We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Big Data Integration Solutions in Organizations: A Domain-Specific Analysis

Sreekantha Desai Karanam,

Rajani Sudhir Kamath, Raja Vittal Rao Kulkarni
and Bantwal Hebbal Sinakatte Karthik Pai

Abstract

Big Data Integration (BDI) process integrates the big data arising from many diverse data sources, data formats presents a unified, valuable, customized, holistic view of data. BDI process is essential to build confidence, facilitate high-quality insights and trends for intelligent decision making in organizations. Integration of big data is a very complex process with many challenges. The data sources for BDI are traditional data warehouses, social networks, Internet of Things (IoT) and online transactions. BDI solutions are deployed on Master Data Management (MDM) systems to support collecting, aggregating and delivering reliable information across the organization. This chapter has conducted an exhaustive review of BDI literature and classified BDI applications based on their domain. The methods, applications, advantages and disadvantage of the research in each paper are tabulated. Taxonomy of concepts, table of acronyms and the organization of the chapter are presented. The number of papers reviewed industry-wise is depicted as a pie chart. A comparative analysis of curated survey papers with specific parameters to discover the research gaps were also tabulated. The research issues, implementation challenges and future trends are highlighted. A case study of BDI solutions implemented in various organizations was also discussed. This chapter concludes with a holistic view of BDI concepts and solutions implemented in organizations.

Keywords: master data management (MDM), internet of things (IoT), business intelligence (BI), software as a service (SAAS), machine learning (ML), artificial intelligence (AI)

1. Introduction

Accenture company has conducted a survey on the implementation of BDI solutions in organizations. The survey outcome revealed that 92% of managers are happy with the results obtained from BDI solutions and 89% of managers agree that big data integration and analytics is very vital for their business planning to leverage competition. The Internet Trends Report from KPCB's by Mary Meeker discovered the decreasing trends in the cost of hardware technology in the past twenty years, the cost of computing has been reduced by 33%, 38% storage cost reduction

and 27% bandwidth costs reduction year after year. The major challenges faced in BDI processing are data selection, gathering, storing, communication, searching, visualization, ensuring privacy, and security of data.

The efficiency in handling big data drives effective decision making. The advancement in computing infrastructures, algorithms and innovative technologies have boosted the big data management and analytics domain and reduced the investment costs to deliver the best value for businesses.

1.1 Motivation and significance of BDI study

The business experts have agreed that big data would mean big value. The digital transformation of business operations is enhancing customer experience and reducing costs. Consumers would like to access personalized data and carry out business on the go. Online processing of bigdata using analytical platforms in the organizations can make the information accurate, standardized, and actionable. Acquiring insights from big data leverage the companies to make more informed business decisions with improved efficiency, and to design more BDI applications. The revolution in computing and digitalization has also increased the potential of cyber-attacks. The cyber threats by hackers are ever increasing and becoming more and more complex day by day. ML and DL techniques have been significantly applied to design intelligent and secure BDI solutions for automating business processes. ML projects are receiving the maximum funding since 2019, compared to all other AI projects combined. Walmart corporation has implemented BDI solutions for acquiring business intelligence and taking real-time business decisions. Many leading fast-food based companies such as McDonald's, KFC, Pizzahut are using BDI solutions for designing their marketing strategies to discover the hanging business trends. The Casinos are also utilizing the BDI solutions to enhance their revenues in the recent years and to attract and inspire customers for regular visits. The hotel industry uses BDI applications to predict customer behavior, food habits and demands. Tourists today are also using digital solutions to collect information on all issues related to tourism. BDI has been applied in the healthcare industry for rendering quality healthcare services, decreasing the wastage of money and time. The governments are using BDI for developing smart city public services. BDI has empowered e-commerce industries such as Amazon, Flipkart, etc. by providing data insights and analytical reports. The integration of AI, BDI and visualization tools helped meteorologists to predict weather conditions precisely. BDI solutions have been applied successfully in modern agriculture. BDI solutions have also empowered digital marketing for the success of every business. The above facts and applications have motivated the researchers to study the BDI in detail.

1.2 International market potential

According to global forecasts BDI solutions market size is estimated to reach US$ 12.24 billion by 2022 at a Compound Annual Growth Rate (CAGR) of 13.7%. The market survey by Dresner Advisory Assc. in the year 2020 has revealed that 80% of organizations are considering BDI solutions as critical for decision-making activities and 60% of them prefer to deploy BDI solutions on cloud platforms. International Data Corporation (IDC) has predicted that the global data-sphere would be about 175 zettabytes by 2025. IDC has estimated that several billion IoT devices and embedded systems would generate, gather, communicate a wealth of IoT data and carryout analytics every day throughout the world. IDC has also predicted that by 2025 about six billion customers or 75% of the global population would be communicated by using online and real-time data every day. The share of real-time data would be about 30% in global data as estimated by IDC.

1.3 Overview of BDI technologies

1.3.1 BDI process types

BDI is the process of consolidating data from multiple applications and creating a unified view of data assets. BDI is the main component of various mission-critical data management projects, such as building an enterprise data warehouse, migrating data from one or multiple databases to another, and synchronizing data among applications. BDI directs at furnishing an integrated and consistent view of data coming from external and internal data sources.

1.3.1.1 Data consolidation

Big data consolidation is the process of consolidating or integrating data from various data sources to make a centralized data store or repository. This is an amalgamated data store used for diverse purposes, such as data analysis and reporting. It can also execute for downstream applications as a data source.

1.3.1.2 Data federation

A Data Federation is a data integration technique. Data federation is used to integrate the data and simplify the approach for consuming by the users and front-end applications. In data federation, distributed data with various data models are combined into a unified data model that features a virtual database.

1.3.1.3 Data propagation

It is another technique for data integration. Data would be propagated from an enterprise data warehouse to different data marts after the needed transformations.

1.3.2 BDI technologies

1.3.2.1 Extract, transform, load (ETL)

ETL is the best-known data integration technology. ETL is a process of data integration that includes extraction of data from a source system and it's loading after transformation to a target destination.

1.3.2.2 Enterprise information integration (EII)

This data integration technology is used to deliver curated data-sets on an on-demand basis. EII is a technology that admits developers and business users alike to treat a range of data sources as if they were one database and represent the incoming data in novel ways.

1.3.2.3 Enterprise data replication (EDR)

EDR is a real-time data consolidation method that includes moving data from one storage system to another. In its simplest form, having the same schema, EDR involves shifting a data-set from one database to another database.

1.3.3 Bigdata integration platforms

1.3.3.1 Adeptia connect

Enterprise BDI tools provided by Adeptia may be utilized by other than technical business users. Adeptia Connect has an easy user interface to coordinate with all data interfaces and external connections. It also involves a no-code approach and self-service partner onboarding that allows partners and users to view, set up and coordinate data connections. The platform brags a suite of Cloud Services Integration and pre-built connections along with protocol support and B2B standards.

1.3.3.2 Alooma platform

Alooma provides a data pipeline service that combines with prevalent data sources. The Alooma platform contains security from end-to-end level, which ascertains that every event is securely moved to a data warehouse (HIPAA, SOC2, and EU-US Privacy Shield certified). The solution reacts to the changes in data in real-time to ascertain that no such events have vanished. Users can select to carry out changes automatically or get notified and do on-demand changes. This tool also automatically reduces the data volume to make control customizable.

1.3.3.3 Boomi AtomSphere

It is a Dell Technologies company's Boomi's flagship product. AtomSphere supports an integration process between cloud platforms, software-as-a-service applications, and on-prem systems. The visual interface is used by AtomSphere is used to configure application integrations. Wherever it is needed, the solution's runtime tool, Boomi Atom, allows integrations to be deployed. Based on use case and functionality, the AtomSphere platform is also available in various editions.

1.3.3.4 Integrator.io

Celigo company provides an Integration Platform called Integrator.io as a service product. The solution enables organizations to synchronize the data, connect applications, and automate processes. Celigo bundles an integration wizard that involves visual field mapping interface, an API assistant, and drop-down menus. This tool also provides integration templates which are reusable pre-configured on the integrator.io marketplace, permitting users to have their library of reusable, standalone flows.

1.3.3.5 Cleo Integration Cloud

The Cleo Integration Cloud accords organizations to connect to SaaS applications and enterprises with a range of connectors and APIs. This tool automatically accepts, transforms, orchestrates, connects and integrates all B2B data types from any source and to any target, and can be implemented via several different methods. Cleo Integration Cloud can also be engrafted for Information Services organizations or SaaS and can be used as an administered service to divest complex integrations to the vendor's experts.

1.3.3.6 Denodo Platform

The Denodo Platform provides data virtualization for integrating multi-structured sources of data from database management systems, a wide variety of other big data, cloud, documents, and enterprise sources. Connectivity support involves legacy data, flat files, relational databases, packed applications, CML, and emerging data types including Hadoop. The only data virtualization solution, Denodo, is to be represented as a virtual image on AWS Marketplace of Amazon.

