

بررسی معماری تکنولوژی های کلان داده و

ضرورت بهره برداری از آن در سیستم های مدیریت زنجیره تامین الکترونیکی

A Survey Big Data Technologies Architecture and The need to exploit it in ESCM Systems

محمدرضا بهبهانی نژاد^۱، حسن رشیدی^۲

چکیده

جهان امروزی با افزایش حجم داده ها و سرعت افزایش و تنوع داده ها با مفهوم جدید "کلان داده"^۳ روبرو است. از طرفی به کمک روشهای معمول نمی توان ذخیره و پردازش چنین داده هایی را انجام داد. به همین دلیل دنیای امروزی نیاز به تکنولوژی های جدیدی برای ذخیره و پردازش و راه حلهایی برای کلان داده می باشد. از طرف دیگر سیستم های کسب و کار مانند سیستم های مدیریت زنجیره تامین الکترونیکی^۴ برای ذخیره و پردازش این داده ها نیز دچار مشکلاتی هستند.

این مقاله ابتدا به معرفی و شرح ویژگیهای کلان داده می پردازد و سپس تکنولوژیهای مختلف ذخیره و پردازش آنرا مورد بررسی قرار می دهد. سپس چالش های سیستمهای مدیریت زنجیره تامین الکترونیکی را در مواجهه با کلان داده ارائه کرده و ضرورت یکپارچگی معماری کلان داده ها و معماری سیستم های مدیریت زنجیره تامین الکترونیکی را برای رفع این مشکلات طرح می کند. مهمترین هدف این مقاله ترویج دیدگاه یکپارچه سازی کلان داده و سیستم های مدیریت زنجیره تامین الکترونیکی در چارچوب فرایند تصمیم گیری برای کمک به مدیران برای ایجاد فرصتهای جدید در حل مشکلات و مسائل خاص می باشد. این یکپارچگی فرصتهای جدیدی را در حل مشکلات مربوط به سیستم های مدیریت زنجیره تامین الکترونیکی و کلان داده ایجاد می کند و کمک شایانی برای مدیران و دست اندرکاران این سیستم ها خواهد کرد.

واژگان کلیدی : کلان داده، ذخیره سازی کلان داده، پردازش کلان داده، معماری کلان داده، سیستم های مدیریت زنجیره تامین الکترونیکی

^۱. کارشناسی ارشد مهندسی فناوری اطلاعات-تجارت الکترونیک، دانشکده برق-رایانه و فناوری اطلاعات، واحد قزوین، دانشگاه آزاد اسلامی، قزوین،

ایران، reza2005nejad@gmail.com

^۲. استاد گروه رایانه، دانشکده علوم ریاضی و رایانه، دانشگاه علامه طباطبائی، تهران، ایران، hrashi@gmail.com

^۳Big Data

^۴Electronic Supply Chain Management

دنیای امروزی با حجم زیاد و تنوع گسترده ای از کلان داده روبرو است. چندین تعریف از کلان داده ارائه شده است. موسسه گارتر و ویژگیهای کلان داده را اینگونه بیان کرده که کلان داده ها دارای حجم بالایی از اطلاعات است که بسرعت تولید می شوند و دارای تنوع بسیار زیادی می باشند و در ادامه این ویژگیها آمده که کلان داده ها نیاز به اشکال جدیدی از پردازش دارند تا بتوانند تصمیم گیری را غنی تر سازند، بینش جدیدی را کشف کنند و نیز فرایندها را بهینه نمایند. کریس لینچ (محقق و کارآفرین) می گوید : کلان داده اولین فناوری پس از شکل گیری اینترنت است که می تواند جهان را تغییر دهد.

یکی از رویکردهای مهم جهت تصمیم گیری، استفاده از سیستمهای مدیریت زنجیره تامین الکترونیکی است. زنجیره تامین، شبکه ای از تسهیلات و امکانات توزیع است که وظایف تهیه مواد اولیه، تبدیل این مواد به محصولات میانی و نهایی و توزیع این محصولات نهایی به مشتریان را برعهده دارد. مدیریت زنجیره تامین به عنوان یک استراتژیک رقابتی برای یکپارچه سازی تامین کنندگان و مشتریان با هدف بهبود پاسخگو بودن و انعطاف پذیری سازمان های تولیدی شناخته شده است. چنانچه یک زنجیره تامین به صورت الکترونیکی و با کمک نرم افزارهای کامپیوتری مبتنی بر وب مدیریت شود، مدیریت زنجیره تامین الکترونیکی نامیده می شود. مدیریت زنجیره تامین الکترونیکی تلاشی در جهت تسهیل جریان اطلاعات درون زنجیره تامین است. از آنجایی که سیستم های مدیریت زنجیره تامین الکترونیکی توانایی ذخیره و پردازش داده های با ویژگی های کلان داده را ندارند و در این زمینه با چالش روبرو هستند باید سعی شود با ادغام معماری کلان داده با سیستم های مدیریت زنجیره تامین الکترونیکی این چالش برطرف شود.

ادامه این مقاله بدین صورت ساختار بندی شده است: در بخش دوم، مسئله و سئولات اساسی مربوط به این موضوع بیان خواهد شد. در بخش سوم ادبیات موضوعی کلان داده و تاریخچه آن و در بخش چهارم ذخیره و پردازش کلان داده مورد بررسی قرار می گیرد. در بخش پنجم سیستم های مدیریت زنجیره تامین الکترونیکی و کلان داده در ارتباط با یکدیگر و بحث یکپارچگی آنها ارائه می شود. در بخش آخر نتیجه گیری و پیشنهاد کارهای آتی ذکر خواهد شد.

۲- مسائل مرتبط با کلان داده

کلان داده چیست؟ برای پشتیبانی سیستم های مدیریت زنجیره تامین الکترونیکی از کلان داده ها، چه باید کرد؟ چه تغییری در معماری سیستم های مدیریت زنجیره تامین الکترونیکی باید صورت گیرد؟ آیا می توان ایندو معماری را با هم تلفیق کرد؟ آیا می توان بجای پایگاه داده های رابطه ای در سیستم های مدیریت زنجیره تامین الکترونیکی از پایگاه داده های غیررابطه ای استفاده کرد؟

برای شناخت بیشتر مفهوم کلان داده، ابتدا ویژگیهای آنرا مورد بررسی قرار می دهیم. البته در مورد تعداد ویژگی های کلان داده نظرات مختلف است برخی ۳۵، عده ای ۴۵ و اخیراً تعدادی از محققین ۵۵ را بعنوان ویژگیهای کلان داده معرفی کرده اند که اشتراک بین تمام این نظرات همان ۳۵ یعنی سه ویژگی حجم^۱ و سرعت^۲ و تنوع^۳ است. علاوه بر این سه ویژگی از دیگر ویژگی هایی که برخی از محققین معرفی کرده اند می توان به این موارد اشاره کرد : ویژگی پیچیده بودن داده ها (منظور داده هایی است که نمی توان آنها را با روشهای معمول مورد پردازش و

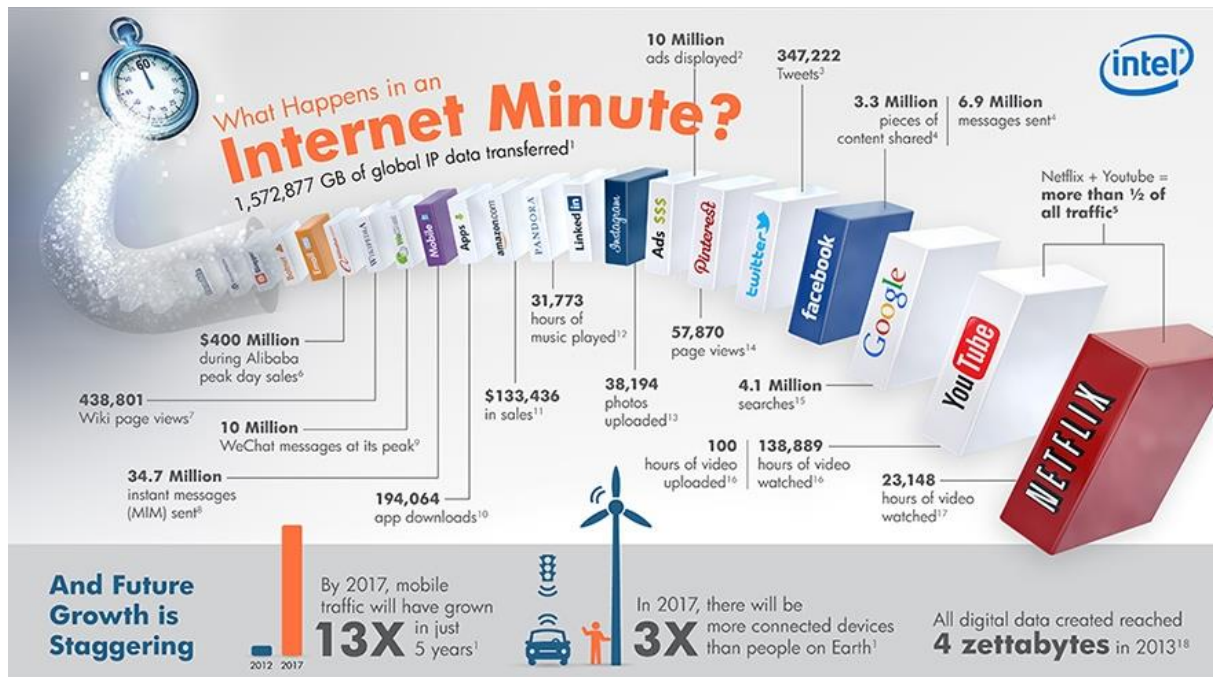
^۱. NOSQL : Not Only SQL

^۲. Volume

^۳. Velocity

^۴. Variety

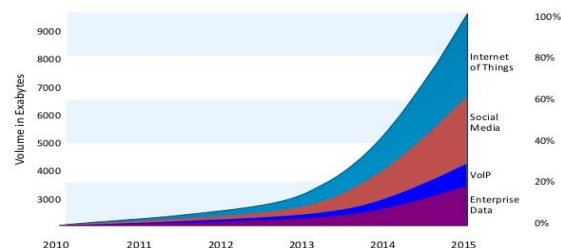
تجزیه و تحلیل قرار داد)، ویژگی صحت داده ها و ویژگی ارزش داده ها. در ادامه توضیحاتی پیرامون سه ویژگی اصلی کلان داده آورده می شود تا این ویژگیها بهتر شناخته شوند.



