أعلى عشرة

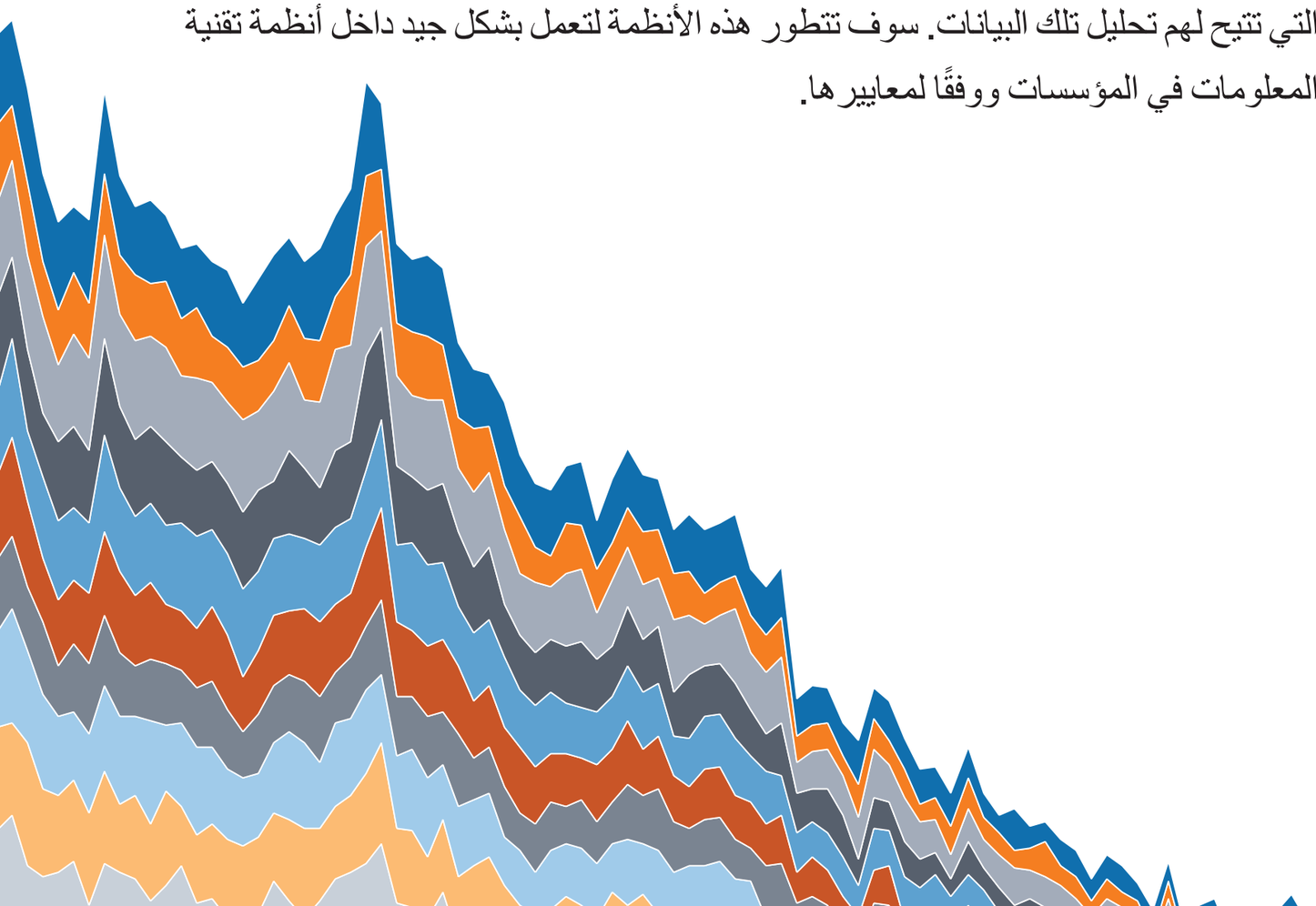
مؤشرات بخصوص البيانات الضخمة

لعام ٢٠١٧

أعلى ١٠ مؤشرات بخصوص البيانات الضخمة لعام ٢٠١٧

كان عام ٢٠١٦ حدثًا بارزًا بالنسبة للبيانات الضخمة لما شهده من قيام المزيد من المؤسسات بتخزين البيانات من كافة الأشكال والأحجام، ومعالجتها، واستخلاص القيمة منها. ومن المتوقع أن يشهد عام ٢٠١٧ تواصل بروز الأنظمة التي تدعم الأحجام الكبيرة لكلّ من البيانات المنظمة وغير المنظمة. وسوف يتواصل طلب السوق على الأنظمة الأساسية التي ستساعد مسؤولي حماية البيانات في إدارة البيانات الضخمة وتأمينها بينما يتم تزويد المستخدمين النهائيين بالإمكانات التي تتيح لهم تحليل تلك البيانات. سوف تتطور هذه الأنظمة لتعمل بشكل جيد داخل أنظمة تقنية المعلومات في المؤسسات ووفقًا لمعاييرها.

كل عام في Tableau، نبدأ حوارًا حول ما يدور في أروقة الصناعة. يقود هذا النقاش قائمتنا التي تضم أعلى المؤشرات بخصوص البيانات الضخمة لدينا للعام التالي. وهذه هي توقعاتنا لعام ٢٠١٧.



البيانات الضخمة تصبح سريعة وسهلة المنال: تتوسع الخيارات لزيادة سرعة Hadoop

١

بالتأكيد، يمكنك تطبيق تعلّم الآلة وتنفيذ تحليل الآراء والعواطف على برنامج Hadoop، ولكن السؤال الأول الذي غالبًا ما يتم طرحه هو: ما مدى سرعة SQL التفاعلية؟ مع ذلك، تمثل SQL السبيل الذي ينبغي أن يرتاده المستخدمون في الشركات ممن يريدون استخدام بيانات Hadoop للحصول على لوحات معلومات أكثر سرعة وقابليةً للتكرار لمؤشر الأداء الرئيسي (KPI)، وكذلك لإجراء التحليل الاستكشافي.

