

به نام خدا



پژوهشکده فناوری اطلاعات

"پروژه تدوین نقشه راه کلان داده‌ها"

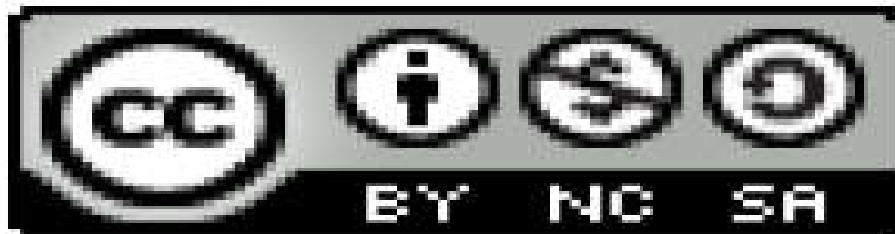
گزارش فاز اول

"کاربردهای رایانش ابری در کلان داده‌ها"

کد پروژه: ۹۰۴۹۵۰۱۰۰

مجری:	محمد شهرام معین
تهیه کننده:	زهرا معز کریمی، سید رضا متقی و محمد قاسم پور
کد گزارش:	
تاریخ ارائه:	۹۶/۱/۲۳
نسخه / وضعیت:	۱ / نهایی

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--



در راستای تحقق مأموریت پژوهشگاه ارتباطات و فناوری در فراهم سازی سکویی برای ارتقاء دانش، انتقال فناوری و بومی سازی محصولات و خدمات حوزه فاوا و با هدف جلب مشارکت علاقه مندان در توسعه و بهره مندی از دستاوردهای پژوهشگاه ارتباطات و فناوری اطلاعات، آزاد رسانی این دستاوردها در زمره برنامه های اولویت دار پژوهشگاه به شمار می آید. به همین منظور مستند حاضر تحت مجوز بین المللی CC-BY-SA-NC نسخه 4 ، در دسترس عموم قرار گرفته است. شایان ذکر است تحت این مجوز، ضمن حفظ مالکیت فکری این مستند برای پژوهشگاه ارتباطات و فناوری اطلاعات، باز انتشار و بکارگیری آن صرفاً برای موارد تحقیقاتی و با ذکر نام پژوهشگاه ارتباطات و فناوری اطلاعات بلامانع است.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## شناسنامه گزارش

عنوان: کاربردهای رایانش ابری در کلان داده ها	شماره نسخه: ۱
کد:	نوع گزارش: راهبردی
نام پروژه: تدوین نقشه راه کلان داده ها	تاریخ ارائه گزارش: ۹۶/۱/۲۳
تاریخ شروع: ۹۵/۷/۶	نوع پروژه: راهبردی
تاریخ پایان: ۹۶/۴/۶	نام گروه: فناوری اطلاعات / سامانه های چندرسانه ای
کد پروژه: ۹۰۴۹۵۰۱۰۰	شماره و تاریخ قرارداد: حکم شماره ۵۰۰/د/۶۷۰۲/پ در تاریخ ۹۵/۸/۱۰
مجری: محمدشهرام معین	ناظر / ناظرین: آقایان دکتر علیرضا یاری، دکتر روح ا... رحمانی، دکتر مجید رسولی دیسفانی، دکتر امین شکری پور و مهندس فرزاد ابراهیمی
تهیه کننده / تهیه کنندگان: زهرا معز کریمی، سیدرضا متقی و محمد قاسم پور	
نام و نشانی مجری: تهران، انتهای خیابان کارگر شمالی، پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران) _ کد پستی: ۱۴۳۹۹۵۵۴۷۱ _ تلفن: ۸۴۹۷۷۵۸۵	
نام و نشانی حمایت کننده: تهران، انتهای خیابان کارگر شمالی، پژوهشگاه ارتباطات و فناوری اطلاعات (مرکز تحقیقات مخابرات ایران) _ کد پستی: ۱۴۳۹۹۵۵۴۷۱ _ تلفن: ۸۰۰۵۵۰۸-۱۰	
ملاحظات:	
چکیده: در این گزارش ابتدا به طور مختصر به مفاهیم اساسی کلان داده ها و رایانش ابری پرداخته می شود. سپس چالش های کلان داده ها و راهکارهای رایانش ابری برای مواجهه با این چالش ها معرفی می گردد. در ادامه، ارتباط رایانش ابری با کلان داده ها و گونه شناسی کلان داده ها از دیدگاه ابر شرح داده می شود. در انتها سرویس های کلان داده های موجود روی بستر ابر معرفی می گردد.	
کلمات کلیدی: رایانش ابری، کلان داده ها، خدمات کلان داده ها	
وضعیت گزارش: نهایی	زبان گزارش: فارسی
وضعیت دسترسی: عادی	تعداد صفحات: ۳۰

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## چکیده

در این گزارش ابتدا به طور مختصر به مفاهیم اساسی کلان داده ها و رایانش ابری پرداخته می شود. سپس چالش های کلان داده ها و راهکارهای رایانش ابری برای مواجهه با این چالش ها معرفی می گردد. در ادامه، ارتباط رایانش ابری با کلان داده ها و گونه شناسی کلان داده ها از دیدگاه ابر شرح داده می شود. در انتها سرویس های کلان داده های موجود روی بستر ابر معرفی می گردد.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

## اطلاعات مرتبط

## مستندات مرتبط

شماره مستند	نوع مستند	نام مستند

## تغییرات اعمال شده در نسخه های پیشین

شماره نسخه	تاریخ	تغییرات اعمال شده

## تأییدکنندگان

نام و نام خانوادگی	تاریخ	امضاء	ملاحظات
محمدشهرام معین			مجری پروژه
زهرا معز کریمی، سیدرضا متقی و محمد قاسم پور			تهیه کننده / تهیه کنندگان
دکتر علیرضا یاری، دکتر روح ا... رحمانی، دکتر امین شکری پور، دکتر مجید رسولی دیسفانی، مهندس فرزاد ابراهیمی			ناظر پروژه
مهندس فرزاد ابراهیمی			مدیر گروه
مانا روزی طلب			مسئول مستندات پژوهشکده
دکتر علیرضا یاری / دکتر کامبیز بدیع			رئیس پژوهشکده / معاون پژوهشی

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

## تقدیر و تشکر

بدین وسیله از ناظرین و مشاورین محترم پروژه قدردانی می شود.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

