

ارزیابی داده کاوی در رایانش ابری

شهرام محمدی^۱، علیرضا جواد کلباسی^۲

^۱ مربی گروه مهندسی کامپیوتر، دانشگاه فنی و حرفه ای دختران اصفهان، اصفهان، ایران.

^۲ کارشناس ارشد فلسفه و کلام اسلامی.

نام نویسنده مسئول:

شهرام محمدی

تاریخ دریافت: ۱۴۰۲/۰۲/۱۰

تاریخ پذیرش: ۱۴۰۲/۰۴/۲۰

چکیده

این مقاله چگونگی استفاده از داده کاوی را در رایانش ابری شرح می‌دهد. داده کاوی برای استخراج اطلاعات مفید بالقوه از داده های خام به کار می رود. یک پارچه سازی تکنیک‌های داده کاوی در فعالیت های روز به روز عادی، موقعیت رایجی پیدا کرده است. مردم هرروز با تبلیغات هدفدار رو به رو می شوند و تکنیک های داده کاوی با کاهش هزینه، به تجارت گران کمک می‌کنند تا کارا تر گردند. تکنیک ها و اپلیکیشن داده کاوی در پارادایم رایانش ابری بسیار مورد نیاز هستند. پیاده سازی تکنیک‌های داده کاوی در رایانش ابری به کاربران اجازه می‌دهد اطلاعات بامعنی را از انبار داده‌ی مجازی یک پارچه سازی شده بازیابی کنند که این کار هزینه‌های مربوط به زیرساخت و ذخیره سازی را کاهش می‌دهد.

واژگان کلیدی: رایانش ابری، داده کاوی.

