5th Conference on Recent Research in Science and Technology
پنجمین همایش پژوهش های نوین در علوم و فناوری
بهار ۱۳۹۵ Spring 2016

Application of big data to fight crime

Mohammad khalili and seyedeh afagh navaei fijani

Department of E-Learning, Shiraz University, Shiraz, Iran.

* ***Corresponding author:mohammad khalili*** **.** Tel: +989173151985 E-mail: mohammadkhalili45@gmail.com

Abstract

Big Data is a massive volume of data that can't be processed easily using traditional techniques. In fact, big data analysis refers to the capability of managing a huge volume of heterogeneous data at the right speed, and with suitable speed to allow real-time processing. The world of Big Data offers a new set of challenges and obstacles. The main challenge in this area is analyzing, processing and searching in large volume of data for the purpose of sharing, storage, and transfer of data in a secure manner. In short, the term Big Data applies to the case of information that can't be processed or analyzed using traditional processes or tools. Increasingly, organizations today are facing more and more Big Data challenges. Big Data analysis and the prediction of future outcomes based on the results of analysis obtained have found applications in business, finance, healthcare, etc. Similarly Big Data also have recently found its way to fight crime. Big companies like IBM, Dell, Google, Microsoft invest billions of dollars in research and development of Big Data technologies. The developers have their own communities for technologies like Hadoop, NoSQL, Real time, machine learning, etc. Much success has been achieved already in terms of developing and deploying technologies which help combat crimes. Big data will increasingly inform everyday policing and cyber security efforts. Law enforcement of all kinds, state and local, are making use of big data practices to pinpoint potential crime hot spots or predict houses that could be burglarized. Besides cutting back on violent crimes and robberies, big data is also being used to identify white collar criminals who are involved in financial crimes such as insurance frauds, insider trading, money laundering and healthcare fraud. Cyber attacks of all kinds are on the rise. One way to defend against these attacks is to use big data to become aware of cyber attacks and to find vulnerabilities to defend against an attack. Researchers in computational science, information systems, social sciences, engineering, medicine, and many other fields have been called upon to help enhance our ability to fight violence, terrorism, cyber crimes, and other cyber security concerns. In this seminar, we tried to introduce Big Data as well as analyzing the application of Big Data and its revolution to combat crimes.

Archive of SID

www.SID.ir

کاربرد ابر داده در مبارزه با جرم و جنایت

محمد خلیلی و سیده افاق نوایی فیجانی

دانشکده آموزشهای الکترونیکی، دانشگاه شیراز، شیراز ،ایران
نویسنده مکاتبه کننده:محمد خلیلی تلفن 00989173151985 ایمیل:mohammadkhalili45@gmail.com

چکیده فارسی:

Big data یا ابر داده به مجموعه ای از داده‌های بزرگ و پیچیده اطلاق می‌شود که پردازش آنها توسط نرم افزارهای سنتی پردازش اطلاعات امکان پذیر نیست یا به سختی انجام می‌شود. در واقع ابر داده قابلیت مدیریت حجم عظیمی از اطلاعات مختلف در سرعت مناسب و در چارچوب زماني مناسب است که اجازه ي تجزیه و تحلیل و واکنش زمان واقعی[1] را میدهد. جهان داده هاي حجیم یک مجموعه ي جدیدي از چالش ها و موانع را ارائه مي کند. چالش اصلی در این زمینه شامل تجزیه و تحلیل، پردازش و جستجو در میان داده ها است. همچنین فرآیندهایی مانند ذخیره سازي، به اشتراک گذاري و انتقال داده‌ها از دیگر چالش های امنیتي است که دربحث پردازش داده‌های حجیم با آنها مواجه هستیم. به طور خلاصه، لغت داده ي حجیم براي اطلاعاتی به کار مي رود که نمی توانند با استفاده از فرآیندهای سنتی یا ابزارها تجزیه و تحلیل شوند. به صورت افزایشي، سازمان ها امروزه بیشتر و بیشتر با چالش هاي داده ي حجیم روبرو هستند. تجزیه و تحلیل داده های حجیم و پیش بینی نتایج آینده، بر اساس نتایج به دست آمده از تجزیه و تحلیل برنامه های کاربردی در حوزه هاي کسب و کار، امور مالی، بهداشت و درمان و غیره انجام میگیرد. ابرداده به طور مشابه راهش را در زمینه ي مبارزه با جرم و جنایت نیز پیدا کرده است. شرکت هاي بزرگ همانند IBM, Dell, Google و Microsoft میلیاردها دلار در زمینه ي تحقیقات و توسعه فنآوریِ داده های حجیم سرمایه گذاري کرده اند. توسعه دهندگان، انجمن هاي[2] خودشان را برای فن آوری هایي Hadoop، NoSQL، Real time، machine learning و غیره دارند. در حال حاضر در زمینه توسعه و استقرار فنآوری، دستاوردهاي زیادي به دست آمده است که به پیکار علیه جرم و جنایت کمک میکنند. ابرداده به طور فزاینده اي از حفظ نظم و آرامش هر روزه و تلاش های امنیتی سایبری خبر مي دهد. هر نوع اجراي قانون، اعم از ایالتي و محلي، از روش هاي ابرداده براي اشاره ي دقیق به نقاط بالقوه جرم و جنایت یا پیش بیني خانه هایي که احتمال حمله مسلحانه به آنها مي رود، استفاده مي کند. حملات سایبری از همه نوع در حال افزایش است. در حوزه ي جرم و جنایت، داده هاي حجیم علاوه بر کاهش در جنایات خشونت آمیز و دزدی، به تشخیص مجرمان یقه سفید که در جرایم مالي از جمله تقلب هاي بیمه، تجارت هاي خودي[3] ، پول شویي و تقلب در مراقبت هاي بهداشتي و درماني درگیر هستند نیز کمک مي کنند. حملات سایبری از همه نوع در حال ازدیاد است و یک راه برای مقابله با این حملات استفاده از ابرداده براي آگاهي از حملات اینترنتي و یافتن نقاط آسیب پذیر براي دفاع در برابر این حملات مي باشد. همچنین از محققان در علوم محاسباتی، سیستم هاي اطلاعاتي، علوم اجتماعي، مهندسي، پزشکي و بسیاري از زمینه هاي دیگر برای افزایش توانایی ما در مبارزه با خشونت، تروریسم، جرایم سایبري و دیگر نگرانی های امنیتی سایبری درخواست کمک شده است. در این مقاله سعي بر این است تا ضمن معرفي ابرداده، به تحلیلی از کاربرد داده‌های حجیم و انقلاب آن در مبارزه با جرم و جنایت بپردازیم.

[1]Real-Time
[2]Communities
[3]Insider Trading

Archive of SID

www.SID.ir

کلمات کلیدی (*Keywords*): ابر داده، داده هاي حجيم، جرم و جنايت، جرايم سايبري، مبارزه

Archive of SID

www.SID.ir

معرفی موضوع (تشریح و بیان مسئله) : (Introduction)

1- نگاهی بر تاریخچه مدیریت داده‌ها

مدیریت و تحلیل داده‌ها همیشه بهترین مزایا و چالش‌ها را برای سازمان‌ها در همه اندازه‌ها و در تمامی صنایع عرضه کرده است. کسب و کارها ، ستیزی طولانی بر سر یافتن روشی واقع‌بینانه به منظور دست‌یابی اطلاعات در مورد مشتری‌ها، محصولات و سرویس ها داشته‌اند. زمانی که یک شرکت تنها چند مشتری داشت ، که همگی محصولی یکسان را با شیوه‌ای یکسان خریداری می‌کردند، همه چیز ابتدایی و ساده بود. اما در گذر زمان، شرکت‌ها و بازارهایی که در این رابطه تشکیل شد، پیچیده‌تر گردیدند. این شرکت ها با هدف بدست آوردن برتری در رقابت با رقبای خود به ناچار خطوط تولید محصول خود را گسترش دادند و راه‌های ارائه محصول به مشتریان خود را متحول کردند. کشمکش بر سر داده، تنها به کسب و کار محدود نمی‌شود. به عنوان نمونه، سازمان‌های تحقیق و توسعه، برای کسب قدرت پردازش کافی برای اجرای مدل‌های پیچیده یا پردازش تصاویر و دیگر منابع داده‌ی علمی، چالش داشته‌اند .

در واقع زمانی که ما با داده سرو کار داریم با پیچیدگی‌های بسیاری مواجه می شویم. برخی از داده‌ها ساختاریافته ، در یک پایگاه داده رابطه ای سنتی ذخیره شده‌اند، در حالی‌که سایر داده‌ها شامل اسناد، سوابق خدمات مشتریان، حتی تصاویر و ویدئوها بی‌ساختار هستند. شرکت‌ها می بایستی منبع جدیدی از داده‌های تولیدشده توسط ماشین‌ها از قبیل حس گرها را نیز در نظر بگیرند. سایر منابع اطلاعاتی جدید، مانند داده‌های موجود در شبکه‌های اجتماعی و داده‌های حاصل از تعاملات با وب سایت‌ها، توسط انسان تولید می‌شود. علاوه بر این میزان دسترسی و تطابق با ابزارهای سیار قدرتمند و مدرن، منابع جدیدی را همراه با دسترسی دائمی به شبکه‌های جهانی برای داده ایجاد می‌کنند .

اگرچه هر منبع داده‌ای می‌تواند به صورت مستقل مدیریت و جست‌وجو شود، چالش امروز این است که چگونه شرکت ها اشتراکاتی میان همه این گوناگونی داده‌ها می‌یابند؟ هنگامی‌که با این حجم اطلاعات در اشکال مختلف سر و کار دارید، فکر کردن به مدیریت داده ها به روش سنتی غیرممکن است. اگر بخواهید نسبت به دیروز قیاسی داشته باشید، باید گفت ، که امروزه به شکل معنا داری داده های ما از گسترش و تنوع بیشتری برخوردار می باشند و سازمان‌های درگیر از راهکارهای بیشتری برای استفاده از این اطلاعات نسبت به گذشته برخوردار می باشیم. بنابراین، شما باید در مورد مدیریت داده ها ی دردسترس، با دیدی متفاوت نسبت به گذشته تفکر کنید[1] .

1-1 معرفي ابرداده ها

بیگ دیتا یا ابرداده‌ها در حقیقت بیانگر حجم زیادی از داده می باشد که اندازه آنها فراتر از حدی است که با نرم افزارهای معمول بتوان آنها را در یک زمان معقول اخذ، دقیق سازی، مدیریت و پردازش کرد و آنالیز و دسته بندی آن ها به وسیله زیرساخت ها و دیتابیس های رایج مشکل می باشد. سازمان های بزرگ با انبوه اطلاعات متنوعی چون ایمیل، عکس، ویدیو، گزارشات روزانه، نوشته و اطلاعات روزانه مواجه هستند که می بایست به صورت هوشمندانه ای طبقه بندی گردند تا به سرعت و در زمان کوتاهی قابل بررسی و دستیابی باشند. مفهوم «اندازه» در داده‌های بزرگ بطور مستمر در حال تغییر است و به مرور بزرگتر می‌شود. داده‌های بزرگ (Big Data) مجموعه از تکنیک‌ها و تاکتیک‌هایی است که نیازمند شکل جدیدی از یکپارچگی هستند تا بتوانند ارزش های بزرگی را که در مجموعه‌های بزرگ، وسیع، پیچیده و متنوع داده، پنهان شده‌اند، آشکار سازند. از این رو با رشد روز افزون داده‌ها و نیاز به بهره برداری و تحلیل از این داده‌ها، بکارگیری زیرساخت‌های داده‌های بزرگ از اهمیت ویژه‌ای برخوردار شده است[1] .

www.SID.ir

عبارت Big Data مدت‌ها است که برای اشاره به حجم‌های عظیمی از داده‌ها که توسط سازمان‌های بزرگ، کارخانجات، سازمان‌های دولتی و ... ذخیره و تجزیه و تحلیل می‌شوند مورد استفاده قرار می‌گیرد. اما به تازگی، این عبارت بیشتر برای اشاره به مجموعه‌های داده‌ای بزرگی استفاده می‌شود که به قدری بزرگ و حجیم هستند که با ابزارهای مدیریتی و پایگاه‌هایِ داده سنتی و معمولی قابل مدیریت نیستند. مشکلات اصلی در کار با این نوع داده‌ها مربوط به برداشت و جمع‌آوری، ذخیره‌سازی، جست‌وجو، اشتراک‌گذاری، تحلیل و نمایش آنها است[2] .

این مبحث، به این دلیل هر روز جذابیت و مقبولیت بیشتری پیدا می‌کند که با استفاده از تجزیه و تحلیل حجم‌های بیشتری از داده‌ها، می‌توان تحلیل‌های بهتر و پیشرفته‌تری را برای مقاصد مختلف، از جمله مقاصد تجاری، پزشکی و امنیتی انجام داد و نتایج مناسب‌تری را دریافت‌کرد[3] . بیشتر تحلیل‌های مورد نیاز در پردازش داده‌های بزرگ توسط دانشمندان در علومی مانند هواشناسی، ژنتیک، مراقبت‌های بهداشتی، شبیه‌سازی‌های پیچیده فیزیک، تحقیقات زیست‌شناسی و محیطی، جست‌وجوی اینترنت، تحلیل‌های اقتصادی و مالی و تجاری مورد استفاده قرار می‌گیرد[4] .

برای نمونه یک فروشگاه اینترنتی را در نظر بگیرید که بتواند با بررسی کلیک بازدیدکنندگان بر روی لینک های مختلف وب سایت، به علایق و سلیقه مشتریان پي ببرد و بدین سان کسب و کار خود را بهبود بخشد .یا دولت یک کشور می تواند با رصد شبکه های اجتماعی به مقابله و پیشگیری از ناهنجاری های جامعه بپردازد. نمونه بارز آن موتور جستجوی گوگل است که به دسته بندی و اولویت بندی علایق کاربران در کشورهای مختلف می پردازد و کاربران را در دستیابی به اطلاعات مورد نظرشان راهنمایی می نماید. همچنین شهرداری یکی از شهرهای بزرگ آمریکا توانسته با جمع آوری داده ها و اطلاعات در بیش از 30 نقطه جغرافیایی شهری و آنالیز این اطلاعات، به منظور کاهش جرم و جنایت و نیز ارایه خدمات بهتر شهری استفاده کند[5] .

مسائل امنيتي يک نگراني عمده براي بسياري از سازمانها هستند. بر طبق تحقيقات شرکت داده هاي بين المللي، از شرکت هاي بزرگ انتظار مي رود که در سال 2012 مبلغ 32.8 ميليون دلار را در بخش امنيت رايانه صرف کنند و شرکت هاي کوچک و متوسط در طول سه سال آينده بيشتر در زمينه امنيت هزينه خواهند کرد تا روي ديگر خريدهاي فناوري اطلاعات [6].

