



امروزه بسیاری از سازمان‌ها به دنبال آن هستند تا از بهترین ابزارهای مرتبط با بزرگ داده‌ها در جهت دستیابی به چشم‌اندازهای از پیش تعیین شده خود استفاده کنند. اما نیل به این هدف معطوف به درک درست پرسش‌هایی است که در ادامه آن‌ها را مشاهده خواهیم کرد.

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.

در جریان برگزاری کنفرانس شرکت Dell EMC در ماه می سال جاری، کوری مینتون مهندس بخش سیستمی شرکت دل به روی صحنه رفت و به رهبران فناوری اطلاعات نشان داد به‌کارگیری بزرگ داده‌ها چگونه می‌تواند چشم‌اندازهای ناپیدا را پیش روی شرکت‌ها قرار دهد و دید بازتری را به رهبران کسب و کار اعطا کند. بزرگ داده‌ها به کسب و کارها وعده داده‌اند که سودآوری آن‌ها را بیش از پیش افزایش دهند، اما برای بسیاری از کسب و کارها دستیابی به این مهم کاری دشوار است، به واسطه آنکه بسیاری از سازمان‌ها به درستی نمی‌دانند باید چه معماری را به کار گیرند یا ابزارهایی را مستقر کنند و مهم‌تر از آن اساساً به چه زیرساختی نیاز دارند. از آمار توصیفی گرفته تا مدل‌های پیش‌بینی کننده و حتی هوش مصنوعی همه این فناوری‌ها نیاز مبرمی به بزرگ داده‌ها دارند. امروزه بسیاری از سازمان‌ها به دنبال آن هستند تا از بهترین ابزارهای مرتبط با بزرگ داده‌ها در جهت دستیابی به چشم‌اندازهای از پیش تعیین شده خود استفاده کنند. اما نیل به این هدف معطوف به درک درست پرسش‌هایی است که در ادامه آن‌ها را مشاهده خواهیم کرد.

افزایش ضریب هوشمندی شرکت‌ها و به‌ویژه بنگاه‌های اقتصادی یکی از مهم‌ترین دلایل به‌کارگیری بزرگ داده است. اما سامانه‌ها، سرویس‌ها یا محصولاتی که از سوی شرکت‌ها خریداری یا ساخته می‌شوند زمانی کارکرد واقعی خود را نشان می‌دهند که ابتدا با داده‌های مناسب تغذیه شوند و در ادامه به پردازش داده‌ها پردازند. در چنین شرایطی است که این بنگاه‌ها می‌توانند از تحلیل‌های هوشمندانه مبتنی بر داده‌ها برای اخذ تصمیمات تطبیق‌پذیر و پیش‌دستانه استفاده کنند. در کنفرانس امسال شرکت دل که ماه می 2017 برگزار شد، کوری مینتون مهندس سامانه‌های تحلیلگر دل به شش پرسشی اشاره کرد که شرکت‌ها و کسب و کارها پیش از آنکه به حوزه بزرگ داده‌ها وارد شوند، بهتر است آن را مطرح کنند. در این مقاله به بررسی این شش پرسش خواهیم پرداخت.

1. خرید کردن در مقابل ساختن

اولین پرسشی که سازمان شما پیش از ورود به حوزه **بزرگ داده‌ها** باید آن را مطرح کند این است که بهتر است یک سرویس/ سامانه بزرگ داده‌ها را خریداری یا خود از ابتدا چنین سامانه/ سرویسی را ایجاد کند؟ محصولات محبوبی

همچون Teradata, SAS, SAP و Splunk را می‌توان خریداری و به‌سادگی پیاده‌سازی کرد. البته سازمان‌ها می‌توانند از Hortonworks, Cloudera, Databricks و Apache Flink به‌منظور ساخت یک سرویس/ سامانه بزرگ داده‌ها استفاده کنند. مینتون می‌گوید: «سازمان‌هایی که تصمیم می‌گیرند محصولات آماده را خریداری کنند به‌دنبال آن هستند تا در کوتاه‌مدت ارزش‌آفرینی کنند. این کار دو مزیت مهم برای آن‌ها به همراه دارد، اول آنکه خریداری یک سامانه/ سرویس یا محصول کار ساده‌تری است و دوم آنکه به‌کارگیری آن در دسر کمتری دارد. با این حال، برای دستیابی به این سادگی باید هزینه مربوط را پرداخت کنید که غالباً این سامانه‌ها با قیمت بالایی در اختیار سازمان‌ها قرار می‌گیرند.

