

This is the author's version of a work that was submitted/accepted for publication in the following source:

**Debortoli, S., Müller, O., & vom Brocke, J. (forthcoming).  
Comparing Business Intelligence and Big Data Skills: A Text  
Mining Study Using Job Advertisements. *Business &  
Information Systems Engineering.***

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source.

The final publication will be available at

<http://link.springer.com/journal/12599>

**Title [English]**

Comparing Business Intelligence and Big Data Skills: A Text Mining Study Using Job Advertisements

**Abstract [English]**

While many studies on big data analytics describe the data deluge and potential applications for such analytics, the required skill set for dealing with big data has not yet been studied empirically. The difference between big data (BD) and traditional business intelligence (BI) is also heavily discussed among practitioners and scholars. We conduct a latent semantic analysis (LSA) on job advertisements harvested from the online employment platform monster.com to extract information about the knowledge and skill requirements for BD and BI professionals. By analyzing and interpreting the statistical results of the LSA, we develop a competency taxonomy for big data and business intelligence. Our major findings are that (1) business knowledge is as important as technical skills for working successfully on BI and BD initiatives; (2) BI competency is characterized by skills related to commercial products of large software vendors, whereas BD jobs ask for strong software development and statistical skills; (3) the demand for BI competencies is still far bigger than the demand for BD competencies; and (4) BD initiatives are currently much more human-capital-intensive than BI projects are. Our findings can guide individual professionals, organizations, and academic institutions in assessing and advancing their BD and BI competencies.

**Keywords (up to 8) [English]**

Big Data, Business Intelligence, Competencies, Latent Semantic Analysis, Text Mining

## 1 Introduction

Big data and big data analytics are among today's most frequently discussed topics in research and practice (Buhl et al. 2013). In loose terms, big data refers to data sets that are too large and complex to be processed using traditional storage (e.g., relational database management systems) and analysis technologies (e.g., packaged software for statistical analysis). More specifically, researchers and practitioners use the term "big data" to refer to the ongoing expansion of data in terms of volume, variety, velocity (Laney 2001), and veracity (IBM 2012).

Given the current excitement around big data, critical voices question whether big data is "really something new or [...] just new wine in old bottles" (Buhl et al. 2013) or postulate that we should "forget big data [because] small data is the real revolution" (Pollock 2013). Others, such as Chen, Chiang, and Storey (2012) and Golden (2013), argue that big data is not a revolution but an evolution of traditional business intelligence (BI). According to this view, big data analytics widen the scope of BI, which focuses on integrating and reporting structured data residing in company-internal databases, by seeking to extract value from semi-structured and unstructured data that originates in data sources like the web, mobile devices, and sensor networks that are external to the company.

Big data offers enormous opportunities for businesses but also poses many challenges (Buhl 2013). A survey of nearly 3,000 executives, managers, and analysts from more than 30 industries and 100 countries conducted by MIT Sloan Management Review and the IBM Institute for Business Value finds that top-performing organizations use analytics five times more often than lower performers do (LaValle et al. 2011), yet not all corporate big data initiatives are successful. Research shows that "inadequate staffing and skills are the leading barriers to Big Data Analytics" (Russom 2011), and a study by the McKinsey Global Institute states that "[t]he United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings" (Manyika et al., 2011, p. 3).

Given these figures, we academics have to ask ourselves to what degree current research agendas and curricula satisfy industry's growing demand for competence in the areas of big data and analytics. Against this background, the objective of this paper is to clarify the competency requirements of the emerging field of big data (BD) and compare them to the requirements of the established field of BI. In particular, we seek to (1) identify and catego-

alize competency requirements for BD professionals and BI professionals from a practitioner's point of view and (2) highlight these requirements' similarities and differences.

The current literature contains only a few contributions on the topic of BI and BD competencies, so we collected and analyzed empirical data from the BI and BD job market. Following the logic of extant studies on information systems competency requirements (e.g., Gallivan et al. 2004; Litecky and Aken 2010; Todd et al. 1995), we used online job advertisements as a data source and performed a quantitative content analysis of 1,357 BI-related and 450 BD-related job advertisements using a text-mining technique called latent semantic analysis (LSA).

Our analysis revealed fifteen distinct areas of competency for BI professionals and fifteen distinct areas of competency for BD professional. On the most abstract level, these areas of competency can be classified into business competencies and IT competencies. The business competencies can be further sub-divided into management and domain competencies, and the IT competencies can be further sub-divided into methodological, conceptual, and product-specific competencies. Comparing and contrasting the competency requirements for BI and BD professionals shows areas of overlap, especially regarding IT concepts and methods and the business domain, as well as clear differences when it comes to IT competencies. While BI requires skills in the area of commercial software platforms, BD largely relies on software engineering, statistics skills, and open-source products.

Our empirically grounded frameworks of BI and BD competencies contribute to the IS body of knowledge by (1) helping professionals to assess and advance their individual competencies, (2) guiding organizations in composing effective portfolios of BI and BD professionals, and (3) informing the development of academic and professional education programs.

The remainder of this paper is structured as follows. The next section provides research background on the topic of BI and BD competencies. Then we introduce our methodology and explain our data-collection and analysis processes. Next, we present our results and discuss our findings against the background of related work. We close by pointing out the limitations of our work and implications for future research.

## **2 Research Background**

The resource-based view (RBV) of the firm, especially the framework by Melville et al. (2004), can be used to evaluate BI / BD implementations' generation of business value and to

assess which resources and competencies are required and may lead to competitive advantage. In the focal firm, IT business value is generated by the deployment of IT and complementary organizational resources (Melville et al. 2004). However, IT affects organizational performance only via intermediate business processes. Melville et al. (2004) operationalize IT based on Barney's (1991) classification of firm resources into physical capital (technological IT resources or TIR, i.e., infrastructure and business applications), human capital (human IT resources or HIR, i.e., technical skills and managerial skills), and organizational capital resources (e.g., organizational structures, policies and rules, workplace practices, culture). Section 2.1 elaborates on the technological IT resources associated with BI and BD, Sections 2.2 and 2.3 discuss required human IT resources, and Section 2.4 addresses complementary organizational capital resources.

## 2.1 Business Intelligence and Big Data

Howard Dresner of the Gartner Group introduced the term "business intelligence" in 1989, describing "a set of concepts and methods to improve business decision making by using fact-based support systems" (Power 2007). The first productive BI systems were implemented at large consumer goods manufacturers like Procter & Gamble and retailers like Wal-Mart for the purpose of analyzing sales data (Power 2007). Although Dresner's original definition of BI, as well as more recent definitions from analysts like Gartner, Forrester, and TDWI, are broad in scope, most practitioners associate with the term a narrow set of capabilities, such as extraction, transformation, and loading (ETL); data warehousing; on-line analytical processing (OLAP); and reporting (Davenport 2006). The focus of these traditional BI solutions is on analyzing historical data in order to answer questions like "how much did we sell in a certain region?" and "how much profit did we make last quarter?"

At the end of the 1990s, the term "big data" started to appear in the scientific literature, referring to data sets that were too large to fit into main memory or even local disks (Cox and Ellsworth 1997; Forbes 2013). The first publications about big data originated from the field of scientific computing, but in 2001 Doug Laney, an analyst with the Meta Group, transferred the concept to the business domain and coined the term "the 3Vs" to stand for volume, velocity, and variety, which quickly became the constituting dimensions of big data (Laney 2001). After the mid-2000s, fueled by Davenport's (2006) seminal article "Competing on Analytics," businesses became increasingly interested in big data, and the focus shifted from technical issues around the storage of big data to its analysis. Internet-based businesses like Google,

Amazon, and Facebook were among the first to exploit big data by applying sophisticated data mining and machine learning techniques. What differentiates today's big data analytics applications from traditional business intelligence applications is not only the breadth and depth of the data processed, but also the types of questions they answer. While BI traditionally focuses on using a consistent set of metrics to measure past business performance (Davenport 2006), big data applications emphasize exploration, discovery, and prediction. As Dhar (2013) states, "Big data makes it feasible for a machine to ask and validate interesting questions humans might not consider."