1.3.3.7 Diyotta data integration suite

A unified data integration platform, Diyotta, that combines with data warehousing environments and modern data lake. The native processing capabilities and drag-and-drop user interface to build this product. Diyotta enables faster data movement, shorter development times, and reusability all over the enterprise to simplify future development. Diyotta touts the industry's first data integration software to leverage modern data processing platforms like Snowflake, Google BigQuery, Hadoop, and Amazon Redshift.

1.3.3.8 IBM products - InfoSphere information server

IBM provides several distinct data integration tools in both cloud and on-prem deployments, and for every enterprise use case virtually. Its on-prem data integration suite has tools for modern data integration synchronization, data virtualization) and traditional (replication and batch processing) requirements.

IBM also provides a range of connectors and pre-built functions. The mega-vendors cloud integration product is considered as one of the most excellent in the marketplace.

1.3.3.9 Informatica products - an intelligent data platform

Informatica's data integration tools portfolio covers both cloud deployments and on-prem for several enterprise use cases. The vendor integrates governance functionality and advanced hybrid integration with self-service business access for different analytical functions. Augmented integration is possible via a metadata-driven AI engine, and Informatica's CLAIRE Engine, that enforces machine learning. Informatica touts interoperability in strong in nature.

1.3.3.10 Microsoft Products - SQL Server Integration Services (SSIS).

The company's SQL Server Integration Services (SSIS), traditional integration tools, is integrated inside the SQL Server DBMS platform. Microsoft also promotes two cloud SaaS products: Microsoft Flow and Azure Logic Apps. Flow is adhoc integrator-centric and integrated into the overarching Azure Logic Apps solution.

1.3.3.11 Oracle products - data integration cloud service

Oracle provides a full spectrum of data integration tools for modern ones as well as conventional use cases, in both cloud and on-prem deployments.

The company's product portfolio includes services and technologies that permit organizations for data enrichment and full lifecycle data movement. Oracle data integration allows permanent and uninterrupted access to data across heterogeneous systems via transformation, bidirectional replication, bulk data movement, data services, metadata management, and data quality for product and customer domains.

1.3.3.12 SAP products - data services

SAP provisions clouds and on-prem integration functionality by two primary channels. Traditional capabilities are provided through a data management platform, SAP Data Services, that gives capabilities for data cleansing, integration, and quality. SAP Cloud Platform provides Integration Platform as a Service features are existing in it. Integration of processes and data between cloud apps, third-party applications, and on-prem solutions are arranged through SAP's Cloud Platform.

1.4 Organization of the chapter

This chapter has been framed into seven sections. Section 1 explains the introduction, sub section 1.1 discusses the motivation and significance of the study. Sub section. 1.2 shows international market potential and 1.3 presents an overview of big data technologies and taxonomy. Sub section 1.4 Organization of the chapter, 1.5 summarizes the authors' research contribution. 1.6 Illustrates the list of acronyms used. The review of recent literature is described in Section 2. Section 2 is further divided into four subsections. 2.1 subsection describes the papers reviewed from one technology domain. The highlights and findings from each paper are tabulated. Sub section 2.2 deals with a comparative analysis of survey papers with specific parameters. Section 3 shows the architecture of BDI. 4th section deals with the research issues, challenges. Section 5. Presents the case studies of BDI solutions from various organizations. Section 6. Outlines the findings and conclusion. Section 7 is the references of the papers reviewed.

1.5 Research contribution

1. This study has revealed that various technologies, systems, techniques, algorithms are applied for implementing business intelligence systems across the world. These papers have been further classified technology-wise and presented as a pie chart in **Figure 1**.

2. A table of acronyms is presented in **Table 1**

3. **Figure 2** presents taxonomy of concepts applied in BDI techniques in various applications

4. The overview of the organization of this chapter, section-wise is shown **Figure 3**.

5. In each concept of taxonomy, the existing literature has been mapped to several issues as shown in **Figure 4**.

6. Research issues, challenges and future directions of BDI technologies are discussed

7. A case study of five BDI based solutions implemented healthcare, retail, finance and tourism domains are discussed

8. The set of a curated survey papers are compared with specific factors such as architecture, open issues and challenges, applications, taxonomy and security to understand the scope of coverage each paper and to understand the research gaps **Tables 2** and **3**

1.5.1 Table of acronyms

This section shows a list of all the acronyms used in this chapter for easy reference is presented **Table 1**.
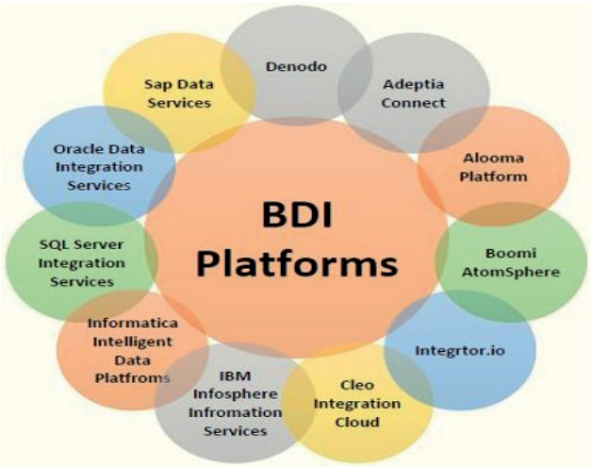

Figure 1.
Bigdata integration platforms.

Acronym	Description	Acronym	Description
AI	Artificial Intelligence	BDI	Big Data Integration
AWS	Amazon Web Services	AIG	American International Group
B2B	Business to Business	CAGR	Compound Annual Growth Rate
CCTV	Closed Circuit TV	ETL	Extract, Transform, Load
EII	Enterprise Information Integration	EDR	Enterprise Data Replication
HIPAA	Health Insurance Portability and Accountability Act	ICT	Information and Communication Technology
IoT	Internet of Things	MDM	Master Data Management
IoMT	Internet of Medical Things		
SOC 2	Service organization control is an auditing procedure	API	Application Programming Interface

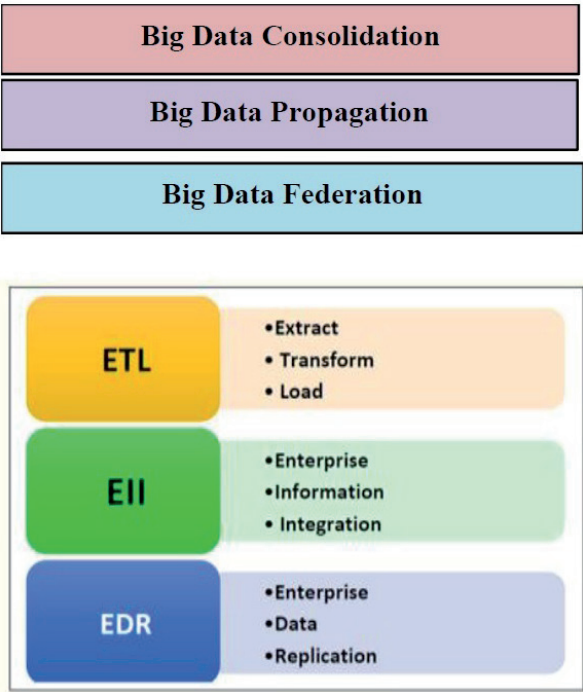**Table 1.**
List of acronyms used in this chapter.
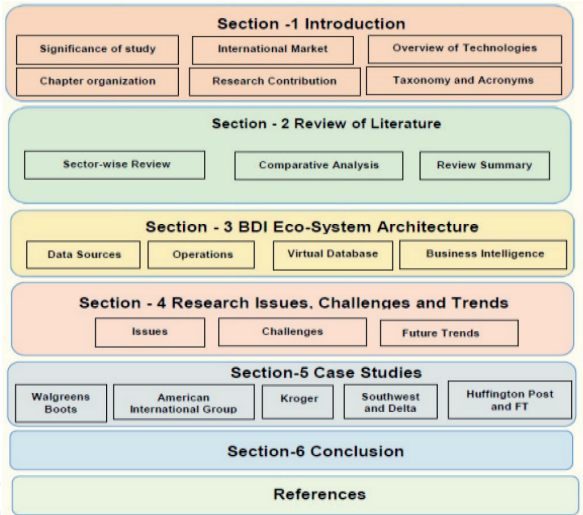
Figure 2.
Taxonomy of big data concepts.
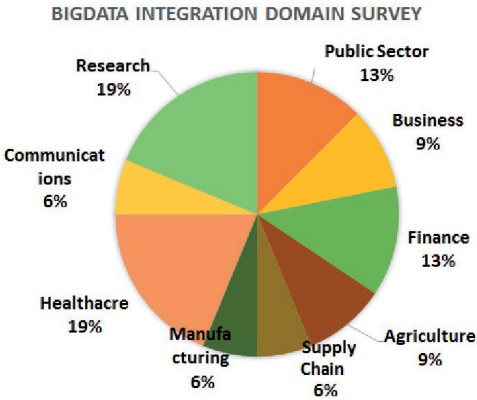

Figure 3.
Organization of the chapter.

