The Usage of Best Practices and Procedures in the Database Community

ABSTRACT

Database Management has an important role to play in the management of data assets which are at the heart of every organization. In a fast moving technological era, where data is rapidly expanding, understanding the current best practices and procedures is important for continuous improvement. This paper investigates how databases are actually administered and identifies what practices and procedures are utilized throughout the database lifecycle. The paper highlights the demographics of people who manage database systems and the diverse requirements of database systems given the wide range of software and hardware available. The results of this paper show the breadth of issues relevant to database management. The paper concludes by showing where existing practice and procedures are not optimal, and by highlighting the complexities in the field.

Keywords: Database Administration; Database Management; Data Management; Database Lifecycle; Best Practice & Procedures; Database Operations

author mail id: holt.victoria@gmail.com

INTRODUCTION

Database management is, at its heart, the administrative tasks associated with the storage, modification and retrieval of data held within a database management system (DBMS). Organizations today require impeccable database management in order to maintain a high quality of data, and for that data to be secure and available whenever it is required. The data for governments, banks, financial institutions etc. must also satisfy statutory legal requirements.

It has been well publicised that problems exist with the setting up and successful operation of important databases. An early report by Blasis (1977) highlighted problems with administration, organizational issues, new technology introduction, control and technical configuration. There are many reasons for these problems. The use of appropriate practices and procedures can have a significant impact on the availability, recoverability and quality of data used in the operations of businesses. The diversity of an organization's domains and strategies can lead to a variety of practices and procedures.

Certain practices used by organizations can be considered 'best practice'. Best practices are frequently described as those which are recommended for carrying out actions for desirable outcomes (see Figure 9 below). Best practices drive operational excellence and effectiveness (Dembowski, 2013). Other key terms used here are: processes – "a series of actions or steps taken in order to achieve a particular end"; procedures – "an established or official way of doing something"; and methodology – "a system of methods used in a particular area of activity" (all definitions from Oxford Dictionaries, 2014).

This paper presents the findings from a comprehensive survey which investigated to what extent best practices and procedures are utilized by the database community.

BACKGROUND TO THE RESEARCH

Database management has evolved over the last five decades, since the first functioning prototype DBMS (Haigh, 2012) , to become an integral part of most organizations' business. The majority of global organizations today cannot operate without a functioning database. Organizations increasingly realize the value of the data that they hold and are beginning to draw more benefit from its analysis and mining ("Data, Data everywhere," 2010). Some examples of usage include shopping history to predict purchases (Cuddeford-Jones, 2013), social media to predict trends (Schoen et al., 2013), live mapping for disaster aid (Gao et al., 2011), the storage of the human genome to aid medical research (Ballew et al., 1998), sensor data from CERN (Segal et al., 2000) and other data intensive scientific discoveries (Hensley et al., 2014; Hey et al., 2009).

Changes in culture and everyday life have brought about the sharing of more information, and this has radically changed the usage of databases. New technologies such as cloud and virtualization enable a different operating model, one that allows organizations to share resources, but these new models add to the complexity of an already complicated activity.

The trend of rapidly increasing data volumes is being driven not only by the requirements of government and business to store more information, but also by the digitization of film and TV and the use of social media. Gantz & Reinsel (2010) have estimated that the volume of all known data will have grown from 0.8 zettabytes (10^{21} bytes) in 2009 to 35 zettabytes in 2020. This is a mixture of both structured and unstructured data. Unstructured data, objects that have little or no identifiable structure, e.g. text, images, audio and video, were not previously considered within the database community. However this view has now changed within the industry as a whole. The Lowell Report (Abiteboul et al., 2005), a summary of a gathering of academic database researchers' discussions on the state of database research, states that "Database needs are changing, driven by the internet and increasing amounts of scientific and sensor data" (Abiteboul et al., 2005, p. 111).

The administration and management of this complex area would benefit from a better understanding of the extent to which best practices and procedures are utilized by the database community. However, that raises the question as to whether the adoption of best practice is constrained by the many interactions between different the interconnected aspects of the management of database systems.

The database management system (DBMS) itself is constructed of many components which can be considered to form a layered technical system. Typically, disparate organizational teams manage each layer independently. These teams have different sets of goals, together with a variety of approaches, and problems in operation which can occur due to the interconnections. This has increased the challenges of overall management.

Many organizations demand low cost infrastructure without jeopardizing functionality or operational ability. Database architecture, design and development are the foundation of any well designed DBMS. The database is continually evolving and adapting to the demands of the users, organizations and the global environment.

The management of the database and the data contained within it are often undertaken by different teams. Two separate functions were identified by Kahn (1983, p. 794) as database administration and data administration. However, they have many interconnected components. Organizations' business requirements for data collection and manipulation appear to be driving the requirements for database management. The fields are gradually merging, and Mullins (2012) proposed data administration practices and procedures to address this, arguing that "when database administration is treated as a management discipline, the treatment of data within your organization will improve" (Mullins, 2012, p. 9). A management discipline is required to manage the successful operation of the databases. The management methods used for development and infrastructure projects has also changed. Historically, development projects have been dominated by Waterfall methodologies, but there has been some shift towards Agile methodologies to overcome Waterfall's inflexibility in relation to rapidly changing business requirements (Verner & Babar, 2004).

Changes to the database engine, structure and hardware all require practices and procedures to ensure that data is protected. The Independent Oracle Users Group (IOUG) has recently completed several surveys that aid in understanding the complexities of database systems. The IOUG report (Mckendrick, 2011a) entitled 'Managing the Rapid Rise in Database Growth' identified the importance of database change management practices.

There are many current complexities related to data management and data administration, and as Aiken et al., (2011) suggest, data management is still evolving. The Claremont report on database research (Agrawal et al., 2009, p. 65) highlighted concerns that are important to the community regarding the increasing technical scope, processes and keeping track of the field that is important to the community. Other surveys previously undertaken provided some insight, and highlighted the rise of database administration, with an unclear direction of the future path (Aiken et al., 2011; Gillenson, 1982, 1985, 1991; McCririck, I.B., & Goldstein, 1980; Mckendrick, 2013).

The survey reported here highlights the real world situation of the database community at present, and shares the current practices and procedures of the respondents.

THE SURVEY

The primary goal of the research was to investigate how databases were actually administered and to identify what practices and procedures were utilized throughout the database lifecycle. Further aims were to understand the demographics of people who manage database systems, and to investigate how they learned about best practice and whether any IT frameworks were used. As the database community is dispersed globally, the survey has sought to reflect this global nature. The anonymized data is on www.sqltoolkit.co.uk/publications.htm.

Sampling

Non-probability convenience sampling was selected to allow all those who were willing and able to participate in the survey. It is impossible to know the size and dispersion of the database-management and data-professionals population. Convenience sampling (Bradley, 1999); (Buckingham & Saunders, 2009); (Denscombe, 2008) is commonly used during preliminary research to gain a summary of interesting information. It allows the data to be collected quickly and inexpensively.

The sample was obtained through advertising the survey via social media such as Twitter, LinkedIn and Facebook groups, email newsletters and blog posts. A strategic decision was taken to gather data as widely as possible across the database population, to include a range of geographic locations, job roles and database software used. For convenience sampling the larger the sample size the better, in order to get an overall picture. The primary disadvantage of this sampling method is that there is no guarantee of a representative sample; hence it is not possible to make generalizations of the entire population from the results. Any patterns found in the data could potentially just reflect patterns in the sample and not be related to the entire population.

Method

A web-based survey was designed to obtain an indication of the current state of practices and procedures within database management. The primary research question under examination was: “To what extent are best practices and procedures utilized by the database community?”

The survey was open for data collection between 13 December 2012 and 6 February 2013. A total of 453 respondents (n = 453) participated from within the global community of database and data professionals.

The questionnaire comprised 83 questions of different types, including dichotomous, nominal (multichotomous), interval level (Likert scale) and multi-option responses. Several free text questions were also included to allow the respondents to explain items in more depth.

Consideration was given to the problem that inconsistency of data might occur within the survey. This was addressed by adding a few questions to the survey which would enable triangulation of some of the data. There are various types of triangulation (Guion, 2002; Jick, 1979; Thurmond, 2001). This research used methodological triangulation to cross check the consistency and reliability of the survey responses. For validation the same people were asked different questions on the same aspect of the research, which provided limited but additional validation of the data.

The survey instrument used was Survey Monkey (www.surveymonkey.com).

FINDINGS

The survey investigated the current practices and procedures that are used, together with respondents' perspectives on 'best practice'. The main areas of the survey comprised: demographics of respondents; respondents' organizations; training; database servers and database demographics; understanding of best practice; database architecture, design and development; database technical practices; data management; application-centric perspective; change management; and organizational culture. Findings related to each of these areas are discussed in turn in this section.

Demographics

The respondents came from a diverse spread of job roles. The three largest groups were database administrators (43%), database developers (15%) and business intelligence (BI) roles, including BI analysts, data analysts, data scientists and BI architects (13%). In all, the population was split into 28 different classes covering all levels of management.

Respondents worked in 40 countries, with the majority based in the USA (40%) and the UK (33%). The remaining 27% were divided amongst 38 countries, with no single country greater than 10%.

The respondents worked in 34 different industry sectors, of which the technology sector accounted for 25%, banking, insurance or financial services 21%, healthcare 12%, professional services 12% and education 10%.

As Table1 shows, 30% of the repondents worked in large organizations with over 2,500 employees. Every category of organizational size was represented among the survey respondents.

TABLE 1 SIZE OF ORGANIZATION'S WORKFORCE

What is the approximate size of your organization's total workforce?

Total Workforce	Response Percent	Number of Responses
Less than 5	4.9%	22
5 – 19	6.9%	31
20 – 49	6.0%	27
50 – 99	4.5%	20
100 – 249	9.6%	43
250 – 499	11.8%	53
500 – 999	9.8%	44
1,000 – 2,500	11.8%	53
Over 2,500	30.5%	137
Self-employed consultant	3.1%	14
Don't know	1.1%	5

Slightly more than half the respondents (53%) had worked in the database field for over 10 years, with 29% working in the field between 5 and 10 years (Figure 1).

