



## داده های بزرگ، مروری بر فرصت ها، کاربردها و ابزارهای مورد استفاده

ناصر اصفهانی پور<sup>۱</sup>، مهدی جوانمرد<sup>۲</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، صندوق پستی ۳۶۹۷-۱۹۳۹۵ تهران، ایران

[esfahani.naser@gmail.com](mailto:esfahani.naser@gmail.com)

<sup>۲</sup> استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، صندوق پستی ۳۶۹۷-۱۹۳۹۵ تهران، ایران [javanmard@pnu.ac.ir](mailto:javanmard@pnu.ac.ir)

### چکیده

امروزه بسیاری از صنایع و دولت می توانند از داده های بزرگ برای حصول به بینشی ارزشمند بهره ببرند. چنین بینشی می تواند به تصمیم گیرندگان کمک کند تا راهبردهای خود را بهبود بخشیده و برنامه های خود را بهینه سازی کنند. سازمان ها با استفاده از داده های بزرگ می توانند به مزیت رقابتی دست یافته و در بسیاری از بخش های اقتصادی و اجتماعی ارزش افزوده داشته باشد. در حقیقت، برخی از دولت ها برنامه هایی با بودجه ی زیاد آغاز کرده اند تا توسعه و تحقیقات در زمینه ی داده های بزرگ را به پیش ببرند. بخش خصوصی نیز سرمایه گذاری های زیادی برای به حداکثر رساندن سود و بهینه سازی منابع انجام داده است. در این مقاله، به بررسی چند پروژه در ارتباط با داده های بزرگ و فرصت ها، نمونه ها و مدل های مرتبط با آن در بسیاری از بخش ها از جمله سلامت، تجارت، گردشگری و سیاست می پردازیم. این مقاله همچنین نمونه هایی از فن آوری ها و راه کارهای مقابله با مسائل داده های بزرگ را ارائه می دهد.

**کلمات کلیدی:** داده های بزرگ، فرصت های داده های بزرگ، کاربردهای داده های بزرگ، فن آوری های داده های بزرگ.

### مقدمه

در حال حاضر بسیاری از صنایع و دولت ها از ارزش داده های بزرگ آگاهی دارند. در واقع، استخراج کارآمد داده های بزرگ باعث بهتر شدن مزیت رقابتی و ارزش افزوده برای بسیاری از بخش ها (اقتصادی، اجتماعی، پزشکی، تحقیقاتی و غیره) می شود. داده های بزرگ بطور اساسی توسط سه ویژگی اساسی<sup>۱</sup> خود مشخص می شوند که عبارتند از سرعت (داده ها با سرعت بالایی افزایش و تغییر می یابند)، گوناگونی (داده ها دارای انواع قالب های متفاوتی هستند) و حجم (مقدار زیادی از داده ها در هر ثانیه تولید می شوند). این سه ویژگی باید همزمان وجود داشته باشند تا اثبات شود یک منبع از نوع منبع داده های بزرگ است. اگر یکی از این سه ویژگی وجود نداشته باشند نمی توان درباره ی داده های بزرگ بحث کرد. برای بیان بهتر داده های بزرگ، ویژگی های دیگری نیز توسط برخی از بازیگران معرفی شده اند از جمله بینایی (یعنی، هدف مشخص داده کاوی های بزرگ)، تایید (یعنی، تطابق داده های پردازش شده با برخی از مشخصات)، اعتبارسنجی (یعنی برآورده شدن هدف مورد نظر)، ارزش (یعنی، حصول به مزایای بسیاری از بخش ها از طریق استخراج اطلاعات مربوطه)، پیچیدگی (سازمان دهی و تجزیه و تحلیل داده های بزرگ بدلیل تکامل روابط مربوط به داده ها دشوار است) و تغییر ناپذیری (با مدیریت مناسب داده های گردآوری و ذخیره سازی شده می توان آن ها را بصورت دائمی حفظ کرد). علاوه بر این، برخی افراد هنگام تعریف داده های بزرگ اشاره می کنند که هر مقدار عظیمی از داده های دیجیتالی که دیگر نمی توانیم از طریق زیرساخت ها و فن آوری های موجود گردآوری و پردازش کنیم ذاتا از نوع داده های بزرگ هستند. این مقاله پروژه ها، فرصت ها و مدل هایی از داده های بزرگ را در ارتباط با بخش های مختلف از جمله سلامت، تحقیقات، تجارت، حمل و نقل، گردشگری، سیاست ارائه می دهد. این مقاله برخی از فن آوری های بکار رفته برای کاربردهای داده های بزرگ را نیز ارائه می دهد.

### ۱. فرصت های داده های بزرگ

بسیاری از محققان فرصت های مرتبط با داده های بزرگ را در زمینه ی مربوط به خود مشخص کرده اند. در این بخش، فرصت ها و مزایای داده های بزرگ را در بخش های زیر بصورت خلاصه بیان می کنیم: سلامت، تحقیقات، تجارت، حمل و نقل، گردشگری، سیاست. هدف ما ارائه ی تصویری بزرگ از این موضوع است.



#### الف) بخش سلامت

تحلیل داده‌های بزرگ بینش ارزشمندی برای بخش سلامت در پی دارد. در حقیقت، چند کاربرد از داده‌های بزرگ برای بهبود خدمات پزشکی خصوصی و دولتی و حمایت بهتر از بیمار و کارکنان بخش پزشکی بررسی شده است. تحلیل داده‌های بزرگ می‌تواند با بهینه‌سازی خدمات عملیاتی، ارائه‌ی ابزارهای پشتیبانی از تصمیم و با کاهش هزینه‌ی بالای این بخش باعث تحولی عظیم در زمینه‌ی سلامت شوند که بصورت زیر به این موارد اشاره می‌شود.

