

## بررسی و ارزیابی فناوری‌های داده‌های عظیم به‌عنوان سرویس

محمد رضا حسین‌زاده<sup>۱</sup>، آرش غفوری<sup>۲\*</sup>، حسن نادری<sup>۳</sup>

۱- دانشجوی کارشناسی ارشد دانشگاه امام حسین<sup>(ع)</sup>، ۲- دانشجوی دکتری و ۳- استادیار دانشگاه علم و صنعت ایران

(دریافت: ۹۷/۱۲/۰۳؛ پذیرش: ۹۸/۰۲/۲۹)

### چکیده

با پدیدار شدن فناوری داده‌های عظیم به‌عنوان سرویس، سازمان‌های کوچک می‌توانند خدمات موردنیاز خود، در مورد داده‌های عظیم را از طریق سرویس‌های شرکت‌های بزرگ تأمین کنند. این کار باعث کاهش منابع موردنیاز برای پروژه‌های آزمایشی و کاهش هزینه‌ها می‌شود. با توجه به لایه‌های چارچوب داده‌های عظیم به‌عنوان سرویس، شرکت‌های پیشرو هرکدام سرویس‌های خود را روی برخی از لایه‌های این چارچوب ارائه می‌دهند و کاربران می‌توانند با توجه به نیازهای خود، سرویس‌دهنده موردنظر را انتخاب کنند. در این مقاله هشت فناوری مهمی که به‌صورت داده‌های عظیم به‌عنوان سرویس توسط شرکت‌های مختلف ارائه می‌شوند مورد بررسی و ارزیابی قرار گرفته و با یکدیگر مقایسه شدند. نکته مهمی که باید به آن توجه شود تنوع سرویس‌هایی است که توسط شرکت‌های مختلف ارائه می‌شوند به‌نحوی که نمی‌توان در مورد برتری یک سرویس نسبت به سرویس‌های دیگر قضاوت کرد بلکه هدف ارائه‌دهندگان سرویس، ارائه خدمات متنوع‌تر و مبتنی بر تقاضا است که کاربران با توجه به نیازهای خود سرویس‌های موردنظر را انتخاب کنند.

### کلیدواژه‌ها:

داده عظیم، داده‌های عظیم به‌عنوان سرویس، BDaaS، ابر

#### ۱- مقدمه

بر اساس پیش‌بینی‌ها حجم داده‌ها با نرخ تقریبی ۲ برابر در هر ۲ سال افزایش می‌یابد. بر اساس آمارهای موجود در سال ۲۰۱۵، حجم داده‌ها به ۹ اگزابایت رسیده است که احتمال دارد تا سال ۲۰۲۰ حجم داده‌ها به ۴۰ اگزابایت برسد<sup>۲</sup>.

داده‌های عظیم در پایگاه داده‌های NoSQL ذخیره می‌شوند و به‌طور نوعی، غیر رابطه‌ای، توزیع‌شده، متن‌باز، مقیاس‌پذیر و حجیم هستند که بیان‌کننده سه ویژگی حجم<sup>۳</sup>، سرعت<sup>۴</sup> و تنوع<sup>۵</sup> داده‌ها است [۲]. بنابراین، مهم‌ترین ویژگی داده‌های عظیم را می‌توان 3V's معرفی کرد که اعم از حجم، سرعت و تنوع هستند [۳]. در اینجا حجم داده مربوط به اندازه نسبی داده از نظر قابلیت پردازش، سرعت مربوط به بسامد تولید، به اشتراک‌گذاری و ذخیره‌سازی داده بوده و منظور از تنوع داده، گسترش و تفاوت

داده‌های موجود در دنیای فناوری اطلاعات روزبه‌روز در حال افزایش بوده و تعداد زیادی از منابع، داده‌های کلانی را تولید و یا مصرف می‌کنند. به‌عنوان مثال استفاده از گوشی‌های هوشمند و شبکه‌های اجتماعی، رشد قابل توجهی از داده‌ها را در چندین سال اخیر پدیدار کرده است. داده عظیم<sup>۱</sup> مجموعه‌ای از داده‌های بسیار بزرگ و پیچیده است که پردازش آن‌ها جهت مدیریت، ذخیره‌سازی، به اشتراک‌گذاری، تجزیه و تحلیل، انتقال و جستجو در یک‌زمان قابل تحمل توسط نرم‌افزارهایی که به‌طور معمول استفاده می‌شوند، دشوار است. اندازه این داده‌ها پیوسته در حال رشد است [۱]. در حال حاضر روزانه حجم بسیار عظیمی از داده در حال تولید است و بیش از ۹۰٪ داده‌های تولیدشده و ذخیره‌شده در دنیای دیجیتال به‌نوعی غیر ساخت‌یافته هستند.

<sup>۲</sup> <http://www.statisticbrain.com/average-cost-of-hard-drive-storage/>.

<sup>۳</sup> volume

<sup>۴</sup> velocity

<sup>۵</sup> variety

\* رایانامه نویسنده پاسخگو: krghafouri@ihu.ac.ir

<sup>۱</sup> Big Data

به مشتریان ارائه می شوند مورد بررسی قرار گرفته و در پایان با یکدیگر مقایسه می شوند.

## ۲- معرفی داده های عظیم به عنوان سرویس

داده های عظیم به عنوان سرویس حاصل تلفیق دو گرایش عمده فناوری اطلاعات یعنی داده های عظیم و محاسبات ابری است. داده های عظیم مربوط به رشد سریع در تنوع، حجم و سرعت اطلاعات تولیدی در شرکتها است و محاسبات ابری به مهاجرت نرم افزارها و زیرساخت های فناوری اطلاعات از مراکز داده ای شرکت های بزرگ به سمت سرویس هایی که قابلیت تحویل از اینترنت را دارند مربوط می شود.

معماری محاسبات ابری مبتنی بر معماری سرویس گرا است لذا جهت درک بهتر معماری های پایه در محاسبات ابری، توجه به رابطه میان معماری سرویس گرا و مفهوم محاسبات ابری حائز اهمیت است؛ با این حال تمامی سرویس های معماری سرویس گرا ممکن است جزئی از سرویس های محیط ابری نباشند.

داده های عظیم مبتنی بر محاسبات ابری در معماری سرویس گرا یک حوزه کاری جدیدی در مهندسی کامپیوتر است که منجر به ایجاد یک شاخه ای از فناوری به نام داده عظیم به عنوان سرویس می شود [۹-۱۰].

BDaaS می تواند به وظایفی نظیر پردازش، ذخیره سازی و بصری سازی پاسخگو باشد که هر سرویس به صورت مستقل است [۱۱]. هدف BDaaS ارائه بینش مقرون به صرفه و با ارزش برای سازمانها و کسب و کارهای کوچک به منظور افزایش رقابت پذیری و نوآوری آنها و در نتیجه افزایش درآمد آنهاست.