همچنین، برای آنکه این ابزارها در شرایطی مطلوب کار کنند باید تا حد امکان از تنوع اطلاعاتی واردشونده به این سامانه‌ها کم کرد. اگر سازمان شما با یک فروشنده شناخته شده در ارتباط است، گزینه ایده‌آل این است که از محصولات و ابزارهای جدیدی که از سوی یک فروشنده معروف در ارتباط با محصولات بزرگ داده طراحی شده استفاده کنید.» اما در طرف مقابل سازمان‌ها گزینه دیگری که همانا ساخت چنین سامانه‌هایی است را در اختیار دارند. بیشتر ابزارهای محبوبی که برای ساخت یک سامانه بزرگ داده‌ها به کار گرفته می‌شوند اغلب قیمت کمی دارند یا رایگان هستند.

به‌کارگیری این ابزارها به سازمان‌ها اجازه می‌دهد روی یک جریان ارزشی منحصر به فرد سرمایه‌گذاری کنند. زمانی که تصمیم می‌گیرد از این ابزارها به‌منظور ساخت سامانه‌های بزرگ داده‌ها استفاده کنید عموماً فرصت‌ها یا به عبارت دقیق‌تر کانال‌های زیادی را به‌لحاظ تنوع و گسترش‌پذیری می‌توانید به وجود آورید، اما این رویکرد می‌تواند کار را بیش از اندازه پیچیده کند. یکی از بزرگ‌ترین مشکلات مدیران با چنین سامانه‌هایی در ارتباط با قابلیت همکاری است.

بیشتر ابزارهای محبوبی که برای ساخت یک سامانه بزرگ داده‌ها به کار گرفته می‌شوند اغلب قیمت کمی دارند یا رایگان هستند. به‌کارگیری این ابزارها به سازمان‌ها اجازه می‌دهد روی یک جریان ارزشی منحصر به فرد سرمایه‌گذاری کنند

20000 داده‌های جریانی در مقابل داده‌های انبوه/ دسته‌ای

مینتون می‌گوید: «داده‌های انبوه/ دسته‌ای از سوی محصولاتی همچون اوراکل، هادوپ، مپ ردیوس و آپاچی اسپارک مدیریت می‌شوند. این ابزارها قادرند حجم بسیار بالایی از داده‌ها را مدیریت کنند.» این ابزارها قابل برنامه‌ریزی هستند و اغلب به‌منظور ساخت مجموعه‌ای از داده‌ها که از سوی متخصصان علم داده برای آزمایش‌ها مورد استفاده قرار می‌گیرند به کار گرفته می‌شوند. محصولاتی همچون Apache Kafka و Flink در ارتباط با داده‌های جریانی به کار گرفته می‌شوند. این ابزارها در تعامل با داده‌های جریانی به‌منظور ایجاد مدل‌های پیش‌بینی کننده به کار گرفته می‌شوند. مشکلی که داده‌های جریانی دارند این است که سرعت آن‌ها بالا است و همین موضوع باعث می‌شود کمتر بتوانیم به آن‌ها اعتماد کنیم. اما ویژگی مثبت داده‌های جریانی در این است که متنوع و وسیع هستند. داده‌های جریانی عمدتاً برای سازمان‌هایی مناسب هستند که فرهنگ دوآپس به‌خوبی در آن سازمان‌ها رعایت می‌شود.

