studiu recent realizat de New Vantage Parteners. Această tendință va continua să crească, deoarece companiile doresc să integreze mai multe surse și să se concentreze pe „coada lungă” a big data. De la formatele de date JSON fără schemă fixă, până la tipurile imbricate din alte baze de date (relaționale și NoSQL) sau la datele non-flat (Avro, Parquet sau XML), există tot mai multe formate de date iar conectorii devin esențiali. În 2017, platformele de analiză vor fi evaluate pe baza abilității de a oferi o conexiune directă, live cu aceste surse disparate.

LECTURĂ SUPLIMENTARĂ: Inițiativele big data sunt motivate de varietate, nu de volum („Variety, not volume, is driving big data initiatives”)

6

Spark și machine learning propulsează big data

Apache Spark, care era odată o componentă a ecosistemului Hadoop, devine acum platforma big data de referință pentru companii. Într-un sondaj la care au participat arhitecți de date, manageri IT și analiști BI, aproape 70% dintre respondenți au ales Spark în defavoarea MapReduce, care este orientat pe prelucrare în bloc și nu este compatibil cu aplicațiile interactive sau cu procesarea în timp real a fluxurilor.

Aceste capacități de calcule mari pentru big-data au propulsat platformele care oferă machine learning cu procese de calcul intensive, AI și algoritmi pe grafuri. Microsoft Azure ML, în special, a evoluat atât de mult datorită faptului că este ușor de folosit de către începători și se integrează ușor cu platformele Microsoft existente. Deschiderea ML către toate tipurile de utilizatori va duce la crearea mai multor modele și aplicații, care vor genera petabytes de date. Pe măsura ce sistemele învață și devin din ce în ce mai inteligente, toată atenția se va concentra asupra furnizorilor de software self-service pentru a observa cum fac ei aceste date abordabile pentru utilizatorii finali.

LECTURĂ SUPLIMENTARĂ: De ce să folosești Spark pentru machine learning („Why you should use Spark for machine learning”)

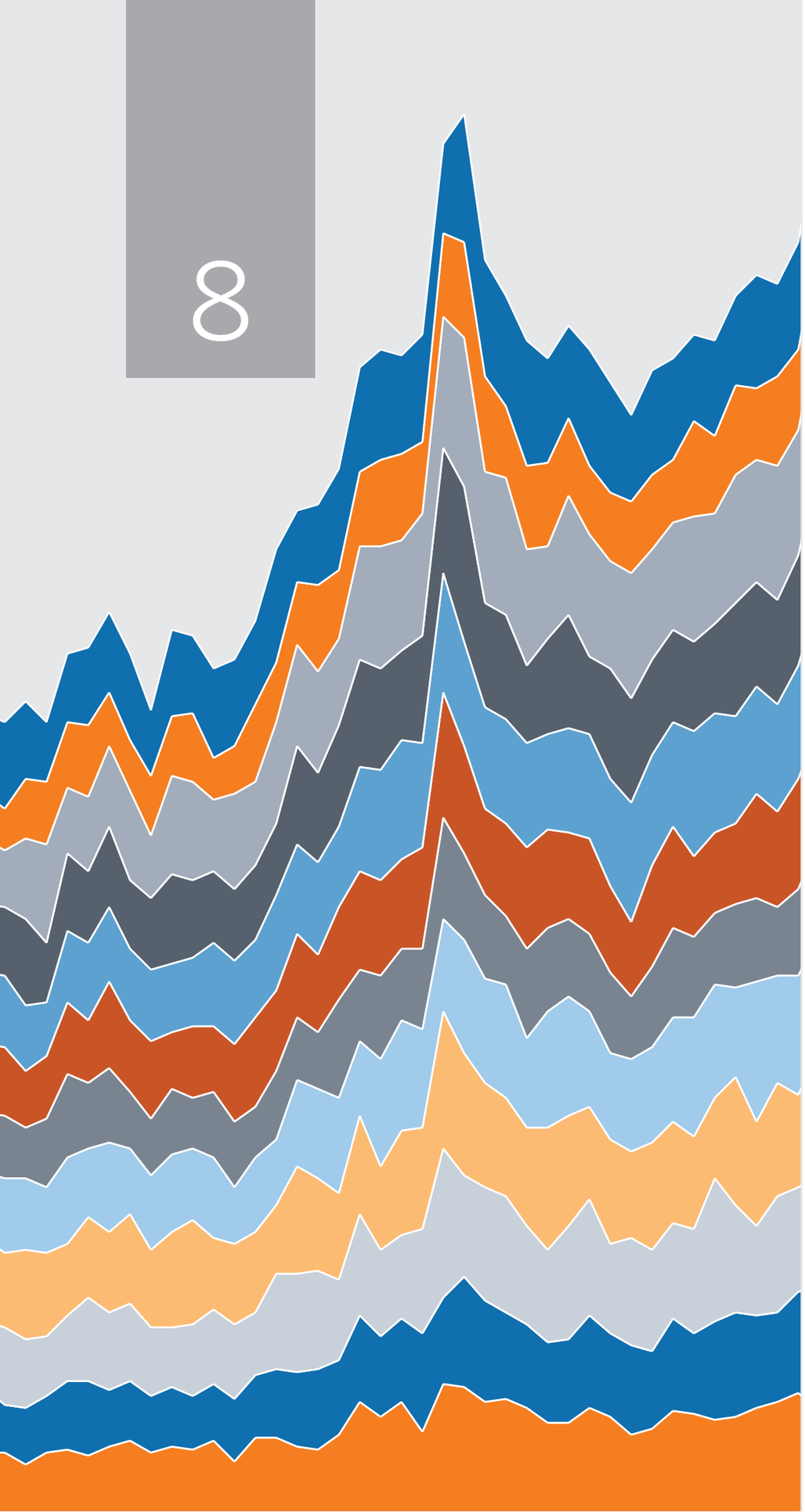
7

Convergența IoT, cloud și big data creează noi oportunități pentru analiza de tip self-service

