اولین کنفرانس ملی تصادفات وسوانح جاده ای و ریلی
ایران- زنجان، دانشگاه آزاد اسلامی –واحد زنجان، آذر ۱۳۸۸

شناسایی نقاط حادثه خیز جاده‌ها با استفاده از انبارداده مکانی

امیرحسین شکوهی [۱]، علی شکوهی [۲]

[۱]کارشناس ارشد مهندسی نرم‌افزار ، مدرس گروه کامپیوتر دانشگاه آزاد اسلامی واحد زنجان؛ shokuhi@comp.iust.ac.ir
[۲]دانشجوی دکترای شهرسازی دانشگاه آزاد اسلامی علوم تحقیقات، عضو هیات علمی دانشگاه آزاد اسلامی واحد زنجانa_shokoohi2002@yahoo.com

چکیده

امروزه انبارهای داده نقش بزرگی را در پشتیبانی از سیستم‌های تصمیم‌گیری فراهم کرده و بستر مناسبی برای پاسخگویی به پرسش‌های تحلیلی از داده‌ها را برای کاربران ایجاد می‌کنند. به همین لحاظ توسعه و گسترش انبارهای داده در ابعاد مختلف در دستور کار مراکز علمی پژوهشی دنیا قرار گرفته است. بخشی از این توسعه شامل سازگارسازی ساختار انبارداده با نوع ساختار و ویژگی‌های داده‌های ورودی است. با افزایش روزافزون منابع عملیاتی مختلف مانند سیستم‌های موقعیت یاب ماهواره‌ای مرکز عملیات پلیس راهنمایی کشور که داده‌های مکانی را با سرعت و حجم زیاد تولید می‌کنند نیاز است که بتوان در زمان کوتاهی حجم زیادی از داده‌ها را تحلیل کرد و با استفاده از نتایج آنها تصمیم‌های مناسبی را در زمان سوانح به منظور ارائه به نیروهای امدادی اتخاذ کرد.

در این مقاله به موضوع طراحی و پیاده‌سازی انبارهای داده مکانی کم‌تأخیر برای سیستم‌های حمل و نقل هوشمند پرداخته شده و مدلی را برای کاهش تأخیر در سیستم‌های حمل و نقل هوشمند برای شناسایی نقاط حادثه خیز با پیاده‌سازی موردی بر روی داده‌های مکانی پایگاه داده پلیس ارائه نموده است.

این مقاله گزارشی است از پروژه طراحی و پیاده سازی انبارداده مکانی بلادرنگ جهت تحلیل تخلفات جاده‌ای است که برای استفاده پلیس راه‌های کشور تهیه شده است. در این پروژه یک انبارداده مکانی برای پلیس طراحی شده که می‌تواند به تحلیل تخلفات خودروهای در حال حرکت در محورهای حمل و نقل کشور پرداخته و از تحلیل داده‌های تخلفات، نقاط حادثه‌خیز را شناسایی نماید.

کلمات کلیـــدی: انبارهـای داده مکـانی، کـاهش زمـان تـأخیر، داده مکانی، سیستم های حمل و نقل هوشمند

مقدمه

دنیای امروز نیازمند اطلاعات مناسب، به شکل مطلوب و در زمان مناسب است و ذخیره‌سازی و بازیابی چنین اطلاعاتی اهمیت دوچندان دارد؛ مدیران و تحلیل‌گران نیازدارند تا اطلاعات جامع‌تری در اختیار داشته باشند و بتوانند شرایط حاصل از تغییرات را به سرعت تجزیه، تحلیل و برنامه‌ریزی کرده و به شیوه‌ای صحیح به آن واکنش نشان دهند.

داده‌ها از منابع متعدد و براساس فناوری‌های مختلف تولید می‌شوند. شکل این داده‌ها و فناوری مولد آنها شیوه‌های خاصی را جهت بهره‌برداری از آنها طلب می‌کند.

در این میان انبارهای داده نقش بزرگی را در پشتیبانی از سیستم‌های پشتیبان تصمیم‌گیری فراهم کرده و بستر مناسبی برای انجام پرسش‌های تحلیلی از داده‌ها ایجاد می‌کنند. انبارهای داده پردازش‌های اطلاعاتی را با ایجاد یک بستر قوی و یکپارچه از داده‌های قابل تحلیل مهیا می‌سازد و امکاناتی را برای یکپارچه‌سازی سیستم‌های غیر یکپارچه فراهم می‌آورد.

با وجود جدید بودن فن‌آوری انبارهای داده در جهان و در ایران، پیشرفت‌های بسیار سریع در این زمینه صورت گرفته و مدام در حال توسعه است. اختراع و بروز دستگاه‌های جدید که می‌توانند داده‌ها را از نقاط بسیار پراکنده و دور جمع‌آوری و در اختیار قراردهند (نظیر گیرنده‌های تعیین موقعیت ماهواره‌ای) انبارهای داده را با موارد مختلف و جدیدی روبرو ساخته است که یک مورد از آنها در این مجموعه مورد بررسی قرار می‌گیرد.