مطلب پیشنهادی



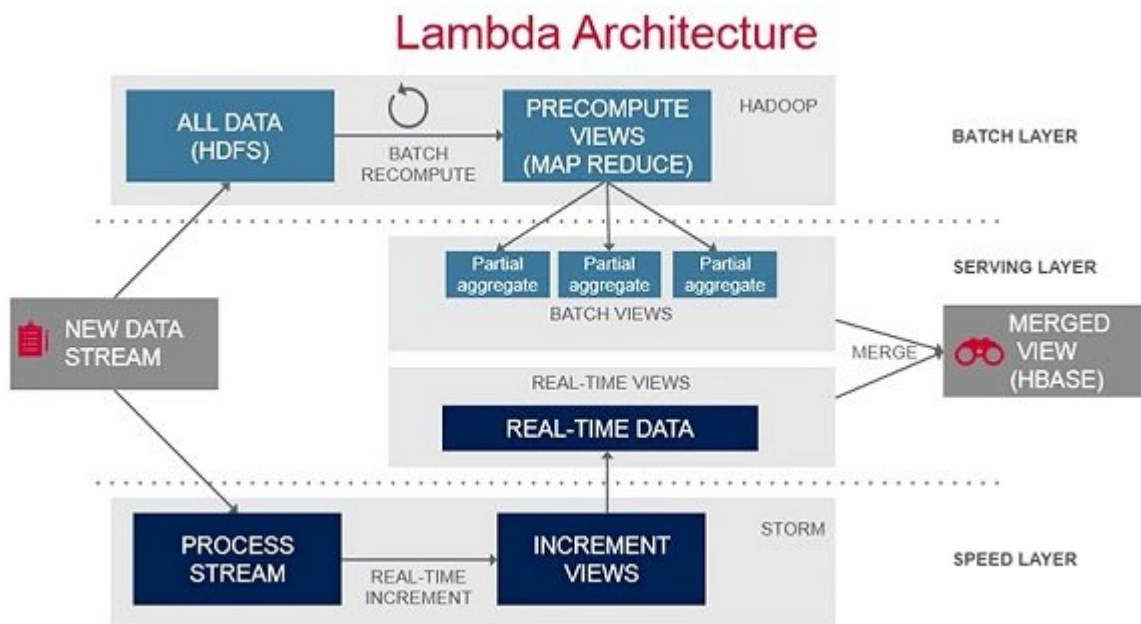
دانلود کنید: پرونده ویژه داده‌های بزرگ؛ فردای بزرگ‌تر

30000 کاپا در مقابل معماری لامبدا

تویتر مثال خوبی از معماری لامبدا است. در این معماری داده‌ها به دو مسیر تقسیم می‌شوند. گروهی از داده‌ها به‌منظور تغذیه لایه سرعت به‌منظور ارائه بینش سریع مورد استفاده قرار می‌گیرند، در حالی که گروه دیگری از داده‌ها به لایه سرویس‌دهی انتقال داده می‌شوند. در این **معماری داده‌های** جریانی واردشونده به سیستم به‌منظور تغذیه دولا به batch و speed مورد استفاده قرار می‌گیرند. (شکل 1) مینتون می‌گوید: «مدل فوق به یک سازمان اجازه می‌دهد به هر دو رویکرد بینش جریانی و انبوه دسترسی پیدا کند و در نتیجه میان استریم‌ها تعادل برقرار می‌کند. چالشی که در ارتباط با این معماری وجود دارد این است که شما هم‌زمان دو نوع کدنویسی و دو برنامه را باید مدیریت کنید.» معماری کاپا همه چیز را در قالب یک جریان نشان می‌دهد، اما جریانی است که هدفش حفظ

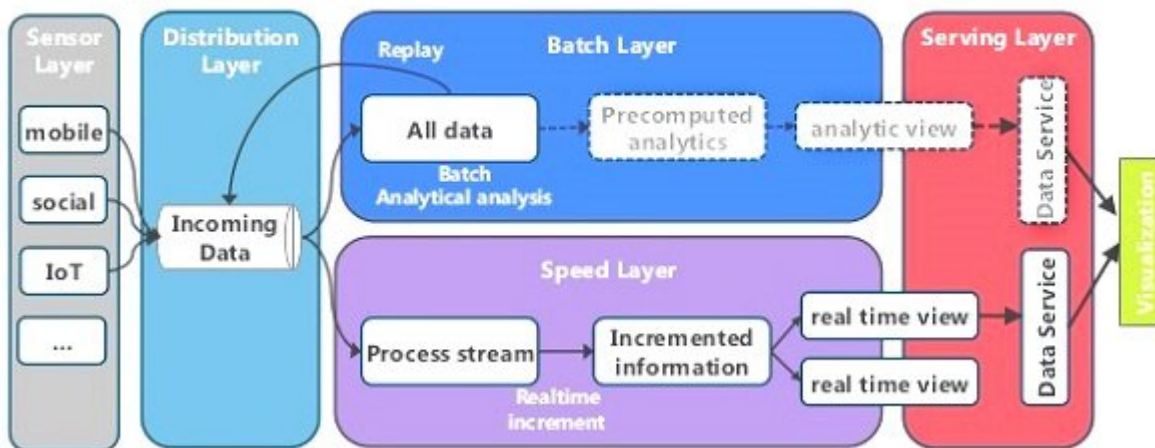
اعتبار داده‌ها و ارائه پردازش بی‌درنگ است (شکل 2). همه داده‌ها به یک ورودی تغییرناپذیر نوشته می‌شوند و در ادامه تغییرات با این داده‌های تغییرناپذیر مورد ارزیابی قرار می‌گیرند. این رویکرد مؤثر است، به واسطه آنکه به کدنویسی کمتری احتیاج دارد و همچنین مدلی را ارائه می‌کند که به اعتقاد میتون برای سازمانی که تازه کار خود را با بزرگ داده‌ها آغاز کرده مناسب است.