## 2.2 Business Intelligence Competencies

As we found no literature that studies individual BI competencies, we gained an overview of individual BI competency requirements by consulting extant work on BI maturity/capability models, reviews of the BI literature, and panel reports.

Both research and practice have engaged in developing BI maturity/capability models. (For an overview, see, e.g., Russell, Haddad, Bruni, & Granger, 2010). The general purpose of such models is to systematize organizational capabilities and outline pathways for advancing them. Models that originate from industry include the TDWI Business Intelligence Maturity Model (Eckerson 2004), Gartner's Maturity Model for Business Intelligence and Performance Management (Hostmann and Hagerty 2010), Gartner's Magic Quadrant for Business Intelligence Platforms (Schlegel et al. 2013), and Logica's Capability/Maturity Model (Van Roekel et al. 2009). Lahrman et al. (2011), Dinter (2012), and Cates et al. (2005) provide examples of academic BI maturity models. Industry maturity models tend to focus on technological capabilities that BI platforms should provide (Russell et al. 2010). For example, Gartner lists thirteen essential capabilities, including reporting, OLAP, and visualization (Schlegel et al. 2013). Such functional IT capabilities provide some guidance for assessing and developing individual-level BI competencies but largely neglect the business-related aspects of BI, such as project management and domain skills. By contrast, the academic models provide a high-level view of strategic BI capabilities like architecture planning, IT-business alignment, and generation of business value. While these topics are key to engaging effectively in BI on an organizational level, we believe that they are too abstract to be useful in assessing and developing individual-level BI competencies.

The purpose of literature reviews is to analyze and synthesize the academic body of knowledge, so it is reasonable to expect that reviews can provide insight into competency requirements by, for example, outlining curricula. We identified one review in the area of BI that explicitly comments on aspects of education. Based on market research results from Gartner, Chen et al. (2012) perform a bibliometric study of academic and industry publications on business intelligence and analytics and structured the business intelligence and analytics (BI&A) discipline into three evolutionary waves—BI&A 1.0 (database-based, structured content), BI&A 2.0 (web-based, unstructured content), and BI&A 3.0 (mobile and sensor-based content)—and five emerging research areas—big data analytics, text analytics, web analytics, network analytics, and mobile analytics. Chen et al. (2012) also outline and map the competency requirements for each of these fields and advocate that higher education should consider these competencies in their curricula. Examples of the competencies Chen et al. (2012) name include relational database management systems (RDBMS), data warehousing, ETL, data mining, statistical analysis, web crawling, recommender systems, social network theories, smartphone platforms, machine learning, process mining, in-memory DBMS, cloud computing, sentiment analysis, and web visualization.

Wixom et al.'s (2011) panel report notes that industry trends raise concerns that “academia may be behind the curve in delivering effective Business Intelligence programs and course offerings to students.” Based on surveys conducted at BI practitioner events, Wixom et al. (2011) formulate four academic BI best practices that would close the gap between BI market needs and the content of IS education programs: (1) provide a broader range of BI skills, (2) take an interdisciplinary approach to BI programs, (3) develop reusable teaching resources, and (4) align with practice. Besides arguing for the need for technical skills, Wixom et al. (2011) argue that a deep understanding of business subjects (e.g., finance, marketing) and strong communication skills are required.

### 2.3 Big Data Competencies

No scientific literature on the topic of BD competences has yet been published, although a number of articles and web resources anecdotally describe the profile of BD specialists or similar jobs, such as those of data scientists.

In an influential *Harvard Business Review* article, Davenport and Patil (2012) describe a data scientist as “a hybrid of data hacker, analyst, communicator, and trusted adviser” (p. 73) and

call the job of the data scientist “the sexiest job of the 21st century” (p. 70). Likewise, Hammerbacher, who created the first data science team at Facebook, portrays a data scientist as “a team member [who] could author a multistage processing pipeline in Python, design a hypothesis test, perform a regression analysis over data samples with R, design and implement an algorithm for some data-intensive product or service in Hadoop, or communicate the results of our analyses to other members of the organization” (as cited in Loukides 2012).

These characterizations seem to call for a hybrid of a computer scientist and statistician, yet many more business-related authors state that, in the world of big data, one cannot separate data processing from analysis or from domain knowledge (e.g., Chen et al., 2012; Davenport & Patil, 2012; Loukides, 2012; Provost & Fawcett, 2013; Waller & Fawcett, 2013). Hence, BD specialists must have substantial industry knowledge in order to make sense of statistical analyses and communicate effectively with business colleagues.

#### 2.4 Organizational Setup of Business Intelligence and Big Data Teams

The differences between BI and BD also have consequences on how they are organized. Traditionally, BI teams are located in internal consulting organizations, centers of excellence, or IT departments, where they provide managers and executives with reports for their well-defined and stable information needs (Burton et al. 2006; Davenport et al. 2012; Varon 2012). However, since most BD initiatives lack predefined questions and are much more experimental in nature (Casey et al. 2013), BD specialists must be organized so they are close to products and processes in organizations, that is, co-located with business units (Davenport et al. 2012).

### 3 Methodology

While the literature provides first insights into the topic of BI and BD competencies, it is not grounded in empirical data. Therefore, we study the competencies required of BI and BD professionals by performing an automated content analysis of job ads using a text mining technique called latent semantic analysis (LSA), a quantitative method for analyzing qualitative data. LSA extracts word usage patterns and their meaning through statistical computations (Landauer et al. 1998) based on the idea that the contexts (e.g., documents, paragraphs, sentences) in which a word appears or does not appear largely determine the word’s meaning. LSA is based on the classical vector space model (Salton et al. 1975), in which documents are



represented as vectors of terms, and a collection of documents is represented as a term-document matrix that contains the number of times each term appears in each document (Manning et al. 2008). In a fashion similar to exploratory factor analysis, LSA performs a matrix operation called singular value decomposition (SVD) on the term-document matrix in order to reduce its dimensionality. The latent semantic factors that are extracted during this process can be interpreted as topics running through the collection of documents analyzed. LSA has received growing attention in the IS discipline for quantitative content analysis of academic papers (e.g., Larsen et al. 2008; Sidorova et al. 2008), social media posts (e.g., Evangelopoulos & Visinescu, 2012), sustainability reports (e.g., Reuter et al. 2014), vendor case studies (e.g., Herbst et al. 2014), and customer feedback (e.g., Coussement & Poel, 2008).

A typical LSA is comprised of three phases. (For a more detailed introduction and numerical examples, see Landauer et al. (1998) and Evangelopoulos, Zhang, & Prybutok (2012)). In the first phase, a collection of documents is transformed into a term-document matrix. This step typically requires pre-processing of documents (e.g., removing irrelevant or duplicate documents) and terms (e.g., uni- and bi-gram tokenization, filtering out uninformative terms, weighting terms according to their relative importance).

In the second phase, the term-document matrix undergoes SVD to reduce the dimensionality of the term-document matrix without losing essential information by identifying groups of highly correlated terms (i.e., terms that co-occur together in documents) and highly correlated documents (i.e., documents that contain similar terms). The result of the SVD is a set of factors (topics) with associated high-loading terms and documents. Together, they form patterns of word use that represent topics in the underlying collection of documents.

