

چالش‌ها و راهکارهای مدیریت داده در سیستم‌های توزیع‌شده

گزارشی از مقالات:

Data Management in Microservices: State of the Practice, Challenges, and Research Directions

و

Data Management in Industry: Concepts, Systematic Review and Future Directions

دانشجو: علی باقری

استاد راهنما: دکتر لطفی

دانشگاه تربیت مدرس

۱. مقدمه

تکامل سریع معماری‌های نرم‌افزاری و سیستم‌های صنعتی، نقش مدیریت داده را به عنوان یکی از عوامل کلیدی در توسعه برنامه‌ها و خدمات مدرن ارتقا داده است. دو حوزه متمایز اما به هم پیوسته (توسعه نرم‌افزار مبتنی بر میکروسرویس‌ها و سیستم‌های تولید صنعتی) با چالش‌های قابل توجهی در مدیریت و بهره‌برداری مؤثر از داده‌ها در محیط‌های فزاینده پیچیده روبرو هستند.

مقاله نخست با عنوان «مدیریت داده در میکروسرویس‌ها: وضعیت فعلی، چالش‌ها و مسیرهای پژوهشی»¹، به بررسی پارادایم نوظهور معماری‌های میکروسرویسی می‌پردازد. این مقاله توضیح می‌دهد که چگونه میکروسرویس‌ها برنامه‌های یکپارچه را به واحدهای کوچک‌تر و مستقل تقسیم می‌کنند که هر یک می‌توانند پایگاه داده مختص خود را داشته باشند. اگرچه این سبک معماری، مقیاس‌پذیری، ایزوله‌سازی خطا و چابکی را نوید می‌دهد، اما چالش‌های قابل توجهی نیز در حفظ انسجام داده‌ها، مدیریت تراکنش‌های توزیع‌شده و هماهنگی بین سرویس‌هایی با نیازهای متنوع مدیریت داده ایجاد می‌کند.

مقاله دوم با عنوان «مدیریت داده در صنعت: مفاهیم، مرور نظام‌مند و مسیرهای آینده»²، به بررسی شیوه‌های مدیریت داده در محیط‌های صنعتی می‌پردازد. با گسترش فناوری‌های صنعت ۴.۰ و رشد نمایی داده‌های تولید شده در خطوط تولید و فرآیندهای صنعتی، سازمان‌ها با فشار فزاینده‌ای برای پیاده‌سازی خطوط پردازش داده مقیاس‌پذیر و مؤثر مواجه هستند. این مقاله مرور نظام‌مندی از روش‌های موجود مدیریت داده صنعتی ارائه می‌دهد و مفاهیم رایجی مانند کلان‌داده، خطوط پردازش داده، ذخیره‌سازی ابری و چارچوب‌های پیش‌پردازش را برجسته می‌کند، در حالی که به فقدان روش‌های استاندارد در صنایع مختلف نیز اشاره دارد.

این دو مطالعه در کنار هم، دیدگاه‌های ارزشمندی درباره وضعیت فعلی مدیریت داده توزیع‌شده ارائه می‌کنند: یکی با تمرکز بر اکوسیستم‌های نرم‌افزاری مبتنی بر میکروسرویس و دیگری بر خطوط پردازش داده در بخش صنعتی. با تحلیل همزمان این یافته‌ها، می‌توانیم درک بهتری از چالش‌ها و فرصت‌های موازی در محیط‌های دیجیتال و فیزیکی کسب کنیم و روشن سازیم که چگونه سازمان‌ها می‌توانند راهبردهای مدیریت داده خود را برای آینده بهینه کنند.

۲. اهداف

هر دو مقاله به دنبال پاسخ به پرسش‌های حیاتی مربوط به مدیریت داده در سیستم‌های توزیع‌شده هستند، اگرچه هر یک بر حوزه‌های کاربردی متفاوتی تمرکز دارند. با وجود این تفاوت، هر دو مقاله هدف مشترکی را دنبال می‌کنند: شناسایی شیوه‌های فعلی، چالش‌ها و مسیرهای آینده مدیریت داده در چارچوب محدودیت‌های معماری و صنعتی مدرن.

