

بررسی اجمالی فرآیند تحلیلی برخط در بانک اطلاعاتی Oracle و

استفاده آن از مدل داده‌ای ابعادی در پاسخ به سؤالات تجاری

آرمان دیدنده، دانشجوی مقطع کارشناسی رشته علوم کامپیوتر

دانشگاه صنعتی امیرکبیر تهران

Arman.Didandeh@gmail.com

مقدمه

این مقاله در دو بخش عمده نگاشته شده است. در بخش اول، ابتدا به معرفی مختصر OLAP می‌پردازیم. سپس به جنبه‌های تجاری آن و نوع به‌کارگیری آن در نرم-افزار بانک اطلاعاتی شرکت ORACLE خواهیم پرداخت. در انتهای بخش اول نیز مخازن داده‌ای چند بعدی را معرفی می‌کنیم. در بخش دوم، مدل داده‌ای ابعادی منطقی را به طور کامل بررسی می‌نماییم.

چکیده

فرآیند تحلیلی برخط، شاخه‌ای از هوش تجاری است که در کنار داده‌کاوی و به‌صورت مکمل به‌عنوان ابزاری برای پاسخ به سؤالات تجاری کاربر سامانه استفاده‌کننده از آن‌ها به‌کار می‌رود. این فرآیند نمودی ایده‌آل و قابل بررسی و بسیار کاربردی در بانک اطلاعاتی تهیه دیده شده توسط شرکت ORACLE دارد، به‌طوری‌که در دنیا از آن در کاربردهای متفاوتی نظیر سامانه‌های مالی و برنامه‌ریزی سازمانی استفاده‌های خوبی شده است. این فرآیند بر مبنای مدل داده‌ای ابعادی منطقی و با استفاده از فضاها کاری چند بعدی تعریف شده در Oracle DB، داده‌های دارای مدل تجمعی چند بعدی را مدیریت نموده و از آن‌ها برای تحلیل و پاسخ به پرسش‌های مورد نظر تحلیل-گران کاربر سامانه استفاده می‌نماید.

کلمات کلیدی

فرآیند تحلیلی برخط (OLAP)، وظیفه‌مندی (Functionality)، انبار داده (Data Warehouse)، فضای کاری (Workspace)، الگو (Schema)، مدل داده‌ای ابعادی منطقی (LDDM)، تحلیل‌گر (Analyst)، جلسه تقابلی (Interactive Session)، تصویر لحظه‌ای (Snapshot)، سلسله مراتب (Hierarchy)، مدل تجمعی (Aggregation Model)

بخش اول: فرآیند تحلیلی برخط (OLAP)

OLAP (On-Line Analytical Processing) وسیله‌ای برای ارزیابی سریع پاسخ به پرس و جوهای است که در طبیعت‌شان چند بعدی هستند. OLAP بخشی از شاخه عمومی‌تری به نام هوش تجاری است که همچنین شامل موارد زیر است:

- Extract Transform Load
- گزارشات رابطه‌ای (Relational Reporting)

• داده کاوی (Data Mining)

کاربردهای معمول OLAP در گزارشات تجاری برای فروش، گزارشات مدیریتی، مدیریت فرآیندهای تجاری (BPM)، پیش‌بینی‌های بودجه، گزارشات مالی و موارد مشابه است. واژه OLAP با تغییر جزئی در عبارت سنتی بانک‌های اطلاعاتی یعنی OLTP (On-Line Transaction Processing) به وجود آمد.

بانک‌های اطلاعاتی طراحی شده برای OLAP از یک مدل داده‌ای چند بعدی برای پرس و جوهای پیچیده تحلیلی در زمان بسیار کوتاه ممکن برای پاسخ‌دهی استفاده می‌کنند. «Nigel Pendse» به عنوان یک جایگزین و شاید عبارتی توصیفی‌تر برای OLAP، FASMI (Fast Analysis of Shared Multidimensional Information) را پیشنهاد می‌کند. بانک‌های OLAP از جنبه‌های بسیار سریع‌تر انواع مختلف بانک‌ها نظیر Navigational و Hierarchical نسبت به خویشاوند رابطه‌ای خود استفاده می‌کنند.