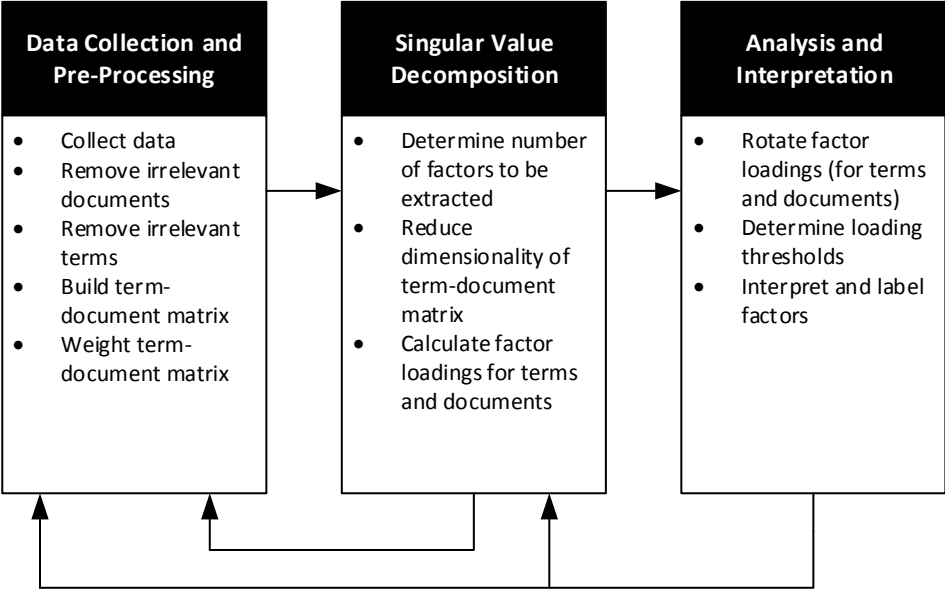
The extracted word-use patterns are interpreted in the third phase, which usually involves additional statistical analyses and, most importantly, expert judgment.

## **4 Data Collection and Analysis**

### **4.1 Overview**

The next sections illustrate how we applied LSA to analyze BI- and BD-related job advertisements. We followed the procedure described in Section 3 and depicted in Figure 1. As Figure 1 indicates, LSA often requires multiple iterations in which experts review statistical

results, and inputs (e.g., documents, terms) and parameters (e.g., term weights, number of factors to be extracted, loading thresholds) are fine-tuned in order to yield optimal results.



**Fig. 1.** Data Collection and Analysis Process

4.2 Data Collection and Pre-Processing

We performed multiple crawls of the global online recruitment website monster.com, downloading job advertisements from the U.S., Canada, Australia, and the U.K. that included either the term “business intelligence” or the term “big data.” We downloaded the data as two single-day snapshots in September 2013 and March 2014. After removing irrelevant hits (e.g., spam, non-English ads), we had an initial pool of 4,246 BI-related job ads and 1,411 BD-related job ads.

Following common text-mining procedures, we reduced the vocabulary in our document collection by removing stop words (e.g., “and,” “or,” “then”) and eliminating terms that occurred in less than 1 percent of the documents (Manning et al. 2008). The remaining vocabulary contained 6,813 terms, which we then manually reviewed to filter out other irrelevant terms while keeping only those terms that describe competencies. In particular, we removed standard human resources terms like “salary,” “bonus,” and “apply.” After this manual data clean-up, the final dictionary that we used as a go-list for the further analysis contained 1,570 terms.

Based on the controlled vocabulary and the two document sets, we built two term-document-matrices, one for BI jobs and one for BD jobs. These matrices contained the number of times a competency-related term appeared in a job ad. Then we weighted terms based on their occurrence in and across documents, applying the commonly used TF-IDF (Term Frequency-Inverse Document Frequency) weighting scheme, which promotes the occurrence of rare terms (e.g., “hadoop”) and discounts the occurrence of more common terms (e.g., “business,” “analysis”) (Manning et al. 2008). The two weighted term-document matrices built the foundation for the subsequent SVD.

### 4.3 Singular Value Decomposition (SVD)

We performed the SVD using the statistical computing software R. The first step of SVD is to define the number of factors (topics) to be extracted. Techniques from exploratory factor analysis, such as scree plots and the Kaiser-Harris criterion, would lead to a high number of factors, so these techniques are not recommended when the goal of LSA is to identify topics in a collection of documents. Since there is no standard procedure for determining an optimal number of topics, we manually explored alternative numbers of factors and qualitatively assessed the results (Evangelopoulos et al. 2012). We tested several dimensionalities, including 2, 5, 10, 15, 20, 30, and 50 factors. For each solution, we performed a SVD to compute term and document loadings for each factor.

### 4.4 Analysis and Interpretation

Following Sidorova et al. (2008), we performed a varimax rotation on the matrices with the term loadings to simplify interpretation of the factors. This procedure rotates the coordinates of the term loadings matrix in a way that maximizes the variance of a factor’s squared loadings on all terms in the matrix. As a result, each factor tends to load either high or low on a particular term; in other words, a term is either descriptive (high-loading) or not descriptive (low-loading) for a particular factor. To maintain the representation of the documents in the same factor space, we performed an identical rotation with the document loadings matrix.

Next, loading thresholds must be defined in order to determine whether a term or document is descriptive for a given factor. Again, no standard rules for setting this thresholds have emerged (Evangelopoulos et al. 2012), so we adopted a heuristic that Sidorova et al. (2008) and Evangelopoulos et al. (2012) apply in their LSA-based literature analyses and set the

threshold based on the probability distribution of term and document loadings. For a k-factors LSA, we retained the top-1/k high-loading documents, so each term and each document loads, on average, on one factor. However, terms and documents that load high on multiple factors or that load on no factor at all are to be expected.

The final step consisted of the manual sense-making and interpretation of the extracted factors and associated high-loading terms and documents. Two researchers independently interpreted and labeled each factor by examining the lists of extracted high-loading terms and documents. In almost all cases, factor interpretation was straightforward, and any minor disagreements in labeling factors were resolved during a final discussion.

**5 Results**

**5.1 Exploratory Data Analysis**

After downloading the job advertisements, we conducted an exploratory data analysis to get a first feeling for the data. We observed that there were about three times more BI-related job advertisements than BD-related job ads on monster.com. As a next step, we conducted a word frequency count, looking for overlaps between job ads (cf. Table 1). The results showed that about 15 percent of the BD jobs also include the term “business intelligence,” while only 5 percent of the BI job ads also included the term “big data,” perhaps an indicator that BD requires some basic BI-related skills, but BI does not necessarily require BD skills.

**Table 1.** Exploratory Data Analysis

<b>Keywords</b>	<b>Big Data (1,411 ads)</b>	<b>Business Intelligence (4,246 ads)</b>
<b>Big Data</b>	100.0% (1,411)	5.2% (221)
<b>Business Intelligence</b>	15.7% (221)	100.0% (4,246)

The word frequency count also showed that the frequency with which the terms “business intelligence” and “big data” appeared in the job ads was unbalanced, as many ads contained the search terms only once. A manual inspection of a sample of these ads revealed that the search terms often occurred only in the company descriptions (e.g., “our company specializes in big data solutions”) and that the companies were not looking for any BD- / BI-related employees but for, for example, a team assistant. Therefore, we filtered out job ads that included

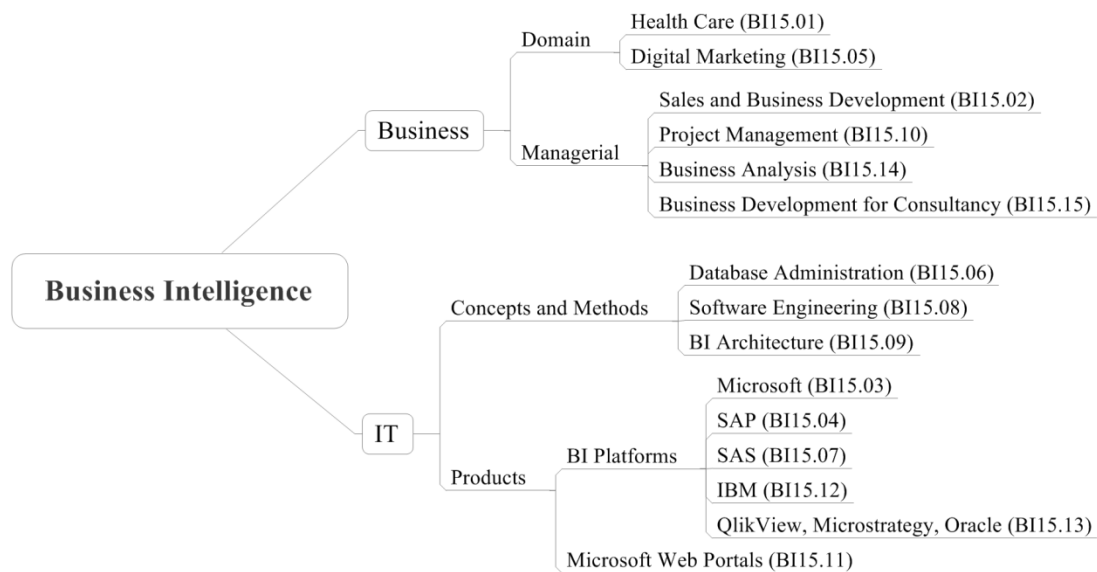
the keywords “big data” or “business intelligence” only once, which narrowed our data set down to 450 BD-related ads and 1,357 BI-related ads. (The ratios displayed in Table 1 were almost unchanged.)

