

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ



همایش اقتصاد ایران؛ اصلاح ساختارها، رفع ناترازی ها

و

نقش آفرینی نوین در عرصه بین الملل



کیفیت داده و شفافیت در ساختار حکمرانی مبنی بر داده

توسط:

سید امیر اصغری

دانشیار و معاون پژوهشی دانشکده فنی و مهندسی دانشگاه خوارزمی





Data as an
asset

همایش اقتصاد ایران

اصلاح ساختارها، رفع ناترازی ها و
نقش آفرینی نوین در عرصه بین الملل



• فهرست مطالب



Data
Quality





پژوهشکده
اقتصادی
وزارت امور اقتصادی و دارایی

همایش اقتصاد ایران

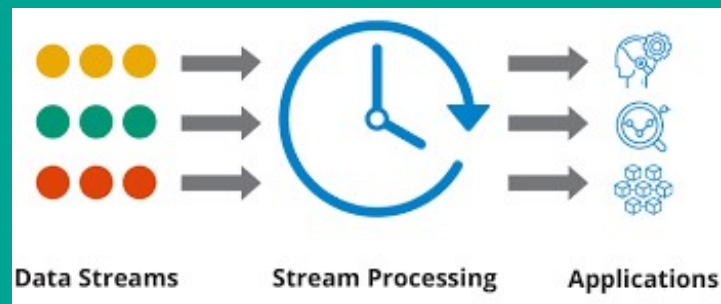
اصلاح ساختارها، رفع ناترازیها و
نقش آفرینی نوین در عرصه بین الملل



تولید داده و پایگاه داده



جریان سازی داده



ذخیره سازی داده

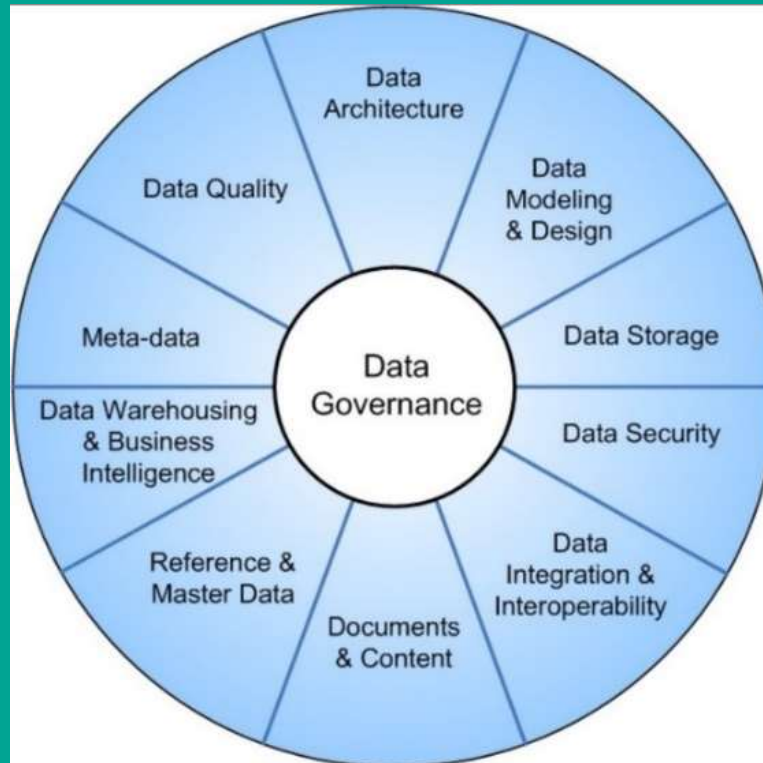




• اطلاعات یک دارایی است

- عبارت "اطلاعات یک دارایی است"، عبارتی معمول و پر استفاده می‌باشد؛ به این معنی که همانطور که دارایی‌ها مدیریت می‌شوند، اطلاعات نیز بایستی مدیریت شود. ارزش داده زمانی مشخص می‌شود که مورد استفاده قرار گیرد؛ برای مثال در تصمیم‌گیری‌ها. متقابلاً، ارزش منفی داده زمانی مشخص می‌شود که یا به غلط از آن استفاده شده یا خود داده نادرست باشد. در نتیجه می‌توان گفت که داده را می‌توان به عنوان یک امتیاز در حکمرانی داده مورد استفاده قرار داد.





• مجموعه دانش مدیریت داده





• ابعاد کیفیت داده





کیفیت داده

- کیفیت داده اندازه گیری می‌کند که چگونه یک مجموعه داده معیارهای زیر را برآورده می‌نماید (مرجع: www.ibm.com):
 - دقت (Accuracy) ؛ استفاده از منبع داده یا Data Source اصلی. از دیگر منابع داده می‌توان برای تایید مضاعف و اطمینان حاصل نمودن استفاده نمود.
 - کامل بودن (Completeness) ، داشتن درصد کمی از مقادیر از دست رفته.
 - اعتبار (Validity) ؛ چه کسری از داده‌ها، با فرمت مناسب ضوابط یا قوانین کسب و کاری منطبق هستند.
 - سازگاری (Consistency) ؛ عدم وجود تضاد داده‌ای در مجموعه داده. برای مثال، تعداد کارمندان در یک دپارتمان نمی‌تواند از تعداد کارمندان کل سازمان بیشتر باشد.
 - منحصر به فرد بودن (Uniqueness) ؛ برای مثال هنگام بررسی داده‌های مشتریان، انتظار بر این است که ID مشتری منحصر بفرد باشد.
 - به موقع بودن (Timeliness) ؛ به آماده شدن داده در یک فریم مورد انتظار از زمان اشاره دارد. برای مثال مشتریان انتظار دارند یک شماره رسید را به محض سفارش دریافت کنند.
 - تناسب با هدف (Fitness for purpose)؛ بررسی این که مجموعه داده برای رفع نیاز یک کسب و کار مشخص تولید شده است.
- اگر چالش‌های مرتبط با داده همانند داده های تکراری (Duplicated Data)، مقادیر از دست رفته (Missing Values) و داده‌های پرت (Outlier Data) به درستی آدرس‌دهی نشده و مورد بررسی قرار نگیرند، خطر خروجی‌های وارونه برای سازمان افزایش می‌یابد.

Garbage in, Garbage out





کیفیت داده

- ایجاد یک مرکز داده متمرکز نیازمند پاکسازی داده می‌باشد که شامل حذف "ضایعات دیجیتالی (Digital Debris)" و پاک کردن داده‌های ناقص است. در تحقیقاتی در سال ۲۰۰۲ توسط شورای انطباق، حکمرانی و نظارت (CGOC: Compliance, Governance and Oversight Council) مشخص شد که حدود ۶۹ درصد از اطلاعاتی که توسط کمپانی‌ها نگه داشته شده است، در واقع آوار دیجیتالی می‌باشد - اطلاعاتی که هیچ ارزش قانونی ندارند. در تحقیقاتی دیگر در سال ۲۰۰۲ توسط موسسه Data Warehousing مشخص شد که ۷۶ درصد از نقص‌های موجود در داده‌های سازمانی به دلیل ورودی داده ضعیف توسط کارمندان بوده است. به طور کلی می‌توان گفت که به جای سعی در یافتن نقص‌های داده و تلاش در رفع آن، بهتر است به سمت جریان کیفیت داده رفته و آن را در فرآیندهای تجاری تعبیه کرد.
- طبق گزارش گارتنر، هزینه کیفیت پایین داده‌ها، سالانه ۱۲.۹ میلیون دلار می‌باشد.





ضایعات دیجیتالی

- هیچ شخص یا سازمانی بصورت عمدی قصد تولید ضایعات دیجیتالی را ندارد. اما هر یک از ما همواره در حال ایجاد فایل، پشتیبان‌گیری، ایجاد پایگاه‌های داده و تولید اپلیکیشن هستیم. در این حین، عکس، موسیقی، برنامه‌های آموزشی و لاگ‌ها و گزارشات را ذخیره می‌کنیم.
- طبق نتایج منتشر شده CGOC در سال ۲۰۱۲:
 - ۱٪ اطلاعات سازمانی مشمول نگهداری قانونی است.
 - فقط ۵٪ بر اساس طرح طبقه‌بندی اسناد نگهداری می‌شود.
 - ۲۵٪ داده‌ها مربوط به یک نیاز تجاری است.
 - ۶۹٪ درصد باقیمانده هیچ ارزش قانونی یا تجاری ندارد.





• ضایعات دیجیتالی (ادامه)

نظرسنجی انجام شده در اجلاس CGOC در سال ۲۰۱۳ نشان داد که سازمان‌ها تمایلی به حذف زباله‌های دیجیتالی ندارند زیرا:

- ✓ نمی‌دانند از کجا شروع کنند
- ✓ نمی‌توان همه ذینفعان اطلاعات را روی میز آورد
- ✓ قادر به نشان دادن فوریت نیستند
- ✓ نمی‌توانند به وضوح اثرات منفی هزینه و ریسک را نشان دهند
- ✓ نمی‌توان یک مورد تجاری قانع‌کننده ایجاد کرد

باید در نظر بگیریم:

- ✓ ما هر روز ۵.۲ کوئینتیلیون بایت داده ایجاد می‌کنیم که این مقدار بطور مستمر در حال افزایش است.
- ✓ مکان‌های ذخیره‌سازی می‌توانند شامل استقرار در محل، خارج از سایت، ابر و نرم‌افزار به عنوان سرویس (SaaS) باشند و در انواع پیکربندی‌های ترکیبی ظاهر شوند.
- ✓ پلتفرم‌های رسانه‌های اجتماعی مانند توییتر، اینستاگرام یا فیس‌بوک حجم زیادی از داده‌ها را با عادات اجتماعی مختلف در نرخ و حجم بالا ترکیب می‌کنند و حجم زیادی از داده‌های بالقوه حساس را ایجاد می‌نمایند.
- ✓ فرآیندهای کشف الکترونیکی منجر به حفظ مقادیر زیادی از داده‌ها، از جمله بسیاری از موارد تکراری می‌شود.





• نمونه‌هایی از ضایعات دیجیتالی

- **Short-term reference files**
 - *SharePoint*, network drives, *Dropbox* and similar repositories used to enable sharing, transferring or temporarily storing files are rarely purged upon task completion.
 - Logs, reports or dumps from applications and database systems all lose their value over time.
 - “Work files” are rarely cleaned up at the end of the project or case.
- **Orphaned files**
 - Departing employees’ data often remains in place and unmanaged. If not specifically reviewed and classified, this data becomes unreferenced, unused and forgotten.
 - Mislabeled, misfiled documents (such as those not conforming to naming standards) will disappear from use but not from storage.
 - Files from old or unsupported applications remain in place because their custodians do not feel empowered to delete them as they are unsure what the files represent.





• نمونه‌هایی از ضایعات دیجیتالی

- **Outdated or superseded files**
 - Outdated versions or draft copies — major sources of near-duplicated analysis in electronic discovery — expose the company to tremendous and unnecessary risk. While initially useful during creation, these files ought to be cleared up or specifically classified as a business record. If the draft has not advanced in 6-12 months, the chances of it doing so diminish.
 - Draft or “old version” folders containing entire working sets of documents or drawings are often saved, superseded and then forgotten





• نمونه‌هایی از ضایعات دیجیتالی

- **Systems upgrades, safety and litigation copies**
 - As systems are replaced or upgraded, the content of the old system is not adequately merged with the new system, causing massive duplication. Similarly, in anticipation of upgrade complications, extra copies are created but never destroyed.
 - Files are often copied because of their importance at the time. However, even as their value decreases, the copies remain.
 - Impending litigation demands that the parties either preserve their data in place or duplicate it to another area. Often, these efforts lead to terabytes of redundant, even superfluous, data. Once the litigation resolves, the parties move on, neglecting to dispose of these “litigation copies”. Then, since they still exist, the same useless information is re-discovered in new cases. This can become a never-ending cycle.