## سرفصل مطالب

۹	۱	مقدمه
۱۰	۲	شناخت مفاهیم اساسی
۱۰	۲-۱	ویژگی های کلان داده ها
۱۲	۲-۲	چرخه حیات کلان داده ها
۱۳	۲-۳	معرفی چالش ها
۱۴	۲-۴	راهکارهای رایانش ابری برای چالشهای حوزه کلان داده ها
۱۵	۲-۵	مفاهیم رایانش ابری
۱۷	۲-۶	انواع سرویسهای رایانش ابری
۱۸	۲-۷	مدل های به کارگیری رایانش ابری
۲۰	۳	ارتباط رایانش ابری با کلان داده ها
۲۱	۳-۱	گونه شناسی کلان داده از دیدگاه ابر
۲۳	۳-۲	بررسی سرویس های موجود کلان داده ها در ابر
۲۳	۳-۲-۱	بررسی سرویس های ذخیره سازی
۲۴	۳-۲-۲	بررسی سرویس های محاسباتی
۲۴	۳-۲-۳	بررسی سرویس های پردازشی
۲۵	۳-۲-۴	بررسی سرویس های تحلیل و بصری سازی داده
۲۷		<b>مراجع</b>
۲۷		<b>واژه نامه</b>
۲۷		انگلیسی به فارسی
۲۹		فارسی به انگلیسی

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## فهرست جداول

جدول ۱: واژه‌نامه انگلیسی به فارسی ..... ۲۹

جدول ۲: واژه‌نامه فارسی به انگلیسی ..... ۳۰

## فهرست اشکال

شکل ۱: چهار ویژگی اصلی کلان داده ها ..... ۱۱

شکل ۲: معماری رایانش ابری ارائه شده توسط سازمان ملی استاندارد و تکنولوژی امریکا [4] ..... ۱۶

شکل ۳: انواع سرویس، مشخصه ها و پیاده سازی رایانش ابری ..... ۱۹

شکل ۴: معماری نمونه استفاده رایانش ابری در اکوسیستم کلان داده ها [۵] ..... ۲۰



نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## ۱ مقدمه

بسیاری از صنایع، مانند مخابرات، بهداشت و درمان و خدمات مالی، حجم زیادی از داده‌ها را تولید می‌کنند. دستیابی به یک دانش در مورد کسب و کار نیازمند تحلیل و بررسی حجم انبوهی از داده‌های تولید شده است. امروزه به دلیل توسعه کسب و کار و افزایش حجم داده‌های تولید شده به همراه افزایش سرعت تولید آنها نیاز به پردازش و ذخیره‌سازی سریع و مقیاس‌پذیر داده‌ها می‌باشد. این در حالی است که هزینه این سیستم‌ها روزبه‌روز از اهمیت بالاتری برخوردار می‌شود. در این مستند در فصل ۲ به تعریف مفاهیم اساسی حوزه کلان داده‌ها می‌پردازیم. همچنین بخشی از مهمترین چالش‌های این حوزه را معرفی و مورد بررسی قرار می‌دهیم و راهکار رایانش ابری را برای این مواجهه با این چالش‌ها معرفی می‌کنیم. همچنین در فصل ۳ جایگاه رایانش ابری را در حوزه کلان داده‌ها بررسی خواهیم کرد.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## ۲ شناخت مفاهیم اساسی

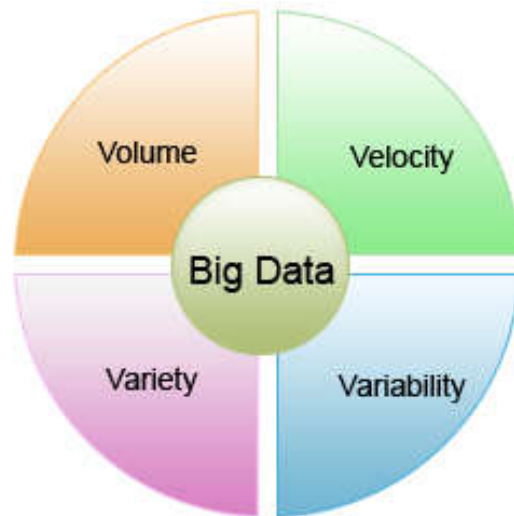
کلان داده ها در واقع تمام داده‌هایی که در اطراف ما در حال تولید هستند می‌باشند. هر فرآیند دیجیتال و هر تبادل در رسانه‌های اجتماعی این داده‌ها را تولید می‌کند. حسگرها، سامانه‌ها، و دستگاه‌های موبایل در حال ارسال این داده‌ها می‌باشند. در حوزه کلان داده ها بر طبق تعریف سازمان ملی استاندارد و فناوری (NIST) چهار ویژگی وجود دارد که کلان داده ها دارای این چهار ویژگی اصلی می‌باشند. در حالی که ویژگی دیگری نیز به کلان داده ها نسبت داده‌اند، ولی سیستم‌های موازی جدید فقط به سمت پیاده‌سازی این چهار ویژگی برای کاربردهای کلان داده ها رفته اند. در ادامه به معرفی این چهار ویژگی به صورت مختصر می‌پردازیم.

### ۲-۱ ویژگی های کلان داده ها

اولین ویژگی که کلان داده ها دارند حجم بسیار زیاد آن‌ها است. با گسترش روزافزون حسگرها و منابع تولید داده، این حجم به صورت نمایی گسترش می‌یابد و کلان داده ها در طول زمان حجم‌های افزایشی را خواهند داشت. [1]

دومین ویژگی کلان داده ها سرعت تولید آن‌ها است. نرخ تولید کلان داده ها می‌تواند متغییر باشد. همان‌طور که اشاره شد، منابع تولید داده‌ها در حال افزایش هستند، با گسترش این منابع و تعداد زیاد آن‌ها سرعت تولید داده‌ها نیز افزایش پیدا می‌کند. حتی برخی داده‌ها نیاز به پردازش بلادرنگ نیز دارند که همین مسئله یکی از چالش‌های مهم در حوزه کلان داده ها می‌باشد.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----



شکل ۱: چهار ویژگی اصلی کلان داده ها

سومین ویژگی کلان داده ها تنوع آن ها می باشد. با توجه به انواع مختلف حسگرها و تولیدکننده ها، داده های مختلف با گوناگون بسیار در حال تولید است. این داده ها می توانند ساخت یافته یا غیر ساخت یافته باشند که در ادامه به بررسی این دو مفهوم خواهیم پرداخت. در ویژگی تنوع داده، چالش بزرگ بروز خطا به خاطر همین تنوع است. بسیاری از اطلاعات مستقیماً از انسان دریافت می شود و بنابراین وجود خطا اجتناب ناپذیر است.

در نهایت چهارمین ویژگی که در تعریف کلان داده ها وجود دارد تغییرپذیری در ویژگی داده ها است. شکل ۱ به صورت کلی این ویژگی ها را نمایش می دهد.