هدف اصلی مقاله Laigner و همکاران، بررسی وضعیت مدیریت داده در معماری‌های میکروسرویسی است. اگرچه میکروسرویس‌ها به دلیل مزایایی همچون مقیاس‌پذیری، ایزوله‌سازی خطا و چابکی به طور گسترده در صنعت نرم‌افزار پذیرفته شده‌اند، اما بعد مدیریت داده در این پارادایم هنوز به اندازه کافی مورد مطالعه قرار نگرفته است. این مقاله با انجام یک مطالعه جامع که شامل مرور نظام‌مند ادبیات، تحلیل برنامه‌های میکروسرویسی متن‌باز³ و نظرسنجی در سطح صنعت است، تلاش می‌کند این خلأ را پر کند. هدف نهایی مقاله، روشن ساختن این است که متخصصان چگونه داده‌ها را در میکروسرویس‌ها مدیریت می‌کنند، از چه الگوها و فناوری‌هایی استفاده می‌کنند و در عمل با چه چالش‌های سیستمی روبرو می‌شوند. نویسندگان همچنین به دنبال ارائه نیازمندی‌هایی برای سیستم‌های مدیریت پایگاه داده⁴ مختص میکروسرویس‌ها هستند که بتوانند این مشکلات را مؤثرتر برطرف کنند.

¹ "Data Management in Microservices: State of the Practice, Challenges, and Research Directions" by Laigner et al.

² "Data Management in Industry: Concepts, Systematic Review and Future Directions" by Freitas et al.

³ Open-Source

⁴ DBMS

از سوی دیگر، مقاله Freitas و همکاران بر بخش صنعتی تمرکز دارد، جایی که مدیریت حجم بالای داده‌های ناهمگن برای تضمین کارایی عملیاتی و تحقق اهداف صنعت ۴۰۰ حیاتی است. این مقاله دو هدف اصلی دارد: (۱) مرور و جمع‌آوری دانش موجود درباره مفاهیم، تکنیک‌ها و معماری‌های مدیریت داده در محیط‌های صنعتی و (۲) انجام یک نظرسنجی نظام‌مند از نشریات علمی برجسته به منظور شناسایی رویکردهای رایج، شکاف‌ها و روندها. این مطالعه بر نیاز به روش‌های مدیریت داده مشخص و استاندارد، به ویژه در زمینه خطوط پردازش داده، مدیریت کلان‌داده و پردازش بلادرنگ داده تأکید دارد.

به طور خلاصه، هر دو مقاله به دنبال ارائه درکی روشن‌تر از نحوه مدیریت داده در سیستم‌های توزیع‌شده امروزی هستند (چه در اکوسیستم‌های نرم‌افزاری مبتنی بر میکروسرویس و چه در صنایع تولیدی) و پیشنهاد مسیرهای پژوهشی و نوآوری‌های فنی جدید که می‌توانند در آینده بهتر به این حوزه‌ها خدمت کنند.

۳. روش‌شناسی

هر دو مقاله از روش‌های پژوهشی نظام‌مند و مبتنی بر شواهد برای بررسی شیوه‌های مدیریت داده در حوزه‌های تخصصی خود استفاده کرده‌اند. اگرچه رویکردهای آن‌ها بر اساس زمینه‌های کاربردی‌شان اندکی متفاوت است، اما هر دو ترکیبی از مرور ادبیات و تحلیل تجربی را برای استخراج بینش‌ها به کار گرفته‌اند.

مقاله اول (با تمرکز بر میکروسرویس‌ها):

در مقاله اول از رویکردی ترکیبی متشکل از سه مرحله کلیدی برای درک شیوه‌های مدیریت داده در معماری‌های میکروسرویسی استفاده شده است:

- **مرور نظام‌مند ادبیات^۵:** نویسندگان یک مرور ساختاریافته از بیش از ۳۰۰ مقاله داور شده مرتبط با پذیرش و شیوه‌های میکروسرویس انجام دادند. از این مجموعه، ۱۰ مطالعه نماینده برای بررسی عمیق‌تر انتخاب شد تا پایه‌ای برای درک روندهای رایج، الگوها و چالش‌های گزارش‌شده در مدیریت داده در سیستم‌های مبتنی بر میکروسرویس فراهم شود.
- **تحلیل برنامه‌های متن‌باز:** نویسندگان ۹ برنامه متن‌باز پرطرفدار میکروسرویسی را از میان بیش از ۲۰ پروژه انتخاب و بررسی کردند. با مطالعه کدبیس‌ها و الگوهای معماری، آن‌ها ارزیابی کردند که در عمل چگونه نگرانی‌های مربوط به مدیریت داده مانند انتخاب پایگاه داده، مدیریت تراکنش‌ها و ارتباط بین سرویس‌ها رفع می‌شود.
- **نظرسنجی صنعتی:** برای اعتبارسنجی یافته‌ها، نویسندگان یک نظرسنجی آنلاین طراحی و توزیع کردند که توسعه‌دهندگان، معماران و پژوهشگران با تجربه در سیستم‌های میکروسرویسی را هدف قرار می‌داد. بیش از ۱۲۰ متخصص پاسخ‌های دقیقی ارائه دادند که بینش‌های عملی درباره استراتژی‌های مدیریت داده در دنیای واقعی، چالش‌ها و فناوری‌های مورد استفاده در میکروسرویس‌ها فراهم کرد.