برای ایجاد یک دید مناسب در خصوص ابرداده و اهمیت آن، جامعه‌ای را تصور کنید که در آن جمعیت بطور نمایی در حال افزایش است، اما خدمات و زیرساخت‌های عمومی آن نتواند پاسخگوی رشد جمعیت باشد و از عهده مدیریت آن برآید. چنین شرایطی در حوزه داده در حال وقوع است. بنابراین نیازمند توسعه زیرساخت‌های فنی برای مدیریت داده و رشد آن در بخش‌هایی نظیر جمع‌آوری، ذخیره‌سازی، جستجو، به اشتراک گذاری و تجزیه و تحلیل می‌باشیم. دستیابی به این توانمندی، معادل است با شرایطی که مثلا بتوانیم "در شرایطی که خطرات امنیتی افزایش پیدا می‌کند، سطح امنیت بیشتری را فراهم کنیم"،"تعداد تراکنش‌های بانکی، بیمه و مالی افزایش پیدا کند، ولی تقلب کمتری را شاهد باشیم"، "هنگامی که با اطلاعات بیشتری در حوزه سلامت مواجه باشیم، با بازدهی بیشتری سلامت را ارتقا دهیم"، "وقتی که با رویدادهای بیشتری از نظر آب و هوایی مواجه باشیم، توان پیش بینی دقیقتر و بهتری بدست آوریم"، "در دنیایی با خودروهای بیشتر، آمار تصادفات و حوادث را کاهش دهیم"، "با منابع طبیعی کمتر، به انرژی بیشتر و ارزان‌تری دسترسی داشته باشیم" و بسیاری موارد دیگر از این قبیل که اهمیت پنهان ابرداده‌ها را نشان می‌دهد[7] .

1-1-1 منابع تولید ابرداده‌ها

حجم داده‌های ذخیره شده در مجموعه‌های داده‌ای ابرداده‌ها، عموماً به خاطر تولید و جمع‌آوری داده‌ها از مجموعه بزرگی از تجهیزات و ابزارهای مختلف مانند گوشی‌های موبایل، حسگرهای محیطی، گزارش نرم‌افزارهای مختلف، دوربین‌ها،

www.SID.ir

میکروفون‌ها، دستگاه‌های تشخیص RFID[4]، شبکه‌های حسگر بی‌سیم وغیره با سرعت خیره‌کننده‌ای در حال افزایش است.

چند مثال از داده هاي ساختر نيافته كه توسط ماشين توليد شده اند:

- ✓ تصاوير ماهواره اي: شامل اطلاعات آب و هوا يا اطلاعاتي است كه دولت در تصاوير دوربين هاي مداربسته ماهواره اي مي گيرد. مثل Google Earth
- ✓ داده هاي علمي: شامل تصاوير زمين لرزه، اطلاعات جوي و فيزيک انرژي بالا است.
- ✓ عكس و ويدئو: شامل ويدئو ترافيک، امنيت و نظارت مي باشد.
- ✓ رادار يا داده هاي صوتي: شامل پروفايل های لرزه ای اقيانوس شناسی، هواشناسي و وابسته به وسايل نقليه مي باشد.

چند مثال از داده هاي ساختارنيافته كه توسط انسان توليد شده اند:

- ✓ متن داخلی به شركت شما: مانند تمام متن ها در اسناد، گزارشات، نتايج تحقيقات و ايميل ها. اطلاعات شركتي در واقع نشان دهنده ي درصد بالايي از اطلاعات متني در دنياي امروزه مي باشد.
- ✓ داده هاي مربوط به رسانه های اجتماعی: اين داده ها توسط پلت فرم هاي رسانه هاي اجتماعي از جمله YouTube, Facebook, Twitter, LinkedIn, and Flickr توليد مي شوند.
- ✓ داده هاي موبايل: شامل داده هايي از قبيل پيام هاي متني و اطلاعات محل مي باشد.
- ✓ محتواي وب سايت: از هر سايتي كه محتواي ساختارنيافته ارائه مي دهد جاري مي شود. مانند YouTube, Flickr, or Instagram.

و اين ليست ادامه دارد[1] .

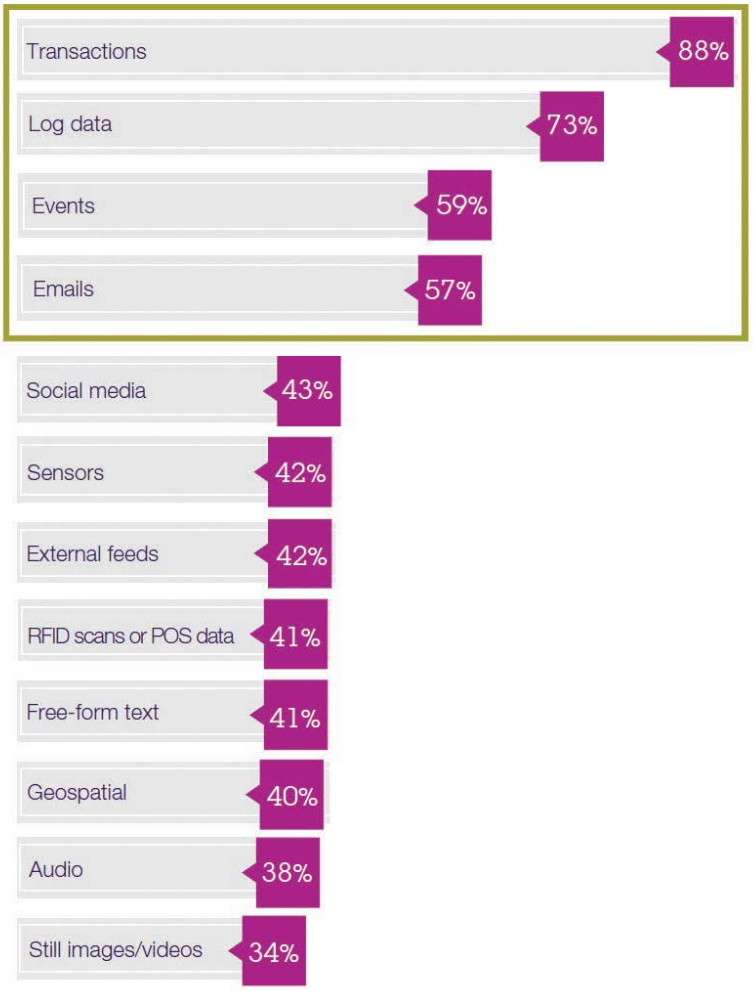
Archive of SID

[4]Radio-frequency Identification

www.SID.ir

سازمان ها به طور عمده از منابع داده داخلي براي تلاش هاي مبتني بر ابرداده استفاده ميكنند. از پاسخ دهندگان با تلاش هاي فعال بر روي ابرداده پرسيده شده است كه در حال حاضر از كدام منابع داده جمع آوري و آناليز مي كنند. هر نقطه داده به صورت مجزا جمع آوري شده بود. مجموع پاسخ دهندگان براي هر نقطه داده در محدوده 557 تا 867 نفر است . [8]

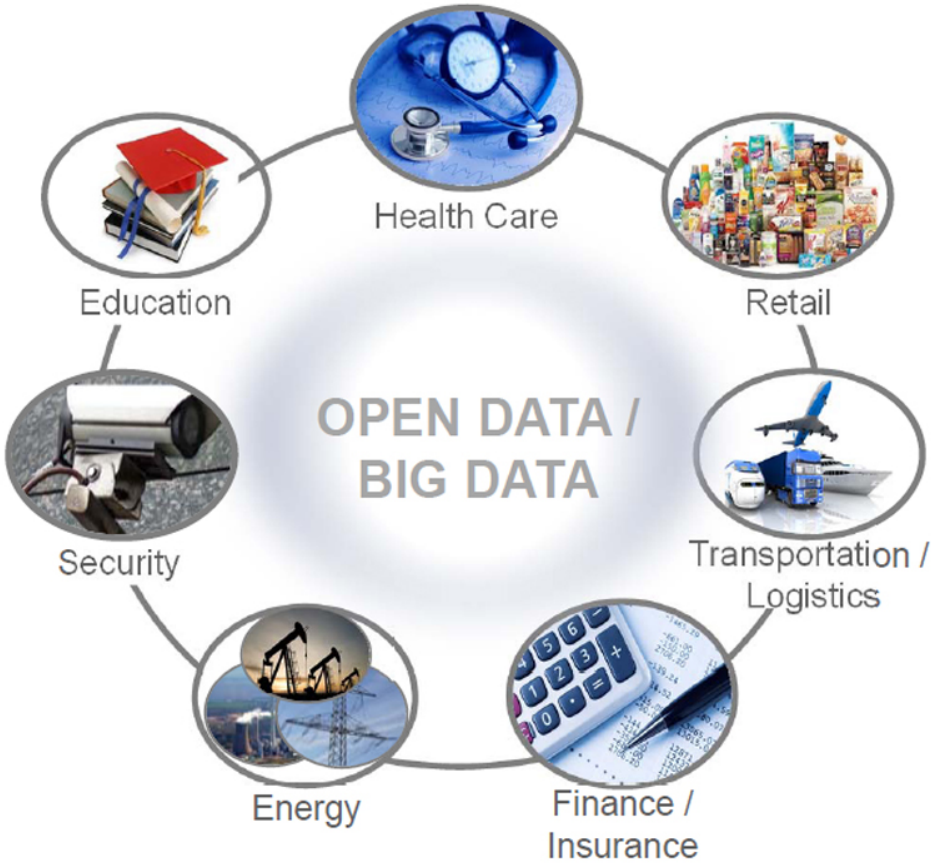
شكل 1: استفاده سازمان ها از منابع داده داخلي [8].



1-1-2 موضوعات مرتبط با ابردادهها

مطالعات فراوان در سطح رسانههای اجتماعی نشان میدهد بیشترین موضوع مرتبط با ابردادهها، در حوزه هاي امنيت، تحصيلات و آموزش هاي الكترونيک، پزشكي و مراقبتهای بهداشتی، خرده فروشي و بازاريابي، تراكنش ها و مديريت آنها، مسائل مالي و بيمه، و انرژي مي باشد كه تجزيه و تحليل آنها بيشتر به وسيله پايگاههای دادهای Oracle، IBM و Apple انجام میشود. (مطابق شكل زير) [13].

www.SID.ir



شكل2: زمينه هاي كاربردي ابرداده ها[13]

Archive of SID

www.SID.ir

1-1-3 ابرداده ها در مبارزه با جرم

امنيت و حفظ حريم خصوصی توسط سرعت، حجم و گوناگونيِ ابر داده ها بزرگنمايي شده اند. مانند زيرساخت هاي ابري در مقياس بزرگ، تنوع در فرمت ها و منابع داده ها، ماهيت جريان مالكيت داده و حجم بالاي جابجاييِ بين ابريِ داده ها. بنابراين مكانيزم هاي سنتي امنيت كه براي برقراري امنيت داده ي ايستا در مقياس كوچک مناسب بودند، ناكافي هستند[9].

برخی از فن آوری هاي مبتني بر داده های حجيم، قابليت استفاده از تحليل های پيش بينی و در نتيجه توانايی "پيش بينی هوشمند" بر اساس فعاليت های جنايی را براي اجراي قانون فراهم مي نمايند. اين توانايي بي نظير پتانسيلي به اجراي قانون مي دهد كه شامل فعال شدن در تلاش ها به منظور جلوگيري از فعاليت هاي جنايي قبل از وقوع آنها مي باشد[10].

علاوه بر اين كاهش در جرايم خشونت آميز و دزدي ها، ابرداده همچنين براي شناسايي مجرمان يقه سفيد كه در جرايم مالي از جمله تقلب هاي بيمه، تجارت خودي، پول شويي وكلاهبرداري در مراقبت هاي بهداشتي دست دارند، استفاده مي شود.

داده ي حجيم اگر به صورت قانونی اعمال شود مي تواند به ايجاد اجتماعي امن تر كمک كند و زيرساخت هاي كشور را منعطف تر كرده و امنيت ملی را قوت بخشد[11] .

www.SID.ir

اهداف و کاربرد موضوع : (*Materials and methods*)

2- ابرداده و تاثیر آن بر روي امنیت و ایمنی عمومی

2-1 برنامه هاي کاربردی:

- ✓ تجزیه و تحلیل جرم و جنایت
- ✓ جرم شناسی محاسباتی
- ✓ انفورماتیک تروریسم
- ✓ هوش منبع باز
- ✓ امنیت سایبری

2-2 داده ها:

- ✓ سوابق جنایی
- ✓ نقشه هاي جرم و جنایت
- ✓ شبکه های جنایی
- ✓ محتویات اخبار و وب
- ✓ پایگاه داده های حادثه تروریسم
- ✓ ویروس ها، حملات سایبری، بات نت ها

2-3 خصوصیات:

اطلاعات هویت شخصی، محتوای ناقص و فریبنده، اطلاعات غني گروه و شبکه، محتواي چند زبانه

2-4 تجزیه و تحلیل:

- ✓ ارتباط مجرمانه قانون کاوي و خوشه بندي
- ✓ تجزیه و تحلیل شبکه های جنایی
- ✓ تجزیه و تحلیل مکانی-زمانی و تجسم
- ✓ تجزیه و تحلیل متن چند زبانه
- ✓ تجزیه و تحلیل احساسات و اثر
- ✓ تجزیه و تحلیل حملات سایبری و اسناد

Archive of SID

www.SID.ir

2-5 اثرات:

- ✓ افزايش امنيت و ايمني عمومي[6] .

تحقيقات اخير روي تكنولوژي هايي براي امنيت سايبري، مبارزه با تروريسم، و نرم افزارهاي كاربردي مبارزه با جرم فاقد يک چارچوب سازگار براي پرداختن به اين چالش هاي داده اي است. تكنولوژي هاي انتخابي BI&A[5] از جمله ارتباط مجرمانه قانون كاوي و خوشه بندي، تجزيه و تحليل شبكه های جنايی، تجزيه و تحليل مكانی-زمانی و تجسم، تجزيه و تحليل متن چند زبانه، تجزيه و تحليل احساسات و اثر، و تجزيه و تحليل حملات سايبری و اسناد بايد براي تحقيقات انفورماتيک امنيتي در نظر گرفته شوند.

Archive of SID

[5]Business Intelligence and Analytics

www.SID.ir

3- مشکلات مربوط به امنیت و حریم خصوصی

داده های حجیم در حال تبدیل شدن به یکی از مهم ترین روندهای تکنولوژی هستند که دارای پتانسیل ایجاد تغییرات چشمگیر در روش استفاده ي سازمان ها از اطلاعات به منظور افزایش تجربه مشتری و دگرگوني مدل های کسب و کار خود، مي باشند[1].

براي مثال موسسات مالي با افزایش تقاضاي مشریان براي خدمات پیشرفته و شخصي تر همراه با افزایش مطالبات قانونی، باید با داده در اندازه هاي پتابایت مقابله کنند[5].

