

## مروری بر چالش‌های متداول در پایگاه‌های داده منبع باز

محسن رباط جزئی<sup>۱</sup>، سیده فاطمه میرصادقی<sup>۲</sup>

<sup>۱</sup> دانشگاه بین المللی امام خمینی، قزوین mohroba@gmail.com

<sup>۲</sup> دانشگاه بین المللی امام خمینی، قزوین f.mirsadeghi@eng.ikiu.ac.ir

### چکیده

پایگاه‌های داده منبع باز در سال‌های اخیر و پیشرفت‌های حاصل شده در حوزه‌های نوین چون داده‌ی حجیم، شبکه‌های اجتماعی مدرن و رایانش توزیع شده، محبوبیت قابل توجهی به دست آورده اند و جایگزین‌های مقرون به صرفه و انعطاف پذیری را برای راه‌های پایگاه داده اختصاصی<sup>۱</sup> ارائه می‌دهند. با این حال، پذیرش و استفاده از پایگاه‌های داده منبع باز با مجموعه‌ای از چالش‌ها همراه است. این مقاله چالش‌های مرتبط با پایگاه‌های داده منبع باز را بررسی می‌کند و راهبردهایی برای کاهش ریسک حاصل از آن‌ها، ارائه می‌دهد. با درک و پرداختن مؤثر به این چالش‌ها، سازمان‌ها می‌توانند از قدرت پایگاه‌های داده منبع باز استفاده کنند، عملکرد را بهینه کنند، از یکپارچگی داده اطمینان حاصل کنند و زیرساخت پایگاه داده خود را ارتقا دهند. این مقاله بینش‌های ارزشمند و راهنمایی‌های عملی را برای سازمان‌هایی ارائه می‌کند که به دنبال استفاده از مزایای پایگاه‌های اطلاعاتی منبع باز و در عین حال کاهش ریسک‌ها و چالش‌های مرتبط با به کارگیری این دسته از نرم افزارهای منبع باز هستند.

واژه‌های کلیدی

پایگاه داده منبع باز، چالش‌ها، مقیاس پذیری، ریسک، مقابله با چالش‌ها.

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

## ۱. مقدمه

پایگاه‌های داده منبع باز به عنوان جایگزین‌های قدرتمندی برای پایگاه‌های داده اختصاصی ظاهر شده اند که مزایای متعددی مانند مقرون به صرفه بودن، انعطاف پذیری و شفافیت را ارائه می‌دهند. این پایگاه‌های داده، با کد منبع در دسترس و جامعه پر جنب و جوش، این فرصت را برای سازمان‌ها فراهم می‌کند تا از فناوری‌های پایگاه داده پیشرفته بدون محدودیت‌های مالی مرتبط با راه‌حل‌های اختصاصی استفاده کنند. [1]

در این مقاله تحقیقاتی، ما به چالش‌ها، بهینه‌سازی‌ها و ملاحظات عملکرد مرتبط با پایگاه‌های داده منبع باز می‌پردازیم. علاوه بر این، ما این جنبه‌ها را در انواع مختلف پایگاه‌های داده، از جمله پایگاه‌های داده رابطه‌ای، مبتنی بر اسناد، کلید - مقدار، ستونی، سری زمانی و گراف بررسی می‌کنیم. با بررسی هر نوع، می‌توانیم درک جامعی از چالش‌ها و بهینه‌سازی‌های خاص مربوط به نمونه‌های مختلف پایگاه داده به دست آوریم. به عنوان مثال، یکی از چالش‌های اصلی مختص پایگاه‌های داده منبع باز، پشتیبانی محدود در مقایسه با پایگاه‌های داده اختصاصی است. در حالی که جوامع منبع باز کمک ارائه می‌دهند، ممکن است فاقد فوریت و پشتیبانی جامعی باشد که معمولاً توسط تیم‌های پشتیبانی فروشنده یا سازنده اختصاص داده می‌شود. این می‌تواند بر حل مشکل و عیب‌یابی تأثیر بگذارد و نیاز به اتکا به انجمن‌ها و منابع جامعه باشد. علاوه بر این، مستندات می‌توانند به دلیل مشارکت‌های غیر متمرکز، چالش‌هایی ایجاد کنند، که منجر به اطلاعات قدیمی یا ناکافی می‌شود که درک و اجرای ویژگی‌های پایگاه داده پیشرفته توسط کاربران را مختل می‌کند. [2]

ملاحظات مقیاس پذیری و عملکرد در مدیریت پایگاه داده بسیار مهم است. پایگاه‌های داده منبع باز ممکن است به بهینه‌سازی و پیکربندی اضافی برای دستیابی به عملکرد و مقیاس پذیری بهینه نیاز داشته باشند. تنظیمات پیش فرض و خارج از جعبه<sup>۲</sup> ممکن است برای بارهای کاری خاص تنظیم نشده باشند، که نیاز به تنظیمات دستی دارد. علاوه بر این، ادغام یکپارچه با ابزارهای شخص ثالث می‌تواند، به دلیل مشکلات سازگاری یا منابع محدود اختصاص داده شده به توسعه اتصالات، چالش برانگیز باشد. این ممکن است مانع پذیرش پایگاه‌های داده منبع باز در محیط‌هایی شود که ادغام با ابزارهای اختصاصی ضروری است. [3], [4]

مدیریت امنیت و آسیب پذیری، از جنبه‌های حیاتی هر سیستم پایگاه داده است. پایگاه‌های داده منبع باز بر تلاش‌های جامعه محور برای ممیزی امنیتی، تشخیص آسیب پذیری و انتشار وصله تکیه دارند. فقدان تیم‌های امنیتی اختصاصی و فرآیندهای حسابرسی<sup>۳</sup> جامع ممکن است منجر به تأخیر در رسیدگی به آسیب پذیری‌های امنیتی شود که به طور بالقوه، سیستم‌ها را در معرض خطر قرار می‌دهد. چالش‌های ایجاد شده با پذیرش پایگاه‌های داده منبع باز، می‌تواند ناشی از پشتوانه تجاری محدود و عدم آشنایی محیط‌های سازمانی باشد. غلبه بر این چالش‌ها، مستلزم ایجاد اعتماد از طریق نشان دادن قابلیت اطمینان، اقدامات امنیتی و داستان‌های موفقیت در اتخاذ پایگاه‌های داده منبع باز است. [5], [6]

یکی دیگر از جنبه‌های مهم پایگاه‌های داده منبع باز پیچیدگی آنها و منحنی یادگیری مرتبط است. ماهیت سفارشی سازی این پایگاه‌های داده، اغلب به کاربرانی نیاز دارد که مهارت‌ها و دانش خاصی را برای استقرار و مدیریت مؤثر آنها کسب کنند. این پیچیدگی ممکن است چالش‌هایی را برای سازمان‌هایی با منابع یا تخصص محدود ایجاد کند که ممکن است منجر به استقرار ناکارآمد یا پیکربندی‌های غیر بهینه شود. سرمایه گذاری در آموزش و توسعه مهارت، برای پیمایش راه پر پیچ و خم موفقیت در بکارگیری پایگاه‌های داده منبع باز، ضروری است. [3]

در قسمت‌های بعدی، به بررسی انواع پایگاه داده می‌پردازیم و چالش‌ها، بهینه‌سازی‌ها و ملاحظات عملکردی را که ارائه می‌دهند، تحلیل می‌کنیم. با به دست آوردن بینشی جامع از این عوامل، سازمان‌ها می‌توانند پیچیدگی‌های پایگاه‌های داده منبع باز را در نظر داشته باشند و از پتانسیل آنها برای برآورده کردن نیازهای مدیریت داده خود، به طور مؤثر، استفاده کنند.