این روش سه‌گانه به نویسندگان اجازه داد تا دانش دانشگاهی، جزئیات پیاده‌سازی عملی و بازخوردهای صنعتی را با هم ترکیب کرده و دیدگاه جامعی از شیوه‌های مدیریت داده در میکروسرویس‌ها ارائه دهند.

مقاله دوم (با تمرکز بر مدیریت داده صنعتی)

در این مقاله یک مرور نظام‌مند جامع (SLR) برای بررسی شیوه‌های مدیریت داده در محیط‌های صنعتی گوناگون انجام شده است:

⁵ Systematic Literature Review (SLR)

- **جستجوی اولیه کلمات کلیدی:** نویسندگان با یک جستجوی اکتشافی با استفاده از Google Scholar برای ترسیم چشم‌انداز کلی مدیریت داده در زمینه‌های صنعتی آغاز کردند. کلمات کلیدی مرتبط با استخراج، ذخیره‌سازی، تحلیل و خطوط پردازش داده در محیط‌های تولیدی مورد استفاده قرار گرفت.
- **مرور نظام‌مند ادبیات (SLR):** نویسندگان SLR متمرکزی را با استفاده از سه پایگاه داده معتبر شامل Web of Science، SCOPUS و IEEE Xplore انجام دادند. این جستجو بر مقالات منتشر شده در مجلات Q1 حوزه مهندسی و علوم کامپیوتر بین سال‌های ۲۰۱۹ تا ۲۰۲۳ متمرکز بود. فرآیند جستجو بیش از ۱۱,۰۰۰ مقاله را شناسایی کرد که پس از اعمال معیارهای سخت‌گیرانه برای گنجاندن یا حذف (مانند تمرکز بر مقالاتی که "خطوط پردازش داده"⁶، "مدیریت داده"⁷ یا "تولید صنعتی"⁸ را در عنوان یا چکیده ذکر کرده بودند) به ۷۶ مطالعه با کیفیت بالا محدود شد.
- **تحلیل موضوعی:** مقالات منتخب برای استخراج الگوهای تکرارشونده، رویکردهای رایج مدیریت داده و الگوهای معماری مورد تحلیل قرار گرفتند. تمرکز ویژه‌ای بر نحوه برخورد صنایع با کلان‌داده، ناهمگونی داده‌ها، پردازش بلادرنگ و رایانش ابری در فرآیندهای مدیریت داده آن‌ها اعمال شد.

هر دو مطالعه از روش‌شناسی‌های ساختاریافته متناسب با حوزه‌های تخصصی خود استفاده کردند تا اطمینان حاصل شود که یافته‌های آن‌ها هم در پژوهش‌های دانشگاهی و هم در پیاده‌سازی‌های عملی ریشه دارند.

۴. مفاهیم و موضوعات کلیدی

اگرچه هر دو مقاله بر حوزه‌های متفاوتی تمرکز دارند (به ترتیب میکروسرویس‌ها و مدیریت داده صنعتی) اما هر دو مقاله چالش‌ها و شیوه‌های اصلی مرتبط با پردازش داده توزیع‌شده و طراحی سیستم را شناسایی می‌کنند. در ادامه مفاهیم و موضوعات اصلی مورد بررسی هر مقاله ارائه شده است.

مقاله اول (میکروسرویس‌ها)

1. **انگیزه‌ها برای پذیرش میکروسرویس‌ها:** مقاله به چند نیروی محرک اصلی پشت تغییر از معماری‌های یکپارچه به معماری‌های میکروسرویسی اشاره می‌کند. مهم‌ترین انگیزه‌ها شامل موارد زیر است:

- مقیاس‌پذیری از طریق تجزیه عملکردی.
- جداسازی خطا برای افزایش تاب‌آوری سیستم.
- پشتیبانی از تکامل طرح و چابکی با استفاده از مدیریت داده غیرمتمرکز.
- پذیرش معماری‌های مبتنی بر رویداد برای ایجاد سرویس‌های loosely coupled.
- پایداری چندزبانه (Polyglot Persistence) که در آن میکروسرویس‌ها از انواع مختلف پایگاه داده (مانند رابطه‌ای و NoSQL) متناسب با بار کاری خاص خود استفاده می‌کنند.