Se pare că, în 2017, totul va avea un senzor care trimite informații înapoi la nava-mamă. IoT generează volume mari de date structurate și nestructurate, iar o parte tot mai mare din aceste date este trimisă către serviciile de tip cloud. Datele sunt deseori eterogene și se găsesc în mai multe sisteme relaționale și non-relaționale, de la clustere Hadoop până la baze de date NoSQL. În timp ce inovațiile în materie de servicii de stocare și servicii administrate au grăbit procesul de captare a datelor, accesarea și înțelegerea datelor pun încă probleme semnificative. Ca urmare, crește cererea pentru instrumente de analiză care se conectează și se îmbină perfect cu o gamă largă de surse de date bazate pe cloud. Aceste instrumente le permit companiilor să exploreze și să vizualizeze orice tip de date, indiferent unde sunt stocate, ceea ce le ajută să descopere o oportunitate nebănuită oferită de investițiile lor în IoT.

LECTURĂ SUPLIMENTARĂ: Tableau despre rezolvarea ultimei provocări pentru IoT („Tableau on solving IoT's last-mile challenge")

8



Pregătirea self-service a datelor devine dominantă, pe măsură ce utilizatorii finali încep să influențeze big data

Una dintre cele mai mari provocări de astăzi este transformarea datelor Hadoop în așa fel încât să fie accesibile utilizatorilor business. Dezvoltarea platformelor care oferă analiză self-service a ușurat puțin această situație. Dar utilizatorii business vor să reducă și mai mult timpul și complexitatea pregătirii datelor pentru analiză, ceea ce este extrem de important când este vorba de o gamă variată de tipuri și formate de date.

Instrumentele rapide de pregătire self-service a datelor nu doar permit datelor Hadoop să fie pregătite la sursă, ci le și fac disponibile sub formă de snapshoturi pentru o explorare mai ușoară și mai rapidă. Am văzut în acest domeniu o mulțime de inovații de la companii care se concentrează pe pregătirea datelor utilizatorilor finali pentru big data, cum ar fi Alteryx, Trifacta și Paxata. Aceste instrumente simplifică adoptarea platformei Hadoop de către cei care nu au adoptat-o încă și ele vor continua să se dezvolte în 2017.

LECTURĂ SUPLIMENTARĂ: De ce pregătirea self-service este o aplicație esențială pentru big data („Why self-service prep is a killer app for big data”)

9

Big data crește: Hadoop contribuie la standardele enterprise

Asistăm la o tendință de creștere a Hadoop, care devine o parte esențială a peisajului IT enterprise. Iar în 2017, vom observa o creștere a investițiilor în componentele de securitate și administrare din jurul sistemelor enterprise. Apache Sentry oferă un sistem pentru autorizarea accesului diferențiat, pe bază de roluri, la datele și metadatele stocate pe un cluster Hadoop. Apache Atlas, creat ca parte a inițiativei pentru administrarea datelor, le permite organizațiilor să aplice clasificări coerente ale datelor în întregul ecosistem de date. Apache Ranger permite administrarea centralizată a securității pentru Hadoop.

Clienții încep să se aștepte ca toate platformele RDBMS enterprise să ofere acest tip de capacități. Aceste capacități ajung treptat în prim-planul tehnologiilor big data în curs de dezvoltare, făcând astfel și mai ușoară adaptarea acestor tehnologii de către companii.

LECTURĂ SUPLIMENTARĂ: Etapele maturizării Hadoop: încotro se îndreaptă? („The phases of Hadoop maturity: Where exactly is it going?”)

10

Progresul cataloagelor cu metadate ajută lumea să găsească big data care merită să fie analizate

Mult timp, companiile au fost nevoite să arunce date pentru că aveau prea multe de procesat. Cu Hadoop, pot procesa o mulțime de date, dar, în general, acestea nu sunt organizate în așa fel încât să fie ușor de găsit.

Cataloagele cu metadate îi pot ajuta pe utilizatori să descopere și să înțeleagă datele relevante care merită analizate cu instrumente self-service. Această necesitate a clienților este acoperită de companii precum Alation și Waterline, care folosesc machine learning pentru a automatiza căutarea datelor în Hadoop. Acestea cataloghează fișierele folosind taguri, descoperă relațiile dintre date și chiar oferă sugestii de interogări prin UI-uri în care se poate căuta. Astfel se reduce timpul necesar pentru căutarea și interogarea corectă a datelor, atât pentru consumatorii, cât și pentru administratorii de date. În 2017, va crește vizibilitatea și cererea pentru capacitățile de descoperire de tip self-service, care se vor dezvolta ca o extensie naturală a analizei self-service.

LECTURĂ SUPLIMENTARĂ: Cataloagele cu date - cerință strategică pentru data lakes („Data catalogs as a strategic requirement for data lakes")

Despre Tableau

Integrarea vizualizării datelor în programele și procesele retail este mai simplă decât credeți.

Tableau ajută oamenii să vizualizeze și să-și înțeleagă datele. Conectați-vă rapid, combinați, vizualizați și distribuiți tablouri de bord pentru date de pe PC pe iPad. Creați și publicați rapoarte și dashboard-uri interactive cu actualizări automate ale datelor și distribuiți-le colegilor, partenerilor sau clienților. Nu sunt necesare cunoștințe în domeniul programării. Încercați-l gratuit chiar de azi.

TABLEAU.COM/TRIAL