Figure 4.
Sector-wise reviewed papers.

Ref. No.	Methods Used	Results	Applications
1	Interviewing of experts and content analysis approach was used as a qualitative technique for data gathering.	The principal problems identified during the study are the hardship of administration, the ineffectiveness of human resource, politics, standards and absence of executives	Data fusion solutions for public sector
2	Comparisons, experiments and questionnaires	The decorum status of high recognition of positive viewpoint is around 65%. Negative perspective has an absence of etiquette knowledge accounted for 56%	Talent refinement of college students
3	Semantic data model, Resource description system, SPARQL, semantic query language	Semantic extract transform load system produces semantic information that would possibly be distributed on the web as Web of data.	Innovative Big data applications in fuel economy, household transportation and vehicles
4	Exploration of BDI based on stage, features and semantic meaning	Big data problems are resolved by appropriate BDI methods.	Open cross-domain Big data
5	A study on analysis of industrial needs and potential applications of BDI in the public sector	A set of open research questions such as scalability of data required in real-time applications	Labor agency, Online gambling operations, Public Safety, and Predictive policing

Table 2.
Public sector review summary.

Ref. No.	Methods Used	Results	Applications
6	Review of literature on BDI, Business Intelligence and Cloud Computing	It is possible to integrate technologies to the need of SMEs	Small and Medium-sized organizations
7	Developing a deployable data integration tool that handles technical issues.	Shortage of machine learning examples, the requirement of clarifying business owners' outcomes and the expense of involving domain experts.	Scalable data integration challenges in the enterprise
8	The amalgamation of diverse sources in the Hadoop environment with HDFS. Spark computation model with a Hive database as a distributed data warehouse	A prototype of a data integration system	E-Commerce Domain
9	A study on data collection, analytics implementation, and benefits of BDI	BDI implementation in business organizations improves business performance	Performance improvement in business organization
10	Randomly selected articles on big data are reviewed to analyze the role of big data in business	As per the study, 63% of business reported that the implementation of big data is useful to business	Decision making in business

Table 3.
Business sector review summary.

2. Review of recent literature

Authors have selected papers from highly reputed research journals from IEEE, Elsevier, Science Direct and Springer publications. About fifty-five papers

covering big data integration concepts and applications are reviewed. This section presents the findings and highlights from each reviewed paper which are organized domain-wise.

2.1 Review of the public sector

Hasliza et al. analyzed the fundamental problems and difficulties encountered by the BDI solutions in the public sector [1]. The discovery of the right dimensions and factors are important to find the solutions to these problems. Zhang has reported the BDI solutions for professional procedure amalgamation in modern decorum [2]. The comparisons, experiments and questionnaires concerning the BDI concepts are discussed. Bansal has proposed the use of semantic technologies for the distribution of information in the contest of semantic ETL [3]. This information is open and the data was gathered from various sources. Zheng et al. have presented significant standards, classification of strategies and models of BDI process [4]. The real BDI issues are discussed using these models.

Authors have classified BDI techniques based on different combinations of strategies such as stage, feature and semantics. Munne has explained the technological trends for current social and economical status. This paper highlighted BDI technologies and applications in the public sector [5]. **Table 2** presents a summary of the highlights of the papers reviewed the public sector.

2.2 Review from business sector literature

The study by Camargo et al. revealed the possibility of incorporating and implementing BDI technologies to the needs of small and medium scale enterprises [6]. Some companies are offering open source BDI tools for organizations for intelligent business decision making. Stonebraker et al. discussed the difficulties in the scalability of BDI solutions today and in near future [7]. This analysis was carried out using the past five years' data from large enterprises. The integration of data from heterogeneous sources in a distributed environment was explored by Sazontev et al. [8]. The authors have explained the process of BDI framework development and its methodology. Alsghaier et al. discussed BDI process in Hadoop platform for business organizations. Authors focused on the implementation and benefits of big data analytics in business organizations [9].

Alam et al. have reviewed the role of BDI in the business sector [10]. **Table 3** presents a summary of highlights from the business sector review.

2.3 Review of the finance sector

Fikri et al. presented a BDI approach combined with distributed datasets of financial ontology and a real-time data stream [11]. This model was associated with classic ETL. This model was suitable for handling BDI in real-time. Bucea-Manea-Tonis illustrated the use of predictive logic in deductive frameworks to integrate different sets of data types [12]. Chen et al. have proposed a framework for managing data with heterogeneity problems [13]. This unified data model was adaptable to different data sources by setting up panoramic data. Hussain and Prieto discussed the analysis of industrial needs, constraints and potential applications of BDI to insurance and finance sectors [14]. Authors have mapped the requirements to research queries. The paper by Avi and Kamaruddin reported the role of BDI in insurance, finance and banking sectors [15]. Authors have highlighted the benefits of cutting-edge technologies associated with BDI in the financial sector. **Table 4** presents a summary of highlights from the finance sector review.

Ref. No.	Methods Used	Results	Applications
11	Combining distributed datasets of financial ontology and real-time stream	The data integration pipeline in real-time. The use of Apache Spark enhances short time frames for quality and availability reporting	Data integration in real-time, Financial reporting
12	Predicate logic in deductive systems	Integrates different kinds of data types	E-Commerce applications
13	Integrating heterogeneous data from multiple sources, Big data ETL in a distributed environment	Better performance in processing data integration from multiple sources	Power dispatching and control system
14	A review of industrial needs, constraints and application	Highlights the challenges in providing an effective technological solution	Manipulation recognition, threat management in finance and insurance sectors
15	A comprehensive review of the banking sector in terms of digital banking, analytics, mobile banking. Use cases of the latest technologies in the banking sector	Various business problems are solved by using the latest technological trends and big data analytics in the banking industry	Big data analytics for the banking industry

Table 4.
Finance sector review summary.

2.4 Review of agriculture sector

BDI and data analytics concepts are emphasized by Nabrzyski et al. [16]. This proposed solution incorporates the execution of complex queries on various datasets. These data sets contain the layers of raster and geospatial data. Kim and Tam (2020) have proposed a data integration estimator [17]. This is a classification technique with non-parametric and overlapping units which recognizes and corrects misclassification errors. Saggi and Jain (2018) have reported a data analytics solution for organizations [18]. This solution performs an exhaustive realistic analysis. The components of BDI application platforms are discussed. Authors have thrown light on past, current research issues and future directions.

Ribarics (2016) explained the importance of big data in agricultural sector [19]. The author has highlighted the need for using technological innovations in farming. Sarker et al. (2020) discussed the impact of BDI in digital farming [20].

The study results showed that big data analytics helps the farmers in crop management and yield forecasting. This study also revealed that BDI in farming is not fully established. **Table 5** presents a summary of highlights from the agriculture sector review.

2.5 Review of literature on BDI in smart cities

Kaur and Kushwaha (2018) were motivated by different applications of BDI and IoT integration in smart cities [21]. The earlier researchers reviewed the critical data analysis issues. Huang et al. (2014) have proposed HiperFuse solution for addressing BDI challenges and automating the BDI process [22]. Nuaimi et al. (2015) reviewed the prospects, issues and advantages of BDI in smart cities. This study discussed the BDI challenges faced in smart cities [23]. Gomes et al. (2016)

Ref. No.	Methods Used	Results	Applications
16	Data acquisition and semantic integration, statistical data analysis, data visualization, data query language, and geospatial data techniques	Data integration and big data analytics solution are discussed	Agriculture decision support system. Helps the policymakers to implement restoration strategies
17	Identifying overlapping units, matching variables, and classification methods	Estimation of the missing data stratum, independent probability of sample infinite population	Agricultural census data analysis of Australia for the year 2015–2016
18	Characteristics of BDA, architecture, technologies, the relationship between value creation and BDA, applications	Big data analytics framework for value creation	Smart city, cybersecurity, agriculture and healthcare domains
19	Summary of Oracle's strategic white paper on Big data applications	Big data analytics as technological innovation in farming	Farming and food production
20	The comprehensive review reveals the impact of big data infarming	Farming is not fully equipped with big data technologies.	Big data analytics helps the farmer in crop management and forecasting

Table 5.
Agriculture sector review summary.

demonstrated a smart city project model using BDI solutions in Brazil [24]. This project proposed a model that can be hosted in big data servers.

Alshawish et al. (2016) discussed the role and potential of BDI solutions in smart cities [25]. The authors have explained the complete process of BDI applications in smart cities.

This study has incorporated some real-world examples of smart city components. **Table 6** presents a summary of highlights from the smart cities review.