2. **الگوهای مدیریت داده:** مطالعه سه الگوی رایج مدیریت داده را شناسایی می‌کند:

⁶ Data Pipeline

⁷ Data Management

⁸ Manufacturing

- جداول خصوصی برای هر میکروسرویس درون یک پایگاه داده مشترک.
 - طرح اختصاصی برای هر میکروسرویس در یک پایگاه داده مشترک.
 - پایگاه داده مجزا برای هر میکروسرویس که رایج‌ترین الگو در عمل است و اغلب در محیط‌های کانتینری پیاده‌سازی می‌شود.
3. مکانیسم‌های هماهنگی: مقاله الگوهای هماهنگی بین میکروسرویس‌ها را بررسی می‌کند، مانند:
- کُرئوگرافی⁹ (ارتباط غیرهمزمان مبتنی بر رویداد).
 - ارکستراسیون¹⁰ (جریان کنترل متمرکز با استفاده از یک هماهنگ‌کننده یا سرویس مرکزی).
 - ساگاها¹¹ (چه به صورت متمرکز و چه غیرمتمرکز). نویسندگان اشاره می‌کنند که صنعت بیشتر به الگوهای شبیه ارکستراسیون تمایل دارد، در حالی که پروژه‌های متن‌باز بیشتر کُرئوگرافی‌های مبتنی بر رویداد را ترجیح می‌دهند.
4. چالش‌های مربوط به انسجام داده: مقاله اول بر مصالحه میان مقیاس‌پذیری و انسجام داده تأکید می‌کند. مدل‌های انسجام ضعیف (مانند انسجام نهایی) غالب هستند، زیرا تراکنش‌های توزیع‌شده (مانند 2PC) به دلیل پیچیدگی کمتر استفاده می‌شوند.
5. پیچیدگی سیستم و بار اضافی در لایه برنامه: توسعه‌دهندگان معمولاً به دلیل نبود پشتیبانی ذاتی از سوی DBMS‌ها برای نیازهای خاص میکروسرویس‌ها، مقدار قابل‌توجهی منطق سفارشی را در لایه برنامه پیاده‌سازی می‌کنند (مانند مدیریت انسجام نهایی و جریان‌های کاری توزیع‌شده).
- مقاله دوم (مدیریت داده صنعتی)**
1. چرخه عمر داده صنعتی: مقاله چهار مرحله کلیدی در خطوط پردازش داده صنعتی را تعریف می‌کند:
- استخراج: جمع‌آوری داده از منابع ناهمگن، معمولاً از طریق سنسورها، دستگاه‌های IoT یا PLC‌ها.
 - پیش‌پردازش: پاک‌سازی، نرمال‌سازی و جایگزینی داده‌ها برای بهبود کیفیت داده.
 - ذخیره‌سازی: انتخاب بین ذخیره‌سازی محلی یا ابری و اتخاذ معماری‌هایی مانند دریاچه داده (Data Lake)، انبار داده (Data Warehouse) و Lakehouse.
 - پردازش: استفاده از تکنیک‌های پردازش دسته‌ای یا بلادرنگ (جریانی) بسته به نوع کاربرد.
2. ویژگی‌های کلان‌داده (مدل 7Vs): در مقاله دوم با اشاره به هفت ویژگی اصلی، دید گسترده‌تری از چالش‌های کلان‌داده در صنعت ارائه می‌دهند:

⁹ Choreography

¹⁰ Orchestration

¹¹ Sagas

- حجم (Volume)، سرعت (Velocity)، تنوع (Variety)، صحت (Veracity)، تغییرپذیری (Variability)، مصورسازی (Visualization) و ارزش (Value). این ابعاد به چارچوب بندی پیچیدگی مدیریت داده های صنعتی کمک می کنند.

3. خطوط پردازش ETL در مقابل ELT: نویسندگان به بررسی مزایا و معایب دو الگوی ETL¹² و ELT¹³ می پردازند:

- ETL برای خطوط پردازش دسته ای سنتی با خروجی های ساختار یافته مناسب است.
- ELT در معماری های مدرن، به ویژه جایی که دریاچه های داده و رویکردهای schema-on-read رایج هستند، ترجیح داده می شود.

4. زمینه Industry 4.0: مدیریت داده در بستر تحول گسترده صنعت با مفاهیمی مانند سیستم های سایبر-فیزیکی و اینترنت اشیا صنعتی (IIoT) بررسی می شود.

5. میان افزار و بروکرهای پیام: یک الگوی تکرار شونده استفاده از میان افزارها یا بروکرهای پیام (مانند Kafka و RabbitMQ) برای ورود بلا درنگ و یکپارچه سازی جریان های داده از سیستم های صنعتی مختلف است.

موضوعات مشترک

با وجود تفاوت در حوزه ها، هر دو مقاله الگوهای مشترکی را شناسایی می کنند:

- نیاز به معماری های توزیع شده و مقیاس پذیر.
- وابستگی روز افزون به ارتباطات غیر همزمان و مبتنی بر رویداد.
- چالش های مرتبط با ناهمگونی و انسجام داده ها.
- نقش فزاینده راهکارهای سفارشی و مهندسی موقتی برای پر کردن شکاف های موجود در فناوری های فعلی.