خروجی یک پرس و جو OLAP معمولاً در یک شکل ماتریسی نمایش داده می‌شود تا مقیاس‌ها و اندازه‌ها را در ابعاد ماتریس نمایان سازد. OLAP در انواع مختلف مطالعه و به کار گرفته می‌شود:

- Multidimensional (MOLAP)
- Relational (ROLAP)
- Hybrid (HOLAP)
- Web-based (WOLAP)
- Desktop (DOLAP)
- Real Time (RTOLAP)
- Spatial (SOLAP)

OLAP در ORACLE

فن‌آوری چند بعدی در بانک‌های اطلاعاتی شرکت ORACLE از نسخه Oracle9i قابل دسترسی است. هر ارایه‌ای پس از آن شامل تمامیت (Integrity)، وظیفه‌مندی (Functionality) و کارایی (Performance) بیشتر بوده است. سازمان‌ها دیگر نیازی به انتخاب بین بانک‌های چند بعدی OLAP و بانک‌های رابطه‌ای ندارند. با یک‌جمله کردن (Integration) جداول چند بعدی و یک موتور تحلیل‌گر درون بانک‌اطلاعاتی، ORACLE یک تحلیل چند بعدی را در کنار قابلیت‌های مدیریت، مقیاس‌بندی و اعتماد بانک‌های Oracle آماده ساخته است. یک‌جمله کردن فن-آوری چند بعدی در یک بانک اطلاعاتی رابطه‌ای بسیار مهم است زیرا برقرار نگه داشتن (Maintenance) یک بانک چند بعدی خودایستا (Standalone) بسیار هزینه‌بر است؛ این مورد به سخت‌افزار اضافی و مدیران بانک اطلاعاتی (Administrator) وارد در به‌کارگیری ابزارهای مدیریتی بانک‌های چند بعدی احتیاج دارد. در کنار این، بانک‌های خودایستای چند بعدی به برنامه‌های واسط کاربری اختصاصی نیازمندند.

این نیازها معمولاً سازمان‌های را وادار می‌کند تا از دو گونه ابزار پرس و جو و گزارش پشتیبانی نمایند؛ یکی برای بانک رابطه‌ای و دیگری برای بانک چند بعدی.

در تقابل، امکان استفاده از فن‌آوری چند بعدی به طور کامل در Oracle نهاده شده است. مدیران بانک اطلاعاتی از همانی ابزاری برای مدیریت این امکان استفاده می‌کنند که برای سایر اجزای بانک نیز از آن‌ها استفاده می‌نمایند.

استفاده از OLAP برای پاسخ به سؤالات تجاری

بانک‌های اطلاعاتی رابطه‌ای، امکان استفاده از OLTP را که برای پیگیری امور تجارت‌ها حیاتی است می‌دهد. بانک‌های اطلاعاتی رابطه‌ای به دلیل طراحی کارا در ذخیره و بازیابی داده‌ها برای نگهداری چندین گیگابایت از اطلاعات تفصیلی

مناسب هستند. موفقیت بانک‌های رابطه‌ای در استفاده از آن‌ها برای ذخیره حیطة بسیار وسیعی از اطلاعات مشهود است. در نتیجه آن‌ها شامل داده‌های بسیار گسترده‌ای می‌باشند که اطلاعات ذی‌قیمتی را درباره یک تجارت ارائه می‌کند. این اطلاعات امتیاز قابل توجهی را در یک بازار رقابتی افزونی (Increasing) برای صاحبان خود به همراه دارد. چالش عمده در استنتاج پاسخ‌های مناسب برای سؤالات تجاری از داده‌های موجود است به نحوی که تصمیم‌گیرندگان در تمامی سطوح، قادر به عکس‌العمل سریع و بهینه در اوضاع تجاری مختلف باشند. یک پرس و جوی تراکنشی استاندارد ممکن است سؤال زیر را مطرح نماید:

«سفارش شماره ۸۴۳۰۵ در چه زمانی فرستاده شده است؟»

این پرس و جو بازتاب پایه سازوکار تجاری را مطرح می‌نماید، زیرا شامل انتخاب و بازیابی داده‌ای یک رکورد- یا حداکثر چندین رکورد مرتبط به هم- شناسایی شده با استفاده از یک شماره سفارش مشخص است. هرگونه سؤال پیگیری در این باره، مانند اینکه «کدام آژانس پستی آن را حمل کرده است؟» و یا «سفارش به کجا حمل شده است؟» را می‌توان با همین رکورد مشخص پاسخ داد. این رکورد دارای یک بازه زندگی (Life Span) مفید در دنیای تراکنشی است: شروع زندگی آن با انجام یک سفارش خرید توسط یک مشتری است و عمر آن پس از پرداخت و ارسال سفارش به اتمام می‌رسد. در این زمان رکورد را می‌توان به آرشیو منتقل نمود.