ویژگی های دیگری نیز برای کلان داده ها ذکر شده است. به عنوان نمونه، یکی از این ویژگی ها صحت داده ها است. بدین معنا که چه مقدار از مقادیر تولید شده دقیق هستند و داده صحیح تولید می شود. برای مثال آیا حسگرهای موجود مقادیر صحیح را به سیستم ارسال می کنند یا به هر دلیلی خطایی در داده های آن ها وجود دارد که صحت داده ها را نامشخص می سازد. در نتیجه ممکن است کلان داده ها لزوماً از صحت کافی برخوردار نباشند.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## ۲-۲ چرخه حیات کلان داده ها

برطبق تعریف سازمان ملی استاندارد و فناوری در علم کلان داده ها چهار فاز اصلی وجود دارد که به نام چرخه حیات داده مشهور است. چرخه حیات داده مجموعه عملیاتی است که یک برنامه کاربردی انجام می دهد تا یک داده خام به یک دانش قابل اقدام تبدیل شود. این چرخه از زمانی که داده توسط فراهم کننده سرویس کلان داده ها دریافت می شود شروع و تا زمانی که روی آن تحلیل انجام می گیرد و نتیجه موردنظر استخراج می گردد ادامه می یابد.

این چهار فاز اصلی (که هر کدام شامل زیرفرایندهایی هستند) عبارتند از: [2]

- ۱) جمع آوری و ذخیره سازی: در این مرحله داده ها دریافت و برای پردازش ذخیره سازی می شوند.
- ۲) محاسبه و پردازش: در این مرحله مجموعه ای از پردازش ها انجام می شود که داده خام به اطلاعات سازماندهی شده تبدیل شوند.
- ۳) تحلیل: در این مرحله الگوهای معنی دار از اطلاعات سازمان دهی شده استخراج می شود.
- ۴) اقدام: در این مرحله از الگوهای معنی دار داده هایی تولید می شود که می توان بر اساس آنها اقدام مهمی انجام داد.

داده ها توسط منابع تولیدکننده به وجود می آیند و می بایست توسط یک شبکه به سیستم پردازشی و ذخیره سازی انتقال پیدا کنند. پس از دریافت داده روی داده ها پردازش و محاسبه انجام می گیرد. ممکن است به برخی از داده ها نیاز نباشد یا داده ها ساختار مشخصی نداشته باشند که باید کارهای مورد نیاز پیش از پردازش انجام پذیرد. گاهی نیز براساس کاربرد و نیاز ممکن است ابتدا داده ها ذخیره و سپس مورد پردازش قرارگیرند. در ادامه یک تحلیل برای استخراج نتایج موردنظر روی داده ها صورت می گیرد و نتایج تحلیل برای اقدام مناسب در اختیار کاربر قرار می گیرد. شرکت های فراهم کننده برنامه های کاربردی کلان داده ها در طول چرخه حیات داده ابزارهای مختلفی را ارائه کرده اند. در بخش اقدام به منظور استفاده مصرف کننده کلان داده ها از این داده ها ابزارهایی برای بصری سازی و دسترسی داده ها وجود دارد که گاهی به جای مرحله اقدام این دو مرحله در چرخه حیات داده ذکر می شود. در ابزارهای بصری سازی هدف تولید یک نمونه بصری از الگوهای معنی دار می باشد. در ابزارهای دسترسی نیز همانند کاری که در ابزارهای جمع آوری داده از تولید کننده داده انجام می شد این بار برای تحویل داده به مصرف کننده داده انجام می شود.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

### ۲-۳ معرفی چالش ها

با توجه به گسترش روزافزون حجم و سرعت تولید داده‌ها، بدیهی است که برای ذخیره‌سازی و پردازش این داده‌ها که هر روز سریع‌تر تولید می‌شوند و حجیم‌تر می‌شوند، نیاز به سیستم پردازشی و ذخیره‌سازی مقیاس‌پذیر است. [3] لزوم پردازش بسیار سریع داده‌ها از یک منظر دیگر نیز اهمیت دارد. برخی داده‌ها ارزش زمانی دارند و منقضی می‌شوند و یا برخی از داده‌ها نیاز به پردازش بلادرنگ دارند. برای مثال کاربران در موتور جستجو کلمه‌ی کلیدی خاصی را جستجو می‌کنند. اگر پردازش بر این کلمه‌های کلیدی و فراوانی آن‌ها به‌موقع انجام شود، می‌توان پیشنهادهای مناسبی را برای کاربران ارائه داد.

منابع تولید داده همواره در حال تولید داده هستند، پس برای ذخیره‌سازی و پردازش نیاز به سیستمی است که همواره بدون بروز اشکال در حال پردازش و ذخیره‌سازی است، که به خاطر محدودیت‌های سخت‌افزاری باید از چندین پردازنده و ذخیره‌ساز جایگزین نیز بهره برد.

در حوزه کلان داده ها یکی از مهم‌ترین مسائل هزینه‌های راه‌اندازی و سخت‌افزاری است. با توجه به نکات ذکر شده نیاز به چندین سیستم پردازشی و ذخیره‌سازی با توانایی بالا است که راه‌اندازی و تهیه این سیستم‌ها هزینه بسیار زیادی را به سازمان تحمیل می‌کند.

همچنین در حوزه کلان داده ها نیاز به دسترسی به چندین تولید کننده داده و چندین مصرف کننده داده داریم. این دسترسی نیازمند ارتباطات است که لازم است فراهم شود.

برای حوزه کلان داده ها چالش‌های دیگری همچون گونه‌های متنوع داده (مانند تصویر، متن و غیره)، بررسی صحت داده‌ها، امنیت و سطوح دسترسی و بصری‌سازی نیز وجود دارد که در این مستند به طور خاص مورد بررسی قرار نمی‌گیرند.

به صورت کلی می‌توان چالش‌های اشاره شده و موردنظر این مستند را به صورت زیر دسته‌بندی نمود:

- نیاز به مقیاس پذیری
- نیاز به تحمل خرابی
- نیاز به پاسخ به موقع
- نیاز به در دسترس بودن

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

- نیاز به ارتباطات میان تولید کننده ها و مصرف کننده های داده
- نیاز به کاهش هزینه ها

#### ۲-۴ راهکارهای رایانش ابری برای چالش های حوزه کلان داده ها

در بخش قبل به برخی از مهمترین چالش های پیش روی حوزه کلان داده ها پرداختیم. در این حوزه به طور قطع چالش های بیشتری وجود دارد، ولی در این گزارش با توجه به هدف تولید این مستند، مهم ترین چالش ها را که می توان با راهکارهای رایانش ابری پاسخگوی آنها بود، مورد بررسی قرار دادیم. با توجه به چالش های ذکر شده نیاز به سیستم های پردازشی و ذخیره سازی با توان بالا و قابلیت مقیاس پذیری و مقاوم در برابر بروز اشکال با هزینه پایین وجود دارد.