در بسیاري از زمینه های مدیریت داده، یک فرض وجود دارد که داده ها برای تجزیه و تحلیل و برنامه ریزی قوی تر به خوبي بررسي و ایمن شده اند. این معمولا از طریق یک فرایند پاکسازی داده و ایجاد پروفایل انجام مي گیرد، بنابراین داده میتواند قابل اطمینان باشد. جهان داده هاي حجیم یک مجموعه ي جدیدي از چالش ها و موانع را ارائه مي کند[1].

3-1 چالش های حوزه ابرداده‌ها

در بحث داده‌های بزرگ، ما نیاز داریم که داده‌ها را به منظور استخراج اطلاعات، کشف دانش و در نهایت تصمیم-گیری در خصوص مسائل مختلفِ کاربردی به صورت صحیح مدیریت کنیم. مدیریت داده‌ها عموما شامل 5 فعالیت اصلی میباشد:

- جمع آوری
- ذخیره سازی
- جستجو
- به اشتراک گذاری
- تحلیل

تاکنون چالش‌های زیادی در حوزه داده‌های بزرگ مطرح شده است که تا حدودی از جنبه تئوری ابعاد مختلفی از مشکلات این حوزه را بیان می‌کنند. این چالش‌ها در ابتدا، سه بعد اصلی حجم داده، نرخ تولید و تنوع به عنوان 3 V’s مطرح شدند:

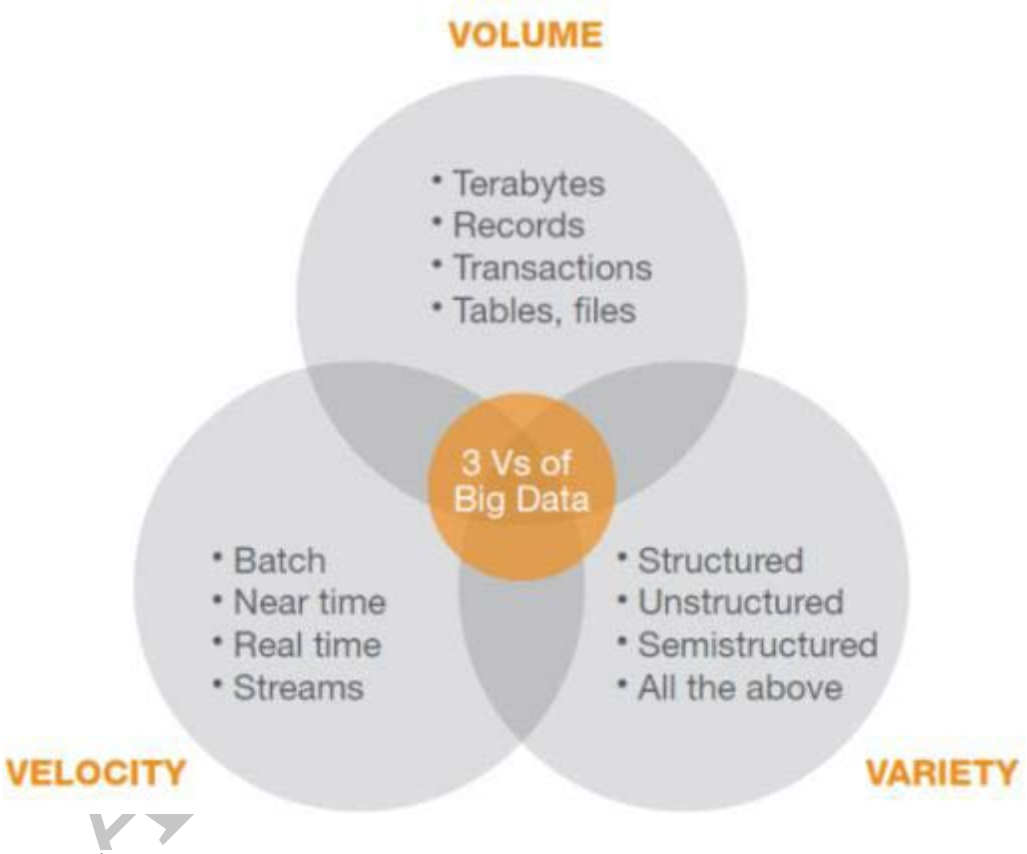
حجم داده (Volume): حجم داده‌ها به صورت نمایی در حال رشد می باشد. منابع مختلفی نظیر شبکه های اجتماعی، لاگ سرورهای وب، جریان‌های ترافیک، تصاویر ماهواره‌ای، جریان‌های صوتی، تراکنش‌های بانکی، محتوای صفحات وب، اسناد دولتی، پرونده‌های پزشکی و ... وجود دارد که حجم داده بسیار زیادی تولید می‌کنند.

نرخ تولید (Velocity): داده‌ها از طریق برنامه‌های کاربردی و حسگرهای بسیار زیادی که در محیط وجود دارند با سرعت بسیار زیاد و به صورت بلادرنگ تولید می‌شوند. بسیاری از برنامه‌های کاربردی لازم است به محض ورود داده، به درخواست کاربر پاسخ دهند. ممکن است در برخی موارد نتوانیم به اندازه کافی صبر کنیم تا مثلا یک گزارش در سیستم برای مدت طولانی پردازش شود.

www.SID.ir

تنوع (Variety): انواع منابع داده و تنوع در نوع داده بسیار زیاد می‌باشد که در نتیجه ساختارهای داده‌ای بسیار زیادی وجود دارد. مثلا در وب، افراد از نرم افزارها و مرورگرهای مختلفی برای ارسال اطلاعات استفاده می‌کنند. بسیاری از اطلاعات مستقیما از انسان دریافت می‌شود و بنابراین وجود خطا اجتناب ناپذیر است. این تنوع سبب می‌شود جامعیت داده تحت تاثیر قرار بگیرد. زیرا هرچه تنوع بیشتری وجود داشته باشد، احتمال بروز خطای بیشتری نیز وجود خواهد داشت.

شکل 3: سه V از ابرداده[3]



در ادامه چالش‌های بیشتری در ادبیات موضوع توسط محققان مطرح شده است که به بررسي آنها مي پردازیم:

صحت (Veracity): با توجه به اینکه داده‌ها از منابع مختلف دریافت می‌شوند، ممکن است نتوان به همه آنها اعتماد کرد. مثلا در یک شبکه اجتماعی، ممکن است نظرهای زیادی در خصوص یک موضوع خاص ارائه شود. اما اینکه آیا همه آنها صحیح و قابل اطمینان هستند، موضوعی است که نمیتوان به سادگی از کنار آن در حجم بسیار زیادی از اطلاعات گذشت. البته بعضی از تحقیقات این چالش را به معنای حفظ همه مشخصه‌های داده اصلی بیان کرده‌اند که باید حفظ شود تا بتوان کیفیت و صحت داده را تضمین کرد. البته تعریف دوم در مولدهای ابرداده صدق می‌کند تا بتوان داده‌ای تولید کرد که نشان دهنده ویژگی‌های داده اصلی باشد.

اعتبار (Validity): با فرض اینکه داده صحیح باشد، ممکن است برای برخی کاربردها مناسب نباشد یا به عبارت دیگر از اعتبار کافی برای استفاده در برخی از کاربردها برخوردار نباشد.

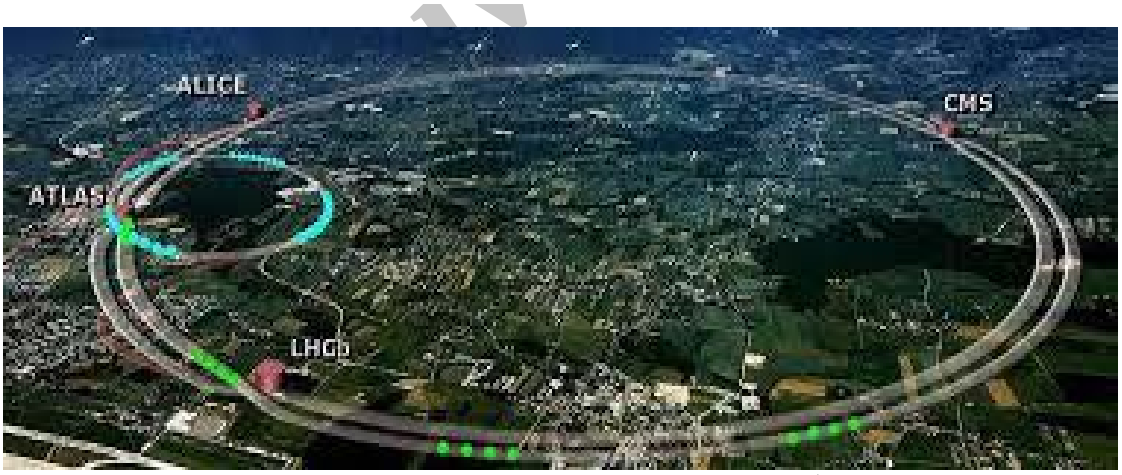
www.SID.ir

نوسان (Volatility): سرعت تغییر ارزش داده‌های مختلف در طول زمان می‌تواند متفاوت باشد. در یک سیستم معمولی تجارت الکترونیک، سرعت نوسان داده‌ها زیاد نیست و ممکن است داده‌های موجود مثلا برای یک سال ارزش خود را حفظ کنند، اما در کاربردهایی نظیر تحلیل ارز و بورس، داده با نوسان زیادی مواجه هستند و داده‌ها به سرعت ارزش خود را از دست می‌دهند و مقادیر جدیدی به خود می‌گیرند. اگرچه نگهداری اطلاعات در زمان طولانی به منظور تحلیل تغییرات و نوسان داده‌ها حائز اهمیت است. افزایش دوره نگهداری اطلاعات، مسلما هزینه‌های پیاده‌سازی زیادی را دربر خواهد داشت که باید در نظر گرفته شود.

نمایش (Visualization): یکی از کارهای مشکل در حوزه ابرداده‌ها، نمایش اطلاعات است. اینکه بخواهیم کاری کنیم که حجم عظیم اطلاعات با ارتباطات پیچیده، به خوبی قابل فهم و قابل مطالعه باشد از طریق روش‌های تحلیلی و بصری سازی مناسب اطلاعات امکان پذیر است.

ارزش (Value): این موضوع دلالت بر این دارد که از نظر اطلاعاتی برای تصمیم‌گیری چقدر داده، حائز ارزش است. به عبارت دیگر آیا هزینه‌ای که برای نگهداری داده و پردازش آنها می‌شود، ارزش آن را از نظر تصمیم‌گیری دارد یا نه. معمولا داده‌ها می‌توانند در لایه‌های مختلف جابجا شوند. لایه‌های بالاتر به معنای ارزش بیشتر داده می باشند. بنابراین برخی از سازمان‌ها می‌توانند هزینه بالای نگهداری مربوط به لایه‌های بالاتر را قبول کنند[12].

اطلاعات ديگر تنها به عنوان چيزي كه درون سيستم مي نويسيد و ذخيره مي كنيد مشاهده نمي شوند. اطلاعات تمام كارهايي است كه شما انجام مي دهيد. وب سايت هاي شبكه هاي اجتماعي مانند Facebook روزانه 500 ترابايت داده توليد مي كنند. مديريت همچين حجم گسترده اي از اطلاعات كار واقعا بزرگ و دشواري است. اين اطلاعات به صورت متن ها، تصاوير، ويدئوها، محل و غيره مي باشد كه آن را يک فرم بسيار متنوع از اطلاعات مي كند. و سرانجام هنگامي كه در مورد سرعت صحبت ميكنيم، اينجا 1155 ميليون كاربر فعال فيس بوک در نيمه سال 2014 وجود دارد. اين بدان معني است كه فيس بوک نه تنها بايد داده هاي بزرگ متنوع را ذخيره كند، بلكه بايد اين كار را با سرعت بسيار بالايي انجام دهد. در واقع آزمایش های انجام شده در آزمایشگاه هاي CERN برای برخورددهنده بزرگ هادرون(LHC)[6] در

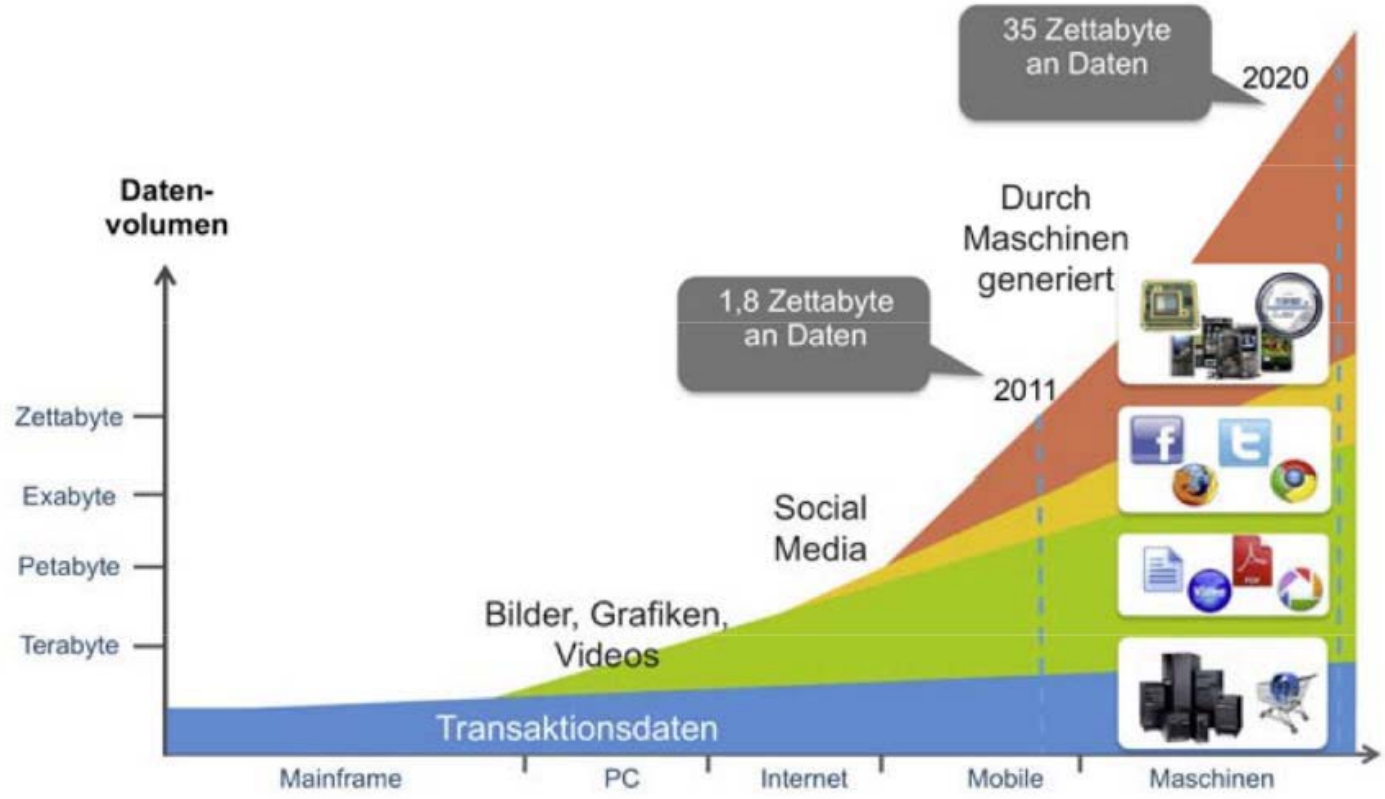


هر ثانیه تقریبا یک پتابایت داده تولید مي كنند. و بازیابی این داده نیز همانقدر اهمیت دارد. و بدین معني است كه ابرداده قابلیت های پردازش موازی در مقیاس بزرگ نیاز دارد.