2.6 Manufacturing

Ahmed et al. (2016) have proposed a Generating Attributes with Rolled Paths (GARP) algorithm that creates a mining table attributes from multiple data sources [26]. The experiments were carried out on the U.S. consumer electric retailer dataset and revealed that classification accuracy was improved by using GAPR. Bennani et al. (2014) have reported a guided BDI solution with Service Level Agreement (SLA) for querying data from multiple clouds [27]. The methodologies and algorithms designed are applied to energy utilization. Product planning, product design, manufacturing and maintenance process are reviewed in terms of concepts and applications. Qi and Tao (2018) provided a 360-degree review of big data in smart manufacturing [28]. Product planning, product design, manufacturing and maintenance processes are reviewed in terms of concepts and applications. Hufnagel et al. (2015) demonstrated a distributed integration model applicable to the manufacturing industry [29]. This research has created the user-oriented integration platform using a modular approach. O'Donovan et al. (2015) reported a detailed review of BDI implementation in the manufacturing sector. This study has provided a detailed review of big data research in manufacturing [30]. **Table 7** presents a summary of highlights from the manufacturing sector review.

Ref. No.	Methods Used	Results	Applications
21	Various technologies for the handling of big data and IoT integration	A new data architecture that supports IoT and other data resources	Critical data analysis solution for IoT and Big Data
22	Data mixing planner, domain-specific data models, robust type inference, and declarative interface	Automates the data integration process and leverages key capabilities	Website visitors income analysis, retail business analytics
23	Literature survey on prospects, issues and advantages of big data technologies in smart cities	Big data applications for smart use of data and operations in smart cities	Effective management of smart city resources
24	Design of smart city project model using big data in Brazil	This software can be used in big data servers	Software for smart city project in Brazil
25	Collecting data from networks, processing data with various stages and visualization data	Big data-driven smart city improves smart city applications	Smart Energy, Smart public safety and Smart traffic systems.

Table 6.
Smart cities sector review summary.

Ref. No.	Methods Used	Results	Applications
26	Automatic generation of discriminant features, aggregation of information from multiple resources	Classification accuracy improvement and discriminant feature generation. Mitigates the impact of class imbalance	Consumer electronic retailer in Circuit City U.S.
27	The economic model of the cloud referred for lookup, aggregation and correlation in SLA data integration, handling SLA interoperability and collaboration	A distributed data as a service for SLA guided data aggregation framework	Energy consumption applications, data integration of political campaign and electronics
28	Compare and contrast of digital twin and big data. Product planning, product design, manufacturing and maintenance process are reviewed in terms of concepts and applications	Digital twin and big data have great significance in smart manufacturing	Smart manufacturing in workshop or factory
29	Featuring missing connection between successful business integration concept and proven graphical description	User-oriented integration platform using a modular approach	Workflows and product life cycles in the manufacturing industry
30	Captured the status of big data research in manufacturing, and compared the secondary research studies	Usage of big data technologies in manufacturing for maintenance and diagnosis	Various manufacturing domain

Table 7.
Manufacturing sector review summary.

2.7 Review of healthcare sector

Hardiman has explored BDI methodologies for Omics data and network algorithm development [31]. The objective was to channel the gap between phenotype and genotype which were not applied earlier. These researchers used spectrometry permitted geneticists, deep sequencing technologies, biostatisticians and biologists. Bhandari et al. have explained HGBEnviroScreen in their paper [32].

This is an EJ mapping tool providing the key services online to local decision-makers and communities. This study has resulted in multiple risk factors leading to the largest vulnerability census tracts. These risk factors lead to natural disaster, social vulnerability and flooding. Shayne et al. have carried out a comprehensive study of integration solutions for big medical data [33]. This study has covered the applications, tools and technologies of BDI in the healthcare domain. Eftekhari et al. have proposed software as a service architecture [34]. This provides backend infrastructure for database access operations on data from different data sources. This methodology was approved with a proof-of-concept prototype developed on the OpenStack cloud architecture. Vidal et al. have presented a knowledge-driven framework [35]. This framework extracts knowledge from short text and unstructured data.

This framework used controlled vocabularies and ontologies to clarify the extracted entities and relations. Husain et al. have reported SOCR data dashboard design, implementation, and testing. SOCR does exploratory questioning of multi-source and heterogeneous and datasets [36]. **Table 8** presents a summary of highlights from the healthcare sector review.

Ref. No.	Methods Used	Results	Applications
31	Network algorithms and Gene ontology path are followed	Chanel the gap between phenotype and genotype on a scale using high throughput techniques	Biomedical, clinical and Omics data integration
32	Five domains data collected at HGB region for the year 1990 and designed EJ mapping tool for community online services	Online services for decision-makers and community by EJ mapping tool	Usage of result in a community action plan by community partners
33	Usage of various tools, techniques and applications of data integration in the healthcare domain. Analysis of integration techniques abilities to handle speed, variety and uncertainty.	Strength and weaknesses of various solutions, and its findings	Healthcare big data integration
34	Designing Big data store by collecting data from multiple sources. Web interface and RESTful APIs for the integration of RDBMSs with non-relational databases. The queries on such remote databases by proof of concept.	SaaS framework for integrating multiple data sources performing operations such as data access, querying and visualization	Ad-hoc querying of health care datasets
35	Data integration of multiple data resources, Knowledge-driven framework for data description that uses knowledge graph	Ontologies and unified schema as a knowledge graph for describing integrated data	Discovery of interactions among drugs in treatments with much faster running time prescribed to lung cancer patients
36	Human-machine interface for integration of data from heterogeneous resources in a secure and scalable way	Human-machine interactions customization	Service-oriented infrastructure for healthcare data.

Table 8.
Healthcare sector review summary.

2.8 Review of communication sector

Cheng et al. proposed a remote sensing data management system [37]. This system is distributed multisource and followed the MongoDB model. The remote sensing, data integration and access are examined by designing a set of experiments.

Wang et al. have described the major aspects of BDI such as characteristics, advantages, platform architecture, and application areas in telecommunication [38]. This research can be extended by improving multiple levels of protection technologies in the big data platform. Yayah et al. explained a few use cases of machine learning implementation in big data platforms [39]. Scalability and extensibility are the parameters used for the evaluation of BDI technologies. Nwanga et al. studied the impact of big data analytics in mobile phone industry [40]. This study has revealed that BDI solutions and big data analytics has an impact on the growth of the telecommunication industry by adding huge data insights. **Table 9** presents a summary of highlights from the communication sector review.

2.9 Review of supply chain sector

Antonio et al. presented a literature survey of simulation techniques in supply chain risks [41]. The authors have highlighted the significance of BDI in supply chain systems. This analysis has concluded that the problem at hand is simplified without complexity in modeling. This study has complied with industry 4.0 standards. Ostrowski et al. have explored the potential of semantic web technologies by demonstrating a case study in the supply chain [42]. Authors have identified the system for supporting data from multiple sources. This study was carried out by semi-automated mapping using shared domain ontology. Awwad et al. provided a review on applications, advantages and issues of BDI technologies for the supply chain management [43]. The supply chain risk management was carried out using data analytics by making a proactive decision. Lia and Liu have illustrated a data-driven framework for supply chain management [44]. The various circumstances of the supply chain

Ref. No.	Methods Used	Results	Applications
37	Multi-Source BDI framework, Spatial Segmentation Indexing Model, integration based on distributed storage	Scalable storage data integration architecture, latest technical support and development	Professional remote sensing big data
38	Introduced and reviewed major aspects of big data such as characteristics, advantages, platform architecture, and application areas in telecommunication	Internal data applications enhance the efficiency of big data applications. External cooperation provides better services.	Development in telecommunication organization
39	Integrating machine learning tools in the Hadoop platform	Adoption and improvement of big data in the telecommunication industry	Telco, Retail, Financial Services and Energy sector
40	Comprehensive study and analysis of the impact of big data in the mobile industry	Big data analytics has impact on the growth of the telecommunication industry	Growth of the telecommunication industry with help of big data

Table 9.
Communication sector review summary.

Ref. No.	Methods Used	Results	Applications
41	Literature survey of simulation techniques for the analysis and synthesizing risks in supply chains	Analyzed the impact of risks in supply chains	Supply chain management
42	Semantics with annotation in Ontology federation process, Integration of data from multiple resources using shared domain ontology	Integration of data from multiple resources using semi-automated mapping	Supply chain risk detection
43	Detailed review on applications, advantages and issues of big data technologies for supply chain management	Infrastructure and human skillset need to be improved, new and effective techniques need to be developed	Supply chain in manufacturing and logistics
44	Design and development of a data-driven framework for supply chain management	Multiple working modes of big data in supply chain management	Power split device in hybrid vehicles
45	A detailed survey of various supply chain operations reference model with opportunities and challenges.	Studies revealed that the supply chain process is having higher importance	The supply chain and manufacturing products

Table 10.
Supply chain sector review summary.

are accommodated by enabling multiple working modes. Benabdellah et al. discussed the impact of big data on supply chain management [45].

The survey of various supply chain operation reference model presented their applications and challenges. **Table 10** presents a summary of highlights from the supply chain sector review.