- **بهینه‌سازی خدمات و هزینه‌های سلامت:** تحلیل داده‌های بزرگ به سازمان‌ها کمک می‌کند تا مشخص شود کدام بخش نیاز به سازمان‌دهی مجدد دارد. این تحلیل کمک می‌کند تا ارزیابی و نظارت فوری بر کیفیت خدمات، عملکرد بخش‌های پزشکی و نیازهای منابع انسانی و تجهیزات پزشکی داشت.
- **درک بهتر از پیشروی بیماری:** تجزیه و تحلیل اطلاعات روی منابع بزرگ اطلاعاتی در مورد ویروس و دی‌ان‌ای‌ها می‌تواند به درک بیماری‌ها کمک کند. چنین دانشی به پزشکان و محققان کمک می‌کند تا شیوه‌های پیشگیری جدیدی را برای بیماری‌های ارثی و ژنتیکی پیدا کنند.
- **پشتیبانی از تصمیم‌گیران پزشکی:** به عنوان مثال، می‌توان تجزیه و تحلیل روی جراحی‌های قبلی در پرونده‌ی بیماران را با تجزیه و تحلیل علائم فعلی یا سوابق بیمار ترکیب شود. چنین همبستگی کمک می‌کند تا بهترین روش‌ها و مداخلات را براساس مشخصات بیمار پیدا کرد.
- **پیشگیری بهتر:** مدل‌های پیش‌بین داده‌های بزرگ می‌توانند داده‌های بالینی در بخش‌های خصوصی و دولتی را برای جلوگیری از گسترش بیماری‌های واگیردار تحلیل کنند. این مدل‌ها براساس تشخیص علائم هشدار دهنده در سلامت جامعه است. با توجه به مناطقی که بیماری در آنجا سرایت کرده است و همچنین علائم جمعیت، تصمیم‌گیران می‌توانند برنامه‌ی پیشگیری موثری ایجاد کرده و از گسترش بیماری‌های واگیردار جلوگیری کنند.
- **سفرهای سازی خدمات پزشکی:** برخی از پروژه‌های پزشکی به منظور بهبود رضایت بیماران، بازخورد بیماران را در زمان واقعی جمع‌آوری و تحلیل می‌کنند. به عنوان مثال، داده‌های پزشکی در زمان واقعی این امکان را فراهم می‌کنند تا شرایط بیماران را برای تطبیق دوز دارو یا ارائه‌ی توصیه‌های پزشکی بر اساس علائم تحلیلی بررسی کرد. بعضی از پروژه‌ها از حس‌گرهای هوشمند استفاده می‌کنند که به گوشی‌های هوشمند یا گلوکومترها وصل شده‌اند. هدف از این کار نظارت آنلاین بر علائم بیمار (سطح قند خون، ضربان قلب، و غیره) است. اطلاع‌رسانی اضطراری و اطلاعات در مورد علائم بلافاصله به پزشک ارسال می‌شود تا بتواند طبق علائم جدید بیمار دارو تجویز کند. بطور کلی، تجزیه و تحلیل داده‌های پزشکی باعث بهبود زندگی بیماران می‌شود و در عین حال بینش ارزشمندی در مورد تاثیرگذاری درمان‌ها و جراحی‌ها در اختیار پزشکان قرار می‌دهد.

یکی از مهم‌ترین چالش‌های داده‌های بزرگ در بخش سلامت یکپارچه‌سازی منابع مجزا و مختلف است. برای انجام این کار، همکاری بین تمام نهادهای پزشکی (یعنی بخش‌های دولتی و خصوصی، پزشکان و کارکنان بخش سلامت) لازم است. چالش دیگر در رابطه با اتصال و یکپارچه‌سازی چند سیستم پزشکی به همدیگر است. این یکپارچه‌سازی نیاز به مدیریت ایمن پرونده‌های پزشکی و احترام به حریم خصوصی بیماران مطابق بر قوانین و مقررات موجود است. چنین ICTهایی حاکی از مدیریت دسترسی گران‌تری به داده‌های پزشکی نیز هستند که در عین حال باید دسترسی سریع به حجم زیادی از داده‌ها و مصورسازی کارآمد نتایج تجزیه و تحلیل داده‌های بزرگ تضمین شود.

#### ب) بخش تجارت

در صورت استفاده و مدیریت مناسب داده‌ها، آن‌ها دارایی ملموسی برای شرکت‌ها هستند. با توجه به فن‌آوری‌های پیشرفته‌ی داده‌های بزرگ، می‌توان بطور موثری یکپارچه‌سازی و تجزیه و تحلیل منابع داده‌های بزرگ و متنوع (مانند داده‌های داخلی تولید شده از فرایندهای داخل سازمانی، داده‌های خارجی گردآوری شده از منابع عمومی، وبسایت‌ها و اطلاعات خریداری شده از شرکت‌های دیگر) انجام داد. تجزیه و تحلیل داده‌های بزرگ در بخش تجارت استخراج بینش ارزشمندی را میسر می‌سازد. این تجزیه و تحلیل درکی بهتر از رفتار و ترجیحات مشتریان فراهم می‌کند. از این تحلیل‌ها برای آزمایش بازدهی راهبردهای تجاری نیز استفاده می‌شود. با داشتن این دانش، سازمان‌ها می‌توانند



محصولات، خدمات و راهبردهای خود را (مانند تبلیغات هدفمند در زمان واقعی) تطبیق و بهبود دهند. بنابراین، این تحلیل‌ها باعث افزایش رضایت مشتری، افزایش سود و افزایش رقابت می‌شوند. بطور کلی، کاوش داده‌های بزرگ باعث نظارت بهتر بر اقتصاد کلان و کوچک می‌شود و در دستیابی به فرصت‌های تجاری و پیش‌بینی رکود به تصمیم‌گیران کمک می‌کند.

به عنوان مثال، فیسبوک، گوگل و آمازون اطلاعات مربوط به رفتار کاربران وب، نظرات، بازخوردها، نظرات و معاملات آنلاین را گردآوری و می‌فروشند. شرکت‌های کارت اعتباری (مانند Equifax و Transunion) همین کار را برای بالا بردن سود و افزایش خدمات خود انجام می‌دهند. علاوه بر این، تکثیر چند ICT و ارتباط بسیار خوبی بین بسیاری از سازمان‌ها (مثلا شرکت‌های تابعه، شرکا، تامین‌کنندگان و مشتریان آنلاین) بوجود می‌آورند، مدل‌های کسب و کار جدیدی را براساس به اشتراک‌گذاری زمان واقعی و کسب درآمد در داده‌های بزرگ ارائه می‌دهند. در واقع، همانطور که در مرجع اشاره شد، شرکت‌ها می‌توانند طی چند مرحله و با توجه به سطح بلوغ خود از داده‌های بزرگ بهره ببرند:

- شرکت‌ها بجای اینکه فقط بر فرآیندهای کسب و کار داخلی خود نظارت داشته باشند می‌توانند از نتایج تجزیه و تحلیل داده‌های بزرگ برای درک رفتار مشتریان و افزایش راهبردهای تجاری و محصولات خود بهره‌برداری کنند.
- با حصول به بلوغی خاص، شرکت‌ها با بهینه‌سازی فرآیندهای خود و شناسایی فرصت‌های تجاری تلاش می‌کنند عملیات خود را بهتر کنند.
- سطح بلوغ بعدی زمانی حاصل می‌شود که سازمان بتواند نه تنها مدل کسب و کار خود را بهینه کند بلکه از داده‌های بزرگ گردآوری شده کسب درآمد کند. می‌توان این کار را با ایجاد سود بیش‌تر از طریق فروش مجدد داده‌ها و نتایج تجزیه و تحلیل انجام داد. روش دیگر این است که از بینش داده‌های بزرگ برای افزایش محصولات و تجربه‌ی مشتریان در فروشگاه‌ها و بصورت آنلاین استفاده کند.
- هرگاه شرکتی به سطح بالایی از بلوغ برسد، می‌داند که دیگر نمی‌تواند بدون توسعه‌ی اکوسیستم خود رشد کند. در این مرحله، سازمان می‌تواند بر مدل تجاری تغییر یافته و جدید بر مبنای به اشتراک‌گذاری داده‌ها و ارتباط بسیار خوب بین همه‌ی طرف‌ها (به عنوان مثال، شرکت‌ها، تامین‌کنندگان، مشتریان، شرکا و اجرای قانون آژانس) تکیه کند. راه‌حل یک پلت‌فرم داده‌های بزرگ مشترک است که باعث می‌شود تمامی طرف‌ها سود خود را به حداکثر برسانند و تجربه‌ی خود را افزایش دهند. چنین پلت‌فرم مشترک می‌تواند سازگاری با قوانین و مقررات را حفظ کند. با دسترسی مبتنی بر نقش‌گرانشی، ICT باید احترام به سیاست‌های امنیتی و حفظ حریم خصوصی تمامی نهادها تضمین کند.

(۱) نمونه‌ی وال‌مارت: وال‌مارت یک شرکت خرده‌فروشی چند ملیتی دارای فروشگاه‌های زنجیره‌ای تخفیف‌دار است. این مثال نمونه‌ی خوبی از مدل کسب و کار تغییر یافته است که از طریق بهره‌برداری از داده‌های بزرگ حاصل شده است. قدرت سامانه‌ی آن بخاطر تجزیه و تحلیل زمان-واقعی و به اشتراک‌گذاری داده‌های جمع‌آوری شده از منبع مختلف زیر است: داده‌های داخلی (مانند اطلاعات مالی، مشتریان، ترافیک، ورود یا محصولات) و داده‌های خارجی (مثلا رفتار مشتریان، نظرات در فیسبوک، داده‌ها تلفن‌های همراه، ایمیل‌ها و یا کلیک کاربران روی وبسایت‌ها).

پلت‌فرم وال‌مارت مبتنی بر اکوسیستم جامع داده‌های بزرگ برای بازاریابی آنلاین با بیش از ۲۰۰ گره هادوپ و موتور تجزیه و تحلیل قدرتمند است. برای فائق آمدن بر جریان داده‌هایی که با سرعت بالایی در حال افزایش بوده و پردازش می‌شوند، وال‌مارت اجزای خود را توسعه داد و سپس (مانند Muppet که ابزاری است که برای پردازش و تجزیه و تحلیل داده‌های بزرگ در زمان-واقعی روی خوشه‌های متعدد طراحی شده است و Thorax که چارچوب طراحی شده برای ساخت اپلیکیشن‌های وب در مقیاس بزرگ است) آن‌ها را بصورت منبع باز درآورد. وال‌مارت به کاوش نظرات و رفتار کاربران آنلاین می‌پردازد تا محصول مورد علاقه‌ی کاربران را ارائه دهد. برای داشتن تجربه‌ی خوب خرید، وال‌مارت گردش در داخل مغازه را از طریق گوشی موبایل برای مشتریانی فراهم می‌آورد که از گوشی‌های هوشمند استفاده می‌کنند. به گفته‌ی مدیر وال‌مارت: «مشتریان می‌توانند به دنبال تخفیف‌ها بگردند، قیمت‌ها را مقایسه کنند و در راهروهای فروشگاه جستجو کنند». همچنین، با بررسی ارزش داده‌های بزرگ، وال‌مارت قدرت تاثیرگذاری بر تصمیمات شرکا برای تطبیق قیمت‌ها، افزایش سود و کاهش هزینه‌ها را دارد. وال‌مارت سود خود را از طریق به اشتراک‌گذاری داده‌های بزرگ و درآمدزایی افزایش می‌دهد.



بهره‌برداری از داده‌های بزرگ تولیدی از بخش کشاورزی می‌تواند بینش ارزشمندی را به ارمغان بیاورد. چنین بینشی برای بهینه‌سازی راهبردهای تولید، تطبیق برنامه‌ها با توجه به پیش‌بینی وضعیت آب و هوا، نظارت بر تقاضای منطقه و مشخصات مشتریان و موارد دیگر کاربرد دارد. می‌توان این کار را با تجزیه و تحلیل منابع نامتجانس و چندگانه انجام داد (به عنوان مثال، شرایط و تاریخ، پیش‌بینی تقاضا و یا حس‌گرهای هوشمند). به عنوان مثال، با توجه به مرجع، هدف پروژه‌ی ژاپنی ایجاد سیستم پیشنهادی پیشرفته بر اساس تجزیه و تحلیل داده‌های بزرگ است. هدف از انجام این کار توصیه‌ی بهترین ترکیب از محصولات، رستوران‌ها و تولیدکنندگان برای کاربران آنلاین است که این محصولات را مطابق بر ترجیحات مشتریان (مانند، محصولات بیولوژیک یا محصولات بدون آلرژیک) ارائه می‌دهد. این سیستم همچنین محصولات مناسب و تامین‌کنندگان آن‌ها را تنها با پر کردن علائم بیمار توصیه می‌کند. هدف این سیستم ایجاد بهم‌پیوستگی در میان چند نهاد (کاربران، رستوران‌ها یا تولیدکنندگان) و تضمین دسترسی گران‌تری به داده‌ها براساس مشخصات است.