همه شرکتها و سازمانها قادر نیستند و یا نیازی ندارند که یک انبار داده و مرکز پردازشی را نصب، راه اندازی و مدیریت کرده و بتوانند تحلیل داده ها را امکان پذیر کنند. BDaaS این امکان را می دهد که شرکتها و سازمانهای متعددی با هزینه پایین این اعمال را به عنوان سرویس خریداری نمایند؛ یعنی ذخیره سازی، مدیریت و تحلیل های خود را روی داده های عظیم انجام دهند [۱۲].

ذخیره سازی و مدیریت حجم کلانی از داده ها نیازمند سرمایه گذاری در زمان و منابع است در حالی که در صورت خرید سرویس از شرکت های دیگر، هزینه نگهداری و حفاظت از داده ها برعهده ارائه دهندگان سرویس است که با ذخیره داده ها در

انواع داده ها است که از منظر منبع ایجادکننده داده ها نیز تفاوت وجود دارد [۴]. سازمانها هزینه های قابل توجهی برای پیاده سازی سامانه های داده عظیم، نصب و راه اندازی زیرساخت آن و نیروهای انسانی مورد نیاز می پردازند. اخیراً بسیاری از سازمانها منابع قابل توجهی برای فناوری ارزیابی داده های عظیم صرف می کنند. چالش عمده پیاده سازی داده های عظیم، تعیین ارزش کسب و تجارت آن است. مطابق با گزارش های ارائه شده بالغ بر ۵۶ درصد چالش های داده های عظیم تعیین ارزش کسب و کار آن است. در یک نتیجه گیری مشابه، ۲۹ درصد پاسخگویان عمده مشکل داده های عظیم را نصب و مدیریت زیرساخت برای مدیریت داده های عظیم تلقی کرده اند [۵]. این چالشها باعث شده تا شاخه جدیدی از داده های عظیم تحت عنوان داده های عظیم به عنوان سرویس<sup>۱</sup> ایجاد شود. در این حالت شرکتها و سازمانها، سرویس هایی ارائه می دهند که داده های عظیم در آن به عنوان سرویس هستند. حال می توان بدون دغدغه مدیریت، پردازش و ذخیره سازی داده های عظیم، سرویس آنها را از شرکتها خریداری کرده و در ساختار سرویس گرا استفاده نمود. تأثیر داده های عظیم روی معماری سرویس گرا که نتیجه تلفیق آنها داده عظیم به عنوان سرویس است، مناسب کاربران نهایی است تا با کمترین هزینه به این سرویسها دسترسی داشته باشند.

در [۶] بازار BDaaS در حدود ۲/۵۵ میلیارد دلار تخمین زده شده که این مقدار در سال ۲۰۲۱ به ۳۰ میلیارد دلار خواهد رسید. از طرفی تخمین زده شده است که ۵۰ درصد داده ها تا آخر سال ۲۰۱۶ در سطح ابر ذخیره سازی خواهند شد؛ بنابراین داده عظیم در ترکیب با ابر و معماری سرویس گرا آینده فناوری اطلاعات را تعریف می کند [۷]. به گزارش گارتنر [۸]، داده عظیم در فاز افزایشی بازار قرار دارد که نسبت هزینه محصولات نرم افزاری برای خدمات فناوری در این فاز بسیار قابل توجه است. از این رو، یک بازار رقابتی بزرگ بین شرکت های پیشرو در دنیا ایجاد شده است و هر شرکت سعی در ارائه محصولات و فناوری های برتر BDaaS است تا مشتریان بیشتری را جذب کند.

در این مقاله ابتدا داده های عظیم به عنوان سرویس معرفی شده و انواع داده های عظیم به عنوان سرویس تشریح می شوند سپس در بخش بعد چارچوب داده های عظیم به عنوان سرویس بیان شده و بر اساس این چارچوب مزایا و معایب داده های عظیم به عنوان سرویس تشریح می شود. در بخش چهارم مقاله، هشت فناوری برتر BDaaS که توسط شرکت های مختلف

<sup>1</sup> Big Data as a Service (BDaaS)

## ۲-۱- انواع داده‌های عظیم به‌عنوان سرویس

با توجه به این‌که داده‌های عظیم به‌عنوان سرویس بر روی زیرساخت ابر بنا شده است، برای BDaaS نیز سه مدل بنیادی سرویس‌های محاسبات ابری مورد استفاده قرار می‌گیرد. بر این اساس داده‌های عظیم به‌عنوان سرویس به سه نوع: زیرساخت به‌عنوان سرویس<sup>۱</sup>، بستر به‌عنوان سرویس<sup>۲</sup> و نرم‌افزار به‌عنوان سرویس<sup>۳</sup> تقسیم می‌شود [۱۴].

IaaS دسترسی به منابع اصلی همانند ماشین‌های فیزیکی، ماشین‌های مجازی و فضاهای ذخیره‌سازی را امکان‌پذیر می‌کند. دو نمونه مهم از انواع IaaS که به‌عنوان BDaaS ارائه می‌شوند ذخیره‌سازی به‌عنوان سرویس و قدرت محاسباتی به‌عنوان سرویس است.

PaaS محیط در حال اجرایی برای برنامه‌های کاربردی و همچنین ابزارهایی برای توسعه و گسترش برنامه‌های کاربردی فراهم می‌کند.

ایجاد یک برنامه با استفاده از خدمات PaaS نسبت به رویکرد سنتی در برنامه‌نویسی و گسترش نرم‌افزار با کار کمتر و در زمان کوتاه‌تری انجام می‌شود، دیگر نیازی به نصب و پیکربندی بسترها و ابزارها و برنامه‌های گسترش نرم‌افزار نیست. دو نمونه مهم از انواع PaaS که به‌عنوان BDaaS ارائه می‌شوند پایگاه داده به‌عنوان سرویس و داده به‌عنوان سرویس است.

SaaS نوعی مدل ارائه نرم‌افزار است به‌طوری‌که نرم‌افزار و اطلاعات مربوط به آن به‌طور کامل و یکپارچه روی ابر مستقر شده‌اند. دسترسی کاربران به نرم‌افزار با کمک یک مرورگر وب امکان‌پذیر خواهد بود.

از آنجایی‌که نرم‌افزار روی سرور متعلق به ارائه‌دهنده، نگهداری می‌شود، مصرف‌کننده نیازی به تهیه سخت‌افزار برای استقرار نرم‌افزار ندارد. همچنین هزینه‌های نصب، تعمیرات، نگهداری و به‌روز رسانی نیز در این مدل برای مصرف‌کننده وجود ندارد. مهم‌ترین سرویسی که به‌عنوان SaaS در BDaaS ارائه می‌شود، تحلیل داده‌های عظیم است.

