



موضوع

**یکپارچه سازی داده (Data Integration)**

ارائه دهنده

**سارا معصومی**

آذر ۱۴۰۱

## یکپارچه سازی داده چیست؟

یکپارچه سازی داده ها، فرآیندی است در هوش کسب و کار / هوش تجاری که در آن داده های موجود در منابع مختلف با یکدیگر ترکیب می شوند؛ این داده ها عموماً با فناوری های متفاوت و در منابع مختلف جدا از هم ذخیره شده و نگهداری می شوند. یکپارچه سازی داده ها منجر به شکل گیری یک دید واحد نسبت به داده های پراکنده می شود. امروزه Data Integration به فرآیندی مهم در ارتباط و اتصال سیستم های شرکت های مختلف با یکدیگر و تحکیم جایگاه برنامه های کاربردی و سیستم های داخلی شرکت ها تبدیل شده است.

یک سازمان داده محور نیازمند اطلاعات یکدست و یکپارچه است. یکپارچه سازی سیلوهای متعدد سازمانی یک بینش جدید در خصوص مشتریان، محصولات، کانال های بازاریابی و ... را ایجاد می نماید و در عمل بستر مناسب جهت نگرش ۳۶۰ درجه ای به موجودیت های اصلی کسب و کار (به عنوان نمونه مشتری) را در یک سازمان فراهم می کند. با نگرش محدود و جزیره ای به موجودیت های اصلی کسب و کار یک سازمان و با فرض صحت داده مرتبط با موجودیت ها، نمی توان در بازار رقابتی شدید امروزی حرف زیادی برای گفتن داشت و این نوع سازمان ها با فرض امکان استمرار حیات خود، ناچار خواهند شد که همواره مهمترین تصمیمات را بر اساس حداقل داده مرتبط با آن موجودیت بخصوص اتخاذ نمایند. این نوع سازمان ها نباید انتظار داشته باشند که سهم آنها در بازار رقابتی روند رو به رشدی را داشته باشد.

ادغام با فرآیند بلع ingestion شروع می شود و شامل مراحل مانند پاکسازی، نقشه برداری ETL و تبدیل می شود. یکپارچه سازی داده ها در نهایت ابزارهای تحلیلی را قادر می سازد تا هوش تجاری مؤثر و عملی تولید کنند.

هیچ رویکرد جهانی برای یکپارچه سازی داده ها وجود ندارد. با این حال، راه حل های یکپارچه سازی داده ها معمولاً شامل چند عنصر مشترک، از جمله شبکه ای از منابع داده، یک سرور اصلی، و مشتریانی هستند که به داده ها از سرور اصلی دسترسی دارند.

به عنوان مثال، برای یک مورد استفاده معمولی از ۳۶۰ مشاهده مشتری، داده هایی که باید ترکیب شوند ممکن است شامل داده های سیستم های CRM، ترافیک وب، نرم افزار عملیات بازاریابی، برنامه های کاربردی روبروی مشتری، فروش و سیستم های موفقیت مشتری و حتی داده های شریک باشد. اطلاعات از همه آن منابع مختلف اغلب برای نیازهای تحلیلی یا اقدامات عملیاتی باید کنار هم جمع شوند، و این کار کوچکی برای مهندسان داده یا توسعه دهندگان نیست که همه آنها را با هم جمع کنند.

## چرا داده هارا یکپارچه میکنیم؟

### ۱. صرفه جویی در زمان و افزایش کارایی:

هنگامی که یک شرکت اقداماتی را برای یکپارچه سازی صحیح داده های خود انجام می دهد، زمان آماده سازی و تجزیه و تحلیل آن داده ها را به میزان قابل توجهی کاهش می دهد. اتوماسیون نماهای یکپارچه نیاز به جمع آوری دستی داده ها را کاهش می دهد و کارمندان دیگر نیازی به ایجاد اتصالات از ابتدا ندارند، هر زمان که نیاز به اجرای گزارش یا ساختن یک برنامه داشته باشند. علاوه بر این، استفاده از ابزارهای مناسب، به جای کدگذاری دستی ادغام، زمان (و در کل منابع) بیشتری را به تیم توسعه دهنده بازمی گرداند.

