

entity resolution for big data

****Entity Resolution for Big Data: Unlocking Accurate Insights from Massive Datasets****

entity resolution for big data is rapidly becoming a crucial process in the world of analytics and data management. As organizations collect immense volumes of information from diverse sources, the need to accurately identify and consolidate records that refer to the same real-world entity—whether it be a person, product, or organization—has never been more important. Without effective entity resolution, data can become fragmented, inconsistent, and misleading, ultimately leading to poor decision-making and lost opportunities.

In this article, we'll explore the fundamentals of entity resolution for big data, its challenges, the latest techniques, and practical tips on implementing it successfully in large-scale environments. If you're navigating the complexities of big data analytics, understanding how to unify your data through entity resolution will empower you to extract clearer, more actionable insights.

What Is Entity Resolution in the Context of Big Data?

Entity resolution (ER), sometimes called record linkage or data matching, is the process of identifying and merging records that correspond to the same entity across one or multiple datasets. Imagine having millions of customer records scattered across various databases, each with slightly different spellings, incomplete information, or outdated details. Entity resolution helps unify these disparate entries into a single, comprehensive profile.

When applied to big data, the scale and complexity increase exponentially. Big data environments typically involve high velocity, volume, and variety, meaning that ER must handle massive datasets streaming in real-time or batch forms, often coming from numerous heterogeneous sources such as social media, customer relationship management (CRM) systems, transactional logs, and more.

Why Entity Resolution Matters in Big Data Analytics

Without entity resolution, organizations face several risks:

- ****Duplicate records**** inflate data size and distort metrics.
- ****Inconsistent data**** leads to unreliable analyses and faulty machine learning models.
- ****Missed connections**** between related data points reduce the ability to personalize services or detect fraud.
- ****Compliance issues**** arise when data about individuals is fragmented and incomplete.

By resolving entities effectively, companies can achieve a 360-degree view of their customers, improve operational efficiency, and unlock deeper insights from their big data.

Key Challenges of Entity Resolution for Big Data

Entity resolution in smaller datasets is already a non-trivial task, but when scaled to big data, several unique challenges emerge.

Volume and Velocity

The sheer volume of records—often billions or more—makes traditional pairwise comparison methods computationally infeasible. Additionally, data may be arriving at high velocity, requiring near real-time entity resolution to keep the data current and useful.

Data Variety and Quality

Big data typically comes from diverse formats and sources, including structured databases, semi-structured logs, and unstructured text. Variations in data quality, such as missing fields, typos, inconsistent formatting, or outdated information, complicate matching efforts.

Scalability and Performance

Algorithms that work well on small data fail to scale efficiently. Entity resolution for big data demands distributed computing frameworks and optimized algorithms to process massive datasets within acceptable timeframes and resource constraints.

Privacy and Security Concerns

When resolving sensitive personal data, organizations must ensure compliance with data protection regulations like GDPR and CCPA. This adds a layer of complexity, requiring secure data handling and potential anonymization techniques.

Techniques and Technologies Used in Entity Resolution for Big Data

Over the years, the field has evolved from simple rule-based systems to sophisticated machine learning approaches. Combining these with big data technologies enables effective entity resolution at scale.

Blocking and Indexing

To reduce the number of record comparisons, blocking methods group records into smaller candidate sets based on shared characteristics. For example, records might be clustered by ZIP code or the first letter of a last name. This drastically cuts down computational overhead in large datasets.

Deterministic vs. Probabilistic Matching

- **Deterministic matching** relies on exact or rule-based matches (e.g., same Social Security Number or email).
- **Probabilistic matching** assigns similarity scores based on multiple attributes and thresholds to decide matches, which is essential when data is noisy or incomplete.

Many big data ER systems combine both approaches for improved accuracy.

Machine Learning and AI Models

Machine learning techniques, especially supervised learning, have transformed entity resolution. Models can be trained on labeled pairs of records to predict whether they represent the same entity. Features might include string similarity metrics (Levenshtein distance, Jaccard similarity), numeric attribute differences, or domain-specific rules.

Deep learning models, including natural language processing (NLP) for unstructured data, enhance the ability to resolve entities even when details vary significantly.

Distributed Computing Frameworks

Big data ER often leverages distributed processing platforms like Apache Spark or Hadoop. These frameworks allow parallel processing of data partitions, enabling scalable execution of blocking, matching, and merging steps.