## 5.2 Competency Requirements for Business Intelligence Professionals

On the most abstract level of the LSA, the two-factor solution, jobs were assigned to only two topics. The first factor was associated with high-loading descriptive terms like “developer,” “sql server,” “data warehouse,” “etl,” and “bi developer.” Associated titles of job ads included “BI Developer SQL Server,” “ETL Developer,” and “SQL Server DBA.” Terms like “sales,” “business development,” “marketing,” “account,” and “new business” described the second group of jobs, with such exemplary associated job titles as “Business Development Manager BI,” “Sales Executive BI,” and “New Business Sales Executive.” We had no difficulty or disagreement in making sense of and interpreting these results and labeled the two areas of competency “BI Architecture” and “Sales and Business Development.”

The fifteen-factor solution revealed clearly distinguishable BI-related topics that were neither too broad nor too specific. Table 2 provides an overview of the results and shows the high-loading terms and job ad titles as well as the manually assigned labels for each of the extracted factors. The terms and job ad titles are presented in order of descriptiveness, as expressed by the factor loadings calculated during SVD. (Uninformative terms and duplicate job titles were removed.) We will refer to these factors as competency requirements or competencies.

Table 2 makes clear that industry demands both business and IT competencies. The group of business-oriented competencies includes those related to specific domains (i.e., healthcare and digital marketing) and those related to managerial competencies (e.g., project management). The IT competencies can be divided into those related to vendor-specific products (e.g., Microsoft, SAP) and those related to general concepts and methods (e.g., database administration, BI architecture). Figure 2 aggregates the fifteen areas of competency in a taxonomy.



**Fig. 2:** Business Intelligence Competency Taxonomy

A more detailed examination of the descriptive terms and job ads associated with each factor gives insights into the corresponding competency requirements. Among the vendor-specific competencies are product and technology names of specific vendors. For example, BI professionals working with SAP technologies (Factor BI15.04) need competencies in SAP BusinessObjects (“business objects”), SAP Business Warehouse (“sap bw”), and/or the SAP High Performance Analytical Appliance (“hana”). Vendors focus on varying aspects of BI, as competencies related to the SAS BI Platform (Factor BI15.07) are described using terms like “statistical,” “analytics,” and “mining,” and important descriptors for IBM BI Platform competencies (Factor BI15.12) are “etl,” “report,” and “query.” The varying foci and strengths of each vendor explain these differences, as SAS is strong in data mining and IBM Cognos is a leader in data warehousing.

Our analysis also produced some generic IT competencies, such as database administration, software engineering, and BI architecture. Database administration requires SQL knowledge as well as knowledge in performance tuning of applications. Typical job ads that include these competencies are titled with “DBA” and its variants, depending on the operating platform (e.g., Oracle or MS SQL). Software engineering describes the competency of building custom pieces of software for data analysis. In particular, Java programming skills and web front-end-development knowledge are demanded. Last, the factor BI architecture describes a demand for expertise along the whole BI stack, from ETL to building reports.

**Table 2. Competency Requirements for Business Intelligence Professionals**

Factor ID (# of jobs)	Factor label	High-loading descriptive terms (excerpt)	Titles of high-loading job ads (excerpt)
BI15.01 (9%) <sup>1</sup>	Healthcare	care, health, systems, reporting, information, analysis	Business Analyst Regulatory Healthcare, Report Writer Business Analyst, Manager Clinical Decision Support
BI15.02 (8%)	Sales and Business Development	sales, business development, executive, legal, sales team	Legal Sales Executive, Business Development Manager, Sales Executive Business Intelligence, Sales Manager Business Intelligence
BI15.03 (16%)	BI Platforms (Microsoft)	sql server, ssis, ssrs, ssas, microsoft, microsoft bi, reporting services, etl	BI Developer SSIS SSAS SSRS SQL, BI Data Warehouse Developer SQL Server, SQL Server Developer, ETL Developer Business Intelligence SSIS SQL SSRS
BI15.04 (5%)	BI Platforms (SAP)	sap, sap bi, hana, business objects, sap bw, erp, consultant, business analyst, crystal	SAP BI Principal Consultant, SAP BI Senior Technical Consultant, SAP BI Report Analyst Developer, Senior Business Objects Consultant
BI15.05 (6%)	Digital Marketing	marketing, digital, campaigns, product, analytics, segmentation, customer	Senior Marketing Executive Online Data Solutions Job, Marketing Database Analyst, Email Marketing Manager, Digital Relationship Marketing Manager
BI15.06 (2%)	Database Administration	dba, database administrator, sql server, oracle, sql, production, developer, tuning	Oracle DBA SQL Server Database Administrator, Senior DBA SQL Server Database Administrator, SQL DBA with BI Business Intelligence, MS-SQL Server DBA
BI15.07 (3%)	BI Platforms (SAS)	sas, studio, analytics, statistical, mining, olap, data mining, data analytics	SAS BI Analyst, Data Analytics Business Intelligence Consultant, Senior SAS Developer, SAS Consultant
BI15.08 (4%)	Software Engineering	java, eclipse, apache, web, linux, engineer, software, javascript, developer, big data	Senior Java Consultant, Senior Java Technical Consultant, Mobile Developer Java jQuery HTML5, Front End Engineer, Senior Backend Engineer

<sup>1</sup> Even though we retained the top-1/k term and document loadings and set the computed threshold value accordingly, we followed Evangelopoulos et al. (2012) in double-checking and manually selecting a threshold for each factor separately based on domain knowledge. As a result, we had to reduce the number of jobs that loaded on the first factor (BI15.01).

Factor ID (# of jobs)	Factor label	High-loading descriptive terms (excerpt)	Titles of high-loading job ads (excerpt)
BI15.09 (5%)	BI Architecture	bi developer, etl, developer, bi stack, organization, report	Business Intelligence Developer, ETL Business Intelligence Developer, BI Developer Excel Microsoft BI SQL Server, Senior BI Developer Architect
BI15.10 (5%)	Project Management	project manager, project, management, head, client, change, agile, planning	Senior Project Manager, Technical Project Manager, BI Project Manager Data Warehouse Implementations, Sr Project Manager Business Intelligence
BI15.11 (3%)	Web Portals (Microsoft)	sharepoint, .net, server, microsoft, administrator, software, web, application	SharePoint Developer, SharePoint 2007-2010 Developer, SharePoint Administrator SharePoint 2010 Server, SharePoint Consultant, SharePoint Architect
BI15.12 (5%)	BI Platforms (IBM)	cognos, studio, manager, report, framework, developer, ibm, query, analyst, etl	Cognos BI Developer, Cognos BI Manager, Cognos Designer, MIS Manager with Cognos BI experience, Cognos 10 Consultant Developer
BI15.13 (15%)	BI Platforms (QlikView, Microstrategy, OBIEE)	qlikview, microstrategy, oracle, obiee, warehouse, etl, architect, consultant	MicroStrategy Business Intelligence Analyst, MicroStrategy Developer, Senior QlikView Developer, BI Visualization Consultant, ETL Specialist
BI15.14 (6%)	Business Analysis	business analyst, data analyst, reporting, excel, organization, specialist, pivot	Business Analyst, Business Analyst SAP APO Excel Expert, Data Analyst, Reporting Data Analyst, BI Report Analyst, Technical Business Analyst
BI15.15 (7%)	Business Development (Consultancy)	consultancy, business development, sales, development manager, account, market	Business Development Manager & Market Intelligence Consultancy, Sales Business Development Manager, Sales Account Manager Research Consultancy