[6]Large Hardon Collider

www.SID.ir

شکل 4: برخورددهنده بزرگ هادرون(LHC) **[3]**



شکل5: توسعه حجم داده ها [13]

سیل داده ها در حال حاضر بیش از ظرفیت ذخیره سازی در دسترس:

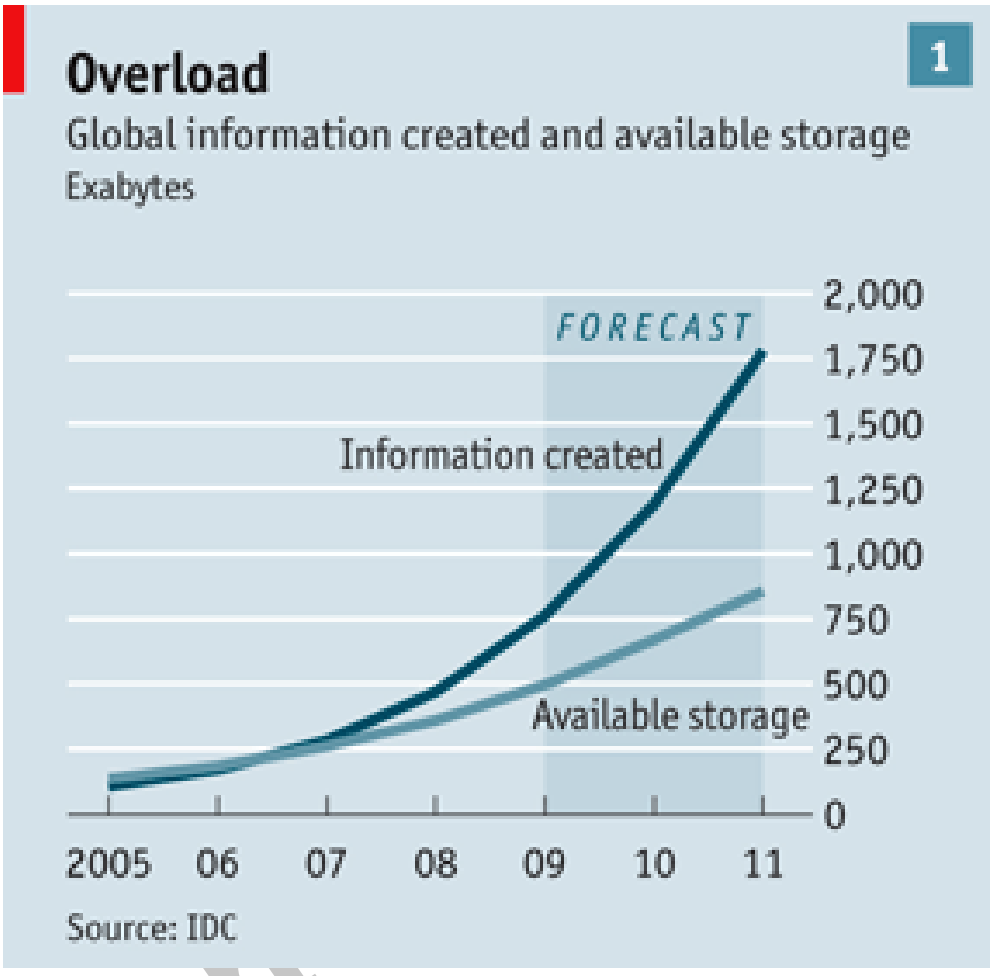
در حال حاضر داده هاي توليد شده در سراسر جهان از انباره موجود بيشتر مي باشند.

منابع داده به خصوص:

- ✓ داده هاي حركتي (GPS, GSM, etc.)
- ✓ اطلاعات موجودی
- ✓ داده های تراکنش
- ✓ داده های سنسور
- ✓ تصویر، ویدئو، فایل های صوتی
- ✓ داده های رسانه های اجتماعی
- ✓ فايل هاي گزارش اينترنت[13]

Archive of SID

www.SID.ir



شکل6: اضافه بار داده ها[13]

www.SID.ir

راه حل هاي ارايه شده به همراه ارجاعات: (*Results*)

4- مبارزه با جرم با تجزیه و تحلیل ابرداده

4-1 پیشبیني جرم با استفاده از ابرداده ها

ابرداده می‌تواند الگوهای رفتاری پنهان مردم و حتی نیت آنها را آشکار سازد [15].

بطور دقیق‌تر، ابرداده می‌تواند میان فاصله آنچه مردم می‌خواهند انجام دهند و آنچه واقعا انجام می‌دهند و اینکه آنها چگونه با دیگران و محیط خود تعامل دارند ارتباط ایجاد کند. این اطلاعات برای سازمان های دولتی و همچنین شرکت های خصوصی برای حمایت از تصمیم گیری در مورد زمینه هایی از اجرای قانون تا خدمات اجتماعی و امنیت میهن مفید است[16].

سازمان ها از تجزیه و تحلیل داده برای پیش بینی فصول مطلوب، روندهای آینده، مشتریان هدف، موقعیت های احتمالی و غیره استفاده می‌کنند. یک چنین رشته ای جرم شناسی و اجرای قانون نامیده می‌شود[3].

4-2 کاربرد ابرداده در مبارزه با جرم و جنایت

4 گام اساسی مهم وجود دارد:
1) جمع آوری اطلاعات.
2) تجزیه و تحلیل اطلاعات.
3) پیش بینی جنایات.
4) جلوگیری از جنایات.

فن آوری و سیستم های مختلف از انواع مختلف اطلاعات استفاده می‌کنند و دستگاه ها / رابط های مورد استفاده برای جمع آوری این اطلاعات نیز در برنامه های مختلف متفاوت است. تجزیه و تحلیل این اطلاعات با استفاده از نمودارهاي هیستوگرام[7]، نمودارهاي پای[8]، نمودارهاي وِن[9]، نمودارهاي میله اي[10] و غیره انجام شده است.

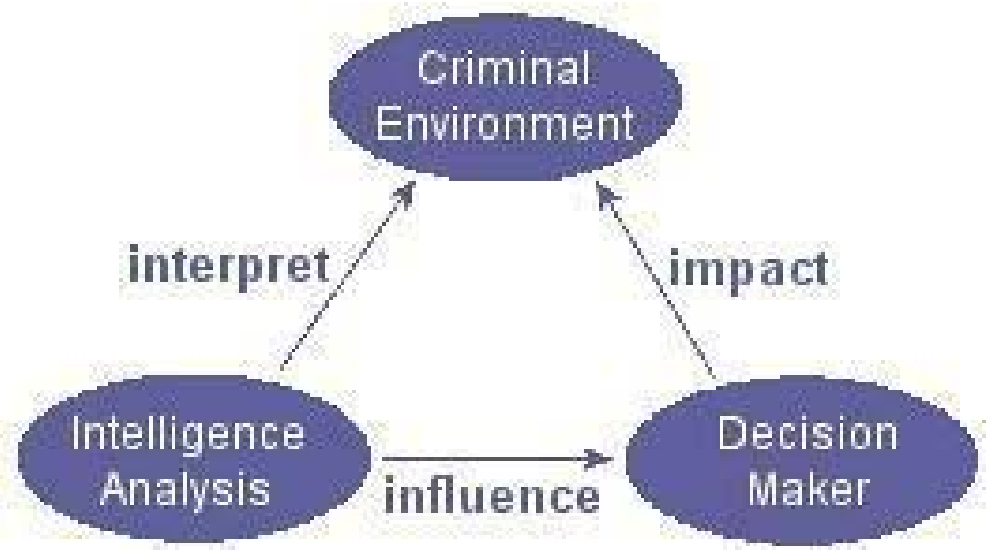
ابزار و یا نرم افزار سبب مراقبت از ترکیب و ارائه اطلاعات می‌شود که تخصص قابل توجه انسانی را در دو مرحله باقی مانده در نظر می‌گیرد. کاربر سیستم نتایج منطقی این ارائه را مشخص می‌کند و احتمال رخ دادن جرم و جنایت را تعیین می‌کند.

[7]Histograms
[8]Pie charts
[9]Venn diagrams
[10]Bar graphs

www.SID.ir

نتایج حاصل از تجزیه و تحلیل داده های بزرگ و سناریوی دنیای واقعی هم باید باهمدیگر در نظر گرفته شوند تا به یک پیش بینی دست یابیم و مرحله آخر انجام هر چیز ممکن در محدوده مرزهای قانونی و امکان پذیر و عملي برای جلوگیری از جرم و جنایت است. این نوع سیستم به نام هوش پلیسی (ILP) [11] نامیده می‌شود. بیایید فن آوری های موجود را ببینیم:

شکل7: هوش پلیسي [3]



4-2-1 نمونه های DNA بدست آمده از صحنه های جرم و جنایت

DNA جنایتکاران از صحنه هاي جرم را می‌توان به غیر معمول ترین روش ها به دست آورد. به طور کلی, DNA از مایعات ترشح شده انسان و یا از قطعات پوست به دست می‌آید. درواقع مایعاتي چون خون، تف، بزاق می‌تواند توسط کارشناسان پزشکی قانونی برای مطالعه DNA استفاده شود. این مایعات از قاب عینک، قاشق، چاقو، ته سیگار و غیره به دست می‌آیند. بخش هایی از پوست را می‌توان از زیر ناخن قربانی یا از لباس، متکا و غیره بدست آورد. نمونه هاي مو ممکن است توسط تیم پزشکی قانونی حفظ شوند. در انگلستان و ولز، هر کسی که به ظن داشتن یک جرم دستگیر شده باشد باید یک نمونه DNA را ارسال کند.

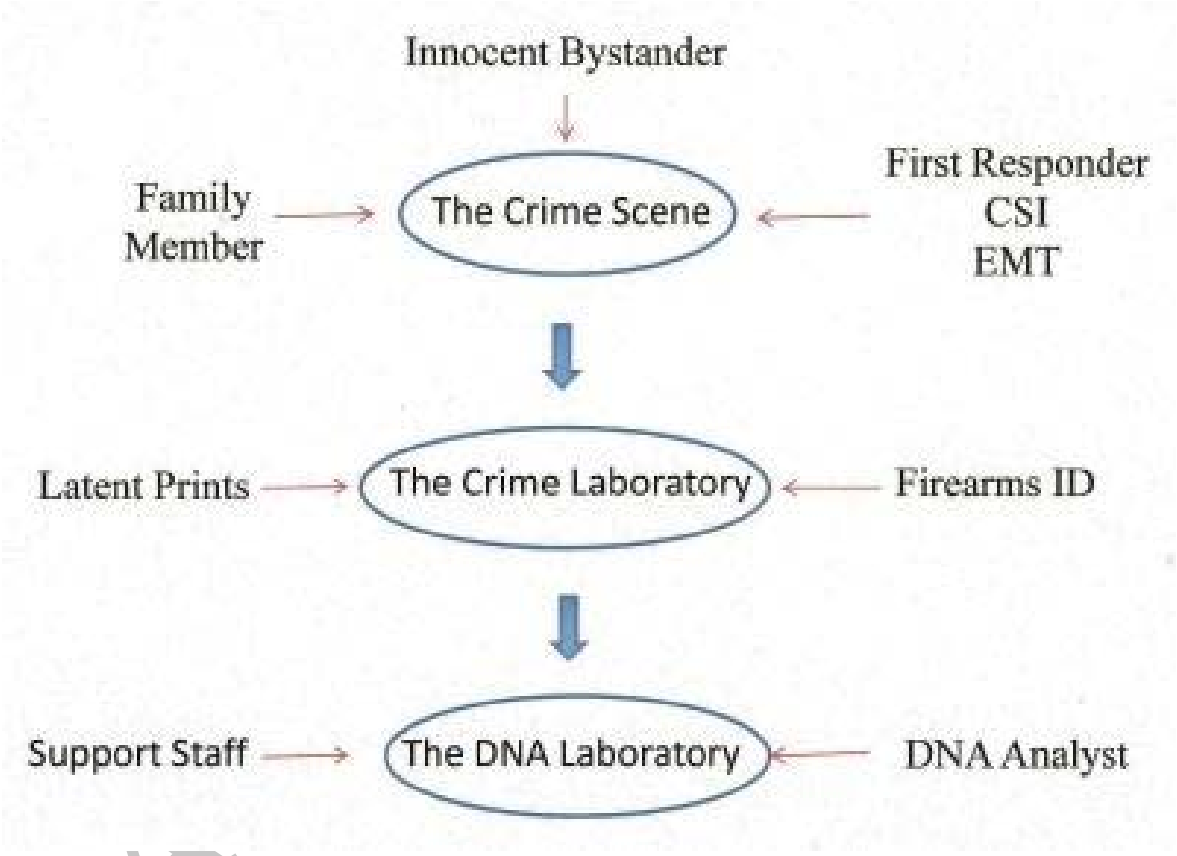
در اسکاتلند، قانون استلزام به حذف پروفایل DNA بسیاری از افرادي که تبرئه شده اند را از پایگاه داده دارد. در سوئد، تنها نمایه DNA جنایتکارانی که بیش از دو سال را در زندان گذرانده اند ذخیره می‌شود. در نروژ و آلمان، برای مجرمان جدی و خطرناک و کسانی که مرتکب جرم خاصي شده اند و احتمال ارتکاب مجدد جرم را دارند، به ترتیب دستور دادگاه و تنها در صورت دسترس بودن مورد نیاز مي باشد.

[11] Intelligence Led Policing

www.SID.ir

چهل و نه ایالت در ایالت متحده آمریکا نمایه DNA مجرمان خطرناک را ذخیره مي کنند. ایالات متحده آمریکا بزرگترین پایگاه داده DNA را در جهان، با سیستم شاخص DNA ترکیبی (CODIS) [12] و با بیش از 9 میلیون پرونده در سال 2011 را داراست.

همه این پرونده ها با استفاده از ابرداده ذخیره می‌شوند. با استفاده از این نمونه ها محققان می‌توانند نتیجه بگیرند که آیا نمونه DNA متعلق به یک تماشاچی بی گناه است و یا کسی که در حال حاضر دارای سابقه کیفری می‌باشد.