شکل ۲. سرعت رشد داده‌ها در اینترنت در یک دقیقه [۱۴]

• **کلان داده از نظر حجم داده:** کلان داده اصطلاحی است که به حجم خاصی از داده اشاره ندارد اما اغلب به مقدار پتابایت یا اگزابایت داده گفته می شود. به عنوان مثال روزانه بیش از ۵۰۰ ترابایت داده در فیس بوک به اشتراک گذاشته می شود. که در یک ماه حدود ۱۵ پتابایت می شود. پس فیس بوک یک کلان داده است. دیگر شبکه های اجتماعی مانند توئیتر، یوتیوب، لینکدین، اینستاگرام نیز از نظر حجم داده وضعیتی شبیه به فیس بوک دارند. پس آنها نیز از جمله کلان داده محسوب می شوند.

طبق شکل ۱ حجم داده ها در اینترنت بسیار بالا رفته است و اکنون دنیا در مقابل انبوهی از اطلاعات مواجه است. همانطور که در تصویر مشاهده می شود از سال ۲۰۱۰ تا ۲۰۱۳ حجم داده ها در اینترنت بصورت خطی افزایش یافته اما از سال ۲۰۱۳ تا ۲۰۱۵ این رشد بصورت نمایی پیش بینی شده است و در این رشد نمایی حجم داده های اینترنت اشیاء و شبکه های اجتماعی که از جمله داده های بدون ساختار و از مصادیق کلان داده هستند بیش از همه خود را نشان می دهد.



شکل ۱. نمودار حجم داده ها در اینترنت [۱۳]

- **کلان داده از نظر سرعت رشد داده :** طبق برآوردهای صورت گرفته داده هایی که در عرض ۵ سال گذشته تولید شده، هم اکنون در عرض ۵ دقیقه تولید می شود که اکثر این داده ها در شبکه های اجتماعی تولید می شوند و شکل ۲ سرعت رشد داده های اینترنتی را در یک دقیقه نشان می دهد. همانطور که مشاهده می شود صفی از سایتهای اینترنتی که بیشترین سرعت را در افزایش داده های خود دارند تشکیل شده و رتبه های اول این صف اغلب مربوط به شبکه های اجتماعی می باشد. نتیجه اینکه سرعت رشد داده های غیرساختاریافته در شبکه های اجتماعی بیشتر از سرعت رشد بقیه داده ها در اینترنت است.

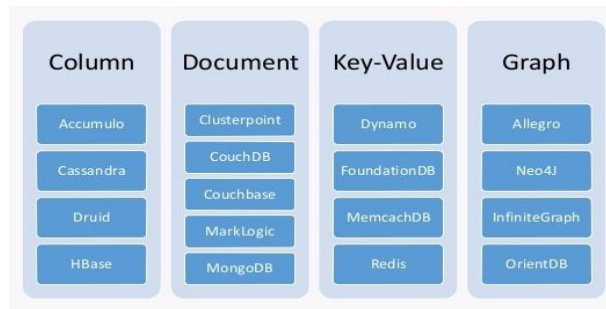
- **کلان داده از نظر تنوع داده :** به طور کلی از نظر ساختاری سه نوع داده وجود دارد :
 - داده های ساختار یافته، که شامل داده های موجود در پایگاه داده بوده که بصورت سطری و ستونی گروه بندی شده اند. این داده ها حدود ۵ درصد از داده های موجود در دنیا را دربر می گیرد.
 - داده های شبه ساختار، که دارای ساختار می باشند اما ساختارشان با ساختارهای رسمی و شناخته شده تطابق ندارد. این داده ها حدود ۱۰ درصد کل داده های موجود را شامل می شود.
 - داده های بدون ساختار، که نمی توان آنها را در پایگاه داده های سطری و ستونی ذخیره سازی کرد مانند داده های صوتی و تصویری و ... این نوع داده ها حدود ۸۰ درصد کل داده ها را شامل می شود و کلان داده اغلب از این نوع داده ها هستند. از آنجایی که اغلب داده های موجود در اینترنت بدون ساختار و یا شبه ساختار هستند پس از این لحاظ کلان داده محسوب می شوند.

گسترش کلان داده و تکنولوژی های مربوط به آن، ویژگیهای جدیدی را به مشخصات این نوع داده ها اضافه نموده است. برخی از این ویژگیها که در سالهای اخیر بیشتر مطرح شده شامل: ویژگی پیچیده بودن داده ها (منظور داده هایی است که نمی توان آنها را با روشهای معمول مورد پردازش و تجزیه و تحلیل قرار داد)، ویژگی صحت داده ها و ویژگی ارزش داده ها.

- **پایگاه داده کلان داده:** از آنجایی که بیش از ۸۰ درصد داده ها در کلان داده بدون ساختار هستند و پایگاه داده های رابطه ای توانایی ذخیره و مدیریت اینگونه داده ها را ندارند. برای ذخیره این داده ها باید از پایگاه داده های غیررابطه ای استفاده کرد، که در ادامه تعریف این نوع از پایگاه داده ها و انواع آن معرفی شده است.

- **انواع پایگاه داده ها :**
 - پایگاه داده های رابطه ای : پایگاه داده هایی هستند که دارای سطرو ستون بوده و از زبان SQL پشتیبانی می کنند مانند: Microsoft Access , MySQL , Oracle , SqlServer و ...
 - پایگاه داده های غیر رابطه ای^۱: پایگاه داده هایی که دارای سطرو ستون نبوده و از زبان SQL پشتیبانی نمی کند. مانند: MongoDB , OrientDB , CouchDB , Hbase و ...

انواع پایگاه داده های غیررابطه ای : در شکل ۳ انواع پایگاه داده های غیر رابطه ای و مثالهایی از هر کدام آورده شده است. همانطور که در شکل مشاهده می شود چهار نوع پایگاه داده غیررابطه ای وجود دارد که بر اساس ستون، سند، ارزش کلیدی و گراف طبقه بندی شده اند و برای هر نوع از این پایگاه داده ها، چهار مثال آورده شده است.



شکل ۳. انواع پایگاه داده غیر رابطه ای [۱۵]

۳- ادبیات موضوعی کلان داده

ادبیات موضوعی کلان داده شامل تعریف کلان داده از دیدگاه های مختلف و معماری کلان داده در زیر شرح داده شده است.

۳-۱. تعریف کلان داده

سایت دانشنامه ویکی پدیا در تعریف کلان داده اینگونه توضیح داده است که کلان داده اصطلاحی است که به مجموعه داده هایی اطلاق می شود که مدیریت، کنترل و پردازش آنها فراتر از توانایی ابزارهای نرم افزاری در یک زمان مورد انتظار است. مقیاس کلان داده، به طور مداوم در حال رشد از محدوده چند ۱۰ ترابایت به چندین پتابایت، در یک مجموعه داده واحد است و از نمونه های آن می توان به گزارش های وبی، سامانه های بازشناسی با امواج رادیویی، شبکه های حسگر، شبکه های اجتماعی، متون و اسناد اینترنتی، نمایه های جستجوهای اینترنتی، نجوم، مدارک پزشکی، آرشیو عکس، آرشیو ویدیو، پژوهش های زمین شناسی و همچنین به تجارت در مقیاس بزرگ اشاره کرد.

کلان داده فرصتهای جدیدی برای محققان در زمینه رسیدن به ارتباطات بهتر و بیشتر در هنگام حضور در شبکه های اجتماعی فراهم می کند از جمله این فرصتها می توان به این موارد اشاره کرد: راه هایی برای مدیریت زنجیره تامین الکترونیکی به کمک شبکه های اجتماعی، رسیدگی به مشکلات ترافیکی در سیستم حمل و نقل هوشمند و پشتیبانی از شهرهای هوشمند علاوه بر این موارد، ارزشمندی کلان داده موجب ایجاد ابزارها و سیستم های جدیدی شده که موجب تسهیل اطلاعاتی در رفتار مصرف کننده و مشتریان شده است و از طرفی امکان پیش بینی های اقتصادی در بازار سرمایه را نیز بیش از پیش فراهم کرده است.

اطلاعات و دانش با ارزش ترین دارایی سازمانها برای تصمیم گیری می باشد که نیاز به یک محیط پردازشی و همچنین نیاز به بارگذاری دارند تا در فرایندهای سازمان مورد استفاده قرار گیرند و در تصمیم گیری های سازمان موثر باشند.

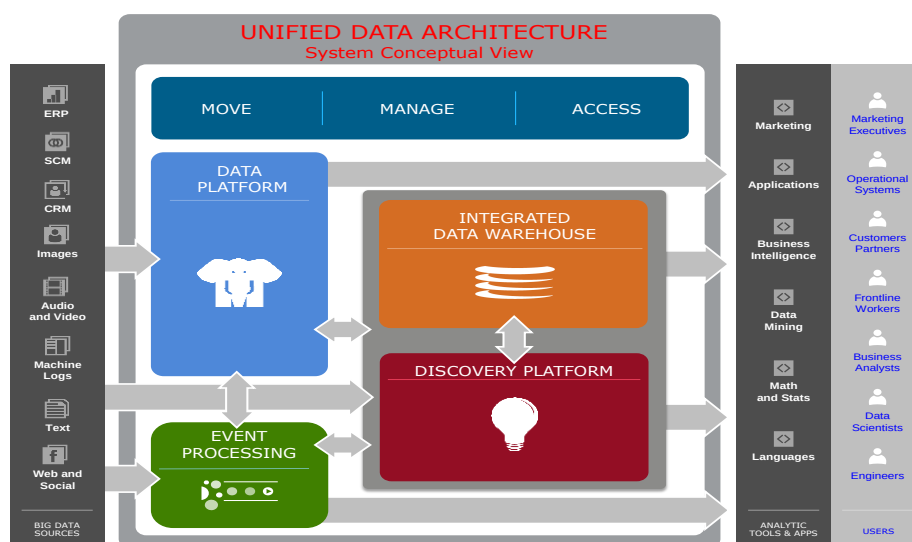
کلان داده می تواند در مواردی چون تمرکز در فرایند تصمیم گیری، بهبود کار سیستم های مدیریت زنجیره تامین الکترونیکی، پردازش، تجزیه و تحلیل، اشتراک گذاری و نمایش بصری اطلاعات به روند جمعیت و تبدیل اطلاعات به دانش و به بهبود سازمان از نظر دانشی کمک کند.

با افزایش حجم داده داخلی و خارجی مفهوم کلان داده مطرح می شود که دارای ویژگی هایی چون حجم، سرعت، تنوع و ارزش می باشد. مفهوم حجم داده ها، به مقدار داده اشاره دارد. مفهوم سرعت، به چابکی داده ها در تجزیه و تحلیل و پردازش اشاره می کند. مفهوم تنوع، توصیفی از انواع مختلف داده ها از نظر ساختار یافتگی است و مفهوم ارزش، اشاره دارد به داده های ارزشمندی که می توان به کمک تجزیه و تحلیل، ارزش پنهان آنها را کشف کرد و کلان داده می تواند بواسطه ارزش این داده ها، فرصتهای جدیدی را ایجاد نماید و بالاخره یافته های جدید به تصمیم گیری مدیران کمک کند.

بین کلان داده و بقیه داده ها پنج تفاوت عمده وجود دارد : الف) هدف داده ها ب) محل داده ها ج) ساختار داده ها د) نحوه آماده سازی داده ها ه) تجزیه و تحلیل داده ها

۳-۲. معماری کلان داده

تصویری از معماری سطح بالا برای استفاده از کلان داده در شکل ۴ نمایش داده شده است. همانطور که در شکل مشاهده می شود منابع کلان داده ها ورودی این معماری هستند در وسط تصویر معماری داده ها را نشان می دهد که در سه مرحله کار یکپارچه سازی داده ها را انجام می دهد (مراحل Access , Manage , Move) و در سمت راست تصویر ابزارهای تجزیه و تحلیل و کاربران قرار دارند که خروجی این معماری محسوب می شوند.



شکل ۴. تصویری از معماری سطح بالا برای استفاده از کلان داده ها [۱۶]

تاریخچه پروژه های مربوط به معماری کلان داده در جدول ابخوبی نمایش داده شده است. در بین معماری های ارائه شده برای کلان داده، هادوپ موفق تر ظاهر شده و این موفقیت به دلیل منبع باز بودن و پشتیبانی آن از طرف متخصصین بوده ، به گونه ای که ارتقا قابل ملاحظه ای در سالهای اخیر داشته و همچنان در حال گسترش است. هادوپ در ابتدا بعنوان زیرپروژه موتور جستجوی ناچ و سپس تحت گواهی آپاچی گسترش فراوانی یافت. در ادامه آنقدر گسترش پیدا کرد که خود شامل تعداد زیادی پروژه شد بطوریکه در حال حاضر هادوپ دیگر نام یک زیرپروژه یا حتی نام یک پروژه نیست بلکه نام مجموعه ای از پروژه ها یا عبارتی نام کتابخانه ای از پروژه هاست.

جدول ۱. تاریخچه پروژه های مربوط به معماری کلان داده ها [۱۷]

سال	نام شرکت	فعالیت عمده
۲۰۰۰	Seisint	ایجاد چارچوب فایل توزیع شده برای ذخیره سازی و جستجو در کلان داده ها در چندین سرور توزیع شده
۲۰۰۴	Google	مدل پردازش موازی نگاشت/کاهش را منتشر کرد که از همان معماری شرکت Seisint تبعیت می کرد و شرکت آپاچی این مدل را گسترش داد.
۲۰۰۶	Yahoo	پروژه موتور جستجوی ناچ ^۱ که برای جستجو در کلان داده ها طراحی شده بود را خرید و از پروژه هدوپ ^۲ تحت گواهی آپاچی حمایت کرد و در سال ۲۰۰۹ اولین کدمنبع خود را که با استفاده از محصولات هدوپ بود منتشر کرد
۲۰۰۸	ChoicePoint	پلتفرم های پردازش موازی را ایجاد کرد
۲۰۱۱	LexisNexis	این شرکت در سال ۲۰۰۴ شرکت Seisint و در سال ۲۰۱۱ شرکت ChoicePoint را خرید و سیستم HPCC را تحت گواهی آپاچی منتشر کرد

- بررسی ذخیره و پردازش داده ها در معماری کلان داده: مهمترین فاکتورهای موفقیت در تجزیه و تحلیل کلان داده در شکل ۵ نمایش داده شده است که عبارتند از: نیاز به یک کسب و کار پاک، پرسنلی با مهارتهای تجزیه و تحلیل پیشرفته، ابزارهای تجزیه و تحلیل درست، زیرساخت داده ای قوی، فرهنگ تصمیم گیری واقعیت محور، انطباق بین استراتژی کسب و کار و فناوری اطلاعات، ضمانت سپرده شده قوی.



شکل ۵. فاکتورهای اصلی موفقیت در تجزیه و تحلیل کلان داده [۱۶]

از آنجایی که موفق ترین معماری از بین معماری هایی که تاکنون برای کلان داده ارائه شده است معماری هدوپ است در اینجا به بررسی ذخیره و پردازش داده ها در معماری هدوپ پرداخته می شود.

هدوپ یک پایگاه داده نیست و همچنین هدوپ یک نرم افزار هم نیست بلکه هدوپ یک چارچوب یا مجموعه ای از نرم افزارها و کتابخانه هایی است که ساز و کار پردازش حجم عظیمی از داده های توزیع شده را فراهم میکند. در واقع هدوپ را می توان به یک سیستم عامل تشبیه کرد که طراحی شده تا بتواند حجم زیادی از داده ها را بر روی ماشین های مختلف پردازش و مدیریت کند.

بهترین مثال برای فهم عملکرد هدوپ، بیان تفاوت نرم افزار وی ام ویر و هدوپ است. وی ام ویر یک سرور فیزیکی را به چند سرور مجازی تبدیل می کند و هدوپ چند سرور فیزیکی را به یک سرور مجازی تبدیل می کند. فریم ورک هدوپ شامل زیر پروژه های مختلفی می شود که در جدول ۲ نمایش داده شده است :

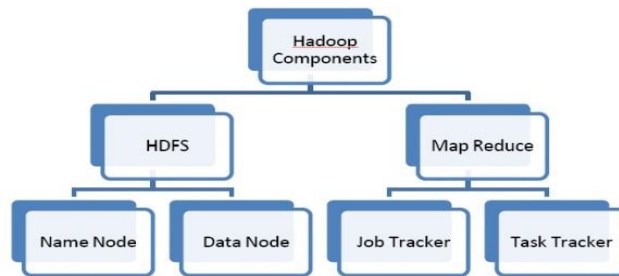
جدول ۲. زیرپروژه های کتابخانه هدوپ [۱۸]

HDFS	Ambari	Hive	Tez	Cassandra
MapReduce	Spark	Mahout	ZooKeeper	Chukwa
YARN	HBase	Pig	Avro	Sqoop

مهمترین زیرپروژه های هدوپ، سیستم فایل توزیع شده هدوپ و نگاشت/کاهش می باشد از آنجایی که کلیه معماریهای کلان داده از ابزارهایی شبیه به ایندو برای مدیریت داده ها استفاده می کنند، در ادامه بصورت مبسوط توضیح داده خواهند شد. برخی دیگر از زیرپروژه های فرعی هدوپ عبارتند از :

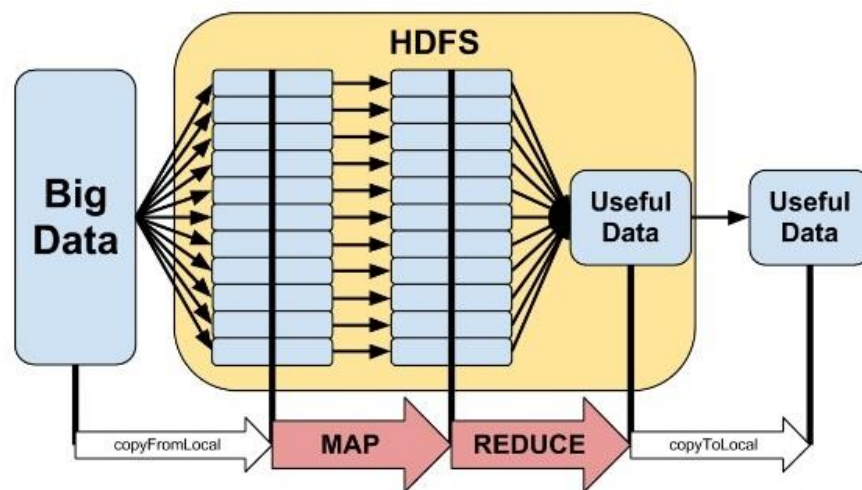
- Hbase و Cassandra که از جمله پایگاه داده های غیر رابطه ای^۵ از نوع ستون محور هستند
- Hive یک انبارداده است که توسط FaceBook بصورت بازمتن، جهت تحلیل حرفه ای داده ها تولید شده
- Mahout یک ابزار داده کاوی است که با اکثر الگوریتم های داده کاوی کار میکند
- Sqoop ابزار مفیدی است که وظیفه انتقال اطلاعات از پایگاه داده های رابطه ای به هدوپ را برعهده دارد

- **مهمترین بخشهای معماری کلان داده:** اکثر معماریهایی که تاکنون برای کلان داده تعریف شده مانند هدوپ، از دو بخش اصلی تشکیل شده اند. بخش اول، سیستم فایل توزیع شده و بخش دوم، نگاشت/کاهش است.
- **کامپوننت های اصلی:** در شکل ۶ کامپوننت های اصلی هدوپ نمایش داده شده است و در ادامه وظیفه هر کدام نیز شرح داده شده است. مهمترین کامپوننت های هدوپ سیستم فایل توزیع شده هدوپ و نگاشت/کاهش می باشد و بقیه کامپوننت ها زیرمجموعه ای از این دو کامپوننت محسوب می شوند.