۵. یافته ها

هر دو مقاله به یافته های مهمی دست یافته اند که نشان می دهد چگونه سیستم های توزیع شده مدرن (چه در قالب برنامه های مبتنی بر میکروسرویس و چه در محیط های صنعتی) در شرایط واقعی با چالش های مدیریت داده روبه رو هستند.

مقاله اول (میکروسرویس ها)

1. میکروسرویس ها به صورت ذاتی داده ها را در سیلوهای جداگانه نگهداری می کنند: بیشتر سازمان ها الگوی "یک پایگاه داده برای هر میکروسرویس" را برای اطمینان از جداسازی خطا، مقیاس پذیری و چرخه های استقرار مستقل پیاده سازی می کنند. با این حال، این امر منجر به افزایش تعداد پایگاه های داده مستقل با مدل های مختلف (رابطه ای، مبتنی بر سند و غیره) می شود که پیچیدگی عملیاتی را بالا می برد.

¹² Extract-Transform-Load

¹³ Extract-Load-Transform

2. **ناهماهنگی در تراکنش‌های بین سرویس‌ها:** یافته‌ها نشان می‌دهد که تراکنش‌های میان میکروسرویس‌ها رایج هستند، اما توسعه‌دهندگان عمدتاً از پروتکل‌های commit توزیع‌شده (مانند 2PC) به دلیل پیچیدگی و نبود انتراعات کاربردی پرهیز می‌کنند. در عوض، الگوهای هماهنگی غیرهمزمان و مبتنی بر رویداد (مانند ساگاها و کرنوگرافی‌ها) به طور گسترده مورد استفاده قرار می‌گیرند.
3. **وابستگی به راهکارهای موقتی:** توسعه‌دهندگان اغلب ابزارهای ناهمگن (بروکرهای پیام، کش‌ها، موتورهای تحلیلی و چندین DBMS) را به صورت موقتی و موقتی برای برآوردن نیازهای مدیریت داده میکروسرویس‌ها ادغام می‌کنند. این موضوع منجر به انتقال منطق بیشتر به لایه برنامه می‌شود، که سربار نگهداری را افزایش داده و قابلیت مشاهده (observability) سیستم را کاهش می‌دهد.
4. **فقدان پشتیبانی DBMS برای میکروسرویس‌ها:** پایگاه‌های داده پیشرفته فعلی در پاسخ به چالش‌هایی مانند مدیریت محدودیت‌های توزیع‌شده، تضمین انسجام میان میکروسرویس‌ها و پشتیبانی از سرویس‌های مستقل اما هماهنگ، ناکارآمد هستند. در نتیجه، سیستم‌های پایگاه داده عمدتاً نقش ذخیره‌سازی صرف را ایفا می‌کنند و در فراهم کردن ایمنی و هماهنگی سطح سیستم ضعیف عمل می‌کنند.
5. **نیازهای نوظهور برای DBMS‌های مناسب میکروسرویس‌ها:** مطالعه نتیجه می‌گیرد که نسل بعدی DBMS‌ها باید برای میکروسرویس‌ها طراحی شوند و پشتیبانی ذاتی از انسجام توزیع‌شده، تراکنش‌های غیرمتمرکز و جریان‌های کاری مبتنی بر رویداد را ارائه دهند.

مقاله دوم (مدیریت داده صنعتی)

1. **فقدان استانداردسازی در صنایع:** این بررسی نشان می‌دهد که اگرچه بیشتر بخش‌های صنعتی خطوط پردازش داده مدرن را پذیرفته‌اند، اما هیچ روش‌شناسی یکسانی در میان صنایع یا حتی فرایندهای مشابه وجود ندارد. رویکردها به‌طور قابل‌توجهی بسته به اندازه سازمان، بخش و سطح بلوغ متفاوت هستند.
2. **رواج معماری‌های ترکیبی داده:** اکثر شرکت‌ها به محیط‌های ترکیبی متکی هستند که ذخیره‌سازی محلی (مانند سرورهای محلی، دستگاه‌های edge) را با خدمات ابری ترکیب می‌کنند. مصالحه میان تأخیر، مقیاس‌پذیری و حاکمیت داده از ملاحظات کلیدی تأثیرگذار بر تصمیمات ذخیره‌سازی است.
3. **چالش‌های مدیریت کلان‌داده در صنعت:** ناهمگونی داده‌ها، جریان‌های پر حجم داده و نیاز به تحلیل‌های بلادرنگ از چالش‌های اصلی هستند. بسیاری از صنایع در یکپارچه‌سازی داده‌های ساختاریافته و غیرساختاریافته در عین حفظ عملکرد و قابلیت اطمینان دچار مشکل هستند.
4. **استفاده گسترده از میان‌افزارها و بروکرهای پیام:** مقاله نشان می‌دهد که پلتفرم‌های میان‌افزار (مانند سرورهای OPC-UA و بروکرهای MQTT) ابزارهای رایجی برای جمع‌آوری و یکپارچه‌سازی بلادرنگ داده‌ها هستند، به‌ویژه در اکوسیستم‌های IIoT.
5. **روندهای ETL در مقابل ELT:** اگرچه ETL همچنان الگوی غالب برای پردازش دسته‌ای سنتی است، اما با ظهور دریاچه‌های داده و نیاز به انعطاف‌پذیری، پذیرش مدل ELT در حال افزایش است، به‌خصوص هنگام کار با مجموعه داده‌های بزرگ و در حال تغییر.

