



# کلان داده در حوزه سلامت دکتر محمد ستاری

دکترای مهندسی کامپیوتر - نرم افزار از دانشگاه اصفهان

عضو هیات علمی گروه مدیریت و فناوری اطلاعات سلامت دانشگاه علوم پزشکی اصفهان

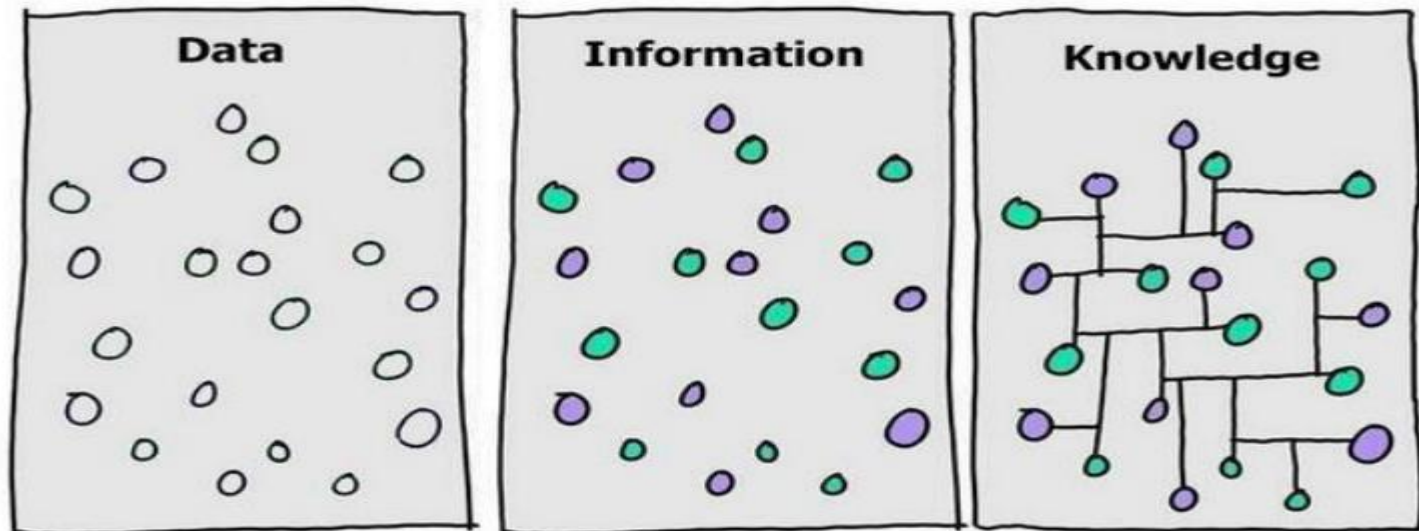
عضو مرکز تحقیقات فناوری اطلاعات در امور سلامت

# داده، اطلاعات و دانش

داده عبارت است از یک یا چند کاراکتر (اعداد، حروف و نشانه ها) که هیچ معنای خاصی ندارد

اطلاعات، داده های پردازش شده و حقایقی با مفهوم هستند که به توصیف و تعریف داده ها می پردازند و توسط گیرنده پیام درک و تفسیر می شوند.

دانش وسیع تر، عمیق تر و غنی تر از داده ها و اطلاعات است، دانش مخلوطی سیال از تجربیات، ارزش ها، اطلاعات موجود و نگرش های کارشناسی نظام یافته است که بعنوان چار چوبی برای ارزشیابی و بهره مندی از تجربیات و اطلاعات جدید بکار برده می شود.





# انواع داده

داده بدون ساختار ▶

داده نیمه ساخت یافته ▶

داده ساخت یافته ▶



## داده بدون ساختار

- ▶ داده های بدون ساختار بر اساس یک مدل داده از پیش تعیین شده، تعریف نشده اند
- ▶ نمی توان آنها را در یک پایگاه داده سنتی رابطه ای ذخیره کرد.
- ▶ متن و چند رسانه ای دو نوع رایج محتوای بدون ساختار هستند.



## داده بدون ساختار

- ▶ از ۸۰٪ تا ۹۰٪ داده‌های تولید و جمع‌آوری شده توسط سازمان‌ها بدون ساختار هستند و حجم آن‌ها به سرعت در حال رشد است - چندین برابر سریع‌تر از نرخ رشد برای پایگاه‌های داده ساختاریافته.
- ▶ داده بدون ساختار حاوی اطلاعات فراوانی هستند که می‌توانند برای هدایت تصمیمات تجاری مورد استفاده قرار گیرند. با این حال، تجزیه و تحلیل داده‌های بدون ساختار بسیار دشوار بوده است. با کمک هوش مصنوعی و یادگیری ماشین، می‌توان تجزیه و تحلیل این نوع از داده‌ها را انجام داد.



## داده بدون ساختار

- ▶ فایل های متنی: فایل های ورد، ایمیل، شرح حال بیماران
- ▶ رسانه های اجتماعی و وبسایت ها: داده های شبکه های اجتماعی مانند توئیتر، لینکدین، و فیس بوک و وبسایت هایی مانند سایت های اشتراک گذاری، یوتیوب.
- ▶ داده های تلفن همراه و ارتباطات: پیام های متنی و چت
- ▶ چندرسانه ای: عکس های دیجیتال، فایل های صوتی، ویدئو



## داده ساختاریافته

- ▶ داده های ساختاریافته داده هایی هستند که در یک مخزن قالب بندی شده، معمولاً یک پایگاه داده، سازماندهی شده اند تا بتوان عناصر آن را برای پردازش و تجزیه و تحلیل موثرتر آدرس دهی کرد.
- ▶ ساختار داده نوعی مخزن است که اطلاعات را برای این منظور سازماندهی می کند. یک زبان پرس و جو پایگاه داده، مانند SQL زبان پرس و جو استاندارد به مدیر پایگاه داده اجازه می دهد تا با پایگاه داده تعامل داشته باشد.