رایانش ابری یکی از راهکارهای این حوزه است که در این مستند مورد توجه می باشد. در ادامه به برخی از ویژگی ها و مزایای رایانش ابری اشاره می کنیم که چگونه چالش های حوزه کلان داده ها را حل می کند.

- مقیاس پذیری
- تحمل پذیری در برابر خرابی
- هزینه کم راه اندازی
- قابلیت ذخیره سازی و پردازش موازی داده ها به منظور تولید پاسخ در زمان مورد نظر
- تداوم و قابلیت اعتماد
- ارتباطات گسترده

همچنین رایانش ابری علاوه بر مزایای فوق دارای نکات مثبت زیر نیز می باشد.

- منابع رایانشی از خارج از شرکت تأمین می شوند.
- منابع رایانشی (معمولاً) از طریق اینترنت قابل دستیابی هستند.
- هزینه استفاده از منابع بر اساس درخواست متغیر است.
- منابع رایانشی از دید کاربر مستقل از وسیله هستند.

با توجه به نکات مثبت و مزایای رایانش ابری در حوزه کلان داده ها، می توانیم راهکارهای مبتنی بر رایانش ابری را مناسب، کم هزینه و قابل اعتماد در نظر گرفت. در ادامه به بررسی مفاهیم رایانش ابری پرداخته خواهد شد.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

## ۵-۲ مفاهیم رایانش ابری

به بیان ساده، مدل رایانشی است که می توان سرویس های پردازشی مانند سرورها، ذخیره سازی، پایگاه داده، نرم افزار و غیره را در اختیار کاربر قرار داد. در بخش قبل به مزایای رایانش ابری اشاره شد. در این بخش به صورت کامل تر به بررسی این مزایا پرداخته می شود.

اولین مزیتی که سرویس های ابری برای کاربران دارند، هزینه پایین آن ها در مقایسه با خرید سخت افزار یا نرم افزار مورد نظر می باشد. خرید سخت افزار مستلزم پرداخت هزینه های دیگری همچون مصرف توان، نگهداری، خنک سازی و غیره نیز می باشد. به کمک رایانش ابری هزینه استفاده از سخت افزار بسیار پایین آمده است و کاربر بر اساس نیاز برای سرویس مورد نظر هزینه می کند.

مزیت دیگر رایانش ابری کارایی آن ها می باشد. سرویس های رایانش ابری به دلیل اینکه برحسب تقاضا می باشند، می توانند برای کاربرد مورد نظر بهینه شوند.

مزیت مهم دیگر رایانش ابری قابلیت مقیاس پذیری آن ها می باشد. این بدین معنا است که با کلیک می توان سیستم و سرویس مورد نظر را ارتقا داد تا پهنای باند، قدرت پردازشی و حجم ذخیره سازی کمتر یا بیشتر بر حسب نیاز آماده باشد.

همچنین رایانش ابری قابلیت اطمینان بالایی نیز دارد به دلیل اینکه سیستم های پشتیبان برای سرویس وجود دارد و از داده ها نیز نسخه پشتیبان تهیه می شود و تمام این کارها از دید کاربر مستقل است.

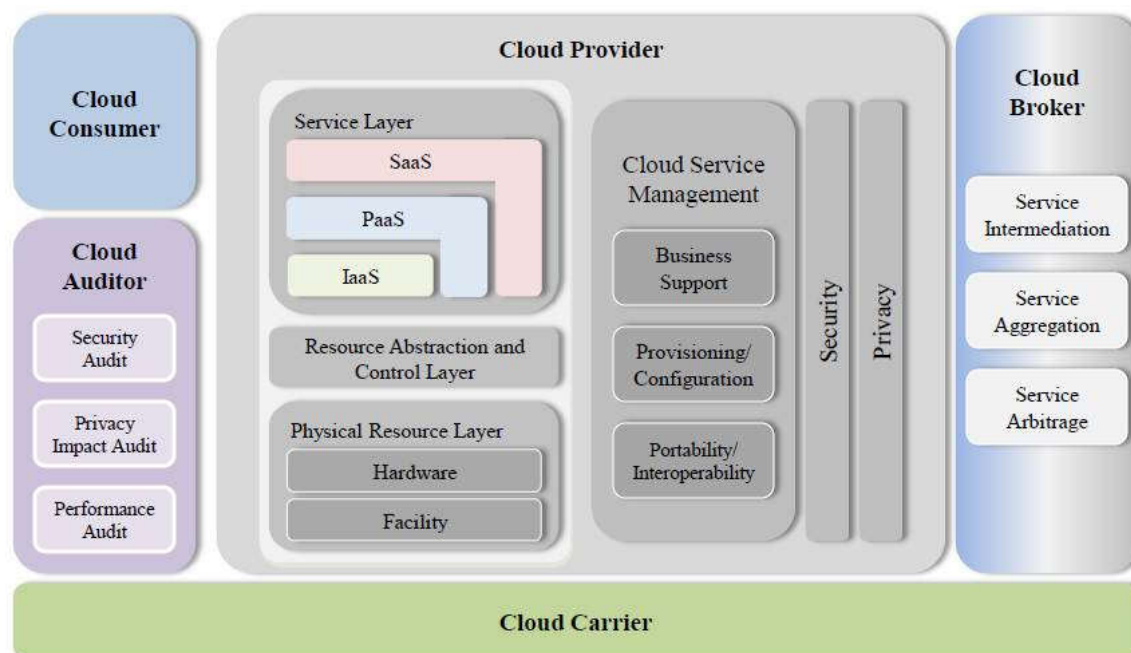
تمام این مزیت ها استفاده از رایانش ابری را در بخش ذخیره سازی و تحلیل و پردازش کلان داده ها را ضروری می سازد. سازمان ملی استانداردها و تکنولوژی امریکا (NIST) رایانش ابری را به صورت زیر تعریف کرده است.

رایانش ابری مدلی است برای فراهم کردن دسترسی آسان بر اساس تقاضای کاربر از طریق شبکه به مجموعه ای از منابع رایانشی قابل تغییر و پیکربندی (مانند شبکه ها، سرورها، فضای ذخیره سازی، برنامه های کاربردی و سرویس ها که این دسترسی بتواند با کمترین نیاز به مدیریت منابع و یا نیاز به دخالت مستقیم فراهم کننده سرویس به سرعت فراهم شده یا آزاد گردد.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

این سازمان همچنین برای رایانش ابری پنج مشخصه را معرفی کرده است که در ادامه به بررسی این مشخصه‌ها می‌پردازیم. شکل ۲ معماری ارائه شده توسط این سازمان برای رایانش ابری را نشان می‌دهد.