بینش‌های کلی

- هر دو حوزه با چالش‌های مربوط به پیچیدگی سیستم و فناوری‌های ناهمگون روبه‌رو هستند.
- هر دو مقاله به شکاف میان فناوری‌های پایگاه داده/ذخیره‌سازی موجود و معماری‌های توزیع‌شده مدرن اشاره می‌کنند که منجر به وابستگی به راهکارهای سفارشی یا موقتی می‌شود.

- روند رو به رشد به سمت سیستم‌های مبتنی بر رویداد در هر دو محیط میکروسرویسی و صنعتی مشاهده می‌شود، هرچند با الگوهای پیاده‌سازی خاص هر حوزه.

۶. چالش‌های شناسایی شده

هر دو مقاله به چندین چالش فنی و عملیاتی اشاره می‌کنند که به‌طور ذاتی در مدیریت داده‌ها در سیستم‌های توزیع‌شده وجود دارد. اگرچه زمینه‌ها متفاوت است (یکی بر معماری میکروسرویس‌ها تمرکز دارد و دیگری بر کاربردهای صنعتی) اما چالش‌ها در موضوعاتی مانند پیچیدگی، انسجام و مقیاس‌پذیری شباهت‌های بنیادی دارند.

مقاله اول (میکروسرویس‌ها)

1. **انسجام داده میان سرویس‌ها:** تضمین انسجام میان چندین میکروسرویس مستقل همچنان یک مانع جدی است. به دلیل وضعیت توزیع‌شده و سرویس‌های loosely coupled، دستیابی به انسجام قوی اغلب غیر عملی است و توسعه‌دهندگان را به سمت مدل‌های انسجام نهایی (Eventual Consistency) سوق می‌دهد.
2. **نبود پشتیبانی از تراکنش‌های توزیع‌شده:** توسعه‌دهندگان از پروتکل‌های commit توزیع‌شده (مانند 2PC) اجتناب می‌کنند، زیرا این پروتکل‌ها پیچیده و کند تلقی می‌شوند. این موضوع تیم‌ها را مجبور می‌کند که تراکنش‌های جبرانی یا الگوهای ساگا را پیاده‌سازی کنند که پیچیدگی منطق برنامه را افزایش می‌دهد.
3. **محدودیت DBMS‌های موجود:** پایگاه‌های داده سنتی با نیازهای خاص میکروسرویس‌ها طراحی نشده‌اند. ویژگی‌هایی مانند پشتیبانی بومی از هماهنگی توزیع‌شده، معناشناسی تکثیر (replication semantics) و پردازش رویداد غیرهمزمان در این سیستم‌ها غایب است و توسعه‌دهندگان را به استفاده از ابزارهای خارجی (مانند بروکرهای پیام، کش‌ها) و کدنویسی سفارشی مجبور می‌کند.
4. **وابستگی میان سرویس‌ها و افزایش coupling:** درحالی‌که میکروسرویس‌ها برای خودمختاری طراحی شده‌اند، مطالعه نشان می‌دهد که بسیاری از سیستم‌ها از طریق جریان‌های کاری مشترک و سازوکارهای هماهنگی، وابستگی‌های پنهانی ایجاد می‌کنند که اهداف اولیه ماژولار بودن و استقرار مستقل را تضعیف می‌کند.
5. **پیچیدگی عملیاتی:** سیستم‌های میکروسرویس اغلب پشته‌های ناهمگن (SQL + NoSQL + کش + موتورهای تحلیلی) را به شکلی غیر استاندارد به هم متصل می‌کنند. این موضوع باعث افزایش پیچیدگی در استقرار، مانیتورینگ و نگهداری می‌شود.