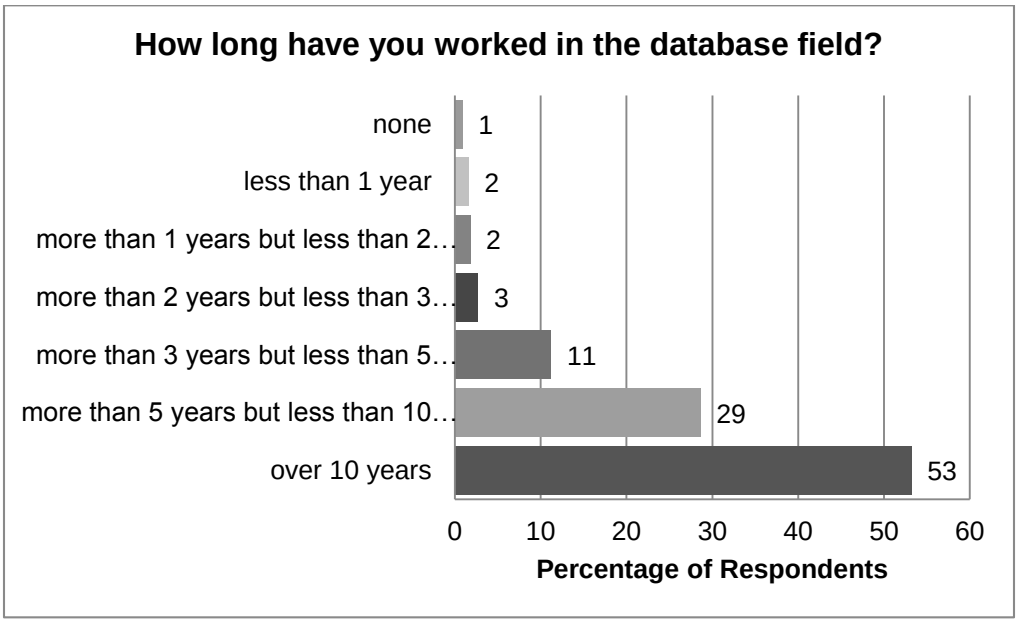

FIGURE 1 EXPERIENCE IN THE FIELD

Respondents' Organizations

As Table 2 shows, 33% of respondents had 10-50 database servers in their organization, with 25% reporting up to 10 database servers.

TABLE 2 ORGANIZATION'S NUMBER OF DATABASE SERVERS

What is the approximate number of database servers in your organization?

Number of Database Servers	Response Percent	Number of Responses
up to 10	25.1%	112
10-50	32.5%	145
51-100	12.8%	57
101-500	11.4%	51
501-1000	3.8%	17
1000+	7.2%	32
Don't know	7.2%	32

As shown in Table 3, 51% of respondents reported that there were between 2 and 5 people who administered databases in their organization, although 18% of respondents reported that there was just one person. (Note that this role may not have been confined to those with a job title of Database Administrator.)

TABLE 3 PEOPLE ADMINISTERING THE DATABASES

How many people administer the databases in your organization?

Number of Administrators	Response Percent	Number of Responses
1	17.8%	79
2 - 5	50.8%	226
6 -10	12.1%	54
11-15	3.8%	17
16-20	2.7%	12
21-30	1.8%	8
31-40	0.0%	0
41-50	0.4%	2
51 or More	5.2%	23
Unknown	5.4%	24

Turning to database-related roles found in respondents' organizations, 78% of the respondents stated that their organizations had the roles of database administrators or database engineers; 58% of organizations had database developers and 44% had staff holding Business Intelligence roles.

Time spent by respondents in managing database servers varied (Table 4), with 7% of respondents spending all their time on these tasks, while 43% spent less than a quarter of their time on this.

TABLE 4 TIME SPENT MANAGING DATABASE SERVERS

What percentage of your time is spent managing database servers?

Percentage of Time	Response Percent	Number of Responses
None	8.4%	35
1 - 25 %	43.3%	180
26 - 50 %	16.8%	70
51 - 75 %	12.0%	50
76 - 99 %	11.5%	48
All	7.0%	29
Unknown	1.0%	4

Other	2

Training

Database training was obtained through a variety of methods. 85% of respondents read articles when required, and 59% of respondents receiving training from external conferences. 12% of respondents reported that no training was provided (Figure 2).

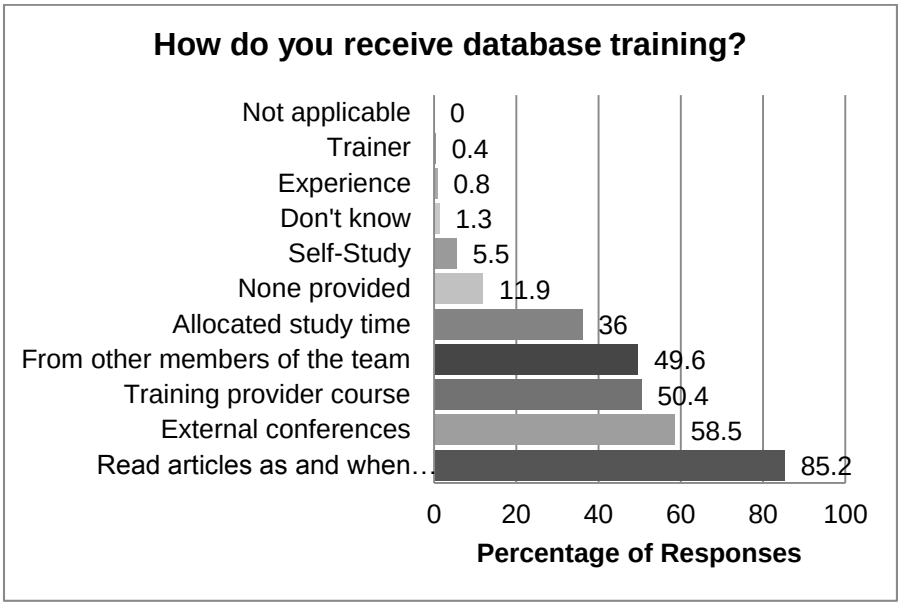

FIGURE 2 RECEIPT OF DATABASE TRAINING (NOTE: PERCENTAGES DO NOT TOTAL 100% BECAUSE RESPONDENTS COULD CHECK ALL THAT APPLY)

Although 46% of the respondents had professional certifications; 53% of respondents said their company never or rarely encouraged professional certification (Figure 3).

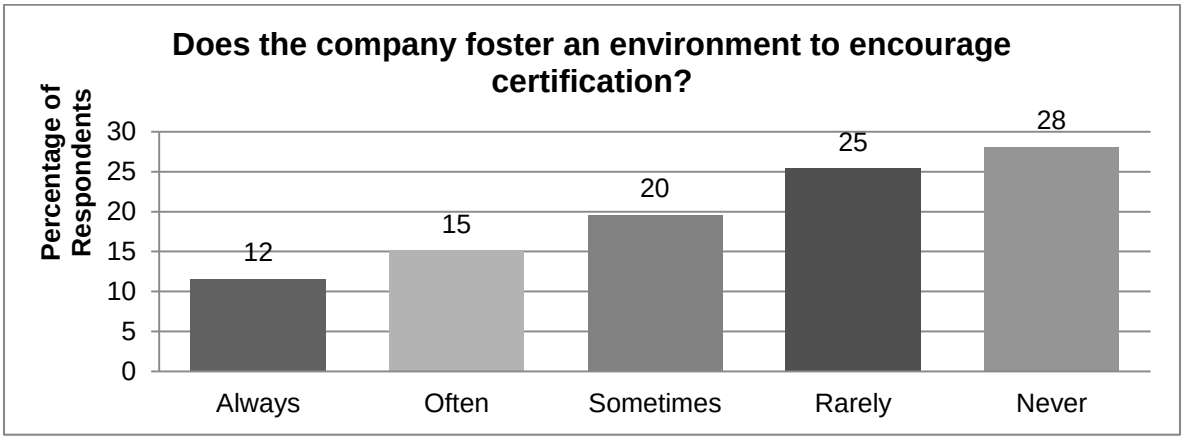

FIGURE 3 ENCOURAGEMENT FOR TAKING CERTIFICATIONS

36% of respondents had the opportunity to undertake formal training once a year (Figure 4) yet 23% never had this opportunity. This compared with 11.9% for whom no training was provided (Figure 2).

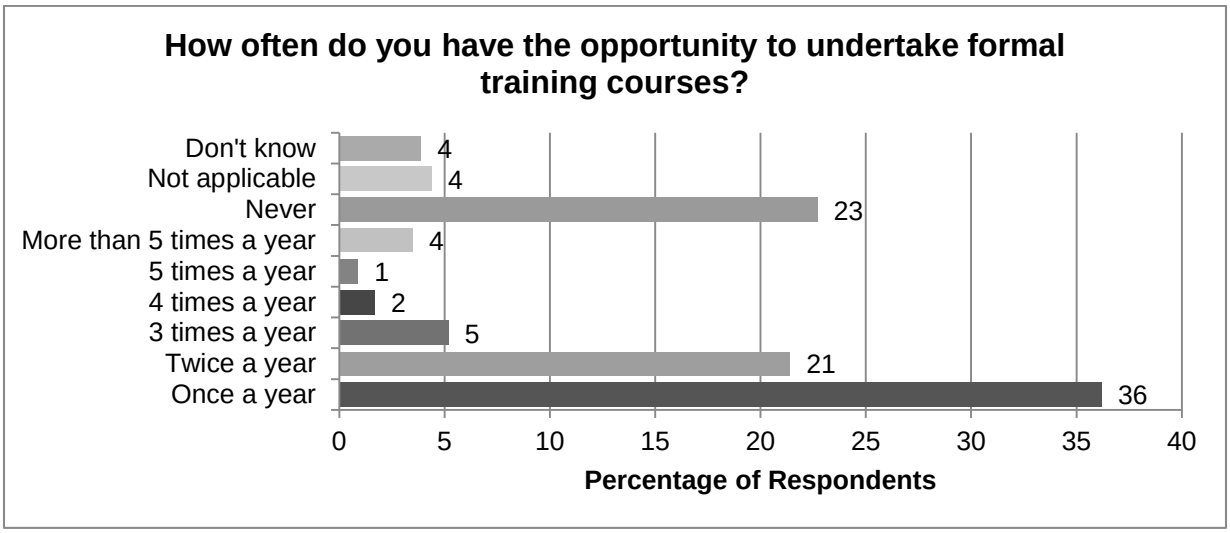

FIGURE 4 OPPORTUNITY FOR FORMAL TRAINING

Many respondents took part in database community associations. These associations help share knowledge in the field through technical tips and training and provide networking connections. Only 23% of respondents were *not* involved in any user groups.

33% of respondents had a chance to attend conferences, workshops or seminars once a year, with others able to attend more often. However 24% were never able to attend conferences (Figure 5). Comments in the free-text section relating to this question included: “once every two years”, “not in a long time”, “irregularly” and “at my own expense”.

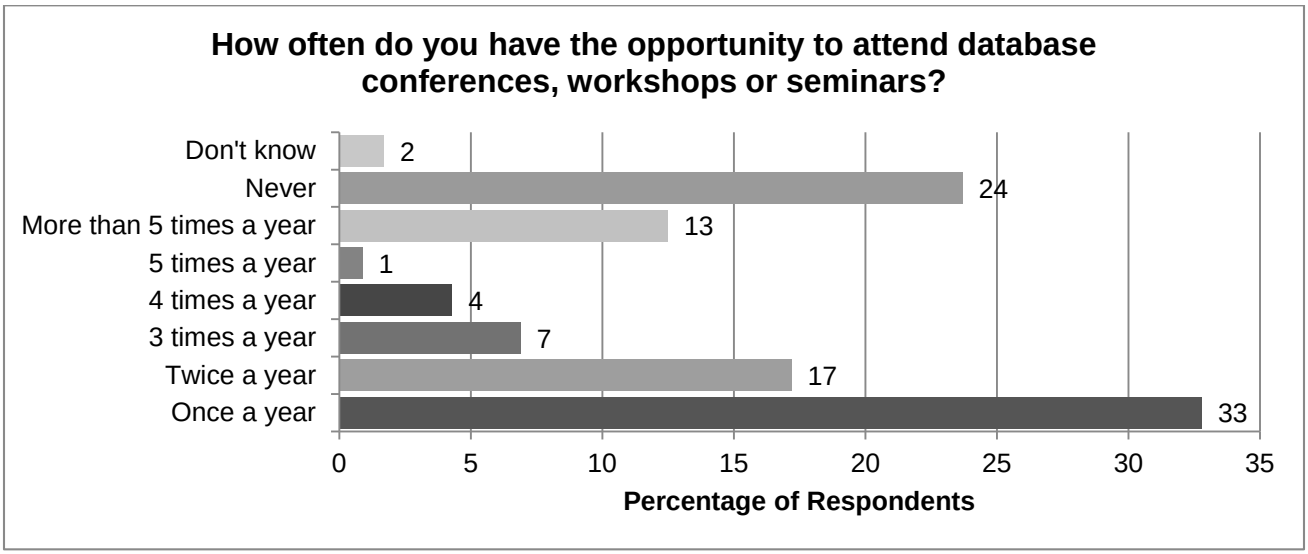

FIGURE 5 OPPORTUNITY TO ATTEND CONFERENCES

Database Servers and Database Demographics

For 55% of the respondents, the largest database that they managed was between 101 GB and 5TB (Figure 6).

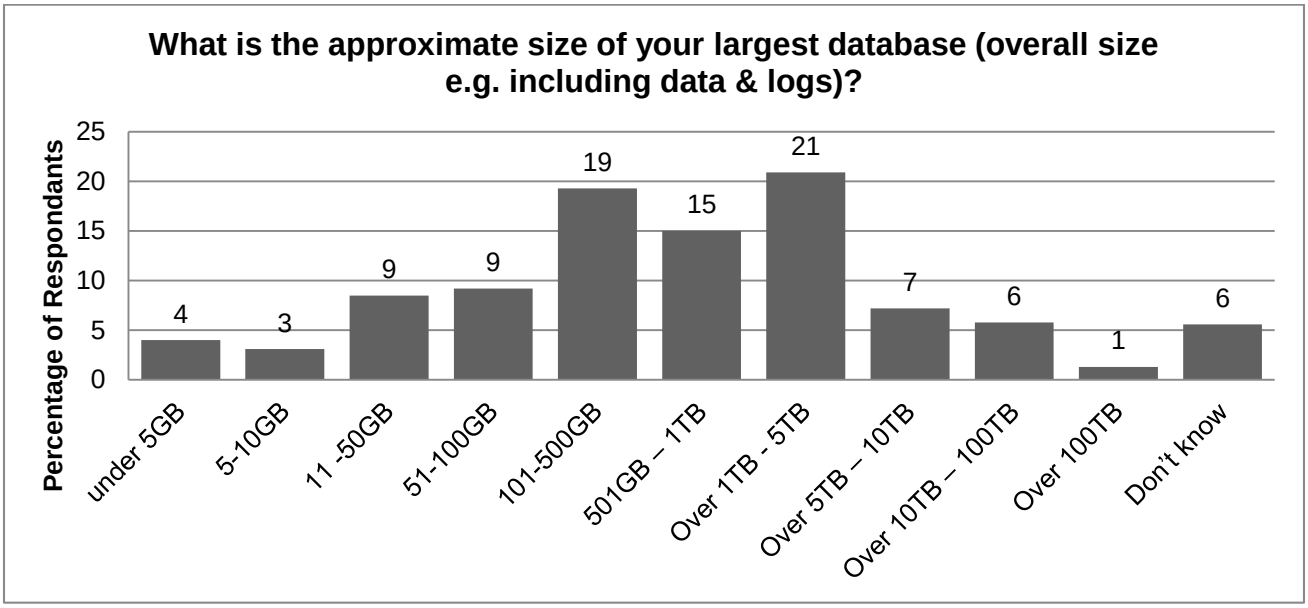

FIGURE 6 SIZE OF THE SINGLE LARGEST DATABASE

51 different database applications were used, including relational, NoSQL, NewSql, InMemory and cloud database applications. Microsoft SQL Server was the most frequently reported database application (89% of respondents), followed by Oracle (44%) and MySQL (40%), (Figure 7).

On average a respondent used 2.6 database applications, while 28% of respondents used just one database application.

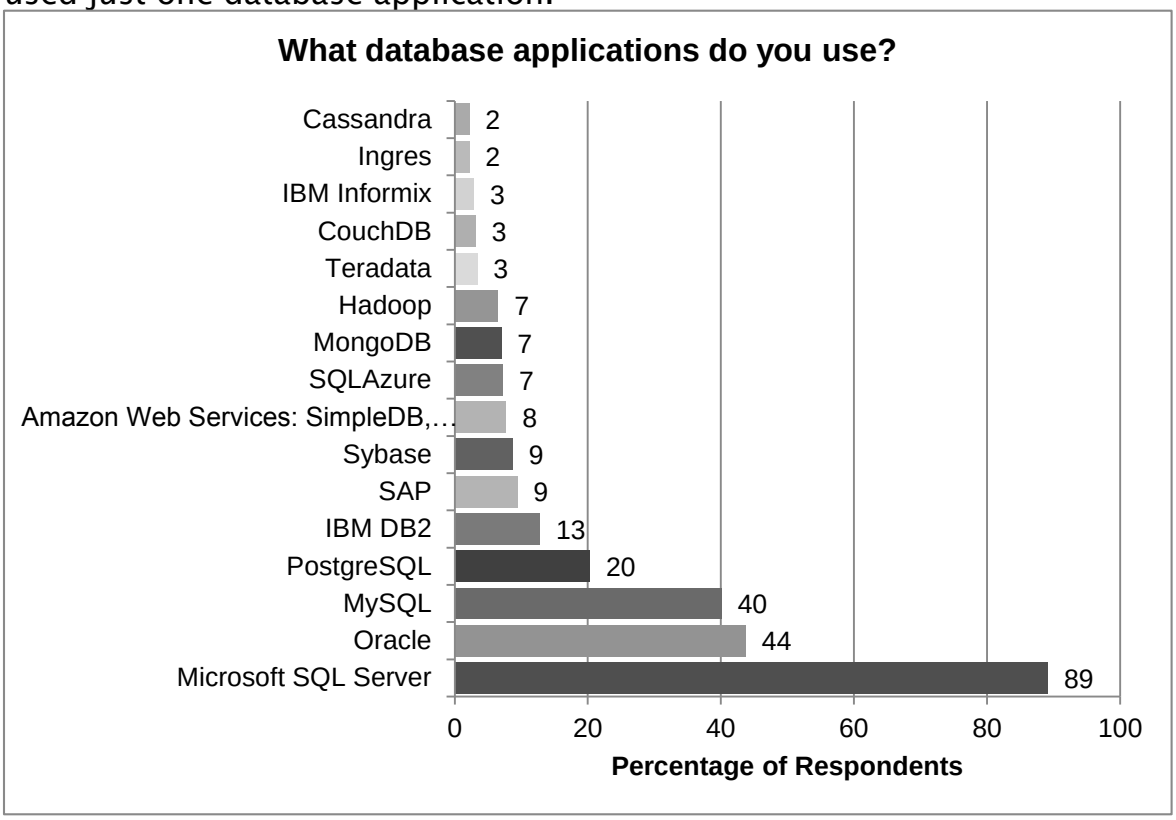

FIGURE 7 DATABASE APPLICATIONS USED (NOTE: PERCENTAGES DO NOT TOTAL 100% BECAUSE RESPONDENTS COULD CHECK ALL THAT APPLY)

89% of respondents used physical database platforms, 83% used virtual database platforms and 18% used cloud consumer services for their database platforms (Figure 8). However, 66% of respondents reported that none of their databases used cloud database services (Table 5).

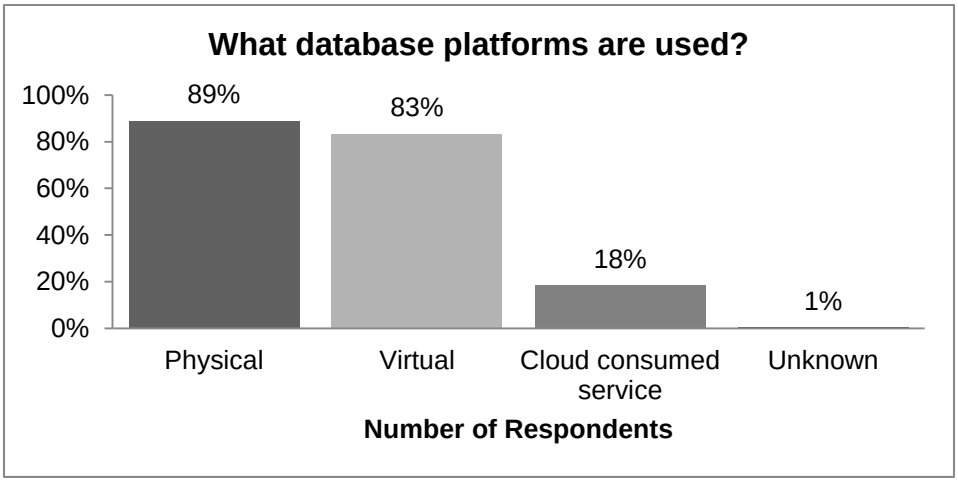

FIGURE 8 DATABASE PLATFORMS USED (NOTE: PERCENTAGES DO NOT TOTAL 100% BECAUSE RESPONDENTS COULD CHECK ALL THAT APPLY)

TABLE 5 CLOUD DATABASE SERVICES USED

What percentage of your databases use cloud database services (databases which are accessible via public, private or hybrid cloud instantly, on-demand, e.g. SQL Azure)?

Percentage of Databases which use Cloud	Response Percent	Number of Responses
None	65.9%	273
1 - 25 %	22.7%	94
26 - 50 %	4.3%	18
51 - 75 %	1.2%	5
76 - 99 %	0.7%	3
All	1.0%	4
Unknown	4.1%	17

Best Practice

Best practice can mean different things to different people. Respondents selected a definition of best practice from a list of definitions (Figure 9). The most frequently selected definition was “recommended practice”.

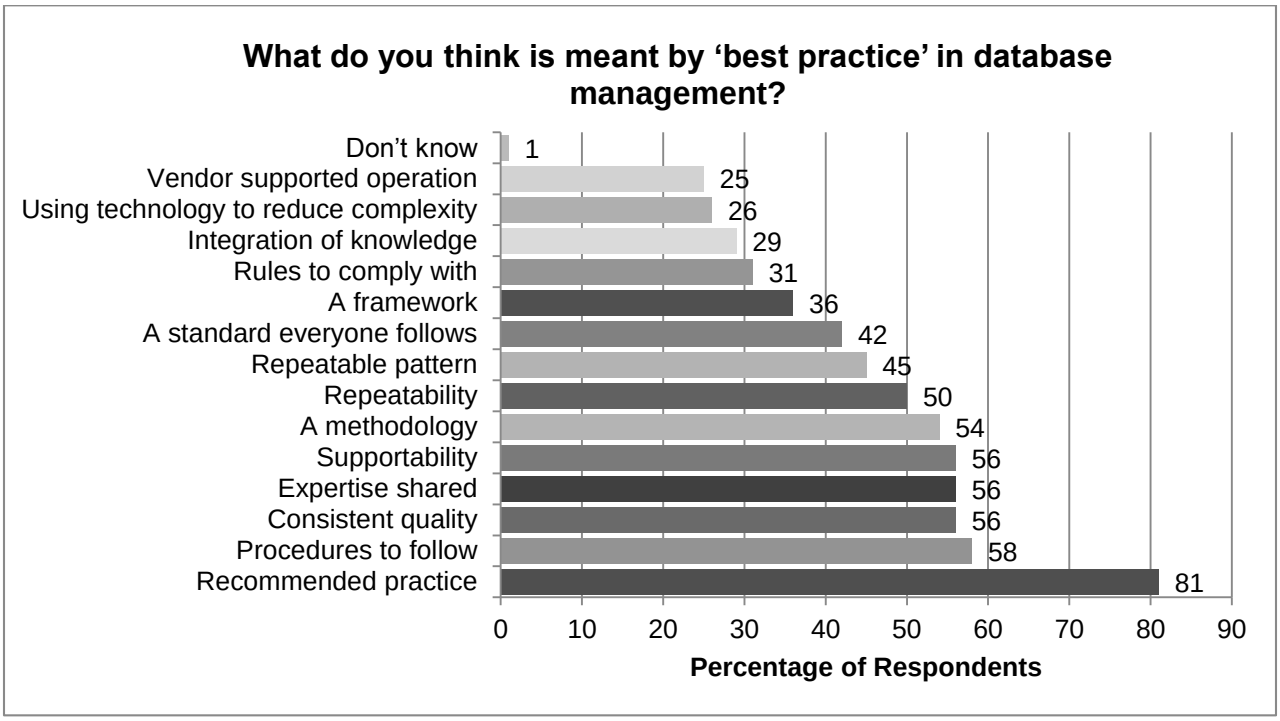

FIGURE 9 FREQUENCY TABLE OF BEST PRACTICE (NOTE: PERCENTAGES DO NOT TOTAL 100% BECAUSE RESPONDENTS COULD CHECK ALL THAT APPLY)

An additional comment from a respondent answering this question highlighted the complexity: "Best practices are NOT set in stone, simply because project requirements are always differing and technologies that sit on top of the database layer change, which can create new best practices."

Best practice guidelines used by respondents originated from a variety of sources including, but not limited to: industry whitepapers, books, presentations, software vendor websites such as Microsoft, organization's documented procedures, current usage patterns, blogs, colleagues, user groups and conferences (Figure 10).

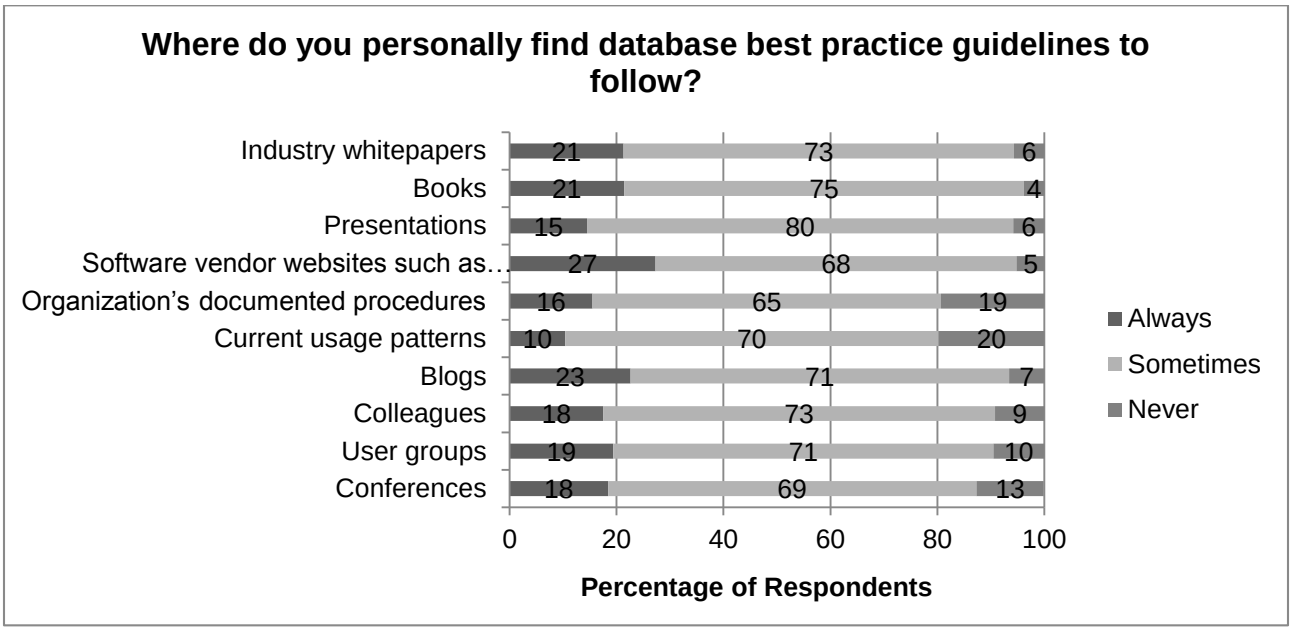

FIGURE 10 WHERE DO YOU FIND BEST PRACTICE

42% of respondents reported that their organizations followed best practice guidelines, through creating their own best practices. 10% said that their organizations did not follow best practice guidelines (Figure 11).

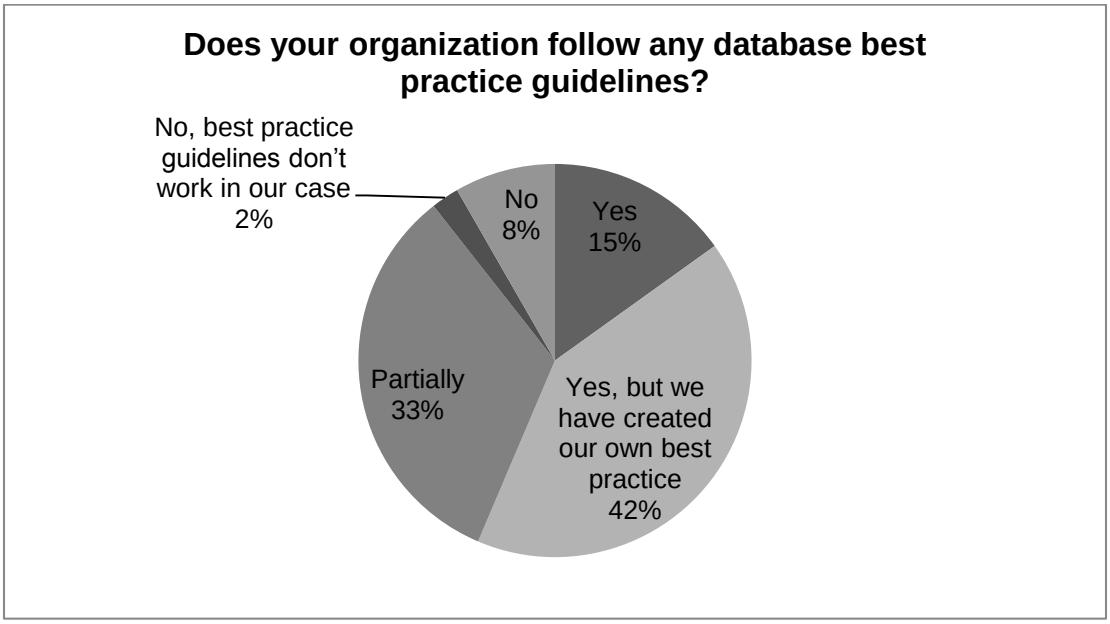

FIGURE 11 ORGANIZATION FOLLOWING BEST PRACTICE

An overwhelming 94% of respondents thought it was important to have best practices (Figure 12). 80% of respondents thought that following best practice was a labour intensive process. 63% of respondents thought that following processes could be obstructive to best practice.