## داده نیمه ساختاریافته

- ▶ داده های نیمه ساختار یافته در بین داده های ساخت یافته و بدون ساختار قرار دارد.
- ▶ یک سند **Word** به طور کلی داده های بدون ساختار در نظر گرفته می شود. با این حال، می توانید تگ های فراداده را در قالب کلمات کلیدی و سایر ابرداده هایی که محتوای سند را نشان می دهند، اضافه کنید و وقتی افراد آن عبارات را جستجو می کنند، یافتن آن سند را آسان تر می کند در این حالت داده ها اکنون نیمه ساختار یافته هستند.
- ▶ داده ی **XML** نمونه ی دیگر از داده های نیمه ساخت یافته است

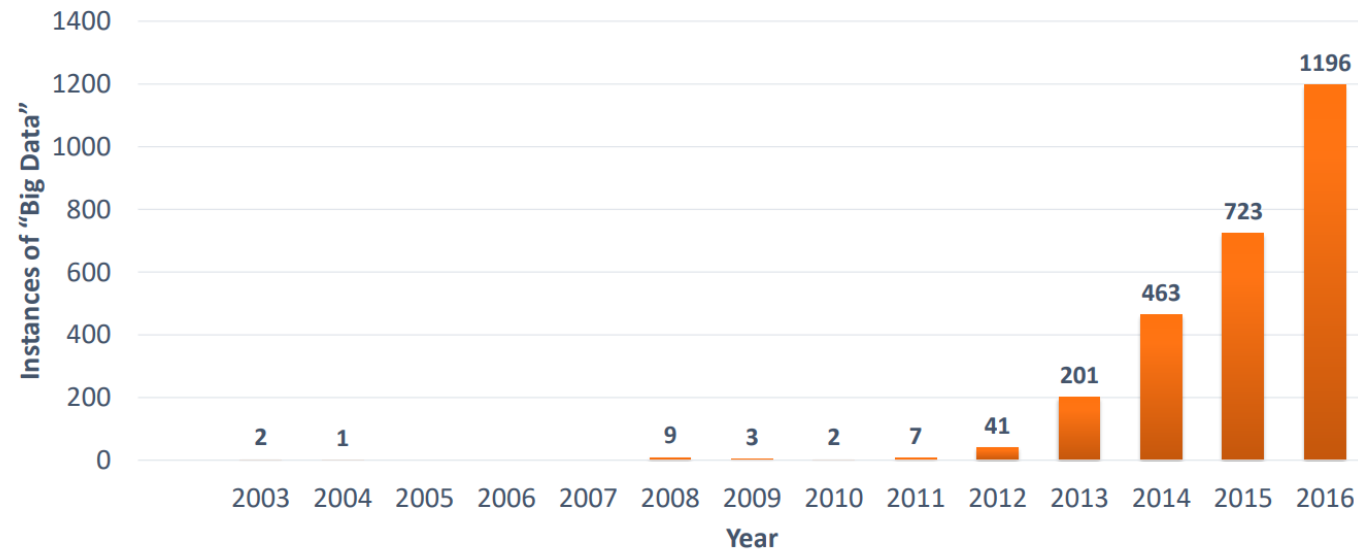


# معرفی کلان داده



کلان داده اصطلاحی است که برای حجم زیادی از داده‌ها استفاده می‌شود که خیلی سریع و پیچیده هستند و پردازش آنها با استفاده از روش‌های سنتی، مشکل است.

## “Big Data” on PubMed





حجم یا Volume :

اولین ویژگی بیگ دیتا حجم و مقدار آن است که با کلمه **Volume** شناخته می شود. سازمان ها و نهادهای گوناگون با جمع آوری اطلاعات از منابع متنوع و متفاوت اقدام به ایجاد یک پایگاه داده می کنند. این اطلاعات می تواند شامل مواردی مثل اطلاعات مربوط به حجم معاملات و تبادلات تجاری، اطلاعات دریافت شده از شبکه های اجتماعی یا اطلاعات ماشینی و طبقه بندی شده دیگر باشد. حجم زیاد داده ها، چالش هایی را ایجاد می نماید که نیاز به سیستم های جدید برای مدیریت این نوع داده ها مطرح می شود.



bit	b	binary digit, a single 0 or 1
byte	B	8 bits
kilobyte	KB	1,024 bytes
megabyte	MB	1,024 KB
gigabyte	GB	1,024 MB
terabyte	TB	1,024 GB
petabyte	PB	1,024 TB
exabyte	EB	1,024 PB
zettabyte	ZB	1,024 EB
yottabyte	YB	1,024 ZB



چه حجمی از داده در بیمارستان های مختلف شهر اصفهان ذخیره می شود؟ ▶

# ویژگی های اصلی کلان داده



سرعت یا Velocity

دومین ویژگی مهم کلان داده بحث سرعت فرآیندهای آن است. مبادلات اطلاعات باید با سرعت بسیار بالایی انجام شود

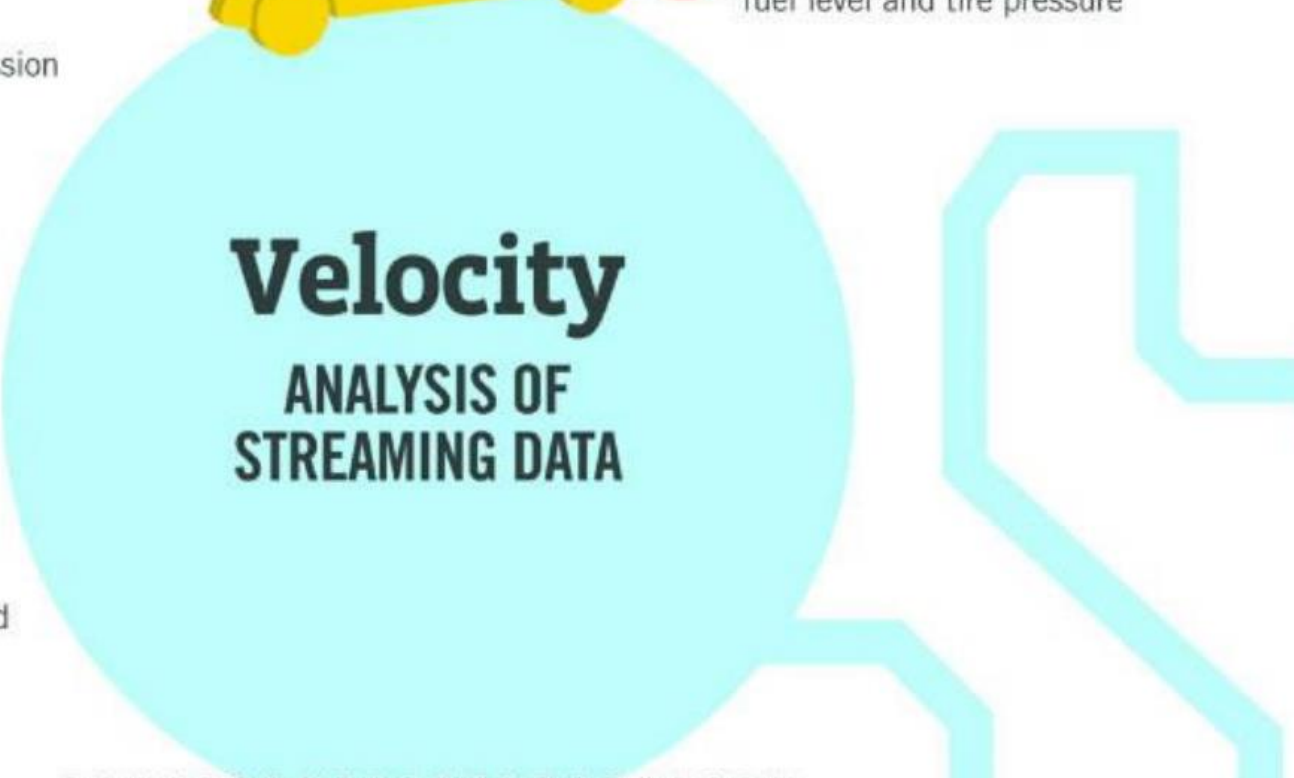
تولید داده‌ها با سرعتی بی سابقه انجام می‌شود و باید به موقع پردازش شود.