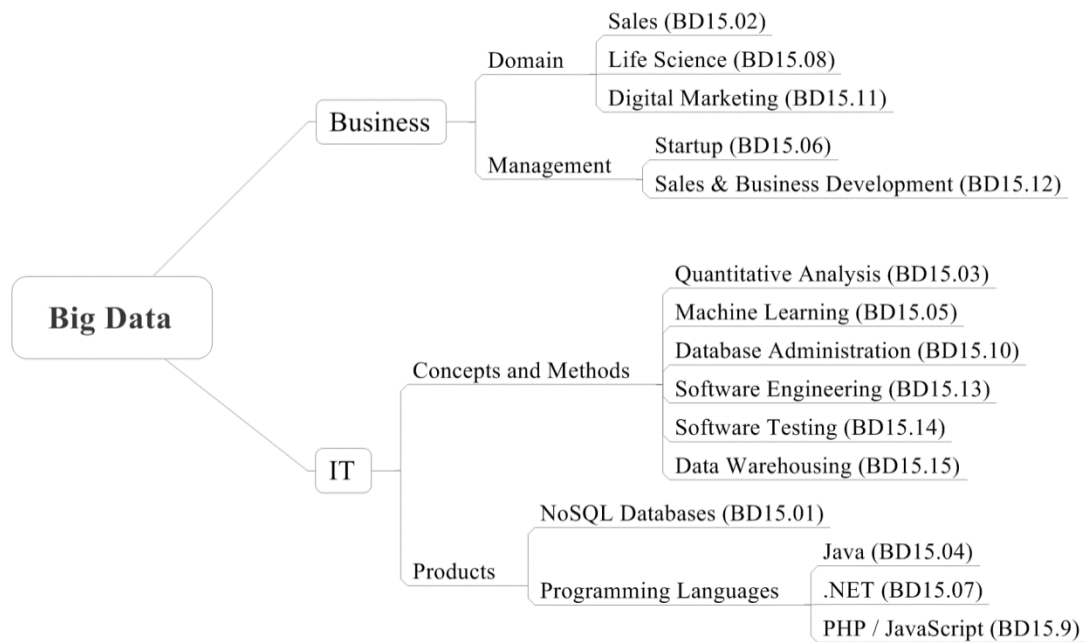
In addition to analyzing single areas of competency, we determined the current demand for each competency by calculating how many job ads loaded high on a factor. The relative numbers of jobs assigned to a factor, displayed in Table 2, indicate that competencies in BI platforms, healthcare, and sales and business development are among the competencies with the highest demand on the BI job market. Table 2 also shows that the demand for business-related jobs and IT-related jobs is almost evenly distributed.



### 5.3 Competency Requirements for Big Data Professionals

To report on the results for BD-related jobs, we conducted the LSA on several levels of abstraction. On the most abstract level, the two-factor solution, we assigned jobs to two topics. The five highest-loading terms for the first topic were “java,” “developer,” “hadoop,” “web,” and “sql,” and exemplary titles of high-loading job ads were “Experienced Java Developer,” “Java Hadoop Developer,” and “Data Scientist Java Hadoop NoSQL.” In contrast, the top five descriptive terms for the second topic were “digital,” “sales,” “manager,” “advertising,” and “marketing,” and frequent job titles included “Digital Sales Executive,” “Sales Manager Big Data,” and “Digital Relationship Marketing Manager.” The examination of the highest-loading terms and job titles for both factors suggests that the first factor describes jobs related to the development of BD solutions (big data developers), while the second factor refers to the use of BD in marketing and sales (big data users).

Table 3 provides an overview of the results of the fifteen-factor solution and shows exemplary high-loading terms and job titles, as well as the manually assigned labels for each of the extracted factors. The inspection of the identified areas of competency shows that, just as for BI jobs, competencies can be clustered into business competencies and IT competencies. The IT competency area can be further broken down into generic concepts and methods like quantitative analysis, machine learning, and database administration, and products for developing big data solutions (i.e., a variety of programming languages and NoSQL databases). The group of business-oriented competencies is made up of domain competencies in the areas of life sciences and digital marketing, as well as managerial competencies in sales and business development and working in start-up companies. Figure 3 summarizes these findings in a big data competency taxonomy.



**Fig. 3:** Big Data Competency Taxonomy

In contrast to the BI competencies, we find no factors related to the technologies of commercial vendors, yet many conceptual and methodological competencies, as well as programming skills in various languages are required. In the factor representing competency in NoSQL (BD15.01), not a single product or technology name of one of the big commercial database vendors appears. Instead, terms referring to open-source technologies from the Apache Foundation are dominating the descriptions (e.g., “hadoop,” “hive,” “pig,” “cassandra”). Furthermore, conceptual and methodological IT skills like quantitative analysis (BD15.03), machine learning (BD15.05), database administration (BD15.10), and software engineering and testing (BD15.13, BD15.14) are in high demand. These findings suggest that the field of BD is not (yet) dominated by big vendors’ standard software but (still) relies largely on open-source technologies and custom-made software solutions.

**Table 3. Competency Requirements for Big Data Professionals**

Factor ID (# of jobs)	Factor label	High-loading descriptive terms (excerpt)	Titles of high-loading job ads (excerpt)
BD15.01 (17%)	NoSQL Data-bases	hadoop, nosql, java, hive, scripting, distributed, database, apache, mapreduce, hbase, pig, cassandra	Java Hadoop Developer, Big Data Solutions Architect, Big Data Consultant, Database Architect, Big Data Scientist, Chief Architect Big Data Guru

Factor ID (# of jobs)	Factor label	High-loading descriptive terms (excerpt)	Titles of high-loading job ads (excerpt)
BD15.02 (6%)	Sales	digital, sales, advertising, manager, media, forecasting, presentation, platforms	Junior Digital Sales Manager, Digital Agency Sales Manager, New Business Sales Manager
BD15.03 (2%)	Quantitative Analysis	quantitative, risk, analyst, models, modeling, matlab, java, algorithms, physics, phd, financial, mathematics, data analyst	Senior Quantitative Analyst, Quantitative Analyst Financial Risk Management, Big Data Business Systems Analyst, Sr Data Analyst
BD15.04 (8%)	Programming (Java)	java developer, junit, tdd, hadoop, maven, git, nosql, hibernate, eclipse, agile, hive, mongodb, apache, pig	Experienced Java Developer Java Multi Thread JUnit TDD, Java Developer Big Data, Java Hadoop Developer, Senior Java Architects Developers Core Java Programming, Senior Java Consultant Java Spring Hibernate Maven
BD15.05 (7%)	Machine Learning	data scientist, machine learning, visualization, statistical, algorithms, mining, predictive, analysis, science, mathematics	Data Scientist Machine Learning C Java Python, Software Engineer Data Scientist Machine Learning, Security Cleared Big Data Scientist, Big Data Architect Hadoop R Machine Learning
BD15.06 (2%)	Startup	startup, sales, analytics platforms, market, data analytics, applications, solutions, information, enterprise, consultant	Sales Big Data Software, Front End Developer for Big Data Startup, Big Data Analytics Sales Consultant, Junior Python Developer Big Data Tech Startup, Java Software Engineers High Profit Big Data Startup
BD15.07 (2%)	Programming (.NET)	net, sql server, microsoft, visual, developer, warehouse, api, front end, scrum, mvc, agile, project, online	High Paid Junior C# ASP.NET Developer, Developer .NET MVC API, C# .NET Developer SQL Server Senior Software Engineer Big Data, Junior C# ASP.NET Developer
BD15.08 (2%)	Life Science	sciences, life, medical, visualization, revenue, health, care, project management, industries, consulting	Strategic Account Manager Big Data Life Sciences, Big Data Engineer Exciting Start Up, Revenue Analyst, Project Manager Big Data, Solutions Consultant Big Data Analytics Visualisation
BD15.09 (8%)	Programming (PHP / JavaScript)	developer, php, web, javascript, front end, user, css, agile, html, jquery, web services, api, mysql, open source, mongodb	Lead PHP Ninja, Front End Developer Start Up, Senior PHP Developer, UI Developer Big Data JavaScript, Front End Developer HTML5 CSS3 JavaScript, PHP Web Developer OOP LAMP