سرورهای خود، مسئولیت نگهداری از آن‌ها را بر عهده می‌گیرند. سرویس تحلیل‌های IBM برای توییت، نمونه‌ای از این سرویس‌ها است که روزانه تحلیل ۵۰۰ میلیون توییت را امکان‌پذیر می‌سازد [۱۳].

BDaaS قابلیت‌های زیادی برای تغییر در بخش‌های مختلف زندگی و فرایندهای آن مانند آموزش، امور مالی، بیمه و مراقبت‌های بهداشتی در اقتصاد نوظهور امروزی دارد که در گذشته این امر میسر نبود. جدول (۱) مقایسه داده‌های عظیم سنتی و امروزی را نشان می‌دهد.

جدول (۱): مقایسه BDaaS با داده‌های سنتی.

محیط‌های سنتی	داده‌های عظیم سنتی	BDaaS
کمبود منابع مانند قدرت محاسباتی و ظرفیت ذخیره‌سازی	مقیاس‌پذیری در پردازش و ذخیره‌سازی از طریق معماری توزیع‌شده	مقیاس‌پذیری مبتنی بر ترکیب محاسبات ابری و معماری توزیع‌شده
رسانه ذخیره‌سازی مجتمع داده شبیه SAN و NAS و دیسک‌های سنتی	ذخیره‌سازی داده روی HDFS یا بسترهای توزیع‌شده	ذخیره‌سازی مجازی داده روی بسترهای توزیع‌شده
داده‌های ساخته‌یافته	داده‌های ساخته‌یافته و بدون ساختار	داده‌های ساخته‌یافته و بدون ساختار
دسترسی محدود	دسترسی محدود	دسترسی کامل
گزارش‌گیری از طریق ابزارهایی مانند OLAP	تجزیه و تحلیل پیشرفته	تجزیه و تحلیل پیشرفته با قدرت محاسباتی مبتنی بر تقاضا

<sup>1</sup> Infrastructure as a service

<sup>2</sup> Platform as a service

<sup>3</sup> Software as a service

نوعاً از هزاران سرور متصل به یکدیگر توسط یک شبکه پرسرعت شکل گرفته‌اند. مراکز داده همچنین شامل مسیریاب‌ها، فایروال‌ها و سامانه‌های پشتیبان برای مقابله با گم شدن داده‌ها هستند. این مراکز داده، سرویس‌های ذخیره‌سازی و مدیریت پر قدرت داده‌های عظیم را میسر می‌سازند. قابل ذکر است هزینه ساخت هر مرکز داده می‌تواند در حدود یک و نیم میلیون دلار برای هر مترمربع از مساحت برآورد شود [۱].

چارچوب BDaaS نیازمند چابکی، قابلیت انعطاف، امنیت، عملی بودن و کارایی است که چابکی برای پیکربندی پویا برای رشد آسان کلاسترهای داده، قابلیت انعطاف برای ذخیره‌سازی انواع مدل‌های داده عظیم و امنیت برای فراهم نمودن تحمل پذیری خطا در همه مراحل BDaaS اعم از ذخیره‌سازی، مدیریت و پردازش داده‌های عظیم است [۱۵].

### ۳- مزایا و معایب داده‌های عظیم به‌عنوان سرویس

با توجه به لایه‌بندی چارچوب داده‌های عظیم به‌عنوان سرویس، می‌توان مزایا و معایب داده‌های عظیم به‌عنوان سرویس را بررسی کرد. لایه تجزیه و تحلیل داده برای پردازش داده و ایجاد گزارش‌های تجاری برای کاربران مفید است.

فناوری‌های لایه مدیریت داده و پردازش، نیازمند دانش فنی و برنامه‌نویسی برای پیاده‌سازی آن است. دولا به زیرین چارچوب BDaaS نیز نیازمند منابع محاسباتی از جمله ماشین‌ها و سخت‌افزارهای مجازی هستند.

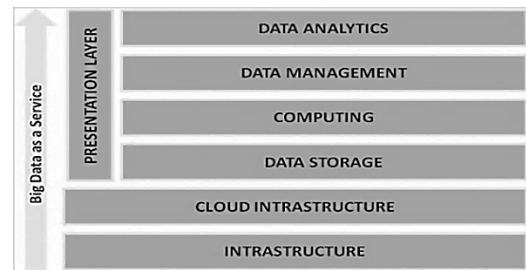
جدول (۲) مزایا و معایب لایه‌های چارچوب BDaaS را به تفکیک لایه‌ها نمایش می‌دهد.

بررسی جدول (۲) نشان می‌دهد که دو نکته مهمی که می‌تواند مزایا و معایب BDaaS را تحت تأثیر قرار دهد بحث امنیت و مسائل مربوط به تخصص است.

در مسائل مربوط به امنیت سامانه باید به‌گونه‌ای باشد که اجازه دست‌کاری منفی داده‌ها توسط اشخاص خارجی داده نشود و از طرف دیگر در مسائل مربوط به تخصص، عدم وجود نیروی کار ماهر یک چالش بزرگ برای توسعه‌دهندگان و استفاده‌کنندگان از BDaaS است که باید افراد به‌گونه‌ای تربیت شوند تا علاوه بر دانش‌های مربوط به توسعه و نگهداری ابر، با سایر دانش‌های مربوط به برنامه‌نویسی و توسعه ابزار کار با داده‌های حجیم آشنا باشند.

### ۲-۲- چارچوب داده‌های عظیم به‌عنوان سرویس

چارچوب BDaaS بسته به نوع فناوری و وظایف مؤلفه‌های آن از لایه‌هایی تشکیل شده است. این چارچوب در شکل (۱) نشان داده شده است.



شکل (۱): لایه‌های چارچوب داده‌های عظیم به‌عنوان سرویس.

لایه‌های چارچوب BDaaS به شرح ذیل است:

- لایه تجزیه و تحلیل داده<sup>۱</sup>: این لایه شامل برنامه‌های کاربردی سطح بالایی هستند که می‌توانند برای تجزیه و تحلیل داده‌های عظیم به‌کار گرفته شوند. کاربران به فناوری‌های تجزیه و تحلیل داده‌های عظیم از طریق یک واسط وب دسترسی دارند که پرس‌وجوها و گزارش‌ها را ایجاد می‌نمایند.
- لایه مدیریت داده<sup>۲</sup>: در این لایه برنامه‌های کاربردی سطح بالایی همانند سرویس پایگاه داده رابطه‌ای آمازون پیاده‌سازی شده‌اند تا سرویس پردازش و مدیریت توزیع شده داده‌های عظیم محقق شود.
- لایه محاسبات<sup>۳</sup>: این لایه مرکب از فناوری‌هایی است که سرویس پردازشی داده‌های عظیم را روی یک بستر وب فراهم می‌سازد. API ها و Map-Reduce نمونه‌ای از آنهاست.
- لایه ذخیره‌سازی داده<sup>۴</sup>: این لایه نوعاً به‌صورت توزیع شده بوده و مبتنی بر تقاضا می‌تواند مقیاس پذیر باشد. کاربر در این لایه می‌تواند مستقیماً با HDFS ارتباط برقرار کرده و داده‌ها را در آن ذخیره‌سازی نماید.
- لایه زیرساخت ابر<sup>۵</sup>: در این لایه بستر ابر، شبیه سرور ماشین مجازی ESX است و مستقیماً از طریق برنامه‌های کاربردی خارجی قابل دسترسی نیستند.
- لایه زیرساخت<sup>۶</sup>: این لایه از سخت‌افزارهایی شبیه مراکز داده واقعی و گره‌های فیزیکی تشکیل شده است که مراکز داده