هذه الحاجة إلى السرعة كانت بمثابة الوقود الذي حفز تبنّي قواعد بيانات أكثر سرعة، مثل Exasol وMemSQL، ومستودعات تخزين قائمة على Hadoop، مثل Kudu، وتقنيات تمكّن الاستعلامات بسرعة أعلى. باستخدام محرّكات SQL-on-Hadoop (Apache Impala، وHive LLAP، وPresto، وPhoenix، وDrill) وتقنيات OLAP-on-Hadoop (AtScale، وJethro Data، وKyvos Insights)، تقوم مسرّعات الاستعلام هذه بتقليل وضوح الخطوط الفاصلة بين مستودعات التخزين التقليدية وعالم البيانات الضخمة.

قراءة المزيد: الاختبار المعياري لذكاء الأعمال على Hadoop من AtScale للربع الرابع من عام ٢٠١٦

٢

لم يعد سوق البيانات الضخمة مقتصرًا على Hadoop: فالأدوات المصممة لإنجاز أغراض معينة لـ Hadoop أصبحت جزءًا من الماضي

وقد لاحظنا، في السنوات السابقة، بروز العديد من التقنيات مع موجة البيانات الضخمة للوفاء بالحاجة إلى التحليلات على Hadoop. لكن المؤسسات ذات البيئات المعقدة غير المتجانسة لم تعد تمتلك الرغبة في تبنّي نقطة وصول منفصلة لذكاء الأعمال لأجل مصدر بيانات واحد فقط (Hadoop). وتكون الإجابات عن أسئلتها مغمورة في مجموعة من المصادر التي تتراوح بين أنظمة السجلات وأنظمة تخزين Cloud، والبيانات محددة البنية والبيانات غير محددة البنية من كلٍ من المصادر التي تنتمي إلى Hadoop والمصادر غير المنتمية له. (حتى قواعد البيانات الارتباطية تصبح دون قصد جاهزة لدعم البيانات الضخمة. أضاف SQL Server ٢٠١٦ مؤخرًا، على سبيل المثال، دعم JSON.)