این مقاله گزارشی است از پروژه طراحی و پیاده سازی انبارداده مکانی بلادرنگ جهت تحلیل تخلفات جاده‌ای است که برای استفاده پلیس راه‌های کشور تهیه شده است. در این پروژه یک انبارداده مکانی برای پلیس طراحی شده که می‌تواند به تحلیل تخلفات خودروهای در حال حرکت در محورهای حمل و نقل کشور پرداخته و از تحلیل داده‌های تخلفات، نقاط حادثه‌خیز را شناسایی نماید.

۱- معرفی انبارداده مکانی و تفاوت آن با پایگاه داده مکانی

کارهایی که در زمینه طراحی و ساخت انبارهای داده تا به امروز انجام شده‌است، چالش‌های منحصر به فرد ساخت یک انبارداده در برخورد با داده‌های مکانی را در نظر نگرفته‌اند. از طرفی پایگاه‌های داده مکانی به عنوان یکی از سیستم‌های مدیریتی که ماهیت خاص داده‌های مکانی را لحاظ کرده‌اند، مبتنی بر تراکنش بوده و قادر به ارایه پرسش‌های تحلیلی و برخط نمی‌باشند، بنابراین اتکا به پایگاه داده مکانی عملیاتی برای پشتیبانی از سیستم‌های

Archive of SID
www.SID.ir

تصمیم‌گیر به تنهایی کافی نبوده و برای این منظور استفاده از انبارداده ویژه داده‌های مکانی پیشنهاد می‌شود.

می‌توان انبارداده مکانی را ترکیبی از سیستم‌های انبارداده و پایگاه‌داده مکانی رایج برای مدیریت مقادیر عظیم داده‌های تاریخچه‌ای که شامل موقعیت‌های مکانی هستند، دانست. ترکیب این دو تکنولوژی اجازه به‌کارگیری قابلیت‌های هر دو سیستم برای بهبود تجزیه و تحلیل‌ها، مصورسازی‌ها و دستکاری داده‌ها به منظور حمایت از تصمیم‌گیری را به کاربران دانش(مدیران) می‌دهد. سیستم‌های انبارداده متُدها و روش‌های مدیریتی حجم عظیم داده‌ها به منظور پاسخگویی سریع به پرس‌وجوها را ارایه می‌کنند و از طرف دیگر پایگاه‌های داده مکانی با یک تجربه قوی و طولانی در مدیریت داده‌های مکانی و ارایه تجزیه تحلیل‌های مکانی می‌توانند به پاسخگویی و مدیریت پرس‌وجوهای مکانی کمک قابل توجهی نمایند. تجارب قبلی بدست آمده برای مدیریت داده‌های سرجمع شده در سیستم‌های پردازش تحلیلی برخط نیز می‌توانند توسط سیستم‌های پردازش تحلیلی بر خط مکانی[1] توسعه داده شوند.

۲- تفاوتهای انبارداده مکانی با پایگاه داده مکانی

انبارداده مکانی ماهیتاً با پایگاه داده مکانی متفاوت است و در واقع جز امکان نمایش کارتوگرافیک داده‌های مکانی و ساختارهای ذخیره و بازیابی و شاخص‌دهی مکانی شباهت دیگری به هم ندارند اما به جهت مشخص کردن قابلیت‌ها و تفاوت‌های این دو سیستم برای کاربران سنتی پایگاه داده مکانی می‌توانیم موارد زیر را به عنوان وجوه تمایز این دو نوع سیستم ذکر کنیم:

انبارداده مکانی ترکیبی از سیستم‌های پایگاه داده مکانی و انبارداده محسوب می‌شود.

سیستم‌های پایگاه داده مکانی مبتنی بر تراکنش بوده ولی سیستم‌های انبارداده مکانی، تحلیلی می‌باشند.

سیستم‌های انبارداده مکانی تاریخچه‌ای از داده‌ها را نگهداری می‌کنند ولی پایگاه‌های داده مکانی داده‌های جاری و به‌روز را نگهداری می‌کنند.

سیستم‌های انبارداده مکانی خلاصه‌ای از داده‌ها را نگهداری می‌کنند ولی پایگاه‌های داده مکانی داده‌ها را به صورت کامل نگهداری می‌کنند.

در پایگاه داده مکانی دو بعد داده‌های مکانی و داده‌های مفهومی (معنایی) مطرح است درحالیکه انبارداده مکانی علاوه بر این دو بعد شامل بعد زمان نیز می‌باشد.

پاسخگویی به پرسش‌ها در انبارداده مکانی همیشه به صورت برخط و سریع می‌باشد اما در پایگاه داده مکانی در پرس‌وجوهای پیچیده این‌گونه نیست.