• نمونه‌هایی از ضایعات دیجیتالی

- **Outdated storage technology**
 - The storage systems used to house data also age: from backup tapes to old disks to forgotten computer systems. Legacy systems in general are often a vast source of digital debris.
 - In the ground-breaking case of *Coleman (Parent) Holdings, Inc. v. Morgan Stanley & Co., Inc.*, 2005 WL 679071 (Fla. Cir. Ct. Mar. 1, 2005), the discovery of responsive but unprocessed emails on several undisclosed, eight-year-old backup tapes, led to a fine (later partially overturned) that exceeded \$1.4B.
 - A Fortune 100 company, performing an internal audit, located more than 33,000 items of abandoned storage media, Blackberry's, disks and computers. These legacy devices held very significant volumes of data — none of it easily accessible.





• نمونه‌هایی از ضایعات دیجیتالی

• Technical duplicates

- On average, 15% of duplicate files are necessary to their context and cannot be deleted without ruining the integrity of a larger set.
 - Examples are logo files that are embedded within other files, database configuration files, or externally referenced files in CAD engineering drawings.
- Technical duplicates are not debris, but they do increase the data set and cost and risk and have to be dealt.
- Duplicate files in existence at the establishment of a legal or preservation hold cannot be deleted.





**"اقیانوسی از زباله‌های دیجیتال در حال حاضر بسیاری از شرکت‌ها را
تحت الشعاع قرار داده است... تا زمانی که داده‌ها پاکسازی نشده و
فرایند پاکسازی ضایعات دیجیتالی، یک برنامه از حکمرانی داده نشود،
به هم ریختگی ادامه می‌یابد."**





• ارزش‌گذاری داده

- ارزش (یا بها) اختلاف میان هزینه یک شیء و منافع مستخرج از آن شیء است. برای برخی از دارایی‌ها، نظیر سهام، محاسبه ارزش آسان است. این ارزش اختلاف بین هزینه سهم در زمان خریداری و ارزش فروش آن است. ولی برای داده، این محاسبات بسیار پیچیده‌تر هستند، زیرا نه هزینه‌ها و نه منافع داده‌ها استاندارد شده نیستند.
- از آنجائیکه داده‌های هر سازمان مختص خود آن سازمان هستند، رویکرد ارزش‌گذاری داده باید با تشریح هزینه عمومی و دسته‌بندی‌های سود و منفعت که درون سازمان قابل اعمال هستند، آغاز شود. دسته‌بندی‌های نمونه شامل موارد زیر هستند:
- هزینه دریافت و ذخیره‌سازی داده
- هزینه جایگزینی داده در صورت از بین رفتن
- تاثیر از بین رفتن داده بر سازمان
- هزینه تقلیل ریسک و هزینه بالقوه ریسک‌های مرتبط با داده
- هزینه بهبود داده
- منافع داده‌های با کیفیت بالاتر
- میزان پرداختی رقبا برای داده
- موارد فروش داده
- درآمدهای مورد انتظار از استفاده‌های نوآورانه از داده





• ارزش‌گذاری داده

- چالش اصلی در ارزش‌گذاری داده این است که ارزش داده زمینه‌ای (چیزی که برای یک سازمان دارای ارزش است شاید برای سازمان دیگر بی‌ارزش قلمداد شود) و گاهی لحظه‌ای (چیزی که دیروز دارای ارزش بوده است شاید امروز ارزش خود را از دست داده باشد) است. با این وجود، درون سازمان انواعی از داده‌ها را می‌توان یافت که ارزش خود را با گذشت زمان حفظ می‌کنند. به عنوان مثال اطلاعات مشتریان را در نظر بگیرید که همواره برای سازمان در درجه اول اهمیت قرار دارند. با انباشته شدن داده‌های مرتبط با فعالیت مشتریان، اطلاعات مشتریان می‌توانند با گذشت زمان ارزش بیشتری نیز بیابند.





• ارزش‌گذاری داده

• داده‌های با کیفیت پایین برای هر سازمانی بسیار هزینه خواهند داشت. تخمین‌ها در این زمینه متفاوت هستند، ولی کارشناسان اعتقاد دارند که ۳۰-۱۰ درصد از هزینه سازمان‌ها صرف جبران مسائل کیفیت داده می‌شود. طبق تخمین IBM، هزینه کیفیت پایین داده‌ها در آمریکا در سال ۲۰۱۶ حدود ۳.۱ تریلیون دلار بوده است. بسیاری از هزینه‌های وجود چنین داده‌های بی‌کیفیتی مخفی، غیرمستقیم و بنابراین غیرقابل اندازه‌گیری هستند. سایر هزینه‌ها نظیر جریمه‌ها، مستقیم و قابل محاسبه می‌باشند. هزینه‌های این بخش شامل موارد زیر هستند:

- خرابی کار و اقدام به کار مجدد
- کارهای متفرقه و فرایندهای اصلاح پنهان
- ناکارآمدی سازمانی یا بهره‌وری پایین
- تضادها و درگیری‌های سازمانی
- رضایت شغلی کم
- نارضایتی مشتریان
- هزینه‌های فرصت شامل عدم توانایی در نوآوری
- هزینه‌ها یا جریمه‌های سازگاری
- هزینه‌های اعتبار





• یک نمونه از استفاده داده برای حکمرانی بهتر و شفاف‌سازی و اهمیت کیفیت داده

← → ↻ setadiran.ir/setad/cms/setadTransactions#top

سامانه تدارکات الکترونیکی دولت | ورود/ثبت نام

صفحه نخست قوانین و مقررات آگهی‌های عمومی شفاف‌سازی معاملات داشبوردهای هوش تجاری پشتیبانی افراد کم‌توان پیش‌نیازهای سیستمی آموزش امضای الکترونیکی سوالات متداول اخبار

۱۴۰۲ آذر ۱۷ ۲۰:۰۲

دستگاه‌های اجرایی

توجه: مرتب‌سازی جدول با کلیک بر روی ستون مورد نظر امکان‌پذیر می‌باشد.

هموطن عزیز، هموطن عزیز، در راستای دسترسی آزاد به اطلاعات غیر محرمانه و همچنین شفاف‌سازی معاملات، اطلاعات معاملات سامانه تدارکات الکترونیکی دولت به شرح زیر، در اختیار شما قرار می‌گیرد.

نام دستگاه مرکزی: وزارت علوم تحقیقات و فناوری

استان: همه موارد

نام سامانه: همه موارد

نوع معاملات: همه موارد

نام دستگاه اجرایی: [خالی]

پاک کردن شرایط جستجو جستجو

نمایش ۱ تا ۵۰ از ۴۲,۵۳۲ رکورد

ردیف	نام دستگاه اجرایی	نام دستگاه مرکزی	نوع معامله	موضوع معامله	شماره معامله	هزینه خرید اسناد	مهلت دریافت اسناد	مهلت ارائه پیشنهاد / ارسال پاسخ	تاریخ بازگشایی پاکتا	مبلغ معامله (ریال)
۱	دانشگاه آبه اله العظمی بروجردی (ره)	وزارت علوم تحقیقات و فناوری	مناقضه-ترک تشریفات	ترک تشریفات مناقصه واگذاری خدمات پشتیبانی دانشگاه	۲۰۰۹۶۴۲۶۸۰۰۰۰۰۰۰۶	۰	-	-	-	۱۶,۸۰۰,۰۰۰,۰۰۰





خداوندا! ما را از علم نافع و عمل صالح بهره‌مند و موفق بدار.