في ٢٠١٧، سوف يزيد طلب العملاء على تحليلات بشأن كل البيانات. سوف تزدهر الأنظمة الأساسية غير محددة المصدر وغير محددة البيانات، بينما تتساقط على جانبي الطريق تلك الأنظمة المصممة لتحقيق أغراض خاصة بـ Hadoop والتي سيفشل نشرها عبر حالات الاستخدام. يمثل خروج Platfora من حلبة المنافسة مؤشرًا مبكرًا لهذا الاتجاه.

قراءة المزيد: إحساس غير مألوف: مستودع البيانات الضخمة

٣

تحقيق المؤسسات لأقصى استفادة من مستودعات البيانات الأولية من البداية لتحفيز القيمة

يشبه مستودع البيانات الأولية خزانًا من صنع الإنسان. ففي البداية، تسدّ الطرف (تبني مقطع تخزين)، ثم تتركه حتى يمتلئ بالمياه (البيانات). بمجرد أن تنشئ البحيرة، تبدأ في استعمال المياه (البيانات) للعديد من الأغراض، مثل توليد الكهرباء، والشرب، والاستجمام (التحليلات التنبؤية، تعلّم الآلة (ML)، أمان تقنية المعلومات، وما إلى ذلك).

حتى الآن، كان تزويد البحيرة بالمياه غاية في حد ذاته. في ٢٠١٧، سوف يتغير هذا الأمر عندما تبلغ الجدوى الاقتصادية لاستخدام Hadoop مبلغ الإحكام. سوف يزيد طلب المؤسسات على الاستخدام المرن والقابل للتكرار لمستودعات البيانات الأولية للحصول على إجابات أسرع. وسوف يضعون المحصلات الاقتصادية في الاعتبار بحذر قبل ضخ الاستثمارات في الأفراد، والبيانات، والبنية الأساسية. وهذا من شأنه أن يعزّز شراكة أقوى بين الشركة وقسم تقنية المعلومات. وسوف تكتسب الأنظمة الأساسية ذاتية الخدمة تقديرًا أعمق بوصفها أداة للاستفادة من أصول البيانات الضخمة.

قراءة المزيد: زيادة قيمة البيانات لأقصى حد باستخدام مستودعات البيانات الأولية

٤

التطور الطارئ على البنى يجعل الحلول ذات الحجم الواحد الذي يناسب كل أُطُر العمل أمرًا مرفوضًا

لم يعد Hadoop مجرد نظام أساسي لمعالجة دفعة من الملفات للاستفادة بها في حالات الاستخدام الخاصة بعلم البيانات. فلقد أصبح محركًا متعدد الأغراض للتحليل المخصص. بل إنه أصبح مُستخدمًا لإعداد التقارير التشغيلية عن أعباء العمل اليومية — وهي المهمة التي كانت تتولاها مستودعات البيانات عادةً.

في ٢٠١٧، سوف تتجاوب المؤسسات مع هذه الاحتياجات المختلطة من خلال مواصلة استخدام تصميم البنى محدد الحالة. كما أنها سوف تجري أبحاثًا لدراسة مجموعة من العوامل، بما في ذلك شخصيات المستخدمين، والأسئلة، والأحجام، ومعدل الوصول، وسرعة البيانات، ومستوى التجميع، قبل الالتزام باستراتيجية بيانات. سوف تكون هذه البنى ذات المرجعية الحديثة مدفوعة بغرض الوفاء بالاحتياجات. وسوف تجمع بين أفضل أدوات تجهيز البيانات ذاتية الخدمة، وHadoop Core، والأنظمة الأساسية لتحليلات المستخدم النهائي بوسائل قابلة لإعادة التهيئة كلما تطورت تلك الاحتياجات. في نهاية الأمر، سوف تقود المرونة التي تتسم بها هذه البنى الاختيارات التقنية.