تمام زمانی که در این وظایف صرفه جویی می شود، می تواند برای استفاده های بهتر، با ساعت های بیشتری برای تجزیه و تحلیل و اجرا در نظر گرفته شود تا سازمانی سازنده تر و رقابتی تر شود.

### ۲. کاهش خطا و جلوگیری از دوباره کاری:

وقتی صحبت از منابع داده یک شرکت می شود، چیزهای زیادی برای پیگیری وجود دارد. برای جمع آوری دستی داده ها، کارمندان باید هر مکان و حسابی را که ممکن است نیاز به کاوش داشته باشند بدانند و قبل از شروع همه نرم افزارهای لازم را نصب کنند تا مطمئن شوند مجموعه داده هایشان کامل و دقیق است. اگر یک مخزن داده اضافه شود و آن کارمند از آن بی خبر باشد، یک مجموعه داده ناقص خواهد داشت. علاوه بر این، بدون راه حل یکپارچه سازی داده ها که داده ها را همگام سازی می کند، گزارش دهی باید به طور دوره ای مجدداً انجام شود تا هرگونه تغییری در نظر گرفته شود. با این حال، با به روزرسانی های خودکار، گزارش ها می توانند به راحتی در زمان واقعی، هر زمان که نیاز باشد، اجرا شوند.

### ۳. ارزشمندی داده ها:

تلاش های یکپارچه سازی داده ها در واقع ارزش داده های یک کسب و کار را در طول زمان بهبود می بخشد. همانطور که داده ها در یک سیستم متمرکز ادغام می شوند، مسائل کیفیت شناسایی می شوند و بهبودهای لازم اجرا می شوند، که در نهایت منجر به داده های دقیق تر می شود که پایه و اساس تجزیه و تحلیل کیفیت هستند.

## انبار داده چیست؟

از طریق فرآیند یکپارچه سازی دید واحدی نسبت به داده های شرکت برای کاربران و کارمندان کسب و کار مربوطه ایجاد می شود. این طرح غالبا تحت عنوان "انبار داده" نامیده می شود.

می توان شناخته شده ترین پیاده سازی از رویکرد یکپارچه سازی داده ها را ساخت انبار داده سازمانی دانست. مزایای ناشی از ساخت انبار داده، کسب و کارها را قادر می سازد تا تجزیه و تحلیل های خود را مبتنی بر داده های جمع آوری شده در انبار داده انجام دهند.

ابتکارات یکپارچه سازی داده ها به ویژه در بین مشاغل بزرگ اغلب برای ایجاد انبارهای داده استفاده می شود که منابع داده های متعدد را در یک پایگاه داده رابطه ای ترکیب می کند. انبارهای داده به کاربران این امکان را می دهند که پرس و جوها را اجرا کنند، گزارش ها را جمع آوری کنند، تجزیه و تحلیل تولید کنند و داده ها را در قالبی ثابت بازایی کنند. به عنوان مثال، بسیاری از شرکت ها به انبارهای داده مانند Microsoft Azure و AWS Redshift برای تولید هوش تجاری از داده های خود متکی هستند.

Extract, Transform, Load استخراج، تبدیل، بارگذاری، که معمولا به عنوان ETL شناخته می شود، فرآیندی است در یکپارچه سازی داده ها که در آن داده ها از سیستم منبع گرفته شده و به انبار تحویل داده می شوند. این فرآیند مستمری است که انبار داده متعهد می شود تا چندین منبع داده را به اطلاعات مفید و سازگار برای هوش تجاری و تلاش های تحلیلی تبدیل کند. ساده سازی هوش تجاری (BI)

با ارائه یک نمای یکپارچه از داده ها از منابع متعدد، یکپارچه سازی داده ها فرآیندهای تحلیل هوش تجاری (BI) را ساده می کند. سازمان ها می توانند به راحتی مجموعه داده های موجود را مشاهده کرده و به سرعت درک کنند تا اطلاعات عملی در مورد وضعیت فعلی کسب و کار به دست آورند. با یکپارچه سازی داده ها، تحلیلگران می توانند اطلاعات بیشتری را برای ارزیابی دقیق تر بدون غرق شدن در حجم بالا جمع آوری کنند.