شکل 1- معماری لامبدا



شکل 2- معماری کاپا

■ "Kappa Architecture"



تویتر مثال خوبی از معماری لامبدا است. در این معماری داده‌ها به دو مسیر تقسیم می‌شوند. گروهی از داده‌ها به منظور تغذیه لایه سرعت به منظور ارائه بینش سریع مورد استفاده قرار می‌گیرند، در حالی که گروه دیگری از داده‌ها به لایه سرویس‌دهی انتقال داده می‌شوند

4. کلاود عمومی در مقایسه با کلاود خصوصی

کلاودهای عمومی و خصوصی یکی از ملزومات بزرگ داده‌ها هستند و از بسیاری جهات به یکدیگر شباهت‌هایی دارند. برای شروع، یک سازمان باید این موضوع را مورد بررسی قرار دهد که چه محیطی بر اساس استعدادها و درون‌سازمانی گزینه مناسبی برای شروع کار است. نیاز به داده‌ها، امنیت، انطباق و الگوهای مصرفی انعطاف‌پذیر از دیگر عواملی هستند که باید مورد توجه قرار گیرند.

50005 مجازی در مقابل فیزیکی

مینتون می‌گوید: «سال‌ها پیش بحث درباره زیرساخت‌های فیزیکی و مجازی به شدت گرم بود. با این حال فناوری مجازی‌ساز به رشد خود ادامه داد و به یکی از رقبای سرسخت تجهیزات فیزیکی تبدیل شد. به طوری که امروزه در زمینه استقرار بزرگ داده‌ها هر دو فناوری تقریباً مشابه یکدیگر هستند.»

اما اینکه کدامیک از این دو مدل برای کسب و کار شما مناسب است به تصمیمات مدیریتی و اینکه مدیران با کدامیک از این فناوری‌ها احساس راحتی می‌کنند و همچنین چه زیرساخت‌هایی در حال حاضر در سازمان شما وجود دارد بستگی دارد.

مطلب پیشنهادی



تصورات اشتباه در ارتباط با مدل‌های داده‌ای با 10 باور نادرست در ارتباط با بزرگ داده‌ها آشنا شوید

60006 DAS در مقابل NAS

مینتون می‌گوید: «سامانه‌های ذخیره‌ساز DAS (سرنام Direct-Attached Storage) که به طور مستقیم به سرور متصل می‌شوند در گذشته تنها راهی بودند که به منظور استقرار یک کلاستر هادوپ به کار گرفته می‌شدند. با توجه به اینکه پنهان باند شبکه‌ها افزایش پیدا کرده است، در حال حاضر گزینه ذخیره‌سازهای تحت شبکه NAS (سرنام Network-Attached Storage) گزینه ایده‌آلی در ارتباط با بزرگ داده‌ها به شمار می‌روند.»

اگر از DAS استفاده کنید، آغاز کار برای شما ساده‌تر خواهد بود و مدلی که طراحی کرده‌اید به خوبی با مفاهیم نرم‌افزارمحور کار خواهد کرد. همچنین به این نکته توجه داشته باشید که DAS به خوبی با داده‌های جریانی کار می‌کند.

اما NAS نیز می‌تواند به خوبی چند پروتکل را مدیریت کند، در زمینه گسترش‌پذیری عملکرد خوبی دارد و همچنین به نیازهای امنیتی و انطباقی به خوبی پاسخ می‌دهد.

منبع:

نشانی منبع:

<https://www.shabakeh-mag.com/cover-story/10793/%D8%B4%D8%B4-%D9%BE%D8%B1%D8%B3%D8%B4%DB%8C-%DA%A9%D9%87-%D9%87%D8%B1-%DA%A9%D8%B3%D8%A8-%E2%80%8C%D9%88-%DA%A9%D8%A7%D8%B1%DB%8C-%D8%A8%D8%A7%DB%8C%D8%AF-%D8%AF%D8%B1%D8%A8%D8%A7%D8%B1%D9%87-%D9%85%D8%B9%D9%85%D8%A7%D8%B1%DB%8C-%D8%A8%D8%B2%D8%B1%DA%AF-%D8%AF%D8%A7%D8%AF%D9%87%E2%80%8C%D9%87%D8%A7-%D9%85%D8%B7%D8%B1%D8%AD-%DA%A9%D9%86%D8%AF>