شکل8: ارزیابي DNA [3]

4-2-2 سنسور / محل یاب تیراندازی

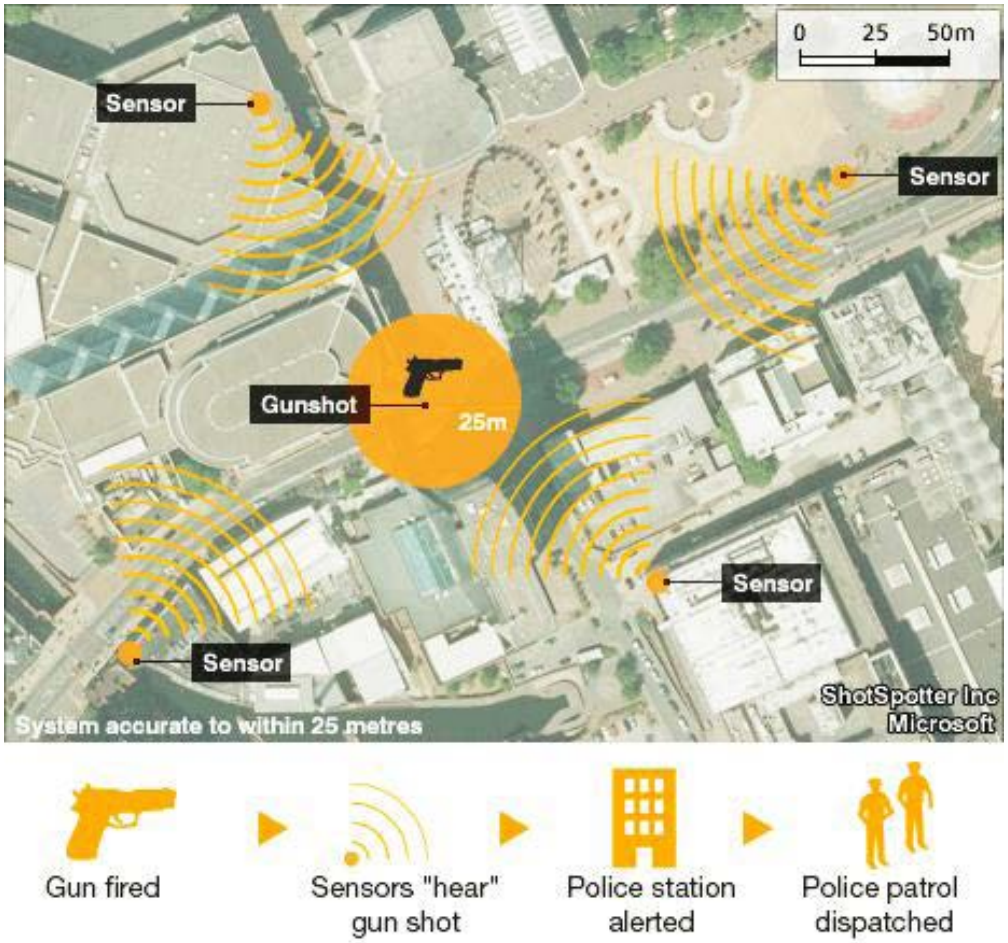
ایالات متحده آمریکا این تکنولوژی را با موفقیت اجرا کرده است. این سنسورهای تیراندازی می‌توانند برای پوشش مکان های عمومی کوچک مانند پارک ها، پارکینگ ها، موزه ها و غیره مورد استفاده قرار گیرند و یا آنها می‌توانند برای

[12]Combined DNA Index System

www.SID.ir

پوشش شهرهای کوچک، مجتمع های برجسته و معروف، پایگاههای عملیاتی نظامی و غیره بکار روند. این سنسورها نه تنها می‌توانند شلیک یک گلوله را تشخیص دهند، بلکه آنها همچنین می‌توانند موقعیت جغرافیایی حادثه را تعیین کنند.

صوت شناسي تیراندازی مورد مطالعه قرار گرفت و سنسورها صدای شلیک گلوله یا خط سیر تیراندازي را تشخیص می‌دهند. شلیک یک گلوله معمولي معادل 120 تا 160 دسی بل صدا را تولید می‌کند. همچنین، اگر گلوله در محدوده مشخص شده سنسورها باشد، سنسور می‌تواند حرکات (نورشناسي) مادون قرمز گلوله را احساس کند.



شکل9:سنسور تیراندازي [3]

هنگامی که یک تیراندازي شناسایی شد، به مقامات اطلاع داده می‌شود و در برخی موارد، دوربین های نزدیک به محل براي ضبط تصویر به سرعت در یک ردیف قرار مي گیرند. این کار با استفاده از سیستم اطلاعات جغرافیایی

www.SID.ir

(GIS) [13] انجام می‌شود. در حال حاضر، مجموعه ای از این همه داده از قبیل صدا، نور، موقعیت و غیره نیاز به تکنولوژی ابرداده ها دارد. چنین سیستمی بخشی از ILP (هوش پلیسی) است.

4-2-3 بیومتریک ها براي مقابله با سرقت

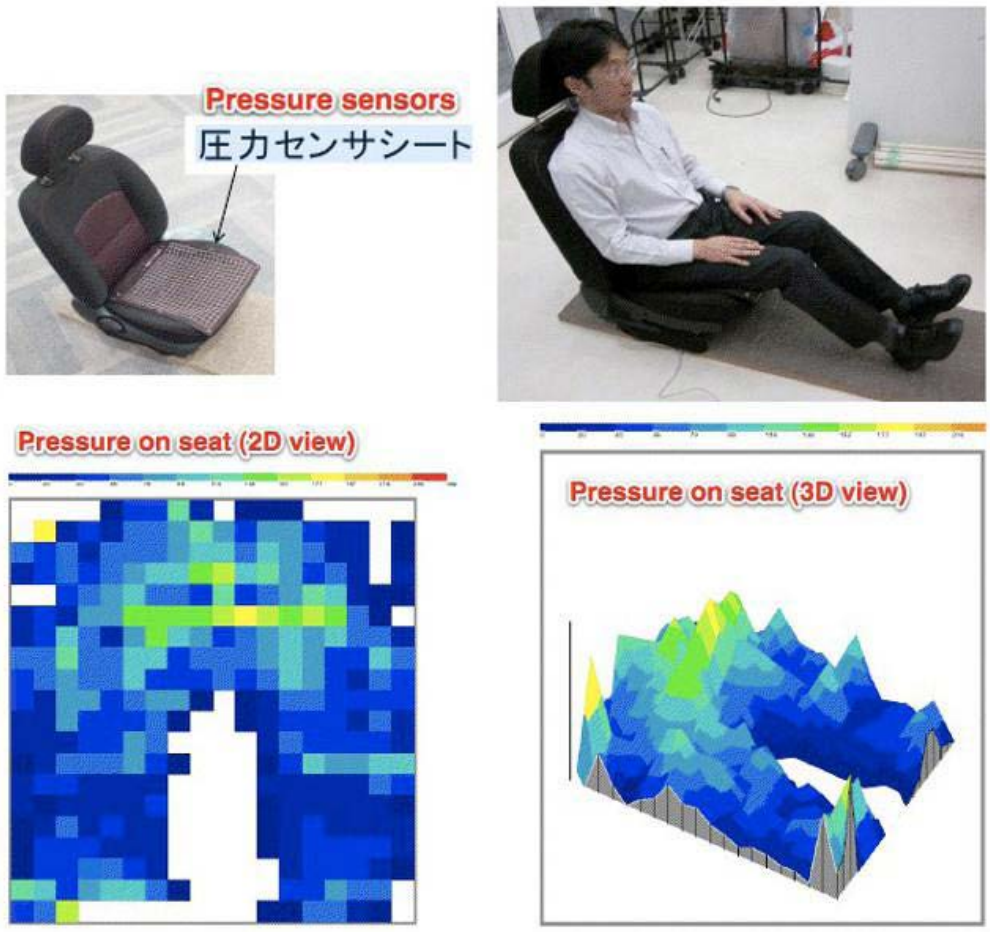
بیومتریک ها را می‌توان به روش های مختلفی برای جلوگیری از سرقت استفاده کرد. محققان مدرن امروزی استفاده از سنسور داخل و اطراف صندلی راننده را پیشنهاد کرده اند. کل اطلاعات بیومتریک راننده ثبت خواهد شد و در سیستم ذخیره می‌گردد. این اطلاعات شامل وزن، دینامیک بدن، اسکن شبکیه چشم، ردیاب اثر انگشت بر روي فرمان اتومبیل و غیره مي باشد.

اگر شخصی تلاش کند ماشینی را براند که اطلاعات بیومتریک او از قبل در سیستم آن ذخیره نشده است، سیستم بلافاصله درهای ماشین را قفل می‌کند و ممکن است با استفاده از سیستم موقعیت یابی جهانی (GPS) [14] به پلیس اعلام خطر کند. راننده داخل ماشین گیر می‌افتد، مگر اینکه یک رمز عبور توسط شخصي که اطلاعات بیومتریک وي توسط سیستم شناسایی شده است، به سیستم وارد شود. نوع مشابهی از سیستم بیومتریک برای اسلحه نیز پیشنهاد شده است. تفنگ شلیک نمی کند مگر آنکه توسط صاحب تفنگ نگه داشته شده باشد. چنین سیستم های بیومتریکي در واقع مقادیر زیادی از اطلاعات را در هر زماني که درخواست کشیدن ماشه داده میشود، تولید می‌کنند. این اطلاعات بسیار به سرعت تولید می‌شوند و یک نوع داده یکسان نیستند.

Archive of SID

[13] Geographic Information System
[14] Global Positioning System

www.SID.ir



شکل10: سیستم ضد سرقت اتومبیل [3]

www.SID.ir

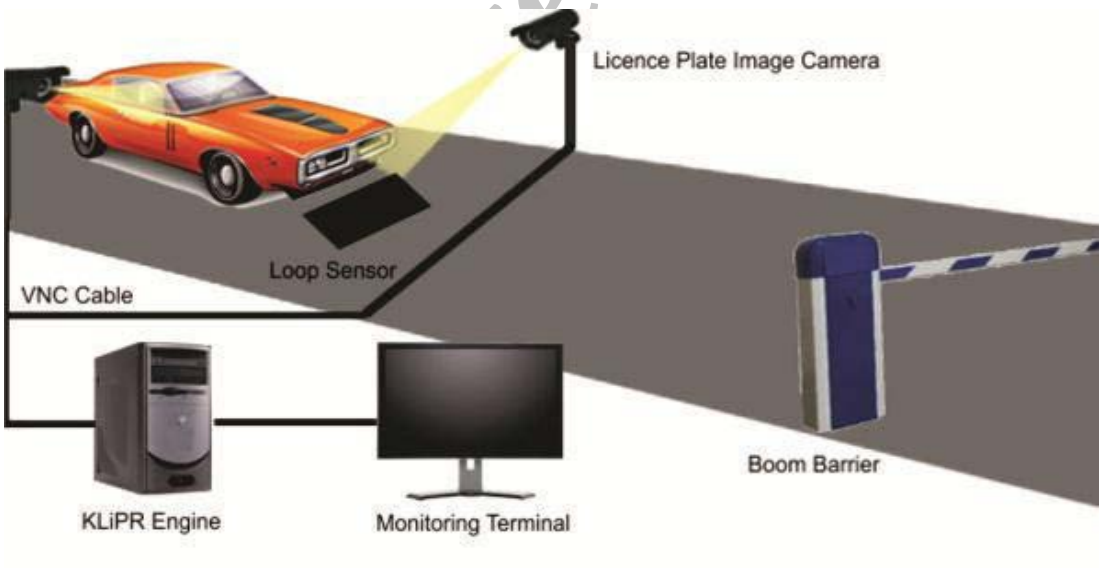
شکل11: سنسورهاي روي يک تفنگ [3]

4-2-4 شناسایی تصویر

شناسایی تصویر به خودی خود یک شاخه گسترده است. شناسایی تصویر شامل تشخیص چهره، تصویربرداری ویدئویی، تشخیص خودکار شماره پلاک (ANPR) [15] و غیره مي باشد. سازمان های اجرای قانون که یک مجموعه دیجیتالی از تصاویر همه شهروندان را دارند می‌توانند از این گونه سیستم ها بهره بسیاري برند.

ویژگی های تشخیص چهره توسط سایت های شبکه های اجتماعی مانند فیس بوک استفاده می‌شوند و در طول سال ها، این سیستم ها نسبتا دقیق شده اند. شناسایی چهره ها از یک پس زمینه در حال حرکت (مانند یک تصویر متحرک) مزیت بزرگ آن است که به تشخیص هویت چهره هاي شناسایی شده کمک می‌کند، اگر و فقط اگر یک پایگاه داده جامع وجود داشته باشد.

این بدین معنی است که عملا هر ویدئو ضبط شده از یک جرم می‌تواند برای شناسایی متخلف مورد استفاده قرار گیرد. همچنین، سیستم هایی مانند ANPR می‌توانند بسیار مفید باشند. این روش مي تواند به پلیس در گرفتن رد اتومبیل هايي که به قلمرو و حوزه ي استحفاظي آنها وارد و خارج میشوند، کمک کند. سیستم های سنتی تنها شماره پلاک خودرو و ماشین را عکس می‌گرفتند. اما ANPR صاحب ماشین را شناسایی کلي می‌کند. چنین سیستم هايي به ما کمک می‌کنند تا بین حرکت خودرو و جنایات ارتباط برقرار کنیم. ابرداده این امر را ممکن می‌سازد.