<sup>6</sup> Infrastructure

<sup>1</sup> Data Analytics

<sup>2</sup> Data Management

<sup>3</sup> Computing

<sup>4</sup> Data Storage

<sup>5</sup> Cloud Infrastructure

جدول (۲): مزایا و معایب BDaaS

پیشنهاد	معایب	مزایا	لایه
مناسب کاربران تجاری و تحلیل‌گران تجاری	تجزیه و تحلیل محدود داده‌ها و عدم دسترسی مستقیم به آن‌ها	دسترسی آسان کاربران به سرویس‌ها بدون نیاز به زیرساخت	تجزیه و تحلیل داده
مناسب برای مدیران پایگاه داده و برنامه‌نویسان	نیاز به دانش برنامه‌نویسی برای فعالیت‌های خاص	دسترسی مستقیم به داده و قابلیت تجزیه و تحلیل یا دست‌کاری داده‌ها	مدیریت داده
مناسب برای برنامه‌نویسان		انعطاف‌پذیری بالا طوری که برنامه‌نویسان می‌توانند برنامه‌هایی برای دست‌کاری داده‌ها بنویسند	محاسبات و پردازش
		دسترسی به ردیف داده در رسانه‌های ذخیره‌سازی توزیع‌شده	ذخیره‌سازی داده
مناسب برای مهندسیین زیرساخت و برنامه‌نویسان	نیاز به دانش و کسب اطلاعات از زیرساخت دارد	قابلیت استفاده در همه زیرساخت‌های فناوری اطلاعات	زیرساخت ابر
مناسب برای مهندسیین زیرساخت		سخت‌افزار عمومی روی زیرساخت‌های پایه محاسباتی	زیرساخت

سرورهای اختصاصی مجازی که سامانه‌عامل‌های لینوکس یا ویندوز را بر روی Xen Hypervisor اجرا می‌کنند، میسر کرده است. در این سرویس نمونه‌های ماشین با توان‌های پردازشی مختلف موجود هستند و بر اساس محاسبات/ساعت اجاره می‌شوند [۱۶]. برنامه‌های مستقر بر روی این ماشین‌ها بسیار توسعه‌پذیر و با تحمل‌پذیری بالا خطا هستند. تعدادی از ابزارهایی که برای پشتیبانی سرویس‌های EC2 استفاده می‌شوند به شرح زیر است:

- سرویس صف ساده آمازون<sup>۱</sup>: یک صف پیام یا سامانه تراکنش برای برنامه‌های مبتنی بر اینترنت توزیع شده است. این سرویس تضمین می‌کند که پیام‌ها حتی در زمانی که مؤلفه‌ای موجود نیست، گم نشود و برای انتقال پیام میان مؤلفه‌های مختلف که هر کدام کار جداگانه‌ای را انجام می‌دهند، بسیار مناسب است.
- سرویس آگاه‌سازی ساده آمازون<sup>۲</sup>: وب‌سرویس است که می‌تواند پیام یک برنامه را منتشر کند و آن‌ها را به برنامه‌ها یا مشترکین دیگر منتقل کند. SNS روشی را برای راه‌اندازی فعالیت‌ها ارائه می‌نماید که برنامه‌ها را قادر می‌سازد تا در مورد اطلاعات جدید یا تغییر یافته از آن‌ها نظرسنجی شود یا به‌روزرسانی‌ها را انجام دهند.

#### ۴- فناوری‌های مختلف داده‌های عظیم به‌عنوان

##### سرویس

با توجه به محدودیت منابع و محدودیت زمانی، محققین نمی‌توانند و یا تمایل ندارند یک زیرساخت مقیاس‌پذیر هادوپ را پیاده‌سازی و اجرا نمایند. یعنی نصب و راه‌اندازی این زیرساخت مستلزم هزینه و زمان است. اما به‌وسیله فناوری‌های BDaaS محققین می‌توانند ذخیره‌سازی، محاسبات، تجزیه و تحلیل و پردازش داده‌های عظیم را به‌صورت جداگانه انجام دهند.

در این حالت به دلیل اینکه چارچوب BDaaS خدمات را در لایه‌های مختلفی دسته‌بندی کرده است، می‌توان برای هر لایه به‌صورت جداگانه کاربرد خاصی را طراحی کرد که کاربران با توجه به نیازهای خود از خدماتی که هر محصول ارائه می‌دهد استفاده کنند. این امر موجب صرفه‌جویی در هزینه‌های کاربران و سادگی کار برای توسعه‌دهندگان می‌شود.

در ادامه مهم‌ترین فناوری‌های طراحی شده برای BDaaS را بررسی کرده و با یکدیگر مقایسه می‌کنیم.

#### ۴-۱- ابر محاسباتی توسعه‌پذیر آمازون

ابر محاسباتی توسعه‌پذیر آمازون یا EC2، قسمتی از پلتفرم رایانش ابری شرکت آمازون است و جزء سرویس‌های مجازی خصوصی محسوب می‌شود. این سرویس، استفاده و مدیریت

<sup>1</sup> Simple Queue Service

<sup>2</sup> Simple Notification Service

برای مقیاس‌های بالا، کاربران می‌توانند توان عملیاتی بالایی خریداری کرده و DynamoDB به صورت خودکار، ترافیک را روی چندین سرور توزیع می‌کند. برای ایجاد و نمونه‌سازی یک جدول NoSQL در DynamoDB کاربران باید ظرفیت توان عملیاتی مشخصی را برای خواندن آن، مدنظر قرار بدهند.

DynamoDB یک سرویس رسانه ذخیره‌سازی مقیاس‌پذیر از آمازون است که سه لایه زیرساخت، زیرساخت ابر و ذخیره‌سازی داده از چارچوب BDaaS را پوشش می‌دهد. DynamoDB پیچیدگی نصب و راه‌اندازی یک NoSQL را سهل و آسان کرده و کاربران را قادر می‌سازد تا از طریق یک لایه نمایش به داده‌ها دسترسی داشته باشند. DynamoDB شبیه به سایر ذخیره‌سازهای کلید-مقدار همانند کاساندر یا MongoDB است، اما به‌عنوان یک سرویس توسط آمازون ارائه شده است.