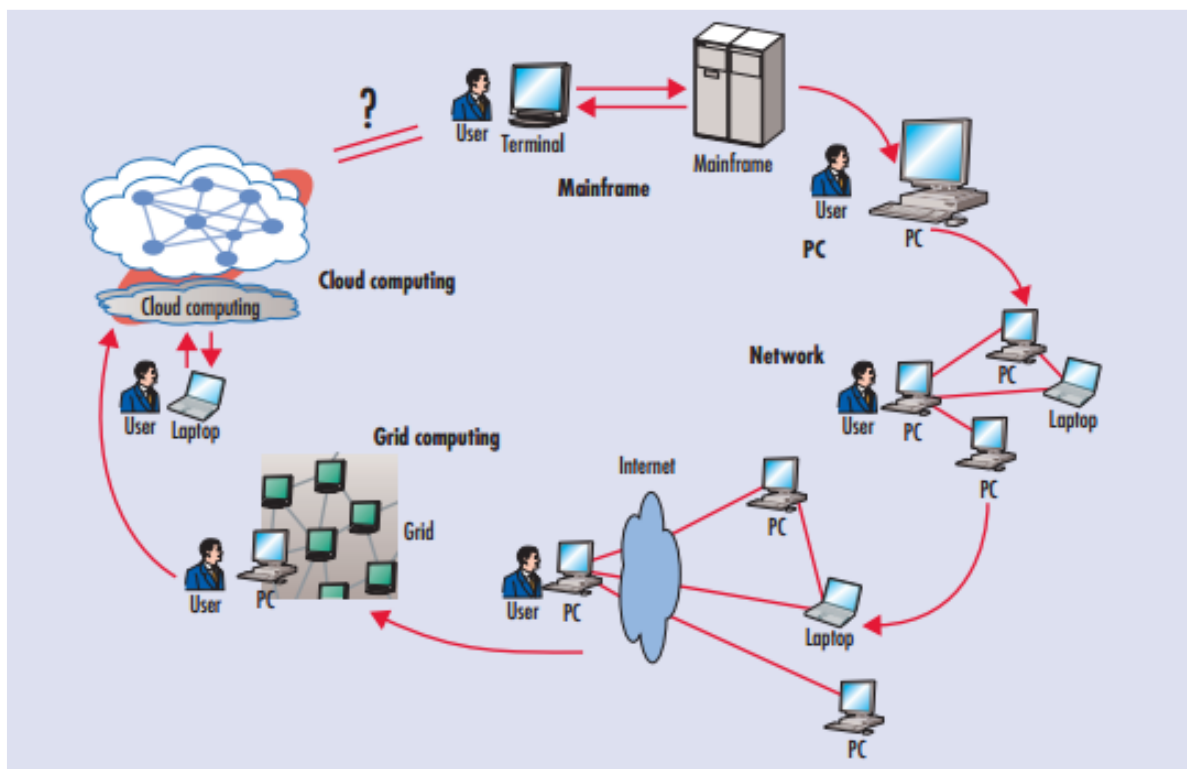
مقدمه

اینترنت دارد یک ابزار بسیار حیاتی تر در زندگی روزمره‌ی ما می‌شود؛ هم زندگی کاری و هم زندگی شخصی؛ زیرا تعداد کاربران آن به شدت در حال افزایش است. تعجب برانگیز نیست که انجام کارهای تجاری در اینترنت در حال افزایش است. شاید یکی از تکاملی‌ترین مفاهیم سال‌های اخیر رایانش ابری باشد. ابر، همان‌طور که اغلب به آن اشاره می‌شود، شامل استفاده از منابع محاسباتی - سخت‌افزار و نرم‌افزار - است که به عنوان یک سرویس بر روی اینترنت تحویل داده می‌شوند (در بیشتر نمودارهای آی تی به صورت ابر نشان داده شده است). شرکت‌های زیادی می‌خواهند که به جای ساختن زیرساخت آی تی خود برای میزبانی پایگاه‌های داده یا نرم‌افزار، یک شخص ثالث را برای میزبانی آن‌ها روی سرویس دهنده‌های بزرگ انتخاب کنند، بنابراین شرکت به داده‌ها و نرم‌افزار خود بروی اینترنت دسترسی خواهد داشت. استفاده از رایانش ابری در حال کسب محبوبیت است و علت این محبوبیت، پویایی، در دسترس بودن و هزینه‌ی پایین است. از طرف دیگر، رایانش ابری تهدیدات بیشتری برای امنیت داده‌ها و اطلاعات شرکت به وجود می‌آورد. تاحد تقریباً برابری، در سال‌های اخیر، تکنیک‌های رایانش ابری تکامل یافته و بیشتر مورد استفاده قرار گرفته‌اند. کشف دانش در پایگاه‌های داده، دارد در زمینه‌های متنوعی حیاتی می‌گردد: تجارت، پزشکی، علم و مهندسی، داده‌های مکانی و غیره. روند نوظهور رایانش ابری، مزیت منحصر به فرد دسترسی بی سابقه به داده‌های ارزشمند را برای کاربران فراهم می‌کند که می‌تواند به آن‌ها در رسیدن به اهداف تجاریشان کمک نماید.

بعضی از جنبه‌های مربوط به رایانش ابری

رایانش ابری، نرم‌افزار و سخت‌افزار تحویل داده شده در قالب سرویس‌ها (خدمات) را از طریق اینترنت نشان می‌دهد. رایانش ابری مفهوم جدیدی است که استفاده از محاسبه را به عنوان یک ابزار سودمند تعریف می‌کند که به تازگی توجه زیادی را به خود جلب کرده است.