Graph-Based Approaches

Representing records as nodes in a graph where edges indicate potential matches helps capture complex relationships and transitive matches. Graph clustering algorithms can then identify groups of records belonging to the same entity.

Best Practices for Implementing Entity Resolution in Big Data Environments

Successfully deploying entity resolution at scale requires more than just picking the right technology. Here are some key tips to keep in mind:

Start with Data Profiling and Cleaning

Understanding your data's structure, quality, and common inconsistencies is essential. Data

cleansing, standardization, and normalization improve matching accuracy significantly.

Define Clear Matching Criteria

Collaborate with domain experts to establish meaningful rules and thresholds. Consider the business impact of false positives (incorrect merges) versus false negatives (missed matches).

Leverage Hybrid Approaches

Combine deterministic rules with probabilistic models and machine learning for the best balance of precision and recall. Iteratively refine models with feedback loops.

Use Scalable Infrastructure

Deploy your ER solution on cloud platforms or distributed clusters that can elastically scale with data size and processing needs.

Monitor and Evaluate Continuously

Entity resolution is rarely a one-off task. Continuously monitor match quality, retrain models as data evolves, and audit results to ensure ongoing reliability.

Address Privacy and Compliance

Implement data governance policies and anonymization techniques when necessary to protect sensitive information during the resolution process.

Real-World Applications of Entity Resolution for Big Data

Entity resolution is foundational to many industries and use cases where big data plays a pivotal role.

- **Customer 360 Views:** Retailers and financial institutions unify customer data from multiple channels to personalize marketing and improve service.
- **Fraud Detection:** Banks and insurers link disparate records to identify fraudulent patterns across accounts and transactions.

- **Healthcare:** Patient records from multiple providers are merged to provide comprehensive medical histories.
- **Supply Chain Management:** Companies match product and vendor information from varied sources to optimize operations.
- **Government and Public Sector:** Agencies consolidate citizen data for social services, tax collection, and law enforcement.

Each of these scenarios benefits tremendously from accurate and scalable entity resolution powered by big data technologies.

The Future of Entity Resolution in Big Data

As data volumes continue to explode and sources diversify, entity resolution will evolve with advances in artificial intelligence, edge computing, and privacy-enhancing technologies. Emerging trends include:

- **Automated feature engineering** that reduces human effort in model tuning.
- **Self-learning ER systems** that improve over time with minimal supervision.
- **Federated entity resolution**, enabling matching across organizations without sharing raw data, preserving privacy.
- **Integration with knowledge graphs** to enrich entity profiles and improve matching context.

These innovations promise to make entity resolution even more accurate, efficient, and privacy-conscious, unlocking greater value from big data than ever before.

Navigating the complexities of entity resolution for big data can be challenging, but it's a vital step toward harnessing the true power of your data assets. By combining the right methodologies, scalable technologies, and continuous refinement, organizations can transform messy, fragmented information into clear, unified insights that drive smarter decisions and better outcomes.

Frequently Asked Questions

What is entity resolution in the context of big data?

Entity resolution in big data refers to the process of identifying and merging records that refer to the same real-world entity across large and diverse datasets, ensuring data quality and consistency.

Why is entity resolution important for big data applications?

Entity resolution is crucial for big data applications because it helps eliminate duplicates, improve data accuracy, and enable more reliable analytics and decision-making by consolidating information

about entities from multiple sources.

What are the common challenges of entity resolution in big data environments?

Common challenges include handling massive data volumes, dealing with noisy or incomplete data, scaling algorithms efficiently, addressing data heterogeneity, and ensuring real-time or near-real-time processing capabilities.

Which algorithms are commonly used for entity resolution in big data?

Popular algorithms include probabilistic matching, rule-based matching, machine learning models (such as clustering and classification), and deep learning approaches, often combined with blocking techniques to reduce computational complexity.

How does blocking improve entity resolution performance in big data?

Blocking groups records into smaller subsets based on shared attributes, reducing the number of record comparisons needed and thus significantly improving the scalability and speed of entity resolution processes on large datasets.

What role does machine learning play in entity resolution for big data?

Machine learning automates and enhances entity resolution by learning complex matching patterns from labeled data, improving accuracy over traditional rule-based systems, and adapting to evolving data characteristics.

Which big data platforms support entity resolution workflows?

Platforms like Apache Spark, Hadoop, Google Cloud Dataflow, and AWS Glue provide scalable frameworks and tools that support the development and execution of entity resolution workflows on large datasets.

How can privacy be maintained during entity resolution in big data?

Privacy can be maintained by applying techniques such as data anonymization, secure multi-party computation, differential privacy, and federated learning to ensure sensitive information is protected while performing entity resolution.

Additional Resources

Entity Resolution for Big Data: Navigating the Complex Landscape of Identity Matching

entity resolution for big data represents a critical challenge and opportunity in today's data-driven world. As organizations collect vast quantities of information from multiple sources—ranging from customer databases and social media feeds to transactional records and IoT devices—the ability to accurately identify and link data entries that refer to the same real-world entity becomes essential. Without effective entity resolution, businesses risk fragmented insights, duplicated efforts, and compromised decision-making. This article offers a comprehensive examination of entity resolution for big data, unpacking its complexities, methodologies, and implications for enterprise-scale data management.

The Growing Importance of Entity Resolution in Big Data Environments

In the age of big data, organizations often grapple with heterogeneous datasets characterized by volume, velocity, and variety. Entity resolution (ER), also known as record linkage or data deduplication, serves as the process of disambiguating and consolidating records that pertain to identical entities—be they individuals, organizations, products, or locations. When dealing with big data, the scale and diversity of records introduce unique technical and operational challenges.

Big data environments typically involve:

- Massive datasets spanning millions or billions of records
- High-velocity data streams requiring near real-time processing
- Data heterogeneity, where records come from varied sources with inconsistent formats
- Data quality issues, including missing values, typographical errors, and conflicting information

The absence of a universal identifier across disparate systems exacerbates the difficulty of linking records reliably. For example, a single customer may be represented differently across sales, support, and marketing databases—using variations in name spelling, address formats, or contact details. Entity resolution for big data addresses this by employing sophisticated algorithms and scalable infrastructure to reconcile these discrepancies.

Why Traditional Entity Resolution Techniques Fall Short

Legacy ER methods, often designed for smaller, structured datasets, rely heavily on exact matching or simple rule-based heuristics. While these approaches can be effective in controlled settings, they tend to falter under big data conditions:

- **Scalability Constraints:** Pairwise comparison of all records is computationally prohibitive at scale, leading to performance bottlenecks.
- **Limited Flexibility:** Fixed rules and deterministic matching cannot accommodate the nuances of noisy or incomplete data.
- **Inability to Handle Diverse Data Types:** Traditional methods struggle with semi-structured or unstructured data common in big data contexts.

Consequently, enterprises have pivoted towards more advanced, probabilistic, and machine learning-driven methods to enhance accuracy and efficiency.

Core Methodologies in Entity Resolution for Big Data

Entity resolution frameworks for big data typically combine multiple strategies to optimize both precision and recall. Below are some of the predominant methodologies:

Blocking and Indexing Techniques

Given the combinatorial explosion in pairwise comparisons, blocking serves as a pre-processing step to reduce the search space. Records are grouped into “blocks” based on shared attributes or hashing functions, with comparisons performed only within each block.

Common blocking methods include:

- **Standard Blocking:** Partitioning data by exact values of one or more fields (e.g., ZIP code).
- **Sorted Neighborhood:** Sorting records by a key attribute and sliding a fixed-size window to compare neighboring records.
- **Canopy Clustering:** Using inexpensive similarity metrics to create overlapping clusters before detailed matching.

These approaches significantly improve computational feasibility while maintaining reasonable linkage quality.

Similarity Metrics and Probabilistic Matching

Matching records involves calculating similarity scores across relevant attributes. Techniques may employ string similarity measures like Levenshtein distance, Jaccard index, or cosine similarity for textual fields, and exact or range comparisons for numerical data.

Probabilistic models, such as the Fellegi-Sunter framework, assign match probabilities based on attribute agreement patterns, balancing false positives and negatives. These models can incorporate attribute-specific error rates and missing data handling, enhancing robustness.

Machine Learning and Deep Learning Approaches

Recent advances have introduced supervised and unsupervised learning techniques to entity resolution. Supervised models train on labeled pairs of matched and unmatched records to learn complex patterns beyond manual rules.

Popular algorithms include:

- Random forests and gradient boosting for feature-based classification
- Support vector machines optimized for pairwise record classification
- Deep neural networks leveraging embeddings to capture semantic similarities

Deep learning models can process unstructured data such as text, images, or audio, widening the scope of ER in big data scenarios. Moreover, active learning strategies help reduce labeling efforts by focusing on uncertain cases.

Technological Infrastructure for Scalable Entity Resolution

Entity resolution for big data demands not only advanced algorithms but also scalable computing resources and optimized workflows.

Distributed Computing Frameworks

Frameworks like Apache Hadoop and Apache Spark enable parallel processing of huge datasets across clusters of machines. Spark's in-memory computation and support for iterative algorithms make it particularly suitable for iterative ER workflows involving blocking, matching, and clustering.

Cloud-Based Solutions and Data Lakes

Cloud platforms offer flexible storage and compute scalability, facilitating the ingestion of heterogeneous data sources into centralized data lakes. They also provide managed services for machine learning and data analytics, reducing operational overhead.

Real-Time Entity Resolution

In contexts such as fraud detection, customer experience personalization, or cybersecurity, real-time or near real-time ER is critical. Stream processing systems combined with incremental matching algorithms focus on promptly reconciling incoming records against existing profiles.

Challenges and Considerations in Implementing Entity Resolution at Scale

Despite technological progress, entity resolution for big data remains an inherently difficult task with several persistent challenges:

- **Data Privacy and Security:** Handling sensitive personally identifiable information (PII) requires compliance with regulations like GDPR and HIPAA, complicating data sharing and processing.
- **Ground Truth Scarcity:** Obtaining accurately labeled data for supervised learning is often expensive and time-consuming.
- **Dynamic and Evolving Data:** Entities may change attributes over time, necessitating continuous updates and re-resolution.
- **Balancing Precision and Recall:** Over-aggressive matching can merge distinct entities erroneously, while conservative settings may miss true matches.

Addressing these issues often involves iterative tuning, domain expertise, and hybrid approaches that combine automated and manual verification.

Emerging Trends and Future Directions

The field of entity resolution for big data is rapidly evolving. Promising developments include:

- **Explainable AI:** Enhancing transparency in machine learning models to build trust and facilitate compliance audits.
- **Graph-Based Resolution:** Leveraging graph databases and network analysis to model complex relationships between entities.
- **Privacy-Preserving ER:** Techniques such as secure multiparty computation and federated learning enable entity resolution without exposing raw data.
- **Integration with Knowledge Graphs:** Enriching entity profiles with semantic context to

improve matching accuracy and downstream analytics.

These innovations aim to make entity resolution more accurate, scalable, and privacy-conscious in increasingly complex data ecosystems.

The ability to conduct effective entity resolution for big data is becoming a cornerstone of modern data strategy. Organizations that master this capability can unlock richer insights, streamline operations, and deliver more personalized experiences. Conversely, failure to resolve entities accurately risks fragmented data silos and impaired analytics. As big data grows ever more vast and varied, entity resolution will continue to be a pivotal discipline at the intersection of data engineering, machine learning, and business intelligence.

[Entity Resolution For Big Data](#)

Find other PDF articles:

<https://lxc.avoicemen.com/archive-top3-11/Book?dataid=CMG63-9364&title=finding-the-epicenter-worksheet-answer-key.pdf>

entity resolution for big data: Entity Resolution in the Web of Data Vassilis Christophides, Vasilis Efthymiou, Kostas Stefanidis, 2022-05-31 In recent years, several knowledge bases have been built to enable large-scale knowledge sharing, but also an entity-centric Web search, mixing both structured data and text querying. These knowledge bases offer machine-readable descriptions of real-world entities, e.g., persons, places, published on the Web as Linked Data. However, due to the different information extraction tools and curation policies employed by knowledge bases, multiple, complementary and sometimes conflicting descriptions of the same real-world entities may be provided. Entity resolution aims to identify different descriptions that refer to the same entity appearing either within or across knowledge bases. The objective of this book is to present the new entity resolution challenges stemming from the openness of the Web of data in describing entities by an unbounded number of knowledge bases, the semantic and structural diversity of the descriptions provided across domains even for the same real-world entities, as well as the autonomy of knowledge bases in terms of adopted processes for creating and curating entity descriptions. The scale, diversity, and graph structuring of entity descriptions in the Web of data essentially challenge how two descriptions can be effectively compared for similarity, but also how resolution algorithms can efficiently avoid examining pairwise all descriptions. The book covers a wide spectrum of entity resolution issues at the Web scale, including basic concepts and data structures, main resolution tasks and workflows, as well as state-of-the-art algorithmic techniques and experimental trade-offs.

entity resolution for big data: Big Data Analytics and Knowledge Discovery Carson K. Leung, Anton Dignös, Gabriele Kotsis, A. Min Tjoa, Ismail Khalil, 2025-09-20 This book constitutes the proceedings of the 27th International Conference on Big Data Analytics and Knowledge Discovery, DaWaK 2025, held in Bangkok, Thailand, during August 25-27, 2025. The 12 full and 14 short papers included in the proceedings were carefully reviewed and selected from 62 submissions. The proceedings also contain one keynote talk, one invited talk and one tutorial. The papers were organized in topical sections as follows: Data mining and knowledge discovery; sequential data analytics and recommendation systems; graph data processing and analytics; data management and

indices; Large language Models (LLMs); and Neural Networks.

entity resolution for big data: *Web and Big Data* Xiangyu Song, Ruyi Feng, Yunliang Chen, Jianxin Li, Geyong Min, 2024-04-27 The 4-volume set LNCS 14331, 14332, 14333, and 14334 constitutes the refereed proceedings of the 7th International Joint Conference, APWeb-WAIM 2023, which took place in Wuhan, China, in October 2023. The total of 138 papers included in the proceedings were carefully reviewed and selected from 434 submissions. They focus on innovative ideas, original research findings, case study results, and experienced insights in the areas of the World Wide Web and big data, covering Web technologies, database systems, information management, software engineering, knowledge graph, recommend system and big data.

entity resolution for big data: The Four Generations of Entity Resolution George Papadakis, Ekaterini Ioannou, Emanouil Thanos, Themis Palpanas, 2022-06-01 Entity Resolution (ER) lies at the core of data integration and cleaning and, thus, a bulk of the research examines ways for improving its effectiveness and time efficiency. The initial ER methods primarily target Veracity in the context of structured (relational) data that are described by a schema of well-known quality and meaning. To achieve high effectiveness, they leverage schema, expert, and/or external knowledge. Part of these methods are extended to address Volume, processing large datasets through multi-core or massive parallelization approaches, such as the MapReduce paradigm. However, these early schema-based approaches are inapplicable to Web Data, which abound in voluminous, noisy, semi-structured, and highly heterogeneous information. To address the additional challenge of Variety, recent works on ER adopt a novel, loosely schema-aware functionality that emphasizes scalability and robustness to noise. Another line of present research focuses on the additional challenge of Velocity, aiming to process data collections of a continuously increasing volume. The latest works, though, take advantage of the significant breakthroughs in Deep Learning and Crowdsourcing, incorporating external knowledge to enhance the existing words to a significant extent. This synthesis lecture organizes ER methods into four generations based on the challenges posed by these four Vs. For each generation, we outline the corresponding ER workflow, discuss the state-of-the-art methods per workflow step, and present current research directions. The discussion of these methods takes into account a historical perspective, explaining the evolution of the methods over time along with their similarities and differences. The lecture also discusses the available ER tools and benchmark datasets that allow expert as well as novice users to make use of the available solutions.

entity resolution for big data: *Database Systems for Advanced Applications* Chengfei Liu, Lei Zou, Jianxin Li, 2018-05-12 This book constitutes the workshop proceedings of the 23rd International Conference on Database Systems for Advanced Applications, DASFAA 2018, held in Gold Coast, QLD, Australia, in May 2018. The 23 full papers presented were carefully selected and reviewed from 44 submissions to the four following workshops: the 5th International Workshop on Big Data Management and Service, BDMS 2018; the Third International Workshop on Big Data Quality Management, BDQM 2018; the Second International Workshop on Graph Data Management and Analysis, GDMA 2018; and the 5th International Workshop on Semantic Computing and Personalization, SeCoP 2018.

entity resolution for big data: Innovative Techniques and Applications of Entity Resolution Wang, Hongzhi, 2014-02-28 Entity resolution is an essential tool in processing and analyzing data in order to draw precise conclusions from the information being presented. Further research in entity resolution is necessary to help promote information quality and improved data reporting in multidisciplinary fields requiring accurate data representation. *Innovative Techniques and Applications of Entity Resolution* draws upon interdisciplinary research on tools, techniques, and applications of entity resolution. This research work provides a detailed analysis of entity resolution applied to various types of data as well as appropriate techniques and applications and is appropriately designed for students, researchers, information professionals, and system developers.

entity resolution for big data: EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing Anandakumar Haldorai, Arulmurugan Ramu, Sudha

Mohanram, Chow Chee Onn, 2019-10-18 This proceeding features papers discussing big data innovation for sustainable cognitive computing. The papers feature detail on cognitive computing and its self-learning systems that use data mining, pattern recognition and natural language processing (NLP) to mirror the way the human brain works. This international conference focuses on cognitive computing technologies, from knowledge representation techniques and natural language processing algorithms to dynamic learning approaches. Topics covered include Data Science for Cognitive Analysis, Real-Time Ubiquitous Data Science, Platform for Privacy Preserving Data Science, and Internet-Based Cognitive Platform. The EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing (BDCC 2018), took place on 13 - 15 December 2018 in Coimbatore, India.

entity resolution for big data: Transactions on Large-Scale Data- and Knowledge-Centered Systems LVII Abdelkader Hameurlain, A Min Tjoa, 2024-10-24 The LNCS journal Transactions on Large-scale Data and Knowledge-centered Systems focuses on data management, knowledge discovery, and knowledge processing, which are core and hot topics in computer science. Since the 1990s, the Internet has become the main driving force behind application development in all domains. An increase in the demand for resource sharing (e.g. computing resources, services, metadata, data sources) across different sites connected through networks has led to an evolution of data- and knowledge-management systems from centralized systems to decentralized systems enabling large-scale distributed applications providing high scalability. This, the 57th issue of Transactions on Large-scale Data and Knowledge-centered Systems, contains five fully revised selected regular papers. Topics covered include leveraging machine learning for effective data management, access control models, reciprocal authorizations, Internet of Things, digital forensics, code similarity search, volunteered geographic information, and spatial data quality.

entity resolution for big data: Entity Information Life Cycle for Big Data John R. Talburt, Yinle Zhou, 2015-04-20 Entity Information Life Cycle for Big Data walks you through the ins and outs of managing entity information so you can successfully achieve master data management (MDM) in the era of big data. This book explains big data's impact on MDM and the critical role of entity information management system (EIMS) in successful MDM. Expert authors Dr. John R. Talburt and Dr. Yinle Zhou provide a thorough background in the principles of managing the entity information life cycle and provide practical tips and techniques for implementing an EIMS, strategies for exploiting distributed processing to handle big data for EIMS, and examples from real applications. Additional material on the theory of EIMS and methods for assessing and evaluating EIMS performance also make this book appropriate for use as a textbook in courses on entity and identity management, data management, customer relationship management (CRM), and related topics. - Explains the business value and impact of entity information management system (EIMS) and directly addresses the problem of EIMS design and operation, a critical issue organizations face when implementing MDM systems - Offers practical guidance to help you design and build an EIM system that will successfully handle big data - Details how to measure and evaluate entity integrity in MDM systems and explains the principles and processes that comprise EIM - Provides an understanding of features and functions an EIM system should have that will assist in evaluating commercial EIM systems - Includes chapter review questions, exercises, tips, and free downloads of demonstrations that use the OYSTER open source EIM system - Executable code (Java .jar files), control scripts, and synthetic input data illustrate various aspects of CRUD life cycle such as identity capture, identity update, and assertions

entity resolution for big data: The Enterprise Big Data Lake Alex Gorelik, 2019-02-21 The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build

one successfully with the best practices in this book. Alex Gorelik, CTO and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries

entity resolution for big data: Advances in Data Science and Information Engineering

Robert Stahlbock, Gary M. Weiss, Mahmoud Abou-Nasr, Cheng-Ying Yang, Hamid R. Arabnia, Leonidas Deligiannidis, 2021-10-29 The book presents the proceedings of two conferences: the 16th International Conference on Data Science (ICDATA 2020) and the 19th International Conference on Information & Knowledge Engineering (IKE 2020), which took place in Las Vegas, NV, USA, July 27-30, 2020. The conferences are part of the larger 2020 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE'20), which features 20 major tracks. Papers cover all aspects of Data Science, Data Mining, Machine Learning, Artificial and Computational Intelligence (ICDATA) and Information Retrieval Systems, Information & Knowledge Engineering, Management and Cyber-Learning (IKE). Authors include academics, researchers, professionals, and students. Presents the proceedings of the 16th International Conference on Data Science (ICDATA 2020) and the 19th International Conference on Information & Knowledge Engineering (IKE 2020); Includes papers on topics from data mining to machine learning to informational retrieval systems; Authors include academics, researchers, professionals and students.

entity resolution for big data: Utilizing Big Data Paradigms for Business Intelligence

Darmont, Jérôme, Loudcher, Sabine, 2018-08-10 Because efficient compilation of information allows managers and business leaders to make the best decisions for the financial solvency of their organizations, data analysis is an important part of modern business administration. Understanding the use of analytics, reporting, and data mining in everyday business environments is imperative to the success of modern businesses. Utilizing Big Data Paradigms for Business Intelligence is a pivotal reference source that provides vital research on how to address the challenges of data extraction in business intelligence using the five "Vs" of big data: velocity, volume, value, variety, and veracity. This book is ideally designed for business analysts, investors, corporate managers, entrepreneurs, and researchers in the fields of computer science, data science, and business intelligence.

entity resolution for big data: New Trends in Database and Information Systems Panos K.

Chrysanthi, Kjetil Nørvang, Kostas Stefanidis, Zheyang Zhang, Elisa Quintarelli, Ester Zumpano, 2025-09-21 This book constitutes short papers, Doctoral Consortium and workshop papers which were presented at the 29th European Conference on New Trends in Databases and Information Systems, ADBIS 2025, which took place in Tampere, Finland, during September 23-26, 2025. This CCIS proceedings book contains 14 short papers and 3 demo papers from the main ADBIS 2025 conference. In addition, 24 out of 48 papers submitted to the workshops are included in this book. The papers have been organized in topical sections as follows: Query Optimization; Spatio-Temporal & Graph Data; Data Sharing & Synthesis; Entity Resolution & Integration; Doctoral Consortium School Invited Talks; MADEISD 2025: 7th Workshop on Modern Approaches in Data Engineering and Information System Design; DOING 2025: 6th Workshop on Intelligent Data - From Data to Knowledge; K-GALS 2025: 4th Workshop on Knowledge Graphs Analysis on a Large Scale; CAIMA 2025: 1st Workshop on Cooperative AI Models and Applications; ERGA 2025: 1st Workshop on Entity Resolution and Graph Alignment; FEHDA 2025: 1st Workshop on Fairness Exploration in Heterogeneous Data and Algorithms; and IT4TOCI 2025: 1st Workshop on Information Technology for Tourism and Culture Industries.

entity resolution for big data: Intelligent Computing Kohei Arai, Supriya Kapoor, Rahul

Bhatia, 2018-11-01 This book, gathering the Proceedings of the 2018 Computing Conference, offers a remarkable collection of chapters covering a wide range of topics in intelligent systems, computing

and their real-world applications. The Conference attracted a total of 568 submissions from pioneering researchers, scientists, industrial engineers, and students from all around the world. These submissions underwent a double-blind peer review process. Of those 568 submissions, 192 submissions (including 14 poster papers) were selected for inclusion in these proceedings. Despite computer science's comparatively brief history as a formal academic discipline, it has made a number of fundamental contributions to science and society—in fact, along with electronics, it is a founding science of the current epoch of human history ('the Information Age') and a main driver of the Information Revolution. The goal of this conference is to provide a platform for researchers to present fundamental contributions, and to be a premier venue for academic and industry practitioners to share new ideas and development experiences. This book collects state of the art chapters on all aspects of Computer Science, from classical to intelligent. It covers both the theory and applications of the latest computer technologies and methodologies. Providing the state of the art in intelligent methods and techniques for solving real-world problems, along with a vision of future research, the book will be interesting and valuable for a broad readership.

entity resolution for big data: *Database and Expert Systems Applications* Christine Strauss, Toshiyuki Amagasa, Gabriele Kotsis, A Min Tjoa, Ismail Khalil, 2023-08-17 The two-volume set, LNCS 14146 and 14147 constitutes the thoroughly refereed proceedings of the 34th International Conference on Database and Expert Systems Applications, DEXA 2023, held in Penang, Malaysia, in August 2023. The 49 full papers presented together with 35 short papers were carefully reviewed and selected from a total of 155 submissions. The papers are organized in topical sections as follows: Part I: Data modeling; database design; query optimization; knowledge representation; Part II: Rule-based systems; natural language processing; deep learning; neural networks.

entity resolution for big data: *Social Networks Analysis and Mining* Luca Maria Aiello, Tanmoy Chakraborty, Sabrina Gaito, 2025-01-23 This LNCS conference 4-volume set constitutes the proceedings of the 16th International Conference on Social Networks Analysis and Mining, ASONAM 2024, in Rende, Italy, during September 2-5, 2024. The 33 full papers together with 36 short papers included in this volume were carefully reviewed and selected from 167 submissions. The conference covers a wide spectrum of research contributions to the foundations and applications of social networks.

