



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

انبار داده ها و OLAP داده های بزرگ: چالشهای موجود و سمت و سوی مطالعات آتی

خلاصه

در این مقاله ما به مسائل موجود و سمت و سوی مطالعاتی در زمینه ی انبار داده ها و OLAP داده های بزرگ میپردازیم. "داده های بزرگ" اصطلاحی است که اخیرا در این حوزه مطرح شده است. همچنین به افق های مطالعاتی پیش روی جدیدی در این حوزه اشاره خواهیم داشت و نتایج مثبتی که طی این نوع مطالعات حاصل خواهد شد را مطرح خواهیم کرد.

این مطالعه زیرمجموعه ی زمینه های زیر است:

مدیریت پایگاه داده ها و انبار داده ها

اصطلاحات کلی

الگوریتم ها، طراحی، مدیریت، عملکرد، تئوری

اصطلاحات کلیدی

داده های بزرگ، داده های چندبعدی بزرگ، انبار داده ها، OLAP

1. مقدمه

در ادبیات تحقیقی این حوزه علاقه ی زیادی به "داده های بزرگ" ایجاد شده است. این اصطلاح به انواع خاصی از مجموعه داده ها اشاره دارد که در اپلیکیشن های محاسبات علمی از آنها استفاده میشود. داده های ذخیره شده در لایه ی اصلی تمامی این اپلیکیشن ها یک سری از مشخصات مشترک دارند مثلا: (1) داده های بزرگ مقیاس که به حجم و توزیع مخازن داده اشاره دارد (2) مسائل مقیاس پذیری که مربوط به قابلیت های اپلیکیشن هایی است که برای مخازن داده ی حجیم و بزرگ مقیاس کاربرد دارند. (3) پشتیبانی از پروسه های ETL (ExtractionTransformation-Loading) (4) طراحی و ارائه ی تجزیه و تحلیل آسان و قابل تفسیر مخازن داده های بزرگ بمنظور حصول آگاهی و دانش مفید.

به سبب مشخصات خاص کاربردهای داده های بزرگ استفاده ی از روشهای انبار داده ها و OLAP با هدف جمع آوری، استخراج، انتقال، بارگذاری، انبار و OLAP این نوع از پایگاه های داده امری کاملا منطقی است. و مبحثی کاملا جدید در حوزه ی مطالعات پایگاه داده و انبار داده ها است.

انبار داده ها و OLAP زمینه های علمی سابقه داری در این رابطه ها هستند که توسط محققان جامعه ی علمی انبار داده ها و پایگاه داده برای دهه های متمادی مورد تحقیق و تفحص واقع شده اند. مشکل اصلی در حوزه ی محاسبه ی داده ها در انواع گسترده ای از مطالعات مطرح شده است.

با ظهور حوزه ی مطالعاتی داده های بزرگ کاملا طبیعی است که مشکل محاسبه ی مکعب های داده را یکی از جالب ترین و موردتوجه ترین چالش های پیش روی جامعه ی علمی محققین داده تلقی کنیم. که باید حصول فناوریهای قدرتمند در اپلیکیشن ها و سیستم های بزرگ مقیاس در زندگی واقعی را نیز در این رابطه مدنظر قرار دهیم.

متأسفانه روشهای کنونی قابلیت های لازم جهت محاسبه ی مکعب های داده ی OLAP را ندارند که عمدتاً به سبب دو مشخصه ی اصلی مخازن بزرگ داده است: (1) حجم، که در این نوع پایگاه های داده خیلی بالاست. (2) پیچیدگی (مدل های داده ی چندبعدی)، که در این نوع پایگاه های داده میتواند خیلی زیاد باشد.

در نتیجه برای طراحی مدل ها، تکنیک ها، الگوریتم ها و پایگاه های محاسباتی جدید جهت مدیریت مشکل محاسبه ی مکعب های داده ی OLAP در داده های بزرگ نیازی مبرم احساس میشود. ضرورت دارد که برای مطالعات آتی انبار داده ها و OLAP از هم اکنون آماده باشیم.