### ۱.۱. شیوه کاری پژوهش

برای این مقاله مروری متمرکز بر چالش‌های معمول پایگاه‌های داده منبع باز، یک رویکرد مرور ادبیات جامع به کار گرفته شد. هدف این تحقیق تجزیه و تحلیل و ترکیب ادبیات موجود برای شناسایی و طبقه بندی چالش‌های مربوط به پایگاه‌های داده منبع باز بود. شیوه کاری شامل مراحل کلیدی زیر است:

<sup>2</sup> Out of the box

<sup>3</sup> Audit

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senacnf.ir

- انتخاب کلمه کلیدی: کلمات کلیدی مرتبط برای دامنه بررسی انتخاب شدند که از جمله آن ها عبارات فارسی مانند «پایگاه های اطلاعاتی منبع باز»، «ترم افزار منبع باز»، «پایگاه داده» و ترجمه انگلیسی آن ها استفاده شد.
- موتور های جستجو: یک جستجوی جامع در پایگاه های اطلاعاتی دانشگاهی از جمله Google Scholar, IEEE Xplore و ACM Digital Library با استفاده از کلمات کلیدی انتخاب شده برای بازیابی مقالات پژوهشی مرتبط انجام شد. همچنین برای اطلاعات بیشتر در مورد پایگاه های داده مطرح شده، به مستندات موجود در وبگاه های موجود برای هر یک مراجعه شد.
- معیارهای انتخاب: مقالات پژوهشی بر اساس ارتباط آنها با موضوع انتخاب شدند. معیارهای انتخاب به جای تحمیل محدودیت های زمانی، بر همسویی محتوا با چالش های پیش روی پایگاه های داده منبع باز متمرکز بود؛ به این دلیل که بحث پایگاه های داده منبع باز موضوع نوینی نیست اما پیشرفت آن ها به مرور زمان، چالش های جدیدی را مطرح کرده است که نیاز به بررسی موارد قدیمی و جدید به صورت جامع دارد.
- استخراج داده: اطلاعات کلیدی از مقالات انتخاب شده، مانند چالش های شناسایی شده، روش های مورد استفاده و یافته ها، برای تجزیه و تحلیل استخراج شد.
- ترکیب: چالش ها به مضامین دسته بندی شدند و زمینه های مشترک مورد توجه در استفاده از پایگاه داده منبع باز مورد اشاره قرار گرفته اند.

## ۲. بررسی ادبیات پژوهش

### ۱.۲. انواع پایگاه داده

پایگاه های داده نقش مهمی در مدیریت و سازماندهی حجم وسیعی از داده در صنایع و برنامه های مختلف دارند. این دسته از نرم افزارها را می توان بر اساس چندین معیار طبقه بندی و دسته بندی کرد که هر کدام ویژگی ها و عملکردهای متمایزی را ارائه می دهند. [7] در ادامه به بررسی هر یک از این معیارهای دسته بندی و عنصرهای آن ها خواهیم پرداخت.

#### مدل داده

بر اساس این معیار، پایگاه های داده بر پایه مدل داده ای که ذخیره می کنند، طبقه بندی می شوند. پایگاه های داده ای پر کاربرد در این دسته عبارتند از [8]:

- پایگاه داده رابطه ای<sup>۴</sup>: داده را در جداول با روابط از پیش تعریف شده سازماندهی کنید
- پایگاه داده مبتنی بر سند<sup>۵</sup>: داده را در قالب های اسناد انعطاف پذیر و نیمه ساختار یافته مانند جی سان<sup>۶</sup> یا اکس ام ال<sup>۷</sup> ذخیره کنید.
- پایگاه داده ذخیره کلید – مقدار<sup>۸</sup>: داده را به صورت جفت کلید-مقدار ذخیره کنید، که امکان بازیابی سریع را فراهم می کند.
- پایگاه داده ستونی<sup>۹</sup>: ذخیره داده در ستون ها به جای ردیف، بهینه سازی برای پرس و جوی های تحلیلی
- پایگاه داده سری زمانی<sup>۱۰</sup>: طراحی شده برای ذخیره و تجزیه و تحلیل داده ای زمان بندی شده
- پایگاه داده گراف<sup>۱۱</sup>: بر روابط بین موجودیت های داده، که به صورت گره ها و لبه ها نشان داده می شوند، تمرکز کنید

#### هدف و کارکرد

- پایگاه های داده را می توان بر اساس هدف و کارکرد آنها دسته بندی کرد. در این دسته بندی موارد زیر قرار می گیرند [9], [8]:

<sup>4</sup> Relational Databases

<sup>5</sup> Document Databases

<sup>6</sup> JSON

<sup>7</sup> XML

<sup>8</sup> Key-Value Databases

<sup>9</sup> Columnar Databases

<sup>10</sup> Time-Series Databases

<sup>11</sup> Graph Databases

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senacnf.ir

- پایگاه‌های داده عملیاتی<sup>۱۲</sup>: برای عملیات تراکنش‌های روزانه و پردازش داده در زمان واقعی استفاده می‌شود.
- پایگاه‌های داده تحلیلی<sup>۱۳</sup>: برای تجزیه و تحلیل داده‌ی پیچیده، گزارش دهی و هوش تجاری طراحی شده است.
- پایگاه‌های داده توزیع شده<sup>۱۴</sup>: شامل چندین پایگاه داده به هم پیوسته است که در مکان‌های مختلف پراکنده شده اند.
- پایگاه‌های داده کاربر نهایی<sup>۱۵</sup>: پایگاه‌های داده‌ای که در ایستگاه‌های کاری یا دستگاه‌های شخصی ذخیره می‌شوند.

## مدل استقرار

پایگاه‌های داده را می‌توان بر اساس مدل استقرار آنها دسته بندی کرد. برای مثال می‌توان به موارد زیر اشاره کرد [10]:

- پایگاه‌های داده در محل<sup>۱۶</sup>: مستقر و مدیریت شده در زیرساخت‌های خود سازمان
- پایگاه‌های داده ابری<sup>۱۷</sup>: میزبانی و مدیریت شده توسط ارائه دهندگان خدمات ابری که مقیاس پذیری و انعطاف پذیری را ارائه می‌کند.
- پایگاه‌های داده ترکیبی: ترکیب هر دو مؤلفه داخلی و مبتنی بر ابر، استفاده از مزایای هر دو

## روش دسترسی

پایگاه‌های داده را می‌توان بر اساس روش دسترسی که به کار می‌برند، دسته بندی کرد که به صورت ذیل می‌باشند [11]:

- دسترسی با زبان ساختار یافته پرس و جو<sup>۱۸</sup>: برای تعریف و دستکاری داده از زبان پرس و جو ساختار یافته استفاده می‌کنند.
- دسترسی بدون زبان ساختار یافته پرس و جو<sup>۱۹</sup>: استفاده از مدل‌های داده غیر رابطه ای و ارائه شیوه‌های انعطاف پذیر ذخیره سازی و بازیابی داده
- پایگاه‌های داده مبتنی بر دسترسی ترکیبی: مقیاس پذیری پایگاه‌های داده غیر رابطه ای را با خاصیت‌های اسید<sup>۲۰</sup> پایگاه‌های داده سنتی و رابطه ای ترکیب می‌کنند.

## ساختار و نحوه فایل بندی

پایگاه‌های داده را می‌توان بر اساس مکانیزم ذخیره سازی آنها دسته بندی کرد. مانند [8]:

- پایگاه‌های داده با فایل مسطح<sup>۲۱</sup>: داده را در فایل‌های متنی ساده ذخیره می‌کند که اغلب برای ذخیره سازی و بازیابی ساده داده استفاده می‌شود.
- پایگاه‌های داده سلسله مراتبی<sup>۲۲</sup>: داده را در یک ساختار درخت مانند با روابط والدین و فرزند سازماندهی می‌کنند.
- پایگاه‌های داده شبکه: نمایش داده با روابط پیچیده و به هم پیوسته، شبیه به ساختار گراف
- پایگاه‌های داده شی گرا<sup>۲۳</sup>: داده را به عنوان اشیا ذخیره کنید، که امکان مدل سازی و ارث بری داده‌ی پیچیده را فراهم می‌کند.