قراءة المزيد: إطار العمل الخامل/متوسط النشاط/النشط وكيفية يمكن تطبيقه على استراتيجية Hadoop لديك

٥

التنوع، وليس الحجم أو السرعة، هو ما يحفز ضخ الاستثمارات في قطاع البيانات الضخمة

يُعرِّف Gartner البيانات الضخمة بأنها عبارة عن أصول المعلومات كبيرة الحجم عالية السرعة فائقة التنوع. ومع تنامي هذه السمات الثلاث، يصبح التنوع المحرك الأكبر الوحيد للاستثمارات في قطاع البيانات الضخمة كما شوهد في نتائج دراسة استقصائية حديثة نفذتها شركة New Vantage Partners. سوف يستمر هذا المؤشر في التنامي بينما تسعى الشركات إلى دمج المزيد من المصادر وتركز على "قائمة مطولة" من البيانات الضخمة. بدءًا من تنسيق JSON الخالي من المخططات، إلى الأنواع المضمّنة في قواعد بيانات أخرى (الارتباطية وNoSQL)، وحتى البيانات غير المُسطحة (XML، Parquet، Avro)، تتضاعف تنسيقات البيانات وتكتسب الموصّلات أهمية بالغة. في ٢٠١٧، سوف يتم تقييم الأنظمة الأساسية للتحليلات استنادًا إلى قدرتها على توفير إمكانية اتصال مباشرة وحيوية بتلك المصادر المتفاوتة.

قراءة المزيد: التنوع، وليس الحجم، هو ما يحفز مبادرات البيانات الضخمة

٦

محرك Spark وتعلّم الآلة يضيئان سماء البيانات الضخمة

الآن يصبح محرك Apache Spark، والذي كان ذات يوم أحد مكوّنات نظام Hadoop للاتصال المشترك، النظام الأساسي للبيانات الضخمة المفضل لدى المؤسسات. في دراسة استقصائية شملت مصمّمي بنى البيانات، ومديري تقنية المعلومات، ومحلّلي أنظمة ذكاء الأعمال، فضّل ما يقرب من ٧٠٪ ممن شملتهم الدراسة Spark على المحرك الحالي MapReduce، والذي يعتمد على معالجة دفعات الملفات ولا يتلاءم مع التطبيقات التفاعلية أو معالجة تدفقات البيانات في الوقت الفعلي.

لقد أدت إمكانات إجراء عمليات الحوسبة الضخمة على البيانات الضخمة إلى الرفع من شأن الأنظمة الأساسية التي تتضمن ميزات تعلّم الآلة كثيف الحوسبة، والذكاء الاصطناعي (AI)، وخوارزميات الرسم البياني. وقد كانت انطلاقة Microsoft Azure ML تحديدًا بفضل سهولة استخدامه بالنسبة للمبتدئين وسهولة تكامله مع أنظمة Microsoft الأساسية القائمة. سيؤدي فتح أبواب تعلّم الآلة (ML) على مصراعيها أمام الجماهير إلى إنشاء المزيد من النماذج والتطبيقات التي تنشئ بدورها كميات هائلة من البيانات تقدر بالبيتابايت. وبينما تتعلّم الآلات وتكتسب الأنظمة الذكاء، ستتوجه كل الأنظار إلى موفري البرامج ذاتية الخدمة لمعرفة كيفية تسهيل مهمة وصول المستخدم النهائي إلى هذه البيانات.