در مقابل، می‌توان یک سری از پرس و جوهای تحلیلی را نیز مطرح نمود:

«نتیجه مقایسه فروش سه‌ماهه جاری در محدوده خلیج پارس با همین محدوده زمانی در سال گذشته چگونه است؟»

«چه پیش‌بینی‌ای برای فروش در سه‌ماهه آینده داریم؟»

«تغییر کدام عوامل باعث پیشبرد فروش پیش‌بینی شده خواهند شد؟»

این‌ها سؤالاتی درباره تراکنش‌های تجاری نیستند، بلکه درباره تحلیل کارایی گذشته و تصمیم‌گیری‌هایی برای بهبود کارایی آینده، دار بودن امتیاز بیشتر در رقابت و تسهیل در بالا بردن سودآوری هستند. بانک اطلاعاتی تحلیلی اطلاعات لازم برای تصمیم‌گیرندگانی را فراهم می‌آورد که قابلیت هدف‌گذاری امروزشان وابسته به پیش‌بینی‌های آن‌ها درباره آینده است. پاسخ به این پرسش‌ها مستلزم محاسبات نقطه سر به سر، آنالیز سری‌های زمانی و دسترسی به داده‌های تجمعی گذشته و حال می‌باشد. تمامی موارد فوق‌الذکر نیازمند استفاده از OLAP هستند.

کاربردهای معمول OLAP

در این جا به ذکر چند مثال کاربردی که از OLAP برای موارد ارزش‌مند در وظیفه‌مندی و بهره‌وری استفاده می‌نمایند، خواهیم پرداخت:

- نرم‌افزارهای برنامه‌ریزی به سازمان‌ها این قابلیت را می‌دهند که درآمدها را پیش‌بینی نمایند. آن‌ها با استفاده از ابزارهای تحلیلی نظیر مدل‌ها، اجتماع داده‌ها، تخصیص منابع و مدیریت سناریوهای مختلف به تولید داده‌های جدید می‌پردازند. نمونه‌هایی از این‌گونه نرم‌افزارها عبارتند از: تحلیل‌گران بودجه و مسایل مالی و سامانه‌های برنامه‌ریزی تقاضا.

- برنامه‌های تحلیل‌گر بودجه و مسایل مالی این قابلیت را به سازمان می‌دهند که کارایی گذشته را تحلیل نموده، طرح مخارج و درآمدها را ساختار بخشند، به سوی اهداف سودآور مدیریت نمایند و اثر تغییرات را بر طرح مالی مدل نمایند. مدیریت این امکان را می‌یابد که سطوحی را برای سرمایه‌گذاری و مخارج را تعیین کند که درخور سطوح درآمدها و سود باشند. تحلیل‌گران مالی می‌توانند

بودجه‌ای جایگزین را مشروط بر عواملی چون نوسانات ارزی برای طرح‌های سرمایه‌گذاری در نظر داشته باشند.

○ سامانه‌های برنامه‌ریزی تقاضا این قابلیت را به سازمان می‌دهند که تقاضای بازار را بر پایه عواملی چون سابقه فروش، طرح‌های تبلیغی و الگوهای قیمت‌گذاری پیش‌بینی نمایند. آن‌ها می‌توانند سناریوهای مختلفی را برای پیش‌بینی تقاضای محصولات مدل کنند و سپس اهداف مناسب و درخور را تعیین نمایند.

آن‌گونه که این بحث مشخص می‌کند، پردازش اطلاعات لازم برای پاسخ‌دهی به سؤالات تحلیلی اساساً با ساختار پردازشی پاسخ‌دهی به سؤالات تراکنشی متفاوت است. کاربران، اهداف‌شان، پرس و جوهای‌شان و انواع داده‌های مورد نیازشان نیز متفاوت است. یک انبار داده (Data Warehouse) رابطه‌ای دارای امکان افزوده OLAP بهترین محیط را برای تحلیل داده‌ای فراهم می‌آورد.