#### د) بخش گردشگری

چند مدل داده‌های بزرگ ایجاد شده است یا در حال ساخت هستند تا فعالیت‌های گردشگری را ارتقا داد و خدمات بهتری در اختیار گردشگران قرار داد. در واقع، شرکت‌ها می‌توانند از فن‌آوری‌های داده‌های بزرگ برای استخراج بینش ارزشمند مانند درک بهتر رفتارهای گردشگران، تعیین ترجیحات و نیازهای در حال رشد، نظارت بر مکان جغرافیایی، فعالیت‌ها و زمینه‌ی گردشگران استفاده کنند. به عنوان مثال، می‌توان بصورت فوری هتل‌ها، رستوران‌ها و فعالیت‌هایی را با توجه به ترجیحات، رفتار آنلاین و موقعیت‌های جغرافیایی گردشگران در اختیار آن‌ها قرار داد. با توجه به این امر، سیستم پیشنهادی گردشگری توسط مرجع ارائه شد. این سیستم بر مبنای ترکیبی از ابزارهای جامع تحلیل و مصورسازی داده‌های بزرگ است (از جمله: ۱) تجزیه و تحلیل الگوهای قبلی فعالیت‌های گردشگر، ۲) تجزیه و تحلیل زمان-واقعی فعالیت‌ها، ترجیحات، مشخصات و بازدید فعلی گردشگر از وبسایت‌ها، ۳) ردیابی موقعیت جغرافیایی گردشگر، ۴) نظارت بر پارامترهای دیگر مانند شرایط آب و هوایی و شدت ترافیک. هدف این است که توصیه‌ی شخصی در زمان-واقعی ایجاد کرد. علاوه بر این، برخی از اپلیکیشن‌های داده‌های بزرگ نیز امکان نظارت بر کارایی شیوه‌ها و مقررات را در این بخش فراهم آورده‌اند. در واقع، این اپلیکیشن‌ها توصیه‌هایی برای بهبود ترافیک شهری یا دسترسی آسان به مکان‌های شلوغ مانند فرودگاه‌ها دارند. آن‌ها همچنین به مسیریابی رانندگان با توجه به موقعیت فعلی، مقصد، زمینه، ترجیحات و سایر پارامترهای در حال تغییر (مانند آب و هوا و یا تصادفات) کمک می‌کنند. برای مثال، مدل‌های مبنی بر داده‌های بزرگ برای رانندگان تاکسی توسط مرجع [۱۰] پیشنهاد شده است. هدف آن مدل‌سازی تصمیم‌گیری رانندگان تاکسی برای توصیه‌های بهتر است که می‌تواند خدماتی در اختیار بسیاری از نهادها (مدیران ترافیک فرودگاه، سیاست‌گذاران و یا رانندگان تاکسی) قرار دهد.

#### ه) بخش سیاست

بسیاری از دولت‌ها منابع مختلف اطلاعاتی را در زمان-واقعی و بصورت در حال حرکت یا استاتیک تحلیل می‌کنند (مانند ورودها (لاگ‌ها)، رویدادهای تاریخی، دوربین‌های نظارت دولتی و خصوصی، نظرات شهروندان در شبکه‌های اجتماعی، معاملات آنلاین، داده‌های GPS و ارتباطات تلفن همراه). آن‌ها همچنین از داده‌های تولید شده توسط بسیاری از ICT‌های دولتی نیز استفاده می‌کنند. در اینجا هدف دستیابی به بینش ارزشمند، الگوها و همبستگی‌ها یا ایجاد مدل‌های پیش‌بین است که این امکان را برای دولت فراهم می‌کند که راهبردهای خود را بهینه‌سازی کرده و خدمات شهری شهروندان را بهتر کنند. یکی دیگر از اهداف مهم، اطمینان از کنترل و نظارت پیوسته برای محافظت از شهروندان و کاهش جرایم است. برای مثال، دولت می‌تواند از الگوریتم‌های پیشرفته‌ای روی داده‌های بزرگ برای پیش‌بینی رویدادهایی استفاده کند که ممکن است بر امنیت کشور یا شناسایی مظنونین، سازمان جنایی و تروریست‌ها تأثیر بگذارد. با این وجود، این‌گونه نظارت بر ارتباطات و رفتار شهروندان باعث بوجود آمدن مسائلی در رابطه با حریم خصوصی شده است که قابل حل نیست. این مسائل در بخش ۳ مورد بحث قرار خواهد گرفت.

علاوه بر این، دانشمندان و کارشناسان سیاسی ممکن است از تحلیل داده‌های بزرگ برای استخراج دانش ارزشمند استفاده کنند. چنین بینشی درک بهتر مسائل سیاسی را امکان‌پذیر کرده است. به عنوان مثال، اپلیکیشن داده‌های بزرگ را بر اساس تجزیه و تحلیل جغرافیایی سیاسی ارائه داده است. این اپلیکیشن، دیدگاه‌های سیاسی اوباما، رئیس‌جمهور آمریکا را برای مدت زمان مشخصی ارزیابی می‌کند. این اپلیکیشن سخنان اوباما را از وبسایت کاخ سفید گردآوری می‌کند، سپس داده‌ها را برای انطباق پاک می‌کند و مجموعه داده‌های مربوطه را برای کاربرد موردی استخراج می‌کند. این اپلیکیشن از تکنیک‌های داده‌کاوی برای ارزیابی تمرکز رییس‌جمهور روی مسائل سیاسی استفاده می‌کند تا احساسات رییس‌جمهور را تحلیل و اختیارات رییس‌جمهور را در مورد رویدادهای مهم سیاسی بررسی می‌کند. می‌توان مدل پیشنهادی را به منظور تشخیص الگوهای



سیاسی، پیش‌بینی تاثیر انتخابات بر تکامل کشور، سنجش عقاید سیاسی، نظارت بر اهداف سیاسی، دسترسی به اعتماد شهروندان نسبت به وضعیت سیاسی فعلی و غیره استخراج کرد.