اطلاعات موجود در انبارداده مکانی در چندین سطح از جزئیات نگهداری می‌شوند به‌طوریکه کاربر قابلیت هدایت میان این سطوح را دارد اما پایگاه داده مکانی اغلب یک سطح از جزئیات را نگهداری می‌کند.

درخواست پرس‌وجوهای پیچیده در انبارداده مکانی با سهولت بیشتری نسبت به پایگاه داده مکانی انجام می‌گیرد.

۳- معرفی نمونه مدل طراحی شده

۳-۱- معرفی سیستم‌های هوشمند حمل و نقل[2] و نیاز به انبارداده مکانی

سیستم‌های هوشمند حمل و نقل دربردارنده کلیه فن‌آوری‌های اطلاعات، مخابرات، کنترل، مهندسی سیستم و نیز استراتژی‌ها، تصمیم‌گیری‌های مدیریت و سازوکارهای هماهنگ کننده‌ای است که در نتیجه به کارگیری آنها، بهبود پارامترهای حمل‌ونقل و ترافیک نظیر کاهش زمان سفر، مصرف سوخت و افزایش ایمنی حاصل شود. تعاریف متفاوتی از ITS ارائه شده است که مفهوم مشترک آنها عمدتاً بر کاربرد هدفمند و هماهنگ فناوری اطلاعات و ارتباطات و استراتژی‌های مدیریتی به شرط ارتقا بهره‌وری و کارآیی و ایمنی سیستم حمل و نقل تأکید دارند.

تاریخچه به‌کارگیری و استفاده از سیستم‌های حمل و نقل هوشمند به اوایل دهه ۹۰ میلادی می‌رسد که در کشورهای توسعه یافته شروع به مطالعه، طراحی و گسترش این سیستم‌ها نموده‌اند. توسعه سیستم‌های مختلف حمل و نقل و افزایش تقاضای سفر منجر به شکل‌گیری شاخه‌های مختلف ITS شده است.

یک شاخه مهم در سیستم‌های هوشمند حمل و نقل به کارگیری سیستم‌های تعیین موقعیت و جایاب نظیر GPS در تجهیزات حمل و نقل می‌باشد که امکان کنترل و ردیابی دقیق، اتوماتیک و آسان و سریع خودروها را در محورهای مواصلاتی فراهم می‌کند.

از آنجا که این نوع سیستم‌های هوشمند نیاز به پردازش اطلاعاتی خودروهای در حال حرکت دارند، با حجم وسیع و گسترده و متغیر داده‌های مکانی مواجه خواهند بود که نیاز به تحلیل‌های پیچیده و متنوع دارند، از اینرو استفاده از انبارداده‌های فضایی در این سیستم‌ها که بتواند در حداقل زمان ممکن تحلیل‌های مورد نیاز را دراختیار تصمیم‌گیران و برنامه‌ریزان ترافیک برون‌شهری قرار دهند ضروری‌است.

۳-۲- قابلیت‌ها و نیازمندی‌های GPS در کنترل حمل و نقل کشور

استفاده از GPS در ناوگان حمل و نقل عمومی در بیشتر کشورهای دنیا معمول است و نصب این سیستم در خودرو به پلیس این امکان را می‌دهد که رفتار رانندگان و تخلفات آنها را به صورت الکترونیکی کنترل کنند. از مزایای استفاده از GPS در ناوگان حمل و نقل امکان ثبت موقعیت خودرو در زمان‌های مشخص است. این

[1] SOLAP: Spatial OLAP

[2] ITS: Information Transportation System

Archive of SID

www.SID.ir

ویژگی امکان محاسبه سرعت لحظه‌ای و متوسط خودرو، انحراف از مسیر، کنترل سرعت مجاز و ... را میسر می‌سازد.

اما استفاده از چنین سیستمی به صورت گسترده و عمومی (ناوگان حمل و نقل اتوبوس‌های بین شهری) نیازمندی‌های خاص

خود را دارد. این سیستم که در حال حاضر به صورت آزمایشی در چند محور جاده‌ای کشور به اجرا درآمده با وجود قابلیت‌ها نواقصی نیز دارد. به عنوان مثال در حال حاضر با نصب دستگاه GPS بر روی اتوبوس‌های مسیر تهران - شیراز پلیس بایستی در پاسگاه‌ها پورت GPS را از راننده‌ها اتخاذ کرده و با تخلیه اطلاعات آن در کامپیوتر پاسگاه با انجام پردازش وضعیت **شکل ۱- GPS نصب شده بر روی اتوبوس‌های کشور**

تخلفات راننده را چک نماید. این موضوع باعث اتلاف وقت رانندگان و مسافران گردیده، علاوه براین احتمال ایجاد ایراد در تجهیزات سخت‌افزاری، نرم‌افزاری و امکان دستکاری غیرمجاز اطلاعات موقعیتی توسط سوء استفاده‌کنندگان وجود دارد. ضمن اینکه در چنین سیستمی که پایگاه داده پلیس در بازه‌های طولانی (پاسگاه تا پاسگاه) به‌هنگام‌سازی می‌شود، نمی‌توان نظارت درستی نسبت به وضعیت خودروها و عملکرد رانندگان داشت. به عنوان مثال چنانچه یک اتوبوس در میانه راه دچار خرابی شود یا اینکه جهت دورزدن عوارضی یا گشت‌های پلیس به مسیرهای فرعی تغییر مسیر دهد امکان آگاهی از آن به صورت آنلاین وجود ندارد. لذا نمی‌توان در این سیستم از تخلفات پیشگیری نمود یا اینکه به خودروهای آسیب دیده در طی مسیر امدادرسانی کرد.