## چالش های پیش روی یکپارچه سازی داده

در نگاه اول، شاید بزرگترین چالش پیش روی Data Integration، چالش های فنی پیاده سازی این رویکرد باشد. اما ذکر این نکته مهم است که چالش اصلی در واقع کل فرآیند یکپارچه سازی داده ها است. یکپارچه سازی داده ها از سه گام اصلی طراحی، پیاده سازی و تست سیستم پیاده سازی شده تشکیل شده است. در ادامه به بررسی این گام ها می پردازیم.

### (۱) طراحی

طراحی سیستم Data Integration شامل چندین مرحله و پاسخ به سوالات و تامین نیازمندی های متفاوتی چون موارد زیر است:

a. تجزیه و تحلیل نیازمندی ها (شناسایی نیازمندی های کسب و کار) در Data Integration و یکپارچه سازی داده ها، شامل پاسخگویی به مسائلی چون موارد زیر است:

- دلایل یکپارچه سازی داده ها و هدف از پیاده سازی آن در شرکت چیست؟
- داده های مورد نیاز از چه منابعی تامین می شوند؟
- آیا داده های موجود، ملزومات فرآیند یکپارچه سازی داده ها را برآورده می کنند؟
- قوانین مرتبط بر کسب و کار مربوطه چه هستند؟
- مدل های پشتیبان و موافقت نامه های سطح خدمات آن کسب و کار چه هستند؟
- انتظار سازمان شما از اجرای پروژه یکپارچه سازی داده چیست ؟
- از چه ابزار و یا پلت فرمی باید استفاده کرد؟

❖ مهمترین هدف اجرای پروژه یکپارچه سازی داده ، حذف سیلوهای اطلاعاتی و حرکت به سمت تعامل منطقی و سیستماتیک سیستم های عملیاتی در یک سازمان است .

b. تجزیه و تحلیل سیستم منابع ذخیره اطلاعات در یکپارچه سازی داده ها شامل سؤالاتی درباره نحوه استخراج داده ها از سیستم (روش هایی چون استخراج همراه با ارائه پیغام بروزرسانی، استخراج افزایشی و full extract) می شود که شامل سؤالاتی همچون سوالات زیر است:

- نرخ استخراج داده های مورد نیاز یا در دسترس چقدر باشد؟

- کیفیت داده ها چگونه و در چه حد است؟
- آیا فیلدهای داده‌ای بدرستی پر شده‌اند؟
- آیا مستندسازی داده ها در دسترس است؟
- حجم داده های پردازش شده چقدر است؟
- مالک سیستم کیست؟

- c. نیازمندی های غیر عملکردی سیستم، مانند data processing window، زمان پاسخ‌دهی سیستم ، تخمینی از تعداد کاربران سیستم (به‌خصوص تعداد کاربرانی که به‌صورت هم‌زمان از سیستم استفاده می‌کنند)، سیاست به کار گرفته شده در تامین امنیت داده ها و بک آپ گیر یاز داده ها نیز باید تامین شوند.
- d. مدل پشتیبان به کار رفته در سیستم جدید چیست؟ نیازمندی های موافقت نامه‌های سطح خدمات چه مواردی را شامل می‌شوند؟
- e. در انتها نیز باید به این مساله پرداخت که مالک سیستم کیست و هزینه‌های نگهداری و به‌روزرسانی این سیستم از چه منبعی تامین می‌شود؟

نتایج گام‌های ذکر شده باید به فرم سند شناسایی ملزومات نرم افزاری [۹] نوشته شده و توسط تمام طرف‌های شریک در پروژه یکپارچه سازی داده ها تایید و امضا شود.

## ۲. پیاده سازی

براساس موارد مطرح شده در اسناد نیازمندی های کسب و کار و اسناد نیازمندی‌های نرم افزاری، برای انتخاب ابزارهای مناسب مورد نیاز در یکپارچه سازی داده ها، بهتر است که مطالعات امکان سنجی [۱۰] انجام شود.

شرکت‌ها و کسب و کارهای کوچکی که در ابتدای راه بهره‌برداری از پروژه های انبارش داده ها قرار گرفته‌اند، با چالش تصمیم‌گیری در مورد ابزارهای مورد نیاز پیاده‌سازی این راهکار مواجه‌اند. کسب و کارهای بزرگتر که در پیاده سازی پروژه های Data Integration تجربه دارند، کار نسبتاً راحت‌تری در انبارش داده ها پیش رو دارند، چرا که همانطور که اشاره شد تجربه پیاده سازی آن رویکردها را داشته و همچنین می‌توانند سیستم‌ها و دانش در اختیارشان را برای پیاده سازی رویکردهای جدید، توسعه داده و مورد بهره‌برداری قرار دهند.