#### ۴-۳- آمازون Elastic MapReduce

EMR یک سرویس آمازون است که از چارچوب پردازش توزیع‌شده مثل هادوپ برای پردازش داده عظیم استفاده می‌نماید. EMR به کاربران این امکان را می‌دهد تا بدون اطلاع از نحوه نصب، راه‌اندازی و مدیریت زیرساخت هادوپ از آن به صورت دلخواه استفاده کنند. علاوه بر اینکه EMR از یک چارچوب توزیع‌شده استفاده می‌نماید، همچنین به سرعت و به صورت مؤثری داده‌های زیادی را مطابق با کلاستر مقیاس‌پذیر پردازش می‌کند. EMR قابل‌استفاده در کاربردهای متعددی از جمله تجزیه و تحلیل، انبار داده و یادگیری ماشین است. نمونه یک کلاستر هادوپ و مدیریت کارهای Map-Reduce در EMR به صورت خودکار هستند. کاربران آپاچی می‌توانند برنامه‌های کاربردی برای دسترسی به داده‌های پردازش‌شده را توسط EMR بی‌افزاینده [۱۸].

EMR از یک سرویس آمازون S3 و هادوپ سفارشی برای انجام عملیات خود استفاده می‌کند. کاربران باید داده‌های مدنظر خود برای پردازش را در یک باکت<sup>۴</sup> آمازون S3 بارگذاری کنند که این کار از طریق رابط آمازون صورت می‌گیرد. از آنجائی که داده در منبع S3 بارگذاری می‌شود، EMR می‌تواند مورد استفاده وظایف Map-Reduce قرار گیرد تا عمل پردازش داده را انجام دهد. برای مثال، اگر داده‌ها در یک فایل متنی جمع‌آوری شده باشند، کاربر می‌تواند آن را به یک باکت آمازون S3 بارگذاری کرده و سپس یک برنامه سفارشی روی سرویس EMR اجرا نماید. پس از انجام عمل پردازش، خروجی این باکت برای استفاده باکت دیگر ذخیره‌سازی می‌شود [۱۸].

• سرویس نظارت ابر آمازون<sup>۱</sup>: کنسولی را فراهم می‌کند که در آن مصرف منابع، شاخص‌های کلیدی عملکرد سایت و نشانگرهای عملیاتی برای عواملی همچون تقاضای پردازشگر، مصرف دیسک و ورودی و خروجی شبکه ارائه می‌شود. نتایج معیارهایی که توسط آن کسب می‌شود برای فعال‌سازی قابلیت به نام Scaling Auto مورد استفاده قرار می‌گیرد که به صورت خودکار می‌تواند یک سایت EC2 را بر مبنای مجموعه‌ای از قوانین که توسعه‌دهنده ایجاد می‌کند، توسعه دهد.

• توازن بار منعطف<sup>۲</sup>: نمونه‌های ماشین آمازون<sup>۳</sup> با استفاده از این قابلیت، دارای امکان توازن بار ترافیکی می‌شوند. این قابلیت هنگامی که نمونه‌ای دچار شکست می‌شود آن را کشف کرده و ترافیک را به یک نمونه سالم حتی نمونه‌ای در محیط‌های دیگر AWS مسیریابی مجدد می‌کند.

آمازون EC2 در حقیقت یک سامانه IaaS است که در لایه‌های زیرساخت و زیرساخت ابر از چارچوب BDaaS ارائه خدمت می‌کند. چون این سامانه فقط زیرساخت‌ها را ارائه می‌کند، کاربران باید ابتدا ماشین‌های مجازی موردنیاز خود را روی این زیرساخت نصب کنند سپس سامانه فایل موردنیاز خود (برای مثال HDFS) را پیکربندی کرده و داده‌های حجیم خود را روی آن پیاده کنند. این امر موجب ارائه انعطاف‌پذیری بالا برای کاربران می‌شود اما نیازمند دانش فنی بالا برای پیکربندی است.

#### ۴-۲- آمازون Dynamo DB

آمازون Dynamo یک سرویس پایگاه داده NoSQL مبتنی بر ابر است که کاربران را قادر می‌سازد تا حجم کلانی از داده‌ها را ذخیره و بازیابی نمایند. آمازون DynamoDB مبتنی بر چارچوب Dynamo برای سامانه‌های ذخیره‌سازی توزیع‌شده کلید-مقدار است. آمازون Dynamo متفاوت‌تر از سایر سرویس‌های آمازون بوده و توان عملیاتی بالایی در ذخیره‌سازی دارد. این سرویس نرم‌افزاری، یک پایگاه داده سریع از نوع NoSQL در اختیار برنامه‌های کاربردی قرار می‌دهد که در عین قابل‌اطمینان بودن، زمان پاسخی در حد تنها چند میلی‌ثانیه (تک‌رقمی) دارد. همچنین نگهداری از آن کاملاً توسط خود آمازون صورت می‌گیرد و مدل داده‌ای آن کاملاً منعطف است. کد نویسی در DynamoDB، ساده‌تر بوده ولی این سرویس محدودیت‌هایی نیز دارد [۱۷].

<sup>۱</sup> Amazon Cloud Watch

<sup>۲</sup> Elastic Load Balancing

<sup>۳</sup> Amazon Machine Image

<sup>۴</sup> Bucket

داده‌ها را به‌طور مستقیم از یک واسط کاربری صفحه گسترده به BigQuery منتقل کنند [۱۹].

داده‌های ذخیره‌شده در BigQuery با چندلایه‌ی امنیتی محافظت می‌شوند و در چندین مرکز داده تکرار شده‌اند تا ضریب اطمینان را در این سامانه بالا ببرند و به‌راحتی نیز می‌توانند به خروجی تبدیل شوند. علاوه بر این، سامانه ذخیره‌سازی به‌شدت هزینه‌ها را کاهش داده و اجرای تجزیه‌وتحلیل داده‌های عظیم به راه‌اندازی مرکز داده نیاز ندارد. این سرویس، لایه محاسبات و مدیریت داده BDaaS را پوشش می‌دهد.

#### ۴-۵ - Splunk Storm

Splunk Storm یک پلتفرم تجزیه‌وتحلیل داده‌ها برای مانیتورینگ برخط و آنالیز داده‌های لاگ تولیدی توسط ماشین‌هاست. این پلتفرم کاربران را قادر می‌سازد که یک مخزن داده ایجاد کنند و سپس داده‌های خود را نمایه کنند تا ارائه نمودارها و دیگر رابط‌های بصری اطلاعات تسهیل پیدا کند. Splunk Storm الگوهای داده‌ها را شناسایی می‌کند و به انتشار اطلاعات درون سازمان از طریق یک رابط مبتنی بر وب کمک می‌کند. این پلتفرم روی ابر Splunk پیاده می‌شود که کاربران می‌توانند داده‌های خود را روی ابر بارگذاری کنند و سپس بر روی داده‌های خود پرس‌وجوهای موردنظر خود را اعمال کنند و داشبوردهای سفارشی برای خود ایجاد کنند. این پلتفرم باعث ایجاد مقیاس‌پذیری، سادگی پیاده‌سازی و پرداخت هزینه به ازای میزان استفاده می‌شود.