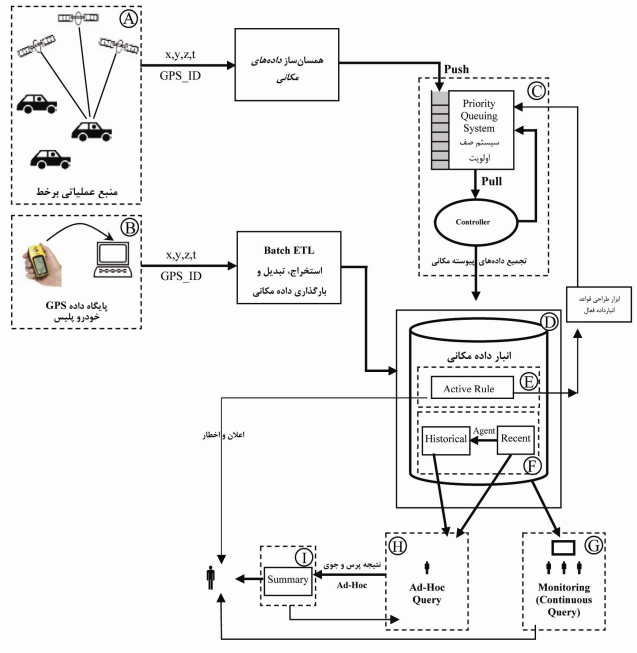
در چنین وضعیتی استفاده از یک انبارداده که بتواند به صورت به هنگام وضعیت‌های مختلف خودروها را تحلیل نماید، پیشنهاد می‌گردد. برای انجام اینکار لازم است که موقعیت خودروها به صورت اتوماتیک به مرکز پلیس ارسال گردیده تا مورد تحلیل قرار گیرد. که در حال حاضر فناوری این نوع GPSها در کشور وجود دارد[3].

در چنین سیستم‌هایی مختصات خودروها به صورت جریانی از داده‌های مکانی پیوسته به پایگاه داده مرکز پلیس گسیل می‌شوند و این موضوع می‌تواند اختلالاتی در عملکرد انبارداده‌های سنتی ایجاد نماید که لزوم طراحی و پیاده‌سازی یک مدل مناسب با کمترین تأخیرات را جهت مواجهه با داده‌های حجیم و جاری گیرنده‌های GPS می‌طلبد که در ادامه به تشریح اقدامات صورت گرفته در این راستا پرداخته می‌شود.

[3] **نمونه هایی از این نوع GPSها با عنوان تجاری AVL یا "رهاوا" بر روی برخی از خودروهای عمومی کشور نصب شده است.**

۳-۳- طراحی و پیاده‌سازی نمونه موردی

براساس مطالعات انجام شده مدل پیشنهادی به صورت معماری شکل ۲ با ارائه نوآوری‌ها و تغییرات پایه‌ای متناسب با ویژگی‌های داده‌های مکانی پلیس‌راه، طراحی گردید که عملکرد بخش‌های مختلف آن در زیر معرفی می‌گردد.



شکل ۲- معماری پیشنهادی

۳-۳-۱- منبع داده عملیاتی برخط

در این بخش سیگنال‌های ماهواره ای GPS توسط گیرنده‌های مستقر بر روی خودروهای حمل و نقل عمومی بین شهری دریافت و مورد پردازش قرارگرفته و مختصات لحظه‌ای خودرو همراه با زمان دقیق اتمی[4] محاسبه و سپس از طریق بسترهای مخابراتی به مرکز پلیس ارسال می‌شود.

داده‌هایی که از گیرنده‌های GPS مستقر بر روی خودروها به مرکز پلیس می‌رسند به لحاظ حجم و سرعت تولید بالا دائماً در حال تغییر سریع هستند.

در این پروژه داده‌های GPS داده‌هایی نامحدود، زودگذر و ناپایداری هستند که به صورت پیوسته، سریع و با حجم بالا توسط گیرنده‌های GPS خودروها تولید می‌شوند. همین موضوع نیاز به انبارداده‌ای را می‌طلبد که بتواند از داده‌های مکانیGPS با این ویژگی‌ها پشتیبانی نماید. داده‌هایی که توسط GPS خودروها تولید می‌شود می‌تواند شامل مختصات دو مؤلفه‌ای (x,y) یا سه مؤلفه‌ای (x,y,z) موقعیت خودرو، زمان لحظه‌ای تعیین موقعیت، کد گیرنده (که می‌تواند همان کد خودرو باشد) و... باشد.