<sup>12</sup> Operational Databases

<sup>13</sup> Analytical Databases

<sup>14</sup> Distributed Databases

<sup>15</sup> End-User Databases

<sup>16</sup> On-Premises Databases

<sup>17</sup> Cloud Databases

<sup>18</sup> Structured Query Language (SQL)

<sup>19</sup> Not-only SQL (NoSQL)

<sup>20</sup> ACID (Atomicity, Consistency, Isolation, Durability)

<sup>21</sup> Flat-file Databases

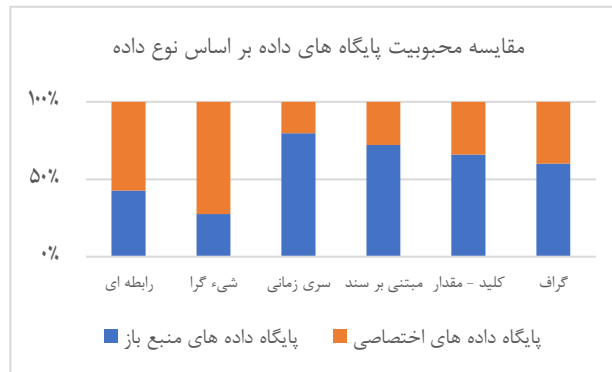
<sup>22</sup> Hierarchical Databases

<sup>23</sup> Object-Oriented Databases

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir



شکل ۱ - نمودار مقایسه پایگاه داده های متن باز و اختصاصی در دسته بندی های مختلف مبتنی بر نوع داده طبق آخرین رتبه بندی های وبگاه DB-Engines

## ۲.۲. بررسی چند نمونه از پایگاه های داده پیشرو

به دست آوردن بینش در قلمرو متنوع ذخیره سازی و بازیابی داده زمانی ارزشمند می شود که ضمن بررسی دقیق دسته بندی های مختلف پایگاه های داده، نمونه های قابل توجه آنها را نیز مطالعه کنیم. با کاوش در پایگاه های داده مطرح و پیشرو در عرصه ذخیره سازی داده و دسته بندی های مربوط به آنها، می توانیم درک خود را از شیوه های عملیاتی و چالش های متمایز آنها عمیق تر کنیم. در ادامه این بخش به معرفی چند پایگاه داده اختصاصی و منبع باز خواهیم پرداخت.

پایگاه داده اوراکل<sup>۲۴</sup> یک نمونه اختصاصی و تجاری پر کاربرد است که در دسته پایگاه داده رابطه ای قرار می گیرد. مجموعه ای جامع از ویژگی ها را برای برنامه های کاربردی در سطح سازمانی ارائه می دهد که از جمله آنها می توان مدیریت تراکنش قوی، در دسترس بودن بالا و مقیاس پذیری اشاره کرد. پایگاه داده اوراکل معمولاً در صنایع مالی، مخابرات و شرکت های بزرگ استفاده می شود.<sup>[12]</sup> کاساندر<sup>۲۵</sup> یک پایگاه داده بسیار مقیاس پذیر و توزیع شده است که به دسته پایگاه های داده ستونی تعلق دارد و طوری طراحی شده است تا حجم زیادی از داده را در چندین سرور با قابلیت دسترسی بالا و تحمل خطا مدیریت کند. کاساندر معمولاً برای موقعیت هایی استفاده می شود که به توان نوشتن بالا و مدیریت داده های غیر متمرکز نیاز دارند، مانند برنامه های مبتنی بر اینترنت اشیا<sup>۲۶</sup>، داده های سری زمانی و تجزیه و تحلیل در لحظه<sup>۲۷</sup>[13].

مای اس کیو ال<sup>۲۸</sup> یک سیستم مدیریت پایگاه داده رابطه ای منبع باز محبوب است که داده را در جداول با روابط از پیش تعریف شده، ذخیره می کند از زبان ساختار یافته پرس و جو برای دستکاری و بازیابی داده بهره می گیرد. مای اس کیو ال به طور گسترده برای برنامه های کاربردی مختلف از جمله توسعه وب، انبار داده و پردازش تراکنش های آنلاین استفاده می شود.<sup>[14]</sup>

مانگو دی بی<sup>۲۹</sup> یک پایگاه داده سند گرا است که به طور گسترده مورد استفاده قرار می گیرد و داده را به جای جداول در اسناد انعطاف پذیر و نیمه ساختار یافته ذخیره می کند. مانگو، برای مدیریت داده های بدون ساختار و به سرعت در حال تحول، طراحی شده است و آن را برای برنامه هایی مانند سیستم های مدیریت محتوا، تجزیه و تحلیل بلادرنگ و برنامه های تلفن همراه مناسب می کند.<sup>[15]</sup>

نئو فور جی<sup>۳۰</sup> نمونه بارز یک پایگاه داده گراف است که به صورت اختصاصی برای مدیریت داده های بسیار متصل طراحی شده است و بر روابط بین موجودیت ها تمرکز دارد. پایگاه های داده گراف از گره ها و لبه ها برای نمایش داده و روابط استفاده می کنند که این امر آنها را برای موقعیت های مربوط به شبکه های پیچیده، شبکه های اجتماعی، سیستم های توصیه و نمودارهای دانش مناسب می کند.<sup>[16]</sup>

<sup>24</sup> Oracle

<sup>25</sup> Cassandra

<sup>26</sup> Internet of Things (IoT)

<sup>27</sup> Real-Time

<sup>28</sup> MySQL

<sup>29</sup> MongoDB

<sup>30</sup> Neo4j

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

همانطور که در شکل ۱ پیداست، محبوبیت پایگاه‌های داده منبع باز، به صورت چشم‌گیری افزایش یافته است. بخش بزرگی از این محبوبیت به دلیل کاربردهای بیشمار و کارایی مناسب در انجام امور علمی و دانش بنیان است که قابلیت توسعه پذیری و سفارشی سازی این دسته از سیستم‌های مدیریت پایگاه داده، نقش به‌سزایی در آن داشته است. [17]

بدیهی است که بررسی مزایای این نوع از پایگاه داده و مقایسه با رقیبان منبع بسته و اختصاصی، درک و بینشی عمیق‌تر نسبت به اهمیت استفاده از این دسته از پایگاه‌های داده را به ما می‌دهد و ما را در فهم بهتر و تلاش برای رفع چالش‌های موجود در استفاده از آن‌ها، یاری می‌کند. از این رو در قسمت آخر این بخش به بررسی مزایای استفاده از پایگاه‌های داده منبع باز به عنوان جایگزینی برای پایگاه‌های داده اختصاصی، خواهیم پرداخت.

## ۳.۲. مزایای استفاده از پایگاه‌های داده منبع باز

پایگاه‌های داده منبع باز معمولاً رایگان در دسترس هستند، که به طور قابل توجهی هزینه‌های اولیه مرتبط با اجرای پایگاه داده را کاهش می‌دهد. سازمان‌ها می‌توانند بودجه خود را به سایر حوزه‌های زیرساخت فناوری اطلاعات خود اختصاص دهند یا روی ویژگی‌ها و پیشرفت‌های اضافی سرمایه گذاری کنند.

دیگر مزیت، انعطاف پذیری و قابلیت سفارشی سازی است. پایگاه‌های داده منبع باز انعطاف پذیری لازم را جهت تغییر دادن و سفارشی کردن کد بر اساس نیازهای خاص فراهم می‌کنند. سازمان‌ها این آزادی را دارند که پایگاه داده را با نیازهای منحصر به فرد خود تنظیم کنند، ویژگی‌های جدید اضافه کنند یا عملکرد موجود را بهینه کنند. [3]

پایگاه‌های داده منبع باز شفافیت را ارائه می‌دهند و به کاربران امکان می‌دهند نحوه عملکرد پایگاه داده را بررسی و درک کنند. این شفافیت اعتماد را تقویت می‌کند و کاربران را قادر می‌سازد تا آسیب پذیری‌های امنیتی یا گلوگاه‌های عملکردی را شناسایی و برطرف کنند. [18]

مهم‌ترین مزیت پایگاه‌های داده متن باز بهره بردن از جامعه بزرگ و فعالی از کاربران و توسعه دهندگان می‌باشد. این انجمن‌ها کار پشتیبانی انجام می‌دهند، دانش متخصصین را به اشتراک می‌گذارند و به بهبود و تکامل مستمر پایگاه داده کمک می‌کنند. کاربران می‌توانند برای جستجوی کمک و راهنمایی به انجمن‌ها، اسناد و منابع، دسترسی داشته باشند. پایگاه‌های داده منبع باز اغلب به دلیل ماهیت مشارکتی جامعه، شاهد نوآوری و توسعه سریع هستند. ویژگی‌های جدید، پیشرفت‌ها و رفع اشکال‌ها به طور منظم توسط توسعه دهندگان در سراسر جهان ارائه می‌شوند و جریان دائمی بروز رسانی‌ها و بهبودها را تضمین می‌کنند.