- در شکل ۱ در زیر، تغییر بزرگ‌در نیمه‌ی دوم قرن اخیر در قالب ۶ مرحله‌ی (فاز) منحصربه‌فرد نشان داده شده است [۱]:
- فاز ۱: مردم از پایانه‌ها برای ارتباط با mainframe های قدرتمندی استفاده می‌کنند که به وسیله‌ی کاربران زیادی به-اشتراک گذاشته شده است.
 - فاز ۲: کامپیوترهای شخصی مستقل به اندازه‌ی کافی قدرتمند شدند که از پس کار روزانه‌ی کاربران بر بیایند.
 - فاز ۳: شبکه‌های کامپیوتری اتصال چندین کامپیوتر را به یکدیگر میسر کردند.
 - فاز ۴: شبکه‌های محلی توانستند به دیگر شبکه‌های محلی متصل شوند تا یک شبکه‌ی سراسری‌تر را بنا نهند.
 - فاز ۵: شبکه یا تور الکترونیکی باعث تسهیل توان محاسباتی و منابع ذخیره‌سازی به اشتراک‌گذاری شده گردید.
 - فاز ۶: رایانش ابری بهره‌برداری از همه‌ی منابع در دسترس در اینترنت را از یک راه ساده و مقیاس‌پذیر ممکن می‌سازد.



شکل ۱: تغییر بزرگ محاسباتی در نیمه‌ی دوم قرن اخیر [۱]

همان‌طور که توسط موسسه‌ی ملی استاندارد و فناوری تعریف شده، رایانش ابری مدلی است برای ممکن ساختن دسترسی‌های اندالزوم، در همه‌جا حاضر و راحت شبکه به یک مخزن به‌اشتراک‌گذاشته‌شده از منابع محاسباتی قابل پیکربندی (مانند شبکه‌ها، سرویس‌دهنده‌ها، محل‌های ذخیره‌سازی، کاربردها و سرویس‌ها) که می‌تواند به سرعت تهیه شده و با کمترین سعی مدیریتی یا فعل و انفعال فراهم‌کننده‌ی سرویس منتشر شود.

این مدل ابری از پنج خصوصیت حیاتی، سه مدل سرویس و چهار مدل توسعه ساخته شده است [۲]. خصوصیات حیاتی رایانش ابری بدین‌گونه است: دسترسی وسیع، خودپایور و اندالزوم به شبکه، pooling منابع، قابلیت ارتجاع سریع و سرویس کنترل شده.

مدل‌های سرویسی که رایانش ابری را می‌سازند عبارتند از نرم‌افزار به عنوان سرویس (SaaS)، بستر نرم‌افزاری platform به عنوان سرویس (PaaS) و زیرساختار به عنوان سرویس (IaaS).

مدل‌های توسعه‌ی رایانش ابری این‌ها هستند: ابر خصوصی، ابر عمومی و ابر مخلوط یا دورگه. جدول ۱ جزئیات را در شرکت‌های رایانش ابری فوقانی و ویژگی‌های کلیدی محصولاتشان را نشان می‌دهد:

جدول ۱: شرکت‌های رایانش ابری فوقانی و ویژگی‌های کلیدی [۳]

نام ابر	خصوصیت کلیدی
Sun Microsystems Sun Cloud	Application های موجود بیشتر از هر سیستم عامل باز open OS دیگری.
IBM Dynamic Infrastructure	مدیریت توان یکپارچه برای اینکه به شما کمک کند برنامه‌ریزی کنید، پیش‌بینی نمایید، بر مصرف توان سرویس دهنده‌های BladeCenter خود نظارت نموده و فعالانه آن‌ها را مدیریت نمایید.
Amazon EC2	طراحی شده تا محاسبات در مقیاس وب را برای توسعه‌دهندگان آسانتر نمایند.
Google App Engine	اگر از سهم اختصاص داده شده به شما بیشتر استفاده نکنید، محدودیتی برای دوره‌ی امتحان آزاد نخواهید داشت.
Microsoft Azure	به‌تازگی یک برنامه‌ی تخفیف "توسعه‌ی شتاب‌دهنده" پیشنهاد داده است. ۱۵ تا ۳۰ درصد تخفیف