درباره مخازن داده‌ای چند بعدی

داده‌های چند بعدی در فضاهای کاری (Workspace) تحلیلی ذخیره می‌شوند؛ جایی که می‌تواند توسط موتور OLAP بانک اطلاعاتی Oracle مدیریت و به کار گرفته شود. فضاهای کاری تحلیلی منحصر به فرد در جداول دارای الگو (Schema)ی رابطه‌ای ذخیره می‌گردند، در نتیجه می‌توانند مانند سایر جداول رابطه‌ای مدیریت شوند. یک فضای کاری تحلیلی متعلق به یک شناسه کاربری خاص است و امکان دادن اجازه دسترسی به سایر کاربران نیز موجود است. در یک بانک اطلاعاتی مشخص، امکان ایجاد فضاهای کاری متعدد و تسهیم (Share) آن‌ها بین کاربران مختلف وجود دارد.

فضاهای کاری صریحاً برای اداره چند بعدیت (Multidimensionality) در مخازن فیزیکی داده‌ای طراحی شده‌اند. فن‌آوری چند بعدی که در زمینه فضاهای کاری تحلیلی قرار دارد، بر پایه یک مدل آرایه‌ای چند بعدی شاخصی (Indexed) طراحی شده است. این ساختار چند بعدی طبیعی موجب می‌شود که بیشتر سرعت و قدرت فضاهای کاری تحلیلی در اجرای تحلیل چند بعدی مصرف شود.

بخش دوم: مدل داده‌ای ابعادی منطقی

در این بخش از مقاله به بررسی جوانب مختلف مدل داده‌ای ابعادی منطقی می‌پردازیم. برای این کار ابتدا نگاهی گذرا بر مدل داده‌ای می‌اندازیم. سپس به مطالعه مکعب‌های منطقی، مقیاس‌های منطقی، ابعاد منطقی، سلسله مراتب و سطوح منطقی و صفات منطقی خواهیم پرداخت.

نگاهی به مدل داده‌ای

مدل داده‌ای ابعادی بخش مورد نیازی از فرآیند تحلیلی برخط (OLAP) می‌باشد. به دلیل برخط بودن OLAP، این فرآیند باید سریع باشد؛ تحلیل‌گران (Analyst) پرس و جوهای مکرری را در ضمن جلسات تقابلی (Interactive Session) - و نه برای یک دسته وظایف شبانه (Overnight batch jobs) - مطرح میکنند.

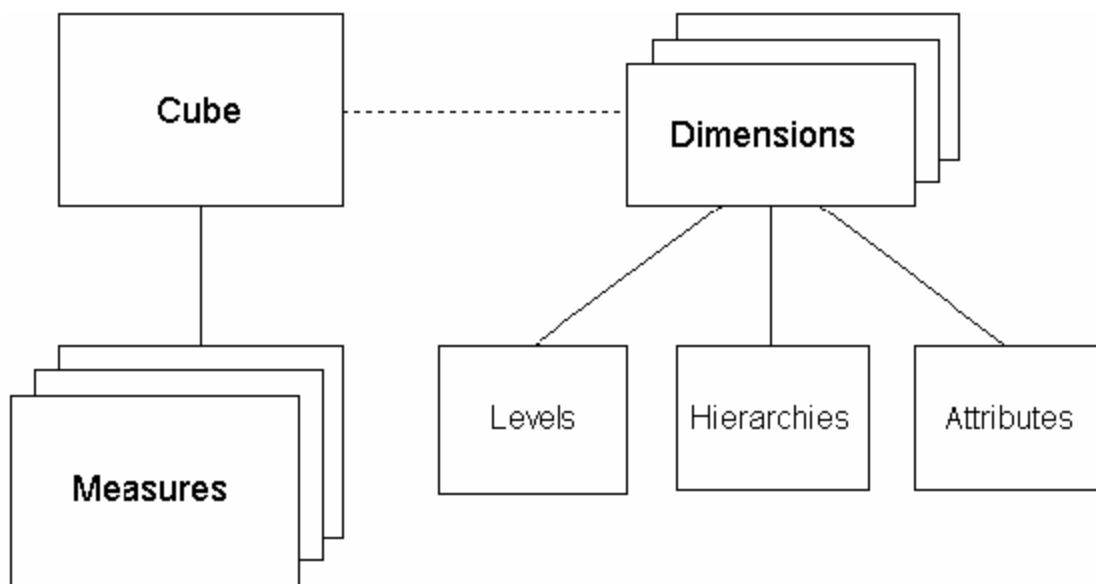
همچنین چون OLAP تحلیلی است، این پرس و جوها پیچیده است.