## ۲. فن‌آوری‌های داده‌ها بزرگ

فن‌آوری‌ها و روش‌های داده‌ای متعارف اکثر اوقات کُند و گران‌قیمت هستند و برای کار روی ذخیره‌سازی و پردازش حجم زیاد و رو به رشد داده‌های نامتجانس مناسب نیستند. یکی از چالش‌ها نحوه‌ی فائق آمدن بر ماهیت پیچیده‌ی داده‌های بزرگ (حجم، سرعت و تنوع) است. چالش دیگر نحوه‌ی نمایش موثر «بینش بزرگ» در حال تغییر برای بسیاری از نهادهای مربوطه بسته به نقش آن‌هاست. علاوه بر این، بازیگران داده‌های بزرگ هنگام تامین داده‌های بزرگ با چالش‌های دیگری روبرو هستند مثلاً نحوه‌ی تامین مجموعه داده‌های بزرگ که در حال تکامل هستند، نحوه‌ی یکپارچه‌سازی لایه‌های امنیتی بدون تاثیر گذاشتن بر عملکرد سیستم‌ها، نحوه‌ی یکپارچه‌سازی سیاست‌های حریم خصوصی در پلت‌فرم‌های مشترک داده‌های بزرگ در حین تامین دسترسی سریع و گران‌ثی به داده‌ها.

جوامع تحقیقاتی در بخش‌های مختلف در تلاش برای ایجاد فن‌آوری‌های داده‌های بزرگ جدید، سریع، پویا و کاربر-دوست هستند. امروزه، بسیاری از راه‌کارهای داده‌های بزرگ منبع باز و اختصاصی موجود هستند. هدف در اینجا کمک به تصمیم‌گیران و دانشمندان داده برای اقدامات بعدی در مبنای الگوهای کشف شده، روابط داده‌ها و دانش کشف شده از داده‌های بزرگ است.

در ادامه برخی از راه‌کارهای ایجاد شده برای رفع چالش‌های داده‌های بزرگ در سطوح مختلف را ارائه می‌دهیم.

### الف) چارچوب‌ها و پلت‌فرم‌های داده‌های بزرگ

چند چارچوب MapReduce (به عنوان مثال، Apache Hadoop، Skynet، Sailfish و FileMap) برای کار روی داده‌های حجیم ساختاریافته و بدون ساختار ایجاد شده‌اند. در واقع، آن‌ها امکان ذخیره‌سازی و پردازش روی حجم زیادی از داده‌های تغییرناپذیر (مانند لاگ‌ها یا اشیاء بزرگ باینری) و همچنین داده‌هایی که بصورت افزایشی گردآوری شده‌اند (مانند خزنده‌های وب، نظرات کاربران در شبکه‌های اجتماعی، داده‌های GPS یا حوادث حس‌گرها) فراهم می‌کنند. آن‌ها در موارد بسیاری (مانند تجزیه و تحلیل فایل‌های ورود، شبیه‌سازی‌های علمی یا پیش‌بینی‌های مالی) کارا هستند. چنین چارچوب‌هایی بر پایه‌ی بسیاری از مفاهیم جامع شامل موارد زیر هستند:

- **ذخیره‌سازی توزیع‌شده:** اکثر پلت‌فرم‌های داده‌های بزرگ، مانند Hadoop و Disco، بر اساس ذخیره‌سازی داده‌های توزیع شده هستند. بر خلاف سیستم‌های قدیمی، آن‌ها بلوک‌هایی از فایل‌های بسیار بزرگ را در گره‌های مختلف ذخیره می‌کنند. آن‌ها برای اجرای بر روی سخت‌افزار ارزان‌قیمت و فراهم کردن دسترسی سریع به مجموعه داده‌ها طراحی شده‌اند.
- **پردازش همزمان و زیاد (MPP):** چند وظیفه‌ی زمان‌بر اپلیکیشن‌های داده‌های بزرگ بصورت همزمان در چند سرور پردازش می‌شوند. MPP به منظور اجرای محاسبات از کپی شدن داده‌های دور جلوگیری می‌کند. MPP وظایفی را انجام می‌دهد که در آنجا داده‌ها ذخیره شده‌اند تا تراکم شبکه را کمینه کرد و از پردازش سریع اطمینان حاصل کرد.
- **سیستم مقاوم در برابر خطا و مقیاس‌پذیر:** برای اطمینان از نداشتن هیچ نقطه‌ی نقص‌دار، سیستم‌های داده‌های بزرگ معمولاً بر اساس معماری قابل اعتماد ارباب-برده و تکرار داده‌ها (مانند Hadoop) یا مفهوم بازیابی همکار (مانند Skynet [۱۶]) استفاده می‌کنند. اکثر راه‌کارها احتمال افزودن خوشه‌ها و مولفه‌ها را برای کار روی داده‌های بیش‌تر و پردازش زیاد فراهم می‌کنند.

بیش‌تر مواقع، مولفه‌های آزاد و اختصاصی روی پلت‌فرم‌های داده‌های بزرگ منبع باز مانند پلت‌فرم داده‌های هورتون‌ورکس (HDP) و IBM InfoSphere BigInsights ساخته می‌شوند. در اینجا هدف ارائه‌ی راه‌کارهای اکوسیستم کامل داده‌های بزرگ است. در حقیقت، آن‌ها قصد دارند از چارچوب‌های MapReduce استفاده کنند تا کاربرد آن را ساده‌تر کنند و قابلیت‌هایشان را افزایش دهند. بسیاری از راه‌کارها و ابزارهای داده‌های بزرگ موجود هستند که در ادامه به آن‌ها اشاره می‌شود:

- ابزارهایی برای یکپارچه‌سازی داده‌های بزرگ (مانند Sqoop، Flume و DataLoader)؛



- ابزارهایی برای مدیریت منابع، گردش کار و خدمات (مانند Zookeeper و Oozie, Yarn, Cloudera Manager)
- ابزارهایی برای کار روی خدمات فراداده (مانند Hcatalog)
- ابزارهایی برای تجزیه و تحلیل داده‌ها (مانند زبان برنامه نویسی Mahout, R, اکوسیستم تحلیلی چاکوا و ترادا)
- ابزارهای جستجوی تعاملی و جستار بومی (مانند جستجوی Cloudera, Sphinx Search Server, و فیسبوک یونی کورن که یک سیستم نمایه‌سازی آنلاین در حافظه است و برای جستجوی تریلیاردها لبه بین ده‌ها میلیارد نهاد روی هزاران سرور کالا طراحی شده است)
- ابزارهایی برای مصورسازی داده‌ها (همانند Advisor, ویزوال آنالیتیکس و سانترفیوژ)

با این فناوری‌ها، مدیران می‌توانند به راه‌کارهای سریع، مقرون بصره، مقاوم در برابر خطا، مقیاس‌پذیر و کاربر-دوست تکیه کنند. مدیران می‌توانند راه‌کار کاملی را انتخاب کنند یا مولفه‌های لازم را به زیرساخت‌های موجود اضافه کنند. بای نمونه، Oracle Big Data Appliance توان سخت‌افزار استانداردهای شبکه‌ی بهینه‌سازی شده را در یک سیستم ترکیب می‌کند، نرم‌افزار Oracle همچنین مزایای مولفه‌های منبع باز هادوپ‌های آپاچی را نیز تجربه می‌کند.