به همین سبب است که در این مقاله به مشکلات و سمت و سوی مطالعات کنونی در حوزه ی انبار داده ها و OLAP داده های بزرگ میپردازیم. تا افق های پژوهشی جدیدی را معرفی کنیم و بر رهیافت های امیدبخش آتی حاصل از این نوع تحقیقات اشاره ای داشته باشیم.

2. چالش های تحقیقی انبار داده ها و OLAP داده های بزرگ

در حین محاسبه ی مکعب های داده ی OLAP در داده های بزرگ با چالش های متعددی روبرو میشویم:

- حجم: جدول های اطلاعات در محاسبه ی مجموعه داده های بزرگ براحتی میتوانند حجیم گردند. وقتی حجم بالا باشد در حین محاسبه با مشکلات عدیده ای مواجه خواهیم بود.
- پیچیدگی: در ساخت مکعب های داده ی OLAP در داده های بزرگ با مشکلاتی مواجه هستیم که در OLAP قبلی وجود ندارند. مثلا با توجه به ماهیت غیرساختارمند پایگاه های داده ی بزرگ، تعداد ابعاد براحتی میتواند بیش از اندازه بوده و دردرساز شود. اقدامات متعددی در رابطه با چنین مکعب های داده ای میتوان انجام داد.
- طراحی: روشهای طراحی مکعب های داده ی OLAP در مطالعات انبار داده و پایگاه داده مورد توجه بالایی بوده است. در روشهای طراحی OLAP در داده های بزرگ به عملکرد باید اهمیت زیادی داده شود. در این موارد طراحان بایستی به سوالات اساسی زیر توجه لازم را داشته باشند: (i) زمان لازم برای طراحی مکعب داده چیست؟ (ii) مکعب داده چگونه بایستی آپ دیت شود؟ در این رابطه چه برنامه ای باید داشته باشیم؟ (iii) چه استراتژی باید برای ساخت و طراحی در پیش بگیریم؟
- روشهای محاسبه: به سبب حجم فوق العاده زیاد محاسبه ی مکعب های داده ی OLAP در داده های بزرگ چالش پژوهشی بزرگی خواهد بود. در این حالت امیدبخش ترین فناوری "الگوی محاسبات ابری" جدید است. شاید بتوان گفت که این فناوری از روشهای محاسباتی موازی قبلی الهام گرفته شده است.
- در حافظه: یک مکعب داده ی OLAP در داده ی بزرگ چگونه باید در حافظه ترسیم شود؟ این مسئله چالشی اساسی تلقی میشود که بسبب تعداد خیلی زیاد ابعاد چنین مکعب هایی است. در نتیجه راه حل های مبتنی بر حافظه ی ثالث بایستی بطور عمیق تحت بررسی قرار بگیرد.
- ساختار سخت افزاری جدید: روشهای سخت افزاری جدید نظیر پردازش مبتنی بر GPU نقشی حائز اهمیت در رابطه با محاسبه ی مکعب های داده ی OLAP در داده های بزرگ بازی خواهند کرد.
- بهینه سازی و زبانهای جستجو: با استفاده از روشهای MDX قدیمی نمیتوان راه حل های بهینه سازی مناسبی را در این حوزه ارائه کرد. در تحقیقات آتی باید به بهینه سازی پردازش داده های بزرگ به حالتی چندبعدی توجه لازم را بکنیم.