مدل ابعادی داده‌ای ترکیبی است از:

- مکعب‌های منطقی

- اندازه‌ها و مقیاس‌های منطقی
- ابعاد منطقی
- سلسله مراتب منطقی
- مراحل و سطوح منطقی
- و صفات منطقی

سادگی مدل ذاتی و ماندگار است زیرا اشیاء تعریف شده نمایان کننده موجودیت‌های دنیای واقعی تجارت هستند. تحلیل‌گران می‌دانند که علاقه‌شان به آزمودن کدام مقیاس‌های تجاری است، کدام ابعاد و صفات داده‌ها را معنی‌دار می‌کند و چگونه ابعاد تجارت‌شان در سطوح و سلسله مراتب متشکل و سازمان‌دهی می‌شود. شکل زیر رابطه عمومی بین اشیاء منطقی را نشان می‌دهد:



مکعب‌های منطقی: مکعب‌های منطقی توانایی سازمان‌دهی مقیاس‌های دارای ساختار و شکل یکسان را می‌دهند؛ بدین معنی که آن‌ها دارای ابعاد دقیقاً یکسانی هستند. مقیاس‌های درون یک مکعب دارای روابط یکسانی با سایر اشیاء منطقی هستند و می‌توان به سادگی آن‌ها را تحلیل نمود و به نمایش گذارد.

اندازه‌ها و مقیاس‌های منطقی: اندازه‌ها درون خانه‌های مکعب منطقی قرار می‌گیرند و دارای حقایق گردآوری شده درباره عملیات تجاری هستند. اندازه‌ها به وسیله ابعاد سازمان‌دهی می‌شوند که به صورت معمول بعد زمان را در بر می‌گیرند. یک بانک اطلاعاتی تحلیلی شامل تصاویر لحظه‌ای (Snapshot) از داده‌های پیشین (Historical) اشتقاقی (Derived) از بانک‌های تراکنشی، سامانه‌های موروثی (Legacy systems)، منابع اتحادیه‌ای (Syndicated sources) و یا سایر منابع داده‌ای است. یک منبع داده‌ای سه ساله برای برنامه‌های کاربردی تحلیلی مناسب خواهد بود. اندازه‌ها و مقیاس‌ها ثابت و پایدار هستند در حالی که تحلیل‌گران از آن‌ها در تصمیم‌گیری‌هایشان استفاده می‌کنند. آن‌ها به صورت دسته‌ای در بازه‌های زمانی نظیر هفتگی، روزانه یا گاه به گاه در طول روز، به‌هنگام‌سازی می‌شوند. برخی

مدیران بانک اطلاعاتی داده‌های خود را با افزودن بازه‌هایی به بعد زمان یک مقیاس تجدید می‌کنند و ممکن است که حتی همان تعداد از بازه‌های قدیمی‌تر را از منابع حذف نمایند. برخی دیگر هم بازسازی کامل داده را به جای روش پیمایش افزایشی ترجیح می‌دهند.

ابعاد منطقی: ابعاد مجموعه‌ای از مقادیر واحد را در بر دارند که باعث شناسایی و طبقه‌بندی داده‌ها می‌شود. آن‌ها لبه‌ای مکعب منطقی را تشکیل می‌دهند و در نتیجه اندازه‌های درون مکعب به ابعاد وابسته است. از آنجایی که اندازه‌ها و مقیاس‌ها نوعاً چند بعدی هستند، مقادیر باید توسط یک‌یک اعضای بعد مورد نظر برای معناداری کنترل شوند. برای نمونه فروش دارای چهار بعد است: زمان، مشتری، محصول و مجرا (Channel). یک مقدار ویژه فروش مثل (۴۳,۶۱۳,۵۰) تنها زمانی معنی خواهد داشت که توسط یک مشتری خاص (آقای مهدی احمدی) در یک زمان خاص (سوم خرداد) برای یک محصول خاص (رایانه خانگی) و یک مجرا (فروش اینترنتی) مشخص شود.