علاوه بر این، به دلیل وجود کاربران گسترده و جامعه فعال، این سیستم‌های مدیریت پایگاه داده، تمایل به دوام طولانی مدت دارند. آنها کمتر مستعد قطع ناگهانی یا عدم پشتیبانی سازنده هستند و سازمان‌ها را نسبت به طول عمر راه حل پایگاه داده انتخابی خود، مطمئن می‌کند. [5]

مزیت دیگر پایگاه‌های داده منبع باز، ویژگی عدم داشتن قفل سازنده<sup>۳۱</sup> می‌باشد که طبق آن، سازمان‌ها به یک فروشنده خاص یا پشته فناوری اختصاصی وابسته نیستند. آن‌ها آزادی انتخاب سخت افزار، سیستم عامل‌ها و ابزارهایی را دارند که به بهترین وجه با نیازهایشان مطابقت دارند، بدون اینکه در زیست‌بوم یک فروشنده یا سازنده قفل شوند. [19]

حال که بررسی پایه ای پایگاه‌های داده منبع باز پرداختیم و اهمیت بررسی چالش‌های موجود در آن‌ها را واکاوی کرده‌ایم، در ادامه مقاله، به بررسی چالش‌های موجود در این نوع از سیستم‌های مدیریت پایگاه داده خواهیم پرداخت.

## ۳. چالش‌های به کار گیری پایگاه‌های داده متن باز

پایگاه‌های داده منبع باز به دلیل مقرون به صرفه بودن، انعطاف پذیری و پشتیبانی فعال جامعه محبوبیت قابل توجهی به دست آورده اند. با این حال، آنها همچنین با چالش‌های خاصی همراه هستند که سازمان‌ها باید در نظر بگیرند.

در این بخش، چالش‌های کلیدی مرتبط با پایگاه‌های داده منبع باز را بررسی خواهیم کرد و در مورد راهبردهای جامع برای کاهش آنها بحث خواهیم کرد.

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

## ۱.۳. بهینه سازی عملکرد و مقیاس پذیری

مقیاس پذیری و عملکرد ملاحظات حیاتی برای هر سیستم پایگاه داده، از جمله پایگاه‌های داده منبع باز هستند. با افزایش حجم داده و بار تراکنش‌ها، سازمان‌ها در حفظ عملکرد بهینه با چالش‌هایی مواجه می‌شوند. پایگاه‌های داده منبع باز ممکن است به توجه و راهبردهای بیشتری برای رسیدگی موثر به نگرانی‌های مقیاس پذیری و عملکرد نیاز داشته باشند. [4] در این بخش، این چالش‌ها را به تفصیل بررسی خواهیم کرد و راهکارهایی را برای کاهش آن‌ها مورد بحث قرار خواهیم داد.

### حجم و رشد داده

یکی از چالش‌های اصلی مقیاس پذیری برای پایگاه‌های داده منبع باز، مدیریت حجم داده در طول زمان است. با افزایش حجم داده، می‌تواند بر زمان پاسخ پرس و جو و عملکرد کلی سیستم تأثیر بگذارد. چندین راهبرد می‌تواند به رفع این چالش کمک کند که عبارتند از [20]:

- **بهینه سازی طرح واره<sup>۳۲</sup> پایگاه داده:** برای این کار، ابتدا طرح واره پایگاه داده را برای اطمینان از ذخیره سازی و بازیابی کارآمد داده تجزیه و تحلیل کنید. برای حذف افزونگی و کاهش تکتیر داده، طرح واره را نرمال سازی<sup>۳۳</sup> کنید. این بهینه سازی عملکرد پرس و جو را بهبود می‌بخشد و نیازهای ذخیره سازی را به حداقل می‌رساند.
- **اجرای راهبردهای فهرست بندی<sup>۳۴</sup>:** فهرست بندی مناسب برای بازیابی کارآمد داده بسیار مهم است. الگوهای پرس و جو باید تجزیه و تحلیل شوند و برای ستون‌هایی که اغلب درخواست می‌شوند، باید فهرست‌های مناسب ایجاد کرد. با این حال، فهرست بندی بیش از حد نیز می‌تواند بر عملکرد نوشتن تأثیر بگذارد، بنابراین ایجاد تعادل و بازبینی دوره ای و بهینه سازی فهرست‌ها ضروری است.
- **حذف نرمال سازی<sup>۳۵</sup> برای بهبود عملکرد:** در برخی موارد، حذف نرمال سازی به وسیله کاهش نیاز به اتصالات پیچیده در چندین جدول، می‌تواند عملکرد را افزایش دهد. با تکتیر و ذخیره داده‌ی مرتبط به شکل نرمال نشده، عملکرد پرس و جو را می‌توان به قیمت افزایش نیازهای ذخیره سازی و ناهنجاری‌های احتمالی بروزرسانی داده، بهبود بخشید.
- **تکنیک‌های قسمت بندی<sup>۳۶</sup>:** قسمت بندی شامل تقسیم یک جدول بزرگ به قسمت‌های کوچکتر و قابل مدیریت بر اساس معیارهای خاص (مانند محدوده، لیست یا هش) است. توزیع بهتر داده، بازیابی سریعتر داده و بهبود عملکرد پرس و جو، توسط این کار امکان پذیر می‌گردد.

### بارهای تراکنش بالا

پایگاه‌های داده منبع باز ممکن است در هنگام برخورد با بارهای تراکنش بالا، به ویژه در موقعیت‌هایی با خواندن و نوشتن همزمان متعدد، با چالش‌هایی مواجه شوند. مدیریت حجم تراکنش‌های بالا مستلزم بررسی دقیق و اجرای راهبردهای مناسبی مانند موارد زیر است [21]:

- **بهینه سازی طرح‌های اجرای پرس و جو<sup>۳۷</sup>:** عملکرد پرس و جو را می‌توان با بهینه سازی طرح‌های اجرای پرس و جو بهبود بخشید. باید اطمینان حاصل شود که پرس و جوها به خوبی نوشته شده اند، از فهرست بندی‌های مناسب استفاده شده است و از اسکن‌های غیر ضروری جداول خودداری می‌گردد. از ابزارهای خاص پایگاه داده یا تحلیلگرهای پرس و جو برای شناسایی و بهینه سازی پرس و جوهای ناکارآمد می‌توان استفاده کرد.
- **شیوه‌های ذخیره موقت (کش)<sup>۳۸</sup>:** پیاده سازی شیوه‌های کش می‌تواند عملکرد را به طور قابل توجهی افزایش دهد. از راه حل‌های برای ذخیره داده‌یی که مداوم به آنها دسترسی دارید، در حافظه، استفاده کنید. این امر نیاز به خواندن و نوشتن دیسک را کاهش می‌دهد و زمان پاسخگویی را به ویژه برای بارهای کاری سنگین، بهبود می‌بخشد.

<sup>32</sup> Schema

<sup>33</sup> Normalization

<sup>34</sup> Indexing

<sup>35</sup> Denormalization

<sup>36</sup> Partitioning

<sup>37</sup> Query Execution Plan

<sup>38</sup> Caching mechanisms

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

- **تکنیک‌های مقیاس بندی افقی:** مقیاس بندی افقی شامل توزیع داده و حجم کار در چندین سرور پایگاه داده است. تکنیک‌هایی مانند خرد کردن و قسمت بندی به توزیع داده و پرس و جویا کمک می‌کند و در نتیجه ظرفیت سیستم را برای مدیریت بار تراکنش‌های بالاتر، افزایش می‌دهد. مقیاس افقی را می‌توان از طریق خوشه بندی پایگاه داده، متعادل سازی بار و یا استفاده از فناوری‌های پایگاه داده توزیع شده به دست آورد.
- **پایگاه‌های داده درون حافظه:** این پایگاه‌های داده، داده را به طور کامل در حافظه ذخیره می‌کنند و امکان خواندن و نوشتن بسیار سریع عملیات را فراهم می‌کنند. پایگاه‌های داده درون حافظه مخصوصاً برای برنامه‌هایی مناسب هستند که به تجزیه و تحلیل بلادرنگ یا دسترسی به داده با تأخیر کم نیاز دارند.