طبق شواهد موجود، می‌توان گفت استفاده از پلت‌فرم‌ها و فناوری‌های جدیدتر و بهتر، در مقایسه با رویکرد تداوم استفاده از استانداردهای فعلی شرکت، موجب عملکرد موثرتر سیستم می‌شود. به طور مثال می‌توان از ابزارهایی نام برد که موجب مقیاس‌گذاری بهتر رشد و توسعه شرکت در آینده می‌شوند. این ابزار، راهکاری را فراهم می‌کند

که موجب کاهش هزینه های پیاده سازی و پشتیبانی، کاهش هزینه های اخذ مجوز و در نهایت تبدیل سیستم قدیمی به پلتفرمی جدید و مدرن می شود.

### ۳. آزمون و تست سیستم

همزمان با پیاده سازی فرآیند یکپارچه سازی داده ها، اطمینان از درستی و کامل و به روز بودن داده های یکپارچه شده از طریق آزمون و تست مداوم سیستم، گامی اساسی است. در مرحله تست، هر دو بخش نیازمندی های کسب و کار و نیازمندی های فنی باید مد نظر قرار گیرند تا از انطباق نتایج به دست آمده با نتایج مورد انتظار و مورد نیاز اطمینان حاصل شود. بنابراین، بهتر است که تست سیستم در قالب ترکیبی از تست ها شامل تست فشار حداقل کارایی (PST) *at Least Performance Stress test*، تست تاییدیه فنی *Technical Acceptance* *Testing (TAT)* و تست تاییدیه کاربر *User Acceptance Testing (UAT)* انجام شود.

### حوزه های کاری مرتبط با یکپارچه سازی داده

پیاده سازی فرایندهای یکپارچه سازی داده ها یا *Data Integration* معمولاً حوزه های کاری زیر را در بر می گیرد:

- انبارش داده ها (*Data Warehousing*)
- انتقال داده ها (*Data Migration*)
- یکپارچه سازی اطلاعات / برنامه های کاربردی کسب و کار با یکدیگر
- مدیریت بالاسری داده ها (*Master Data Management*)

## تکنیک های مختلف یکپارچه سازی داده

- یکپارچه سازی دستی یا مبتنی بر رابط کاربری مشترک Common User Interface کاربران از طریق دسترسی به سیستم های منبع و رابط های مبتنی بر وب، اطلاعات خود را در اختیار داشته و می توانند آن ها را تغییر دهند. در این روش، دید واحدی برای تمامی کاربران بر روی تمامی داده ها در نظر گرفته نشده است.
- یکپارچه سازی مبتنی بر نرم افزار Application Based Integration در این روش، نیازمند ایجاد نرم افزاری برای پیاده سازی فرآیند یکپارچه سازی است. این رویکرد، در تعداد محدودی از کاربردها قابل پیاده سازی است.
- یکپارچه سازی مبتنی بر میان افزار Middleware Data Integration این شیوه، منطق حاکم بر فرآیند یکپارچه سازی داده ها را از سطح برنامه های کاربردی به سطح جدیدی که مبتنی بر میان افزار است، تغییر می دهد. اگرچه بر خلاف روش مبتنی بر نرم افزار، پایه یکپارچه سازی داده ها مبتنی بر برنامه های کاربردی شکل نمی گیرد، اما همچنان به برنامه هایی برای انجام فرآیند یکپارچه سازی داده ها نیاز است.
- دسترسی واحد به داده Uniform Data Access یا یکپارچه سازی مجازی Virtual Integration داده ها همچنان در سیستم های منبع نگهداری می شوند، اما برای کاربران و کارکنان مختلف در سطوح متفاوت کسب و کار، سطوح متفاوتی از دسترسی به آن داده ها تعریف می شود. به طور مثال تنها وقتی که کاربر به اطلاعات مشتری های کسب و کار اجازه دسترسی دارد، جزئیات اطلاعات آن ها در اختیار وی قرار خواهد گرفت. مزیت اصلی یکپارچه سازی مجازی داده ها، نبود فاصله بین به روزرسانی دیدهای تعریف شده برای کاربران و در نتیجه کاهش ریسک عدم تطابق داده ها در سیستم منبع و دیدهای تعریف شده برای کاربران است. علاوه بر این، نیازی به منابع اضافی برای ذخیره داده های مربوط به دیدهای تعریف شده برای هر یک از کاربران به صورت مستقل نیست. در مقابل، نقطه ضعف این روش محدودیت در دسترسی به تاریخچه داده ها، محدودیت در مدیریت نسخه های مختلف داده ها و محدودیت در اعمال این روش بر طیف وسیع و متفاوتی از منابع و پایگاه های داده ای اشاره کرد.
- یکپارچه سازی فیزیکی داده ها Physical Data Integration یا مبتنی بر ذخیره سازی مشترک داده ها Common Data Storage، یک کپی از داده های موجود در سیستم منبع، در یک سیستم جداگانه نگهداری شده و تغییرات انجام شده از سوی کاربران، بر داده های ثانویه اعمال می شود. از شناخته شده ترین نمونه های این رویکرد می توان انبار داده را نام برد. از مزایای این روش می توان به امکان مدیریت نسخه های مختلف داده ها و ترکیب داده های موجود در منابع متفاوت (کامپیوترهای بزرگ، پایگاه های داده ای، یا فایل های ساختار نیافته) با یکدیگر اشاره کرد. باید به این نکته نیز اشاره کرد که یکپارچه سازی فیزیکی داده ها نیازمند سیستمی مستقل برای کنترل حجم بالای داده ها می باشد.