شکل ۶. نمایش درختی کامپوننت های اصلی هادوپ [۵]

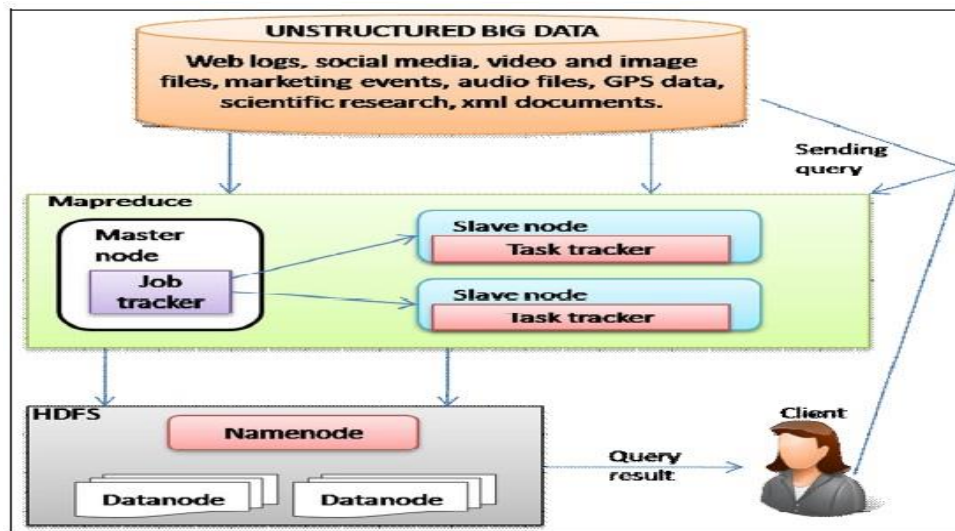
- چارچوب معماری کلان داده: در شکل ۷ چارچوب معماری کلان داده بطور خلاصه نمایش داده شده است. همانطور که مشاهده می شود ابتدا کار نگاشت کلان داده انجام می شود و سپس فرایند مربوط به سیستم فایل توزیع شده هادوپ صورت می گیرد و در پایان کار کاهش انجام می شود و در پایان اطلاعات مفید به کاربر تحویل داده می شود.



شکل ۷. چارچوب کلی معماری کلان داده [۲۰]

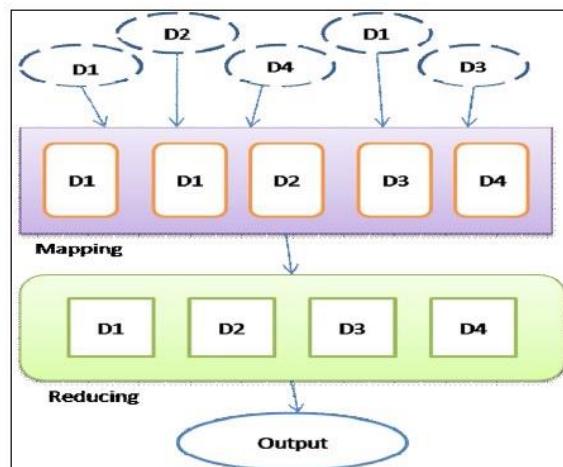
- شکل ۸ معماری کلان داده را با جزئیات بیشتری نمایش می دهد. قسمتهای مختلف نگاشت/کاهش و سیستم فایل توزیع شده هادوپ به همراه توابع مربوط به هر قسمت آورده شده است. گره نام 'داده های مربوط به محل قرارگرفتن فایلها و ویژگیهای اسناد ذخیره شده در فایلها را در خود نگهداری می کند. گره داده محتوای فایلها را بصورت بلوک در گره های مختلف نگهداری می کند. ردیاب شغل گره در قسمت گره اولیه واقع شده است داده ورودی را به چند قسمت تقسیم کرده و هر قسمت را در یک گره ثانویه قرار می دهد. ردیاب شغل پاسخ زیرمسائل و داده های مرتب شده را از گرههای ثانویه به گره اولیه ارسال میکند.

.Name Node
 .DataNode
 .JobTracker
 .MasterNode
 .SlaveNode



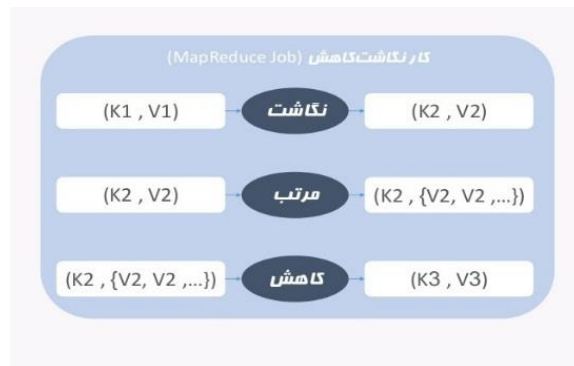
شکل ۸. چارچوب دقیقتری از معماری کلان داده ها [۴]

- ویژگیهای نگاشت/کاهش: الف) به معنای نگاشت/کاهش است. ب) یکی از قسمت های مهم هadoop است. ج) برای اولین بار توسط گوگل ارائه شد. د) برای پشتیبانی از رایانش توزیع شده و موازی است. ه) کار پردازش محاسباتی را بر روی کلیه داده های با ساختار (ذخیره شده در پایگاه داده های رابطه ای) و یا داده های بی ساختار (ذخیره شده در پایگاه داده های غیررابطه ای) انجام می دهد. یکی از مهمترین مزیت های نگاشت/کاهش این است که اجازه می دهد پردازش عملیاتی نگاشت/کاهش بصورت توزیع شده، موازی و همزمان انجام گیرند و مستقل بودن نگاشتها متضمن اجرای موازی آنهاست. نگاشت/کاهش شامل دو گام است: گام نگاشت^۱ و گام کاهش^۲ شکل ۹ فرایندهای نگاشت/کاهش را بصورت نمادین نمایش داده است بطوریکه ابتدا عملیات نگاشت بر روی داده ورودی صورت می گیرد و سپس عملیات کاهش انجام می شود و در پایان اطلاعات خروجی برای کاربر به نمایش در می آید.



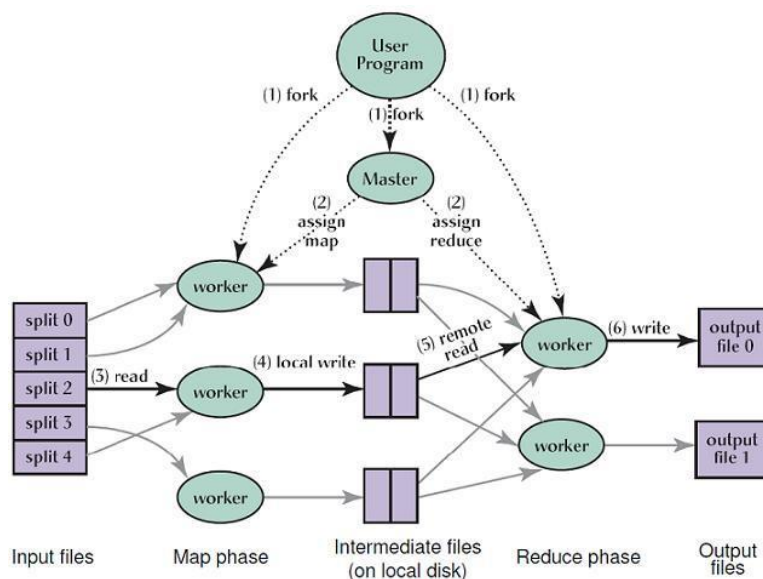
شکل ۹. فرایندهای نگاشت و کاهش [۴]

- شکل ۱۰ فرایند نگاشت/کاهش را به نوع دیگری نمایش می دهد و عملیات تجزیه و تحلیل و مرتب سازی داده ها را که بین فرایندهای نگاشت و کاهش صورت میگیرد، بخوبی نمایش می دهد.



شکل ۱۰. نمایش وظیفه نکاشت/کاهش [۲۲]

شکل ۱۱ جزئیات عملیات نکاشت/کاهش را بطور دقیقتر همراه با اولویت انجام فرایندهای داخلی آن نمایش می دهد بطوریکه این فرایندها با دستور کاربر برای دریافت اطلاعات مورد نیازش آغاز می شود و تا تحویل این اطلاعات در فایل خروجی ادامه می یابد.



شکل ۱۱. نمایش جزئیات عملیات نکاشت/کاهش [۱۹]

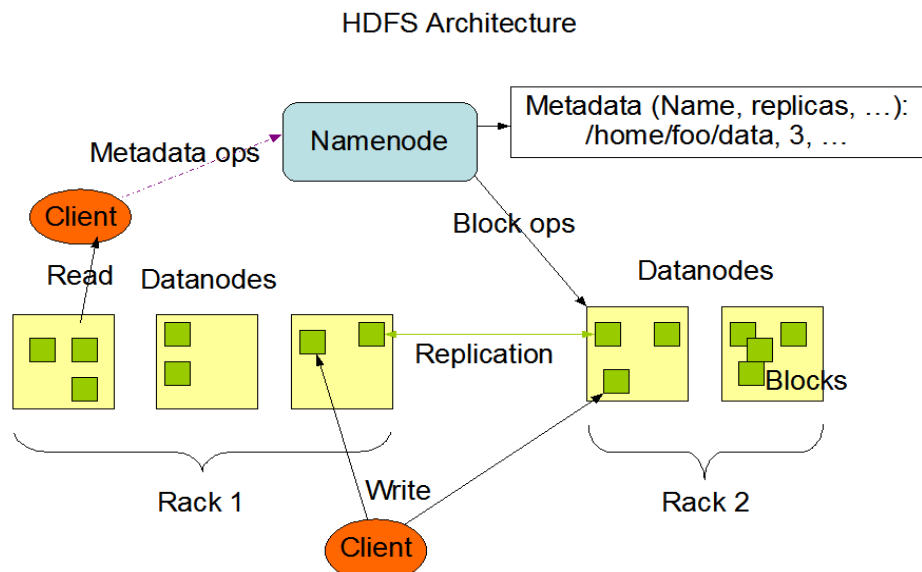
- ویژگیهای سیستم توزیع شده فایل هدوپ^۱: الف) به اختصار و بصورت مخفف به آن HDFS گفته می شود ب) بمعنای سیستمی است در هدوپ که شامل فایلهای توزیع شده می باشد. ج) وظیفه آن ذخیره سازی، جمع آوری و تجزیه و تحلیل داده های با ساختار، نیمه ساختار و بدون ساختار است.

سیستم فایل توزیع شده هدوپ مانند قسمت زباله دانی شهرداری است. در آنجا ابتدا کلیه زباله ها در یکجا ریخته می شود و سپس به زباله های قابل تجزیه و غیرقابل تجزیه تقسیم می شوند آنگاه برای بازیافت ارسال می شوند.

سیستم فایل توزیع شده هدوپ فایلهای بزرگ را در چندین سرور ذخیره می کند تا از یک طرف قابلیت اطمینان را افزایش دهد و از طرف دیگر نیاز به ریدبندی هاردهای سرور را مرتفع کند تا فضای دیسک سرور بطور کامل برای ذخیره داده ها در اختیار باشد.