## نظارت منظم و تنظیم عملکرد

برای اطمینان از مقیاس پذیری و عملکرد بهینه، سازمان‌ها باید به طور مستمر پایگاه داده منبع باز خود را نظارت و تنظیم کنند. راهبردهای ذیل مثالی از این موارد می‌باشند [20], [9]:

- **نظارت بر عملکرد:** به طور منظم معیارهای عملکرد کلیدی مانند استفاده از پردازنده، استفاده از حافظه، ورودی و خروجی دیسک، زمان پاسخ پرس و جو و توان عملیاتی را باید مورد بررسی قرار داد. از ابزارهای نظارتی مانند پرومتهوس<sup>39</sup> و دیتادوگ<sup>40</sup> برای جمع آوری و تجزیه و تحلیل داده‌ی عملکرد می‌توان بهره گرفت. معیارهای خاص پایگاه داده و طرح‌های اجرای پرس و جو را برای شناسایی گلوگاه‌های عملکرد می‌بایست تحت نظارت و بررسی دقیق قرار داد.
- **بهینه سازی پیکربندی پایگاه داده:** بررسی و بهینه سازی پارامترهای پیکربندی پایگاه داده بر اساس ویژگی‌های بار کاری، تنظیم پارامترهایی مانند تخصیص حافظه، ادغام اتصالات و پیکربندی‌های حافظه پنهان می‌تواند عملکرد را تحت تأثیر قرار دهد. درک گزینه‌های پیکربندی خاص پایگاه داده و در نظر گرفتن بهترین شیوه‌های توصیه شده توسط جامعه پایگاه داده منبع باز ضروری است.
- **بهینه سازی پرس و جو:** تجزیه و تحلیل و بهینه سازی پرس و جوهای اغلب اجرا شده و منابع فشرده، شناسایی پرس و جوهای ناکارآمد و بازنویسی آنها، اضافه کردن فهرست‌های مناسب برای بهبود عملکرد ضروری هستند.
- **محک زدن<sup>41</sup> و تست بار:** برای شبیه سازی بارهای کاری واقع بینانه و شناسایی محدودیت‌های عملکرد، محک گذاری و تست بار را انجام دهید. با درک مرزهای عملکرد سیستم، سازمان‌ها می‌توانند به طور فعال برای مقیاس پذیری برنامه ریزی کنند و تصمیمات آگاهانه ای در مورد ارتقاء سخت افزار یا تغییرات معماری بگیرند.

## ۲.۳. امنیت و انطباق با استانداردها و مقررات

اطمینان از امنیت و انطباق پایگاه‌های داده منبع باز برای سازمان‌ها، به ویژه زمانی که با داده‌ی حساس سروکار دارند، از اهمیت بالایی برخوردار است. در حالی که پایگاه‌های داده منبع باز ویژگی‌های امنیتی قوی ارائه می‌دهند، آنها همچنین چالش‌های خاصی را ارائه می‌دهند که باید مورد توجه قرار گیرند. در این بخش، چالش‌های امنیت و انطباق را بررسی خواهیم کرد و به راهبردهای کاهش موثر آنها اشاره می‌شود. [22], [23]

### شیوه‌های کنترل دسترسی محدود

برخی از پایگاه‌های داده منبع باز ممکن است از نظر شیوه‌های کنترل دسترسی در مقایسه با همتایان اختصاصی خود محدودیت‌هایی داشته باشند. سازمان‌ها باید ویژگی‌های امنیتی موجود را به دقت ارزیابی کنند و اقدامات اضافی را برای تقویت کنترل‌های دسترسی در نظر بگیرند. برای این منظور، راهکارهای زیر را باید در نظر داشت [24]:



# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

- مکانیزم‌های احراز هویت و مجوز قوی: برای اطمینان از اینکه فقط کاربران مجاز می‌توانند به پایگاه داده دسترسی داشته باشند، شیوه‌های احراز هویت قوی، از جمله احراز هویت چند عاملی مورد نیاز هستند. از شیوه‌های کنترل دسترسی مبتنی بر نقش<sup>۴۲</sup> یا کنترل دسترسی مبتنی بر ویژگی<sup>۴۳</sup> برای اعمال سیاست‌های مجوز دقیق بر اساس نقش‌ها یا ویژگی‌های کاربر می‌توان استفاده کرد.
- سیاست‌های اتصال ایمن: باید اطمینان حاصل کرد که اتصالات پایگاه داده با استفاده از سیاست‌های مناسب ایمن سازی شده اند. رمز گذاری داده‌ی در حال انتقال، از آن‌ها در برابر افشا یا دسترسی غیرمجاز محافظت می‌کند.
- ویژگی‌های امنیتی در سطح پایگاه داده: ویژگی‌های امنیتی ارائه شده توسط پایگاه داده منبع باز، مانند مدیریت کاربر، افزایش اختیارات<sup>۴۴</sup> و ثبت لاگ<sup>۴۵</sup> برای رفع این چالش مناسب هستند. به طور منظم، باید حساب‌های کاربری را تحت کنترل و مدیریت قرار داد و اختیارات غیر ضروری را لغو کرد. ثبت لاگ برای نظارت و پیگیری فعالیت‌های پایگاه داده نیز ضروری می‌باشد.

## مدیریت وصله‌های امنیتی<sup>۴۶</sup> و آسیب پذیری‌ها

پایگاه‌های داده منبع باز نیاز به مدیریت فعال وصله‌های امنیتی و پاسخ مناسب به آسیب پذیری‌های یافت شده در سیستم، دارند. به روز ماندن با آخرین توصیه‌های امنیتی و اعمال سریع وصله‌ها برای کاهش خطرات احتمالی ضروری است. راهبردهای کاهش آسیب پذیری شامل موارد زیر می‌باشد [23]:

- ایجاد فرآیند مدیریت وصله‌های امنیتی: یک رویکرد سیستماتیک برای نظارت بر توصیه‌های امنیتی و اعمال سریع وصله‌ها، با تعریف یک خط مشی مدیریتی ایجاد می‌شود که مسئولیت‌ها، رویه‌ها و جدول زمانی لازم جهت به روز رسانی پایگاه داده و جلوگیری از آسیب پذیری‌های امنیتی را تعریف می‌کند. سیستم‌های مدیریت وصله خودکار می‌توانند آسیب پذیری‌ها را اسکن کنند، وصله‌ها را دانلود و اجرا کنند و گزارش‌هایی درباره وضعیت وصله ایجاد کنند.
- پایش آسیب پذیری‌ها و تست امنیت سیستم: نظارت منظم بر پایگاه‌های داده آسیب پذیری و اشتراک در فهرست‌های اطلاع رسانی امنیتی برای اطلاع از خطرات احتمالی، بسیار مهم است. انجام تست آسیب پذیری و تست نفوذ برای شناسایی و رفع نقاط ضعف امنیتی به طور فعال، برای هر سازمانی که قصد استفاده از پایگاه‌های داده منبع باز را دارد، ضروری است.

## رعایت مقررات صنعت

سازمان‌ها باید اطمینان حاصل کنند که پایگاه‌های داده منبع باز آنها با مقررات و استانداردهای خاص صنعت مانند مقررات عمومی حفاظت از داده<sup>۴۷</sup>، قانون قابلیت انتقال و مسئولیت بیمه سلامت<sup>۴۸</sup> یا استاندارد امنیت داده‌ی صنعت کارت پرداخت<sup>۴۹</sup> مطابقت دارد. الزامات انطباق ممکن است شامل رمز گذاری داده، ثبت لاگ‌ها، کنترل دسترسی و سیاست‌های حفظ داده باشد. راهبردهای مقابله با چالش‌ها به صورت زیر تعریف می‌گردند [25], [22]:

- رمز گذاری داده: استفاده از تکنیک‌های رمز گذاری مانند رمز گذاری شفاف داده<sup>۵۰</sup>، برای محافظت از داده‌ی ذخیره شده روی دیسک می‌تواند ریسک چالش‌های ذکر شده و افشای اطلاعات را کاهش دهد. رمز گذاری تضمین می‌کند که حتی اگر ذخیره سازی فیزیکی به خطر بیفتد، داده رمز گذاری شده و غیرقابل دسترسی باقی می‌ماند.
- ثبت لاگ و نظارت بر آن: ویژگی‌های گزارش و لاگ ارائه شده توسط پایگاه داده منبع باز را برای ردیابی و نظارت بر دسترسی به داده‌ی حساس، باید فعال نمود و به طور منظم لاگ‌ها را بررسی کرد تا هرگونه فعالیت مشکوک یا تلاش برای دسترسی غیرمجاز شناسایی و خنثی گردد.