۱. ارائه خدمات بر اساس تقاضای مشتری، به صورت سلف سرویس: قابلیت ارائه سرویس‌ها بر اساس نیاز مشتری بدون نظارت نیروی انسانی
۲. دسترسی گسترده از طریق شبکه: قابلیت دسترسی به سرویس‌ها از طریق شبکه در هر لحظه
۳. اشتراک منابع: به اشتراک گذاری منابع و قابلیت ارائه سرویس به صورت مشترک از طریق منابع به مشتریان مختلف
۴. قابلیت ارتجاع سریع: مقیاس‌پذیری سریع منابع و قابلیت ارائه سرویس بر حسب نیاز مشتری، به صورتی که کاربر تصور می‌کند منابع بینهایت هستند
۵. قابلیت محاسبه میزان مصرف و تضمین کیفیت سرویس: قابلیت اندازه‌گیری سرویس‌ها بر اساس میزان استفاده و سطح کیفی سرویس وجود دارد.



شکل ۲: معماری رایانش ابری ارائه شده توسط سازمان ملی استاندارد و تکنولوژی آمریکا [4]



نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## ۶-۲ انواع سرویس های رایانش ابری

رایانش ابری سرویس های متنوعی در سطوح مختلف را در اختیار می گذارد که در این بخش به معرفی این خدمات می پردازیم.

### • زیرساخت به عنوان یک سرویس

پایه ای ترین خدمات زیرساخت مانند تهیه ی سرور، پردازنده، فضای ذخیره سازی و دیگر منابع براساس نیاز در این مدل ارائه می شود. این سرویس امکانی است که برای مشتری فراهم آوری شده به گونه ای که مشتری می تواند نرم افزار دلخواه خود را قرار داده و اجرا کند. مشتری زیرساخت ابری زیرین را مدیریت یا کنترل نمی کند ولی بر روی سیستم های عامل، فضای ذخیره سازی، برنامه های قرارداده شده و احتمالاً گزینش اجزا شبکه بندی (مثل: دیوارهای آتش، متعادل کننده بار) کنترل دارد.

### • سکو به عنوان یک سرویس

این خدمات شامل نرم افزار و سرویس هایی است که به کاربران اجازه می دهد با استفاده از ابزارهای عرضه شده توسط ارائه دهنده، برنامه های کاربردی و نرم افزار ایجاد کنند. این خدمات می تواند شامل ویژگی های از پیش پیکربندی شده ای باشد که مشترکین می توانند به عضویت آن در بیایند و از آن ها استفاده کنند.

### • نرم افزار بعنوان یک سرویس

این نوع رایانش ابری، خدمتی را ارائه می نماید که در آن کاربران می توانند از یک نرم افزار از طریق مرورگر خود و از بستر اینترنت استفاده نمایند. از دیدگاه کاربران سرمایه گذاری جهت خرید نرم افزار و سخت افزار و پشتیبانی آن توجیه ندارد. آنها تنها از امکانات نرم افزار استفاده می نمایند و مطابق منابع مصرف کرده هزینه می پردازند.

خدمات این گروه تنوع زیادی دارند؛ چرا که بیشتر خدمات ارائه شده از یک برنامه کاربردی می تواند تحت عنوان خدمات نرم افزاری در اختیار کاربر قرار گیرد. در این روش کاربر با یک حساب کاربری و بر بستر اینترنت می تواند از نرم افزار استفاده کند. همچنین به روزرسانی و نگهداری این نرم افزارها به عهده ارائه دهنده خدمت است.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

## ۲-۷ مدل های به کارگیری رایانش ابری

در این بخش به معرفی انواع روش های به کارگیری رایانش ابری از دید کاربران می پردازیم. شکل ۳ به طور کلی این مدل ها را به همراه انواع سرویس ها و مزایای رایانش ابری نمایش می دهد.

### • ابر عمومی

در این روش، مالکیت زیرساخت و منابع محاسباتی در اختیار یک شرکت است و از طریق شبکه ی عمومی، خدمات خود را ارائه می کند. در این نوع ابر، برنامه های مشتریان مختلف در سرورها و سیستم های ذخیره ساز و شبکه ابری در کنار یکدیگر قرار دارند.

### • ابر خصوصی

در این روش زیرساخت خدمات در داخل سازمان قرار دارد. به بیان دیگر، ابر خصوصی به معنی طراحی یک زیرساخت محاسباتی با بکارگیری تکنولوژی های مجازی سازی و راه اندازی مرکز داده محلی داخل خود سازمان است.

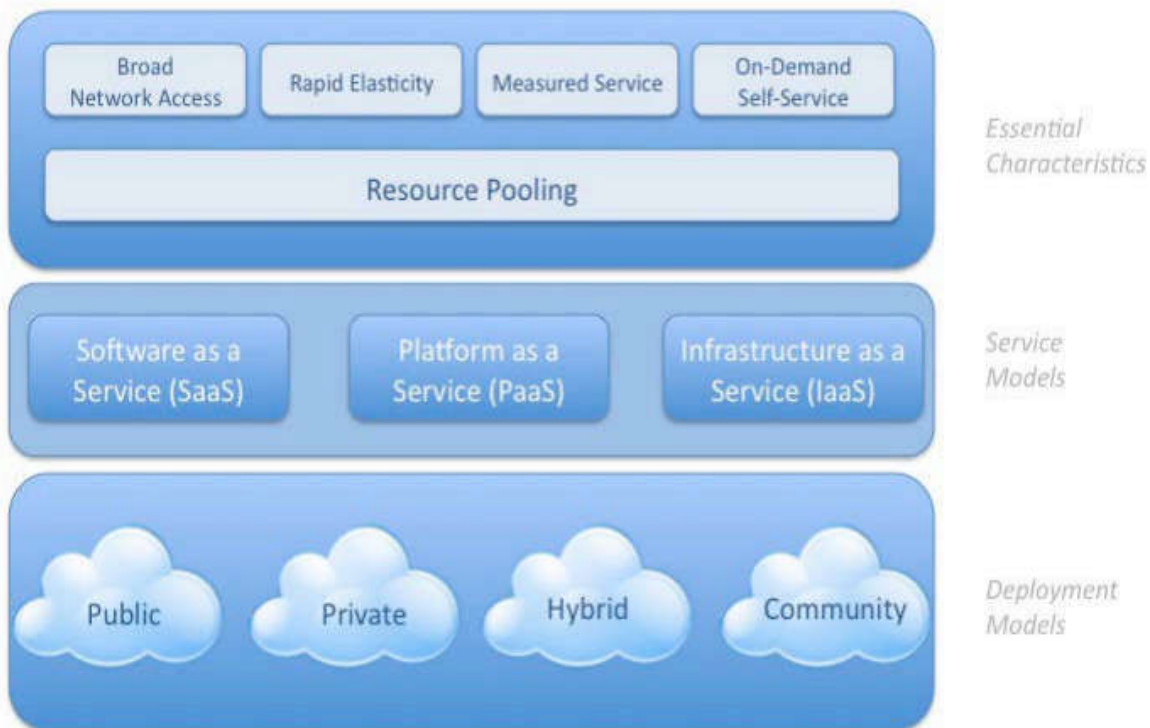
### • ابر گروهی

زیرساخت ابری که بین چند سازمان دارای اهداف و ماموریت های مشترک، به اشتراک گذاشته شده می شود.