2.10 Review of research domain

A review by Li presented BDI technology applications for the analysis of Chinese and Russian dance components with modern features [46]. Arputhamary and Arockiam presented the prominence of BDI by identifying the open problems and the same is extended to proceed with future research in the big data environment [47]. Kadadi et al. have surveyed BDI methods and their interoperability [48]. Authors have also explored its usage in big data setup and the corresponding challenges. Ostrowski and Kim have presented a BDI strategy based on ontology [49]. BDI strategy was implemented in Apache Spark prototyping environment that generates ontology versions using rule-based translations. Sottovia et al. have described the Research Alps project pipeline. This project was funded by the EU Commission [50]. They have created an open dataset providing Alpine area research centre details. Portugal et al. have presented a high-level spatial–temporal architectural framework for massive data integration, analysis and provenance management [51]. This methodology was applied for BDI analysis. **Table 11** presents a summary of the highlights of the research review.

2.11 Review of recent advancements in BDI

Large scale implementation of BDI solutions is a very complex and difficult process than automating data transformation processes. To reduce the complexity, the organizations should implement the procedures for data discovery, semantic or

Ref. No.	Methods Used	Results	Applications
46	Big Data Technology	Combination of Chinese and Russian cultural and modern features	Dance elements
47	Importance of BDI issues and challenges are identified	The existing techniques and approaches are inefficient to handle the problems.	Possible research directions
48	Addressing challenges of BDI such as Data accommodation, Data irregularity, Query optimization, Extensibility, ETL processing	Big data integration architecture	BDI within the organization and inter organizations
49	Ontology-based data integration, creation of new ontology versions by using rule-based translation	Multiple data sources 'Semi-automated mapping	Large scale Big data applications
50	M-STEP and entity matching method and functional framework to deal with hierarchical data instances	Open dataset providing Alpine area research centres details	Research Alps project funded by EU commission
51	Domain experts focusing on appropriate analysis steps, high-level models linked with code produces middleware	Model-driven techniques resulting in data integration and analysis	Provenance information

Table 11.
Research domain review summary.

business comprehension of data, metadata management, structured and unstructured data management, and transformation. Integrating unstructured and semi-structured data enables organizations to manage modern data sources containing text, images, and video. A survey was conducted by AtScale Inc. in collaboration with Cloudera and ODPi.org reveals that most of the organizations are selecting multi-cloud strategies for BDI implementation. Data virtualization and data governance are their top priorities [52]. This survey has collected data from 150 data practitioners where the respondents are from multiple industries around the world. The online magazine "Smarter with Gartner" has reported that top ten technology trends in data analytics require essential investments [53].

This article revealed that the combination of machine learning algorithms and data technologies could help the medical and public health experts to discover new possible treatments. The article entitled "2020 CRN Big Data 100" published in "Data Integration Solutions Review" enlists the emerging big data tool vendors [54]. This list provides the details of data integration software, tools, platforms and vendors. A data-driven technique for a hybrid BDI using multilayer perceptron was discussed in this research [55]. A customized multilayer perceptron model was constructed using time-based parameters. The fields applied in optimization analysis are also used in the error matrix through additional neural network model. Research results revealed that this solution captures the variations in state variables. BDI project implementation for COVID-19 analytics was discussed [56]. This project was funded by the European Union research fund. This platform combines information from multiple sources such as world news, social media, published science and health data from healthcare institutions. The project design was co-created with industry, academia, health professionals, and policymakers to align with innovative technologies. This project successfully provides useful and actionable information to public health authorities.

Authors	Year	Study Objectives	Advantages	Disadvantages	1	2	3	4	5	6
Fikri et al	2019	To get solutions for financial data real-time integration issues and interpretation of data	This real-time data integration solution resolves earlier issues of classic ETL tools	This solution cannot be integrated with a hot production setting	Y	Y	Y	Y	N	Y
Cheng et al	2020	To design BDI distributed architecture for remote sensing data, where these data from multiple sources	Improvement in performance with a distributed architecture for remote sensing data	Time and resource complexity to handle various pre-processing steps in data integration	Y	Y	Y	N	N	Y
Authors	Year	Study Objectives	Advantages	Disadvantages	1	2	3	4	5	6
Bhandari et al	2020	To develop EJ screening, an adaptable and community-based tool for the region Houston Galveston Brazoria	Risk factor identification and understanding among the communities	reducing environmental disparities and improving their health and well-being	Y	Y	N	N	N	Y
Vieira et al	2020	To conduct a literature survey of simulation methods used for handling risks in the supply chain with an emphasis on data integration	Simplification of the problem in the absence of complex modeling	It is required to focus on supply chain real cases	N	Y	Y	Y	N	Y
Stonebraker et al	2018	To explore issues of BDI related to scalability in enterprises at Tamr region	Automation by machine learning and rule-based approach for augmenting	Involves high cost for domain experts, shortage of training data	N	Y	Y	Y	N	Y
Ahmed et al	2016	To aggregate data from local and external resources, to generate mining table from these, automatic generation of potential discriminant features	Classification accuracy improvement and thus mitigates the impact of class imbalance	Time complexity is linear and it is required to reduce computation time with an efficient method	Y	Y	N	Y	N	Y

Authors	Year	Study Objectives	Advantages	Disadvantages	1	2	3	4	5	6
Bansal	2014	BDI by designing a Semantic Extract-Transform-Load architecture	Publishing semantic data on the internet and thus contribute to the web of data	It is required to understand the heterogeneity of data i.e. ontology engineering	Y	Y	Y	N	N	Y
Dhayne et al	2019	To study healthcare data integration methods, tools, and applications	Wide range of healthcare data integration concepts, techniques and tools are covered	Data integration in the healthcare sector could not be done efficiently using traditional way	Y	Y	Y	Y	Y	Y
Sazontev et al	2019	To develop a prototype of a big data integration system	Useful for e-commerce data integration domain	Lacks in methods for schema alignment	Y	Y	N	N	N	Y
Chen et al	2015	To accomplish data integration of back-end datasets in a complete manner	Data movement is faster than that of Spark thus achieved optimization	Integration of more Spark modules is not supported	Y	Y	N	N	N	Y
Zheng et al	2015	To summarize categories and its subcategories of data integration techniques	Extensive details of big data integration solutions for communities	Since BDI methods behave differently in different applications it's difficult to select the best data fusion technique	Y	Y	Y	Y	N	Y
Huang et al.	2014	To automate the data integration process and leverage key capabilities	A more agile process for compelled analysis by generating a subset of data	HiperFuse modules are implemented separately yet to be integrated	Y	Y	N	N	N	Y
Portugal et al	2016	To perform spatial and temporal data analysis for assisting domain experts	High-level representations by domain specific languages, data analysis and integration by model-driven techniques	provenance technologies need to be used in related spatial–temporal approaches	Y	Y	Y	N	N	Y

Authors	Year	Study Objectives	Advantages	Disadvantages	1	2	3	4	5	6
Saggi et al	2018	To bridge the gap by big data processing and analytics	A comprehensive review of big data projects in terms of analytics, management, and machine learning	It is required to carry out empirical research based on qualitative and quantitative methods	Y	Y	N	Y	N	Y
Kim et al	2020	Survey sample data approach to handle big data integration	Recognition of overlapping units and correction of misclassification errors	Statistical inference variance estimation with non-parametric propensity score tuning is not covered	Y	Y	N	Y	N	Y

Table 12.
Comparative Analysis of curated survey papers with specific parameters.

2.12 Comparative analysis of survey papers with specific parameters

Authors have selected fifteen BDI survey papers for comparative study. The specific feature parameters such as 1. Architecture, 2. Applications, 3. Open Issues and Challenges 4. Taxonomy, 5. Security, 6. Future Directions are used for comparing these survey papers. The Comparative analysis results are shown in **Table 12**.

3. The architecture of the BDI ecosystem

The outline architecture of the BDI ecosystem is shown in **Figure 5**. This architecture has four major components. These components are Data Sources, Data Operations, Virtual Databases and Business Intelligence. This architecture also shows the operations performed by each of these components. The business Big data would be collected from various distributed sources in different formats and sizes in Data Sources component. The Data Operations component shows the different operations which are performed on this heterogeneous Big data.

The Big data gathered from various types of physically distributed databases are integrated to form a unified logical virtual database. Business intelligence information is extracted from this virtual data source by performing the operations stated in Business Intelligence Component. This intelligent information would be used for real-time intelligent business decision-making process across the organization.
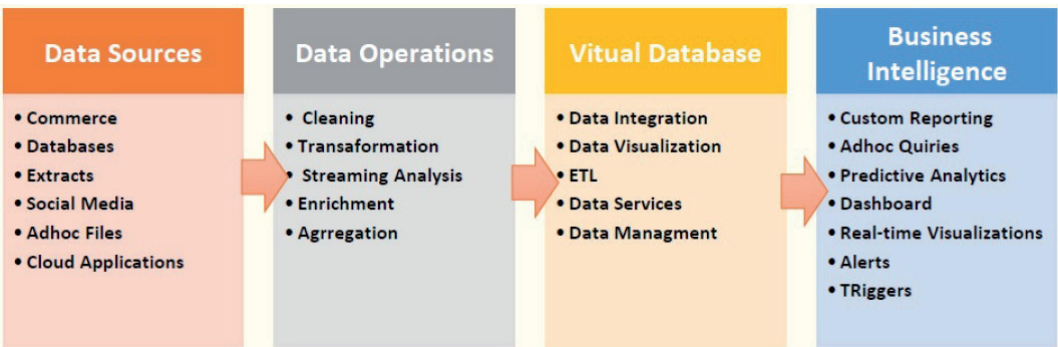

Figure 5.
The architecture of the BDI ecosystem.