<sup>42</sup> Role-Based Access Control

<sup>43</sup> Attribute-Based Access Control

<sup>44</sup> Privilege

<sup>45</sup> Log

<sup>46</sup> Security Patches

<sup>47</sup> General Data Protection Regulation

<sup>48</sup> Health Insurance Portability and Accountability Act

<sup>49</sup> Payment Card Industry Data Security Standard

<sup>50</sup> Transparent Data Encryption

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

- **سیاست‌های حفظ داده:** سیاست‌های حفظ داده از این جهت حائز اهمیت هستند که برای اطمینان از انطباق با الزامات نظارتی ایجاد شده اند. دستورالعمل‌هایی را برای مدت زمان نگهداری داده تعریف می‌کند. همچنین، باید شیوه‌هایی برای پاکسازی یا بایگانی خودکار داده پس از انقضای دوره نگهداری تعبیه شود.
- **ارزیابی منظم تطابق با مقررات:** انجام ارزیابی برای اطمینان از پایبندی به مقررات صنعت ضروری است و شرکت‌ها باید حساب رسان شخص ثالث یا کارشناسان امنیتی را برای انجام ارزیابی‌های مستقل و شناسایی هرگونه شکاف در رعایت مقررات در نظر بگیرند.
- **پیروی از شیوه‌های کد نویسی ایمن:** برنامه نویسان باید در مورد شیوه ی کد نویسی امن آموزش ببینند تا خطر تزریق پرس و جو، تزریق کد از طریق ویگاه<sup>51</sup> و سایر آسیب پذیری‌های رایج را کاهش دهند.

## ۳.۳. سطح پشتیبانی و تخصص جامعه

پایگاه‌های داده منبع باز در حمایت و مشارکت جامعه رشد می‌کنند. با این حال، سازمان‌ها ممکن است از نظر دسترسی به پشتیبانی قابل اعتماد و مشارکت فعال در جامعه با چالش‌هایی مواجه شوند. همانطور که در شکل ۲ - مقایسه اثر چالش‌ها در عدم به کار گیری پایگاه داده های منبع باز قابل مشاهده هست، بیشترین نگرانی موجود در مورد چالش‌های استفاده از پایگاه‌های داده منبع باز، مربوط به پشتیبانی است. [1] در این بخش، چالش‌های حمایت و مشارکت جامعه را بررسی خواهیم کرد [5].

## دسترسی به پشتیبانی قابل اعتماد

یکی از چالش‌هایی که سازمان‌ها ممکن است با پایگاه‌های داده منبع باز مواجه شوند، در دسترس بودن کانال‌های پشتیبانی قابل اعتماد است. بر خلاف پایگاه‌های داده اختصاصی که پشتیبانی اختصاصی از فروشنده ارائه می‌دهند، پایگاه‌های داده منبع باز عمدتاً به پشتیبانی جامعه محور متکی هستند. این موضوع می‌تواند به صورت همزمان، هم یک امتیاز و هم یک چالش محسوب شود. برای رفع مشکلات احتمالی ناشی از چنین چالش‌هایی، می‌توان راهبردهای زیر را به کار گرفت [26]:

- **استفاده از انجمن‌ها و لیست‌های پستی جامعه:** به انجمن‌ها و لیست‌های پستی اختصاصی پایگاه‌های داده منبع باز بپیوندید. با طرح سؤال، به اشتراک گذاری دانش و ارائه بازخورد، فعالانه شرکت کنید. تعامل با جامعه می‌تواند بینش‌های ارزشمندی را ارائه دهد تا افراد از تخصص جمعی استفاده کنند. افراد می‌توانند با گزارش اشکالات، ارائه وصله‌ها یا توسعه ویژگی‌های اضافی، فعالانه به جامعه پایگاه داده منبع باز کمک کنند. با مشارکت فعال در پروژه، سازمان‌ها می‌توانند یک جامعه حمایتی قوی ایجاد کنند و از پایداری بلند مدت پایگاه داده اطمینان حاصل کنند.
- **گزینه‌های پشتیبانی تجاری:** گزینه‌های پشتیبانی تجاری ارائه شده توسط فروشندگان یا سازمان‌های متخصص در پایگاه‌های داده منبع باز را در نظر بگیرید. پشتیبانی تجاری می‌تواند منابع اضافی، زمان پاسخگویی تضمین شده و تخصص برای استقرار حیاتی را فراهم کند. گزینه‌های موجود را بر اساس نیازها و الزامات خاص سازمان ارزیابی کنید.

## مستندات ناکافی یا قدیمی

اسناد و به اشتراک گذاری دانش نقش مهمی در زیست‌بوم منبع باز ایفا می‌کند. با این حال، سازمان‌ها ممکن است در مواجهه با مستندات قدیمی یا ناقص با چالش‌هایی مواجه شوند. برای عدم مواجهه با این مشکل موارد زیر را باید در نظر داشت:

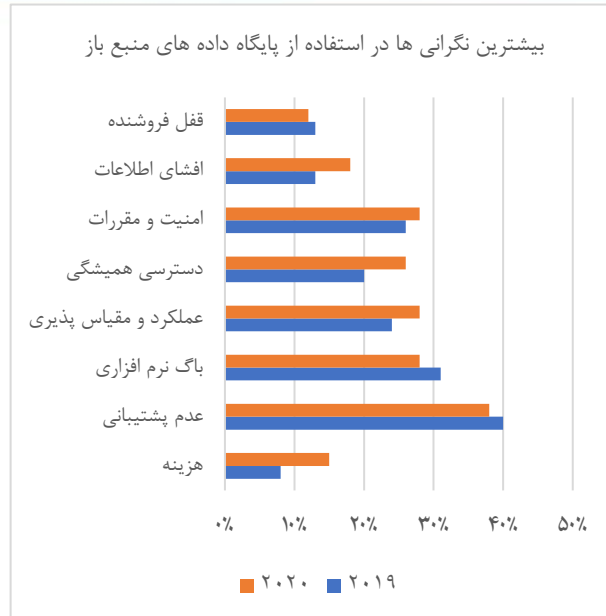
- **کمک به ساخت مستندات:** اگر سازمان‌ها شکاف‌ها یا اطلاعات قدیمی را در مستندات موجود شناسایی کنند، می‌توانند با ارائه بهبودها، شفاف‌سازی‌ها یا مثال‌های اضافی کمک کنند. با مشارکت فعال در بهبود اسناد، سازمان‌ها می‌توانند زیست‌بوم حمایتی جامعه را ارتقا دهند.

<sup>51</sup> Cross Site Scripting (XSS)

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir



شکل ۲ - مقایسه اثر چالش ها در عدم به کارگیری پایگاه داده های منبع باز. اطلاعات نمودار مبتنی بر مقالات سایت Percona می باشد.

- به اشتراک گذاری دانش و تجربیات: باید اعضای تیم را تشویق کرد تا دانش و تجربیات خود را از طریق وبلاگها، مقالات یا مطالعات موردی به اشتراک بگذارند. با به اشتراک گذاشتن بینشها و بهترین شیوهها، سازمانها به جامعه کمک می کنند و خود را به عنوان منابع قابل اعتماد تخصص معرفی می کنند.
- همکاری در آموزشها و راهنماها: برای ایجاد آموزشها، راهنماها یا منابع یادگیری که به افراد تازه وارد کمک می کند تا پایگاه داده منبع باز را به طور موثر درک کنند و از آنها استفاده کنند، با جامعه همکاری کنید. با به اشتراک گذاشتن دانش به شیوه ای ساختار یافته و در دسترس، سازمانها رشد جامعه را تقویت می کنند و امکان پذیرش هموار تر پایگاه داده را فراهم می کنند.