### • ابر ترکیبی

ترکیب مدل های خصوصی، گروهی و عمومی مدل ترکیبی نامیده می شود.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--



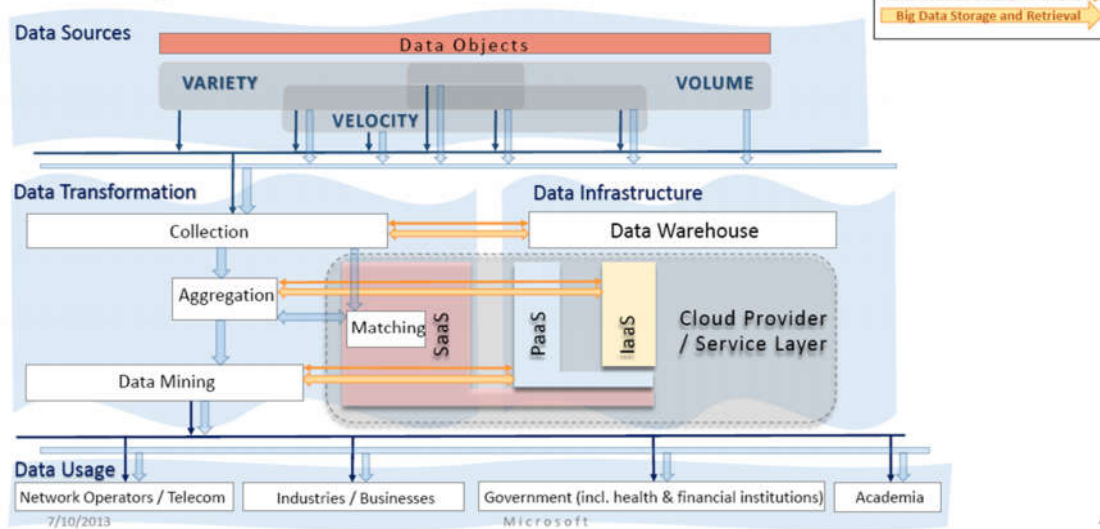
شکل ۳: انواع سرویس، مشخصه ها و پیاده سازی رایانش ابری

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

### ۳ ارتباط رایانش ابری با کلان داده ها

سازمان ملی استانداردها و فناوری (NIST) یک معماری نمونه برای مورد استفاده رایانش ابری در اکوسیستم کلان داده ها ارائه داده است. شکل ۴ این معماری را نمایش می دهد.

#### An Example of Cloud Computing Usage in Big Data Ecosystem



شکل ۴: معماری نمونه استفاده رایانش ابری در اکوسیستم کلان داده ها [۵]

بر طبق این معماری که روند حرکت داده از تولید تا استفاده را نمایش می دهد دو بستر زیرساخت داده و تبدیل داده وجود دارد. بر اساس این دو بستر و لایه سرویس ابر که در بخشهای قبلی به آن اشاره شد، خدمات جمع آوری، ذخیره سازی و محاسبه داده ها به عنوان سرویس از ابر دریافت می شود. در بخش ذخیره سازی داده ها از سرویس سکوی کاری بهره می برد و برای زیرساخت ذخیره سازی داده از سرویس زیرساخت و سکوی ابری استفاده می کند. در بخش تبدیل و پردازش داده نیز این مثال از سرویس برنامه کاربردی ابر استفاده می کند. لازم به ذکر است این یک نمونه از کاربرد و ارتباط رایانش ابری در حوزه کلان داده ها می باشد و برای مثال قسمت پردازش و تبدیل داده می تواند از سرویس سکوی کاری بهره بگیرد. توضیحات تکمیل تر پیرامون استفاده رایانش ابری در اکوسیستم کلان داده ها در ادامه گزارش ارائه شده است.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

### ۱-۳ گونه شناسی کلان داده از دیدگاه ابر

همانطور که در بخش بالا توضیح داده شد لایه ارتباطی بین کلان داده ها و رایانش ابری در سطح سرویس است پس ما برای نمایش گونه شناسی مورد نظر باید کلان داده ها و رایانش ابری را از دیدگاه سرویس مورد بررسی قرار دهیم. کلان داده ها را از دیدگاه ارائه کننده سرویس می توان به ۵ دسته زیر تقسیم کرد:

- محاسبات
- ذخیره سازی
- پردازش
- تحلیل
- بصری سازی

که تعریف و شرح هر سرویس در بخش مفاهیم اساسی آمده است. رایانش ابری طبق استاندارد NIST از لحاظ

سرویس به ۳ دسته تقسیم می شود :

- زیرساخت
- سکو
- نرم افزار

که تعاریف و شرح هر سرویس در بخش مفاهیم اساسی آمده است. در ادامه به بررسی ارتباط بین سرویس های

کلان داده ها و رایانش ابری می پردازیم.

در معماری سرویس های رایانش ابری سلسله مراتبی وجود دارد که در لایه زیرین باید سرویس زیرساخت قرار گیرد و بر روی آن سرویس سکو و بر روی آن سرویس نرم افزار قرار گیرد و این ترتیب اجتناب ناپذیر است یعنی اگر زیرساخت وجود نداشته باشد قادر به ارائه هیچ یک از سرویس های دیگر نمی باشیم.

در معماری کلان داده ها سرویس ها طبق استاندارد خاصی دسته بندی نشده است و مانند رایانش ابری معماری

سرویس های آن ارائه نشده است و دسته بندی ما براساس سرویس های موجود در کلان داده ها می باشد.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

همانطور که پیشتر اشاره شد رایانش ابری و کلان داده ها در سطح سرویس باهم ارتباط دارند و می توان با مشخص کردن سرویس ها ارتباط آنها را با یکدیگر مشخص نمود.

**سرویس محاسبات.** سرویس محاسبات در کلان داده ها سرویسی است که سرورها برای انجام کارهای مربوط به کلان داده ها بهبود داده شده اند (چه از جنبه پیکربندی و چه از جنبه سخت افزاری). ارایه این سرویس بر روی ابر در سطح سرویس زیرساخت می باشد.

**سرویس ذخیره سازی.** سرویس ذخیره سازی در کلان داده ها سرویسی است که به ما امکان می دهد که داده ها را مستقل از درگیری با زیرساخت، ذخیره سازی کنیم و این دسترسی از طریق API و SDK امکان پذیر می باشد. سرویس های ذخیره سازی خود به ۳ دسته کلی ذخیره سازی شیء (object)، بلاک (block) و فایل (file) دسته بندی می شوند که در کلان داده ها بیشتر از ذخیره سازی شیء استفاده می شود. این سرویس در لایه سکو سرویس های ابری قرار می گیرد.

در خدمات سرویس های ذخیره سازی در کلان داده ها، سرویسی به نام ذخیره سازی warehouse وجود دارد که چون زیرساخت آن با ابر متفاوت است در اینجا مورد بررسی قرار نمی گیرد.

همچنین در سرویس ذخیره سازی می توان به سرویس های پایگاه داده هم اشاره نمود که در آن به کاربر امکان استفاده از پایگاه داده را به صورت سرویس ابری می دهد. نکته ای که در سرویس های ابری باید مورد توجه قرار گیرد و پیشتر هم اشاره شد این است که تا وقتی سرویس زیرساخت وجود نداشته باشد نمی توان سرویس سکو ارائه نمود و از طرف دیگر هر سرویس زیرساخت نیز نمی تواند هر نوع سرویس سکو را ارائه دهد. به عنوان نمونه در مورد سرویس ذخیره سازی شیء که در کلان داده ها کاربرد دارد نمی توان از هر ارائه دهنده زیرساخت سرویس گرفت و باید ارائه دهنده زیرساخت این قابلیت را در زیر ساخت ایجاد نماید.

**سرویس پردازش.** سرویس پردازش در کلان داده ها به این معنا است که بتوان فرایندهایی که در کلان داده ها مورد استفاده قرار می گیرد را از طریق روش هایی مانند موازی سازی، توزیع شدگی و غیره تسریع بخشید. این سرویس در معماری ابر در لایه سکو قرار می گیرد.

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

**سرویس تحلیل.** سرویس تحلیل در کلان داده ها به ارائه روش های تحلیل و پردازش داده می پردازد و در لایه سرویس سکو ابر قرار می گیرد.

**سرویس بصری سازی.** سرویس بصری سازی در کلان داده ها نیز به ارائه سرویسی برای مشخص کردن ارتباط و نمایش داده ها می پردازد. این سرویس نیز در لایه سرویس سکو قرار می گیرد.

## ۳-۲ بررسی سرویس های موجود کلان داده ها در ابر

در این قسمت از مستند به معرفی سرویس های ارائه شده ابری داده در شرکت های معتبر دنیا پرداخته می شود.

### ۳-۲-۱ بررسی سرویس های ذخیره سازی

در این بخش به بررسی سرویس های ذخیره سازی در چرخه حیات داده پرداخته می شود.

#### ۳-۲-۱-۱ سرویس S3 شرکت آمازون [6]

در این سرویس، ذخیره سازی به صورت ذخیره سازی شیء صورت می گیرد و بر روی ابر می باشد. در کنار این سرویس خدماتی از قبیل پشتیبان گیری، بازیابی و مهاجرت داده صورت می گیرد و برای مشتری قابلیت data lake را نیز فراهم می سازد. طریقه ارتباط با سرویس از طریق SDK و API می باشد. البته آمازون سرویس های دیگر ذخیره سازی مانند ذخیره سازی فایل، ذخیره سازی بلاک و غیره را نیز ارائه می دهد ولی چون کاربرد کمتری در کلان داده ها دارد مورد بررسی قرار نمی گیرند. آمازون سرویسی از پایگاه داده های مختلف از قبیل NOSQL و پایگاه داده های رابطه ای ارائه می دهد.

#### ۳-۲-۱-۲ سرویس ذخیره سازی [9] Microsoft Azure

این سرویس هم همانند سرویس S3 بر روی ابر قرار دارد. در این سرویس داده ها به صورت شیء و فایل ذخیره می شوند و ارسال و دسترسی به آن از طریق API می باشد. از قابلیت های دیگری که در کنار این سرویس می باشد محدودیت دسترسی های برنامه ها به داده می باشد. این شرکت سرویس های پایگاه داده نیز ارائه می دهد که شامل انواع پایگاه داده های NOSQL و رابطه ای می باشد.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

### ۳-۲-۱-۳ سرویس ذخیره سازی گوگل [7]

این سرویس هم همانند سرویس های دیگر بر روی ابر قرار دارد و داده را به صورت شیء ذخیره می کند. از قابلیت هایی که در این سرویس بیان شده است مدیریت دسترسی ها به داده می باشد. این شرکت سرویس های پایگاه داده از دسته پایگاه داده های NOSQL ارائه می دهد.

### ۳-۲-۱-۴ سرویس ذخیره سازی Oracle [10]

این سرویس هم همانند سرویس های دیگر بر روی ابر قرار دارد و داده را به صورت شیء ذخیره می کند و دسترسی به آن از طریق API است. شرکت اوراکل انواع پایگاه داده را به عنوان سرویس بر روی ابر ارائه می دهد.

### ۳-۲-۱-۵ سرویس ذخیره سازی EMC [8]

این سرویس هم همانند سرویس های دیگر بر روی ابر قرار دارد، داده را به صورت شیء ذخیره می کند و تمام قابلیت های موجود ذخیره سازی شیء را دارا می باشد. این شرکت انواع سرویس های پایگاه داده را نیز ارائه می دهد.

### ۳-۲-۲ بررسی سرویس های محاسباتی

در این قسمت به بررسی سرویس های محاسباتی که در کلان داده ها به کار گرفته می شود و بر روی ابر سرویس می دهند، پرداخته می شود.

#### ۳-۲-۲-۱ سرویس محاسباتی شرکت آمازون

شرکت آمازون با بهره گیری از تعادل بار و گذاشتن یک تسریع دهنده در سرور به نام سرویس EC2 بهبود سرعت در عملیات محاسبات را ارائه نموده است. سایر شرکت ها سرویس محاسباتی مانند آمازون ارائه نمی دهند اما محصولاتی در راستای بهبود محاسبات دارند. مانند شرکت EMC که سروری را برای محاسبات کلان داده ها معرفی کرده است.

#### ۳-۲-۳ بررسی سرویس های پردازشی

در این قسمت به بررسی سرویس های پردازشی که در کلان داده ها به کار گرفته می شود و بر روی ابر سرویس می دهند، پرداخته می شود.



کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

### ۳-۲-۳-۱ سرویس های پردازشی شرکت آمازون

این شرکت سرویس های مختلف پردازشی بر روی ابر ارائه می دهد و یکی از پرکاربردترین آنها سرویس EMR می باشد که در آن معماری هدوپ را به عنوان یک سرویس ارائه می دهد. البته شرکت آمازون سرویس های پردازشی دیگری نیز مانند elasticsearch و جستجوی تعاملی ارائه می نماید.