The New York Stock Exchange captures  
**1 TB OF TRADE INFORMATION**  
during each trading session



Modern cars have close to  
**100 SENSORS**  
that monitor items such as  
fuel level and tire pressure



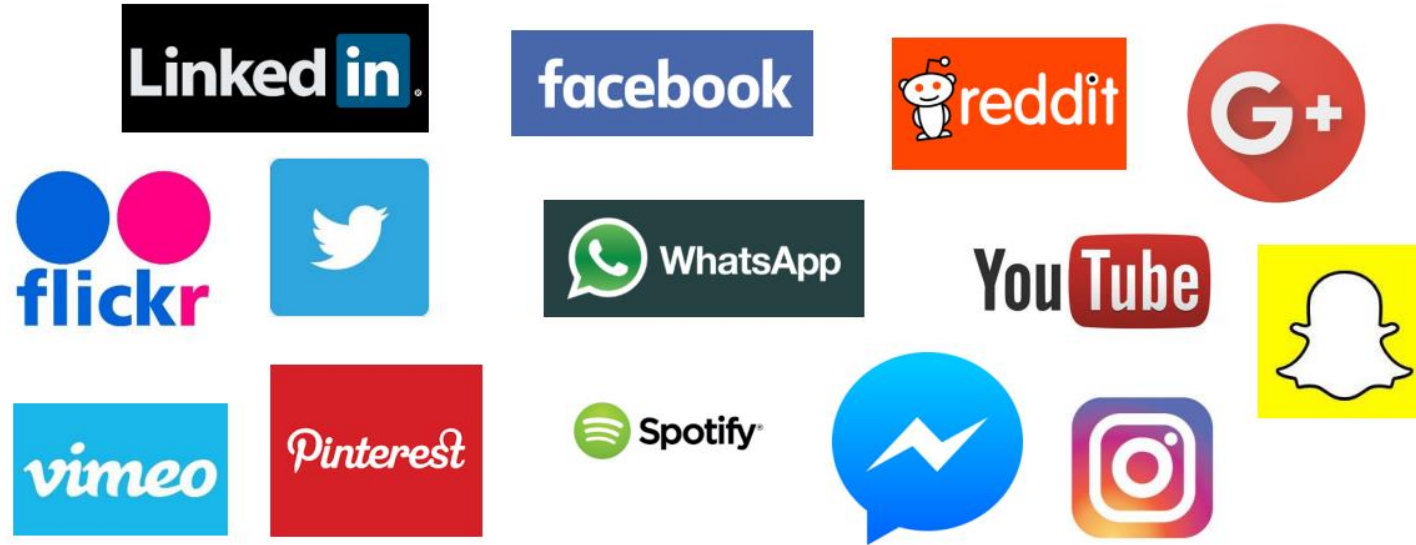
By 2016, it is projected  
there will be  
**18.9 BILLION NETWORK CONNECTIONS**  
– almost 2.5 connections  
per person on earth





تنوع یا Variety :

فرمت و نوع اطلاعات با یکدیگر تفاوت دارند. این اطلاعات می توانند به صورت متنی، تصویری، صوتی یا حتی کدهای صفر و یک باشند! این تنوع باعث شده تا دسته بندی های متعددی برای جمع آوری و ذخیره سازی اطلاعات در نظر گرفته شود

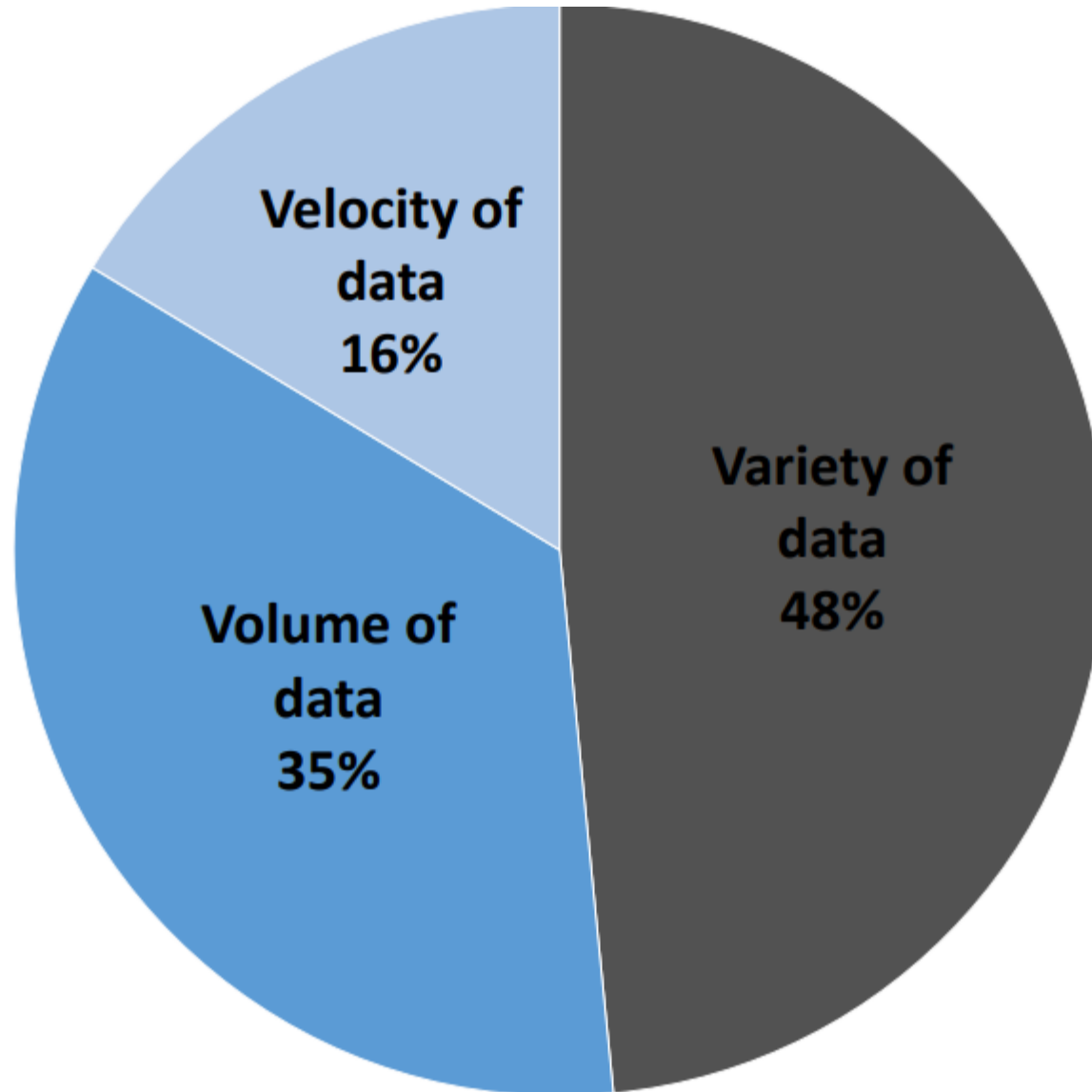


تنوع یا Variety :



# میزان اهمیت هر یک از V ها

17 / 52





چه تنوعی از داده در بیمارستان های مختلف شهر اصفهان ذخیره می شود؟ ▶

# دو ویژگی دیگر کلان داده



ارزش: مهم‌ترین  $V$  از دیدگاه کسب‌وکار، ارزش کلان داده‌ها معمولاً از کشف و شناسایی الگو ناشی می‌شود که منجر به عملیات مؤثرتر، روابط قوی‌تر با مشتری و سایر مزایای تجاری واضح و قابل سنجش می‌شود.

صحت: صحت، در واقع به چه میزان داده‌ها با واقعیت سازگار هستند. میزان اعتماد به کلان داده را تعیین می‌نماید.



## پیچیدگی یا Complexity:

ویژگی دیگری به موضوع پیچیدگی کلان داده مربوط می شود. از جایی که اطلاعات از منابع متعدد و گوناگونی به دست می آیند، فرآیند مرتبط سازی، تطابق و هماهنگ سازی آنها با یکدیگر پیچیده می شود. این مسئله باعث شده تا خیلی از سازمانها با مشکلات زیادی رو به رو شوند. برای جلوگیری از این مسئله باید سیستم های جامع و یکپارچه ای ترتیب داد که مانع ایجاد این مسائل شود.



## صرفه جویی در هزینه:

برخی از ابزارهای کلان داده می‌توانند مزایای هزینه‌ای را برای کسب و کار به همراه داشته باشند بدین صورت که مقادیر زیادی داده ذخیره شود و این ابزارها به شناسایی روشهای کارآمدتر کسب و کار نیز کمک می‌کنند.

## کاهش زمان:

سرعت بالای ابزاری مانند تجزیه و تحلیل می‌تواند به راحتی منابع جدید داده را شناسایی کند که به مجموعه‌ها در تجزیه و تحلیل سریع داده‌ها و تصمیم‌گیری سریع بر اساس آموخته‌ها کمک می‌کند.



## درک شرایط بازار:

با تجزیه و تحلیل کلان داده‌ها می‌توانید درک بهتری از شرایط فعلی بازار داشته باشید. به عنوان مثال، با تجزیه و تحلیل رفتارهای خرید مشتریان، یک شرکت می‌تواند محصولی را که بیشترین فروش را دارد پیدا کند و مطابق این روند محصول تولید کند. با این کار می‌تواند از رقبای خود پیشی بگیرد.



## چه سازمان هایی می توانند از کلان داده استفاده نمایند

سازمان های بهداشتی

بانکداری

سیستم های آموزشی

شرکت های تجاری

بیمه ها



## کاربردهای کلان داده در حوزه سلامت

▶ تیم تحقیقاتی سلامت قلب و عروق در مراقبت‌های سرپایی CANHEART، یک ابتکار تحقیقاتی مشاهده‌ای مبتنی بر جمعیت منحصر به فرد با هدف اندازه‌گیری و بهبود سلامت قلبی عروقی و کیفیت مراقبت‌های قلبی عروقی ارائه نموده است.

▶ این شامل داده های ۹.۸ میلیون بزرگسال انتاریو بود.

▶ داده‌ها با پیوند دادن پایگاه‌های اطلاعاتی متعدد، مانند نظرسنجی‌های الکترونیکی، مدیریت بهداشت، پایگاه‌های داده‌های بالینی، آزمایشگاهی، دارو و پرونده الکترونیک پزشکی با استفاده از شناسه‌های شخصی رمزگذاری شده جمع‌آوری شدند.





## کاربردهای کلان داده در حوزه سلامت

پیام های ارسال شده در رسانه های اجتماعی می تواند برای غربالگری و تشخیص احتمالی افسردگی استفاده شود. تجزیه و تحلیل آنها بر اساس تحقیقات قبلی در مورد ارتباط بین اختلالات افسردگی و افکار تکراری است

ابزارهای تجزیه و تحلیل داده های بزرگ با استخراج الگوهای رفتاری و احساسی پنهان در پیام ها یا «توئیت ها» که در توییتر ارسال می شوند، الگوی احساسی مرتبط با بیماری را می توانند شناسایی نمایند



## کاربردهای کلان داده در حوزه سلامت

افراد با سه یا بیشتر از این عوامل خطر سندرم متابولیک دارند: مقاومت به انسولین، فشار خون بالا، تری گلیسیرید بالا، HDL پایین

بیش از یک نفر از هر سه بزرگسال در ایالات متحده به سندرم متابولیک مبتلا هستند سندرم متابولیک خطر بیماری قلبی، سکته مغزی و دیابت را افزایش می دهد. درمان آن هزینه بر است.

منابع داده ای

پایگاه داده ایتنا از ۳۷ هزار نفر

جمعیت شناختی (سن، شاخص توده بدنی و غیره)

تست آزمایشگاهی و نتایج مهندسی بیومتریک

پلتفرم GNS Healthcare Analytics برای پیش بینی دقیق نتایج مداخله



## زیرساخت های سخت افزاری کلان داده

زیرساخت سخت افزاری کلان داده (بیگ دیتا)، وظیفه ذخیره سازی فیزیکی اطلاعات جمع آوری شده را دارا است. زیرساخت ذخیره سازی می تواند از ابعاد مختلف بررسی شود. ابزارهای ذخیره سازی بر اساس فناوری هایی که از آن بهره می برند، قابل دسته بندی هستند. به طور مثال، رم، آرایه ای از دیسک ها، دیسک های مغناطیسی و حافظه های فلش از جمله فناوری های ذخیره سازی هستند



## زیرساخت های سخت افزاری کلان داده

زیرساخت ذخیره سازی از دید معماری شبکه

**DAS:** نوعی فضای ذخیره سازی است که بدون استفاده از شبکه و به صورت مستقیم به کامپیوتر متصل می شود.

Direct-attached storage

**NAS:** ذخیره سازی در سطح فایل، شامل چندین دیسک است که به شبکه ای از رایانه ها متصل شده و وظیفه ذخیره سازی داده ها را بر عهده دارد.