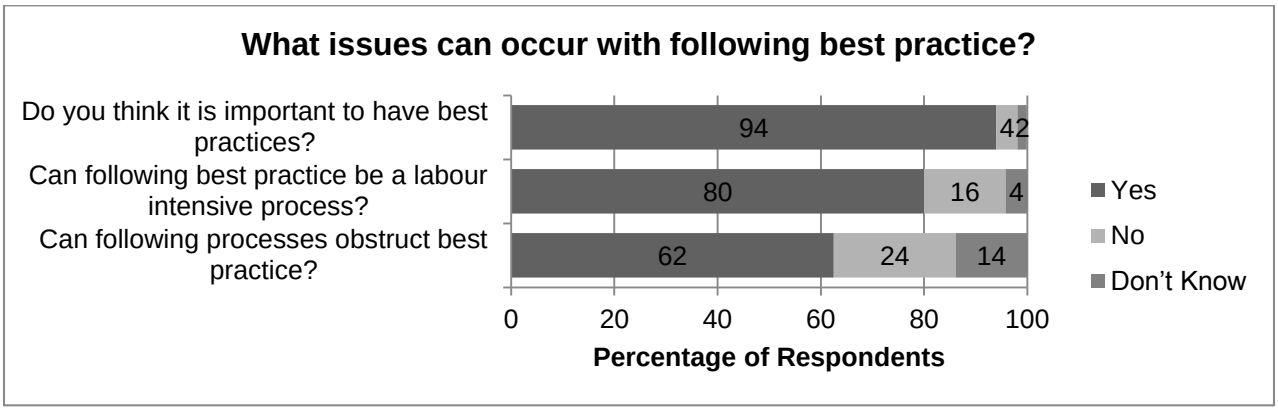

FIGURE 12 ISSUES WITH BEST PRACTICE

The extent to which best practice was controlled within the respondents' organizations is shown in Figure 13. The most highly controlled areas were: Database security; and High availability resilience and disaster recovery.

The areas the participants reported as most uncontrolled were: Cloud database design and security; and Cloud database service management.

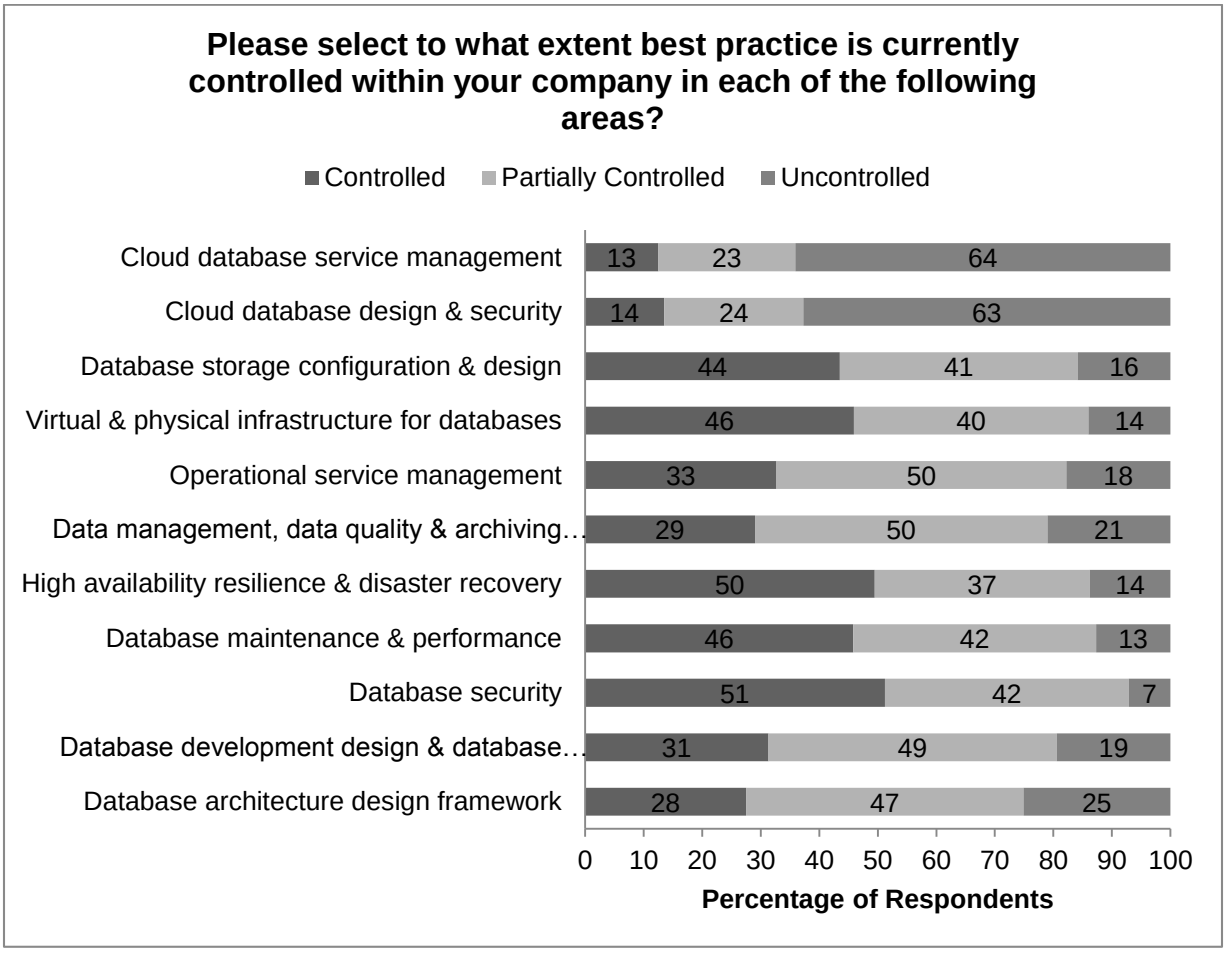

FIGURE 13 CONTROL OF BEST PRACTICE

Database Architecture, Design and Development

The results in Figure 14 indicate that 60% of the respondents do not use common industry standard architecture frameworks for database design. However 18% do use documented design patterns.

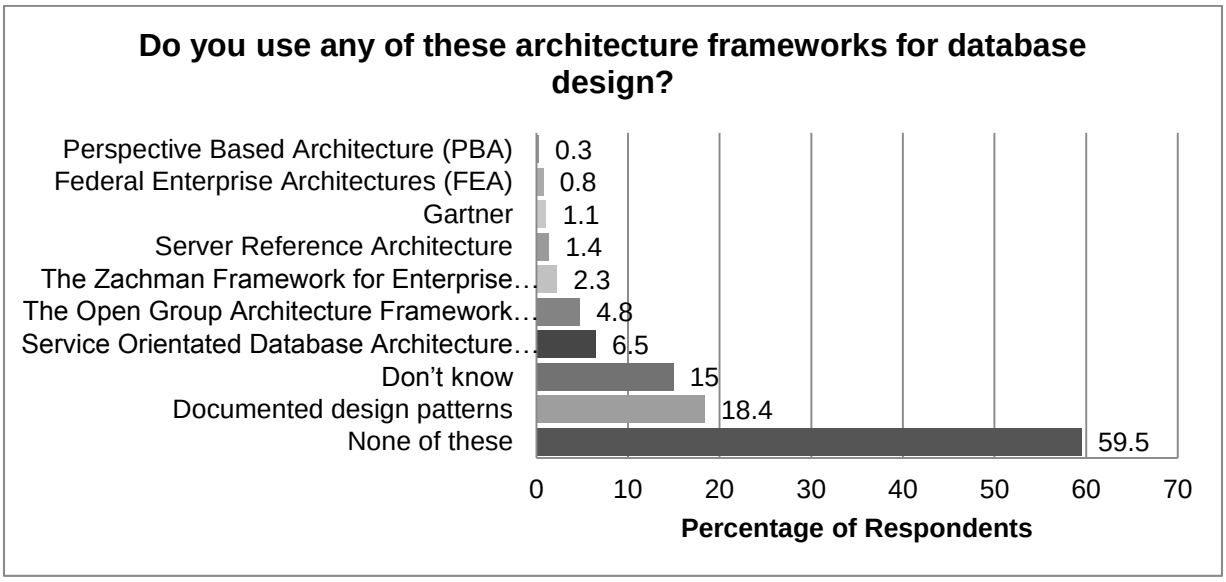

FIGURE 14 ARCHITECTURE FRAMEWORKS (NOTE: PERCENTAGES DO NOT TOTAL 100% BECAUSE RESPONDENTS COULD CHECK ALL THAT APPLY)

Most respondents reported that, at the architectural stage, high level and low level designs are created and the database solution documented (Figure 15). However, 41% of respondents reported that no set process is employed for requirements gathering.

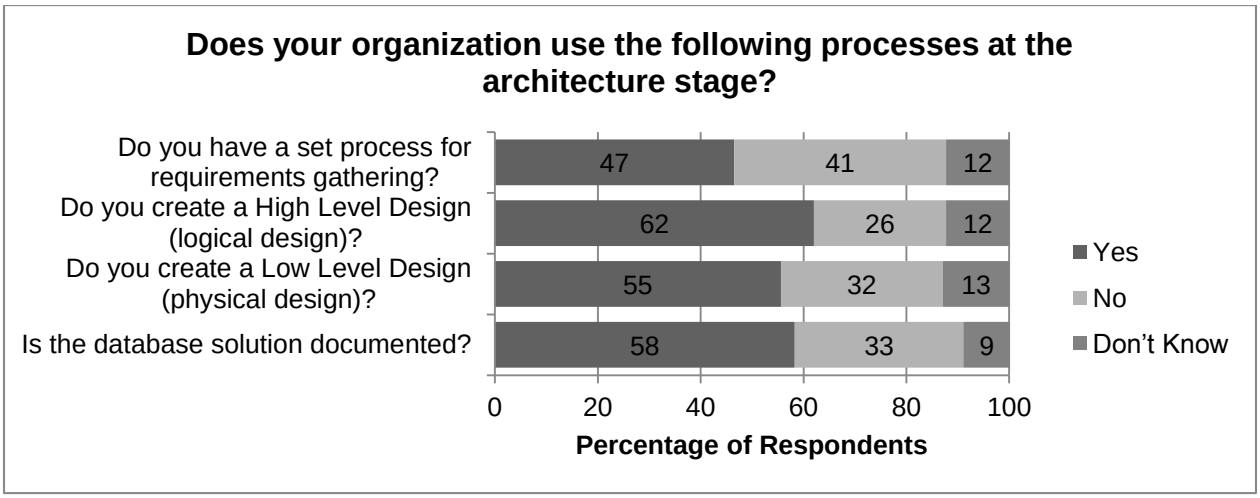

FIGURE 15 PROCESSES AT THE ARCHITECTURAL STAGE

Figure 16 shows processes followed at the design stage. For relational database design and data warehousing the majority of respondents reported that some form of process was followed. For all other database engines there was an overwhelming lack of use of design processes. Design process usage was slightly higher for data structure and hardware manageability.

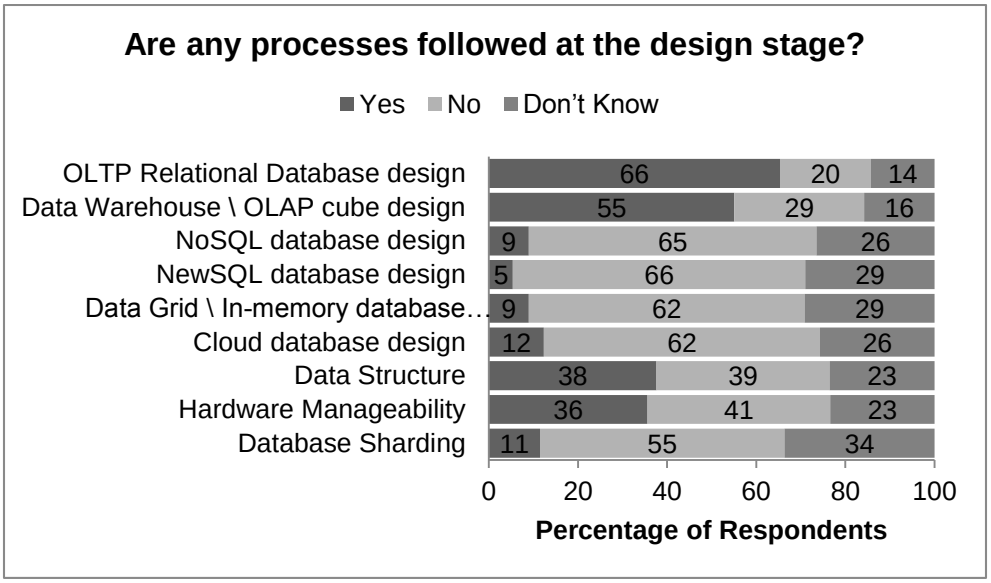

FIGURE 16 DESIGN PROCESSES

The development methodologies followed by respondents are shown in Figure 17. 64% reported use of Agile development methods and 35% reported use of Waterfall development methods.

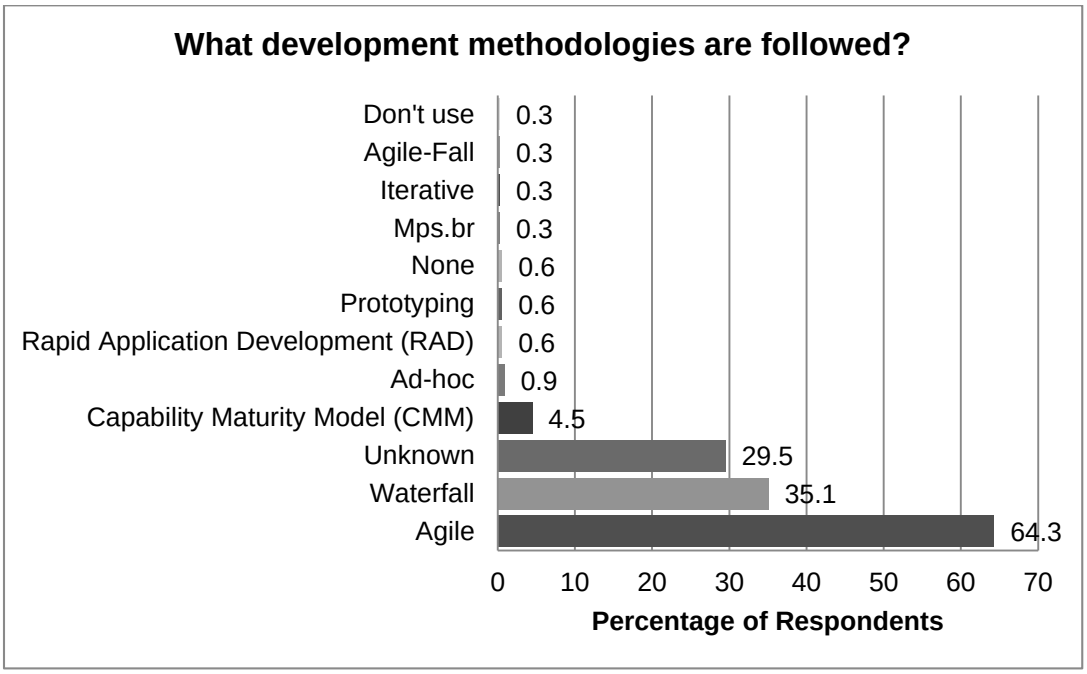

FIGURE 17 DEVELOPMENT METHODOLOGIES (NOTE: PERCENTAGES DO NOT TOTAL 100% BECAUSE RESPONDENTS COULD CHECK ALL THAT APPLY)

Figure 18 shows the extent to which processes are used at the development stage. 50% of respondents did not have a standard database testing process and 48% of respondents did not follow a defined database development lifecycle. In comparison 70% used a source control system and 63% had standard database coding practices.

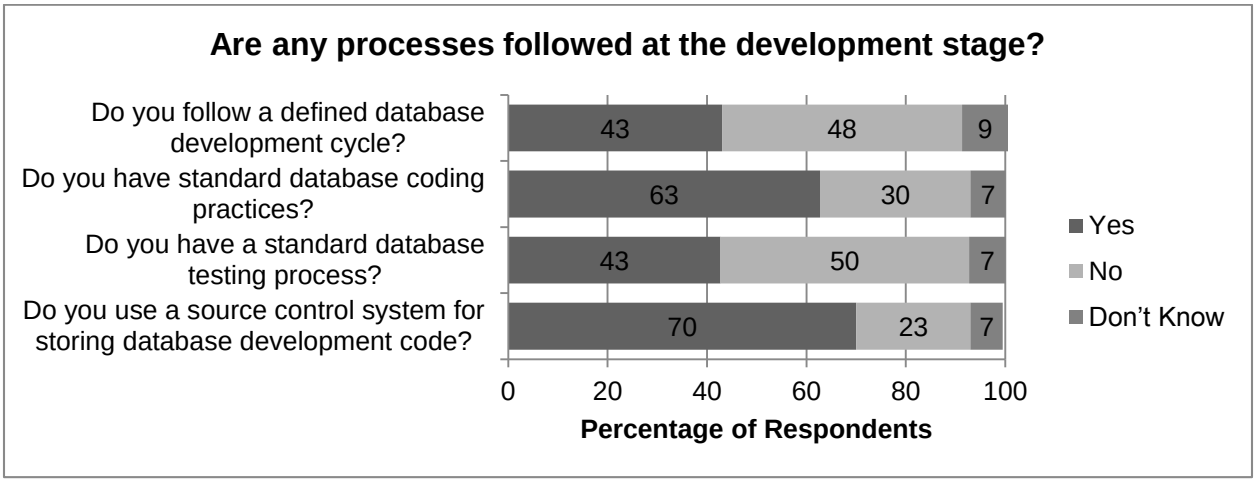

FIGURE 18 DEVELOPMENT STAGE PROCESSES

Database Technical Practices

Practices and procedures for database management were undertaken for day to day database maintenance, in particular: processes for monitoring regular maintenance, automated procedures to issue alerts, and processes for managing database server performance (Figure 19). The only area where practices were not reported as optimal was the recording of a performance baseline, undertaken by only 42% of respondents. 45% of respondents did not have recovery time objectives (RTO – the time it takes to restore the data), and 43% respondents did not have recovery point objectives (RPO – the amount of data loss allowable).

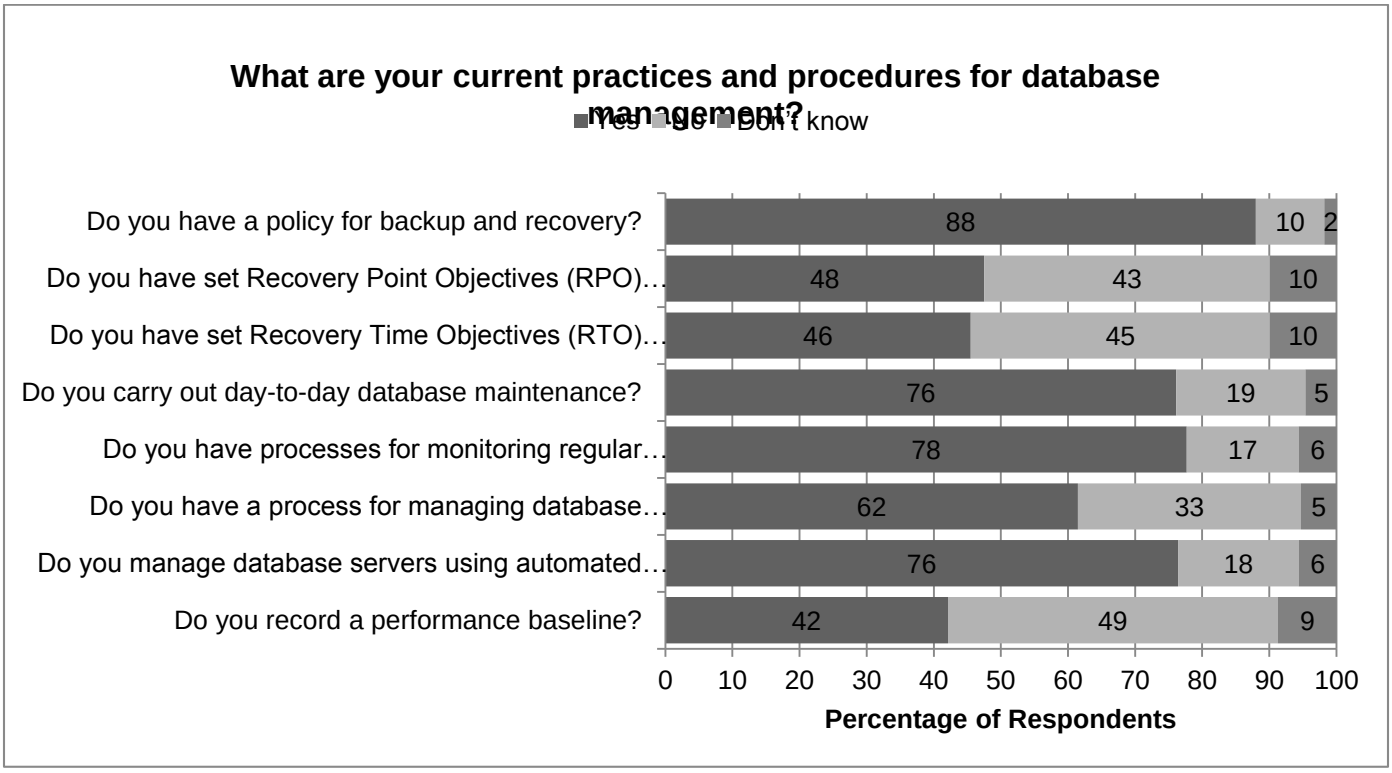

FIGURE 19 PRACTICES AND PROCEDURES FOR DATABASE MANAGEMENT

In terms of the use of IT service management frameworks: 42% of respondents used the IT Infrastructure Library (ITIL) framework, while 35% of respondents used no framework (Figure 20).

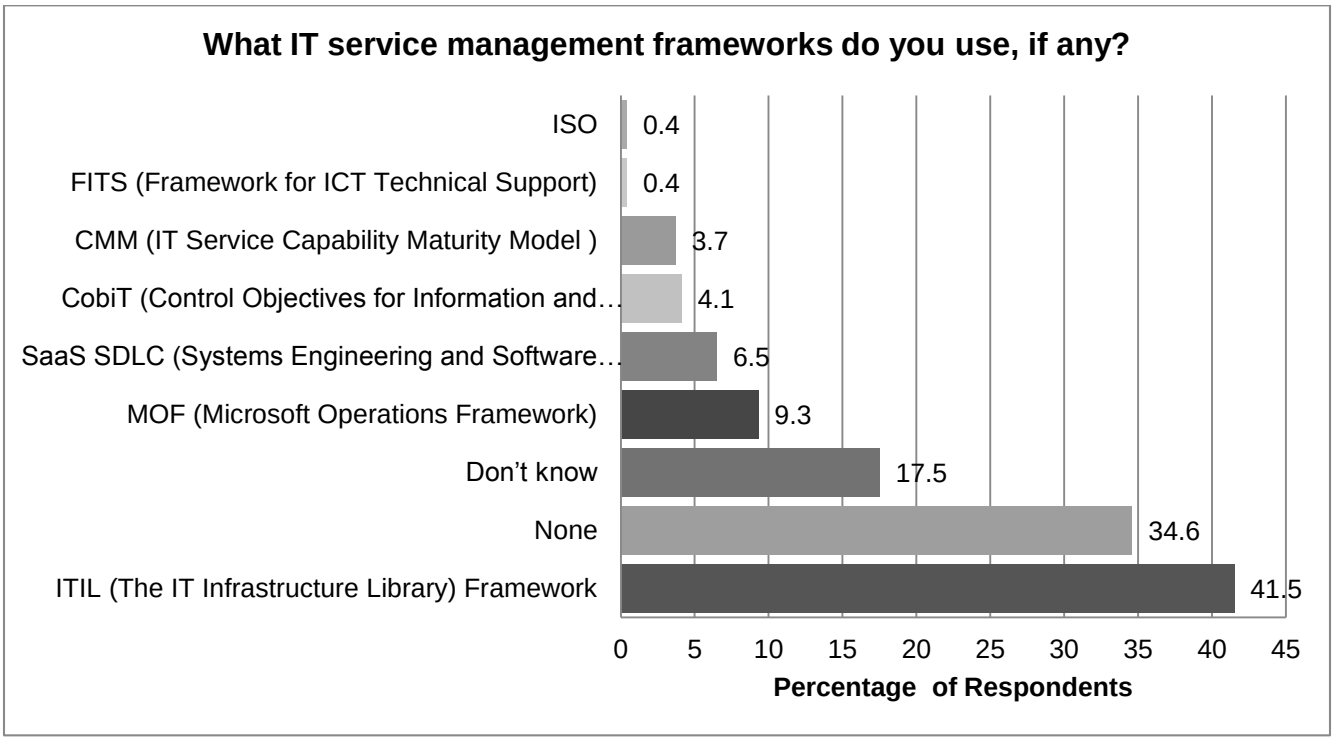

FIGURE 20 IT SERVICE MANAGEMENT FRAMEWORK

Problem management methods are not widespread: almost half of respondents did not use one, while a further quarter did not know whether they did (Figure 21).

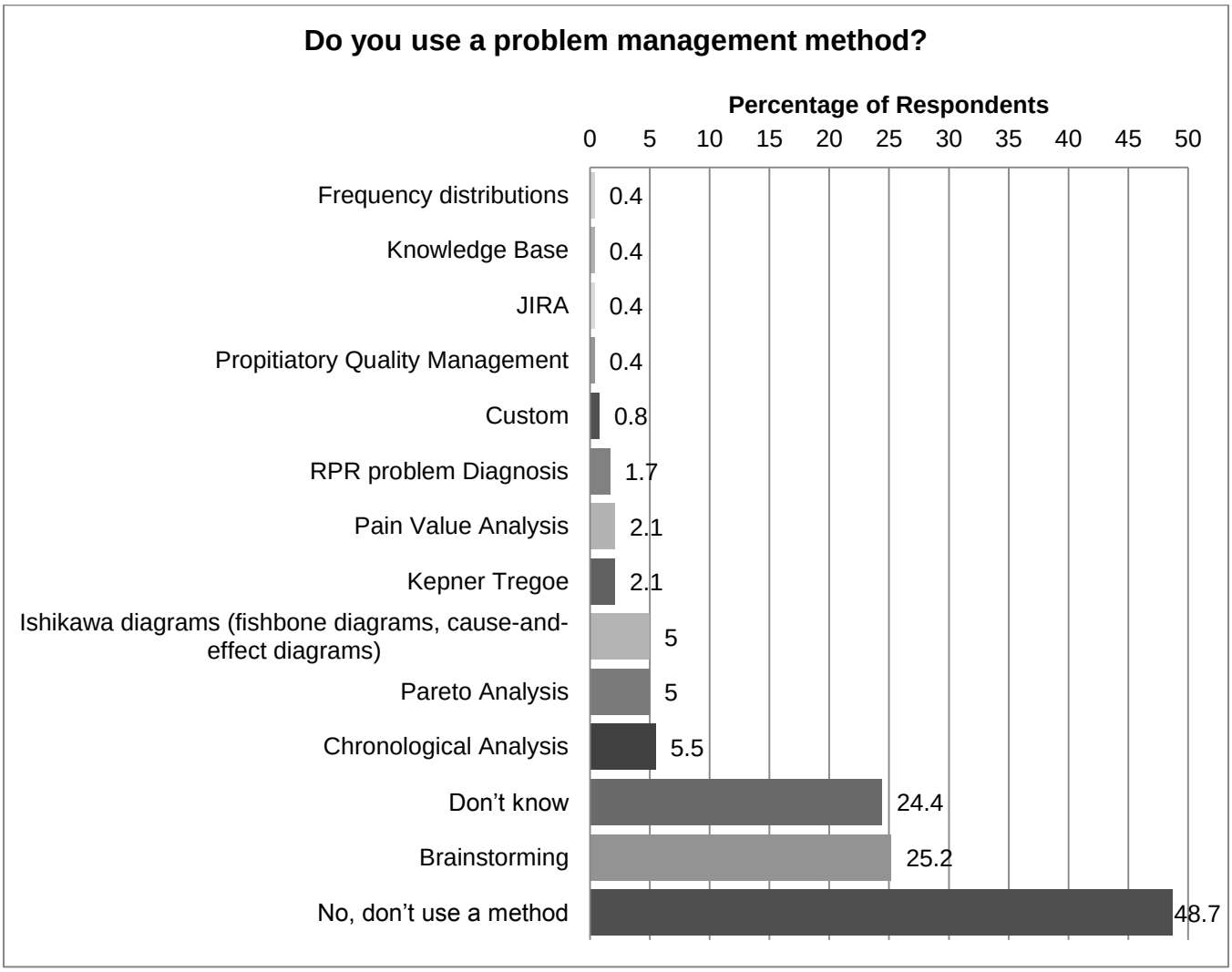

FIGURE 21 PROBLEM MANAGEMENT METHOD

In reporting responses to malfunctions, almost half the respondents reported that malfunctions were always or often dealt with re-actively (Figure 22).

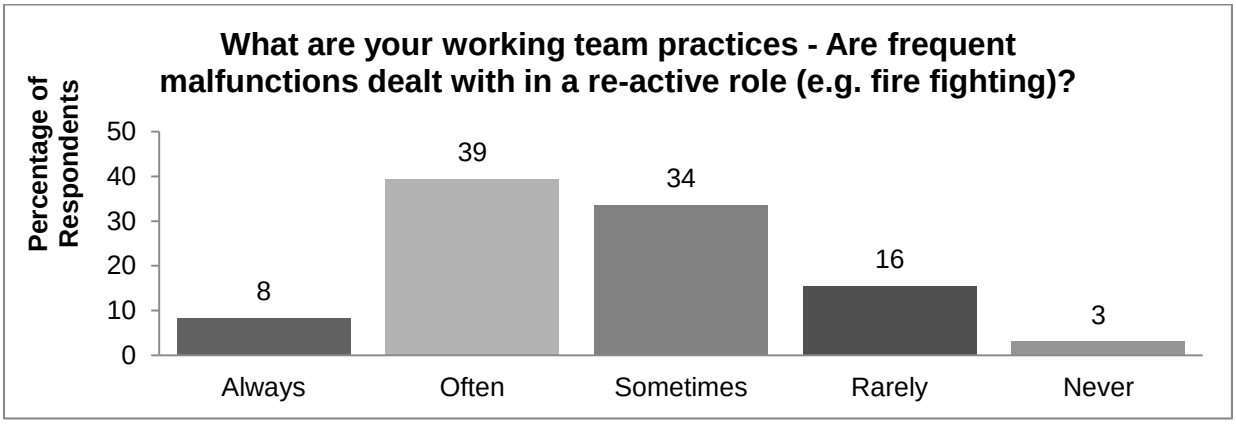

FIGURE 22 FREQUENT MALFUNCTIONS

However, two-thirds of respondents (Figure 23) always or often put long-term fixes in place for regularly occurring issues, to future proof the database application.