## ویژگی ابزارهای یکپارچه سازی داده

ابزارهای یکپارچه سازی داده ها این پتانسیل را دارند که این فرآیند را تا حد زیادی ساده کنند. ویژگی هایی که باید در ابزار یکپارچه سازی داده ها جستجو کنید عبارتند از:

- Open source

معماری های منبع باز معمولاً انعطاف پذیری بیشتری را ارائه می دهند و در عین حال به جلوگیری از قفل شدن سیستم کمک می کنند.

- Portability

قابل حمل بودن: از آنجایی که شرکت ها به طور فزاینده ای به سمت مدل های ابری ترکیبی می روند، مهم است که بتوانیم یکبار ادغام داده های خود را بسازیم و آنها را در هر جایی اجرا کنیم.

- Ease of use

راحتی در استفاده: ابزارهای یکپارچه سازی داده ها باید آسان برای یادگیری و استفاده آسان با یک رابط رابط کاربری گرافیکی باشد تا تجسم خطوط لوله داده شما را ساده تر کند.

- A transparent price

قیمت شفاف ارائه دهنده: ابزار یکپارچه سازی داده شما نباید از شما برای افزایش تعداد رابط ها یا حجم داده ها هزینه ای درخواست کند.

- Cloud compatibility

سازگاری با محیط ابر: ابزار یکپارچه سازی داده شما باید به صورت بومی در یک محیط ابری واحد، چند ابری یا ترکیبی کار کند.

## معرفی یک نرم افزار قدرتمند برای یکپارچه سازی داده

برای شرکت شما بسیار مهم است که به هر مجموعه داده از هر منبع دسترسی کامل داشته باشد. پلتفرم یکپارچه سازی Talend Cloud به کسب و کارها کمک می کند تا داده ها را تقریباً از هر منبعی یکپارچه کنند و آن ها را برای تجزیه و تحلیل با هر انبار داده آماده کنند.

### Talend چیست؟

Talend یک پلت فرم یکپارچه سازی داده های منبع باز است. این نرم افزار و خدمات مختلفی را برای یکپارچه سازی داده ها، مدیریت داده ها، یکپارچه سازی برنامه های کاربردی سازمانی، کیفیت داده، ذخیره سازی ابری و Big Data ارائه می دهد. Talend اولین بار در سال ۲۰۰۵ به عنوان اولین فروشنده نرم افزار منبع باز تجاری نرم افزار یکپارچه سازی داده ها وارد بازار شد. در اکتبر ۲۰۰۶، اولین محصول خود را راه اندازی کرد - Talend Open Studio که در حال حاضر به عنوان Talend Open Studio برای یکپارچه سازی داده ها شناخته می شود. از آن زمان، طیف وسیعی از محصولات را عرضه کرده است که در بازار بسیار مورد استفاده قرار می گیرند.