- عملکرد کاربر نهایی: مکعب های داده ی OLAP محاسبه شده در داده های بزرگ حجیم هستند از اینرو خیلی طبیعی است که کاربر نهایی خصوصا در پروسه های جستجو و .. عملکرد ضعیفی را داشته باشد. بنابراین در حین طراحی مکعب های داده ی OLAP در داده های بزرگ بایستی عملکرد کاربر نهایی را مدنظر داشته باشیم.
- کیفیت: در نسل بعدی انبار داده ها و روشهای OLAP در داده های بزرگ کیفیت بیشتر و بیشتر اهمیت مییابد. در واقع به سبب ماهیت کاملا غیرساختارمند منابع داده های بزرگ براحتی این امکان هست که محاسبات ضعیفی صورت پذیرد. از اینرو بسادگی میتوان به اهمیت کنترل کیفی مکعب های داده ی نهایی پی برد.
- قابلیت استفاده: مکعب های داده ی OLAP در داده های بزرگ باید بطور اصولی پردازش شده و مدیریت شوند تا بتوان تجزیه و تحلیل مفیدی صورت داد. از این جهت گستره ی وسیعی از مسائل مطالعاتی مطرح خواهند شد نظیر متدهایی جهت "ارزیابی" نحوه ی پردازش و مدیریت مکعب های داده در مخازن بزرگ داده
- تصویرسازی: با توجه به اینکه داده های بزرگ حجم فوق العاده زیادی دارند تصویرسازی مکعب های داده ی OLAP در داده های بزرگ در این گونه از مطالعات مطرح بوده و حائز اهمیت فراوان است. از اینرو روشها، متدها و رویکردهای تصویر سازی جدیدی بایستی ارائه شود تا بتوانیم چالش های پیش رویمان در رابطه با تصویرسازی مکعب های OLAP حجیم در داده های بزرگ، تصویرسازی زمان واقعی داده های استخراج شده، تصویرسازی داده های ترکیبی و تصویرسازی موثر در موبایل ها را به نحو احسن مدیریت کنیم.
- اکتشاف تعاملی: در کنار مسائل مرتبط با تصویرسازی مشکلات اکتشاف تعاملی در حوزه ی پژوهشی مکعب های داده ی OLAP در داده های بزرگ نیز حائز اهمیت فراوان بوده و باید مدیریت و حل شوند. واقعیت این است که پویای در مکعب های داده ی فوق العاده حجیم و استخراج دانش مفید بهمراه با مدیریت محتوا و.. کاری شاق میباشد.

- تجزیه و تحلیل: تجزیه و تحلیل داده های بزرگ زمینه ای است که جامعه ی محققین انبار داده ها و پایگاه داده روز به روز تمایل بیشتری به آن سمت پیدا میکند. در این زمینه مشکلات عدیده ای وجود دارد که بایستی تحت بررسی قرار گیرند. مشکلاتی از قبیل نحوه ی طراحی پروسه های تحلیلی در مکعب های داده ی OLAP محاسبه شده روی داده بزرگ و نحوه ی بهینه سازی اجرای این نوع پروسه های تحلیلی.
- ادغام با پایگاه های داده ی قبلی: یکی از مسائل مهم در این حوزه نحوه ی ادغام مدلها، تکنیک ها، الگوریتم ها و پایگاه های محاسباتی طراحی شده برای OLAP در داده های بزرگ با پایگاه های داده ی قبلی به گونه ای بدون ایراد و صحیح است.
- ابزارهای پیشرفت در این حوزه: ابزارهای مناسب برای پشتیبانی طراحی و پیشرفت در زمینه ی مکعب های داده ی OLAP با ادغام تمامی جوانب بحث شده در بالا چالشی حائز اهمیت در این رابطه است.

3. سمت و سوی مطالعات آتی انبار داده ها و OLAP روی داده های بزرگ

- با بررسی مشکلات و مسائل انبار داده ها و OLAP در داده های بزرگ سمت و سوی مطالعات آتی تعیین خواهد شد. در این میان ما به موارد زیر اشاره میکنیم:
- روشهای مبتکرانه جهت طراحی مکعب های داده ی OLAP در داده های بزرگ: نیازی شدید به ارائه ی روشهایی وجود دارد که جهت مدیریت الزامات طراحی و مدلینگ مکعب های OLAP در داده های بزرگ مفید واقع میشوند.
 - روشهای مبتکرانه جهت محاسبات: محاسبات که مسئله ای حائز اهمیت در مطالعات قبلی انبار داده ها و OLAP بود در رابطه با مکعب های داده ی OLAP در داده های بزرگ مشکلی مضاعف تر خواهد بود. از اینرو به روشهایی مبتکرانه نیاز است که با استفاده از آنها بتوان مسائل مرتبط با مکعب های داده ی بزرگ OLAP نظیر بلای ابعاد، مجموعه داده های "نامنظم"، تراکم های چندجانبه و .. را حل کرد.