4. BDI research issues, challenges and future directions

The research issues, challenges and future directions related to BDI implementation are discussed in the following sections.

4.1 BDI research issues

1. Scalability - Scalable architectures for parallel big data processing

2. Real-time big data analytics - Stream big data processing of text, image, and video

3. Deployment of the IoMT, IoT and CCTVs systems in smart environments would capture big data continuously. Processing multimedia big data in real-time with low latency and high accuracy

4. The balancing of big data processing load at the edges and distributed to the hybrid cloud securely

5. Implementing real-time, complex big data analytics in the cloud by reducing the cost of operations

6. Ensuring authorization, authentication, security and privacy at the edges and cloud.

7. Efficient storage and transfer of big data in real-time

8. Efficient modeling of uncertainty with unlabeled big data

9. Management of graphical big databases

10. Social media analytics using efficient graphical processing.

11. Quantum computing for big data analytics

12. Building context-sensitive large scale systems

4.2 BDI challenges

1. Extracting actionable information from BDI solutions

2. Synchronization of data across heterogeneous data sources

3. Lack of comprehension and management of uncertainty

4. Effective anonymization of sensitive fields in the largescale data systems

5. Support for scalable privacy preservation during BDI processing

6. Generating process models that learn with a smaller number of data samples

7. Building context-sensitive large scale systems:

8. BDI Talent shortage

4.3 BDI trends

1. Everyone is adopting Software as a Service (SAAS)

2. Self-service has evolved to self-sufficiency

3. Shared data, visualizations and storytelling are consumed by all

4. Now constant updating of business-ready data is very vitaa

5. Support for advanced analytics with different perspectives

6. It is critical to gather and create alternative big data

7. Every business is undergoing re-engineering process

8. The measures for competition, surveillance and security are constantly redefined

9. Collaboration has to coalesce earlier in the chain

10. The great digital switch may force a generational shift in analytics.

4.4 BDI advantages

1. Improved e-commerce sales and operations efficiency

2. Creating efficient marketing strategies

3. Increased security enforcement

4. Improving fraud prevention;

5. Enhancing user experience

6. Increased profits

5. BDI organizational case studies

Authors have presented a set of real-life case studies of BDI solutions implemented successfully across the business domains in organizations. **Figure 6** shows the domains and tolls used in that domain for illustration.

5.1 Walgreens Boots Alliance Company

It is a global leader in retail and wholesale pharmacy business operating in the U.S. and Europe and has more than 170 successful business years of serving humanity. Walgreens Boots Alliance, Inc., declared its IT collaboration with Microsoft and Adobe to introduce a world's best digital platform for enhanced customer experience and data insights to offer truly customized healthcare, adhere to their healthcare plans and shopping services as stated by their global chief marketing officer Vineet Mehra. The BDI systems can manage 7.5 billion medical transactions 100 million citizens providing a singular, unified view of the customer information about demographics, registration, diagnoses, procedures, and data from managed-care plans.

This BDI digital platform helps the customers to access key services of pharmacy, beauty and other categories on daily basis. Data security and privacy are important principles in the design of Microsoft's trusted cloud platform. Walgreens has introduced personally customized prescription understanding for patients at Walgreens, Dynamics 365 Customer Insights would serve as WBA's Customer Data Platform (CDP) provided by Microsoft. CDP provides a unified, 360-degree perceptions of the customer and reveals the details to leverage personal experience. Adobe's Customer Experience Management (CXM) solutions leverage Walgreens to offer supreme customer experience, with end-to-end platform for analysis, managing content, customization, campaign composition and many more. Walgreens Company also extends collaboration with Tata Consultancy Services to build highly scalable, maintainable and world-class unified IT operating platform to enable digital transformation, innovation and automation of services offerings. Walgreens Boots Alliance also collaborates with Hortonworks to offer excellent customer satisfaction.
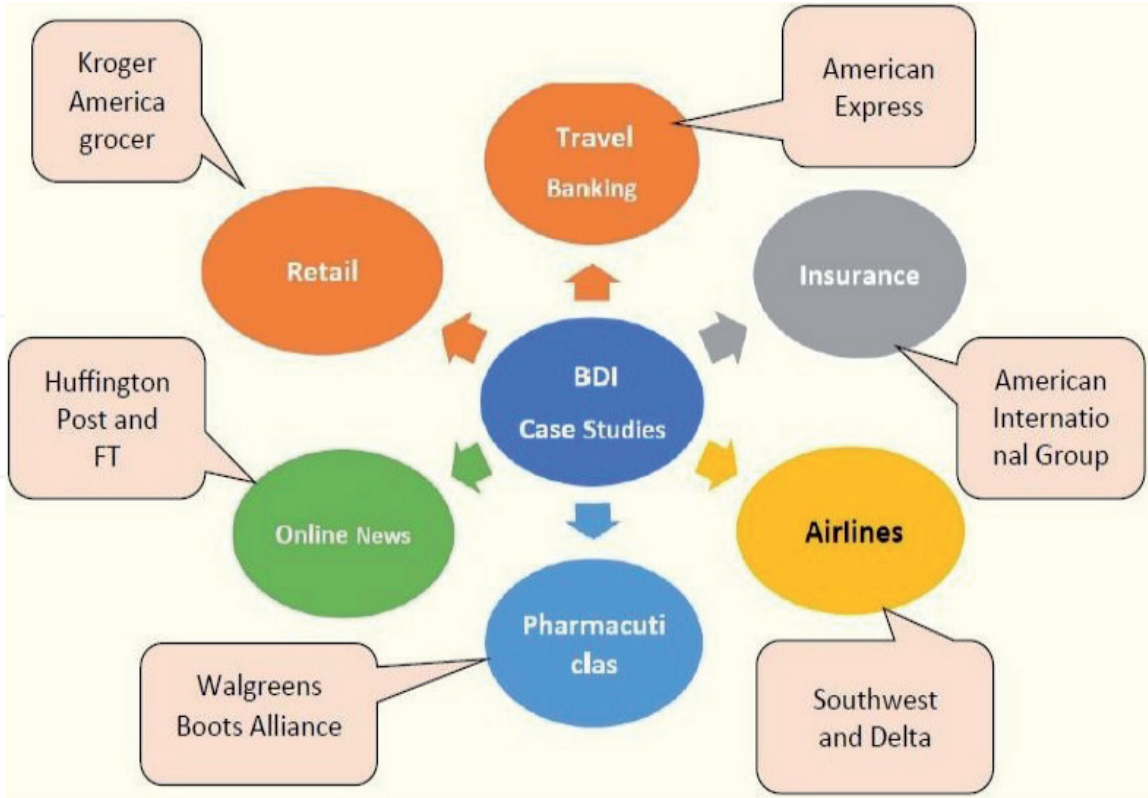
Figure 6.
Case study domains.

5.2 The American International Group (AIG)

AIG Data Services Pvt. Ltd. is a 100% owned subsidiary of American International Group Inc. It is a Fortune 500 company with revenues of the US $70 billion. AIG drives the best decision-making through BDI solution sutilizing business and customer big data across 130+ countries and 64,000 employees which is ever-growing. AIG has implemented sophisticated prediction models with 115 variables to analyze the past business transactions to forecast the potential trends. AIG identified 24% accounts in the Australian market that are about close in next four-month time. AIG has applied BDI tools and visualization systems to discover the frauds by detecting the false claims and adjuster handwritten notes to detect probable frauds. These tools offer insights into insurance claims and enhance machine learning algorithms. AIG creates data profiles and assesses vital data elements against pre-defined data quality standards on important business data for important applications. Today big data is distributed across the globe and facts are available across multiple sources. The team responsible for data sourcing uses ETL tools to provide a unified virtual version of these facts collected from various data sources. AIG has implemented Netezza and data virtualization technologies on Cisco Information Server. AIG also utilizes Hadoop, R, Python, SAS and other open-sourced/licensed tools to implement BDI solutions to beat the competitors. AIG uses tools such as QlikView, Tableau, Cognos and Micro strategy for data visualization.

5.3 Kroger - America's grocer company

Kroger has nearly 2,800 stores in 35 states under twenty-four banners with and annual sales exceeding 121.1 billion. Kroger today ranks as one of the world's largest retailers. Kroger with its joint venture with Dunnhumby is leveraging BDI solutions. Dunnhumby is a technology solutions provider company for retail industry.