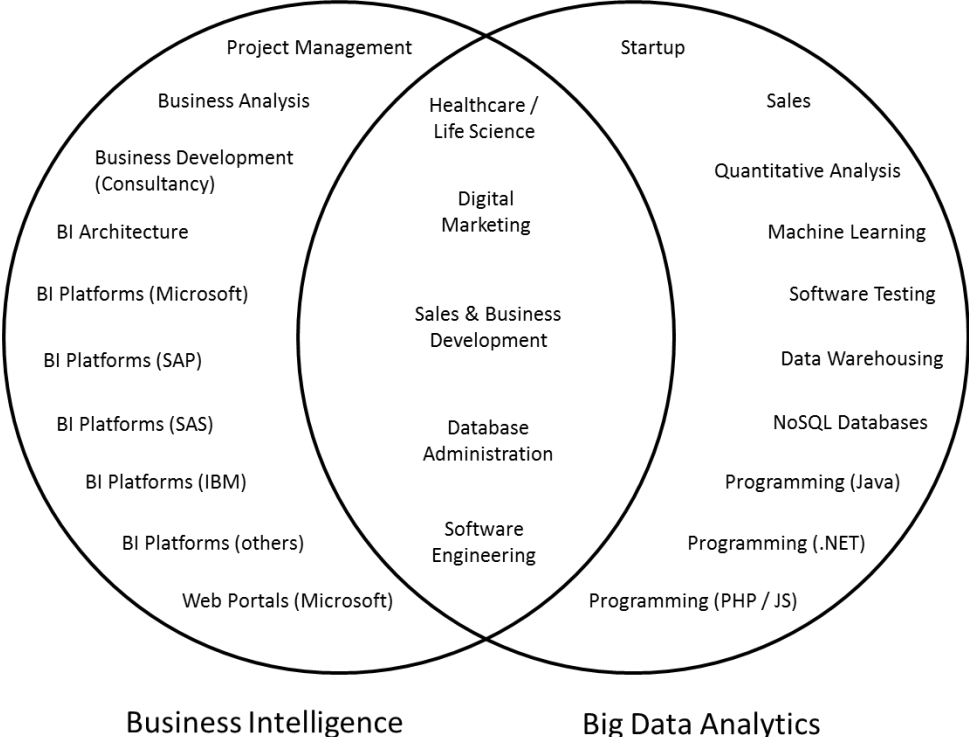
Factor ID (# of jobs)	Factor label	High-loading descriptive terms (excerpt)	Titles of high-loading job ads (excerpt)
BD15.10 (4%)	Database Administration	dba, mysql, oracle, high availability, sql server, linux, database, senior, consultancy	MySQL DBA Big Data High Availability Replication, DBA Data Modeler, Junior Database Administrator, DBA Systems Engineer MySQL NoSQL Big Data Unix Linux
BD15.11 (10%)	Digital Marketing	marketing, digital, analytics, media, insights, information, social, research, strategy	Associate Director Digital Media Analytics, Marketing Director, Senior Analyst Big Data Digital Media, Digital Relationship Marketing Manager
BD15.12 (11%)	Business Development	sales, customer, revenue, account, management, executive, business development, marketing, relationships	Business Development Manager Big Data Technology, Business Development Manager Cloud Computing
BD15.13 (9%)	Software Engineering	software engineer, linux, data engineer, online, professional, open, product, natural language, systems, agile, distributed	Senior Engineer Big Data, Principle Software Engineer Head of Software Development, Senior Big Data Engineer, Senior Software Engineer
BD15.14 (9%)	Software Testing	testing, software, engineer, product, machine learning, automated, development, open source, agile, building	Python Test Engineer, Software Test Engineer, Java Tester, Software Design Engineer in Test, Test Lead
BD15.15 (3%)	Data Warehousing	etl, data warehouse, business intelligence, manager, project, technical, modeling, agile	Data Warehouse Product Owner, Director of Data Engineering, Snr Business Intelligence Developer, ETL Engineer, DWH Delivery Manager

Comparing the relative demand between business and IT competencies reveals that almost 70 percent of the posted BD-related job ads seek technical skills. Knowledge in NoSQL databases and software engineering and programming are the most highly demanded areas of technical competency. Digital marketing, business development, and sales constitute highly demanded business competencies.

#### 5.4 Comparison

We identified a number of similarities between the fields of BI and BD. Especially when it comes to generic IT concepts and methods and business skills, we observed a considerable overlap between BI and BDA (cf. Figure 4). For example, working in either field requires a certain amount of software engineering and database competency. Sales and business devel-

opment skills for managing BI and BD solutions also overlap. Finally, domain knowledge overlaps in healthcare / life sciences and digital marketing, domains known to be especially data-driven. The absence of other domain skills is a result of the level of analysis we chose; a more granular LSA on BI and BD job ads (e.g., 50 instead of 15 factors) would reveal the additional domains of banking, finance, insurance, and supply chain management.



**Fig. 4.** Similarities and differences in BI and BD areas of competency

The major differences between BI and BD competencies are discussed in the next section.

**6 Discussion**

Our research revealed highly demanded BI and BD skills in at least two areas, business and IT. This first finding empirically grounds the ongoing discussion about business knowledge’s being as important as technical skills for working successfully on BI and BD initiatives (e.g., Chen et al., 2012; de Lange, 2013; Waller & Fawcett, 2013; Wixom et al., 2011). For example, De Lange (2013) sees programming and statistical expertise as the foundation for data scientists but also states that “a strong background in business and strategy can help jettison a younger scientist’s career to the next level.” Chen et al. (2012) argue that BI and analytics professionals “must be capable of understanding the business issues” and at the same time capable of “framing the appropriate analytical solutions” to provide useful decision-making

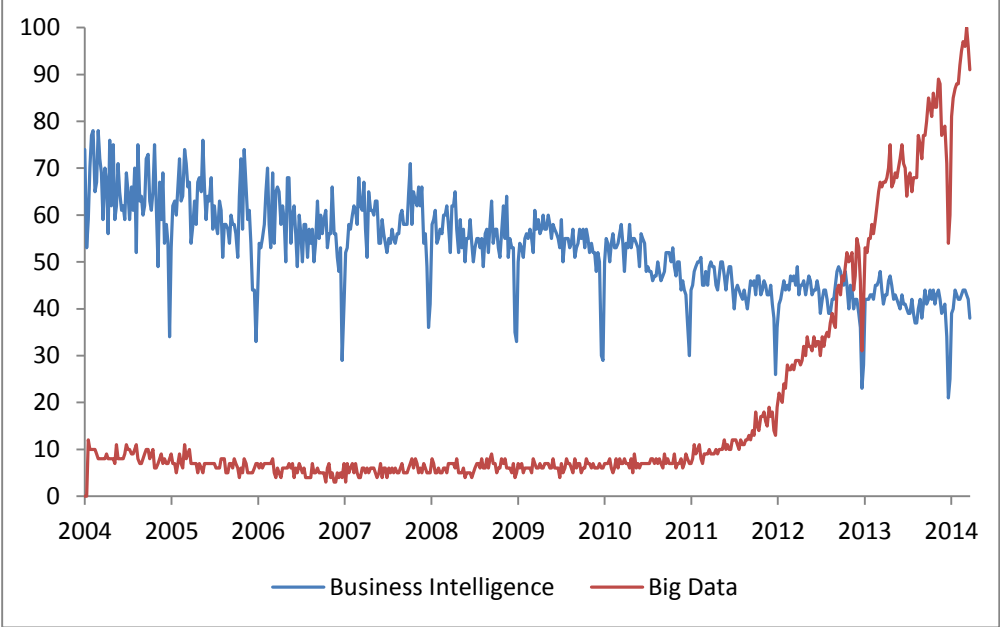
support. Wixom et al. (2011) analyze existing BI-related university programs and courses and conclude that the BI program of the future should include both business and technical courses, including at least an understanding of data management, functional business knowledge, statistics and quantitative analysis, and communication and visualization skills, in order to address the widest scale of industry needs. The empirical evidence we provide with this study underscores these arguments and should encourage IS scholars to develop inter-disciplinary programs and courses to prepare “the next generation of analytical thinkers” (Chen et al. 2012).

We also showed that there are considerable differences between BI skills and BD skills. The extracted BI competency requirements feature skills related to commercial products of large software vendors, whereas no BD skills descriptors refer to one of the large BI vendors. In addition, almost 70 percent of the BD jobs we analyzed asked for strong software development skills and statistical knowledge, whereas BI jobs required much less “programming” and statistical knowledge. While BD jobs demand quantitative analysis and machine learning skills, there is no explicit mentioning of such terms in BI jobs.

Why did we find such differences, although both BI and BD focus on supporting decision-making through quantitative analysis of data? There are two possible explanations for this finding: First, the emerging literature on big data consistently emphasizes its variety, suggesting that big data does not refer to relational data managed in enterprise systems or data warehouses but to streams of data in various formats and from various sources (Davenport et al. 2012), mostly the Internet. Because of this variety of data, big data analytics solutions rely less on standard software products than they do on custom-made solutions. Second, current big data projects seek answers to highly specialized questions and are often more comparable to research projects than to traditional IT projects (Marchand and Peppard 2013). Because of this variety of questions, big data analytics solutions require more tailored software tools and better methodological skills than traditional BI does. Whatever the explanation, our observation is in line with Golden’s (2013) argument that big data investments will be open-source. Even though large vendors like SAP are developing analytical solutions like SAP HANA (vom Brocke et al. 2014) that will offer “predictive analytics, text and big data in a single package” (SAP 2014), our analysis shows that the BD job market does not yet ask for experts in the use of these tools.

We also found that the demand for BI competencies is still far bigger than that for BD competencies, as we found three times more job ads containing the term “business intelligence” than

we did job ads containing the term “big data.” This finding might be surprising given the current media excitement around big data (cf. Figure 5), but our empirical results suggest that most companies are still working on advancing the maturity of their internal BI and are not yet seeking to exploit big data.



**Fig. 5.** Google search volume for the search queries “business intelligence” and “big data” (Source: Google Trends)

Highlighting our results against the background of the resource-based view of the firm (i.e., the TIR and HIR mentioned in Section 2), we argue that BI implementations rely heavily on well-established TIR, as BI platform vendors already provide them. Significant amounts of knowledge have already been built into the technology itself, and it is at a mature state and easily deployed in a company. HIR are only required for customizing and adapting the technology to the organizational context. However, BD still relies on basic TIR, such as programming languages and plain database technologies, which require extensive HIR in order to build the sophisticated, company-specific big data solutions that may lead to temporary competitive advantage. Therefore, we can conclude that BD initiatives are currently much more human-capital-intensive than BI projects are, so we call for further action in educating current and future employees.

Contrasting our findings against the three evolutionary BI&A waves Chen et al. (2012) identify, we observe that the skills that are related to the first wave (structured content residing in databases) are still the most frequently demanded. Examples include conceptual knowledge

about data warehousing and practical skills concerning major BI platforms. Supporting decision-making by extracting knowledge from web-based and unstructured content (i.e., second-wave BI&A) still seems to be in its infancy, as we found no factor related to text mining, web mining, or social network analytics, although some of these terms were scattered among the high-loading terms that described BD factors. These results are surprising, as many experts point out that these techniques are at the core of big data analytics. Finally, our analysis did not produce evidence that industry currently demands third-wave BI&A competencies (i.e., mobile and sensor data), a finding that disagree with Chen et al. (2012).

## **7 Conclusion**

This paper set out to shed light on the topic of individual-level BI and BD competencies. Given the lack of empirical research in this area, we conducted an LSA of 1,357 BI-related and 450 BD-related job ads harvested from the online employment platform monster.com. By analyzing and interpreting the statistical results of the LSA, we developed BI and BD competency taxonomies. Our major findings are that (1) business knowledge is as important as technical skills for working successfully on BI and BD initiatives; (2) BI competency is characterized by skills related to commercial products of large software vendors, whereas BD jobs ask for strong software development and statistical skills; (3) the demand for BI competencies is still far bigger than the demand for BD competencies; and (4) BD initiatives are currently much more human-capital-intensive than BI projects are.

Our research contributes to the scientific body of knowledge on BI and BD and has several implications for practitioners. By uncovering highly demanded skill sets for BI and BD experts, we complement existing scientific work on BI / BD maturity models. The empirically grounded taxonomies we developed can be used as a foundation for future empirical studies on BI and BD, such as efforts to develop measurement instruments for studying BI and BD professionals or teams. Our findings also inform the assessment and development of BI and BD curricula. As numerous practitioners and researchers have pointed out, undergraduate and graduate programs should be created or modified in order to satisfy industry's high demand for analytical skills, especially in the areas of software engineering, statistics, and business skills. Practice may benefit from this study in two ways. At an individual level, our results provide guidance for individuals' professional development by, for example, outlining pathways for career choices and decisions about continuing education. At an organizational level,



the identified competencies can be used to inform strategic HR management (e.g., establishment of a BI/BD Center of Excellence) and staffing decisions (e.g., for BI/BD projects). In particular, we advise organizations that want to engage in big data analytics either to invest in building in-house software engineering and statistical skills or to collaborate with third parties (e.g., universities) in order to obtain the required competencies.

As in all research, this study is not without several limitations. First, our findings are based on snapshots of the BI and BD job market taken in September 2013 and March 2014. To gain a more reliable picture of knowledge and skill requirements and track their development over time we plan to repeat the study presented here regularly in the future. Second, our data analysis used job advertisements to elaborate on the differences between BI and BD competencies, as it is reasonable to assume that job advertisements act as proxies for a demand for human capital in industry and that they can provide insights into competency requirements. However, one must be aware that job ads do not always reflect an employer's true requirements, as the employer may ask for more competencies than can be reasonably expected from an applicant, or they may use a specific vocabulary to polish job ads so they are appealing to a certain group of candidates. Such may be the case especially in the area of BI and BD, which lacks clear-cut definitions and is full of industry jargon. While we acknowledge that such biases may exist in our data, we believe that the number of job ads that we examined should be sufficient to minimize the effect of biases in a few ads. The processing of such a broad data source as that used in this research gives a particular advantage to the approach we used over other research methods, such as interviews, because it diminishes the risk of biases caused by specific contextual backgrounds. Third, our findings are limited to job markets in English-speaking countries because of the nature of the text mining technique we applied, which cannot process multilingual texts. Future studies may look at job markets in other major language regions (e.g., Spanish, French, Portuguese, German, Russian, Hindustani, Mandarin Chinese). Finally, our study is inductive and exploratory in nature, so future confirmatory research (e.g., surveys) is needed in order to test and refine our results.