Network-attached storage

**SAN:** شبکه هایی اختصاصی هستند که ذخیره سازی در سطح بالا را برای گروهی از کامپیوترها فراهم می کنند.

Storage Area Network



## انباره داده

- ▶ انباره داده به مجموعه‌ای از داده‌ها گفته می‌شود که از منابع مختلف اطلاعاتی سازمان جمع‌آوری، دسته‌بندی و ذخیره می‌شود.
- ▶ انباره داده یک مجموعه از داده‌های که پشتیبان فرآیندهای تصمیم‌سازی مدیریتی است.



## خصوصیات اصلی انبار داده

### موضوع‌گرا Subject Oriented

انبار داده برای پاسخ‌گویی به پرسش‌های مختلف در مورد یک موضوع خاص بهینه‌سازی می‌شود. مثلاً یک فروشگاه می‌خواهد رفتار خریداران خود را بررسی کند. پس موضوع در این جا رفتار خریداران است. باید داده‌های مرتبط با رفتار مشتریان را در خود نگاه دارد و این داده‌ها را به منظور استخراج مفاهیم و نتایج خاصی به شکلی ویژه سازماندهی کند.

### مجتمع Integrated

در سیستم‌های مختلف داده‌ها ممکن است از جنبه‌های مختلفی با هم نامتناسب باشند در واقع انبار داده به صورت تکه تکه در جاهای مختلف نیست و به صورت مجتمع در یک منبع ذخیره شده است. مثلاً منابع داده در کشورهای مختلف با زمان و تاریخ‌های مختلف ذخیره شده‌اند. با توجه به مختلف بودن منابع انبار داده، قبل از ذخیره سازی آن‌ها در انبار داده برای تأمین یکپارچگی تکنیک‌های پاکسازی داده‌ها و مجتمع‌سازی به کار می‌رود.



## خصوصیات اصلی انبار داده

### نگهدار زمان‌های مختلف Time varying

هر ساختار کلیدی در انبار داده شامل عنصر زمان یا همان مهر زمانی است این مهر زمانی به عنوان کلید به سایر جداول عمل می‌کند. به این معنا که در انبار داده، داده‌های قدیمی‌تر نیز ذخیره می‌شوند. مثلاً برای بدست آوردن اطلاعات فروش ۱ ماه اخیر، می‌توانیم انبار داده را جستجو نماییم. سیستم انبار داده داده‌ها هرگز به روزرسانی نمی‌شوند. بلکه داده‌های جدید و یا تغییر یافته با مهرهای زمانی جدید به انبار اضافه می‌شوند.

### غیر فرار none-volatile

داده‌های انبار داده همیشه از لحاظ فیزیکی مجزا هستند ، و هیچگاه نیازی به تغییر و به به روز رسانی نخواهند داشت. به این معنا که اگر داده‌ای در انبار داده ثبت شد، دیگر امکان تغییر آن وجود ندارد. برای مثال در یک پایگاه داده عادی (غیر انبار داده‌ای) اگر یک شخص آدرس خود را عوض کرد، آدرس جدید جایگزین آدرس قبلی می‌شود، ولی در انبار داده سابقه آدرس‌های شخص ذخیره می‌شود و تغییرات آدرس یک شخص به صورت جدیدتر در انبار داده ثبت می‌شود و آدرس‌های قبلی نیز در انبار داده جهت بررسی موجود می‌باشد.



# ETL

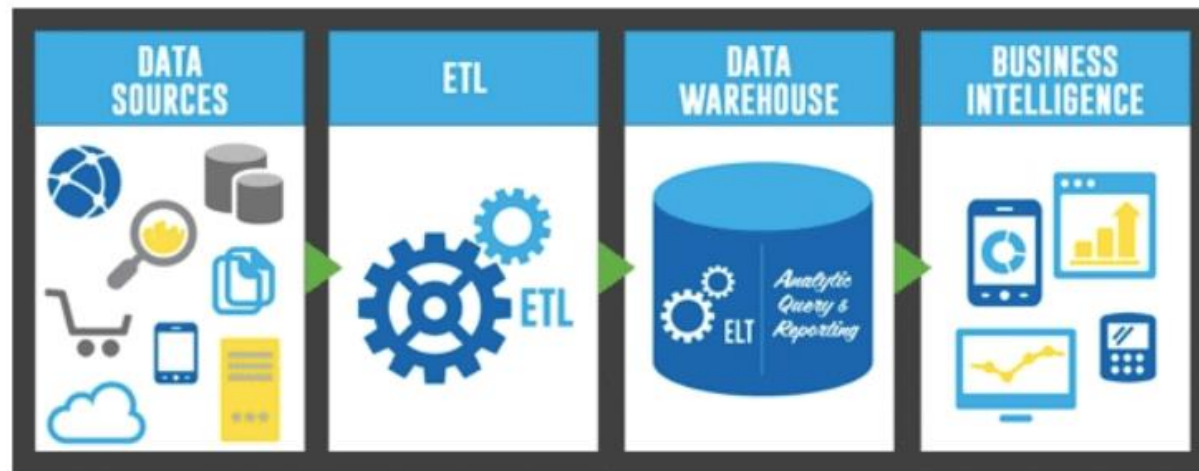
ETL فرایند واکنشی اطلاعات از سیستم های اطلاعاتی، تبدیل و بارگذاری در انبار داده است. برای انبار داده، داده های مفیدی که ارزش تحلیلی دارند را از سیستم های اطلاعاتی مختلف استخراج می شوند. این داده های خام نیاز به پالایش دارند و باید ساختار آنها به گونه ای که با انبار داده و تحلیل کسب و کار متناسب باشد، تغییر کند. این تغییر شامل تغییر نوع و مقدار و ترکیب و یا جداسازی داده است.





# ETL

با توجه به اینکه امکان اشتباه بودن یا نامعتبر بودن برخی داده‌ها در سیستم‌های اطلاعاتی وجود دارد، باید داده‌های استخراج شده را در طی فرایندی پالایش **Data Cleaning** نمود و اطلاعاتی که باعث بروز اشتباه می‌شوند را حذف و یا تغییر داد. در این فاز بررسی صحت داده‌ها (**Data Validation**) و مقایسه با سیستم‌های عملیاتی صورت می‌پذیرد. پس از این مراحل، داده‌ها در انبار داده بارگذاری می‌شوند.