شکل12: تشخیص خودکار شماره پلاک خودرو [3]

[15] Automated Number Plate Recognition

www.SID.ir

4-2-5 برنامه های منطقه بندی

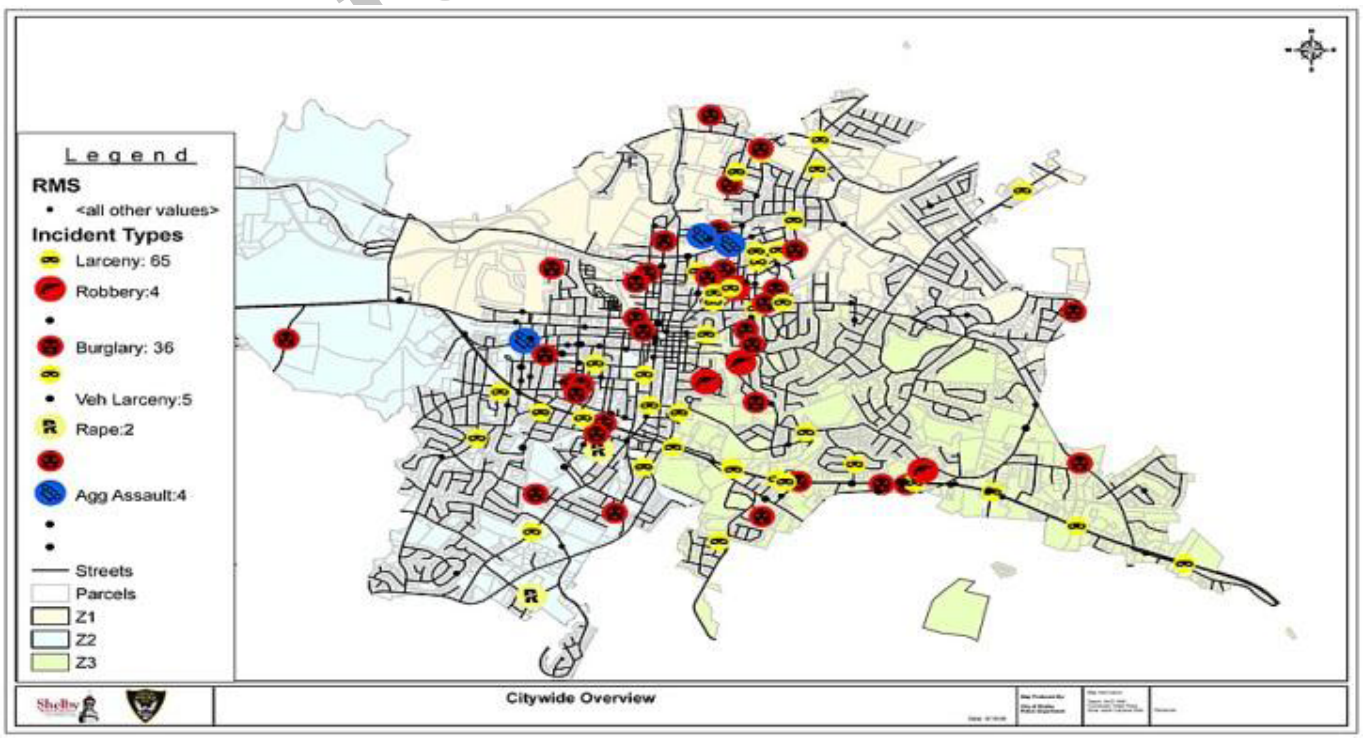
برنامه های منطقه بندی واقعا ترکیب مرحله پیش بینی هستند. تمام برنامه های قبلی که ما مورد مطالعه قرار دادیم جمع آوری، تجزیه و تحلیل و روش های پیشگیری بودند. برنامه های منطقه بندی موفق هستند، اگر نهادهای مجری قانون (LEA) [16]، یک GPS قوی و یا هر سیستم GIS اي را در اختیار داشته باشند.

برنامه های منطقه بندی داراي پارامترهای تنظیم شده اي مي باشند که در آنها حوادث خاص ثبت می‌شوند. یک برنامه منطقه بندی داده شده ممکن است توسط پلیس راهنمایی و رانندگی برای ذخیره اطلاعات در مورد تصادفات ثبت شده، انبوه ترافیک، تخلفات قانون ترافیک و غیره مورد استفاده قرار گیرد.

تمام این وقایع همراه با محل اتفاق به صورت منظم ثبت شده اند. با استفاده از این اطلاعات، تحلیلگران ممکن است نمایشي ارائه دهند از اینکه چه نوع جرم هایي، در کدام بخش از شهر و در کدام روز از هفته و چه ساعتي رخ می‌دهند. با استفاده از این، عیب یابي برای مقامات دولتی و گاهی هم براي سیاست گذاران آسان می‌شود که به حل مسائل کشوري، در صورت وجود، بپردازند. لازم به ذکر است که راهبندان در طول زمان مدرسه گزارش مي شود. در این حالت، شاید از اتوبوس های خواسته شود که مسیرهای جایگزین را پیدا کنند. یا اینکه ممکن است مشاهده کنیم که قوانین راهنمایی و رانندگی بیشتر در طول صبح و در روزهای هفته نقض مي شوند.

زمانی که سیستم ها با چنین نتایج آماری وارد میشوند، مقامات می‌توانند قدم هاي مورد نیاز را با سهولت و حس مدیریت بیشتر بردارند. به طور مشابه، ما می‌توانیم سیستمی را داشته باشیم که محل همه جنایت ها را به صورت معقولانه در تمام طول سال ردیابي کند.

در حال حاضر ممکن است در طول جشنواره ها، مغازه داران تعداد بیشتری از موارد دزدی را گزارش کنند و یا ممکن است متوجه شویم که حوادث رانندگی نوجوانان تحت نفوذ (DIU) [17] در یک منطقه خاص پس از نیمه شب و زمانیکه نیمسال تحصیلي پایان یافته، افزایش می‌یابد. با استفاده از این داده ها، تحلیل گران می‌توانند نتایج را بیرون کشیده و اقدامات مبتکرانه اي را بکار ببرند. اینگونه است که ما می‌توانیم جنایات آینده را پیش بینی کنیم و به جلوگیری از خسارات بالقوه کمک کنیم[3].



شکل13: مناطق جرم برجسته [3]

4-3 استفاده از تحليل هاي ابرداده درمبارزه با جرايم مالي

تبديل حجم، سرعت، تنوع و گوناگوني داده به بصيرت و درون بيني براي حفاظت از کسب و کار خود.

موسسات مالي چگونه مي توانند از اين داده ها براي جلو بردن مديريت جرايم مالي استفاده کنند وقتيکه شرکت ها به طور سنتي از کمتر از 5% داده هاي موجود براي جلوگيري از فعاليت هاي مالي غيرقانوني استفاده مي کنند؟

4-3-1 مدیریت ابرداده در خدمات مالی

موسسات مالی، در مواجهه با رشد خواسته های مشتری برای خدمات پیشرفته و شخصی تر همراه با افزایش خواسته هاي نظارتي، مي بايست با پتابايت داده مقابله کنند. تجزیه و تحلیل ابرداده ها گام بعدی در تکامل ابزار BI[18] است. انتظار مي رود که اين ابزارها براي تحليلگران جرائم مالی، توانایی بازجویی سریع مجموعه هاي وسيعي از داده را فراهم کنند که منجر به تشخیص زود هنگام و حتی پیشگیری از جرایم مالی می‌شود.

این قابلیت هزینه ها را بوسیله حذف تکیه بر فروشندگان نرم افزار مدیریت جرائم مالی، کاهش خواهد داد و امکان بازجویی کوئري هاي پيچيده روي مجموعه داده های تصادفی را به صورت سریع تر و کارآمدتر براي شرکت فراهم می‌کند. منفعت تجزیه و تحلیل ابرداده فراتر از داشتن حجم زیادی از داده های واضح بوده، در واقع می‌تواند در به حداقل رساندن تقلب کمک کند.

ابرداده به ساخت یک مدل کمک می‌کند. یک مدل ساخته شده بر اساس هر واقعه کلاهبرداري، به جای محدود کردن آن به یک گروه نمونه، برمي گردد به سال هاي خيلي قبل براي تک تک افراد. برای کمک به جلوگیری از تقلب در آینده، هر بار که انحراف از مدل اتفاق مي افتد، می‌تواند دوباره ساخت شود.

4-3-2 مدیریت جرائم مالی و تجزیه و تحلیل داده ها

همانطوریکه موسسات مالی اجازه دسترسی به خدمات را در سراسر کانال های جدید می‌دهند، ریسک مسیرهاي حملات جديد افزايش مي يابد. افزایش بانکداری تلفن همراه مجموعه ای جدید از دستگاه ها را معرفي کرد وسازندگان ابزارهای بانکی و موبايلي بايد بر ضعف ذاتي طراحی این دستگاه ها چیره شوند.

همانطوریکه بازار گوشی های هوشمند توسعه می‌یابد و تولید کنندگان شروع به انجام اقدام دفاعی در برابر هکرها و تقویت امنیت دستگاه های خود می‌کنند، اقلام داده های جدید در دسترس قرار خواهند گرفت که موتورهای تشخیص و

[18]Business Intelligence

www.SID.ir

پیشگیری از تقلب را تقویت خواهد کرد. تجزیه و تحلیل داده برای تعیین قوانین حل مدیریت جرائم مالی (FCM)[19] توسط تشخیص ارتباط بین جرم های مالی و ویژگی های معامله یا مجموعه ای از معاملات استفاده می‌شود.

- دفتر مواد مخدر و جرایم سازمان ملل اخیرا بیان کرده است که 1.6 تریلیون دلار کثیف شناور در اقتصاد جهانی وجود دارد و این مقدار هر ساله در حال افزایش است.
- آخرین آمار سالانه سازمان ملی تقلب انگلستان نشان می‌دهد که بیش از 38 بیلیون پوند در 12 ماه گذشته به دلیل تقلب از دست رفته است. این مقدار به بیش از 25٪ افزایش یافته است.
  - بخش عمومی(21 بیلیون پوند) بزرگترین بخش از دست رفته را گزارش کرده است، در حالی که بخش خصوصی هزینه 12 بیلیون پوند را براي دولت در بر داشته بعلاوه 4 بیلیون پوند دیگر در ضرر و زیان تقلب علیه افراد.
  - صنعت خدمات مالی در سال 2010 مبلغ 3.6 بیلیون پوند در ضرر و زیان های جعلی را ثبت کرده است.

4-4 راه حل ابرداده براي پردازش سریع

با تکامل فن آوري محاسبات،این امکان بوجود آمده تا بتوان حجم عظيمي از داده ها را که قبلا فقط توسط ابر رایانه ها و با هزینه کلان به کار گرفته میشد را مدیریت کرد. قیمت سیستم های کاهش یافته و در نتیجه تکنیک هاي جدید براي محاسبات توزیع شده، جریان اصلي مي باشند. دستیابی به موفقیت واقعی در ابرداده زماني اتفاق افتاد که شرکت هاي بزرگ مانند yahoo!، Google و Facebook به این واقعیت رسیدند که مقادیر زیاد اطلاعاتي که با عرضه هایشان توليد مي کنند را باید به پول تبدیل کنند، و براي این کسب درآمد نیاز به کمک دارند. این شرکت هاي نوظهور نیاز به یافتن تکنولوژي هاي جدید داشتند که به آنها اجازه ذخیره، دسترسي، و تجزیه و تحلیل مقدار زیادي از داده ها را نزدیک به زمان واقعي بدهد. بنابراین آنها میتوانستند از مزایای داشتن این همه داده درباره شرکت کنندگان در شبکه هایشان کسب درآمد کنند. راه حل هاي منتج شده آنها تبدیل بازار مدیریت داده ها مي باشد. به طور خاص، نوآوري MapReduce، Hadoop ، و Big Table، ثابت کرد که مي تواند جرقه ای باشد که به نسل جدیدی از مدیریت داده ها منجر مي شود. این فن آوري ها رسیدگی به یکی از اساسی ترین مشکلات یعني قابلیت پردازش حجم انبوهی از داده ها به طور کارآمد، به صورت مقرون به صرفه و در زمان سریع، مي باشد.

در پایگاه داده ابزارهاي تجزیه و تحلیل منطق کسب و کار را که بوسیله تجارت توسعه و پشتیبانی شده اند را فعال مي کنند، تا به طور مستقیم برای مجموعه هاي بزرگي از داده با حداقل دخالت بخش IT به کار برده شوند. پردازش موازی انبوه (MPP)[20] پایگاه داده ها، مدیریت مرکزی را قادر به محافظت و نگهداري از داده ها مي کند. بسته به اینکه چگونه اجرا می‌شوند، دسترسی سریع تر و کارآمدتر به مجموعه هاي بزرگ داده با موازي سازي کوئري ها و یا وظایف، فراهم می‌شود. پایگاه های داده MPP اجازه می‌دهند که تیم مدیریت جرائم مالی مقادیر زیادی از اطلاعات کسب و کار را به سرعت و کاراتر از پایگاه داده های استاندارد استخراج کنند.

هرچند که دخالت IT برای پیاده سازی منابع جدید داده برای اطمینان از موازی سازی بهینه کوئري ها و یا وظایف مورد نیاز است.، ترکیبی از تجزیه و تحلیل هاي درون پایگاهي و MPP، به تیم FCM این اختیار را مي دهد که به سرعت پاسخي ارائه دهند و منطق کسب و کار اضافه شده موثري را به منظور مبارزه با حملات مسیرهاي جدید و سپس کاهش تاثیر آن حملات بدون نیاز به تکیه بر شخص ثالث فروشندگان محصول FCM، فراهم آورند.

[19] Financial Crime Management
[20] Massively Parallel Processing

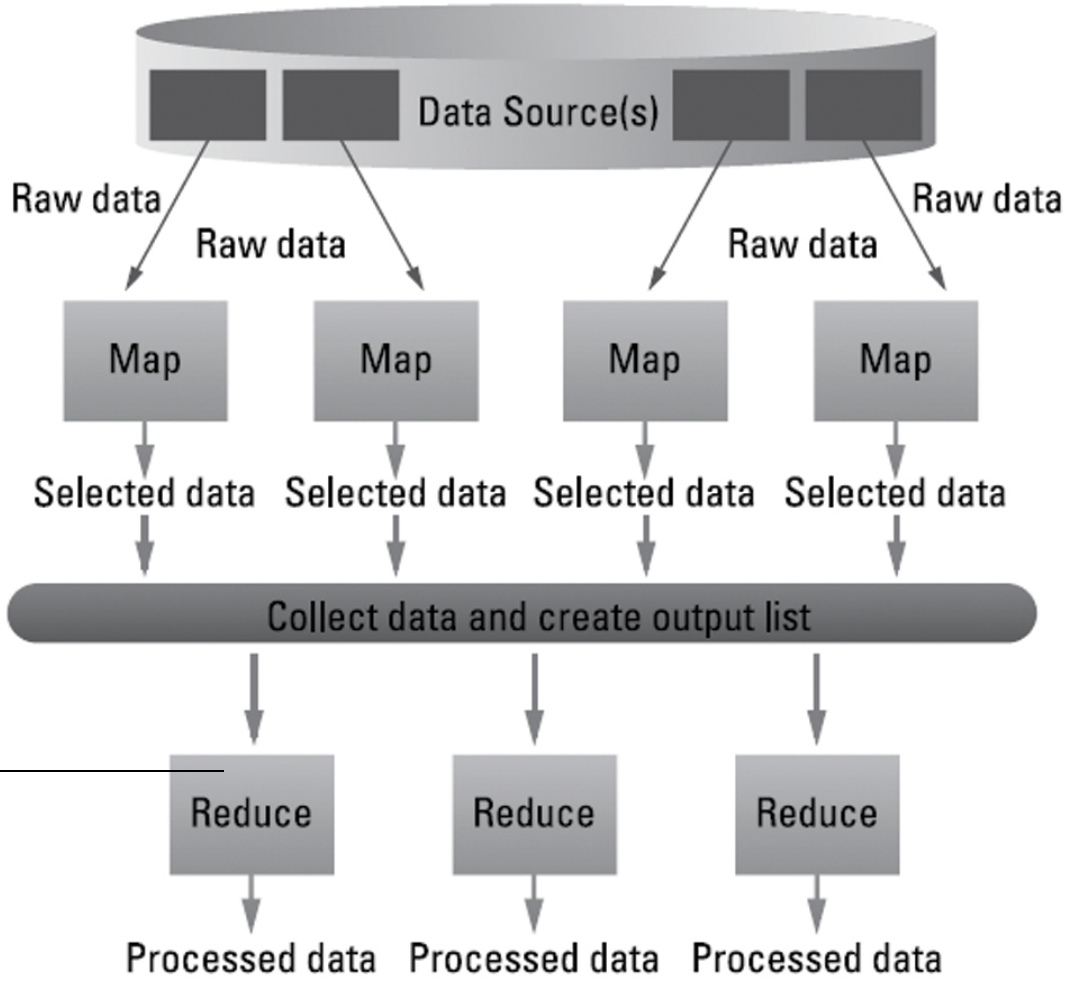
www.SID.ir

MapReduce 4-4-1

نقشه/کاهش[21]، تکنولوژی اي است که توسط گوگل طراحی شده و چارچوبي نرم افزاري مي باشد براي پشتيباني از پردازش توزيع شده بر روي مجموعه داده های بزرگ در سراسر خوشه هاي رايانه اي و برای افزایش مقیاس از یک ماشین به چند هزار ماشین، طراحي شده که توسط شرکت هایی مانند گوگل[22] و یاهو[23] برای کمک به جستجو استفاده مي شده است. آمازون[24]، به عنوان بخشی از ارائه دهنده محاسبات ابری، قابلیت های داده کاوی را برای مشتریان EC2 فراهم می‌کند.