## References

- Barney J (1991) Firm Resources and Sustained Competitive Advantage. *J Manage* 17:99–120.
- Vom Brocke J, Debortoli S, Müller O, Reuter N (2014) How In-Memory Technology Can Create Business Value: Insights from the Hilti Case. *Commun Assoc Inf Syst* 34:151–168.
- Buhl HU (2013) Interview with Martin Petry on “Big Data.” *Bus Inf Syst Eng* 5:101–102.
- Buhl HU, Röglinger M, Moser F, Heidemann J (2013) Big Data. *Bus Inf Syst Eng* 5:65–69.
- Burton B, Geishecker L, Hostmann B, et al. (2006) Organizational Structure: Business Intelligence and Information Management. 1–11.
- Casey T, Krishnamurthy K, Abezgauz B (2013) Who Should Own Big Data? strategy+business
- Cates J, Gill S, Zeituny N (2005) The Ladder of Business Intelligence (LOBI): a framework for enterprise IT planning and architecture. *Int J Bus Inf Syst* 1:220–238.
- Chen H, Chiang R, Storey V (2012) Business intelligence and analytics: from big data to big impact. *MIS Q* 36:1165–1188.
- Coussement K, Poel D Van den (2008) Improving customer complaint management by automatic email classification using linguistic style features as predictors. *Decis Support Syst* 44:870–882.
- Cox M, Ellsworth D (1997) Application-controlled demand paging for out-of-core visualization. *Proc 8th Conf Vis* 235–244.
- Davenport TH (2006) Competing on analytics. *Harv Bus Rev* 84:98–107.
- Davenport TH, Barth P, Bean R (2012) How “big data” is different. *MIT Sloan Manag Rev* 54:22–24.
- Davenport TH, Patil D (2012) Data Scientist. *Harv Bus Rev* 90:70–76.
- Dhar V (2013) Data Science and Prediction. *Commun ACM* 56:64–73.
- Dinter B (2012) The Maturing of a Business Intelligence Maturity Model. *Proc. Am. Conf. Inf. Syst.* Seattle, pp 1–10
- Eckerson W (2004) Gauge Your Data Warehouse Maturity. <http://www.information-management.com/issues/20041101/1012391-1.html>. Accessed 29 Apr 2013
- Evangelopoulos N, Visinescu L (2012) Text-mining the voice of the people. *Commun ACM* 55:62–69.

- Evangelopoulos N, Zhang X, Prybutok VR (2012) Latent Semantic Analysis: five methodological recommendations. *Eur J Inf Syst* 21:70–86.
- Forbes (2013) A Very Short History Of Big Data. <http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/>. Accessed 23 Mar 2014
- Gallivan MJ, Truex DP, Kvasny L (2004) Changing Patterns in IT Skill Sets. A Content Analysis of Classified Advertising. *DATA BASE Adv Inf Syst* 35:64–87.
- Golden B (2013) Does Big Data Spell the End of Business Intelligence As We Know It? In: CIO. <http://www.cio.com/article/print/730774>. Accessed 4 Oct 2013
- Herbst A, Simons A, vom Brocke J, et al. (2014) Identifying and Characterizing Topics in Enterprise Content Management: A Latent Semantic Analysis of Vendor Case Studies. *Proc. 22nd Eur. Conf. Inf. Syst.*
- Hostmann B, Hagerty J (2010) ITScore Overview for Business Intelligence and Performance Management. <http://www.gartner.com/id=1433813>. Accessed 29 Apr 2013
- IBM (2012) Analytics: The real-world use of big data. <http://www-935.ibm.com/services/us/gbs/thoughtleadership/ibv-big-data-at-work.html>. Accessed 4 Oct 2013
- Lahrman G, Marx F, Winter R, Wortmann F (2011) Business intelligence maturity: Development and evaluation of a theoretical model. *Proc. Hawaii Int. Conf. Syst. Sci. Koloa*, pp 1–10
- Landauer TK, Foltz PW, Laham D (1998) An introduction to latent semantic analysis. *Discourse Process* 25:259–284.
- Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety. In: META Gr. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Accessed 4 Oct 2013
- De Lange C (2013) So you want to be a data scientist? In: *Nat. Jobs Blog*. <http://blogs.nature.com/naturejobs/2013/03/18/so-you-want-to-be-a-data-scientist>. Accessed 3 May 2013
- Larsen KR, Monarchi DE, Hovorka DS, Bailey CN (2008) Analyzing unstructured text data: Using latent categorization to identify intellectual communities in information systems. *Decis Support Syst* 45:884–896.
- LaValle S, Lesser E, Shockley R, et al. (2011) Big Data , Analytics and the Path From Insights to Value Big Data , Analytics and the Path From Insights to Value. *MIT Sloan Manag. Rev.* 52:
- Litecky C, Aken A (2010) Mining for Computing Jobs. *IEEE Softw* 27:78–85.
- Loukides M (2012) What Is Data Science? O'Reilly Media

- Manning CD, Raghavan P, Schütze H (2008) Introduction to information retrieval. Cambridge University Press, New York
- Manyika J, Chui M, Brown B, et al. (2011) Big data: The next frontier for innovation, competition, and productivity. 1–156.
- Marchand D, Peppard J (2013) Why IT Fumbles Analytics. *Harv Bus Rev* 91:104–112.
- Melville N, Kraemer KL, Gurbaxani V (2004) Review: Information technology and organizational performance: An integrative model of IT business value. *MIS Q* 28:283–322.
- Pollock R (2013) Forget Big Data, Small Data is the Real Revolution. In: Open Knowl. Found. Blog. <http://blog.okfn.org/2013/04/22/forget-big-data-small-data-is-the-real-revolution/>. Accessed 4 Oct 2013
- Power D (2007) A brief history of decision support systems.
- Provost F, Fawcett T (2013) Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data* 1:51–59.
- Reuter N, Vakulenko S, vom Brocke J, et al. (2014) Identifying the Role of Information Systems in Achieving Energy-Related Environmental Sustainability Using Text Mining. *Proc. 22nd Eur. Conf. Inf. Syst.*
- Van Roekel H, Linders J, Raja K, et al. (2009) The BI Framework: How to Turn Information into a Competitive Asset. [http://www.tdwi.eu/wissen/whitepaper/?no\\_cache=1&tx\\_mwknowledgebase\\_pi1\[showU id\]=45](http://www.tdwi.eu/wissen/whitepaper/?no_cache=1&tx_mwknowledgebase_pi1[showU id]=45). Accessed 29 Apr 2013
- Russell S, Haddad M, Bruni M, Granger M (2010) Organic Evolution and the Capability Maturity of Business Intelligence. *Proc. Am. Conf. Inf. Syst.*
- Russom P (2011) Big Data Analytics. *TDWI Res.*
- Salton G, Wong A, Yang CS (1975) A vector space model for automatic indexing. *Commun ACM* 18:613–620.
- SAP (2014) SAP HANA integrates predictive analytics, text and big data in a single package. <http://www.sap.com/pc/tech/in-memory-computing-hana/software/analytics/big-data.html>. Accessed 23 Mar 2014
- Schlegel K, Sallam RS, Yuen D, Tapadinhas Y (2013) Magic Quadrant for Business Intelligence Platforms. <http://www.gartner.com/technology/reprints.do?id=1-1DZLPEP&ct=130207&st=sb>. Accessed 29 Apr 2013
- Sidorova A, Evangelopoulos N, Valacich JS, Ramakrishnan T (2008) Uncovering the intellectual core of the information systems discipline. *MIS Q* 32:467–482.

- Todd PA, McKeen JD, Gallupe RB (1995) The evolution of IS job skills: a content analysis of IS job advertisements from 1970 to 1990. *MIS Q* 19:1–27.
- Varon E (2012) Rethink Your Org Chart for Big Data Analytics Teams. <http://data-informed.com/rethink-your-org-chart-for-big-data-analytics-teams/>. Accessed 13 Mar 2014
- Waller M a., Fawcett SE (2013) Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *J Bus Logist* 34:77–84.
- Wixom B, Ariyachandra T, Goul M, et al. (2011) The current state of business intelligence in academia. *Commun Assoc Inf Syst* 29:299–312.