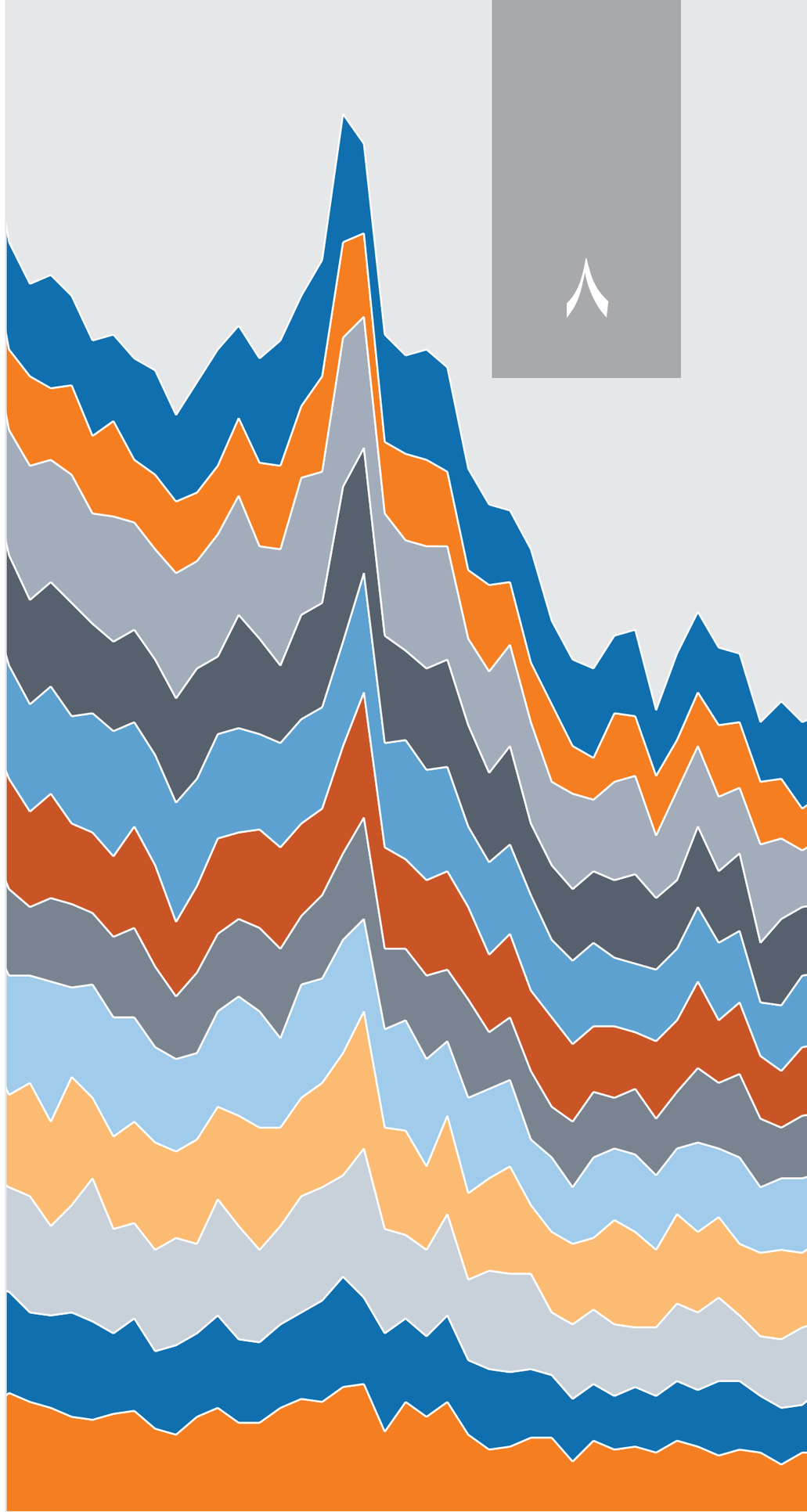
قراءة المزيد: الأسباب التي توجب عليك استخدام Spark لتعلم الآلات