These solutions are implemented using the latest techniques, algorithms, procedures and applications. The Kroger company gathers and processes the data from about 770 million consumers. Kroger has implemented BDI solutions for extracting more actionable information for profitability, customer loyalty. Kroger claims that 95% of sales are from the loyalty card. Kroger achieved about $12 billion in revenues by BDI implementing and analytics solutions since 2005.

5.4 Southwest and Delta Airline company

This company has encashed on customer loyalty and relationships by providing boundless service through social channels and other data exchange mechanisms. Southwest utilizes speech analytics to help and enhance the exchanges between service professionals and customers. Southwest applied BDI solutions to understand customer online behavior and activities, increasing offers for customers and driving growth in customer satisfaction year after year. Delta has applied BDI solutions to support most painful travel condition that results in lost baggage. This company tracks the data about baggage and became the first airline company to permit customers to trace their baggage from smartphones. This company checks about 130 million baggage every. Delta is branding its self as a customer friendly services by permitting customers to download their apps over 11 million times and provides best customers with baggage secure services company.

5.5 Huffington Post and FT is an online news service company

This company has become number one online news site in the United States. According to this report, the company's leadership believes in running the business based on big data. This involves enhancing the user experience in real-time through recommendations, moderation, social trends, and personalization. This company optimizes its portal in many ways, and its analytics platform powers the entire analytical process. Huffington Post utilizes data to comprehend and serve the customers well, make targeted advertising, and design innovative products based on information gathered. Their CEO informed that BDI solutions have transformed its business by intelligent and real-time decision making. This company utilized many data points to enhance relevance in their communications, analyze customer content preferences, and personalize the content all to keep traffic and visitors always. The BDI also benefits the company to comprehend the time of day consumption based on both mobile channels and PC.

6. Results discussion

This chapter reviewed the literature on BDI tools and applications in diverse industries and presented the highlights from each domain. All most all organizations are gathering a huge quantity of big data in real-time, Online and offline modes. Managing, real-time processing this big data to extract useful business information for intelligent decision making is the real challenge. The big data processing systems are empowered by big data integration and analytics platforms. BDI systems are facing the challenges in integrating and synchronization of heterogeneous big data from multiple distributed sources. The lack of comprehension and management of uncertainty in big data is another challenge faced in the big data processing. BDI processing should ensure context-sensitivity and extracting the semantics in the distributed data processing. Research on designing effective machine and deep learning algorithms is going on in the BDI domain. BDI Processing uses Hadoop environment with HDFS,

Spark computation model with a Hive database as a distributed data warehouse. The use of Apache Spark enhances short time frames for quality and availability reporting. BDI processing involves data acquisition, semantic integration, statistical data analysis, data visualization, data query language, geospatial data techniques. Big data analytics framework enables us to create business value.

The economic model of the cloud promotes BDI processing by providing online services for decision-makers and the business community. Usage of various tools, techniques and applications of BDI leverages the ability to handle speed, variety and uncertainty. Knowledge-driven framework for BDI describes a knowledge graph. Human-machine interface for BDI integrates data from heterogeneous resources in a secure and scalable way. Ontologies and unified schema as a knowledge graph for describing integrated data. Multi-Source BDI is a framework for integrating data in the distributed storage environment. Authors have discussed BDI research issues and challenge data accommodation, data irregularity, query optimization, extensibility, ETL processing. Remote sensing data. Real-time big data analytics processes stream of big text, image, and videos generated from IoMT, IoT and CCTVs systems. Implementing real-time, complex big data analytics in the cloud using BDI process reduces the cost of operations. This paper discussed five case studies of BDI applications implemeneted in the in world-class organizations.

7. Conclusion

This chapter discussed the importance of BDI process implemented in diverse organizations for providing valuable insights into business data. These insights into the data enable the manager to take intelligent and well-informed rational decisions. An extensive study of literature on BDI applications deployed in diverse domains across the world was carried out and highlights are discussed. The intelligent and autonomous BDI systems are designed using AI, Blockchain, Big data, 5G, Fog and cloud technologies. The comparative analysis of specific parameters was carriedout on curated to survey papers to identify the research gaps and future opportunities in the BDI domain. The five case studies from fortune 500 companies have discussed the insights about how BDI is empowering business decision making leveraging quality, trust, security, flexibility, efficiency and also reduce the cost of operations. The authors attempted to provide a holistic view of BDI concepts and applications. Authors concluded that BDI plays a vital role in the diverse organizations at present and in near future also.

Author details

Sreekantha Desai Karanam[1*], Rajani Sudhir Kamath[2], Raja Vittal Rao Kulkarni[2] and Bantwal Hebbal Sinakatte Karthik Pai[1]

1 NMAM Institute of Technology, Nitte, Karnataka, India

2 CSIBER, Kolhapur, Maharastra, India

*Address all correspondence to: sreekantha@nitte.edu.in

IntechOpen

References

[1] Hasliza, N., Hassana, M., Ahmada, K. & Salehuddina, H. (2020). Diagnosing the Issues and Challenges in Data Integration Implementation in Public Sector, International Journal Advanced Science Engineering Information Technology, 10(2).

[2] Zhang, Y. (2020). The Integration of Professional Ethics of Modern Etiquette Students under the Background of Big Data, Journal of Physics: Conference Series 1574.

[3] Bansal, S. K. (2014). Towards a Semantic Extract-Transform-Load (ETL) framework for Big Data Integration, IEEE International Congress on Big Data, 978-1-4799-5057-7/14 © 2014 IEEE, DOI 10.1109/BigData.Congress.2014.82

[4] Zheng, Y. (2015). Methodologies for Cross-Domain Data Fusion: An Overview. IEEE Transactions On Big Data, 1(1).

[5] Munné R. (2016). Big Data in the Public Sector. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_11.

[6] Camargo-Perez, J. A., Puentes-Velasquez, A. M., & Sanchez-Perilla, A. L. (2019). Integration of big data in small and medium organizations: Business intelligence and cloud computing, J. Phys.: Conf. Ser. 1388 012029.

[7] Stonebraker, M., & Ilyas, I. F. (2018). Data Integration: The Current Status and the Way Forward, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.

[8] Sazontev, V., & Stupnikov, S. (2019). An Extensible Approach for Materialized Big Data Integration in Distributed Computation Environments, Ivannikov Memorial Workshop (IVMEM), 978-1-7281-4623-2/19/ ©2019 IEEE DOI 10.1109/IVMEM.2019.00011

[9] Alsghaier, H., Akour, M., Shehabat, I., & Aldiabat, S. (2017). The Importance of Big Data Analytics in Business: A Case Study. American Journal of Software Engineering and Applications, 6(4), 111-115.

[10] Alam, J. R., Sajid, A., Talib, R., & Niaz, M. (2014). A Review on the Role of Big Data in Business. International Journal of Computer Science and Mobile Computing, 3(4), 446-453.

[11] Fikri, N., Rida, M., Abghour, N., Moussaid, K., & Omri, A. I. (2019). An adaptive and real-time based architecture for financial data integration. Journal of Big Data, 6(97).

[12] Bucea-Manea-Tonis, R. (2018). Deductive systems for Big data integration, Journal of Economic Development, Environment and People, 7(1).

[13] Chen, W., Wang, R., Wu, R., Tang, L., & Fan, J. (2016). Multi-source and Heterogeneous Data Integration Model for Big Data Analytics in Power DCS [Paper Presentation]. International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery.

[14] Hussain K., Prieto E. (2016). Big Data in the Finance and Insurance Sectors. In: Cavanillas J., Curry E., Wahlster W. (eds) New Horizons for a Data-Driven Economy. Springer, Cham. https://doi.org/10.1007/978-3-319-21569-3_12

[15] Avi V., Kamaruddin S. (2017). Big Data Analytics Enabled Smart Financial Services: Opportunities and

Challenges. In: Reddy P., Sureka A., Chakravarthy S., Bhalla S. (eds) Big Data Analytics. BDA 2017. Lecture Notes in Computer Science, vol 10721. Springer, Cham. https://doi.org/10.1007/978-3-319-72413-3_2

[16] Nabrzyski, J., Liu, C., Vardaman, C., Gesing, S., & Budhatoki, M. (2014). Agriculture Data for All - Integrated Tools for Agriculture Data Integration, Analytics and Sharing. IEEE International Congress on Big Data. 978-1-4799-5057-7/14 © 2014 IEEE DOI 10.1109/BigData.Congress.2014.117

[17] Kim, J. K., & Tam, S. (2020). Data integration by combining big data and survey sample data for finite population inference. arXiv:2003.12156v3.

[18] Saggi, M. K., & Jain, S. (2018). A survey towards the integration of big data analytics to big insights for value-creation. Information Processing & Management, 54.

[19] Ribarics, P. (2016). Big Data and its impact on agriculture. Eco cycles, 2(1), 33-34.