Splunk Storm یک رابط مبتنی بر وب در اختیار کاربران خود می‌گذارد تا از طریق اعمال پرس‌وجوهای خود، تجزیه‌وتحلیل موردنظر را انجام دهند. این امر موجب می‌شود تا کاربران بتوانند اطلاعات موردنیاز خود را از فایل‌های لاگ، بدون نیاز به نوشتن یک برنامه مخصوص، استخراج کنند. برای مثال اگر مهندس‌ان بخواهند یک خطای به‌خصوص را در فایل‌های لاگ جستجو کنند که توسط یک شناسه خاص قابل‌شناسایی است، به‌سادگی می‌توانند از داشبورد به‌خصوص برای این کار که در Splunk Storm تعبیه‌شده است استفاده کنند.

Splunk بر روی سه لایه از چارچوب BDaaS ارائه خدمات می‌کند که دو لایه محاسبات و تجزیه تحلیل توسط خود Splunk ارائه می‌شود و لایه مربوط به ذخیره‌سازی داده‌ها از آمازون AWS سرویس می‌گیرد. لایه تجزیه‌وتحلیل برای مدیریت و پردازش فایل‌های لاگ استفاده می‌شود که نتایج بعد از تجزیه‌وتحلیل به لایه ذخیره‌سازی داده فرستاده می‌شوند. قیمت‌گذاری بر روی خدمات نیز بر اساس مقدار داده‌ای که کاربر هر روز بر روی Splunk نمایه می‌کند محاسبه می‌شود.

همان‌طور که گفته شد آمازون EMR از S3 برای ذخیره‌سازی داده‌ها استفاده نموده و توسط چارچوب Map-Reduce قابلیت پردازش داده‌های عظیم را نیز دارد. این سرویس آمازون چهار لایه زیرین BDaaS یعنی زیرساخت، زیرساخت ابر، ذخیره‌سازی داده و محاسبات را پوشش می‌دهد.

#### ۴-۴ - Google BigQuery

شرکت گوگل یک سرویس ابر برای اجرای پرس‌وجو روی مجموعه‌های داده عظیم فراهم کرده است که به کاربران این امکان را می‌دهد تا بدون نصب و راه‌اندازی زیرساخت با هزینه گزاف، به داده عظیم دسترسی بلادرنگ داشته و با این کار صرفه‌جویی در منابع انجام گیرد. سرویس Google BigQuery از طریق یک رابط کاربری مبتنی بر وب یا واسط کاربری مستقل قابل دسترسی است. این سرویس با دیگر بسترهای گوگل یکپارچه بوده و می‌تواند برای سازمان‌هایی که علاقه‌مند به پردازش داده‌های عظیم بدون هزینه سربار زیاد هستند، مفید باشد. درواقع این سرویس گوگل، امکانی فراهم می‌کند تا حجم کلانی از داده‌ها با استفاده از SQL در ابر مورد تجزیه‌وتحلیل قرار گیرند. برخی از شرکت‌ها از جمله Claritics (شرکت تجزیه‌وتحلیل‌های اجتماعی و سیار) از این سرویس برای ساخت نرم‌افزارهای موردنیاز خود استفاده کرده‌اند که به تولیدکنندگان امکان می‌دهد رفتار کاربران را به‌صورت بلادرنگ مشاهده کنند.

BigQuery برای پردازش داده‌ها از فناوری Dremel استفاده می‌کند که این فناوری یک سامانه پرس‌وجوی تعاملی، ادهاک و مقیاس‌پذیر بوده و توانایی کار با داده‌های عظیم را دارد. Dremel با ترکیب درخت‌های اجرای چند سطحی و طرح داده‌های ستونی به کاربران اجازه می‌دهد از پرس‌وجوهای در یک‌زبان شبیه SQL برای کار بر روی جداول داده عظیم استفاده کنند که این جداول می‌توانند حاوی میلیاردها خط داده باشند.

برای استفاده از داده‌ها در سرویس BigQuery، ابتدا باید داده‌ها در سامانه ذخیره‌سازی گوگل ذخیره شوند. برای این کار، باید داده‌ها از حالت نرمال دربیابند و همه فیلدها در یک جدول قرار گرفته و سپس در سامانه ذخیره‌سازی گوگل بارگذاری شوند. همچنین BigQuery نیازمند شناسایی و تصدیق همه‌ی درخواست‌ها است. BigQuery گوگل دارای ویژگی‌های مختلفی است که برخی از آنها اعم از مقیاس‌پذیری، سرعت، انعطاف، یکپارچگی و قابلیت دسترسی هستند. مقیاس‌پذیری بدین معناست که سامانه ذخیره‌سازی داده‌ها می‌تواند خیلی راحت به سدها ترابایت داده ارتقا پیدا کند، بدون آن‌که نیازی به مدیریت آن‌ها باشد. یکپارچگی این سامانه با صفحه گسترده‌ی گوگل باعث شده است که تحلیل‌گران داده‌های عظیم بتوانند حجم کلانی از

انجام می‌دهد و سپس نتایج تجزیه تحلیل داده هار را روی سامانه‌های Tibco بارگذاری می‌شوند. از این رو این فناوری روی دو لایه مدیریت داده و تجزیه تحلیل داده از چارچوب BDaaS ارائه خدمات می‌کند.

#### ۴-۸- QuBole

QuBole یک شرکت ارائه‌دهنده داده‌های عظیم به‌عنوان سرویس است که چارچوب هادوپ را بر روی پلتفرم ابر ارائه می‌کند. QuBole پیچیدگی‌های مدیریت خوشه‌های هادوپ را کاهش داده و مقایسه پذیری را به‌صورت بلادرنگ ارائه می‌کند و به کاربران اجازه می‌دهد به داده‌های ذخیره‌شده روی پلتفرم ابر مانند آمازون S3 دسترسی پیدا کنند. این پلتفرم دارای یک ماشین هادوپ داخلی است که برای اجرای سریع‌تر، تخصیص منابع را بهینه‌سازی می‌کند. کاربران می‌توانند داده‌های خود را به‌صورت مستقیم از آمازون S3 بارگذاری کرده و جداول خود را توسط واسط GUI مربوط به QuBole ایجاد کنند سپس نتایج مربوط به پردازش‌ها می‌توانند به‌طور مستقیم به منبع ذخیره‌سازی کاربر بازگردند. QuBole به‌وسیله فعال کردن تدریجی گره‌های موردنیاز هنگام آغاز شدن هر کار و ادامه آن، کاربران را قادر می‌سازد تا هزینه‌های خود را کاهش دهند در این حالت سامانه دارای مقیاس پذیری مبتنی بر تقاضا است که به‌صورت اتوماتیک در صورت نیاز به گره‌های بیشتر آن‌ها را فعال می‌کند. این پلتفرم توجهی به زیرساخت مورد استفاده کاربر ندارد و کاربر می‌تواند از ابر یا حتی سامانه محلی خود استفاده کند.