نگاشت‌کاهش چارچوبی برای پردازش مجموعه‌های عظیمی از داده‌ها بر روی رایانه‌ها(گره‌ها) که بر روی موضوعی خاص فعالیت می‌کنند. این مجموعه روی‌هم رفته به عنوان خوشه شناخته می‌شود(در صورتی که از سخت‌افزاری یکسان بهره برند). پردازش محاسباتی بر روی دادهای ذخیره شده درون سامانه فایل (ساختار نیافته) یا بر روی پایگاه داده (ساختاریافته) قابل اجراست. نقشه/کاهش با اختصاص زیر نمایش ها به گره های کارگر و سپس با کاهش چندین نتیجه به یک نتیجه، می‌تواند به محققان اجازه کشف روابط بین داده های ساختار نیافته و معاملات غیر متشابه در داده های ساختار یافته را بدهد[5].

شکل 14: مدل نگاشت/کاهش [14]



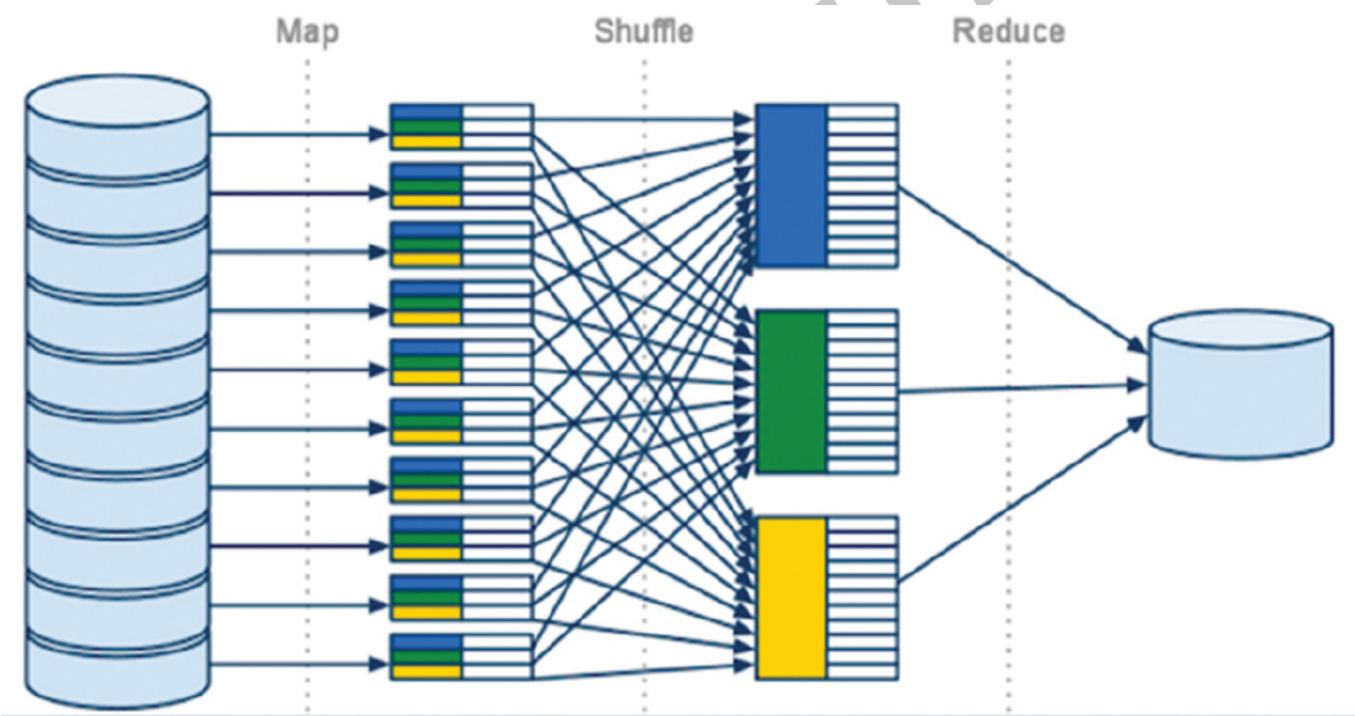
[21] Map/Reduce
[22] Google
[23] Yahoo
[24] Amazon

SID.ir

در این روش، دو گام اصلی به نام Map و Reduce وجود دارد:

گام Map: گره اصلی، ورودی را گرفته، و آن را به زیر مسائل کوچکتری تقسیم می‌کند. سپس آنها را بین گره‌هایی که وظیفه انجام کارها را دارند، توزیع می‌کند. ممکن است این نود نیز همین کار را تکرار کند که در این حالت یک ساختار چند سطحی داریم. در نهایت این زیر مسائل پردازش شده و پاسخ به گره اصلی ارسال می‌شود.

گام Reduce: حال، گره اصلی که پاسخ ها و نتایج را دریافت کرد، آنها را برای ارائه خروجی، ترکیب می‌کند. در این میان ممکن است اعمالی مانند فیلتر کردن، خلاصه کردن و یا تبدیل کردن، بر روی نتایج انجام دهد.



شکل 15: پردازش مجموعه داده ها در مراحل: مدل نگاشت/کاهش در موتور برنامه گوگل [14]

به عنوان مثال ,ما برنامه ای می نویسیم که داده آب و هوا را کاوش می کند. جمع آوری داده سنسورهای آب و هوا در هر ساعت در نقاط بسیاری از سراسر جهان حجم زیادی از داده گزارش را انباشته می کند, که کاندید مناسبی برای آنالیز با MapReduce است، چون نیمه ساخت یافته و رکوردگرا است.

www.SID.ir

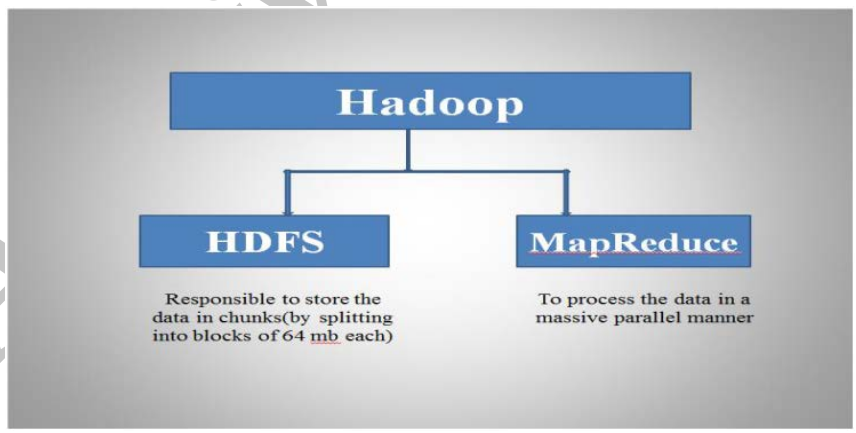
Big Table 4-4-2

جدول بزرگ توسط گوگل طراحی شده است که بتواند یک سیستم ذخیره سازی توزیع شده به منظور مدیریت داده های ساخت یافته در مقیاس بالا باشد. داده به جداول و سطرها و ستون ها سازمان یافته است. برخلاف مدل پایگاه داده رابطه ای سنتی، جدول بزرگ یک نقشه ی چند بعدی دائمی، توزیع شده و پراکنده است. این جدول برای ذخیره حجم عظیمی از داده ها در سراسر سرور کالا در نظر گفته شده است.

Hadoop 4-4-3

هدوپ[25] يك پروژه مبتني بر برنامه نويسي متن باز است كه توسط سازمان نرم افزاري آپاچي ايجاد شده است و از MapReduce و Big Table مشتق شده است. ايده اوليه هدوپ اولين بار در شركت گوگل رقم خورد اما خيلي ها باور به پياده سازي اين سيستم نداشتن و در چند سال اول اين ايده تنها بصورت تئوري مطرح بود. هدوپ امكان ذخيره سازي اطلاعات را در چندين سرور با هزينه اي پايين فراهم مي آورد. كلودرا شركتي است كه بصورت فعال در اين زمينه فعال مي باشد و بسته نرم افزاري بي نظير هدوپ را ايجاد كرده و آن را انتشار داده و پشتيباني مي كند.

تكنولوژي هدوپ از دو بخش كلي سيستم فايل انتشاري هدوپ يا HDFS[26] و همچنين تكنيك با كيفيت پردازي اطلاعات به نام نگاشت/كاهش[27] استفاده مي كند. ساختار كلي اطلاعاتي در هدوپ بدينگونه مي باشد كه اطلاعات توسط سيستم هدوپ شكسته شده و به چندين سرور فرستاده مي شود.سرورها بسته به نوع اطلاعات كه ممكن است پردازشي يا ذخيره اي باشد اطلاعات را پردازش يا ذخيره سازي مي كنند. در هنگام در خواست اطلاعات مجدد سيستم اطلاعات را از سرور هاي مختلف گرفته ، مونتاژ كرده و در خروجي نمايش مي دهد خوبي اين سيستم تهيه نسخه پشتيبان از اطلاعات بصورت خودكار است. هر تكه از اطلاعات در چندين سرور ذخيره مي شود و در صورت آسيب ديدن يكي از سرورها ، سرور ديگر قادر است مسؤوليت را بر عهده گرفته و اطلاعات مورد نظر را جايگزين كند.[14] .



شكل 16: مدل هدوپ [14]

[25]Hadoop
[26]Hadoop Distribition File System
[27]MapReduce

www.SID.ir

5- بررسي و مقايسه استراتژي ها

5-1- تجزيه و تحليل پيشگويانه[28] در مقابل استراتژي پيشگويانه[29]

مانند هر مفهوم جديدي، لغت تجزيه و تحليل پيش بينانه زمانيكه براي تجزيه تحليل و پيش بيني جرم بكار گرفته شد، مبهم بوده است. بسياري ازنمايندگي ها نا مطمئن هستند از اينكه كدام نرم افزار، از جمله ابزارهاي تجزيه و تحليل خودشان، واقعا پيشگو است و قادر به كمک به آنها براي به حداكثر رساندن بهره وری خود در پیشگیری از جرم و جنايت مي باشد. از آنجاييكه بودجه اجرای قانون كاهش يافته و يا ثابت مانده، چالش براي ادامه روند نزولی در ميزان جرم و جنايت با وجود چنين محدوديت های بودجه همچنان باقی مانده است. تحقيقات اخير نشان مي دهند كه 49% آژانس هاانتظار اين را دارند كه بودجه سال 2014 آنها ثابت باقي بماند. سوال برجسته بسياری از روسای پليس و كلانتری ها كه هنوز هم در حال دست و پنجه نرم كردن با پاسخ آن هستند، اين است: "نمايندگي ها چگونه مي توانند با بودجه كمتر، كاركنان كمتر و يا منابع كمتر، موثرتر و كاراتر باشند؟"

بين تجزيه و تحليل پيشگويانه و استراتژي پيشگويانه تفاوت وجود دارد.

"استراتژي پيشگويانه اشاره دارد به هر استراتژی پليس و يا تاكتيكي كه اطلاعات را توسعه داده و از آن استفاده مي كند و تجزيه و تحليل پيشرفته به منظور اطلاع از تفكر رو به جلو پيشگيري از جرم." با اين تعريف، استراتژي پيشگويانه مفهومي است كه سياست گذاري ها، استراتژي ها و يا تاكتيک ها، و تقريبا هر چيز ديگري را شامل مي شود از جمله:

- ✓ يک رويكرد استراتژيک يا يک ابتكار مانند استراتژي هوش پليسي
- ✓ يک سياست همچون استراتژي جامعه محور يا مسئله محور
- ✓ يک ابزار تجزيه و تحليل مانند Hot Spot يا تجزيه و تحليل روند خطي[30]

تجزيه و تحليل پيش گويانه، در مقابل، تركيبی از دو مفهوم ملموس مي باشد: پيش بينی و تجزيه و تحليل (الگوريتم های پيشرفته آماری رياضی). بنابراين تعريف كار زير را اتخاذ گرديد:

" تجزيه و تحليل پيش گويانه اشاره دارد به استفاده از ابزارهاي آماري مبتني بر رايانه به منظور كاوش داده هاي جرم و جنايت و رفتار مجرمانه براي ساخت يک پيش بيني در خصوص احتمالات جنايي آينده كه مشخص، به موقع، و قابل تعقيب قانوني مي باشند."