entity resolution for big data: *Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0* M. Peruzzini, M. Pellicciari, C. Bil, 2018-09-14 The concept of concurrent engineering (CE) was first developed in the 1980s. Now often referred to as transdisciplinary engineering, it is based on the idea that different phases of a product life cycle should be conducted concurrently and initiated as early as possible within the Product Creation Process (PCP). The main goal of CE is to increase the efficiency and effectiveness of the PCP and reduce errors in later phases, as well as incorporating considerations – including environmental implications – for the full lifecycle of the product. It has become a substantive methodology in many industries, and has also been adopted in the development of new services and service support. This book presents the proceedings of the 25th ISPE Inc. International Conference on Transdisciplinary Engineering, held in Modena, Italy, in July 2018. This international conference attracts researchers, industry experts, students, and government representatives interested in recent transdisciplinary engineering research, advancements and applications. The book contains 120 peer-reviewed papers, selected from 259 submissions from all continents of the world, ranging from the theoretical and conceptual to papers addressing industrial best practice, and is divided into 11 sections reflecting the themes addressed in the conference program and addressing topics as diverse as industry 4.0 and smart manufacturing; human-centered design; modeling, simulation and virtual design; and knowledge and data management among others. With an overview of the latest research results, product creation processes and related methodologies, this book will be of interest to researchers, design practitioners and educators alike.

entity resolution for big data: *Bioinformatics and Biomedical Engineering* Ignacio Rojas, Francisco Ortuño, Fernando Rojas, Luis Javier Herrera, Olga Valenzuela, 2024-08-27 This volume

CPA ACCA CPA Reporting entity entity
Model Entity - Model Entity MVC Model
Model
substance entity - 5 substance “ ” “ ”
essence substance
- 2011 1
- 2011 1
- Unity Dots GameObject Entity - load prefab Entity GameObject Entity
GameObjectEntity Co
“ACAD_PROXY_ENTITY” - “ACAD_PROXY_ENTITY” []
CAD
Cesium | 04 Entity 23 Jul 2024 Entity Cesium API Entity
Entity Graphics 17
Busca de dados com Entity Framework - Stack Overflow em 17 Feb 2016 Tenho um projeto em Asp.Net MVC onde preciso trabalhar com um grande volume de dados. Para obtê-los do Banco de Dados (Microsoft SQL Server) usando Entity Framework
Reporting entity entity Reporting entity entity - CPA ACCA CPA Reporting entity entity
Model Entity - Model Entity MVC Model
Model
substance entity - 5 substance “ ” “ ”
essence substance
- 2011 1
- 2011 1
- Unity Dots GameObject Entity - load prefab Entity GameObject Entity
GameObjectEntity Co
“ACAD_PROXY_ENTITY” - “ACAD_PROXY_ENTITY” []
CAD
Cesium | 04 Entity 23 Jul 2024 Entity Cesium API Entity
Entity Graphics 17
Busca de dados com Entity Framework - Stack Overflow em 17 Feb 2016 Tenho um projeto em Asp.Net MVC onde preciso trabalhar com um grande volume de dados. Para obtê-los do Banco de Dados (Microsoft SQL Server) usando Entity Framework
Reporting entity entity Reporting entity entity - CPA ACCA CPA Reporting entity entity

Related to entity resolution for big data

Sigma360 Unveils Real-Time Transaction Screening for High-Volume Payments with Unprecedented Accuracy (5d) Sigma360®, the definitive AI-powered risk screening and monitoring platform, announced the development of a new product line, Transaction Screening, which can screen hundreds of millions of

Sigma360 Unveils Real-Time Transaction Screening for High-Volume Payments with Unprecedented Accuracy (5d) Sigma360®, the definitive AI-powered risk screening and monitoring platform, announced the development of a new product line, Transaction Screening,

which can screen hundreds of millions of

Quantexa Launches Cloud AML Product for U.S. Mid-size and Community Banks (13d) SaaS solution delivers power of contextual data, analytics, & AI to financial crime detection and prevention, resulting in 75% fewer false positives and 80% faster investigationsLAS VEGAS, Sept. 16,

Quantexa Launches Cloud AML Product for U.S. Mid-size and Community Banks (13d) SaaS solution delivers power of contextual data, analytics, & AI to financial crime detection and prevention, resulting in 75% fewer false positives and 80% faster investigationsLAS VEGAS, Sept. 16,

Back to Home: <https://lxc.avoiciformen.com>