٧

الجمع بين إنترنت الأشياء، وCloud، والبيانات الضخمة يخلق فرصًا جديدة للتحليلات ذاتية الخدمة

يبدو أن كل شيء في عام ٢٠١٧ سيحتوي على ما يشبه المستشعر الذي يعيد إرسال المعلومات إلى المركبة الأم. فنظام إنترنت الأشياء (IoT) ينتج كميات هائلة من البيانات محددة البنية والبيانات غير محددة البنية، وحصة متزايدة من هذه البيانات يجري نشرها على خدمات Cloud. تكون البيانات غالبًا غير متجانسة ويتم استيعابها عبر العديد من الأنظمة الارتباطية والأنظمة غير الارتباطية، بدءًا من مقاطع تخزين Hadoop وحتى قواعد بيانات NoSQL. بينما زادت الابتكارات في قطاع التخزين والخدمات المدارة من سرعة عملية جمع البيانات، فإن الوصول إلى البيانات نفسها وفهم مضمونها لا يزال يشّكل تحديًا كبيرًا في المراحل الأخيرة. ونتيجة لذلك، يتزايد الطلب على الأدوات التحليلية التي تتصل بمجموعة واسعة من مصادر البيانات التي تستضيفها Cloud وتقوم بدمجها بسلاسة. تتيح مثل هذه الأدوات للشركات إمكانية استكشاف أي نوع من البيانات المخزّنة وعرضها من أي مكان، مما يساعدها على اكتشاف الفرصة الكامنة في الاستثمار في إنترنت الأشياء (IoT).

قراءة المزيد: جهود Tableau في التغلب على تحدي المرحلة الأخيرة لإنترنت الأشياء (IoT)



٨

تجهيز البيانات ذاتية الخدمة يصبح اتجاهًا سائدًا نظرًا لبدء المستخدمين النهائيين في إنشاء البيانات الضخمة

يمثل تزويد المستخدمين في الشركات بإمكانية الوصول إلى بيانات Hadoop أحد أكبر التحديات التي يشهدها عصرنا. وقد ساعد بروز نجم الأنظمة الأساسية الخاصة بالتحليلات ذاتية الخدمة في تحسين هذه المهمة. لكن المستخدمين في الشركات يرغبون في اختصار المزيد من الوقت وتخفيف التعقيد الذي يتميز به تجهيز البيانات للتحليل، وهو ما يعد أمرًا هامًا بشكل خاص عند التعامل مع مجموعة متنوعة من أنواع البيانات والتنسيقات.

لا تسمح أدوات تجهيز البيانات ذاتية الخدمة فائقة المرونة فقط بالتجهيز المُسبق لبيانات Hadoop عند المصدر، لكنها أيضًا تجعل البيانات متوفرة على هيئة لقطات لمنحك إمكانية استكشافها بسرعة أكبر وسهولة أكثر. لقد شهدنا مجموعة من الابتكارات في هذه المساحة من الشركات التي تركّز على تجهيز بيانات المستخدم النهائي للبيانات الضخمة، مثل Alteryx، وTrifacta، وPaxata. تحدّ هذه الأدوات من عوائق الدخول بالنسبة إلى الشركات التي تبنت Hadoop مؤخرًا والأخرى التي تتباطأ عن تبنيه وستستمر في اكتساب شعبية خلال عام ٢٠١٧.