همانطور که شکل ۱۲ مشاهده می شود سیستم فایل توزیع شده هادوپ دارای یک قسمت گره نام و تعدادی گره داده می باشد. کلان داده به بلوک هایی تقسیم شده و هر یک از این بلوکها در یک یا چند گره داده ذخیره می شود و اطلاعات مربوط به این بلوکها شامل متا دیتای آن بلوک در گره نام ذخیره می شود.

کلاینت ها بلوکهای داده ای مورد نظر خود را از گره های داده می خوانند و یا اطلاعات مورد نظر خود را در آنها می نویسند.



شکل ۱۲. نمایش قسمتهای مختلف سیستم فایل توزیع شده هادوپ [۵]

در معماری کلان داده اهمیت نگاشت/کاهش از سیستم فایل توزیع شده هادوپ بیشتر است و در نگاشت/کاهش مهمترین موضوع زمانبندی است. اگر بتوان زمان خواندن و نوشتن از حافظه و زمان انجام یک کار یا انجام یک پردازش را به حداقل رساند، عملکرد سیستم بهبود می یابد و در نتیجه بهره وری سیستم افزایش چشم گیری خواهد یافت.

۵- سیستم های مدیریت زنجیره تامین الکترونیکی و کلان داده ها

امروزه سازمان های تجاری با محیط های رقابتی تر و پیچیده تری نسبت به قبل مواجه هستند [۵۳]. مدیریت زنجیره تامین یک استراتژی عملیاتی برای کسب موفقیت در رقابتهای سازمانی در سطح جهان است. این رقابت در پاسخگویی به نیاز مشتریان و سودآوری بیشتر انجام میگردد. بالا رفتن سطح تقاضای مشتریان و افزایش دانش و آگاهی آنها منجر به کاهش چرخه عمر محصول شده است. این امر سازمان ها را وادار می کند که زمان را به عنوان یک عنصر مهم در دنیای رقابتی امروز در نظر بگیرند، سطح خدمت به مشتریان را افزایش دهند و محصولات را مطابق با خواسته مشتریان تهیه نمایند [۵۴]. خواسته و نیاز مشتریان (که یکی از ارکان اصلی زنجیره به شمار می آید)، به مرور زمان، جهت گیری اصلی بازار را شکل می دهد [۵۵]. امروزه مدیران ارشد شرکت های تولیدی از روش ها و ابزارهای گوناگونی جهت دستیابی به اهداف و طرح های تجاری خود که مبتنی بر کسب سهم بیشتری از بازار می باشد استفاده می کنند. در این راستا، راهبردهایی مدنظر آنها خواهد بود که به ارائه محصولات با کیفیت تر و با قیمت پایین تر و دسترسی بیشتر منجر می شود. بنابراین همواره در تلاشند تا با کاهش هزینه، افزایش انعطاف پذیری به مزیت رقابتی نسبت به سایر رقبا دست یابند.

اهمیت فناوری اطلاعات و سیستم های اطلاعاتی مبتنی بر زنجیره تامین غیرقابل انکار است. اینترنت به عنوان یک ابزار قدرتمند رقابتی جدید وارد کسب و کار امروزی شده است. ابزارهای الکترونیکی می توانند انعطاف و چابکی شرکت ها را از طریق شناسایی نیازهای بازار، ایجاد تغییرات لازم در سازمان و تعیین مدل کسب و کار افزایش دهند و باعث حضور موفق آنها در بازار شوند.

دلایل اینکه چرا دیگر نمی توان به روش سنتی عمل کرد و از ابزارهای الکترونیکی بهره برداری نکرد عبارت است از:

- پراکندگی تامین کنندگان در نقاط مختلف جهان و نیاز به فناوری های الکترونیکی برای تعامل با آنها
- نیاز به ایجاد سیستم های اطلاعاتی یکپارچه و بروز جهت پیاده سازی سیستم های مدیریت زنجیره تامین
- بکارگیری روشهای نوین تجارت الکترونیک جهت ارائه خدمات و کالا به مشتریان و تعامل موثر با مشتریان

اینترنت، فن آوری اطلاعات (IT) و ارتباطات الکترونیکی، نه تنها فرصتهای گستردهای را در محیط رقابتی کسب و کار به وجود آورده است، بلکه تغییرات قابل توجهی نیز در فعالیتهای کسب و کار ایجاد نموده است؛ اینترنت یکی از مؤثرترین ابزار برای ارائه ارتباطات در میان مردم است و تمام فعالیتهای کسب و کار را تغییر داده است.

بحث مدیریت زنجیره تامین الکترونیکی (e-SCM) یکی از حوزههای رقابتی صنعت کشور است. وجود شبکه اینترنت و اطلاعات در مدیریت زنجیره تامین، ضمن توانمندسازی و رقابت پذیر ساختن شرکت های کوچک و متوسط (SMEs)، باعث تداوم آنها در صحنه رقابت امروزی است. مدیریت زنجیره تامین الکترونیکی در بهبود عوامل زنجیره تامین، صحت اطلاعات، کاهش اثر شلاق چرمی و رساندن محصول در زمان و مکان مناسب با سرعت و دقت بالا به دست مشتری، نقش بسیار مهمی ایفا می کند. توجه به این عوامل می تواند در اتخاذ تصمیم های مناسب برای ارتقای شرکت و یا سازمان توسط مدیران مربوطه، سودمند باشد.

۱-۵. تعریف سیستم های مدیریت زنجیره تامین الکترونیکی

زنجیره تامین شامل همه مراحل (اعضای زنجیره) است که چه مستقیم و چه غیرمستقیم، در برآورده سازی درخواست یک مشتری نقش دارند. در یک زنجیره تامین معمولی، مواد خام از تامین کننده به کارخانه ها ارسال می شوند، سپس محصولات تولید شده در کارخانه ها به انبارهای میانی و انبارهای توزیع کننده ها ارسال می شوند و از آنجا به سمت خرده فروش ها و در نهایت به دست مصرف کننده می رسند [۵۷]. مدیریت زنجیره تامین شامل مجموعه عناصر تامین کنندگان، تهیه کنندگان سرویس لجستیک، سازندگان، توزیع کنندگان و فروشندگان است که جریان نهایی مربوط به مواد خام، محصولات و جریان اطلاعاتی در بین این عناصر وجود دارد [۵۸].

مدیریت زنجیره تامین عبارت است از انتقال و ذخیره سازی مواد خام، موجودی در جریان ساخت و کالای تمام شده از نقطه تولید به نقطه مصرف. مدیریت زنجیره تامین مجموعه ای است از راهکارها جهت یکپارچه سازی اعضای زنجیره تامین (تامین کنندگان، تولید کنندگان، توزیع کنندگان، خرده فروشی ها و مشتری نهایی) است که هدف آن کاهش هزینه های سیستم و نیز افزایش سطح خدمت دهی به مشتریان است [۵۷]. e-SCM یک سری عملیات، راه حل ها و طرحی جامع و کلی بوسیله اینترنت است که با یکپارچه سازی روند محصول از تامین کننده، تولید کننده، عمده فروش، خرده فروش به مشتری به عنوان یک زنجیره تامین به سیستم اجازه می دهد اطلاعاتی همانند تولید، سهام و فروش در میان شرکت بلافاصله به اشتراک گذاشته شود [۵۵] و در جای دیگر آمده است که e-SCM همان تاثیر اینترنت بر یکپارچگی فرایندهای کلیدی کسب و کار تا مصرف نهایی محصول یا خدمت که از طریق تولید محصولات، خدمات و اطلاعات از سوی تامین کنندگان اصلی، ارزش افزوده ای را برای مشتریان و دیگر ذینفعان فراهم می نماید [۵۹].

تعاریف دیگری که در منابع مختلف از سیستم های مدیریت زنجیره تامین الکترونیکی عبارتند از:

- چنانچه یک زنجیره تامین به صورت الکترونیکی و با کمک نرم افزارهای کامپیوتری مبتنی بر وب مدیریت شود، مدیریت زنجیره تامین الکترونیکی نامیده می شود.
 - مدیریت زنجیره تامین الکترونیکی تلاشی در جهت تسهیل جریان اطلاعات درون زنجیره تامین است.
 - تدارکات الکترونیکی به معنی یکپارچگی و مدیریت الکترونیکی در انجام کلیه فعالیت های مربوط به تدارکات می باشد که شامل درخواست خرید، تایید، سفارش، تحویل و پرداخت بین خریدار و تامین کننده با بهره گیری از شبکه های ارتباطی است.
 - **نقش اطلاعات و مدیریت اطلاعات در زنجیره تامین:** مدیریت اطلاعات یعنی گردش و انتقال صحیح اطلاعات، که باعث می شود فرایندهای موثر در زنجیره تامین کارا تر و آسان تر شود. مدیریت اطلاعات و مجموعه سیستم های اطلاعاتی زنجیره تامین می توانند بر روی بسیاری از تصمیم گیری های داخلی بخش های مختلف زنجیره تامین موثر باشند که این موضوع ناشی از اهمیت بالای این مولفه در مدیریت زنجیره تامین می باشد.
- از جمله کاربردهای اطلاعات در زنجیره تامین شامل وارد نمودن سفارشات، پردازش سفارشات، کنترل انبار موجودی کالا، اندازه گیری عملکرد، نظارت بر حمل و نقل کالا، انجام پرداخت های مالی، ذخیره سازی کالا است [۶۰].

۲-۵. چالش و شکاف موجود در سیستم های مدیریت زنجیره تامین الکترونیکی

- **چالش های سیستم های مدیریت زنجیره تامین الکترونیکی و کلان داده:** چنانچه فعالیت افراد موثر در زنجیره تامین از طریق اینترنت و ابزارهای مبتنی بر وب مانند اپلیکیشن های تعاملی و شبکه های اجتماعی صورت گیرد مهمترین چالش نحوه ذخیره سازی و پردازش ساختارهای مختلف داده و برخورد با کلان داده ها است. این چالش ها را می توان به سه دسته تقسیم کرد:
 - ذخیره و پردازش داده های چند ساختاری و کلان داده ها
 - یکپارچه سازی داده های ساختاریافته و داده های چندساختاری
 - تجزیه و تحلیل داده های چند ساختاری و کلان داده ها

سیستم های مدیریت زنجیره تامین الکترونیکی فعلی توانایی پشتیبانی از کلان داده را بنحو مطلوبی ندارند و دارای نواقصی هستند که ازجمله این نواقص می توان به کندی عملکرد و هزینه های بالای آنها اشاره کرد. تجزیه و تحلیل پیشرفته که در معماری کلان داده ابزارهای خاصی برای آن طراحی و پیاده سازی شده شامل داده کاوی، خوشه بندی، تحلیل های آماری و تحلیل احساسات می باشد که سیستم های مدیریت زنجیره تامین الکترونیکی توانایی انجام چنین تجزیه و تحلیلهایی را برای داده های بدون ساختار و کلان داده ها ندارد.