در مورد هزینه‌های مصرف برای ۶ ماه اول	
زیرساختار کاملاً عندالزوم را استفاده کنید یا آن را با اجزاء اختصاص داده شده ترکیب نمایید تا به نیازهای تخصصی دست یابید.	AT&T Synaptic Hosting
تعادل بار و پشتیبانی رایگان ۲۴ ساعته در تمام طول هفته.	GoGrid Cloud Computing
ایستگاه‌های ابری را به ترتیب برای اتوماسیون، سرویس مشتری و بستر نرم‌افزاری پیشنهاد می‌دهد. شفافیت از طریق اطلاعات بلادرنگ در مورد کارایی سیستم و امنیت در trust.salesforce.com	Salesforce

رایانش ابری همه‌ی منابع ممکن بر روی اینترنت را نشان می‌دهد و توان محاسباتی نامتناهی ارائه می‌نماید. از آن جایی که رایانش ابری دارد یک روند فناوری مهم‌تر می‌شود، می‌تواند شکلی دوباره به بخش IT و بازار IT بدهد.

بعضی از جنبه‌های مربوط به داده کاوی

داده کاوی پیدا کردن الگوها یا گرایش‌های مفید از میان مقادیر زیاد داده است.

داده کاوی به عنوان یک "نوع تحلیل پایگاه داده که سعی می‌کند الگوها یا روابط مفید را در گروهی از داده‌ها کشف کند" تعریف شده است. این تحلیل روش‌های پیشرفته‌ی آماری مانند تحلیل خوشه را به کار می‌برد و گاهی اوقات از تکنیک‌های هوش مصنوعی یا شبکه عصبی استفاده می‌کند. یک هدف عمده‌ی داده کاوی یافتن روابط ناشناخته بین داده‌هاست، مخصوصاً زمانی که داده‌ها از پایگاه‌های داده‌ی مختلفی می‌رسند [۴]. مهم‌ترین تکنیک‌های داده کاوی و توضیحاتشان در جدول ۲ در زیر نشان داده شده است:

جدول ۲: تکنیک‌های داده کاوی [۵]

نام ابر	ویژگی کلیدی
خوشه‌بندی	مفید برای اکتشاف داده و یافتن گروه‌بندی‌های طبیعی. اعضای یک خوشه بیشتر از این که شبیه اعضای یک خوشه‌ی دیگر باشند، به هم شبیه هستند. مثال‌های رایج شامل پیدا کردن بخش‌های مشتری و کشف علوم زیستی است.
طبقه‌بندی	پرکاربردترین تکنیک برای پیش‌بینی یک خروجی خاص مانند پاسخ/عدم پاسخ، مشتری با ارزش بالا/متوسط/پایین، احتمال خریدن/نخریدن.
مشارکت	قوانین مربوط به itemهایی را پیدا می‌کند که با بسامد (فرکانس) بیشتری رخ می‌دهند، برای تحلیل سبد خرید، فروش متقاطع یا تطبیقی، تحلیل ریشه‌یابی علت به کار می‌رود. مناسب برای ساختمان محصول، جایابی در فروشگاه و تحلیل کاستی.
رگرسیون	تکنیکی برای پیش‌بینی یک خروجی عددی پیوسته مانند ارزش مشتری، ارزش موسسه و میزان سوددهی فرایند.
اهمیت ویژگی	ویژگی‌ها را بر اساس قوت رابطه با ویژگی هدف، رتبه‌بندی می‌کند. مواردی که استفاده می‌کند شامل پیدا کردن عواملی که بیشترین رابطه با مشتریانی دارند که به یک پیشنهاد یا عرضه پاسخ می‌دهند، عواملی که با افراد موفق ارتباط دارند.
تشخیص ناهنجاری anomaly detection	موارد غیرمعمول یا مشکوک را مبتنی بر انحراف از هنجار تشخیص می‌دهد. مثال‌های رایج آن کلاهبرداری مربوط به بهداشت و درمان، کلاهبرداری مربوط به گزارش هزینه و پیروی از مالیات است.
استخراج ویژگی	ویژگی‌های جدیدی به عنوان ترکیب خطی ویژگی‌های موجود تولید می‌کند. قابل کاربرد در داده‌های متنی، تحلیل معنایی پنهان، تجزیه‌ی داده و projection و تشخیص الگو.

با توجه به تکنیک‌های داده کاوی گوناگون و نیاز شدید به کشف الگوها و گرایش‌ها در داده‌ها که به دانشی منجر خواهد شد که در غیر این صورت فراهم نخواهد شد، تعجبی نیست که داده کاوی در متنوع‌ترین زمینه‌ی فعالیت استفاده شده است.

داده کاوی، استخراج اطلاعات قابل پیش‌بینی پنهان از پایگاه‌های داده‌ی بزرگ، یک فناوری قدرتمند جدید با پتانسیل زیاد برای کمک به شرکت‌ها است تا بر روی مهم‌ترین اطلاعات در انبارهای داده‌شان تمرکز کنند. ابزارهای داده کاوی گرایش‌ها و رفتارهای آینده را پیش‌بینی می‌کنند و شرکت‌های تجاری را قادر می‌سازند تصمیمات کنش‌گرا و مبتنی بر دانش اتخاذ کنند. تحلیل‌های خودکار و آینده‌نگری که داده کاوی ارائه می‌کند فراتر از تحلیل وقایع گذشته قرار می‌گیرد که توسط ابزارهای پس‌نگرانه (مبتنی بر گذشته) معمول سیستم‌های پشتیبانی تصمیم فراهم می‌گردد [۶]. شرکت‌های تجاری می‌توانند در مورد این‌که یک محصول چقدر خوب خواهد فروخت پیش‌بینی کنند یا کمپین‌های تبلیغاتی جدید را با استفاده از این روابط جدیدی که توسط الگوریتم‌های داده کاوی منعکس شده‌اند، توسعه دهند. بخش پزشکی نیز مانند داده‌های جغرافیایی که با استفاده از داده کاوی بهتر تحلیل می‌شود، از تکنیک‌های داده کاوی سود می‌برد. دولت‌ها می‌توانند فعالیت‌های غیرقانونی یا ممنوع افراد، شرکت‌ها یا دولت‌های دیگر را با پیاده‌سازی تکنیک‌های داده کاوی تشخیص دهند. به‌طورمختصر، استفاده از داده کاوی در اکثر زمینه‌های فعالیت توسعه پیدا کرده است.

داده کاوی در رایانش ابری

تکنیک‌ها و کاربردهای داده کاوی در الگوواره‌ی رایانش ابری بسیار مورد نیاز هستند. از آن‌جا که رایانش ابری در حال رخنه کردن بیشتر و بیشتر در همه‌ی محدوده‌های تجارت و محاسبات علمی است، زمینه‌ی خوبی خواهد شد که با داده کاوی بر آن تمرکز شده است. رایانش ابری گرایش جدیدی را در سرویس‌های اینترنتی نشان می‌دهد که برای اداره کردن وظایف بر روی ابرهایی از سرویس‌دهندگان تکیه می‌کند.

داده کاوی در رایانش ابری، فرایند استخراج اطلاعات ساخت‌یافته از منابع داده‌ی غیرساخت‌یافته یا نیمه‌ساخت‌یافته‌ی وب است. داده کاوی در رایانش ابری به سازمان‌ها اجازه می‌دهد مدیریت نرم‌افزار و محل ذخیره‌سازی داده را با اطمینان نسبت به سرویس‌های کارا، قابل اعتماد و امن برای کاربران‌شان متمرکز کنند [۶]. چون رایانش ابری به نرم‌افزار و سخت‌افزار تحویل داده شده به عنوان سرویس از طریق اینترنت اشاره می‌کند، به‌همین ترتیب در رایانش ابری نرم‌افزار داده کاوی نیز فراهم شده است.