سلسله مراتب و سطوح منطقی: سلسله مراتب روش‌هایی برای سازمان‌دهی داده‌ها در سطوح مختلف تجمعی (Aggregation) هستند. در مشاهده داده‌ها، تحلیل‌گران از سلسله مراتب ابعادی برای شناسایی روندها در یک سطح استفاده می‌کنند، به سطوح پایین‌تر مراجعه می‌نمایند تا دلایل این روندها را بفهمند و به سطوح بالاتر می‌روند تا تأثیرات این روندها را بر قسمت‌های وسیع‌تر تجارت بررسی نمایند. دو گونه سلسله مراتب را بررسی می‌کنیم:

- سلسله مراتب بر مبنای سطوح (Level-Based): هر سطح یا درجه یک استقرار را در سلسله مراتب مشخص می‌کند. هر سطح بالای سطح پایه (Base Level) شامل مقادیر داده‌های تجمعی سطوح زیرین خود است. اعضای سطوح مختلف دارای یک رابطه پدر-فرزندی یک به چند هستند. برای نمونه «بهار ۸۶» و «زمستان ۸۶» فرزندان «۱۳۸۶» هستند.

در نظر بگیرید که یک انباره داده شامل تصاویر لحظه‌ای از داده با برنامه زمانی سه بار در روز، هر هشت ساعت یک‌بار است. تحلیل‌گر به صورت طبیعی علاقه‌مند است که داده‌های را به صورت تجمعی در روزها، هفته‌ها، فصول و یا شاید سال‌ها ببیند. در نتیجه بعد زمان نیازمند سلسله مراتب با حداقل پنج سطح است.

به طور مشابه یک مدیر فروش با یک هدف مشخص در سال آینده ممکن است بخواهد آن میزان هدف خود را در بین نمایندگان فروش محدود خود تخصیص دهد. این تخصیص نیازمند یک سلسله مراتب سطحی است که در آن نمایندگان فروش مقادیر فرزند محدود فوق‌الذکر باشند.

سطوح سلسله مراتب دارای روابط چند به چند می‌باشند. یک سلسله مراتب به طور نمونه شامل چندین سطح است و یک سطح ممکن است در چندین سلسله مراتب به کار گرفته شود.

- سلسله مراتب بر مبنای مقادیر (Value-Based): اگرچه سلسله مراتب معمولاً متشکل از سطوح هستند، هیچ لزومی برای این موضوع وجود ندارد. روابط پدر-فرزندی بین اعضای ابعاد مختلف ممکن است سطوح مختلف را معناداری ننماید. برای مثال در یک بعد کارمند، هر مدیر یک یا چند گزارش خواهد داشت که باعث ایجاد یک رابطه پدر-فرزندی می‌شود. دید سطح-نگری بر این روابط (مانند شرکت‌کنندگان مجزا، مدیران سطح اول، مدیران سطح دوم مانند این‌ها) برای تحلیل معنایی در بر نخواهد داشت. همچنین بعد خط تولید در داده‌های مالی دارای سطوح نخواهد بود. این گونه سلسله مراتب را بر مبنای مقدار نام می‌دهیم.

صفات منطقی: یک صفت اطلاعات افزوده درباره داده را ارایه می‌کند. برخی صفات برای نمایش و ارایه کاربرد دارند؛ برای نمونه صفاتی چون رنگ، طعم و یا اندازه. این انواع از صفات می‌تواند برای انتخاب داده‌ها و پاسخ به پرسش‌هایی نظیر موارد زیر به کار گرفته شود:

«کدام رنگ‌ها در پیراهن‌های جوانان در تابستان ۱۳۸۶ رایج‌تر و مورد پسندتر بوده است؟»

«این مورد در قیاس با سال گذشته چگونه بوده است؟»

صفات زمانی نیز می‌توانند اطلاعاتی درباره بعد زمان ارایه دهند که در انواعی از تحلیل قابل به‌کارگیری است؛ مانند مشخص کردن آخرین روز یا تعداد روزهای هر بازه زمانی خاص.