### ۳-۲-۳-۲ سرویس های پردازشی شرکت Microsoft Azure

سرویس پردازشی ارائه شده توسط این شرکت بیشتر به پردازش های دسته ای اختصاص دارد و از طریق API ارائه می شود.

### ۳-۲-۳-۳ سرویس های محاسباتی شرکت گوگل

نوع سرویس های پردازشی ارائه شده در این شرکت شامل پردازش های دسته ای و جاری (جریان) و نیز شامل زیرساخت هایی مانند هدوپ می باشد که روی ابر ارائه شده است.

### ۳-۲-۴ بررسی سرویس های تحلیل و بصری سازی داده

چون در ارائه سرویس، بیشتر شرکت ها سرویس مشترکی بین تحلیل و بصری سازی داده ارائه داده اند، در این گزارش هر دو سرویس با هم بررسی می شوند.

### ۳-۲-۴-۱ سرویس های تحلیل و بصری سازی شرکت آمازون

سرویس آمازون QuickSight یک سرویس تحلیل گر ابری بسیار سریع است که قابلیت انجام آنالیزهای تک کاره و بصری سازی را دارا می باشد. با این سرویس امکان تحلیل های پیشرفته، بصری سازی بسیار خوبی را دارا خواهیم بود. این سرویس یک داشبورد کامل در اختیار کاربر می گذارد که قابل دسترسی از هر نوع مرورگر و دستگاه موبایل می باشد. سرویس دیگری توسط آمازون به نام یادگیری ماشین ارائه می شود که در آن به توسعه دهندگان در تمام سطوح از توانایی های فنی اجازه می دهد تا از یادگیری ماشین استفاده کنند. این سرویس ابزاری برای بصری سازی و رابطی که کمک می کند بدون دانش در باره الگوریتم های پیچیده یادگیری ماشین از آن استفاده کنند، ارائه

نام گزارش: کاربردهای رایانش ابری در کلان داده ها	وضعیت: نهایی	کد
--	--------------	----

می دهد. این سرویس قابلیت مقیاس پذیری بالایی دارد. سرویس های دیگری هم توسط آمازون برای پردازش صوت و تصویر ارائه می شود.

### ۳-۲-۴-۲ سرویس های تحلیل و بصری ساز oracle

یک پلتفرم واحد برای درخواست های تک کاره و تحلیل ها می باشد. همچنین بصری سازی داده ها برای ایستگاه کاری یا دستگاه موبایل را نیز انجام می دهد. این سرویس برای اکتشاف و تبدیل داده هایی که در هدوپ وجود دارند می باشد. این سرویس قابلیت بصری سازی داده های موجود در هدوپ و امکان دسترسی مستقیم به هدوپ را نیز می دهد.

### ۳-۲-۴-۳ سرویس های تحلیل و بصری ساز گوگل

این سرویس برای اکتشاف، تحلیل و بصری سازی داده ها می باشد. که با google cloud machine learning و big query مجتمع می شود تا یک دسترسی آسان به سرویس های داده های پردازشی مهم را به کاربر ارائه دهد.

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

## مراجع

- [1] <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf>
- [2] <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6.pdf>
- [3] <https://www.nist.gov/baldrige/real-challenge-big-data>
- [4] [https://bigdatawg.nist.gov/\\_uploadfiles/M0450\\_v1\\_3857254727.pdf](https://bigdatawg.nist.gov/_uploadfiles/M0450_v1_3857254727.pdf)
- [5] [https://bigdatawg.nist.gov/\\_uploadfiles/M0017\\_v2\\_4175727836.pdf](https://bigdatawg.nist.gov/_uploadfiles/M0017_v2_4175727836.pdf)
- [6] <https://aws.amazon.com/>
- [7] <https://cloud.google.com/solutions/big-data/>
- [8] <https://www.emc.com/en-us/big-data/solutions.htm>
- [9] <https://azure.microsoft.com/en-us/solutions/big-data/>
- [10] <https://cloud.oracle.com/bigdata>

## واژه نامه

### انگلیسی به فارسی

واژه انگلیسی	برگردان فارسی
Analytics	تحلیل
Availability	دسترس پذیری
up Back	نسخه پشتیبان
Band width	پهنای باند
Collection	جمع آوری
Configuration	پیکربندی
consumer	مصرف کننده
Cooling	خنک سازی

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

Data transformation	تبدیل داده
Elasticity	ارتجاع پذیری
Expired	منقضی
Fault tolerance	تحمل پذیری خطا
Infrastructure as a service	زیرساخت به عنوان سرویس
Life cycle	چرخه حیات
Load balancer	متعادل کننده بار
National Institute of Standards and Technology - NIST	سازمان ملی استانداردها و تکنولوژی امریکا
On demand	بر اساس تقاضا
Pattern	الگو
Performance	کارایی
Platform as a service	سکو به عنوان سرویس
Private Cloud	ابر خصوصی
Producer	تولید کننده
Public cloud	ابر عمومی
Real-time	بلادرنگ
Recommendation	پیشنهاد
Reliability	قابلیت اطمینان
Scalability	مقیاس پذیری
Scalable	مقیاس پذیر
Social Media	رسانه اجتماعی
Software as a service	برنامه کاربردی به عنوان سرویس

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

بصری سازی	Visualization
-----------	---------------

جدول واژه نامه انگلیسی به فارسی

### فارسی به انگلیسی

واژه فارسی	برگردان انگلیسی
ابر خصوصی	Private Cloud
ابر عمومی	Public cloud
ارتجاع پذیری	Elasticity
بر اساس تقاضا	On demand
برنامه کاربردی به عنوان سرویس	Software as a service
بصری سازی	Visualization
بلادرنگ	Real-time
پهنای باند	Band width
پیشنهاد	Recommendation
پیکربندی	Configuration
تبدیل داده	Data transformation
تحلیل	Analytics
تحمل پذیری خطا	Fault tolerance
تولید کننده	Producer
جمع آوری	Collection
چرخه حیات	Life cycle
خنک سازی	Cooling
دسترسی پذیری	Availability

کد	وضعیت: نهایی	نام گزارش: کاربردهای رایانش ابری در کلان داده ها
----	--------------	--

Social Media	رسانه اجتماعی
Infrastructure as a service	زیرساخت به عنوان سرویس
National Institute of Standards and Technology - NIST	سازمان ملی استانداردها و تکنولوژی امریکا
Platform as a service	سکو به عنوان سرویس
Reliability	قابلیت اطمینان
Performance	کارایی
Pattern	الگو
Load balancer	متعادل کننده بار
consumer	مصرف کننده
Scalability	مقیاس پذیری
Scalable	مقیاس پذیر
Expired	منقضی
up Back	نسخه پشتیبان

جدول ۲ واژه نامه فارسی به انگلیسی