禪

رشد بزرگ دادهها، تبدیل هادوپ به یک استاندارد سازمانی

10 روند برتر بزرگ دادهها در سال 2017 که تاثیرشان را در سالهای آتی نشان خواهند داد

2017
THE YEAR OF INTELLIGENCE

2016 را میتوان سالی پررونق در عرصه دادههای حجیم دانست، زیرا سازمانهای بیشتری به ذخیره، پردازش و استخراج همه نوع فرمت و اندازهای از داده پرداختند . در سال 2017 نیز سیستمهایی که از دادهها در مقیاس بزرگ در هر دو شکل ساختیافته و بدون ساختار پشتیبانی میکنند، در حال اضافه شدن هستند. این بازار به سکوهایی نیاز خواهد داشت که از دادهها نگهداری کند و آنها را در زمان استفاده کاربر نهایی برای تجزیه و تحلیل امن باقی نگه دارد . این سیستمها بهخوبی در داخل سیستمهای فناوری اطلاعات و استانداردهای سازمانی کار میکنند . اما برای آنکه با کم و کیف این تغییرات در سال جاری بیشتر آشنا شوید، 10 روند برتر بزرگ دادهها در سال جاری را مورد بررسی قرار دادهایم .

این مطلب یکی از مقالات پرونده ویژه «دادههای بزرگ؛ فردای بزرگتر» شماره **197** ماهنامه شبکه است. علاقهمندان میتوانند کل این پرونده ویژه را از روی سایت شبکه دانلود کنند .

1- دادههای حجیم سریع و قابل دسترس خواهند شد

همین حالا هم میتوانید در هادوپ (Hadoop) از یادگیری ماشینی استفاده کنید و به تجزیه و تحلیل دادهها بپردازید، اما اولین سؤالی که مردم اغلب میپرسند این است که SQL تعاملی تا چه میزان سریع

است؟ SQL کانالی برای کاربران تجاری است که میخواهند با استفاده از هادوپ دسترسی سریعتر و تکرارپذیرتری به دادهها داشته باشند . این نیاز بیشتر از طریق پایگاههای داده سریعتر و سازگار مثل Exasol و MemSQL، ذخیرهسازهای مبتنی بر هادوپ مثل Kudu و فناوریهایی که امکان کوئریگیری سریعتری را فراهم میکنند امکانپذیر خواهد شد . استفاده از SQL روی موتورهای هادوپ (Apache Impala, Hive LLAP, Presto, Phoenix و Drill) و فناوریهای OLAP روی هادوپ (AtScale, Jethro Data و Kyvos Insights) باعث میشوند این شتابدهندههای کوئری فاصله و شکاف بین ذخیرهسازهای سنتی و دنیای دادههای حجیم را از بین ببرند .

2- دادههای حجیم تنها در اختیار هادوپ نخواهند ماند

ابزارهای هدف ساخته شده برای هادوپ در حال منسوخ شدن هستند. طی سالهای گذشته شاهد حضور چند فناوری در زمینه دادههای حجیم بودهایم که نیاز به تجزیه و تحلیل در هادوپ را محقق میکردند . اما سازمانهایی متشکل از محیطهای پیچیده و ناهمگن دیگر نمیخواهند تنها به یک منبع داده متکی باشند . آنها برای رفع نیاز خود به منابعی احتیاج دارند که به سیستمهای متکی به کلاود متصل هستند و میتوانند دادههای ساختیافته و بدون ساختار را از منابع هادوپ و غیرهادوپ مدیریت کنند . از طرفی، حتی پایگاههای داده رابطهای نیز برای دادههای حجیم آماده شدهاند . برای نمونه، SQL Server 2016 اخیراً پشتیبانی از JSON را به قابلیتهای خود اضافه کرده است . در سال 2017 مصرفکنندگان متقاضی تجزیه و تحلیل بر روی تمام انواع داده خواهند بود . سکوهایی که توانایی مدیریت داده و منابع agnostic را داشته باشند رونق زیادی پیدا خواهند کرد و آنها که تنها برای هادوپ ساخته شدهاند و موارد استفاده محدودتری دارند، از رده خارج خواهند شد . میتوان از Platfora بهعنوان شاخص اولیه این روند یاد کرد .

3- سازمانها از مخازنی شبیه به دریاچههای داده بهره خواهند برد

یک دریاچه داده شبیه به یک مخزن آب ساخته دست انسان است. اول شما سد را بنا میکنید (ساخت یک کلاستر)، بعد آب (داده) را به آن وارد میکنید . هنگامی که دریاچه احداث شد، شروع به استفاده از این آب (داده) برای مقاصد مختلف مثل تولید برق، نوشیدن و بازسازی (تجزیه و تحیل داده، ML، امنیت سایبری وغیره) میکنید . تا همین چند وقت گذشته پر کردن این دریاچه آخرین کاری بود که باید انجام میشد. اما در سال 2017 توجیه کسب و کار برای هادوپ پراهمیتتر شده و خواهد شد . سازمانها برای دریافت پاسخ سریعتر درخواستهای مکرر و سریعتری را برای استفاده از این دریاچه خواهند داشت . آنها قبل از سرمایهگذاری روی کارکنان، داده و زیرساختها بهدقت نتایج کسب و کار را بررسی میکنند . چنین کاری یک مشارکت قویتر بین کسب و کار و صنعت آیتی را فراهم میکند .