از طرف دیگر سیستم های مدیریت زنجیره تامین الکترونیکی، داده های چند ساختاری بخصوص داده های با محوریت فایل^۱ را به خوبی پشتیبانی نمی کند. این چالشها را به کمک معماری کلان داده تا حد زیادی می توان مرتفع کرد.

سازمانها باید نقش کلان داده را در ارتباط با تصمیم گیری درک نمایند با توجه به اینکه کلان داده می تواند فرصت های جدیدی را برای تصمیم گیری پیش روی ما قرار دهد. در جهانی که مصرف کنندگان آن هر ساعت تغییر می کنند در این حالت تجزیه و تحلیل کلان داده از چندین کانال ارتباطی می تواند پروفایلهای مشتریان و رفتارهای تصمیم گیری را بصورت همزمان ردیابی نماید. پیشرفت ابزارهای کلان

داده و نیازمندی به آنها به حدی است که طبق یک نظر سنجی معتبر پیش بینی می شود تا سه سال آینده بیش از ۵۰ درصد شرکتها و سازمانها از معماری کلان داده استفاده خواهند کرد.

● **مقایسه سیستم مدیریت پایگاه داده^۱ و سیستم فایل توزیع شده هدوپ^۲:** مهمترین ابزار معماری کلان داده که تا کنون طراحی و پیاده سازی شده است سیستم فایل توزیع شده هدوپ می باشد.

سیستم فایل توزیع شده هدوپ دارای یک سیستم فایل توزیع شده است نه یک سیستم مدیریت پایگاه داده. اما برخی از قابلیت های سیستم مدیریت پایگاه داده را نیز دارا می باشد. قابلیت هایی چون فهرست بندی ، دسترسی تصادفی به اطلاعات ، پشتیبانی از زبان SQL و بهینه سازی پرس و جو. البته عملکرد برخی از قابلیت های سیستم فایل توزیع شده هدوپ بهتر از سیستم مدیریت پایگاه داده می باشد قابلیت هایی چون مدیریت حجم گسترده ای از داده های مبتنی بر فایل و مدیریت داده های بدون ساختار و آرشیو و دسته بندی کردن داده ها.

در سیستم های مدیریت زنجیره تامین الکترونیکی از سیستم مدیریت پایگاه داده در جهت مدیریت و ذخیره داده ها استفاده می شود و در کلان داده از سیستم توزیع شده فایل هدوپ برای اینکار استفاده می گردد. البته سیستم توزیع شده فایل هدوپ در کنار ابزارهای دیگری همچون Impala ، HBase مانند یک سیستم مدیریت پایگاه داده عمل میکند.

از جمله ضعف های سیستم مدیریت پایگاه داده عدم توانایی ذخیره کلان داده است و از جمله ضعف های سیستم توزیع شده فایل هدوپ دقت پایین تر نسبت به سیستم مدیریت پایگاه داده می باشد به این ترتیب بهترین ابزار برای ذخیره کلان داده سیستم توزیع شده فایل هدوپ و بهترین ابزار برای داده های ساختار یافته سیستم مدیریت پایگاه داده می باشد پس ایندو می توانند مکمل خوبی برای ذخیره انواع داده ها باشند و نه جایگزینی برای یکدیگر.

اگر از ابزارهای موجود در معماری کلان داده بعنوان مکمل سیستم مدیریت پایگاه داده استفاده کنیم عملکرد سیستم مدیریت پایگاه داده نیز بهتر خواهد شد زیرا سیستم مدیریت پایگاه داده در برخی از قسمتها بسختی کار میکند بخصوص اگر داده ها، بدون ساختار یا نیمه ساختار باشند، اما سیستم توزیع شده فایل هدوپ در چنین قسمتهایی عملکرد خوبی دارد. قطعاً اگر منابع سیستم مدیریت پایگاه داده برای انجام کار بر روی داده های ساختار یافته متمرکز شود و کلیه منابع آن برای این نوع داده ها آزاد شود عملکرد و کارایی بهتری خواهد داشت نسبت به زمانی که منابع آن برای ذخیره و مدیریت همه نوع داده ای صرف شود. از طرف دیگر، در اغلب موارد، داده های عظیم ذاتاً فایل هستند مانند لاگهای وب که در قالب فایل های مستند هستند و اسناد شخصی و یا داده های ناشی از حسگرهای موجود در ماشین آلات کارخانجات، فروشگاه ها و یا داده های موجود در سازمانهایی مانند سازمان آب و برق و...

بسیاری از سازمانها سیستم مدیریت پایگاه داده خود را برای ساختار پایگاه داده رابطه ای طراحی کرده اند بطوریکه ذخیره و مدیریت داده های بدون ساختار برای آنها دشوار است.

از جمله داده هایی که سیستم توزیع شده فایل هدوپ می تواند بخوبی ذخیره و مدیریت نماید فایل های مربوط به جزئیات داده، انبوهی از داده های کوچک و جزئی، همچنین تجزیه و تحلیل این نوع داده هاست.

از آنجا که منابع سیستم مدیریت پایگاه داده هزینه بر هستند و از طرفی سیستم توزیع شده فایل هدوپ بصورت کد باز^۱ رایگان (یا ارزان قیمت) عرضه می شود. با استفاده از منابع سیستم توزیع شده فایل هدوپ و آزاد کردن منابع سیستم مدیریت پایگاه داده می توان صرفه جویی قابل توجهی در هزینه ها را به همراه داشت.

سیستم توزیع شده فایل هدوپ یک ابزار پایه ای برای کتابخانه هدوپ (بزرگترین مجموعه ابزار پیاده سازی شده برای معماری کلان داده) محسوب می شود که بقیه ابزارهای هدوپ روی آن سوار می شوند.

می توان از سیستم توزیع شده فایل هدوپ بعنوان مکمل سیستم مدیریت پایگاه داده استفاده کرد بطوریکه در جاهایی که سیستم توزیع شده فایل هدوپ از سیستم مدیریت پایگاه داده عملکرد بهتری دارد از آن بهره گرفت. در سیستم توزیع شده فایل هدوپ داده ها بصورت خام و بشکل منبع اولیه خودشان نگهداری می شوند در حالی که در سیستم مدیریت پایگاه داده چنین نیست. سیستم توزیع شده فایل هدوپ از قابلیت مقیاس پذیری و چند ساختاری بودن داده ها بخوبی پشتیبانی می کند. سیستم مدیریت پایگاه داده دارای دقت بالایی است از طرفی سیستم توزیع شده فایل هدوپ میتواند امکانات تجزیه و تحلیل سازمانی را ارتقا دهد .

بسیاری از سازمانها در حال حاضر داده هایی دارند که سیستم مدیریت پایگاه داده موجود در سیستم های مدیریت زنجیره تامین الکترونیکی سازمان نمی تواند آنها را پردازش کند ازجمله تماسهای ضبط شده مربوط به روابط عمومی سازمان، اسناد XML مربوط به مدیریت زنجیره تامین الکترونیکی سازمان ، فایلهای حاوی لاگ سنسورها و ماشین آلات، و دیگر داده های بدون ساختار که در سازمان ممکن است وجود داشته باشد و این درحالی است که معماری کلان داده بخوبی می تواند این داده ها را ذخیره و پردازش نماید.

در یک نظر سنجی تمایل شرکتها و سازمانها به استفاده از قابلیتهای مختلف سیستم توزیع شده فایل هدوپ بررسی شده که نتایج آن در شکل ۱۴ آمده است.

همانطور که در شکل مشاهده می شود ۷۸ درصد از شرکتها و سازمانهای شرکت کننده در این نظرسنجی تمایل دارند از سیستم توزیع شده فایل هدوپ بعنوان مکمل سیستم مدیریت پایگاه داده بویژه برای تجزیه و تحلیل پیشرفته استفاده کنند.

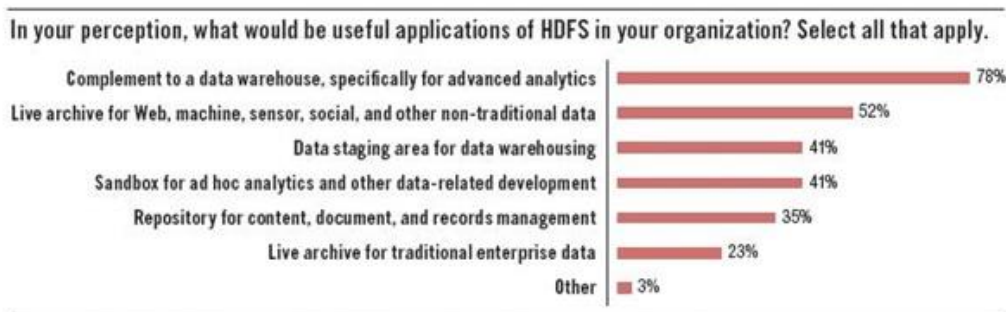
۵۲ درصد، بعنوان آرشیو داده های وب ، داده های ناشی از ماشین آلات و سنسورها و داده های غیرسنتی

۴۱ درصد، بعنوان پایگاه داده های رابطه ای ها

۴۱ درصد، فقط بعنوان تجزیه و تحلیل پیشرفته

۳۵ درصد، بعنوان ذخیره و مدیریت محتوا و رکوردها

۲۳ درصد، بعنوان بایگانی فعال برای ذخیره داده های سنتی



شکل ۱۴. تمایل سازمانها در استفاده از قابلیتهای هدوپ [۲۱]

معماری کلان داده اغلب از روشهای تجزیه و تحلیل اکتشافی، داده کاوی، تحلیل های آماری و پرس و جوهای پیچیده SQL و ... برای تجزیه و تحلیل کلان داده استفاده می کند مهمترین مزیت معماری کلان داده نسبت به سیستم های مدیریت زنجیره تامین الکترونیکی مقیاس پذیری و تنوع داده ها است.