از آنجایی که پلتفرم QuBole اساساً سرویس‌های پردازش داده را به‌وسیله محصولات آپاچی ارائه می‌دهد بنابراین روی دو لایه‌ی محاسبات و مدیریت داده از چارچوب BDaaS ارائه خدمت می‌کند همچنین این پلتفرم به دلیل استفاده از زیرساخت‌های آمازون S3 برای ذخیره‌سازی داده‌ها، در لایه ذخیره‌سازی داده ارائه خدمتی ندارد.

#### ۴-۹- ارزیابی و مقایسه فناوری‌های مختلف

در این قسمت فناوری‌های مختلف BDaaS معرفی شده را بررسی می‌کنیم. این بررسی به‌منظور تفکیک بهتر هر کدام از فناوری‌ها از یکدیگر صورت گرفته و مشخصات هر کدام در جدول (۳) جمع‌بندی شده است. نکته قابل ذکر این است که با توجه به این که فناوری‌های مختلف معرفی شده هر کدام برای ارائه سرویس‌های خاصی طراحی شده‌اند و هدف خاصی را دنبال می‌کنند بنابراین نمی‌توان در مورد برتری یکی بر دیگری اظهار نظر کرد بلکه کاربر بر اساس نیاز خود به سرویس‌های به‌خصوص، هر کدام از این فناوری‌ها را انتخاب می‌کند.

#### ۴-۶- Microsoft Azure-HDInsight

HDInsight یک راه‌کار ارائه داده‌های عظیم به‌عنوان سرویس توسط شرکت ماکروسافت است که آپاچی هادوپ را به‌صورت سرویس به کاربران ارائه می‌دهد. این امر موجب می‌شود که کاربران یک محیط مقیاس‌پذیر و کارا از نظر هزینه داشته باشند. HDInsight بهترین راهکار مبتنی بر ابر ماکروسافت برای تجزیه و تحلیل داده‌های عظیم است. این فناوری پلتفرم ابر ماکروسافت را با معماری پردازش توزیع‌شده هادوپ ترکیب می‌کند تا یک زیرساخت مبتنی بر تقاضا برای داده‌های عظیم به کاربران ارائه کند. HDInsight از مزایای Power View و PowerPivot در Excel برای دریافت اطلاعات از چندین منبع به‌منظور تجزیه و تحلیل داده‌ها بهره می‌گیرد. کلاسترهای Hadoop در چند لحظه و به‌منظور کمک برای تحلیل و درک بهتر داده‌ها ایجاد می‌شوند و زمانی که کار به اتمام رسید می‌توان این کلاسترها را بست.

HDInsight در عین حال به‌شدت قابل سفارشی شدن است و به کاربر اجازه می‌دهد از یک زبان اختصاصی مثل .net یا java به‌منظور تطبیق نتایج داده‌های بزرگ با سامانه استفاده کند. این فناوری منبع ذخیره‌سازی و سرویس‌های محاسباتی را روی زیرساخت‌های ماکروسافت به کاربران ارائه می‌دهد بنابراین مبتنی بر چهار لایه ابتدایی زیرساخت، زیرساخت ابر، ذخیره‌سازی داده و محاسبات از چارچوب BDaaS است [۲۰].

#### ۴-۷- Tibco Silver Spotfire

Tibco Spotfire یک پلتفرم برای تجزیه و تحلیل داده است که به کاربران اجازه می‌دهد تا کارهای تجزیه تحلیلی سنگین خود را روی داده‌ها انجام داده و نتایج را به‌صورت گزارش دریافت کنند. Spotfire همچنین می‌تواند با سایر محصولات Tibco همچون MDM<sup>۱</sup> و Rendezvous تعامل برقرار کند. MDM به کاربران اجازه می‌دهد تا داده‌های ناهمگن را از مراکز مختلف به‌منظور حذف داده‌های تکراری جمع‌آوری کنند و به‌صورت یک مرکز داده یکپارچه ذخیره کنند. Rendezvous Tibco نیز یک واسط ارتباطی بین برنامه‌های کاربردی مختلف فراهم می‌کند به‌نحوی که این برنامه‌ها می‌توانند داده‌های خود را از طریق این واسط با یکدیگر مبادله کنند.

Tibco Spotfire دارای اجزا و واسط‌هایی برای کاربران است. واسط‌ها مربوط به کاربر بر روی زیرساخت کاربر نصب می‌شود و کاربر تجزیه و تحلیل داده‌ها را بر روی زیرساخت و سامانه خود

<sup>۱</sup> Master Data Management



جدول (۳): مقایسه فناوری‌های مختلف داده‌های عظیم به‌عنوان سرویس.

نام فناوری	شرکت ارائه‌دهنده	لایه‌های BDAAS	مأموریت اصلی	مزایا	معایب
EC2	AMAZON	زیرساخت - زیرساخت ابر	ارائه زیرساخت به کاربران	انعطاف پذیری بالا	نیاز به دانش فنی بالا
DYNAMO DB	AMAZON	زیرساخت - زیرساخت ابر - ذخیره‌سازی داده	ارائه سرویس پایگاه داده	توان عملیاتی بالا برای ذخیره‌سازی داده - قابلیت اطمینان بالا	توانایی گسترش تنها روی AMAZON AWS
EMR	AMAZON	زیرساخت - زیرساخت ابر - ذخیره‌سازی داده - محاسبات	پردازش داده‌ها	مقیاس پذیری خودکار کلاسترها	نداشتن مدیریت داده‌ها
BigQuery	GOOGLE	محاسبات - مدیریت داده	اجرای پرس‌وجو روی داده‌های عظیم	استفاده از فناوری DREMEL - امنیت بالا	خارج کردن داده‌ها از حالت نرمال هنگام ذخیره‌سازی
Splunk Storm	Splunk	ذخیره‌سازی داده - محاسبات - مدیریت داده	مانیتورینگ برخط داده‌های ماشین	استفاده از نمایه‌سازی برای ذخیره‌سازی	نداشتن ذخیره‌سازی اختصاصی
Azure-HDInsight	Microsoft	زیرساخت - زیرساخت ابر - ذخیره‌سازی داده - محاسبات	ارائه آپاچی هادوپ به‌عنوان سرویس	قابلیت بالای سفارشی‌سازی	قابلیت کار فقط با محصولات ماکروسافت
Silver Spotfire	Tibco	مدیریت داده - تجزیه و تحلیل داده	تجزیه و تحلیل داده‌ها	سادگی کار برای کاربران	عدم توانایی انجام محاسبات بر روی داده‌ها
QuBole	QuBole	محاسبات - مدیریت داده	ارائه چارچوب هادوپ به‌عنوان سرویس	برخورداری از یک ماشین هادوپ داخلی	نداشتن ذخیره‌سازی اختصاصی