## تجزیه و تحلیل کلان داده

### تحلیل پیش‌بین Predictive Analytics

تحلیل پیش بین فناوری است که از تجربه (داده) می‌آموزد رفتار آینده افراد را به منظور اتخاذ تصمیمات بهتر پیش‌بینی کند (منظور از افراد الزاما انسان‌ها نیست. بنابراین، الزاما رفتار انسانی پیش‌بینی نمی‌شود و امکان انجام انواع پیش‌بینی‌ها بر اساس داده‌های کنونی موجود وجود دارد).

تحلیل‌های پیش‌بین از مدل‌های "پیش‌بین predictive models" استفاده می‌کنند. مدل پیش‌بین، مکانیزمی است که رفتار یک فرد مانند کلیک، خرید، دروغ یا مرگ را پیش‌بینی می‌کند. این مدل، ویژگی‌های (مشخصه‌ها) افراد (موجودیت‌ها) را به عنوان ورودی دریافت کرده و یک امتیاز پیش‌بینی به عنوان خروجی فراهم می‌کند. هرچه امتیاز پیش‌بینی بالاتر باشد، احتمال آنکه آن موجودیت رفتار پیش‌بینی شده را از خود بروز دهد، بیشتر است.



## سوال

چند نمونه از تحلیل پیش بین در سلامت را نام ببرید؟



## تجزیه و تحلیل کلان داده

### تحلیل توصیفی Descriptive Analytics

این نوع از تحلیل همان گونه که از نام آن بر می آید ذاتا توصیفی است. تحلیل های توصیفی داده ها را خلاصه سازی کرده و کمتر بر جزئیات دقیق هر بخش از اطلاعات تمرکز می کند و در عوض بر روایت کلی متمرکز می شود.

### تحلیل تجویزی Prescriptive Analytics

تحلیل های تجویزی عموما به دنبال جنسی از پیش بینی هستند که بر اساس خروجی مدل پیش بین بتوان اقداماتی را تجویز کرد.



## تجزیه و تحلیل کلان داده

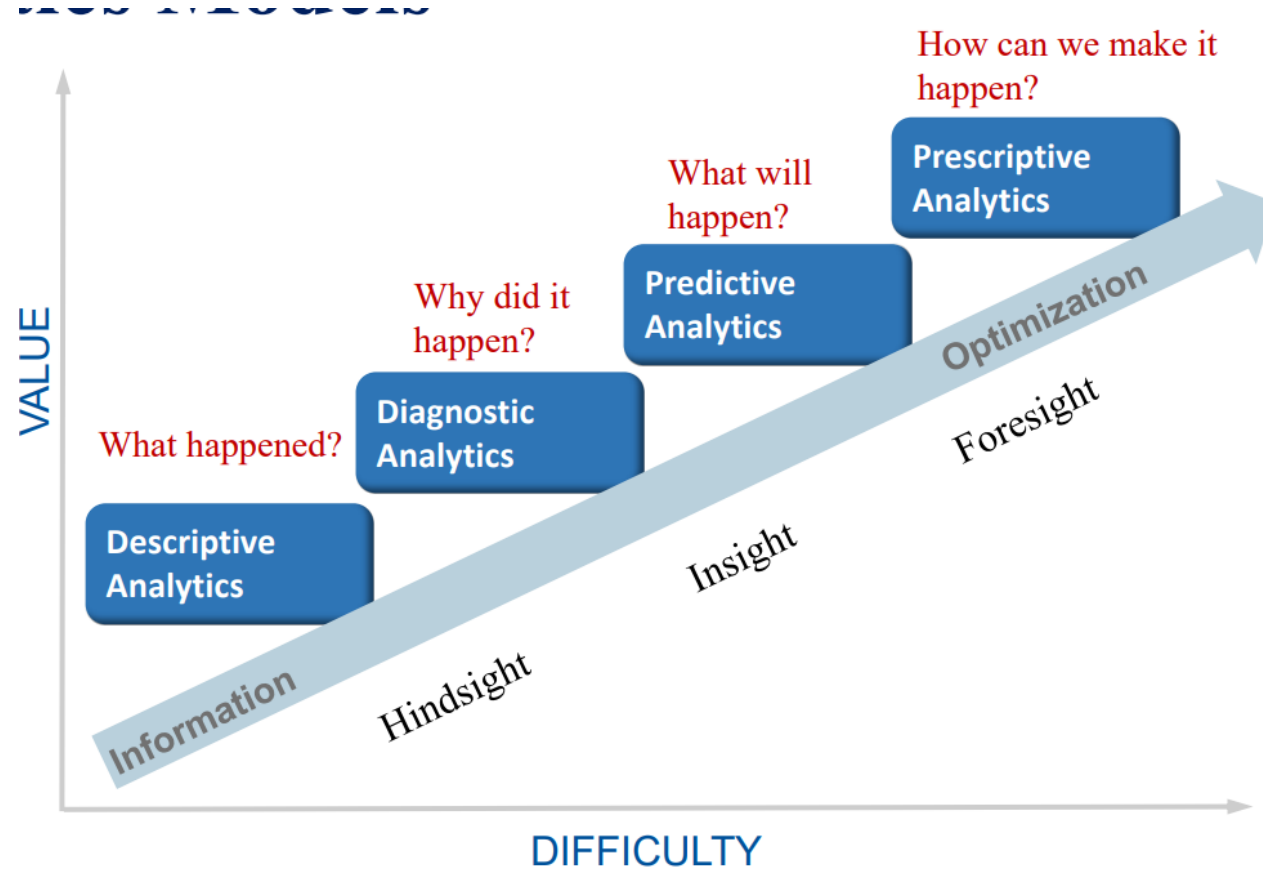
### Diagnostic predictive

تجزیه و تحلیل تشخیصی به پاسخ به سؤالات در مورد چرایی رویدادها کمک می کند.

تکنیک های تجزیه و تحلیل تشخیصی، تجزیه و تحلیل های توصیفی اولیه را تکمیل می کنند و از یافته های تجزیه و تحلیل توصیفی برای کشف علت این رویدادها استفاده می کنند.



# تجزیه و تحلیل کلان داده





## تکنولوژی مدیریت داده در کلان داده

Hadoop یک چارچوب نرم افزاری منبع باز برای ذخیره سازی داده ها و اجرای برنامه ها است. هادوپ ذخیره سازی گسترده برای هر نوع داده، قدرت پردازش عظیم و توانایی انجام وظایف یا کارهای همزمان بی حد و حصر را فراهم می کند.

هادوپ یک چارچوب نرم افزاری منبع باز است که پردازش توزیع شده داده های بزرگ را بر روی خوشه هایی از سرورها ممکن می سازد. برای انجام پردازش توزیع شده بر روی هزاران ماشین با تحمل پذیری خطای بالا طراحی شده است. به جای تکیه بر سخت افزارهای گران قیمت، تحمل پذیری در مقابل خطا در این خوشه ها از توانایی نرم افزاری در تشخیص و اداره ی خرابی در لایه کاربرد می آید. استفاده کنندگان برجسته هادوپ، فیسبوک و یاهو هستند.