## ۴.۳. مهاجرت<sup>۵۲</sup> و سازگاری

مهاجرت به پایگاههای داده منبع باز و اطمینان از سازگاری با سیستمهای موجود می تواند چالشهای مهمی را برای سازمانها ایجاد کند. انتقال داده، سازگاری برنامهها و قفل شدن فروشنده برخی از ملاحظات کلیدی در این فرآیند هستند. در این بخش، چالشهای انتقال و سازگاری را بررسی خواهیم کرد و در مورد راهبردهایی برای کاهش موثر آنها بحث خواهیم کرد.

### انتقال یا مهاجرت داده

انتقال داده از پایگاههای داده موجود به پایگاههای داده منبع باز نیاز به برنامه ریزی و اجرای دقیق دارد. چالشهایی که سازمانها ممکن است در طول انتقال داده با آن مواجه شوند عبارتند از فرمت و ساختار داده، یکپارچگی و سازگاری داده، پنجره زمانی انتقال و مهاجرت، اعتبار سنجی و تایید داده که برای مقابله و رفع موارد ذکر شده، باید راهبردهای زیر را اتخاذ نمود [27]:

- برنامه ریزی و ارزیابی کامل: یک ارزیابی کامل از پایگاه داده موجود انجام دهید و یک برنامه مهاجرتی دقیق را تعریف کنید. چالشهای بالقوه، وابستگیهای داده و اصلاحات لازم را از قبل شناسایی کنید تا فرآیند مهاجرت روان را تضمین کنید.
- تحلیل و پاکسازی داده: برای به دست آوردن بینش در مورد ساختار و کیفیت داده، باید به تحلیل داده پرداخت. برای اطمینان از انتقال دقیق، پاکسازی و نرمال سازی داده و رفع هرگونه ناسازگاری یا مغایرت باید قبل از مهاجرت صورت پذیرد.

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

- **ابزارها و چارچوبهای انتقال داده:** از ابزارها و چارچوبهای انتقال داده که به طور خاص برای پایگاه داده منبع باز طراحی شده اند، استفاده کنید. این ابزارها می توانند فرآیند مهاجرت را خودکار کنند، تبدیل داده را مدیریت کنند و از یکپارچگی و سازگاری داده اطمینان حاصل کنند.
- **مهاجرت تدریجی:** به جای انتقال یکباره کل مجموعه داده، راهبردهای مهاجرت افزایشی را در نظر بگیرید. این رویکرد امکان همگام سازی مداوم داده، به حداقل رساندن خرابی و اطمینان از انتقال یکپارچه را فراهم می کند.

## سازگاری برنامه

اطمینان از سازگاری برنامه با پایگاههای داده منبع باز برای جلوگیری از اختلال و اطمینان از انتقال روان بسیار مهم است. مهم ترین چالشهای این بخش عبارتند از: تفاوت در گویشهای اس کیو ال<sup>53</sup> مخصوص پایگاه داده، سازگاری رابط برنامه نویسی کاربردی<sup>54</sup> و درایور<sup>55</sup>، مدیریت تراکنش و انطباق با خواص اسید، نیازمندی رویههای ذخیره شده و توابع به تغییر. راهبردهای لازم جهت مقابله بدین شرح است [9]:

- **استفاده از لایههای انتزاعی:** استفاده از لایههای انتزاعی یا چارچوبهای نگاشت شی - رابطه<sup>56</sup> را برای انتزاع عملیات پایگاه داده زیربنایی در نظر بگیرید. این چارچوبها یک رابط برنامه نویسی کاربردی یکپارچه ارائه می کنند و تفاوتهای خاص پایگاههای داده مختلف را در سطوح مختلف مدیریت می کنند و فرآیند مهاجرت را ساده می کنند.
- **بررسی و اصلاح کد:** یک بررسی جامع از کد برنامه انجام دهید و مناطقی را که برای اطمینان از سازگاری با پایگاه داده منبع باز هدف نیاز به اصلاح دارند، شناسایی کنید. پرس و جوها، تعاملات پایگاه داده و منطق مدیریت تراکنش را در صورت لزوم به روز کنید.
- **تست سازگاری پایگاه داده:** برای اطمینان از سازگاری برنامه با پایگاه داده منبع باز هدف، آزمایش دقیق انجام دهید. عملکردهای مختلف برنامه، پرس و جوها و رفتار تراکنشها را برای شناسایی و حل مشکلات سازگاری، قبل از استقرار در محیط کاربردی، آزمایش کنید.

## ۴. جمع بندی مقایسه جهت انتخاب بهترین پایگاه داده

بحث بین نرم افزار منبع باز و پایگاه داده اختصاصی در درک چشم انداز مدیریت داده بسیار مهم است. نرم افزار منبع باز، که آزادانه در دسترس و قابل تغییر است، مزیت سفارشی سازی و مقرون به صرفه بودن را فراهم می کند. این به توسعه دهندگان اجازه می دهد تا کد منبع را برای رفع نیازهای منحصر به فرد کسب و کار تنظیم کنند و انعطاف پذیری لازم برای استقرار پایگاههای داده در محیطهای مختلف را ارائه می دهد. با این حال، پایگاههای داده منبع باز ممکن است فاقد پشتیبانی در دسترس باشند و به تخصص اضافی یا کمک شخص ثالث نیاز داشته باشند، که ممکن است هزینههایی را به همراه داشته باشد. [4], [8], [19]

از سوی دیگر، نرم افزار پایگاه داده اختصاصی راه حل های متمرکز و پشتیبانی کارآمد را از یک فروشنده ارائه می دهد. این می تواند هنگام رسیدگی به نگرانیهای فوری پایگاه داده و ساده سازی محیط کلی مفید باشد. با این حال، نرم افزار اختصاصی می تواند گزینههای خلاقانه و مقیاس پذیری را محدود کند و ممکن است منجر به ایجاد قفل فروشنده شود. ارتقاها و تغییرات ممکن است بیشتر با علایق فروشنده هماهنگ باشد تا نیازهای در حال تحول مشتری. همچنین می تواند به شدت از بودجه فناوری مشتری استفاده کند. [5], [18], [19]

برای ایجاد تعادل بین مزایای منبع باز و پایگاههای داده اختصاصی، راه حل های منبع باز درجه بالای سازمانی وارد بازی می شوند. اینها مقرون به صرفه بودن و مقیاس پذیری نرم افزار منبع باز را با سادگی، انسجام و امنیت نرم افزار اختصاصی متمرکز بر وظیفه، ترکیب می کنند. با افزونههای مناسب، یک راه حل منبع باز می تواند برنامههای موجود را گسترش دهد و بارهای کاری حیاتی را مدیریت کند. انعطاف پذیری را ارائه می دهد و می تواند در محل، در فضای ابری یا در محیطهای ترکیبی کار کند. با این حال، حتی با نرم افزار درجه یک سازمانی، برخی از سطوح پشتیبانی ممکن است هنوز لازم باشد، که به فروشنده ای با پشتیبانی چند پایگاه داده، تخصص در زمینه فناوری و قراردادی انعطاف پذیر نیاز دارد. [10], [14]

<sup>53</sup> SQL Dialects

<sup>54</sup> Application Programming Interface

<sup>55</sup> Driver

<sup>56</sup> Object Relational Mapper

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senaconf.ir

هنگام انتخاب یک پایگاه داده منبع باز، چندین فاکتور باید در نظر گرفته شود. عملکرد بسیار مهم است و اطمینان حاصل می‌کند که پایگاه داده با نیازهای خاص سازمان، مانند پردازش تراکنش، تجزیه و تحلیل کسب و کار، یا ردیابی دارایی هماهنگ است. [4] مقیاس پذیری برای تطبیق با رشد داده، انتظارات مشتری و تغییر الزامات انطباق مهم است، در حالی که توانایی کاهش مقیاس در شرایط اقتصادی چالش برانگیز را نیز در نظر می‌گیرد. [3] پشتیبانی جامعه برای دسترسی به منابع، آموزش‌ها و انجمن‌ها، که به حفظ عملکرد بهینه و حل مسائل فنی کمک می‌کنند، ارزشمند است. ملاحظات هزینه شامل هزینه‌های پیاده سازی، نگهداری و پشتیبانی مرتبط با پایگاه داده منبع باز انتخاب شده است. در نهایت، ارزیابی مهارت‌ها و منابع فنی موجود در سازمان برای اطمینان از اجرای یکپارچه و فرآیند مدیریت مداوم حیاتی است. [19], [10], [3]