سیستم توزیع شده فایل هدوپ می تواند علاوه بر داده های بدون ساختار و نیمه ساختار، داده های ساختار یافته را همراه با عملیات ETL در خود ذخیره نماید اما ذخیره و مدیریت داده های ساختار یافته در سیستم مدیریت پایگاه داده با عملکرد بهتری همراه است از طرف دیگر سیستم مدیریت پایگاه داده هم می تواند داده های بدون ساختار و نیمه ساختار را علاوه بر داده های ساختار یافته، مدیریت و ذخیره نماید اما عملکرد سیستم توزیع شده فایل هدوپ برای اینکار از سیستم مدیریت پایگاه داده بهتر است و از طرفی هزینه کمتری را نیز بدنال دارد.

سیستم توزیع شده فایل هدوپ با کمک ابزارهایی چون HBase کار ذخیره و مدیریت داده ها را انجام می دهد. جهت نمایش اطلاعات و گزارشگیری می توان از رابطهای کاربری سیستم های مدیریت زنجیره تامین الکترونیکی یا ابزارهای طراحی شده در معماری کلان داده استفاده کرد. تفاوت ابزارهای بصری کردن اطلاعات در کلان داده و سیستم های مدیریت زنجیره تامین الکترونیکی این است که در کلان داده ابزارها بطورمستقیم با داده ها در ارتباطند اما در سیستم های مدیریت زنجیره تامین الکترونیکی اینگونه نیست. یکی از مهمترین ابزارها در تجزیه و تحلیل کلان داده ابزار نگاشت/کاهش است که هم میتوان آنرا روی سیستم توزیع شده فایل هدوپ سوار کرد و هم می توان بصورت مستقل از آن استفاده کرد.

۳-۵. ضرورت ادغام معماری سیستم های مدیریت زنجیره تامین الکترونیکی و معماری کلان داده

هر سازمان بمنظور بهبود فرایند تصمیم گیری نیاز به استفاده از یک ساختار اطلاعاتی یکپارچه و پویا دارد برای رسیدن به چنین ساختاری مواردی چون جمع آوری و ذخیره داده ها، تجزیه و تحلیل و تبدیل نتایج به اطلاعات مفید و با ارزش از اهمیت بالایی برخوردار است. برای انجام چنین فرایندهای تحلیلی لازم است از یک محیط مناسبی شامل یک مخزن بزرگ و سراسری برای داده ها، یک پردازش قوی و هوشمند و یک رابط کاربری کاربرپسند استفاده شود.

مخزن داده ها از انواع داده ها، اعم از منابع داخلی و خارجی سازمان، پر می شود و با توجه به فعالیتهای شرکت یا سازمان این داده ها بصورت موضوعی بسته بندی شده و در مخازن و پایگاه داده های رابطه ای ذخیره می شود و اخیراً کلان داده نیز به این داده ها باید اضافه شود.

سازمان با ادغام سیستم توزیع شده فایل هدوپ و سیستم مدیریت پایگاه داده می تواند از تمام داده ها برای افزایش ارزش کسب و کار استفاده کرده و هزینه های مدیریت داده ها را کاهش دهد. بعبارت دیگر هدف از ادغام سیستم های مدیریت زنجیره تامین الکترونیکی و معماری کلان داده ، ارتقای سطح ارزش در کسب و کار سازمان یا شرکت، ایجاد مزیت رقابتی، کاهش هزینه ها، بهره وری بیشتر و ارتباط نزدیک تر با مشتریان است و همچنین استفاده از ارزش تمام داده ها نه فقط استفاده از ارزش بخشی از داده ها می باشد.

مهمترین فرصتهایی که معماری کلان داده برای ما فراهم می کند عبارت است از: مقیاس پذیری، تجزیه و تحلیل اکتشافی، هزینه کم، پشتیبانی از داده های چند ساختاری.

با ادغام و یکپارچه سازی سیستم توزیع شده فایل هدوپ و سیستم مدیریت پایگاه داده، این فرصتها مشترک خواهند شد از طرف دیگر برخی از قابلیتهای سیستم های مدیریت زنجیره تامین الکترونیکی و کلان داده ها نیز ارتقا خواهند یافت از جمله: رده بندی داده ها، آرشیو و بایگانی داده های منبع، مدیریت داده های غیرساختار یافته، مدیریت داده های مبتنی بر فایل، ارتقا قدرت پردازش داده ها، مدیریت و ذخیره بهتر داده ها.

از آنجایی که ابزارهای پیاده سازی شده در معماری کلان داده زیاد هستند زمانی که می خواهیم کار یکپارچه سازی و ادغام معماری کلان داده و معماری سیستم مدیریت زنجیره تامین الکترونیکی را انجام دهیم باید با توجه به نیازهایمان ابزارهای مورد نیاز معماری کلان داده را انتخاب نماییم سپس کار ترکیب آنها را با لایه های معماری سیستم های مدیریت زنجیره تامین الکترونیکی انجام دهیم. این ادغام می تواند منجر به بهبود فرایندهای کسب و کار و بهبود طرح های کسب و کار برای رسیدن به اهداف سازمان و شرکت باشد.

بهره برداری از معماری مشترک می تواند منجر به تصمیمات دقیقتری در سازمان شود و می تواند از شکست سیستم در آینده جلوگیری کند. ادغام معماری کلان داده و سیستم های مدیریت زنجیره تامین الکترونیکی امکان بهره برداری بیشتری از داده ها را فراهم می کند زیرا این فرصت را ایجاد می کند که مقدار داده بیشتری مورد تجزیه و تحلیل قرار گیرد و به این ترتیب نتایج بهتری نسبت به زمانی که از معماری کلان داده استفاده نمی شود، بدست می آید.

معماری کلان داده این امکان را فراهم می کند که ارزش بیشتری از داده ها برگردانده شود. شرکت و سازمانها باید درک کنند که این ادغام می تواند فرصتهای جدیدی را پیش روی آنها بگشاید. علاوه بر آن کلان داده این امکان را فراهم می کند که داده های بیشتری مورد ارزش گذاری و بهره برداری قرار گیرند و همچنین می تواند در فرایندهای مدیریتی، تصمیم گیری، یادگیری سازمان و در پیش بینی آینده کسب و کار و در ارتباط با مشتریان تاثیرگذار باشد و با این ادغام می توان دانش ضمنی بیشتری را به دانش صریح و آشکار تبدیل کرد.

کلان داده در تمام رشته ها به یک مسئله پیچیده ای تبدیل شده زیرا در هر رشته ای می تواند این امکان را فراهم کند که از تمام داده های موجود در آن رشته با هر ساختاری، بهره برداری شود بعبارت دیگر امکان بهره برداری از تمام پتانسیل هر رشته را می دهد. برخی از نتایج ناشی از ادغام کلان داده و سیستم های مدیریت زنجیره تامین الکترونیکی عبارت است از :

- بهبود فرایندهای زنجیره تامین الکترونیکی، طرح های کسب و کار و دستیابی به اهداف کسب و کار در سازمان هدف (سازمان هدف می تواند یک شرکت، صنعت، آموزش و پرورش، یک سیستم مالی یا سیستم جهانی باشد)
- کاهش داده های پراکنده و استفاده از داده های کوچکتر اما با ارزش بیشتر بخصوص داده های مربوط به مشتریان
- امکان اخذ تصمیم گیریهای دقیق تر
- پیش بینی آینده جهت جلوگیری از شکست سیستم

۵- نتیجه گیری و پیشنهاد کارهای آتی

ما در این تحقیق ابتدا معماری کلان داده را تشریح کردیم و سپس معماری سیستم های مدیریت زنجیره تامین الکترونیکی را بیان کردیم. در ادامه با بررسی چالشهای موجود در سیستم های مدیریت زنجیره تامین الکترونیکی در برخورد با کلان داده ها و داده های چند ساختاری به این نتیجه رسیدیم که به کمک معماری کلان داده می توان بر بسیاری از این چالشها غلبه کرد. همچنین به این نتیجه رسیدیم که ادغام معماری کلان داده و معماری سیستم های مدیریت زنجیره تامین الکترونیکی برای کلیه شرکتها و سازمانهایی که می خواهند از تمام داده های خود ارزش آفرینی کنند و بیشترین بهره را از داده های خود ببرند تا در عرصه های رقابتی موفق تر باشند، یک ضرورت محسوب می شود. مهمترین هدف از این مقاله ترویج دیدگاه یکپارچه سازی کلان داده ها، هوش کسب و کار و سیستم های مدیریت زنجیره تامین الکترونیکی در چارچوب فرایند تصمیم گیری برای کمک به مدیران جهت ایجاد فرصتهای جدید در حل مشکلات و مسائل خاص می باشد.

پیشنهاد می شود در کارهای آتی بدنبال ساختار جدید ناشی از یکپارچه سازی معماری کلان داده ها، هوش کسب و کار و سیستم های پشتیبانی از تصمیم برای رسیدن به یک هدف خاص باشیم مثلاً ادغام معماری کلان داده و سیستم های پشتیبانی از تصمیم جهت رسیدن به تصمیمات بهتر و دقیقتر یا برای رسیدن به پاسخ های بلادرنگ و ...

مراجع

- [۱] K. Merja¹, Prof. Rupesh G. Vaishnav , *SURVEY ON IMPROVE JOB PERFORMANCE USING MAPREDUCE IN HADOOP* , International Journal of Advance Engineering and Research Development , Volume 1, Issue 12, December ,2014 , pp.37-41.
- [۲] Seyed Reza Pakize, *A Comprehensive View of Hadoop MapReduce Scheduling Algorithms*, International Journal of Computer Networks and Communications Security, VOL. 2, NO. 9, SEPTEMBER 2014, pp.308–317.
- [۳] Revathi S, Rajkumar N, Sathish S, *A Text Sentimental Approach for Online Portals Using Hadoop* , International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February 2015, pp.1023-1028.
- [۴] Kankana Kashyap, Champak Deka, Sandip Rakshit, *A Review on Big Data, Hadoop and its Impact on Business*, INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT, Volume 3 Issue 12, November 2014, pp.78-82.
- [۵] Shraddha Thakkar, Prof. Sanjay Patel, *Improving Map Reduce Performance in Heterogeneous Distributed System using HDFS Environment-A Review*, International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 3, 2015, pp.903-9۱۰.
- [۶] DR. A. N. Nandakumar¹, Nandita Yambem, *A Survey on Data Mining Algorithms on Apache Hadoop Platform*, International Journal of Emerging Technology and Advanced Engineering , Volume 4, Issue 1, January 2014 , pp.263-265.
- [۷] Mehdi Bahrami and Mukesh Singhal , *The Role of Cloud Computing Architecture in Big Data* , Information Granularity, Big Data, and Computational Intelligence Studies in Big Data Volume 8, 2015, pp 275-295.
- [۸] 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, December 2012, pp. ۱۱۶۵-۱۱۸۸.