مطلب پیشنهادی

دادههای بزرگ؛ فردای بزرگتر
BIG DATA

دانلود کنید : پرونده ویژه دادههای بزرگ؛ فردای بزرگتر

4- معماریهای پخته و کامل ساختار متناسب برای تمام چهارچوبها را رد میکنند

هادوپ دیگر تنها یک سکوی پردازش دستهای برای استفاده در علوم داده نخواهد بود . هادوپ به یک موتور چندمنظوره برای تجزیه و تحلیل داده تبدیل شده است و حتی برای گزارشهای عملیاتی چرخهای کاری روزانه نیز از آن استفاده میشود . در سال 2017، سازمانها با پیگیری استفاده از طراحی معماری مورد خاص به این نیازهای ترکیبی پاسخ خواهند داد . آنها قبل از هرگونه اقدام به انجام یک استراتژی برای

[illegible] Hadoop [illegible] Core [illegible].

[illegible]

[illegible]

[illegible] JSON [illegible] Avro, Parquet, XML [illegible] 2017 [illegible].

[illegible]

[illegible]

[illegible]

[illegible] Spark

[illegible] Apache Spark [illegible] BI [illegible] MapReduce [illegible] Spark [illegible] 70 [illegible].
[illegible] Azure ML [illegible] ML [illegible].

[illegible]

2017 [redacted]

7- [redacted]

[redacted] Apache Atlas [redacted]

8- [redacted]

[redacted] Alteryx, [redacted] Paxata и Trifacta [redacted] 2017 [redacted]

[redacted]

[10 ابزار متن‌باز برای مدیریت امنیت داده‌ها در محیط‌های کلان داده](#)

9- پروژه‌های متن‌باز برای مدیریت امنیت داده‌ها در محیط‌های کلان داده

در محیط‌های کلان داده به دلیل حجم بالای داده‌ها و تنوع منابع داده‌ها، مدیریت امنیت داده‌ها بسیار پیچیده است. در سال 2017 نیز پروژه‌های متن‌باز زیادی برای مدیریت امنیت داده‌ها در محیط‌های کلان داده ارائه شده‌اند. Apache Sentry یکی از این پروژه‌ها است که به مدیریت دسترسی به داده‌ها در محیط‌های کلان داده کمک می‌کند. Apache Atlas نیز یکی دیگر از این پروژه‌ها است که به مدیریت متادیتا و مدیریت دسترسی به داده‌ها کمک می‌کند. Apache Ranger نیز یکی دیگر از این پروژه‌ها است که به مدیریت دسترسی به داده‌ها در محیط‌های کلان داده کمک می‌کند. این پروژه‌ها از RDBMS برای ذخیره‌سازی متادیتا استفاده می‌کنند و به مدیران سیستم امکان می‌دهند تا دسترسی به داده‌ها را به صورت دقیق کنترل کنند.

10- ابزارهای متن‌باز برای مدیریت متادیتا در محیط‌های کلان داده

ابزارهای متن‌باز برای مدیریت متادیتا در محیط‌های کلان داده نیز بسیار مهم هستند. این ابزارها به کاربران کمک می‌کنند تا داده‌های خود را به راحتی پیدا کنند و از آن‌ها استفاده کنند. Informatica, Alation و Waterline از جمله ابزارهای متن‌باز برای مدیریت متادیتا در محیط‌های کلان داده هستند. این ابزارها از RDBMS برای ذخیره‌سازی متادیتا استفاده می‌کنند و به کاربران امکان می‌دهند تا داده‌های خود را به صورت دقیق کنترل کنند. این ابزارها در سال 2017 نیز پیشرفت‌های زیادی داشته‌اند.

:نویسنده
[مجید صفاری]
:منبع
Tableau
:مترجم
[مجید صفاری]
:تاریخ انتشار
11:55 - 28/10/1396
:برچسب‌ها
[کلان داده] - [داده‌های بزرگ] - big data - [داده‌های بزرگ] - [کلان داده]

:آدرس
https://www.shabakeh-mag.com/cover-story/10790/10-%D8%B1%D9%88%D9%86%D8%AF-%D8%A8%D8%B1%D8%AA%D8%B1-%D8%A8%D8%B2%D8%B1%DA%AF-%D8%AF%D8%A7%D8%AF%D9%87%E2%80%8C%D9%87%D8%A7-%D8%AF%D8%B1-%D8%B3%D8%A7%D9%84-2017-%DA%A9%D9%87-%D8%AA%D8%A7%D8%AB%DB%8C%D8%B1%D8%B4%D8%A7%D9%86-%D8%B1%D8%A7-%D8%AF%D8%B1-%D8%B3%D8%A7%D9%84%E2%80%8C%D9%87%D8%A7%DB%8C-

%D8%A2%D8%AA%DB%8C-%D9%86%D8%B4%D8%A7%D9%86-
%D8%AE%D9%88%D8%A7%D9%87%D9%86%D8%AF