Data Warehousing and Business Intelligence Skills for Information Systems Graduates: Analysis Based on Marketplace Demand

A. Shirani
San Jose State University

Maria Malu H. Roldan
San Jose State University, malu.roldan@sjsu.edu

Follow this and additional works at: https://scholarworks.sjsu.edu/mis_pub

Part of the Management Information Systems Commons

Recommended Citation

This Article is brought to you for free and open access by SJSU ScholarWorks. It has been accepted for inclusion in Faculty Publications, Information Systems & Technology by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.
DATA WAREHOUSING AND BUSINESS INTELLIGENCE SKILLS FOR INFORMATION SYSTEMS GRADUATES: ANALYSIS BASED ON MARKETPLACE DEMAND

Ashraf Shirani, San Jose State University, shirani_a@cob.sjsu.edu
Malu Roldan, San Jose State University, roldan_m@cob.sjsu.edu

ABSTRACT

In an effort to keep information systems (IS) curriculum current with emerging technologies, this research reports the result of a survey of marketplace demand for database, data warehousing and business intelligence skills. Sample job and skills data were obtained from Dice.com’s online postings for jobs that required a bachelor’s degree in IS or equivalent qualifications. Cluster analysis revealed some skill overlap among the three job categories. Significant skills for each category of jobs, along with skill and vendor rankings, are reported.

Keywords: Information Systems (IS) Skills, IS Curriculum, IT Workforce Trends, Data Warehousing and Business Intelligence Skills

INTRODUCTION

A recent report by a panel on workforce trends in the information technology industry, held at the AMCIS meeting in 2008 concluded that there was an ominous crisis in the making in several areas of demand and supply of information systems (IS) skills, among them: “… graduates who are not trained in areas that the marketplace is seeking; thin pipeline for specific technical skills; and lag in university responsiveness to the needs of the marketplace.” [2]. Separately, the joint committee of the Association for Information Systems (AIS) and the Association for Computing Machinery (ACM) [5], in its recommendations for an IS model curriculum and guidelines noted that today’s global firms were rapidly introducing emerging technologies such as ERP and data warehousing in their operations. Thus one of the objectives of the committee’s latest IS curriculum guidelines was to offer education in topics such as data warehousing, data mining and other emerging technologies suitable for the current millennium. The committee identified eight possible career paths for IS graduates, five with managerial and three with systems and technology emphasis. The technical paths included (1) data administration; (2) networking, telecom, and infrastructure; and (3) software development. As a step towards addressing some of the concerns expressed by the two professional groups noted above, this research attempted to identify market-relevant skills in a set of popular emerging technologies – data warehousing (DW) and business intelligence (BI). Focus of this research is on the first of the three technical paths emphasized by the AIS-ACM curriculum committee – data administration. We report the results of analysis conducted on data from a recent survey of online job market in these technologies.

PRIOR RESEARCH

The importance of identifying market-relevant skills and any gaps that may exist in demand and supply of these skills is not lost on educators and curriculum committees in many disciplines. The issue is much more compelling in the information systems field, which continues to experience unparalleled advances in technology and shifts in business practices. Data warehousing and business intelligence are among the technologies that represent such advances. In a recent survey of 1500 organizations conducted by Gartner, a leading research firm, the CIOs reported (as they consistently did for the past four years) that BI was their top investment priority [6]. With organizations deploying data warehousing and business intelligence solutions in increasing numbers, they require people who understand how relational and dimensional data are organized, cleansed, transformed, stored, and accessed with a variety of tools including dashboards and scorecards. Some of the leading educators and researchers in the field (e.g., [16]) are concerned that many business schools are not adequately preparing their graduates with DW and BI skills that the industry needs.

Reflecting the concerns of stakeholders including educators, students, and employers, a regular stream of skill related research has appeared in recent years. Two broad categories of such research have portrayed the findings from the perspectives of employers and job seekers. As noted by some researchers, [10, 13, 17], surveys of managers tend to emphasize managerial and communication knowledge more than technical skills. In surveys of software and systems client and provider companies [1, 2, 4], for example, executives felt that more
project management and business functional area skills were needed.

From the U.S. job-seekers’ perspective, however, a somewhat different picture emerges— a balance of technical and so-called ‘soft’ skills are sought by employers. Hopek et al. [7] identify five key factors in implementing BI solutions that heavily rely on soft skills: strategic alignment of BI projects to overall enterprise strategy, obtaining upper management support, communicating change management, user buy-in, and user training. Although soft skills such as good communication and teamwork, along with business domain knowledge appear to be necessary conditions, they are by no means sufficient for securing or holding an IS job. Proficiency in job-related technical skills is an essential requirement for most IS jobs. Recent surveys and analyses of online job market for IS graduates [8, 13, 17], for example, have reported strong demand for the traditional IS skill sets (database, systems analysis, networking, and programming), along with written and verbal communication. A study by Simon [15] segregates skill demand by entry- vs. mid-level positions. Their analysis revealed that though as a whole, a balance of domain knowledge, and technical and communication skills was desirable, technical skills were more important for entry level positions. Mid-level positions, on the other hand, emphasized both managerial and technical skills.

Compared with the aforementioned and other similar studies, the focus of this research is on a set of emerging technologies—data warehousing and business intelligence—rather than the overall IS discipline. DW and BI are among the leading information and business performance management technologies in terms of investment priority [6]. Through this study we hope to help reduce the proverbial lag in academic responsiveness to the needs of the industry. Findings of the study should be of use for various stakeholders including IS educators and curriculum committees in curriculum and course design so as to stay current with the emerging technologies.

**RESEARCH METHODOLOGY**

A sample of 180 job posting (60 postings in each of the following three areas: database, data warehousing, and BI) was collected from Dice.com during the period March – June 2009. Dice.com was chosen for its focus on information technology jobs; it was recently ranked first in this category by PC Magazine [3]. The search was conducted for nationwide jobs by specifying “Match all words” using the following terms:

- “Database Management”;
- “Database Administration”; “Junior DBA”; DBA;
- “Data Warehousing”; “Business Intelligence”; BI

Only those ads were used for this study that required a degree in information systems, MIS, or a related filed. Skills listed (both required and desired) for each job posting were recorded in an Excel worksheet. Some of the ads provided a structured list of skills (such as that in Figure 1), while others were more informal or unstructured.

![Figure 1: A sample listing of skills](image)

We conducted *k*-means clustering analysis to determine skill similarity and overlap among the three categories of jobs—database, DW, and BI. *K*-means clustering groups observations into preselected number of non-overlapping clusters [11]. The resulting clusters contain observations that exhibit high within-cluster similarity on selected attributes and high between-cluster heterogeneity. Separate analyses were performed on data combined from two, as well as all three categories. For
example, by using combined DW and BI positions
data and specifying three clusters, one of the clusters
identified by the software contained skills that were
common to both DW and BI categories. Conducting
the analysis by specifying two clusters, on the other
hand, resulted in some observations (job postings)
that were miscategorized by the software because
some of their attribute values were too similar to
another category. A number of DW positions, for
instance, had skill requirements that were quite
similar to BI position requirements and hence were
grouped together with the latter category. Such
postings helped identify skill overlap among
categories, as shown in Figure 2.

Figure 2. Skill Overlap Between Database, Data
Warehousing, and Business Intelligence

RESULTS

Results are summarized in tables 1-6 and discussed in
the following subsections.

Relational Database Skills

Almost all database job descriptions appear to imply
knowledge of foundational relational database
concepts and skills (including database design and
modeling, and SQL programming), though some of
the postings do not explicitly mention them. Some of
these skills are therefore somewhat underrepresented
in terms of their percentages and rankings. For more
advanced database jobs, including those for the
database administration (DBA) positions (both junior
and senior levels DBAs), additional skills are
required. These skills typically include knowledge of
one or more database management systems (DBMS),
a procedural database language such as PL/SQL,
security, performance tuning and back-up and
recovery.

Besides DBA, database development and application
DBA are two other subcategories in this group. These
positions require strong knowledge of programming
and scripting languages, along with skills in some
flavor of UNIX. Table 1 shows top ten relational
database knowledge areas frequently requested.

Using paired samples test, the last column in tables 1,
2, and 5, labeled 'Sig. (p-value)', reports two-tailed
significance in the difference between one ranked
skill and the next. For example, the first p-value
(.003) in Table 1 indicates a significant difference in
ranks between the first two skills, SQL and database
design. The next p-value (.208), suggests that there is
a statistical tie between database design and backup
and recovery skills.

Table 1: Top 10 Database Skills

<table>
<thead>
<tr>
<th>Rank</th>
<th>Knowledge/Skill</th>
<th>Percent</th>
<th>Sig. (p-Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SQL</td>
<td>83.3</td>
<td>.003</td>
</tr>
<tr>
<td>2</td>
<td>Database design (data modeling, and logical and physical database design)</td>
<td>60.0</td>
<td>.208</td>
</tr>
<tr>
<td>3</td>
<td>Backup and recovery</td>
<td>53.3</td>
<td>.850</td>
</tr>
<tr>
<td>4</td>
<td>T-SQL</td>
<td>51.7</td>
<td>.252</td>
</tr>
<tr>
<td>5</td>
<td>PL/SQL</td>
<td>40.0</td>
<td>.858</td>
</tr>
<tr>
<td>6</td>
<td>Performance tuning</td>
<td>38.3</td>
<td>.708</td>
</tr>
<tr>
<td>7</td>
<td>Database security</td>
<td>35.0</td>
<td>.734</td>
</tr>
<tr>
<td>8</td>
<td>UNIX, Linux</td>
<td>31.7</td>
<td>1.00</td>
</tr>
<tr>
<td>9</td>
<td>User requirements</td>
<td>31.7</td>
<td>.708</td>
</tr>
<tr>
<td>10</td>
<td>Programming/scripting languages</td>
<td>28.3</td>
<td></td>
</tr>
</tbody>
</table>

n = 60
Due to significant overlap in skill requirements, the
percentages add to more than 100.

Data Warehousing Skills

Data warehousing skill set builds from the foundation
of relational database knowledge, though sometimes
this fact may not be explicitly noted in a job listing.
Data warehouse or data mart (DM) designs, for
instance, are typically de-normalized or hybrid (with
normalized and de-normalized tables) form of
relational data models, known as star and snowflake
schemas. One specific aspect of this group of skill set
is that skill requirements are more software-specific.
This, in part, is due to the fact that most DBMS vendors such as Oracle, Microsoft, and IBM provide tools for DM/DM design and development as part of their DBMS software. Proficiency in the underlying DBMS is thus essential for any DW job.

As listed in Table 2, extract, transform, and load (ETL), and dimensional modeling emerged among the two most important skill sets in the DW category.

Table 2: Top 10 Data Warehousing Skills

<table>
<thead>
<tr>
<th>Rank</th>
<th>Knowledge/ Skill</th>
<th>Percent</th>
<th>Sig. (p-Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Extract, transform, and load (ETL)</td>
<td>81.7</td>
<td>.024</td>
</tr>
<tr>
<td>2</td>
<td>Dimensional modeling; DW methodologies; DW architecture</td>
<td>65.0</td>
<td>.568</td>
</tr>
<tr>
<td>3</td>
<td>SQL; PL/SQL; other programming languages (Perl, PHP)</td>
<td>61.7</td>
<td>.520</td>
</tr>
<tr>
<td>4</td>
<td>Project management</td>
<td>55.0</td>
<td>.876</td>
</tr>
<tr>
<td>5</td>
<td>Data security; backup; recovery; performance tuning</td>
<td>53.3</td>
<td>.104</td>
</tr>
<tr>
<td>6</td>
<td>Master data management (MDM); data quality; metadata management</td>
<td>36.7</td>
<td>.552</td>
</tr>
<tr>
<td>7</td>
<td>UNIX; Linux</td>
<td>31.7</td>
<td>.718</td>
</tr>
<tr>
<td>8</td>
<td>User requirements</td>
<td>28.3</td>
<td>.854</td>
</tr>
<tr>
<td>9</td>
<td>Analytics</td>
<td>26.7</td>
<td>.320</td>
</tr>
<tr>
<td>10</td>
<td>Reporting</td>
<td>18.3</td>
<td></td>
</tr>
</tbody>
</table>

n = 60

Most DW jobs require vendor-specific skills in DW and DBMS. Tables 3 lists top five DW vendors along with their DW/DB software.

Table 3: Top Five Data Warehouse Vendors /Products

<table>
<thead>
<tr>
<th>Rank</th>
<th>DW Vendor/ Product</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Microsoft: SQL Server</td>
<td>69.2</td>
</tr>
<tr>
<td>2</td>
<td>Oracle: Oracle 10g &amp; 11g</td>
<td>13.7</td>
</tr>
<tr>
<td>3</td>
<td>IBM: DB2</td>
<td>11.3</td>
</tr>
<tr>
<td>4</td>
<td>Teradata</td>
<td>3.5</td>
</tr>
<tr>
<td>5</td>
<td>SAP BW (Business Warehouse)</td>
<td>2.3</td>
</tr>
</tbody>
</table>

n = 60

Since ETL tools emerged as one of the frequently sought skills in the DW and BI categories, we conducted a separate job search for each of the leading ETL tools identified in the sampled job listings. Table 4 shows relative market share in terms of jobs posted for each of these tools.

Table 4: Top Seven ETL Tools

<table>
<thead>
<tr>
<th>Rank</th>
<th>ETL Tool</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Informatica</td>
<td>40.5</td>
</tr>
<tr>
<td>2</td>
<td>Microsoft: SQL Server: SSIS</td>
<td>32.4</td>
</tr>
<tr>
<td>3</td>
<td>Oracle: Oracle Warehouse Builder (OWB); Oracle Data Integrator</td>
<td>12.0</td>
</tr>
<tr>
<td>4</td>
<td>IBM: DataStage; IBM Ascential</td>
<td>5.0</td>
</tr>
<tr>
<td>5</td>
<td>Ab Initio</td>
<td>4.8</td>
</tr>
<tr>
<td>6</td>
<td>Business Objects: Data Integrator</td>
<td>4.7</td>
</tr>
<tr>
<td>7</td>
<td>SAS: Data Integrator</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Business Intelligence Skills

Table 5 lists top BI skills along with their rankings. As summarized in Figure 2, there is considerable overlap among foundational database, DW, and BI skills. Also, as the p-values indicate, the difference in ranks between any two consecutively ranked skills is statistically insignificant.

Table 5: Top Ten Business Intelligence Skills

<table>
<thead>
<tr>
<th>Rank</th>
<th>Skill</th>
<th>Percent</th>
<th>Sig. (p-Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OLAP; querying; reporting</td>
<td>68.3</td>
<td>.260</td>
</tr>
<tr>
<td>2</td>
<td>Extract, transform, and load (ETL)</td>
<td>58.3</td>
<td>.764</td>
</tr>
<tr>
<td>3</td>
<td>User requirements; project management</td>
<td>55.0</td>
<td>.276</td>
</tr>
<tr>
<td>4</td>
<td>Dashboards; scorecards; metrics</td>
<td>43.3</td>
<td>.880</td>
</tr>
<tr>
<td>5</td>
<td>Dimensional modeling, DW architecture</td>
<td>41.7</td>
<td>.740</td>
</tr>
<tr>
<td>6</td>
<td>Analytics; data mining</td>
<td>38.3</td>
<td>.848</td>
</tr>
<tr>
<td>7</td>
<td>UNIX, Linux</td>
<td>36.7</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>SQL; PL/SQL; other programming languages</td>
<td>36.7</td>
<td>.616</td>
</tr>
<tr>
<td>9</td>
<td>Security</td>
<td>31.7</td>
<td>.718</td>
</tr>
<tr>
<td>10</td>
<td>Metadata management</td>
<td>28.3</td>
<td></td>
</tr>
</tbody>
</table>

There are a number of distinct subcategories within the BI category including business analytics and data
Data Warehousing & BI Skills for IS Graduates

Volume X, No. 2, 2009

337

Issues in Information Systems

mining, querying and reporting, and designing, building, and operating data visualization and business performance management (BPM) tools such as dashboards and scorecards. Popular BI software such as IBM Cognos and SAP Business Objects are well-equipped for performing these activities, excluding data mining. Data mining tools and software packages are generally available separately from many different vendors. Skills required of IS graduates in these tools often include knowledge of input data formats, types of analysis to perform, and interpreting and using the output for decision support.

A number of position announcements required skills in specific BI tools. It therefore makes sense that besides learning general concepts and methodologies IS graduates are also trained in popular BI software. Table 6 lists top five vendors along with their BI software products and relative share in the job market. This list does not include Microsoft SQL Server since essential DB, DW, and BI features of SQL Server are well integrated and it’s hard to estimate demand for each tool/component separately.

Table 6: Top Five BI Vendors and Their Products

<table>
<thead>
<tr>
<th>Rank</th>
<th>BI Vendor/Product</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SAP: Business Objects</td>
<td>65.3</td>
</tr>
<tr>
<td>2</td>
<td>IBM: Cognos</td>
<td>15.3</td>
</tr>
<tr>
<td>3</td>
<td>Oracle: Business Intelligence Enterprise Edition (OBIEE); Siebel Analytics; Hyperion Essbase</td>
<td>12.4</td>
</tr>
<tr>
<td>4</td>
<td>MicroStrategy</td>
<td>5.1</td>
</tr>
<tr>
<td>5</td>
<td>SAS BI</td>
<td>1.9</td>
</tr>
</tbody>
</table>

The rankings in Table 6 appear to be consistent with those published by industry groups. For example, the latest rankings of BI platforms based on vendor surveys and customer perceptions of vendors, Gartner’s ‘magic quadrant’ [14] lists the following six BI platforms as the leaders and the most visionary in the field: IBM, Microsoft, SAP, Oracle, Information Builders, SAS, and MicroStrategy. Similarly, market share based ranking of BI vendors by OLAP Report [12] lists the following vendors in descending order of the market share: Microsoft, Oracle, SAP, IBM, and MicroStrategy.

Qualitative Analysis

Qualitative analysis of the job listings reveals three patterns: (1) there is a significant overlap among the three categories of skills in database management, DW, and BI; (2) there is some degree of sequentiality implied in acquiring these skills – from foundational database management to data warehousing, and then to business intelligence; and (3) two natural groupings of skills are clearly identifiable: software- or vendor-specific, and generic skills. Figures 2 and 3 portray an integrated view of these findings. While Figure 2 lists significant skills and skill overlaps in the three job categories, Figure 3 shows sequentiality and interrelationships between the job categories, and lists generic skill sets along with vendor-specific software tools.

This analysis should be of help in curriculum and course design (and for prospective advanced data management students) as it identifies topic areas that should be covered in one course that forms a prerequisite to another.

The interrelationships between various groups of skills in Figure 3 can be described in terms of their progressively advanced nature as one moves from the foundational database skills to those required for DW and BI jobs. Also, higher level skills are more and more tied with their underlying technologies, i.e., they are vendor-specific.

Figure 3. A Comprehensive View of Database, Data Warehousing, and Business Intelligence Skill Sets

CONCLUSIONS

Using online job listings in data management fields, including relational database management, data
warehousing, and business intelligence, this study has attempted to document required and desired skills, their interrelationships, and interdependence. Data warehousing, and business intelligence courses can be offered as senior level undergraduate electives and elective or required courses in various graduate IS programs. This study should be of use to IS students, educators, and curriculum committees. Some of the important findings of the study are highlighted below.

- **Structured query language (SQL)** is one of the skills that permeate throughout the field of data management – from transactional databases to data warehousing, to online analytical processing (OLAP) and BI. Most major DBMS, including those from Oracle, IBM, Microsoft, and Teradata, emphasize SQL skills that build from the basic SQL-3 and 2006 standards to their customized versions, specifically tailored for multidimensional queries and OLAP. Microsoft’s multidimensional extensions (MDX) is such an example.

- **Skills in data modeling and relational database design** are another set of skills that are used successively in DW and DM design and when planning for DW/BI drill-down and rollup capabilities. Popular DW/DM designs today are based on Kimball methodology [9] that extends entity-relationship modeling to de-normalized or hybrid designs in the form of star and snowflake schemas.

- **For BI and business performance management (BPM) application design or development, one needs a set of skills that are not typically taught in database courses. Some of these skills include knowledge of methodologies and tools for balanced scorecards, six sigma, and performance dashboards. For instance, designing and building dashboards using one of the rapid application development (RAD) tools, requires knowledge of appropriate metrics along with some level of dexterity in the underlying programming language. As an example, one can use Oracle Application Express (APEX) along with PL/SQL to design and build a dashboard.**

- **Due to the fast evolving nature of DW/BI technologies, skills associated with some of the latest software tools were not mentioned in job listings frequently enough to be included in one of the clusters and were treated as outliers. APEX is one such example.**

- **Although ETL skill requirements are almost always vendor-specific, there is an underlying common body of knowledge including integration concepts, data types and formats based on multiple platforms and data sources, units of measures, and underlying business processes. For example, what should a global organization do to represent its revenue in a single currency at a specific point in time? Questions like these require decisions about which common currency and exchange rates to use and whether the company plans to take advantage of ‘transfer pricing’ for tax purpose. In short, knowledge of the underlying business domain is essential along with any technical skills highlighted in this study.**

- **Most job listings include verbal and written communication skills as required or desired qualifications. This study focused on technical skills with an underlying assumption that communication skills are essential for all jobs.**

**REFERENCES**