مقاله دوم (مدیریت داده صنعتی)

1. **ناهمگونی داده:** محیط‌های صنعتی با داده‌های بسیار ناهمگن از سنسورها، ماشین‌ها، اپراتورهای انسانی و سیستم‌های سازمانی مواجه هستند. یکپارچه‌سازی منابع داده ساختاریافته، نیمه‌ساختاریافته و غیرساختاریافته چالش فنی بزرگی است.
2. **پردازش داده بلادرنگ:** بسیاری از فرآیندهای صنعتی نیاز به تصمیم‌گیری بلادرنگ یا نزدیک به بلادرنگ دارند، اما دستیابی به خطوط پردازش با تأخیر کم دشوار است، به‌ویژه زمانی که داده‌ها از مکان‌های غیرمتمرکز مختلف تأمین می‌شوند.
3. **روش‌شناسی‌های پراکنده:** اجماعی در صنعت برای ساختاربندی خطوط داده، معماری‌های ذخیره‌سازی یا راهبردهای پیش‌پردازش وجود ندارد. این پراکندگی منجر به ناکارآمدی و تلاش‌های مهندسی تکراری می‌شود.

4. **مقیاس‌پذیری در برابر محدودیت‌های محلی:** اگرچه خدمات ابری قابلیت‌های ذخیره‌سازی و پردازش مقیاس‌پذیر ارائه می‌دهند، اما صنایع اغلب با محدودیت‌هایی مانند نگرانی‌های مربوط به حاکمیت داده، تأخیر شبکه یا نیاز به رعایت مقررات خاص مواجه هستند که چالش‌های استقرار ترکیبی (محلی/ابری) را ایجاد می‌کند.

5. **امنیت و حاکمیت داده:** خطوط پردازش داده صنعتی اغلب بهترین شیوه‌های امنیت و حاکمیت داده را نادیده می‌گیرند، به‌ویژه هنگام کار با سیستم‌های ابری یا زیرساخت‌های IoT توزیع‌شده، که می‌تواند منجر به آسیب‌پذیری‌های بالقوه شود.

چالش‌های مشترک در هر دو مطالعه

- **پیچیدگی در سیستم‌های توزیع‌شده:** هر دو مقاله به پیچیدگی ذاتی مدیریت سیستم‌های داده توزیع‌شده با اجزا و فناوری‌های مختلف اشاره می‌کنند.
- **مدل‌های انسجام ضعیف:** هر دو حوزه به دلیل دغدغه‌های مربوط به مقیاس‌پذیری و خودمختاری، مدل‌های انسجام نهایی و ارتباطات غیرهمزمان را به مدل‌های قوی ترجیح می‌دهند.
- **فقدان ابزارها یا چارچوب‌های استاندارد:** هر دو حوزه میکروسرویس و صنعت از نبود راهکارهای جامع و آماده برای نیازهای خاص مدیریت داده خود رنج می‌برند.
- **بار عملیاتی بالا:** تیم‌ها در هر دو زمینه معمولاً بار اضافی طراحی و نگهداری منطق‌های مدیریت داده موقتی یا سفارشی را به دوش می‌کشند.

۷. توصیه‌ها و مسیرهای آینده

هر دو مقاله با ارائه توصیه‌های عملی و شناسایی مسیرهای پژوهشی آینده به پایان می‌رسند تا به شکاف‌ها و چالش‌های مشاهده‌شده در حوزه‌های مربوطه پاسخ دهند. اگرچه این توصیه‌ها به‌طور خاص برای محیط‌های میکروسرویس و صنعتی مطرح شده‌اند، اما هر دو بر بهبود پشتیبانی در سطح سیستم، کاهش پیچیدگی و افزایش انعطاف‌پذیری در مدیریت داده‌های توزیع‌شده تأکید دارند.

مقاله اول (میکروسرویس‌ها)

1. **طراحی DBMS‌های ویژه میکروسرویس:** نویسندگان خواستار نسل جدیدی از سیستم‌های مدیریت پایگاه داده هستند که به‌طور خاص برای محیط‌های میکروسرویس طراحی شده باشند. این سیستم‌ها باید شامل موارد زیر باشند:

- پشتیبانی داخلی از تراکنش‌های توزیع‌شده و غیرمتمرکز.
- ویژگی‌های بومی مبتنی بر رویداد (مانند جریان‌های کاری غیرهمزمان، event sourcing).
- کنترل‌های دقیق بر سطح انسجام داده که تعادلی میان خودمختاری و ایمنی ایجاد کند.

2. **افزایش آگاهی DBMS از جریان داده‌ها:** DBMS‌های آینده باید با درک عمیق‌تری از جریان‌های داده میان میکروسرویس‌ها طراحی شوند تا امکان هماهنگی بهتر، نظارت و اعمال محدودیت‌ها در مرزهای سرویس‌ها فراهم شود.