[4] **گیرنده‌های GPS و ماهواره‌ها برای محاسبه فاصله به صورت یک‌طرفه عمل می‌کنند، به همین لحاظ در آنها ساعت‌های اتمی دقیق برای تعیین زمان طی مسافت موج تعبیه شده است.**

Archive of SID
www.SID.ir

۳-۳-۲- پایگاه داده عملیاتی پلیس

از آنجاکه ممکن است برخی از گیرنده‌های مستقر بر روی خودروها نتوانند موقعیت محاسبه شده خود را به دلیل اختلالات بستر مخابراتی به مرکز پلیس ارسال کنند بنابراین لازم است که داده‌های آنها نیز به نحوی به انبارداده برسد بدین منظور لازم است که در این خودروها سیستمی تعبیه شود تا مختصات‌های گیرنده را بر روی یک حافظه قابل حمل نظیر یک flash memory ذخیره و در اولین پاسگاه پلیس‌راه در پایگاه داده پلیس تخلیه نماید. این داده‌ها می‌توانند از طریق ساختار Batch ETL پیش‌بینی شده به روش انبارداده‌های مرسوم به داخل انبارداده مکانی بارگذاری شوند.

۳-۳-۳- پیاده‌سازی سیستم صف اولویت:

جریان داده‌های GPS با سرعت و حجم بالا تولید و به مرکز پلیس ارسال می‌شوند بنابراین سیستم انبارداده با فشار[5] انبوهی از جریان داده‌های ورودی مواجه خواهد بود که منتظر تحلیل در انبارداده هستند. برای اینکه سیستم داده‌ای را از دست ندهد و بتواند کنترلی بر ترتیب تحلیل داده‌ها داشته باشد لازم است که یک سیستم صف طراحی گردد تا انبارداده بتواند داده‌ها را بر اساس اولویت آنها فراخوانی[6] کند.

صف یک ساختار داده‌ای چند عنصری است که عناصر آن به ترتیب ورود قابل استخراج هستند[7]. دو عملیات اساسی صف، افزایش (اضافه کردن عناصر جدید به صف) و کاهش (خارج کردن قدیمی‌ترین عناصر صف) می‌باشند. البته در این پروژه باید نوع دیگری از صف استفاده شود که به "صف اولویت" موسوم است. صف اولویت ساختمان داده‌ای که در آن ترتیب طبیعی عناصر نتایج اعمال ابتدایی آن را تعیین می‌کند.

کلیه گیرنده‌های GPS دارای یک ساعت اتمی دقیق هستند و به هنگام تعیین موقعیت وسیله نقلیه زمان دقیق را نیز ثبت می‌کنند. آنچه که توسط گیرنده‌ها به صورت رادیویی به مرکز پلیس ارسال می‌شود شامل ۳ عنصر موقعیت وسیله نقلیه، زمان ثبت موقعیت و کد شناسه وسیله نقلیه (شناسه گیرنده GPS) می‌باشد. این داده‌ها که با حجم و سرعت بالا به سیستم می‌رسند ممکن است که در زمان انتقال ترتیبشان برهم خورده و پس و پیش گردد. علاوه بر این بایستی در جایی به صورت موقت در انتظار تحلیل نگهداری شوند. سیستم صفی که به این منظور طراحی می‌شود باید براساس اولویت زمانی (مهر زمانی داده‌های GPS) آنها را صف‌بندی نماید.

بخش Controller هم وظیفه کنترل ورودی و خروجی صف یعنی Push و Pull را برعهده دارد.

۳-۳-۴- پیاده‌سازی مکعب داده متناسب با ابعاد تحلیلی مورد نیاز حمل و نقل جاده‌ای:

پیش از شکل‌گیری انبارداده لازم است که شاخص‌ها[8] توسط تحلیل‌گران به دقت مشخص شود، چون در مرحله کار با انبارداده اطلاعات اساسی هر تحلیل بر اساس همین شاخص‌ها شکل می‌گیرد. شاخص تقریباً همیشه مقدار عددی است. در نمونه موردی، شاخص‌ها عبارتند از: موقعیت دو بعدی و سرعت که با نامهای (x,y,speed) مشخص شده‌اند.

جدول fact که قلب مکعب داده را تشکیل می‌دهد شامل دو سری فیلد است یکی کلیدهای خارجی به ابعاد که در پروژه عبارتند از: Time و hoze و دیگری شاخص‌ها که در فوق ذکر گردید. که پس از آن می‌بایست شِمای لازم برای مکعب داده مشخص گردد. از آنجا که متداول‌ترین شِما برای انبارداده، شمای ستاره‌ای[9] است، از این شِما، جهت شکل دهی مکعب داده استفاده شد.

۳-۳-۵- Active Rule Engine

برای اینکه مدل طراحی شده بتواند در قبال رویدادهای پیش‌بینی نشده واکنش مناسب و بدون دخالت کاربر نشان دهد این بخش براساس سه‌تایی ECA که در آن E: رویداد، C: شرط، A: اقدام است، پیاده‌سازی شده است.

با داشتن چنین بخشی نظارت بر عملکرد سیستم برعهده خود سیستم بوده و هرگاه وضعیت خاص پیش آید مدل می‌تواند به طور خودکار اقدام مناسب را انجام دهد. عملکرد این بخش را با استفاده از مثالی در رابطه با پروژه پیاده‌سازی شده تشریح می‌گردد. فرض کنیم در سیستم انبارداده کنترل تخلفات جاده‌ای یکی از خودروهای در حال حرکت در مسیر جاده دچار سانحه شده و دیگر اطلاعاتی از موقعیت آن به سیستم نرسد (این سانحه یک Event محسوب می‌شود). برای تشخیص اینکه چنین رویدادی غیر منتظره است باید پایگاهی جهت ذخیره‌سازی شرایط غیر منتظره (Condition) وجود داشته و یک کنترکننده آنرا بررسی کند و سپس یک اقدام (Action) متناسب با شرایط (Condition) از مجموعه Actionها (پایگاه نگهداری Actionها) اختیار گردد. در این مثال اقدامات مناسب می‌تواند شامل ارسال اعلان[10] به کاربران و همچنین ارسال یک اقدام عملیاتی به سیستم صف باشد. بنابراین ملاحظه می‌گردد که داده‌ها پیش از ورود به سیستم صف توسط کنترل کننده وقوع رخدادهای خاص ارزیابی شده و چنانچه براساس پایگاه قواعد دارای شرایط خاص باشند اقدام مناسب شامل ارسال اعلان‌ها به کاربران و اقدام عملیاتی مناسب به سیستم صف را اتخاذ می‌کند.

۳-۳-۶- جداول Recent , Historical

با ایجاد سیستم صف می‌توان انبارداده را قادر ساخت که بتواند بر حجم و سرعت داده‌های ورودی کنترل داشته و داده‌های جدید را در زمان‌های فراغت از تحلیل پردازش کند. اما داده‌های GPS پس از پردازش در سیستم و انجام تحلیل‌ها زمان اعتبار[11] خود را از دست

---

[5] push
[6] pull
[7] FIFO: First Input First Output

[8] Measure
[9] Star schema
[10] Notification
[11] Validity Time

Archive of SID

www.SID.ir

داده و چون فضای زیادی را در جداول مربوطه اشغال می‌کنند و سرعت تحلیل‌ها و پاسخگویی به پرس‌و جوها را پایین آورده و موجب تاخیر می‌شوند باید به دور ریخته شوند البته ممکن است این مشکل به وجود آید که یک "پرس‌و جوی موردی[12]" مطرح شود تا جهت تحلیل آن سیستم نیازمند داده‌های گذشته و به دور ریخته شده باشد بنابراین لازم است که با اعمال روش‌های خاص براین مشکل فائق آمد.

برای طراحی این قسمت لازم است که مجدداً به نوع پرس‌وجوهایی که ممکن است از سیستم مطرح شوند، نگاهی بیندازیم. گفته شد که دسته‌ای از پرس‌وجوهایی که در این پروژه داریم پیوسته (Continuous) بوده و امکان مانیتورینگ و پایش تخلفات سرعتی رخ داده در مسیر را برای پلیس فراهم کند. بدین صورت که در هر لحظه هرگونه تخطی از سرعت مجاز را برروی محور جاده نشانه‌گذاری و ثبت می‌کند و پلیس می‌تواند تراکم تخلفات را درطول یک محور رصد کند. و نقاط حادثه‌خیز را در طول یک جاده شناسایی کند. چنین تحلیلی مستلزم محاسبه سرعت لحظه‌ای است که فقط به دو مختصات آخر[13] هر خودرو و اختلاف زمانی میان آنها نیاز دارد. چنانچه این مقدار بالای ۱۲۰ کیلومتر بر ساعت فرضاً باشد یک تخلف سرعت محسوب شده و مکان آن برروی محور جاده نشانه‌گذاری می‌شود. لذا چنین پرس‌وجویی همواره آخرین موقعیت خودرو و موقعیت ماقبل آنرا جهت تحلیل نیاز دارد و پس از آن نیازی به نگهداری آنها نیست اما بایستی در جای دیگری جهت تحلیل‌های موردی که به داده‌های گذشته هم رجوع می‌کنند نگهداری شوند.

براساس آنچه گفته شد داده‌ها را بایستی در قالب جداولی به دو بخش تقسیم کنیم. یک بخش "آخرین داده‌ها" که در جدولی به نام Recent نگهداری شده و مورد استفاده "پرس‌وجوهای پیوسته" هستند و دیگری "داده‌های انقضا یافته‌ای" که در جدول History قرار می‌گیرند برای انجام این فرایند نیاز به یک "واسطه انتقال[14]" داریم که با استفاده از تریگر داده‌هایی را که در جدول Recent استفاده شده و تاریخ مصرف آنها می‌گذرد به جدول History انتقال می‌دهد.

بدیهی است پرس‌وجوهای موردی ad-hoc، هردو جدول Recent, History را با هم به کار می‌گیرند. داده‌های جدول History نیز با آمدن داده‌های جدید زمان اعتبار خود را از دست می‌دهند و بایستی جهت جلوگیری از انباشته شدن داده‌های بی مورد به دور ریخته شوند.

جهت افزایش کارایی سیستم و ارتقا توان انبارداده در تحلیل‌های جامع‌تر استراتژی دیگری را می‌توان به کار برد و آن اینکه نتایج پرس‌و جوهای موردی را در جدول دیگری به صورت خلاصه شده با عنوان Summary-table نگهداری کرد تا برروی آنها هم بتوان تحلیل‌های پیوسته و موردی دیگری را که به اطلاعات جامع‌تر و کلی‌تر نیاز دارند، به کار برد.

جهت روشن شدن موضوع و تشریح فرایند انتقال اطلاعات میان جداول Recent, History, Summary مثال نمونه موردی سیستم حمل و نقل جاده‌ای در ذیل تشریح می‌گردد:

جریان داده صف شده جهت تحلیل به داخل جدول Recent انتقال می‌یابد. چون داده‌های GPS خودروها به صورت جریانی به داخل این جدول سرازیر می‌شوند حجم جدول در حد انفجاری افزایش یافته و سرعت تحلیل‌ها پایین می‌آید. برای رفع این مشکل دو رکورد آخر هر خودرو را در این جدول باقی گذاشته و رکوردهای قدیمی را توسط واسطه انتقال به جدول Historical انتقال می‌دهیم. با این کار در جدول Recent صرفاً دو مختصات آخری هر خودرو را داریم که می‌توانیم سرعت لحظه‌ای هر خودرو را محاسبه کرده و نتیجه را برای کاربرانی که پرس‌وجوی پیوسته‌ای نظیر اینکه موقعیت تخلفات سرعتی بیش از ۱۲۰ کیلومتر برساعت کجاهاست بر روی مانیتور نمایش دهیم. اما ممکن است برخی دیگر از کاربران پلیس به دنبال تحلیل‌های موردی نظیر این باشند که تعداد تخلفات سرعت به تفکیک محدوده استحفاظی به تفکیک روزهای گذشته در یک جاده خاص چقدر بوده است؟ سیستم بایستی پاسخگوی چنین تحلیلی علاوه برداده‌های جدول Recent به داده‌های قدیمی خودروها نیز در جدول Historical نیاز خواهد داشت. چون داده‌های جدول Historical نیز بیش از یک هفته نگهداری نمی‌شوند لازم است که نتایج پرس‌وجوهای موردی را که خیلی مختصر و کم حجم هستند در جدولی به نام Summary نگهداری کنیم تا در پرس‌وجوهای موردی یا پیوسته آینده به کار گرفته شوند.

۳-۳-۷- پرس‌وجوهای پیوسته، موردی و جدول Summary

در این بخش پیش از تشریح نحوه طراحی معماری History, Recent، لازم است که سوالات و پرس وجوهای مدنظر تعریف گردد؛ (چون انبارداده‌ها متناسب با queryها شکل گرفته و با تغییر آنها تغییر می‌یابند لازم است که تعریف queryها پیش از طراحی انبارداده انجام پذیرد.)

پرس‌وجوهایی که در انبارداده‌ها طرح می‌شوند به دو دسته ad-hoc (موردی) و Continuous (پیوسته) تقسیم می‌شوند. پرس‌وجوهای ad-hoc پرسش‌های از پیش تعریف شده‌ای هستند که مقدار یک شاخص را با توجه به ابعاد مختلف در مکعب داده استعلام می‌کنند. در حالیکه پرسش‌های Continuous جهت مانیتورینگ (پایش) آخرین وضعیت داده‌ها به‌کار می‌روند.

پرس‌و جوهای پیوسته همانگونه که از نامشان پیداست برای مانیتورینگ و پایش آخرین تغییرات داده‌ها در انبارداده صورت می‌پذیرند.