[20] Sarker, M. N., Islam, M. S., Murmu, H., & Rozario, E. (2020). Role of Big Data on Digital Farming. International Journal of Scientific & Technology Research, 9(04).

[21] Kaur, H., & Kushwaha, A. S. (2018). A Review on Integration of Big Data and IoT. 4th International Conference on Computing Sciences. 978-1-5386-8025-4/18/$31.00 ©2018 IEEE DOI 10.1109/ICCS.2018.00040.

[22] Huang, E., Quiroz, A., & Ceriani, L. (2014). Automating Data Integration with HiperFuse [Paper Presentation] 2014 IEEE International Conference on Big Data

[23] Nuaimi, E. A., Neyadi, H. A., Mohamed, N., & Jaroodi, J. (2015). Applications of big data to smart

cities. Journal of Internet Services and Applications, 6(25).

[24] Gomes, E., Dantas, M. A., Macedo, D. D., Rolt, C. D., Brocardo, M. L., & Foschini, L. (2016). Towards an Infrastructure to Support Big Data for a Smart City Project [Paper Presentation]. 2016 IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Paris, 2016, pp. 107-112, DOI: 10.1109/WETICE.2016.31

[25] Alshawish, r. A., Alfagih, S. M., & Musbah, M. S. (2016). Big data applications in smart cities. 2016 International Conference on Engineering & MIS (ICE), Agadir, 2016, pp. 1-7, DOI: 10.1109/ICEMIS.2016.7745338

[26] Ahmed, F., Samorani, M., Bellinger, C., & Zaiane, O. R. (2016). Advantage of Integration in BigData: Feature Generation in Multi-Relational Databases for Imbalanced Learning, 2016 IEEE International Conference on Big Data (Big Data), 978-1-4673-9005-7/16/$31.00 ©2016 IEEE

[27] Bennani, N., Ghedira-Guegan, C., Musicante, M. A., & Vargas-Solar, G. (2014). SLA-Guided Data Integration on Cloud Environments [Paper Presentation]. 2014 IEEE International Conference on Cloud Computing, Alaska, United States. 934-935.

[28] Qi, Q., & Tao, F. (2018). Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. IEEE Access, 6, 3585-3593.

[29] Hufnagel, J., & Vogel-Heuser, B. (2015). Data integration in manufacturing industry: Model-based integration of data distributed from ERP to PLC [Paper Presentation]. 2015 IEEE 13th International Conference on Industrial Informatics (INDIN),

Cambridge, 2015, pp. 275-281, DOI: 10.1109/INDIN.2015.7281747.

[30] O'Donovan, P., Leahy, K., Bruton, K., & T. J. O'Sullivan. (2015). Journal of Big Data, 2(20). DOI 10.1186/ s40537-015-0028-x

[31] Hardiman, G. (2020). An Introduction to Systems Analytics and Integration of Big Omics Data, Genes, 11(245).

[32] Bhandari, S., Lewis, P., Craft, E., Marvel, s. W., Reif, D. M., & Chiu, W. A. (2020). HGBEnviroScreen: Enabling Community Action through Data Integration in the Houston–Galveston– Brazoria Region, Int J Environ Res Public Health, 17(4): 1130.

[33] Dhayne, H., Haque, R., Kilany, R., & Taher, Y. (2019). In Search of Big Medical Data Integration Solutions - A Comprehensive Survey. IEEE Access, 7.

[34] Eftekhari, A., Zulkernine, F., & Martin, P. (2016). BINARY: A Framework for Big Data Integration for Ad-hoc Querying, 2016 IEEE International Conference on Big Data (Big Data), 978-1-4673-9005-7/16/©2016 IEEE

[35] Vidal, M., & Sakor, A. (2019). Semantic Data Integration Techniques for Transforming Big Biomedical Data into Actionable Knowledge, 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS).

[36] Husain, S., Kalinin, A., Truong, A., & Dinov, D. (2015). SOCR Data Dashboard: An integrated Big Data archive mashing Medicare, labour, census and econometric information. Journal of Big Data, 2(13).

[37] Cheng, Y., Zhou, K., Wang, J., & Yan, J. (2020). Big Earth Observation Data Integration in Remote Sensing Based on a Distributed Spatial Framework. Remote Sens. 12, 972.

[38] Wang, Z., Wei, G., Zhan, Y., & Sun, Y. (2017). Big data in telecommunication operators: data, platform and practices. Journal of Communications and Information Networks, 2(3). DOI: 10.1007/ s41650-017-0010-1

[39] Yayah, F. C., Ghauth, K. I., & Ting, C. (2017). Adopting Big Data Analytics Strategy in the Telecommunication Industry. Journal of Computer Science & Computational Mathematics. 7(3). DOI: 10.20967/jcscm.2017.03.002

[40] Nwanga, M. E., Onwuka, E. N., Aibinu, A. M., & Ubadike, O. C. (2015). Impact of Big Data Analytics to the Nigerian Mobile Phone Industry. Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management Dubai, United Arab Emirates (UAE), March 3-5, 2015.

[41] Antonio, A. C., Luis, M. S., Santos, M. Y., Guilherme, A. B., & Jose, A. O. (2020). Supply chain data integration: A literature review. Journal of Industrial Information Integration 19 100161.

[42] Ostrowski, D., & Kim, M. (2017). Semantic-Based Framework for Big Data Integration [Paper Presentation]. 2017 IEEE 11th International Conference on Semantic Computing

[43] Awwad, M., Kulkarni, P., Bapna, R., & Marathe, A. (2018). Big Data Analytics in Supply Chain: A Literature Review. Proceedings of the International Conference on Industrial Engineering and Operations Management, Washington DC, USA, September 27-29.

[44] Lia, Q., Liu, A. (2019). Big Data-Driven Supply Chain Management, Procedia CIRP 81 ScienceDirect 52nd CIRP Conference on Manufacturing Systems, 1089-1094.

[45] Benabdellah, A. C., Benghabrit, A., Bouhaddou, I., & Zemmouri, E. M. (2016). Big Data for Supply Chain

Management: Opportunities and Challenges. International Journal of Scientific & Engineering Research, 7(11).

[46] Li, J. (2020). Research on the Integration of Chinese and Russian Original Ecological Dance Elements and Modern Elements Based on Computer Big Data Analysis. Journal of Physics: Conference Series 1578.

[47] Arputhamary, B. & Arockiam, L. (2015). Data Integration in Big Data Environment. Bonfring International Journal of Data Mining, 5(1), 1-5.

[48] Kadadi, A., Agrawal, R., Nyamful, C., & Atiq, R. (2014). Challenges of Data Integration and Interoperability in Big Data. 2014 IEEE International Conference on Big Data, 978-1-4799-5666-1/14/$31.00 ©2014 IEEE

[49] Ostrowski, D., Rychtyckyj, N., MacNeille, P., Kim, M. (2016). Integration of Big Data Using Semantic Web Technologies. 2016 IEEE Tenth International Conference on Semantic Computing, 978-1-5090-0662-5/16 © 2016 IEEE DOI 10.1109/ICSC.2016.101

[50] Sottovia, P., Paganelli, M., Guerra, F., & Vincini, M. (2019). Big Data Integration of Heterogeneous Data Sources: the Research Alps CaseStudy. 2019 IEEE International Congress on Big Data (BigData Congress), 978-1-7281-2772-9/19 ©2019 IEEE DOI 10.1109/BigDataCongress.2019.00027

[51] Portugal, I., David, P. A., & Cowan, D. (2016). Towards a Provenance-Aware Spatial-Temporal Architectural Framework for Massive Data Integration and Analysis, 2016 IEEE International Conference on Big Data (Big Data).

[52] AtScale Inc, Big Data & Analytics Maturity 2020 Survey Results, https://www.atscale.com/wp-content/uploads/2020/02/2020-Big-Data-Analytics-Survey-Results.pdf

[53] Laurence Goasduff, Gartner Top 10 Trends in Data and Analytics for 2020, https://www.gartner.com/smarterwithgartner/gartner-top-10-trends-in-data-and-analytics-for-2020/ posted on 19th October 2020 and retrieved on 24th December 2020

[54] Timothy King, Data Integration Solutions News, 2020 CRN Big Data 100: 14 Data Integration Tools Companies to Consider, https://solutionsreview.com/data-integration/2020-crn-big-data-100-data-integration-companies-to-consider/ posted on 30th April 2020 and retrieved on 24th December 2020

[55] Lilan Huang, Hongze Leng, Xiaoyong Li, Kaijun Ren, Junqiang Song, Dongzi Wang, A Data-Driven Method for Hybrid Data Assimilation with Multilayer Perceptron, Big Data Research 23 (2021) 10017, https://doi.org/10.1016/j.bdr.2020.100179

[56] Joao Pita Costa, Marko Grobelnik, Flavio Fuart, and Luka Stopar, Meaningful Big Data Integration for a Global COVID-19 Strategy, IEEE Computational Intelligence Magazine, November 2020