3. **ساده‌سازی هماهنگی داده میان سرویس‌ها:** مقاله پیشنهاد می‌کند کتابخانه‌ها و انتزاع‌هایی برای ساده‌سازی الگوهای رایج مانند ساگا، event sourcing و choreography/orchestration ایجاد شود تا بار منطق برنامه برای توسعه‌دهندگان کاهش یابد.

4. **مدل‌های تراکنشی ترکیبی:** پژوهش‌ها باید مدل‌هایی را بررسی کنند که میان تضمین کامل ACID و انسجام نهایی صرف قرار می‌گیرند و به میکروسرویس‌ها اجازه دهند بر اساس نیازهای کسب‌وکار یا عملیات، سطوح انسجام را به‌صورت پویا انتخاب کنند.

5. **ابزارهای بهتر برای مشاهده‌پذیری:** نیاز به ابزارهای مشاهده‌پذیری و دیباگینگ یکپارچه برای مدیریت داده در میکروسرویس‌ها وجود دارد تا تیم‌ها بتوانند وابستگی‌های داده و سلامت جریان‌های کاری را در میان سرویس‌های توزیع‌شده ردیابی کنند.

مقاله دوم (مدیریت داده صنعتی)

1. **توسعه معماری‌های مرجع برای خطوط داده صنعتی:** مقاله بر ایجاد مدل‌ها یا چارچوب‌های مرجع استاندارد برای خطوط داده صنعتی تأکید می‌کند. این چارچوب‌ها باید شامل موارد زیر باشند:

○ بهترین شیوه‌ها برای یکپارچه‌سازی منابع داده ناهمگن.

○ طراحی‌های ماژولار که هم پردازش دسته‌ای (batch) و هم بلادرنگ (real-time) را پوشش دهند.

2. **استفاده از استانداردهای Industry 4.0:** پژوهش‌های آینده باید بررسی کنند که چگونه معماری‌های مرجع Industry 4.0 (مانند RAMI 4.0) و استانداردها (مانند OPC-UA) می‌توانند به طراحی زیرساخت‌های مدیریت داده مقیاس‌پذیر و قابل تعامل کمک کنند.

3. **راهکارهای ترکیبی edge-cloud:** با توجه به محدودیت‌های رویکردهای صرفاً ابری یا کاملاً محلی، صنایع باید در معماری‌های ترکیبی سرمایه‌گذاری کنند که رایانش edge (برای وظایف حساس به تأخیر) را با منابع ابری (برای تحلیل‌ها و ذخیره‌سازی در مقیاس بالا) تلفیق کند.

4. **چارچوب‌های بهبود یافته پیش‌پردازش داده:** به پژوهش بیشتری برای ایجاد چارچوب‌های پیش‌پردازش انعطاف‌پذیر نیاز است که بتوانند با مجموعه داده‌های متنوع صنعتی سازگار شوند، افزونگی داده را کاهش دهند و کیفیت داده را در نقاط ورود بهبود بخشند.

5. **ادغام امنیت و حاکمیت داده:** مقاله بر اهمیت یکپارچه‌سازی اقدامات قوی‌تر در زمینه حاکمیت داده و امنیت سایبری در خطوط داده صنعتی تأکید دارد، به‌ویژه با حرکت سازمان‌ها به سمت محیط‌های بازتر و به‌هم‌پیوسته‌تر.

مسیرهای مشترک آینده

- **نوآوری در سطح سیستم:** هر دو مقاله پیشنهاد می‌کنند بخشی از پیچیدگی مدیریت داده‌ها از لایه برنامه به ابزارهای سطح سیستم (مانند DBMSها، چارچوب‌های خطوط داده) منتقل شود.
- **افزایش خودکارسازی و استانداردسازی:** هر دو مقاله خواستار الگوهای استانداردتر و ابزارهای خودکارسازی بیشتری هستند تا تلاش‌های مهندسی دستی و موقتی در سیستم‌های توزیع‌شده کاهش یابد.
- **پژوهش در مدل‌های انسجام تطبیقی:** هم میکروسرویس‌ها و هم سیستم‌های صنعتی می‌توانند از پژوهش در مدل‌های انسجام تطبیقی که با توجه به بار کاری یا دامنه کسب‌وکار تنظیم می‌شوند، بهره‌مند شوند.
- **تقویت معماری‌های مبتنی بر رویداد:** هر دو حوزه بر اهمیت فزاینده معماری‌های مبتنی بر رویداد تأکید می‌کنند و پیشنهاد می‌دهند پیشرفت بیشتری در ابزارها و میان‌افزارها برای مدیریت داده مبتنی بر رویداد صورت گیرد.