- [٩] Mrs. Malini M Patil ,*Challenges and Issues in Handling Big Data* , International Journal of Innovations & Advancement in Computer Science IJIACS , Volume 4, Special Issue, March 2015 , pp 620 – 625.
- [١٠] Songting Chen , *Cheetah: A High Performance, Custom Data Warehouse on Top of MapReduce*, *Proceedings of the VLDB Endowment*, Vol. 3, No. 2 , 2010 , pp 1459-1468.
- [١١] Rogério Rossi & Kechi HIRAMA, *Characterizing Big Data Management*, Issues in Informing Science and Information Technology Volume 12, 2015 , pp 165-180.
- [١٢] Thiago Poletto, Victor Diogho Heuer de Carvalho, and Ana Paula Cabral Seixas Costa ,*The Roles of Big Data in the Decision-Support Process: An Empirical Investigation*, B. Delibašić et al. (Eds.): IC DSS T 2015, LN DSS P 216, 2015, pp10–21.
- [١٣] IBM. (2012). *IBM Global Technology Outlook-2012*. Retrieve from http://www.progressivepolicy.org/wp-content/uploads/2013/09/09.2013-Mandel_Can-the-Internet-of-Everything-Bring-Back-the-High-Growth-Economy-1.pdf on September 2013: ١٢٠٠٠٠٠٠
- [١٤] Intel. (2014). *what happens in an internet minutes site?*. Retrieve from <http://www.intel.com/content/dam/www/public/us/en/images/illustrations/embedded-infographic-600-logo.jpg> on 2014: 1 page
- [١٥] Jing Han, Haihong E, Guan Le, Jian Du. (2011). "Survey on NoSQL Database". Pervasive Computing and Applications (ICPCA), 2011 6th International Conference on, 26-28 Oct., 2011, Beijing, China, pp. 363 – 366.
- [١٦] Sharda R, Delen D, Turban E. (2015). "*Business Intelligence and Analytics: Systems for Decision Support*". 10/E, Prentice Hall, New Jersey, USA: 688 pp
- [١٧] Wikipedia. (٢٠١٥). "*Apache Hadoop*". Retrieve from https://en.wikipedia.org/wiki/Apache_Hadoop
- [١٨] Dwivedi, K, Dubey, S.K. (2014). "*Analytical Review on Hadoop Distributed File System*". Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference, 25-26 Sept. 2014, Noida, India, pp. 174 – 181.
- [١٩] Dean J, Ghemawat S, "*MapReduce: Simplified Data Processing on Large Clusters*", Communications of the ACM - 50th anniversary issue: 1958 – 2008, Volume 51 Issue 1, January ٢٠٠٨, pp 107-113.

- [20] Glenn K. Lockwood.(2015). “*Conceptual Overview of Map-Reduce and Hadoop*“. Retrieve from <http://www.glennklockwood.com/data-intensive/hadoop/overview.html> on **October 9, 2015**.
- [21] Russom F. (2013). “*INTEGRATING HADOOP INTO BUSINESS INTELLIGENCE AND DATA WAREHOUSING*”, Retrieve from www.sas.com/resources/asset/Hadoop.pdf
- [22] Apache. (2015). “*Apache DataFu Hourglass*”. Retrieve from <http://datafu.incubator.apache.org/docs/hourglass/concepts.html>
- [23] P. A. Goloboff, ”*Techniques for analyzing large data sets*”, Techniques in Molecular Systematic and Evolution, ISBN 978-3- 0348-8125-8, pp(ⅴ-ⅴⅳ)
- [24] S. Madden, MIT, “*From databases to big data*”, IEEE Computer Society, 1089-7801/12, 2012, (pp ⅴ)
- [25] R. Gupta, “*Journey from Data Mining to Web Mining to Big Data*”, International Journal of Computer Trends and Technology (pp 18).
- [26] *what are big data techniques and why do you need them*, FCW, big data techniques, <http://www.fcw.com>,(pp 2)
- [27] *Hadoop: What it is and why it matter*, SAS, <http://www.sas.com>, (pp 1/12)
- [28] T. White, “*Hadoop: The Definitive Guide*”, Third Edition, O’rielly|yahoo press, ISBN 978-1-449-31152-0.
- [29] J. Dittrich, Jorge-Arnulfo Quiñe-Ruiz, “*Efficient Big Data Processing in Hadoop Map Reduce*”, (pp 2014)
- [30] B. Proffitt , “*Hadoop: What It Is and How It Works*”, website <http://www.readwrite.com,2013>.
- [31] *Big Data & Hadoop Solutions*, Attunity website, <http://www.attunity.com>.
- [32] Top 14 hadoop technology companies, technavio’s blog-website, <http://www.technavio.com>
- [33] M. Nemschoff,” *5 major advantages of Hadoop*”, <http://www.itproportal.com>.
- [34] S. Prakash, “*Storing and querying big data in Hadoop distributed file system (HDFS)*.”,sri’s technology blog, website <http://www.ecomcanada.wordpress.com>.
- [35] M. Whitehorn, “*What is the difference between DBMS and RDBMS?*” university of Dundee, website <http://www.searchdatamanagement.techtarget.com>.

- [३६] D. Borthakur, “*The Hadoop Distributed File System: Architecture and Design.*” The Apache Software Foundation, 2007.
- [३७] B. Butler, “*Nine Hadoop companies you should know*”, network world Inc., 2014.
- [३८] S. Ozcan, Database In Action -Integrated Approach to Oracle, “*Difference between Hadoop and RDBMS*”, website <http://www.oraclesys.com>.
- [३९] D. Weldon, “*Examining business cases for Hadoop*”, industry insider M. Nemschoff, 2014.
- [४०] Chen, M., Mao, S., Liu, Y.: *Big data: a survey*. Mob. Netw. Appl. **19**, 171–209 (२०१४)
- [४१] Simon, H.A.: *The New Science of Management Decision*. Harper and Row, New York (१९६०)
- [४२] Renu, R.S., Mocko, G., Koneru, A.: *Use of big data and knowledge discovery to create data backbones for decision support systems*. Procedia Comput. Sci. **20**, 446–453 (२०१३)
- [४३] Berman, J.J.: *Principles of Big Data Preparing, Sharing, and Analyzing Complex Information*. Elsevier, Waltham (२०१३)
- [४४] Klein, D., Tran-Gia, P., Hartmann, M.: *Big data*. Informatik-Spektrum **36**, 319–323 (२०१३)
- [४५] Zikopoulos, P.C., Eaton, C., deRoos, D., Deutsch, T., Lapis, G.: *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. McGraw Hill, New York (२०१२)
- [४६] Grillenberger, A., Fau, F.E.: *Big data and data management: a topic for secondary computing education*. In: ICER 2014 – Proceedings of the 10th Annual International Conference on International Computing Education Research, pp. 147–148 (२०१४)
- [४७] Chang, R.M., Kauffman, R.J., Kwon, Y.: *Understanding the paradigm shift to computational social science in the presence of big data*. Decis. Support Syst. **63**, 67–80 (२०१४)
- [४८] Dobre, C., Xhafa, F.: *Intelligent services for big data science*. Futur. Gener. Comput. Syst. **37**, २६७–२८१ (२०१४)
- [४९] Liu, S., Duffy, A.H.B., Whitfield, R.I., Boyle, I.M.: *Integration of decision support systems to improve decision support performance*. Knowl. Inf. Syst. **22**, 261–286 (२००९)

- [۵۰] Dong, C.-S.J., Srinivasan, A.: *Agent-enabled service-oriented decision support systems*. *Decis. Support Syst.* **55**, 364–373 (۲۰۱۳)
- [۵۱] Daas, D., Hurkmans, T., Overbeek, S., Bouwman, H.: *Developing a decision support system for business model design*. *Electron. Mark.* **23**, 251–265 (۲۰۱۲)
- [۵۲] Popovič, A., Hackney, R., Coelho, P.S., Jaklič, J.: *Towards business intelligence systems success: effects of maturity and culture on analytical decision making*. *Decis. Support Syst.* **54**, 729–739 (۲۰۱۲)
- [۵۳] Lancioni Richard, Hope Jensen Schau and Michael F. Smith, "Internet impacts on supply chain management," *Industrial Marketing Management*, pp. 32,173-175, 2003.
- [54] Patterson Krik A, Grimm Curtis M and Corsi Thomas M, "Adopting new technologies for supply chain management," *Transportation Research part 39*, pp. 95-121, 2003.
- [55] S. I. Chang and Y. F. Tsia, "Critical Factors of E-SCM Adoption on Small- and Medium-Sized Enterprises," *International Journal of Electronic Business Management*, vol. 1, no. 2, p. 1, 2006.
- [56] Kopezak, L.R., "Logistic Partnership and Supply chain restructuring: survey results from the US Computer industry," in *Production and Operation Management* 6(3), US, 1997, pp. 226-247.
- [۵۷] تیموری، ابراهیم، احمدی، مهدی. "زنجیره تامین چیست؟"، در مدیریت زنجیره تامین، تهران، انتشارات دانشگاه علم و صنعت، ۱۳۸۸، 4-5 PP.
- [58] Kopezak, L.R., "Logistic Partnership and Supply chain restructuring: survey results from the US Computer industry," in *Production and Operation Management* 6(3), US, 1997, pp. 226-247.
- [59] M. C. Cooper , D. M. Lambert and e. a. , "Supply Chain Management: More than a new name for Logistics," *The International Journal of Logistics Management*, vol. 8, no. 1, pp. 1-13, 1997.
- [۶۰] سلیمانی سه دهی، مجتبی. "اهمیت فناوری اطلاعات در توسعه لجستیک و زنجیره تامین"، ماهنامه علمی-تخصصی لجستیک و زنجیره تامین، شماره ۱۸، صفحه ۳. زمستان ۱۳۹۳