ب) پایگاه‌داده‌هایی برای داده‌های بزرگ

۱) NoSQL: RDMS (سیستم منطقی مدیریت پایگاه داده) مستلزم آن است که داده‌ها در قالب‌های مشخص ایجاد شوند که در زمینه‌ی داده‌های بزرگ که حجم زیاد و سریعی از داده‌های بدون ساختار ایجاد می‌شوند کافی نیست. برای روبرو شدن با این چالش، چند پایگاه‌داده‌ی NoSQL ایجاد شده‌اند تا روی داده‌های غیر منطقی و بدون ساختار مانند HBase, Cassandra, DynamoDB, Riak, MongoDB, Couchbase, Accumulo, Redis, و غیره کار کرد. این پایگاه داده‌ها از یک یا چند مدل از جمله جفت ارزش-کلیدی، گراف‌های سند-محور (همانند اسناد JSON, BSON, XML, HTML)، که برای داده‌های با اتصال قوی طراحی شده‌اند، ستون‌های عریض و داده‌های جغرافیایی-فضایی پشتیبانی می‌کنند.

پایگاه‌داده‌های NoSQL روشی ارزان‌قیمت (نسبت به RDMS) برای کار روی ذخیره‌سازی و مدیریت داده‌های بزرگ در محیط توزیعی ارائه می‌دهد. این پایگاه‌داده‌ها سطوح مختلفی از تحمل‌پذیری خطا و دسترس‌پذیری داده‌ها را ارائه می‌دهند.

۲) NewSQL: پایگاه‌داده‌های NoSQL معایب زیادی دارند که باعث ایجاد پایگاه‌داده‌های NewSQL شده است. در حقیقت، پایگاه‌داده‌های NoSQL از نمایه‌سازی و جستار SQL پشتیبانی نمی‌کنند. همچنین، برخلاف RDBMS، آن‌ها اصول ACID را برای مبادلات قابل اعتماد تضمین می‌کنند. برای رفع این ایرادات، NewSQL برای اپلیکیشن‌های داده‌های بزرگ ایجاد شد. NewSQL از سیستم‌های مدیریت پایگاه‌داده‌ی منطقی جدیدی بر مبنای معماری توزیعی تشکیل می‌شود.

پایگاه‌داده‌های NewSQL بهترین نسخه از فن‌آوری‌های قبلی RDBMS و NoSQL را ترکیب می‌کنند. NewSQL هر دو ویژگی ACID و جستار SQL را ارائه می‌دهد. NewSQL دسترس‌پذیری خوب به داده‌ها و عملکرد پردازش مبادلات آنلاین را میسر می‌سازد.

۳) جستجو و نمایه‌سازی: روش‌های جستجوی متعارف روی محیط توزیعی و پیچیدگی داده‌های بزرگ تطبیق نمی‌یابند. شرکت‌ها باید جستارهای زمان-واقعی جامعی را از طریق حجم بسیار زیادی از مجموعه داده‌های بدون ساختار و ساختار یافته اجرا کنند. این امر به ایجاد موتورهای جستجوی مقیاس‌پذیر بر مبنای فن‌آوری‌های جستجو و نمایه‌سازی همانند زبان پردازش Lucene و Splunk منجر می‌شود.

ج) امنیت در زمینه‌ی داده‌های بزرگ

امنیت در زمینه‌ی داده‌های بزرگ شامل سه جنبه‌ی اساسی می‌شود: امنیت اطلاعات، نظارت ایمن و امنیت داده‌ها. برای مدیریت امنیت در محیط توزیعی به معنای تضمین موارد زیر است: مدیریت داده‌های بزرگ، یکپارچگی سیستم و امنیت فضای سایبری.

۱) نیاز به ابزارهای امنیتی سریع‌تر: فناوری‌های داده‌های بزرگ با پردازش سریع منابع نامتجانس امنیتی (معاملات آنلاین، دوربین‌های نظارت، رویدادهای کارت‌های دسترسی، رویدادهای تولیدی توسط فرآیندها و اپلیکیشن‌های کسب و کار، فایروال و غیره) امنیت را مدیریت می‌کنند. در حال حاضر پردازش رخدادهای امنیتی از طریق پلت‌فرم هادوپ (بجای ۲۰ تا ۶۰ دقیقه از طریق SIEM متعارف) حدوداً یک دقیقه زمان می‌برد.



اکنون استفاده از الگوریتم‌های داده‌کاوی در رخدادهای امنیتی و ترکیب نتایج در آمار امنیتی و رخدادهای بایگانی‌شده امکان‌پذیر است. چنین همبستگی‌هایی تشخیص کارآمد رخدادهای امنیتی و رفتارهای غیرعادی را می‌سازند. علاوه بر این فن‌آوری‌های داده‌های بزرگ ذخیره‌سازی طولانی‌تر و قابل اعتماد را ممکن می‌سازد. بنابراین، بجای حذف رویدادهای امنیتی پس از چند روز، اکنون می‌توان چنین رخدادهایی را برای مدت زمان‌های طولانی‌تر ذخیره کرد.

۲) *چالش‌های امنیتی داده‌های بزرگ*: داده‌های بزرگ چالش‌های امنیتی زیادی در پی دارند: (۱) نحوه‌ی مدیریت ایمن انواع مجموعه داده‌های بزرگ بدون ساختار و نامتجانس، (۲) نحوه‌ی افزودن ساز و کارهای امنیتی در پلت‌فرم‌های توزیعی و در عین حال تضمین سطح خوبی از عملکرد (مثلاً ذخیره‌سازی کارآمد، پردازش سریع، تجزیه و تحلیل زمان-واقعی، (۳) نحوه‌ی تجزیه و تحلیل جریان‌های داده‌ای عظیم بدون آسیب رساندن به محرمانگی و حریم خصوصی داده‌ها.