- مدل های محاسباتی جدید برای محاسبه ی موثر و کارآمد مکعب های داده ی OLAP در داده های بزرگ: محاسبه ی مکعب های داده ی OLAP در داده های بزرگ زمان بر و هزینه بر است از اینرو ارائه ی مدل های جدید محاسباتی در این رابطه امری ضروری است.
- معماری های عملکردی قدرتمند برای اجرای مکعب های داده ی OLAP در داده های بزرگ: بهره گیری از طراحی های سخت افزاری مناسب برای پشتیبانی از اجرای مکعب های OLAP در داده های بزرگ سمت و سویی امیدبخش در طراحی نسل جدید اپلیکیشن های محاسباتی است.
- مکعب های داده ی OLAP پیچیده در داده های بزرگ: بسبب پیچیده بودن مجموعه داده های بزرگ نیاز به تعریف و استفاده ی از مکعب های داده ی ترکیبی OLAP است بگونه ای که جهت استفاده ی در اپلیکیشن های علمی بزرگ مقیاس مفید واقع شده و مناسب باشند.
- گزاره های MDX ای که میتوان شخصی سازی کرد: مکعب های داده ی OLAP در داده های بزرگ باید انعطاف پذیر باشند که به سبب الزامات کاربردهای عملی جدید است. از این جهت هست که زبان MDX کلاسیک برای پویا سازی داده های چندبعدی باید گسترش داده شود تا گزاره های "قابل شخصی سازی" را نیز در بر بگیرد.
- مکعب های داده ی OLAP ای که به لحاظ متنی غنی هستند: بهره گیری از تکنیک های معنامحور و متنی جهت مدلینگ مکعب های داده ی OLAP تجربه ای موفق در نسل جدید سیستم های پیچیده ی اطلاعات بوده است. باور ما بر این است که این تکنیک ها در رابطه با مکعب های داده ی بزرگ OLAP نیز میتوانند موفقیت هایی را حاصل دهند. چراکه با استفاده از متدهای معنامحور میتوان پروسه های دسترسی، مرور و ارائه را در این زمینه بهبود بخشید.
- زبانهای تعریف پروسه ای برای تجزیه و تحلیل: حتی با استفاده ی از پایگاه های داده ی NoSQL مدرن، تجزیه و تحلیل توابع پیچیده ای را در مقادیر خیلی بزرگ داده تعریف میکند. امکان این وجود دارد که این پروسه تبدیل به عملی پیچیده و مشکل زا گردد. از این جهت نیاز است که زبانهای تعریف پروسه ای قدرتمندی برای تجزیه و تحلیل در مکعب های داده ی بزرگ OLAP ارائه گردد.

- مسائل حریم شخصی و امنیت: جوانب امنیتی و حریم شخصی که در رابطه با مکعب های داده ی OLAP قبلی مطرح بود در رابطه با مکعب های داده در داده های بزرگ نیز مدنظر بوده و حائز اهمیت است. که به سبب این واقعیت است که این ساختارها "باز" هستند. در نتیجه در کارهای آتی باید بر امنیت و حریم شخصی OLAP مکعب های داده ی بزرگ در سیستم های اطلاعات باز و روی ابر اهمیت زیادی داده بشود.

- اپلیکیشن ها: به سبب ماهیت خاص مکعب های داده ی OLAP اپلیکیشن ها در این ساختارها نقش حائز اهمیتی را بازی میکنند. در کارهای مطالعاتی آتی باید به کارایی مکعب های داده ی بزرگ OLAP در گستره ی وسیعی از اپلیکیشن ها از نرم افزارهای پزشکی گرفته تا شبکه های اجتماعی، از سیستم های ابری گرفته تا چارچوب های مبتنی بر سنسور و.. توجه داشته باشیم.

4. نتیجه گیری

در این مقاله بحثی جدی در رابطه با مسائل موجود در رابطه با انبار داده ها و OLAP در داده های بزرگ و سمت و سوی مطالعات آتی داشتیم و بر محدودیت های موجود و راه حل های ممکن اشاره داشتیم. در کل افق های مطالعاتی متعددی را مطرح کردیم تا بلکه سرمشق مطالعات آتی باشند.



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی