

## کاربرد روش یکپارچه سازی داده مبتنی بر CDC در شهر الکترونیک

مرتضی سرگلزایی جوان  
دانشگاه صنعتی امیرکبیر، تهران، ایران  
msjavan@aut.ac.ir

سید علیرضا هاشمی گلپایگانی  
دانشگاه صنعتی امیرکبیر، تهران، ایران  
Sa.hashemi@aut.ac.ir

### چکیده

در دنیای سریع امروزی، اهمیت و فوریت اطلاعات به یک موضوع حیاتی برای رقابت در کسب و کار تبدیل شده و رشد نمای حجم اطلاعات خام (که تقریباً هر ۱۸ ماه دوبرابر می‌شود) باعث شده است که فرایند قدیمی انتقال داده‌های انبوه بین منابع مبدأ و مقصد بسیار نامناسب و زمان بر شود. تکنولوژی ثبت تغییرات داده (Change Data Capture) یک جزء اصلی در بسیاری از سیستم‌های دارای Data Warehouse می‌باشد. تکنولوژی CDC نوعی روش یکپارچه سازی است که بر اساس شناسایی، ثبت و ارائه تغییرات اعمال شده در سیستم، عمل می‌کند. در صورتی که تنها تغییرات را پردازش کنیم، واضح است که مرحله استخراج اطلاعات در ETL بسیار بهینه‌تر انجام خواهد شد. این کار در سازمان‌های تجاری و یا دولتی در خصوص مدت زمان مورد نیاز جهت دستیابی به داده‌های جدیدتر و به روزتر بسیار ضروری است.

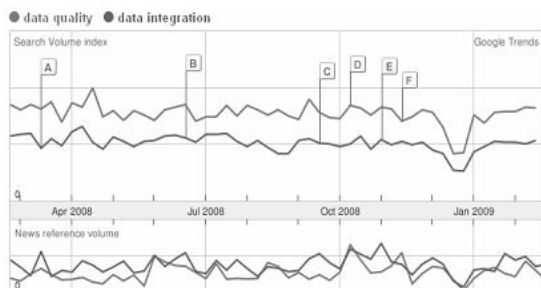
### واژگان کلیدی

یکپارچه سازی داده، یکپارچه سازی اطلاعات، تکنولوژی ثبت تغییر داده، پایگاه داده تحلیلی، پایگاه داده

### ۱- مقدمه

است (شکل ۱). علاوه بر این نشان دهنده میزان اهمیت داده و حجم تقاضا در خصوص یکپارچه سازی و رعایت کیفیت داده در سطح جهانی می‌باشد. در این مقاله پس از بررسی مختصر مفاهیم یکپارچه سازی داده، به بررسی و ارزیابی تکنولوژی CDC و راه حل‌های ارائه شده توسط آن، برای رفع مهم‌ترین چالش موجود در خصوص یکپارچه سازی داده خواهیم پرداخت. سپس کاربرد این تکنولوژی در سازمان‌ها و نقش موثر آن را در تصمیم‌گیری، تشریح خواهیم کرد.

یکپارچه سازی داده (Data Integration) به فرایند ترکیب داده‌های موجود در منابع داده مختلف اطلاق می‌شود. در سازمانها معمولاً داده‌ها توسط منابع مختلف و به شکل‌های مختلفی تهیه و نگهداری می‌شود. با توجه به اینکه مجموعه کاربردهایی که به Data Warehouse وابسته هستند اغلب به تمام این داده‌ها نیاز دارند بنابراین برای اینکه بتوان اطلاعات را از منابع داده مختلف با ساختارهای متفاوت و ناهمگن در Data Warehouse قرار داد، می‌بایست عملیات یکپارچه سازی داده را بر روی آنها انجام داد. با توجه به حجم عظیم اطلاعات تجاری و غیرتجاری در یک شهر الکترونیک و نیاز سازمانها برای دسترسی به آخرین اطلاعات جهت اخذ تصمیمات اثر بخش، لزوم بکارگیری تکنیک‌های موثر یکپارچه سازی برای یکپارچه کردن سریع اطلاعاتی که از منابع مختلف در اختیار سازمان قرار می‌گیرد به شدت احساس می‌شود. نتایج بررسی‌ها در طول یکسال گذشته نشان دهنده رویدادهای متنوعی است که در زمینه کیفیت داده (Data Quality) و یکپارچه سازی داده (Data Integration) اتفاق افتاده



شکل ۱- حجم جستجوهای انجام شده در مورد data integration در یکسال گذشته [6]