قراءة المزيد: لماذا تعدّ أدوات التجهيز ذاتي الخدمة بمثابة البرنامج التطبيقي الأنسب للبيانات الضخمة

٩

البيانات الضخمة آخذة في التزايد: Hadoop يؤثر بشكل متزايد على المعايير المؤسسية

نلاحظ مؤشرًا متزايدًا على أن Hadoop يصبح جزءًا جوهريًا من المشهد في قطاع تقنية المعلومات داخل المؤسسات. وفي عام ٢٠١٧، سنشاهد المزيد من الاستثمارات التي تُضخ في مكوّنات الأمان والتحكم تحيط بأنظمة المؤسسات. تقدم Apache Sentry نظامًا لتطبيق المصادقة المتطورة والقائمة على الأدوار على البيانات والبيانات الوصفية المُخزّنة على مقطع التخزين في Hadoop . يقوم Apache Atlas، الذي تم إنشاؤه في إطار مبادرة التحكم في البيانات، بمنح المؤسسات القدرة على تطبيق تصنيف متناسق للبيانات عبر نظام الاتصال المشترك للبيانات. يوفر Apache Ranger إدارة أمان مركزية لـ Hadoop.

يبدأ العملاء في توقع هذه الأنواع من الإمكانات من أنظمة RDBMS الأساسية من فئة المؤسسات. تستحوذ هذه الإمكانات على موضع الصدارة بين تقنيات البيانات الضخمة الناشئة، وبذلك تقضي على عائق آخر يحول دون تبنّي المؤسسات لهذه الحلول.

قراءة المزيد: مراحل نضج Hadoop: إلى أين يتجه تحديدًا؟

١٠

بروز كتالوجات البيانات الوصفية يساعد المستخدمين في العثور على بيانات ضخمة جديرة بالتحليل

لفترة طويلة، كانت الشركات تتخلص من البيانات نظرًا لأن معالجتها كانت تستهلك أكثر مما ينبغي.
وقد أصبحت قادرة، بفضل Hadoop، على معالجة كميات هائلة من البيانات، لكن البيانات بصفة عامة ليست في صيغة منظمة تسمح بالعثور عليها.

ومن هنا يمكن لكتالوجات البيانات الوصفية أن تساعد المستخدمين في اكتشاف البيانات ذات الصلة والجديرة بالتحليل وفهمها باستخدام أدوات ذاتية الخدمة. هذه الفجوة في احتياجات العميل يجري سدّها بجهود شركات، مثل Alation وWaterline واللتين تستخدمان تعلّم الآلة لأتمتة مهمة العثور على البيانات في Hadoop.
فهما تقومان بإنشاء كتالوجات للملفات باستخدام بطاقات التعريف، وتميطان اللثام عن العلاقات التي تربط بين أصول البيانات، بل وتوفران أيضًا اقتراحات للاستعلامات عن طريق واجهات المستخدم القابلة للبحث. من شأن هذا أن يساعد كلاً من مستهلكي البيانات والمسؤولين عنها في اختصار الوقت الذي يلزم للوثوق بالبيانات، والعثور عليها، والاستعلام الدقيق عنها. في عام ٢٠١٧، سنشهد المزيد من الوعي والطلب على الاكتشاف ذاتي الخدمة، والذي سيتزايد كامتداد طبيعي للتحليلات ذاتية الخدمة.

قراءة المزيد: كتالوجات البيانات كمطلب استراتيجي لمستودعات البيانات الأولية

حول Tableau

إن توظيف عرض البيانات في برامج البيع بالتجزئة وعملياته أسهل مما تعتقد.

تساعد برامج Tableau المستخدمين في مشاهدة البيانات وفهم مضمونها بعض النظر عن ضخامتها، أو عدد الأنظمة المُخزّنة فيها. يمكنك الاتصال السريع بلوحات معلومات البيانات ومزجها وعرضها ومشاركتها عبر تجربة سلسة من الكمبيوتر الشخصي إلى جهاز iPad. يمكنك إنشاء لوحات معلومات تتضمن تحديثات تلقائية للبيانات ونشرها ومشاركتها مع الزملاء أو الشركاء أو العملاء — دون الحاجة لمهارات برمجة. ابدأ تجربة مجانية اليوم.

TABLEAU.COM/TRIAL