Talend به عنوان رهبر نسل بعدی در نرم افزار یکپارچه سازی ابر و Big Data در نظر گرفته می شود. این به شرکت ها کمک می کند تا در زمان واقعی تصمیم گیری کنند و بیشتر داده محور شوند. به کمک این فناوری، داده ها در دسترس تر می شوند، کیفیت آن افزایش می یابد و می توان سریعاً به سیستم های مورد نظر منتقل شد.

### برخی از ویژگی های محصولات Talend:

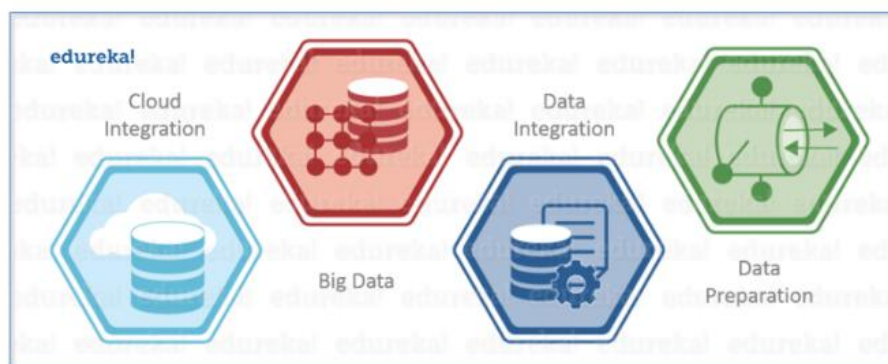
- توسعه و استقرار سریعتر: Talend وظایف را خودکار می کند و بیشتر آنها را برای شما حفظ می کند.
- هزینه کمتر: Talend ابزارهای منبع باز را ارائه می دهد که می توانند رایگان دانلود شوند. علاوه بر این، با افزایش سرعت فرآیندها، هزینه های توسعه به طور قابل توجهی کاهش می یابد.
- ثبات آینده: Talend شامل همه چیزهایی است که برای برآورده کردن الزامات بازاریابی امروز و همچنین در آینده نیاز داشته باشید. بنابراین، احتمال اینکه به زودی از بازار خارج شود بسیار کمتر است.
- پلتفرم یکپارچه: Talend همه نیازهای ما را تحت یک پایه مشترک برای محصولات بر اساس نیازهای سازمان برآورده می کند.

نگاهی به محصولات مختلف عرضه شده توسط Talend.

پلتفرم های Talend



پروژه های Talend



Talend Open Studio



در میان تمامی این محصولات، پرکاربردترین محصولات Talend Open Studios است که به صورت رایگان در دسترس هستند و هرکسی می تواند به راحتی آنها را دانلود و استفاده کند.

## مزایای Talend

- جهانی: تقریباً هر نوع داده را از هر منبع داده به هر مقصد داده، در محل یا در فضای ابری ادغام کنید
- قابل انعطاف: خطوط لوله داده را یک بار بسازید و آنها را در هر مکانی اجرا کنید، از جمله Apache Spark و آخرین فناوری‌های ابری، بدون قفل کردن فروشنده یا پلتفرم
- متحد: یکپارچه‌سازی داده‌ها، کیفیت داده‌ها و اشتراک‌گذاری داده‌ها را در یک راه‌حل واحد و یکپارچه که به کارگیری و استفاده آسان است، ترکیب کنید.
- پاکسازی اولیه
- طراحی سریعتر
- مدیریت کارآمد
- مقیاس پذیری آسان
- آمار زمان واقعی



با Talend، میتوان ۱ ترابایت داده مشتری را در زمان واقعی تجزیه و تحلیل می کنیم که به ما کمک می کند شناسایی کنیم روندهایی که ظاهر می شوند.

1. <https://dataintegration.info/data-integration>
2. <https://www.talend.com/resources/what-is-data-integration/>
3. <https://www.edureka.co/blog/what-is-talend-tool/>
4. Download Talend: <https://www.talend.com/>
5. Install Talend: <https://www.tutorialgateway.org/talend-download-and-install/>

<https://github.com/Sara0M> دانلود مقاله های بیشتر در