# تکنولوژی مدیریت داده در کلان داده

## معماری هادوپ







## اهمیت هادوپ

- ▶ امکان ذخیره و پردازش مقادیر عظیم هر نوع داده به سرعت: با افزایش حجم و انواع مختلف داده ها، به خصوص از طریق رسانه های اجتماعی و اینترنت اشیاء این یک نکته مهم است.
- ▶ قدرت پردازش: مدل محاسباتی توزیع شده هادوپ داده های بزرگ را به سرعت پردازش می کند. هرچه گره محاسباتی بیشتر استفاده کنید، قدرت پردازش بیشتری نیز خواهید داشت.
- ▶ میزان خطا: پردازش داده ها و برنامه ها در برابر خرابی سخت افزار محافظت می شوند. اگر یک گره از کار بیفتد، کارها به طور خودکار به گره های دیگر هدایت می شوند تا مطمئن شوید محاسبات توزیع شده از کار نمی افتند. همچنین چندین نسخه از تمام داده ها بطور خودکار ذخیره می شوند.
- ▶ .



## اهمیت هادوپ

▶ **انعطاف پذیری:** برخلاف بانک های اطلاعاتی سنتی، نیازی به پردازش داده ها قبل از ذخیره کردن آن نیست. شما می توانید اطلاعات را به همان اندازه که می خواهید ذخیره کنید و تصمیم بگیرید که بعداً چگونه از آنها استفاده کنید. این موضوع شامل داده های بدون ساختار مانند متن، تصاویر و فیلم ها است.

▶ **مقیاس پذیری:** به راحتی می توانید سیستم خود را برای مدیریت داده های بیشتر، به سادگی با افزودن گره ها رشد دهید.



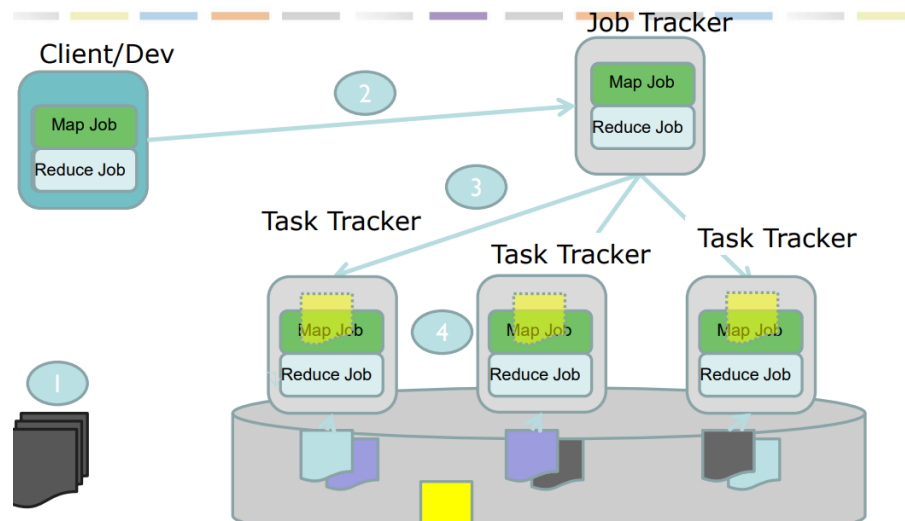
## مدل های برنامه نویسی برای کلان داده

- ▶ نگاهت - کاهش Map-Reduce
- ▶ روشی برای توزیع یک کار در چندین گره
- ▶ هر گره داده های ذخیره شده اش را پردازش می کند
- ▶ از دو مرحله ایجاد شده توسط توسعه دهنده تشکیل شده است
- ▶ ۱. نگاهت ۲. کاهش



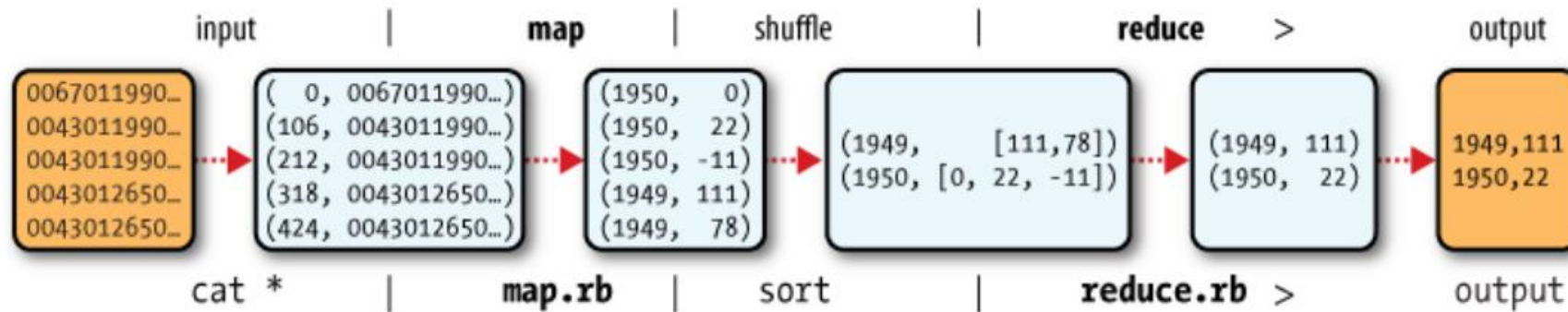
## مدل های برنامه نویسی برای کلان داده

- ▶ یک کاربر یک برنامه کلاینت را بر روی کامپیوتر مشتری اجرا می کند
- ▶ برنامه مشتری job را به Hadoop ارسال می کند
- ▶ job به فرآیند JobTracker در گره اصلی ارسال می شود
- ▶ هر Slave Node فرآیندی به نام TaskTracker را اجرا می کند
- ▶ JobTracker به TaskTrackers دستور می دهد تا وظایف را اجرا و نظارت کنند





# عملیات Map/Reduce

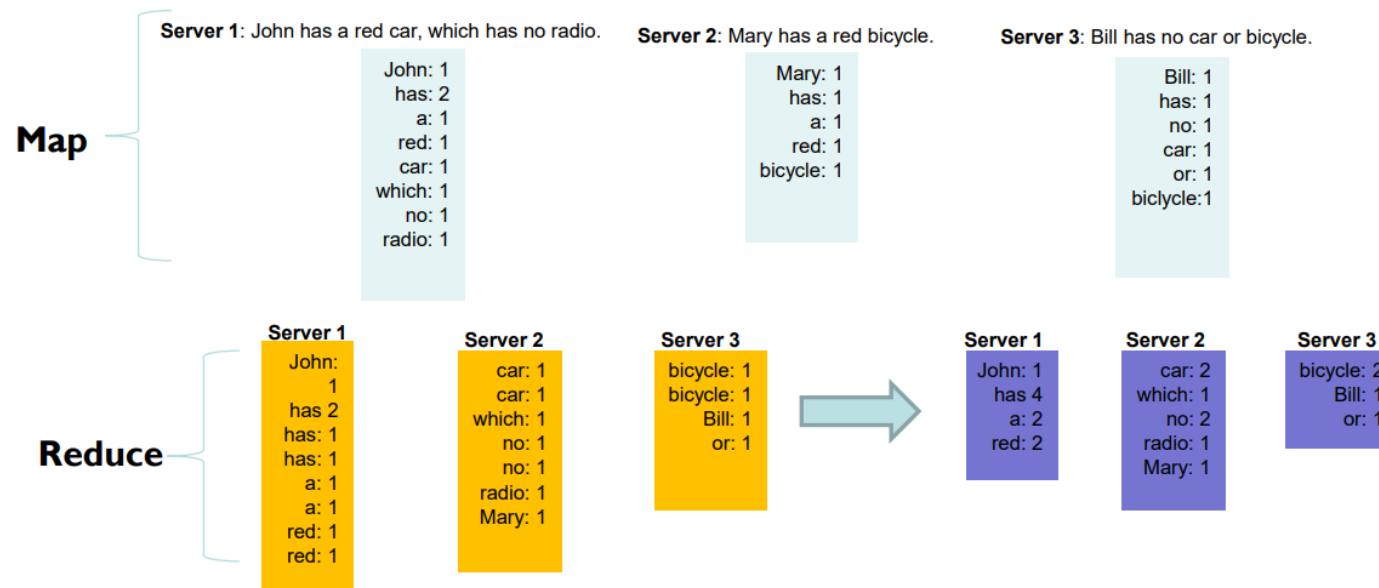




# مثال از Map/Reduce

Education

**Problem:** Count the number of times that each word appears in the following paragraph:  
John has a red car, which has no radio. Mary has a red bicycle. Bill has no car or bicycle.





## زبان برنامه نویسی کلان داده

زبان پایتون: **python** یکی از زبان‌های چندمنظوره شی گراست که در سال ۱۹۹۱ برای اولین بار طراحی و منتشر شد. یکی از مهم‌ترین دلایل ایجاد زبان برنامه‌نویسی پایتون، خوانایی بالای کدها است. پایتون به صورتی طراحی شده است که برنامه نویس می‌تواند به صورت کاملاً واضح و بدون ابهام پروژه‌های برنامه‌نویسی بزرگ و کوچک ایجاد کند.

حوزه تحلیل داده و علم داده نیز شاهد استفاده گسترده از پایتون است. در حال حاضر یعنی زمانی که ما هر روز داده‌های بیشتری تولید می‌کنیم، پلتفرم‌های کارآمد مانند پایتون برای تجزیه و تحلیل و مدیریت داده‌ها تقاضای زیادی دارند.



## استراتژی کلان داده در سازمان

استراتژی کلان داده سازمانی به چشم‌انداز جامع یک سازمان در مورد استفاده از قابلیت‌ها و منابع وابسته به داده و منابع مستقل از داده اشاره دارد

یکی از کاربردهای اصلی تجزیه و تحلیل بیگ دیتاها برای هوش تجاری است که به منظور بهبود قابلیت‌های تصمیم‌گیری، تصمیم‌گیری سریع‌تر، درک نیازهای مشتری، توسعه استراتژی برای راه‌اندازی محصولات و خدمات جدید، کاوش در بازارهای جدید، بهبود گردش موجودی، کاهش شکایات مشتریان، و افزایش بهره‌وری و کارایی کارکنان استفاده می‌شود.





## مثال از کلان داده در سازمان

والمارت هر ساعت بیش از ۱ میلیون تراکنش مشتری را انجام می دهد که به پایگاه های داده ای وارد می شود که تخمین زده می شود حاوی بیش از ۲,۵ پتابایت داده باشد - معادل ۱۶۷ برابر اطلاعات موجود در تمام کتاب های کتابخانه کنگره ایالات متحده.

سیستم تشخیص تقلب در کارت اعتباری FICO از ۲,۱ میلیارد حساب فعال در سراسر جهان محافظت می کند

بر اساس برآوردها، حجم داده های تجاری در سراسر جهان، در همه شرکت ها، هر ۱,۲ سال دو برابر می شود



## اهمیت کلان داده در سازمان

شرکت ها به طور فزاینده ای به دنبال یافتن بینش عملی در مورد داده های خود هستند. بسیاری از پروژه های کلان داده از نیاز به پاسخگویی به سوالات خاص کسب و کار سرچشمه می گیرند. با وجود پلتفرم های مناسب تجزیه و تحلیل کلان داده، یک شرکت می تواند فروش را افزایش دهد، کارایی را افزایش دهد و عملیات، خدمات مشتری و مدیریت ریسک را بهبود بخشد. □

شرکت **QuinStreet, Webopedia**، از ۵۴۰ تصمیم گیرنده سازمانی درگیر در خرید کلان داده ها نظرسنجی نمود تا بداند شرکت ها در کدام حوزه های تجاری قصد دارند از تجزیه و تحلیل داده های بزرگ برای بهبود عملیات استفاده کنند. حدود نیمی از پاسخ دهندگان گفتند که از تجزیه و تحلیل داده های بزرگ برای بهبود حفظ مشتری، کمک به توسعه محصول و کسب مزیت رقابتی استفاده می کنند. □

نکته قابل توجه، حوزه کسب و کار که بیشترین توجه را به خود جلب می کند به افزایش کارایی و بهینه سازی عملیات مربوط می شود. به طور خاص، ۶۲ درصد از پاسخ دهندگان گفتند که از تجزیه و تحلیل داده های بزرگ برای بهبود سرعت و کاهش پیچیدگی استفاده می کنند..



## نمونه کاربرد کلان داده در سازمان

- ▶ پیش بینی رفتار مشتری.
- ▶ پیش بینی فروش.
- ▶ تجزیه و تحلیل داده های دستگاه هایی مانند سیستم های POS و سنسورهای کارخانه
- ▶ پیش بینی ریسک های مالی.
- ▶ شناسایی فعالیت ها یا سوابق متقلبانه.
- ▶ تجزیه و تحلیل نظرات شبکه های اجتماعی، تعامل با مشتری و معیارهای بازاریابی آنلاین.
- ▶ جستجوی همبستگی در چندین منبع داده متفاوت.



با تشکر از همراهی شما  
؟