## ۵- نتیجه‌گیری

با رشد روزافزون داده‌ها نیاز به ابزارهای کار با داده‌های عظیم نیز پررنگ‌تر شده است از طرف دیگر فناوری اطلاعات در حال حرکت به سمت برون‌سپاری فعالیت‌ها و استفاده از سرویس‌های ابری است. این رویکردها موجب شده است شرکت‌های پیشرو در فناوری اطلاعات از قابلیت ترکیب و تلفیق داده‌های عظیم و محاسبات ابری استفاده کنند تا

فناوری‌های جدیدی ایجاد نمایند که بتواند مشکلات حوزه داده عظیم را حل نماید. سازمان‌ها و شرکت‌های کوچک می‌توانند در مواجهه با داده‌های عظیم با استفاده از فناوری‌های داده عظیم به‌عنوان سرویس، هزینه‌های نصب و راه‌اندازی سامانه‌های جمع‌آوری، ذخیره‌سازی، جستجو و پردازش داده عظیم را کاهش داده و به‌آسانی و بدون نیاز به دانش خاصی از این سرویس‌ها استفاده نمایند. در این مقاله داده‌های عظیم به‌عنوان سرویس بررسی شده و لایه‌های

- [10] S. Chethan, "Compendious study of Big Data as a Service (BDaaS)," *International Journal of Control Theory and Applications*, vol. 10(14), pp. 59-63, 2017.
- [11] E. Xinhua, J. Han, Y. Wang, and L. Liu, "Big Data-as-a-Service: Definition and architecture." In *Communication Technology (ICCT)*, " 2013 15th IEEE International Conference on, IEEE, pp. 738-742, November 2013.
- [12] Z. Zheng, J. Zhu, and M. R. Lyu, "Service-generated big data and big data-as-a-service: an overview," In *Big Data (BigData Congress)*, 2013 IEEE International Congress on, pp. 403-410, June 2013.
- [13] B. Marr, "The Next Big Thing In Big Data: BDaaS, Retrieved from linkedin," july 2015. <https://www.linkedin.com/pulse/next-big-thing-data-bdaas-bernard-marr>
- [14] Y. Kim, Y. H. Kim, G. W. Lee, and E. N. Huh, "Survey of BigData-as-a-Service Type," In *High Performance Computing and Communications (HPCC)*, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), 2015 IEEE 12th International Conferen on Embedded Software and Systems (ICSS), 2015 IEEE 17th International Conference on, pp. 1573-1578, August 2015.
- [15] Q. H. Vu and R. Asal, "A framework for big data as a service," In *Digital Signal Processing (DSP)*, 2015 IEEE International Conference on, pp. 492-496, July 2015.
- [16] A. E. C. Cloud, "Amazon web services," Retrieved November 2011.
- [17] S. Sivasubramanian, "Amazon dynamoDB: a seamlessly scalable non-relational database service," In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pp. 729-730, May 2012.
- [18] A. Singh and V. Rayapati, "Learning Big Data with Amazon Elastic MapReduce," Packt Publishing Ltd, 2014.
- [19] S. B. Siewert, "Big data in the cloud. Data velocity," variety, veracity, pp. 4-21, 2013.
- [20] A. Chauhan, V. Fontama, M. Hart, W. H. Tok, and B. Woody, *Introducing Microsoft Azure HDInsight*, Microsoft press, 2014.

چارچوب آن بیان شد سپس هشت فناوری مهم در این حوزه تشریح شده و این فناوری ها با یکدیگر مقایسه و ارزیابی شدند. نتایج بررسی ها نشان می دهد نمی توان در مورد برتری یک سرویس نسبت به سرویس های دیگر قضاوت کرد بلکه هدف ارائه دهندگان سرویس، ارائه خدمات متنوع تر و مبتنی بر تقاضا است که کاربران با توجه به نیازهای خود سرویس های موردنظر را انتخاب کنند.

## ۶- مراجع

- [1] C. Snijders, U. Matzat, and U. D. Reips, "Big Data, Big gaps of knowledge in the field of internet science," *International Journal of Internet Science*, vol. 7(1), pp. 1-5, 2012.
- [2] P. A. Sria and M. Anusha, "Big Data-Survey," *Indonesian Journal of Electrical Engineering and Informatics (IJEI)*, vol. 4(1), pp. 74-80, 2016.
- [3] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation," *Competition, and Productivity*, 2011.
- [4] L. Liu, "Computing infrastructure for big data processing," *Frontiers of Computer Science*, vol. 7(2), pp. 165-170, 2013.
- [5] Everything You Need to Know About Big Data as a Service (BDaaS), 2015. Available at <http://www.cleverism.com/everything-need-know-big-data-as-a-service-bdaas>.
- [6] Oracle Information Architecture: An Architect's Guide to Big Data, Oracle, August 2012.
- [7] A. Rangarajan and D. Batts, "Small Steps to Big Data-An SOA-Driven Framework for Meaningful Healthcare Analytics Implementation," 2015.
- [8] M. Beyer, J. D. Lovelock, D. Sommer, and M. Adrian, "Big Data Drives Rapid Changes in Infrastructure and \$232 Billion in IT Spending Through 2016," Gartner Research, 2012.
- [9] X. Yang and H. Zhang, "Cloud computing and SOA convergence research," In *Computational Intelligence and Design (ISCID)*, 2012 Fifth International Symposium on, IEEE, vol. 1, pp. 330-335, October 2012.

---

## **Review and Evaluation of Big Data Technologies as a Service**

**M. R. Hoseinzadeh, A. Ghafouri\*, H. Naderi**

Iran University of Science & Technology

### **Abstract**

With the advent of big data technology as a service, small organizations can provide services they need in the case of big data through corporate services. This work reduces the resources required for pilot projects and reduces costs. Due to the layers of the big data framework as a service, leading companies each offer their services on some layers of this framework, and users can choose the service provider according to their needs. In this paper, eight important technologies that are presented as big data as a service by different companies were reviewed and compared with each other. An important issue to consider is the variety of services provided by different companies so that one can not judge the superiority of one service over other services, but the goal of service providers is to provide more diverse and demand-based services that users according to their needs Choose them.

**Keywords:** Big Data, Big Data as a Service, BDaaS, Cloud

---

\* Corresponding author E-mail: krghafouri@ihu.ac.ir