اثرات اصلی ابزارهای داده کاوی که به وسیله‌ی Cloud تحویل می‌شود این‌ها هستند:

- مشتری تنها برای ابزارهای داده کاوی که نیاز دارد هزینه پرداخت می‌کند، که این امر مخارجش را کاهش می‌دهد زیرا او مجبور نیست برای برنامه‌های پیچیده‌ی داده کاوی که به‌طور کامل استفاده نمی‌کند هزینه پرداخت نماید.
- مشتری مجبور نیست یک زیرساختار سخت‌افزاری را نگه دارد زیرا او می‌تواند داده کاوی را از طریق یک مرورگر اعمال کند، این بدان معنی است که او تنها مجبور است هزینه‌هایی را پرداخت کند که از رایانش ابری ایجاد شده‌اند. استفاده از داده کاوی از طریق رایانش ابری، موانعی را کاهش می‌دهد که شرکت‌های کوچک را از سود بردن از ادوات داده کاوی دور نگه می‌دارد.

رایانش ابری گرایش جدیدی را در سرویس‌های اینترنتی نشان می‌دهد که برای handle کردن وظایف بر روی ابرهایی از سرویس‌دهندگان تکیه می‌کند. داده کاوی در رایانش ابری، فرایند استخراج اطلاعات ساخت‌یافته از منابع داده‌ی غیرساخت‌یافته یا نیمه‌ساخت‌یافته‌ی وب است.

داده کاوی در رایانش ابری به سازمان‌ها اجازه می‌دهد مدیریت نرم‌افزار و محل ذخیره‌سازی داده را با اطمینان نسبت به سرویس‌های کارا، قابل اعتماد و امن برای کاربران‌شان متمرکز کنند [۶]. پیاده‌سازی تکنیک‌های داده کاوی از طریق رایانش ابری به کاربران اجازه خواهد داد تا اطلاعات بامعنی را از انبارهای داده‌ای که به صورت مجازی یکپارچه شده‌اند، بازیابی نمایند که این امر هزینه‌های مربوط به زیرساختار و ذخیره‌سازی را کاهش می‌دهد.

نتیجه‌گیری

فناوری داده کاوی فراهم شده از طریق رایانش ابری یک خصوصیت بسیار لازم برای شرکت‌های تجاری امروز است تا تصمیمات مبتنی بر دانش و کنش‌گرا بگیرند زیرا به آن‌ها کمک می‌کند که گرایش‌ها و رفتارهای آتی را پیش‌بینی نمایند. این مقاله یک دید کلی نسبت به لزوم و استفاده از داده کاوی در رایانش ابری فراهم می‌کند. از آن‌جا که نیاز به ابزارهای داده کاوی هر روز در حال افزایش است، توانایی یکپارچه‌سازی آن‌ها در رایانش ابری مهم‌تر و مهم‌تر می‌شود.

منابع و مراجع

- [1] Jeffrey Voas and Jia Zhang, "Cloud Computing: New Wine or Just a New Bottle?", IEEE Internet Computing Magazine, 2009, Link.
- [2] Peter Mell, and Timothy Grance, "The NIST Definition of Cloud Computing", The National Institute of Standards and Technology, USA, 2011, Link.
- [3] IT Strategists, "Top Cloud Computing Companies and Key Features", Link: <http://www.itstrategists.com/Top-Cloud-Computing-Companies.aspx>.
- [4] Merriam-Webster Dictionary, "Definition of data mining", Link: <http://www.merriamwebster.com/dictionary/data%20mining>.
- [5] ORACLE, "Oracle Data Mining Mining Techniques and Algorithms", Link: <http://www.oracle.com/technetwork/data-base/options/advancedanalytics/odm/odm-techniquesalgorithms-097163.html>.
- [6] Bhagyashree Ambulkar and Vaishali Borkar, "Data Mining in Cloud Computing", MPGI National Multi Conference 2012 (MPGINMC-2012), 7-8 April 2012, Link.