۴- نتیجه گیری

[12] Ad-Hoc
[13] Recent
[14] Transfer Agent

Archive of SID

www.SID.ir

با توجه به ارزیابی و قابلیت‌های معماری‌های مختلف مدلی متناسب با کارکرد مورد نیاز این پروژه (سیستم تحلیل‌گر حمل و نقل جاده‌ای) طراحی و پیاده‌سازی گردید. توانایی پاسخگویی به هر دو نوع پرس‌وجوهای موردی و پیوسته از ویژگی‌های خاص معماری مدل پیشنهادی است. چون پرس‌وجوهای پیوسته در این پروژه به منظور مانیتورینگ و پایش تخلفات سرعتی در جاده‌ها صورت می‌پذیرد بنابراین امکان کاهش حجم جداول recent وجود دارد بدین معنی که در محاسبات سرعت لحظه‌ای خودروها دو رکورد آخر موقعیت هر خودرو کافی است. این کاهش حجم در اثر فیلتراسیون سرعت پرس‌وجوهای لحظه‌ای به تعداد قابل توجهی افزایش یافت.

برای انجام پرس‌وجوهای موردی نیاز به طراحی مجدد مکعب داده می‌باشد که در این پروژه صرفاً نیاز به تغییر ابعاد مکعب بود و لزومی به تغییر شمای ستاره‌ای با توجه به ویژگی‌های داده مکانی نیست.

از قابلیت‌های دیگر مدل پیشنهادی استفاده از «صف اولویت» بود که بر اساس مهرزمانی داده‌های GPS آنها را صف‌بندی می‌کرد و چون این مهرزمانی یکی از اجزای اصلی تاپل‌های ارسالی از GPS است، سیستم صف پیاده‌سازی شده به خوبی پاسخگوی Push جریان داده ورودی و تقاضای اخذ داده‌ها بود.

در اینجا یک نتیجه مهم نیز به دست آمد و آن این بود که به‌هنگام پیاده‌سازی سوالی مطرح گردید: اگر گیرنده‌های GPS و سیستم مخابراتی جانبی آن دچار اختلال شوند به گونه‌ای که GPS موقعیت‌ها را ثبت کند اما امکان ارسال آنها به مرکز پلیس به صورت مخابراتی امکان‌پذیر نباشد (نظیر آنچه که در حال حاضر، در سیستم فعلی پلیس‌راه وجود دارد) چگونه می‌توانیم از داده‌های این چنین خودروهایی استفاده کنیم؟ پاسخی که به دست آمد این بود که از داده‌های این GPS نمی‌توانیم به صورت بلادرنگ جهت پرس‌وجوهای پیوسته و مانیتورینگ رخدادها استفاده کنیم، اما بایستی امکاناتی از مدل ایجاد می‌شد که در پرس‌وجوهای موردی به کار گرفته شوند. به همین لحاظ بخشی پیاده سازی گردید که بتواند به روش مرسوم اینچنین داده‌ها را از منابع عملیاتی پلیس در پاسگاه‌ها جمع‌آوری و جهت تحلیل در سیستم ETL کند که این موضوع موجب توانایی مدل در مواجهه حالت‌های مختلف و استفاده از حداکثر داده‌ها در فرایند تحلیل گردد.

مراجع

[1]: Inmon, William H., *Building the Data Warehouse*, fourth edition, WILEY, USA, 2005.

[2]: Nguyen, Tho,M., *Zero-Latency Data Warehousing for Heterogeneous Data Source and continuous Data Stream*, Vienna University, 2006.

[3]: T. Pedersen and N. Tryfona, "*Pre-aggregation in spatial data warehouses*," In Proc. of the 7th Int. Symposium on Advances in Spatial and Temporal Databases, pages 460–478,2001.

[4]: Tjoa, A.Min, *"Zero-Latency Data Warehousing (ZLDWH): the State-of-the-art and experimental implementation approaches"*, institute of software technology and Interactive system, Austria, 2006.

[5]: Bruckner1 Robert M. Beate List1, and Schiefer2 Josef, *"Striving towards Near Real-Time Data Integration for Data Warehouses"*, DaWaK 2002, LNCS 2454, pp. 317-326, 2002.

[6]: Kimball Design group. *"Designing A Real Time Partition"*, Kimball Group: www.kimballgroup.com.

[7]: Pedersen,Torben Bach. *"Business Intelligence, Data Warehousing and Multidimensional Databases"*, DAT5 course, September 24, 2007.

[8]: Nguyen Tho Manh, Josef Schiefer, A Min Tjoa. *"Sense & Response Service Architecture (SARESA): An Approach towards a Real-time Business IntelligenceSolution and its use for a Fraud Detection Application"*, Copyright ACM 1-59593-162-7/05/0011,2006.

[9]: Rivest, S. Y. B´edard, and P. Marchand, *"Toward better support for spatial decision making: Defining the characteristics of spatial on-line analytical processing (SOLAP),"* Geomatica, 55(4) pages:539–555,2001.

[10]: Stefanovic, N., J. Han, and K. Koperski, *"Object-based selective materialization for efficient implementation of spatial data cubes,"* IEEE Trans. on Knowledge and Data Engineering, 12(6) pages:938–958, 2001.

Archive of SID

www.SID.ir