اين مهم است كه توجه داشته باشيد كه تجزيه و تحليل پيش بينی، خلاصه ای از داده های گذشته يا گزارشي از اينكه چه اتفاقي براي داده افتاده نيست. بلكه يک پيش بيني از جرايم آينده است، شامل مكان و زماني كه احتمال اتفاق جرم وجود دارد، پس قابل پيگرد قانوني مي باشد. بدين صورت پيش بيني شامل موارد زير مي شود:

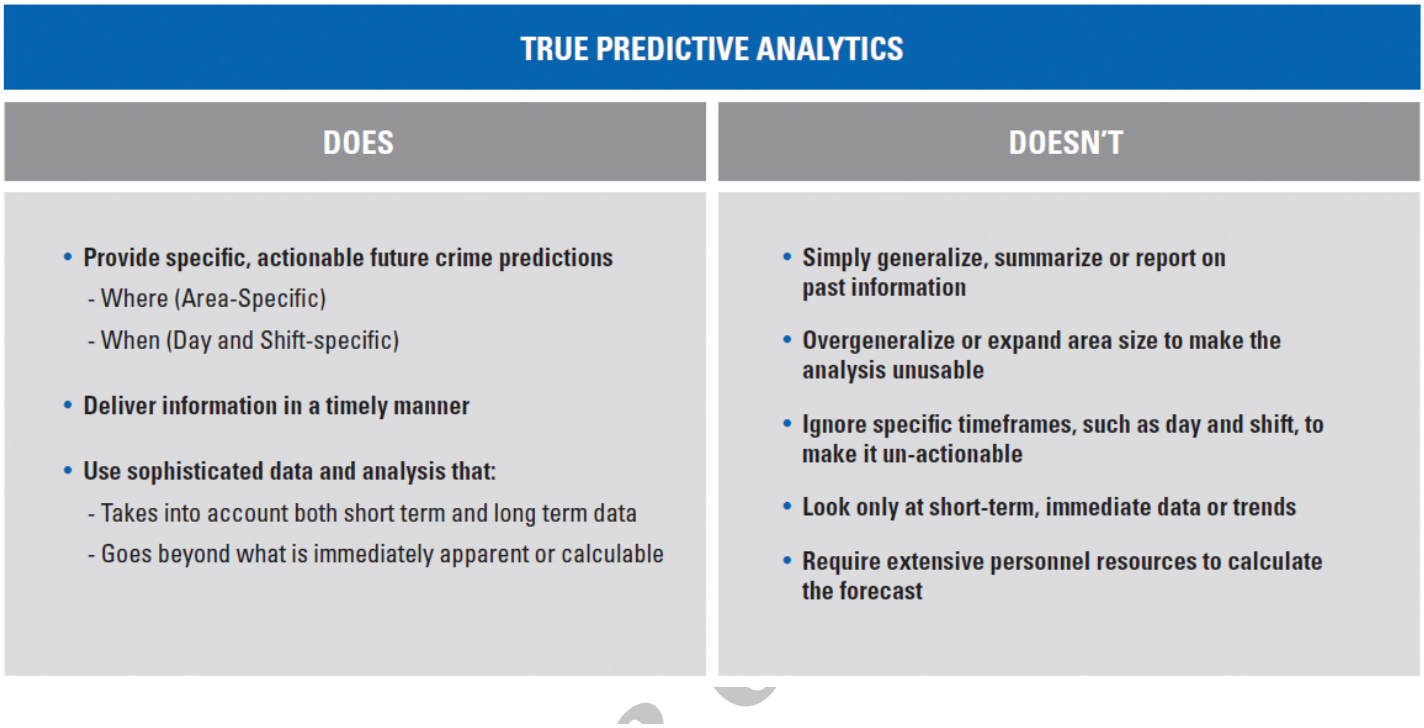
- ✓ از كجا (منطقه خاص)
- ✓ چه هنگام (روز و نوبت خاص)

[28]Predictive Analytics
[29]Predictive Policing
[30]Linear Trending Analysis

www.SID.ir

✓ اطلاعات اضافي (از قبيل انواع جنايات) براي توانمند سازي افسران به اقدام

شكل 17: بررسي پيش بينانه بودن نرم افزار تجزيه و تحليل [17]



در حال حاضر ابزار های بسیاری به عنوان ابزار پیش بینی به بازار عرضه شده اند. - چرا كه آنها از استراتژي پيش بينانه حمايت مي كنند، اما ويژگي هاي تجزيه و تحليل پيش بينانه از جمله: ارائه پیش بینی های جرم و جنایت كه خاص، به موقع، و قابل تعقيب قانوني باشند را ندارند. نمونه هایی از این ابزارهاي شبه پیش بینی عبارتند از: سري هاي زماني و تجزيه و تحليل روند خطي، "منطقه پیش بینی"[31] و "نگاه به نقاط آینده"[32]، جستجوي متنی و ارتباط آثار کلیدی، و مدل هاي ساده زماني/مكاني. مادامیکه این ابزارها در تجزیه و تحلیل جرم و جنایت مفید هستند، اغلب به صورت خيلي ساده خلاصه یا گزارشي روي داده های تاریخی انجام مي دهند یا اینکه فاقد دقت آماري زیاد در فراهم کردن تجزیه و تحلیل صحيح مي باشند[17].

[31]Predictive Zone
[32]Look ahead hotspots

www.SID.ir

نمونه هايي نوعي از ابزارهاي تجزيه و تحليل جرم و اينكه چرا براي تحليل پيش گويانه واجد شرايط نيستند:

FEATURES SUPPORTED						
	Specific Location	Specific Time	Factors in short and long term data	Exposes underlying data	Deliverable in timely manner	Provides actionable underlying
TRUE PREDICTIVE ANALYTICS	**YES**	**YES**	**YES**	**YES**	**YES**	**YES**
TYPICAL NON-PREDICTIVE CRIME ANALYTIC TOOLS						
Time Series	YES	NO	YES	NO	NO	NO
Trend Lines	YES	NO	YES	NO	NO	NO
Hot Spots	YES	NO	NO	NO	NO	NO
Textual Searching and Linking	YES	NO	NO	NO	NO	YES
Simplistic Geospatial/ Temporal Models	YES	NO	NO	NO	YES	NO

شكل 18: نمونه ابزارهاي تجزيه و تحليل جرم [17]

2-5- آمار و ارقام استفاده از داده‌های بزرگ

ابرداده‌ها در حال تغییر جهانی است که می‌شناسیم. اگر باور نمی‌کنید، آمارهای باورنکردنی زیر را بخوانید:

۶۰ درصد: بررسی‌ها نشان داده‌اند که خرده فروشان با استفاده از قدرت داده‌های بزرگ توانسته اند ابعاد فروش محصولات خود را تا ۶۰ درصد افزایش دهند [18].

۸ میلیون دلار: در مطالعه‌ای تجارت‌ها در سال جاری به طور میانگین ۸ میلیون دلار صرف پروژه‌های ابرداده‌ها و مربوط به داده‌های بزرگ کرده‌اند[19].

50/1 میلیارد دلار : داده‌های بزرگ که در سال ۲۰۱۴ ، 28/5 میلیارد دلار ارزشگذاری شده بود قرار است تا سال ۲۰۱۷ ارزش خود را به 50/1 میلیارد دلار برساند [20].

۵ برابر بیشتر : کمپانی‌هایی که تا به حال حداقل یک بار از قدرت تجزیه و تحلیل داده‌های بزرگ استفاده کرده اند، ۵ برابر بیشتر از دیگر کمپانی‌ها علاقه به تصمیم‌گیری سریع‌تر نسبت به کمپانی‌های رقیب خود دارند[21].

۹۰ درصد : نود درصد داده‌های موجود در جهان در مدت تنها دو سال اخیر ایجاد شده‌اند [22].

Archive

www.SID.ir

5-3- نتيجه گيري

سیستم های پیشرفته از ابزارهای پیشرفته برای پیش بینی جرایم استفاده می‌کنند. آنها همچنین از مقايسه هاي سال به سال یا ماه به ماه را برای تعیین اینکه آیا نرخ جرم و جنایت بالا یا پایین رفته است، استفاده مي كنند. شرکت های بزرگ مانند IBM، Dell، گوگل یا مایکروسافت، میلیاردها دلار را در بخش تحقیق و توسعه فن آوری ابرداده سرمایه گذاری مي كنند. توسعه دهندگان جوامع خودشان را براي فن آوری هایی مانند Hadoop، NoSQL، زمان واقعی[33]، یادگیری ماشینی[34] و غیره دارند. دستاوردهایی از نظر توسعه و گسترش فن آوری ها تا کنون به دست آمده است که به مبارزه با جنایات کمک می‌کند.

اما چالش هايي براي مواجهه وجود دارند. هنوز هم ابعاد بسیاری در حوزه بکارگيري اطلاعات وجود دارند که منتظر بهره برداري شدن مي باشند که به ثبت، شناسایی، پیش بینی و جلوگیری از جنایات کمک خواهند کرد. همچنین، با بهبود فن آوری، سازمان ها باید برای انطباق با راه های جدیدتر برخورد با جرم و جنایت آموزش ببینند، زیرا تکنولوژی به شدت سبب کاهش هزینه های عملیاتی و افزایش تحقیقات می‌شود[3].

مزاياي تجزيه و تحليل پيش بينانه پیشگیری از جرم مي باشد. هنگامی که معنای تجزیه و تحلیل پیش بینانه روشن شد، دیدن ارزش تجزیه و تحلیل پیش بینانه واقعی در ارائه اطلاعات به موقع، دقیق و قابل پیگرد قانوني بر روي جرايم آينده، آسان مي شود[17].

[33] Real Time
[34] Machine Learning

www.SID.ir

مراجع (*References*) (طبق استانداردIEEE)*

[1] Judith Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman, *Big Data for DUMMIES, A Wiley Brand*, John Wiley & Sons Inc., 2013. pp. 1-233, [E-book] Available: 115.28.209.159/bigdata/Big%20Data%20for%20Dummies.pdf

[2] Chris Eaton, Dirk Deroos, Tom Deutsch, George Lapis, Paul Zikopoulos. "What is Big Data? Hint: You're a Part of It Every Day" in *Understanding Big Data*, 1st ed. Paul Carlstroem, Steven Sit, Ed. New York: McGraw-Hill, 2012, pp. 31-50.

[3] Rishikesh C. Muchhala, Gaurang A. Shah, Prayag Shah. "Big Data to Combat Crime", *International Journal of Current Engineering and Technology*, Vol.4, No.5, pp. 1-4, Oct. 2014. [Online] Available: http://inpressco.com/category/ijcet

[4] Neil M. Richards, Jonathan H. King, "Big Data Ethics". Vol.49, pp. 407-408, May 2014. [Online] Available: http://ssrn.com/abstract=2384174

[5] Unisys, "Using big data analytics to fight financial crime", White Paper, pp. 1-10, 2011. [Online] Available: http://www.sourcingfocus.com/uploaded/documents/Unisys_-_Using_big_data_analytics_to_fight_financial_crime.pdf

[6] Hsinchun Chen, Roger H. L. Chiang, Veda C. Storey, "Business Intelligence And Analytics: From Big Data To Big Impact", MIS Quarterly, Vol. 36, No. 4, pp. 1165-1188, Dec. 2012. [Online] Available: http://misq.org/misq/downloads/download/editorial/567/

[7] Bernard Marr, "How Big Data Is Changing Healthcare", forbes.com, Apr. 21, 2015. [Online] Available: http://www.forbes.com/sites/bernardmarr/2015/04/21/how-big-data-is-changing-healthcare.

[8] Michael Schroeck, Rebecca Shockley, Dr. Janet Smart, Professor Dolores Romero-Morales and Professor Peter Tufano, "Analytics: The real-world use of big data", Executive Report, pp. 3-12, Oct. 2012. [Online] Available: https://www.ibm.com/smarterplanet/global/files/se__sv_se__intelligence__Analytics_-_The_real-world_use_of_big_data.pdf

[9] Cloud Security Alliance (CSA) Big Data Working Group Co. "Top Ten Big Data Security and Privacy Challenges". isaca.org, November 2012. [Online] Available: www.isaca.org/groups/professional.../big-data/.../big_data_top_ten_v1.pdf

[10] IBM Corporation Software Group, "Extending & Enhancing Law Enforcement Capabilities", 2011. [Online] Available: www.ibm.com/software/jstart/downloads/bd_LawEnforcement.pdf

[11] Executive Office of the President, "Big Data: Seizing Opportunities, Preserving Values", May 2014. [Online] Available: https://www.whitehouse.gov/.../big_data_privacy_report_may_1_2014.pdf

Archive of SID

www.SID.ir

[12] Bill Hamilton, “Big Data is the Future of Healthcare”, Cognizant 20-20 insights, USA, Sep, 2012. [online] Available: http://www.cognizant.com/InsightsWhitepapers/Big-Data-is-the-Future-of-Healthcare.pdf

[13] Arnold Picot, “From Open Data To Big Data”, 2nd International Open Data Dialog, pp. 1-34, Nov. 2013. [Online] Available: http://open-data.fokus.fraunhofer.de/wp-content/uploads/sites/3/2013/11/Picot_From-Open-Data-to-Big-Data.pdf

[14] ITU-T Technology Watch Report, “Big Data: Big Today, Normal Tomorrow”, 2nd International Open Data Dialog, pp. 20-25, Nov. 2013. [Online] Available http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000220001PDFE.pdf

[15] R. Abbas, “The Social Implications of Location-Based Services: An Observational Study of Users,” *J. Location-Based Services*, vol. 5, nos. 3-4, 2011, pp. 156-181.

[16] J. Pitt, “Design Contractualism for Pervasive/Affective Computing,” *IEEE Technology and Society Magazine*, vol. 31, no. 4, 2012, pp. 25-28.

[17] Motorola Solutions, “ Is My Crime Analytics Software Predictive?”, White Paper, pp. 1-5, Aug. 2015. [Online] Available: https://www.motorolasolutions.com/content/dam/msi/docs/products/smart-public-safety-solutions/ilps/Is-My-Crime-Analytics-Software-Predictive-White-Paper.pdf

[18] David Court, “ Getting Big Impact From Big Data”, 1996-2015 McKinsey & Company, Jan, 2015. [online] Availble: http://www.mckinsey.com/insights/business_technology/getting_big_impact_from_big_data?cid=other-eml-alt-mkq-mck-oth-1501

[19] C. Hagen, J. Raudabaugh, A. Prasad, Kh. Khan, “Beyond Big: The Analytically Powered Organization”, ATKearney, atkearney.com, Apr, 2014. [online] Available: https://www.atkearney.com/analytics/featured-article/-/asset_publisher/FNSUwH9BGQyt/content/beyond-big-the-analytically-powered-organization/10192

[20] Jeff Kelly, “Big Data Vendor Revenue and Market Forecast 2013-2017”, Wikibon professional community, wikibon.org, Feb. 12, 2014, [online] Available: http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017

[21] Louis Columbus, "Roundup Of Analytics, Big Data & Business Intelligece Forecasts And Market Estimates,2014”, forbes.com, Jun. 25, 2014, [online] Available: http://www.forbes.com/sites/louiscolumbus/2014/06/24/roundup-of-analytics-big-data-business-intelligence-forecasts-and-market-estimates-2014/

[22] Dennis McCafferty, “Surprising Statistics About Big Data”, BaseLine Magazine, Driving Business Success With Technology, baselinemag.com, Feb. 02, 2014, [online] Available:

Archive of SID

www.SID.ir

http://www.baselinemag.com/analytics-big-data/slideshows/surprising-statistics-about-big-data.html

Archive of SID

www.SID.ir