ارزش داده‌های بزرگ بر روی خوشه‌ها و مراکز داده متمرکز شده است. چنین منابع داده‌ای غنی برای تجارت، دولت‌ها و صنایع بسیار جذاب هستند. داده‌ها در بسیاری از شبکه‌ها به اشتراک گذاشته می‌شوند. این حقایق خطرات مرتبط با امنیت و حفظ حریم خصوصی را افزایش می‌دهند. بنابراین، مهم است که ساز و کارهای پیشرفته‌ی امنیتی را برای محافظت از داده‌های بزرگ مبادله یا ذخیره‌شده در چند خوشه پیاده‌سازی کنیم. با این حال، به دلیل سرعت و حجم بالای داده‌های بزرگ، محافظت از تمام مجموعه داده‌های بزرگ دشوار است. لذا، حفاظت از ارزش داده‌ها و ویژگی‌های کلیدی آن بجای خود داده‌ها عملی‌تر است. علاوه بر این، اضافه کردن لایه‌های امنیتی ممکن است عملکرد سیستم را کاهش دهد و بر تحلیل‌های پویای حجم روز افزون داده‌ها تاثیر منفی بگذارد. بنابراین، کنترل دسترسی و حفاظت از داده‌ها دو مسأله‌ی امنیتی بزرگ هستند. برای اطمینان از حفظ حریم خصوصی و امنیت، ناشناس‌سازی داده‌ها را باید بدون تاثیر بر عملکرد سیستم (تجزیه و تحلیل زمان واقعی) یا کیفیت داده‌ها بدست آورد. با این حال، روش‌های ناشناس‌سازی موجود بر مبنای چند تکرار و محاسبات زمان‌بر هستند. این کار باعث کاهش سرعت عملکرد سیستم بخصوص در زمینه‌ی داده‌های بزرگ می‌شود. علاوه بر این، تجزیه و تحلیل داده‌های بزرگ ترکیب شده دشوار است

۳) *راه‌کارهای امنیت داده‌های بزرگ*: صاحبان داده‌ها مسئول تعیین مقررات و سیاست‌های امنیتی مشخص هستند که باید توسط منابع خارجی رعایت شوند. آن‌ها همچنین باید ریسک‌های امنیتی را در رابطه با ترکیب فن‌آوری‌های مختلف داده‌های بزرگ تحلیل کنند. ساز و کارهای امنیتی تطبیق یافته با پیچیدگی داده‌های بزرگ و محیط‌های امنیتی وجود دارد. چند راه‌کار امنیتی در رابطه با داده‌های بزرگ ایجاد شده‌اند از جمله:

- روش‌های ناشناس‌سازی: برای بهبود ناشناس‌سازی مجموعه داده‌های بزرگ، روشی مبتنی بر یک تکرار را برای تعمیم عملیات پیشنهاد می‌دهد. هدف بهبود ظرفیت‌های موازی‌سازی، عملکرد و مقیاس‌پذیری روش‌های ناشناس‌سازی است.
- روش‌های حفظ حریم خصوصی: در حال حاضر، بسیاری از پروژه‌ها روی ایجاد روش‌های حفظ حریم خصوصی بر مبنای تحلیل آگاهی از حفظ حریم خصوصی و روش‌های زمان‌بندی مجموعه داده‌های بزرگ کار می‌کنند.
- رمزنگاری تک‌ریختی: پیشنهاد شده است تا از محرمانگی اطلاعات اطمینان حاصل شود. برخلاف برخی از تکنیک‌های متعارف در رمزگذاری، این روش محاسبات حتی بر روی داده‌های رمزگذاری شده را می‌سازد. یک پلت‌فرم تطبیقی را برای کار روی محاسبات MapReduce پیشنهاد می‌دهد.
- ساز و کارهای تأیید هویت: کار با آن‌ها در میان خوشه‌های توزیعی و مجموعه داده‌های بزرگ اغلب اوقات پیچیده و سنگین است. در اینجا چند راه‌حل امنیتی برای کار روی این مسأله وجود دارد:
  - Kerberos یک پروتکل تأیید هویت شبکه است که بسیاری از شرکت‌ها از آن برای تضمین تأیید هویت استفاده می‌کنند. این پروتکل با استفاده از رمزنگاری کلید-مخفی تأیید هویت قدرتمندی برای اپلیکیشن‌های مشتری/سرور فراهم می‌کند.
  - سیستم آپاچی سنتری سیستم تأیید هویت مناسب و اجرای قوانین خط مشی را برای کنترل دسترسی به داده‌ها و کنترل دسترسی مبتنی بر نقش را هم روی داده‌ها و هم فراداده‌ها ارائه می‌دهد.
  - ناکس تأیید هویت یکبار ورود و سناریوهای امنیتی تأیید توکن را برای کنترل دسترسی به خدمات هادوپ آپاچی فراهم می‌کند. ناکس همچنین نقطه پایانی خوشه‌ی واحد را برای کاربران فراهم می‌کند که مستلزم دسترسی به داده‌ها و اجرای وظایف در چارچوب هادوپ است.



### نتیجه گیری

همانطور که در این مقاله مشاهده کردیم، نمونه‌های گوناگونی از فرصت‌ها و راه‌حل‌هایی برای داده‌های بزرگ وجود دارد. می‌توانیم نتیجه بگیریم که تجزیه و تحلیل داده‌های بزرگ برای بدست آوردن بینش قابل اعتماد ارزشمند است. چنین بینشی به تصمیم‌گیران اجازه می‌دهد که تصمیم خوبی بگیرند، سیاست‌ها و راهبردهای خود را ارتقاء بخشند، سود بیشتری به دست آورند، و قابلیت رقابت شرکت‌ها را بهبود بخشند. علاوه بر این، این تحول داده‌های بزرگ با تحلیل پیچیده و پیشرفته در میان منابع مختلف موجب غنی‌سازی زمینه‌های مختلف علمی می‌شود.