## ۵. نتیجه گیری

پایگاه‌های داده منبع باز ارائه دهنده مزایای گوناگونی مانند هزینه‌های کمتر، انعطاف پذیری و یک جامعه فعال هستند. با این حال، همراه با این مزایا، چالش‌هایی نیز وجود دارند. در این مقاله، چالش‌های اصلی مرتبط با پایگاه‌های داده‌ی منبع باز مورد بررسی قرار گرفته و راهبردهای مؤثر برای کاهش ریسک مواجهه با این چالش‌ها مورد بحث قرار گرفته است. درک و مدیریت این چالش‌ها، استفاده از سیستم‌های مدیریت پایگاه داده منبع باز را به عنوان جایگزینی مناسب برای پایگاه‌های داده تجاری و اختصاصی، ممکن می‌سازد. [18]

با برنامه ریزی دقیق، مشارکت فعال در جامعه کاربری و اجرای راهبردهای مؤثر، می‌توان این چالش‌ها را پشت سر گذاشت. سازمان‌هایی که از قدرت پایگاه‌های داده‌ی منبع باز استفاده می‌کنند و در عین حال به چالش‌های مطالعه شده در این مقاله توجه دارند، در موقعیت مناسبی قرار دارند تا از مزایای فناوری منبع باز بهره برداری کنند. پایگاه‌های داده منبع باز پایه‌ای قوی برای سازمان‌ها جهت ایجاد راه‌حل‌های مقیاس پذیر، قابل اعتماد و سودمند برای مدیریت داده، فراهم می‌کنند و با مواجهه مؤثر با چالش‌ها، سازمان‌ها می‌توانند پتانسیل کامل پایگاه‌های داده‌ی منبع باز را بهره برداری نمایند.

## ۶. مراجع

- [1] P. Scott, "The Ultimate Guide to Open Source Databases," *Percona Database Performance Blog*, Mar. 30, 2023. <https://www.percona.com/blog/ultimate-guide-open-source-databases/> (accessed Aug. 16, 2023).
- [2] L. F. Da Silva and J. V. F. Lima, "An evaluation of relational and NoSQL distributed databases on a low-power cluster," *J Supercomput.*, vol. 79, no. 12, pp. 13402–13420, Aug. 2023, doi: 10.1007/s11227-023-05166-7.
- [3] S. Butler *et al.*, "Considerations and challenges for the adoption of open source components in software-intensive businesses," *Journal of Systems and Software*, vol. 186, p. 111152, Apr. 2022, doi: 10.1016/j.jss.2021.111152.
- [4] W. Khan, T. Kumar, C. Zhang, K. Raj, A. M. Roy, and B. Luo, "SQL and NoSQL Database Software Architecture Performance Analysis and Assessments—A Systematic Literature Review," *BDCC*, vol. 7, no. 2, p. 97, May 2023, doi: 10.3390/bdcc7020097.
- [5] I. Skoulis, P. Vassiliadis, and A. V. Zarras, "Growing up with stability: How open-source relational databases evolve," *Information Systems*, vol. 53, pp. 363–385, Oct. 2015, doi: 10.1016/j.is.2015.03.009.
- [6] H. Sun, B. Xiao, X. Wang, and X. Liu, "Adaptive trade-off between consistency and performance in data replication: Adaptive trade-off between consistency and performance for data replication," *Softw. Pract. Exper.*, vol. 47, no. 6, pp. 891–906, Jun. 2017, doi: 10.1002/spe.2462.
- [7] M. J. Aqel, A. Al-Sakran, and M. Hunaity, "A Comparative Study Of NoSQL Databases," *Biosci. Biotech. Res. Comm*, vol. 12, no. 1, pp. 17–26, Feb. 2019, doi: 10.21786/bbr/12.1/3.
- [8] A. Petrov, *Database internals: a deep dive into how distributed data systems work*, First edition. Sebastopol, CA: O'Reilly Media, 2019.
- [9] C. Coronel and S. Morris, *Database systems: design, implementation, and management*, 11e [edition]. Australia ; United States: Course Technology Cengage Learning, 2015.
- [10] Y. Nah, B. Cui, S.-W. Lee, J. X. Yu, Y.-S. Moon, and S. E. Whang, Eds., *Database Systems for Advanced Applications: 25th International Conference, DASFAA 2020, Jeju, South Korea, September 24–27, 2020, Proceedings, Part I*, vol. 12112. in *Lecture Notes in Computer Science*, vol. 12112. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-59410-7.
- [11] A. Meier and M. Kaufmann, *SQL & NoSQL Databases: Models, Languages, Consistency Options and Architectures for Big Data Management*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019. doi: 10.1007/978-3-658-24549-8.

# دوازدهمین کنگره ملی سراسری فناوریهای نوین در حوزه توسعه پایدار ایران

12<sup>th</sup> National Congress of  
the New Technologies in Sustainable Development of Iran

senacnf.ir

- [12] "Oracle Database Documentation - Oracle Database." <https://docs.oracle.com/en/database/oracle/oracle-database/> (accessed Aug. 16, 2023).
- [13] "Apache Cassandra Documentation." <https://cassandra.apache.org/doc/latest/> (accessed Aug. 16, 2023).
- [14] "MySQL :: MySQL Documentation." <https://dev.mysql.com/doc/> (accessed Aug. 16, 2023).
- [15] "MongoDB Documentation." <https://www.mongodb.com/docs/> (accessed Aug. 16, 2023).
- [16] "Neo4j documentation - Neo4j Documentation." <https://neo4j.com/docs/> (accessed Aug. 16, 2023).
- [17] "DB-Engines Ranking Open Source vs. Commercial DBMS." [https://db-engines.com/en/ranking\\_osvsc](https://db-engines.com/en/ranking_osvsc) (accessed Aug. 16, 2023).
- [18] N. Purohit, "A Comparative Study on Open Source Database Management System," *IJRASET*, vol. 6, no. 4, pp. 792–795, Apr. 2018, doi: 10.22214/ijraset.2018.4133.
- [19] E. Law, "Open-Source Databases, Part III: Choosing a Database," 2007.
- [20] D. Colley and C. Stanier, "Identifying New Directions in Database Performance Tuning," *Procedia Computer Science*, vol. 121, pp. 260–265, 2017, doi: 10.1016/j.procs.2017.11.036.
- [21] K. Zhao, J. X. Yu, Y. Rong, M. Liao, and J. Huang, "Towards Expectation-Maximization by SQL in RDBMS," in *Database Systems for Advanced Applications*, C. S. Jensen, E.-P. Lim, D.-N. Yang, W.-C. Lee, V. S. Tseng, V. Kalogeraki, J.-W. Huang, and C.-Y. Shen, Eds., in *Lecture Notes in Computer Science*, vol. 12682. Cham: Springer International Publishing, 2021, pp. 778–794. doi: 10.1007/978-3-030-73197-7\_53.
- [22] S. M. Sterett, "Data Access as Regulation," *American Behavioral Scientist*, vol. 63, no. 5, pp. 622–642, May 2019, doi: 10.1177/0002764218797383.
- [23] Y. Wang, J. Xi, and T. Cheng, "The Overview of Database Security Threats' Solutions: Traditional and Machine Learning," *JIS*, vol. 12, no. 01, pp. 34–55, 2021, doi: 10.4236/jis.2021.121002.
- [24] I. Basharat, F. Azam, and A. Wahab Muzaffar, "Database Security and Encryption: A Survey Study," *IJCA*, vol. 47, no. 12, pp. 28–34, Jun. 2012, doi: 10.5120/7242-0218.
- [25] "(GDPR) – Official Legal Text," *General Data Protection Regulation (GDPR)*. <https://gdpr-info.eu/> (accessed Aug. 19, 2023).
- [26] D. Nagy, A. M. Yassin, and A. Bhattacharjee, "Organizational adoption of open source software: barriers and remedies," *Commun. ACM*, vol. 53, no. 3, pp. 148–151, Mar. 2010, doi: 10.1145/1666420.1666457.
- [27] T. Bircan *et al.*, "Gaps in Migration Research. Review of migration theories and the quality and compatibility of migration data on the national and international level," pp. 1–104, 2020.