تجزیه و تحلیل داده‌های بزرگ مدل‌های کسب و کار جدیدی مانند ارزش‌گذاری داده‌ها و دگرگونی کسب و کار را به ارمغان آورده است. در واقع، سازمان‌ها با توجه به سطح بلوغ خود از داده‌های بزرگ استفاده می‌کنند. با رشد و بلوغ سازمان‌ها، آن‌ها نه تنها به تجزیه و تحلیل زمان-واقعی داده‌های بزرگ متکی می‌شوند تا راهبردها و فرآیندها را بهینه‌سازی کنند، بلکه پا را فراتر گذاشته و از ارزش داده‌های بزرگ کسب درآمد می‌کنند. سپس، آن‌ها نه تنها می‌توانند روی افزایش خدمات و محصولات بلکه روی توسعه‌ی اکوسیستم خود تمرکز داشته باشند. برای حفظ توسعه‌ی سازمان‌ها، پلت‌فرم مشترکی لازم است که تمام اعضای اکوسیستم را بهم متصل می‌کند. این پلت‌فرم باید به تجزیه و تحلیل و مدل‌سازی داده‌های بزرگ برای پشتیبانی از نیازهای طرف‌های مختلف (به عنوان مثال دولت‌ها، شرکت‌ها، مشتریان، ادارات، تامین‌کنندگان، جوامع شبکه‌های اجتماعی و کاربران) متکی باشد. این پلت‌فرم باید نهادها را برای کسب دانش عمیق‌تر، بازخورد فوری و توصیه‌های شخصی آماده کند. هدف این است که حداکثر سود برای تمامی نهادها حاصل شود.

با این حال، فن‌آوری‌های متعارف نمی‌توانند روی چالش‌های داده‌های بزرگ (یعنی، سرعت، حجم، تنوع، و پیچیدگی) کار کنند. در حقیقت، مدل‌سازی و کاوش داده‌های بزرگ مستلزم فن‌آوری‌های قدرتمند و روش‌های پیشرفته برای تضمین عملکرد، قابلیت اطمینان نتایج، دسترس‌پذیری داده‌ها و مقیاس‌پذیری است. موضوع دیگر یافتن تعادلی بین الزامات مختلف امنیتی و حریم خصوصی، پردازش قابل اعتماد و سریع، و دسترسی مبتنی بر گرانش برای چند منبع داده‌های بزرگ با اتصال شدید است. بسیاری از فن‌آوری‌ها برای مقابله با این چالش‌ها ایجاد شده‌اند اما هنوز هم معایبی وجود دارد. تحقیقات در بسیاری از زمینه‌ها برای بهبود ویژگی‌ها و قابلیت‌های اپلیکیشن‌های داده‌های بزرگ هنوز در حال انجام است.

### منابع

1. R. Nambiar, R. Bhardwaj, A. Sethi, and R. Vargheese, "A look at challenges and opportunities of big data analytics in healthcare," in IEEE International Conference on Big Data, Oct 2013, pp. 17–22.
2. D. F. Nettleton, Commercial Data Mining: Processing, Analysis and Modeling for Predictive Analytics Projects, 1st ed. Boston, United States: Morgan Kaufmann Publishers-Elsevier, 2014.
3. B. Schmarzo, Understanding How Data Powers Big Business, Wiley, Ed., September 2013.
4. R. Ludena and A. Ahrary, "A big data approach for a new ict agriculture application development," in International Conference on CyberEnabled Distributed Computing and Knowledge Discovery (CyberC), Oct 2013, pp. 140–143.
5. M. Yazici, C. Kamga, and A. Singhal, "A big data driven model for taxi drivers' airport pick-up decisions in new york city," in IEEE International Conference on Big Data, Oct 2013, pp. 37–44.
6. D. Che, M. Safran, and Z. Peng, "From big data to big data mining: Challenges, issues, and opportunities," in Database Systems for Advanced Applications, ser. Lecture Notes in Computer Science, B. Hong, X. Meng, L. Chen, W. Winiwarer, and W. Song, Eds. Springer Berlin Heidelberg, 2013, vol. 7827, pp. 1–15.
7. M. Jensen, "Challenges of privacy protection in big data analytics," in Big Data (BigData Congress), 2013 IEEE International Congress on, June 2013, pp. 235–238.
8. A. Benshir, "Big data for geo-political analysis: Application on barack obama's remarks and speeches," in ACS International Conference on Computer Systems and Applications (AICCSA), May 2013, pp. 1–4.
9. B. Lublinsky, K. T. Smith, and A. Yakubovich, Professional Hadoop Solutions. John Wiley & Sons, 2013.
10. "Skynet." [Online]. Available: <http://skynet.rubyforge.org/doc/>
11. "Product documentation." [Online]. Available: [docs.hortonworks.com/](http://docs.hortonworks.com/)
12. "Ibm infosphere biginsights." [Online]. Available: <http://www-01.ibm.com/>





13. A. Cardenas, P. Manadhata, and S. Rajan, "Big data analytics for security," Security Privacy, IEEE, vol. 11, no. 6, pp. 74–76, Nov 2013.
14. "Oracle white paper: Big data for the enterprise." [Online]. Available: <http://www.oracle.com>
15. T. Lu, X. Guo, B. Xu, L. Zhao, Y. Peng, and H. Yang, "Next big thing in big data: The security of the ict supply chain," in International Conference on Social Computing (SocialCom), Sept 2013, pp. 1066–1073
16. B. Holtsnider and B. D. Jaffe, "Chapter 8 - security and compliance," in {IT} Manager's Handbook (Third Edition), third edition ed., B. Holtsnider and B. D. Jaffe, Eds. Boston: Morgan Kaufmann, 2012, pp. 205-246.
17. T. Ring, "It's megatrends: the security impact," Network Security, vol. ۲۰۱۳no. 7, pp. 5 – 8, 2013.
18. X. Zhang, C. Liu, S. Nepal, C. Yang, W. Dou, and J. Chen, "A hybrid approach for scalable sub-tree anonymization over big data using mapreduce on cloud," Journal of Computer and System Sciences, vol. 80, no. 5, pp. 1008 – 1020, 2014, special Issue on Dependable and Secure Computing The 9th IEEE International Conference on Dependable, Autonomic and Secure Computing.
19. "Apache Knox gateway 0.4.x users guide." [Online]. Available: <http://knox.apache.org/books/knox-0-4-0/knox-0-4-0.html>
20. T. T. Zin, P. Tin, T. Toriu, and H. Hama, "A big data application framework for consumer behavior analysis," in IEEE 2nd Global Conference on Consumer Electronics (GCCE), Oct 2013, pp. 245–246.
21. B. Chandramouli, J. Goldstein, and S. Duan, "Temporal analytics on big data for web advertising," in 28th International Conference on Data Engineering (ICDE), April 2012, pp. 90–101.