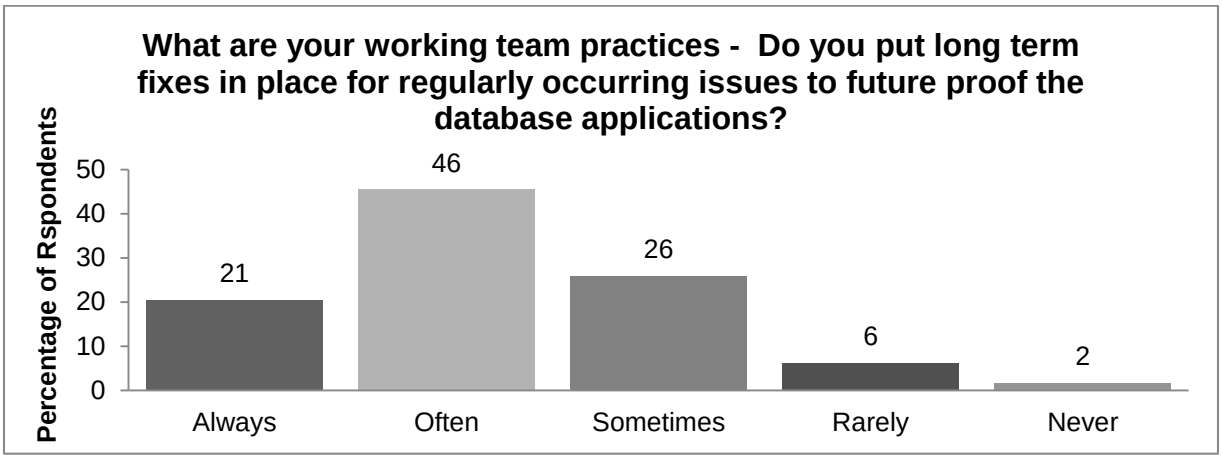

FIGURE 23 LONG TERM FIXES FOR REGULAR ISSUES

Database storage configuration practices are an important part of the management of database systems. Three main sources for configuration practices were reported (Figure 24): storage array manufacturers, manufacturers of databases and database administrators.

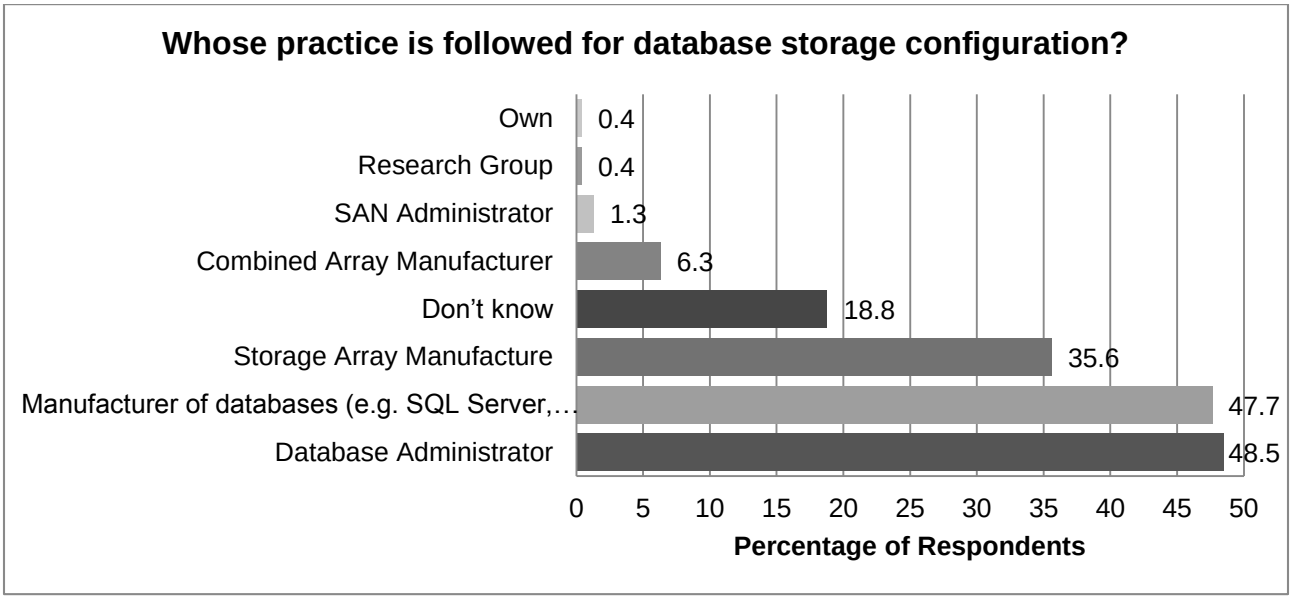

FIGURE 24 PRACTICE FOLLOWED FOR DATABASE STORAGE CONFIGURATION

Very few practices and procedures were used by the respondents to manage any area of cloud databases (Figure 25).

One of the survey questions asked if cloud database services were used, and what was the reason behind this. Very few respondents reported using cloud database

services (see Figure 8). The reasons given included: lack of trust of cloud vendors; not meeting security policies; lack of methods; and lack of appropriateness for their specific uses.

FIGURE 25 PRACTICES AND PROCEDURES TO MANAGE CLOUD

Several respondents regarded cloud services as requiring less in-depth database management. One respondent commented: "From the point of view of the client, the problem of database management, maintaining in-house skills etc. just goes away".

Data Management

Respondents were asked about their practices and procedures for data management. Policies were in place in the majority of respondents' organizations for keeping data for legal reasons, historical data storage and long term preservation (Figure 26). For the majority of respondents (78%), crowdsourcing was not used for predictive analysis, 69% of respondents had no master data management policy, 68% of respondents did not have processes in place for predictive analysis and 45% of respondents did not have data governance policies.

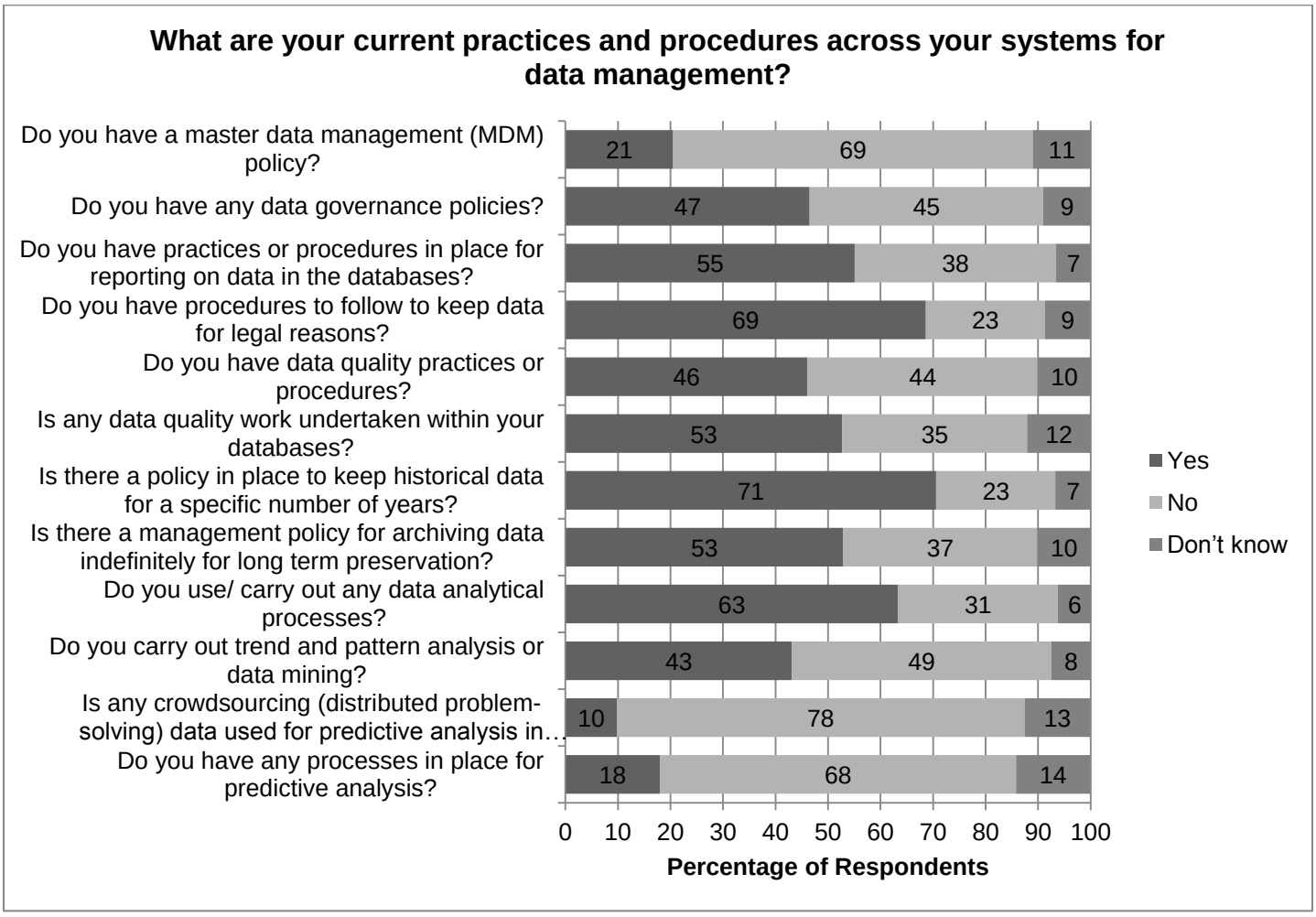

FIGURE 26 CURRENT PRACTICES AND PROCEDURES FOR DATA MANAGEMENT

Moving to specific practices for data management (Figure 27), 52% of respondents did not follow data lifecycle management policies. For each of two standard frameworks (DAMA-BOK from the data management association and the open source MIKE2.0 standard), 78% of respondents reported that they did not follow that framework. However 48% of respondents stated they had their own data management practices and procedures.

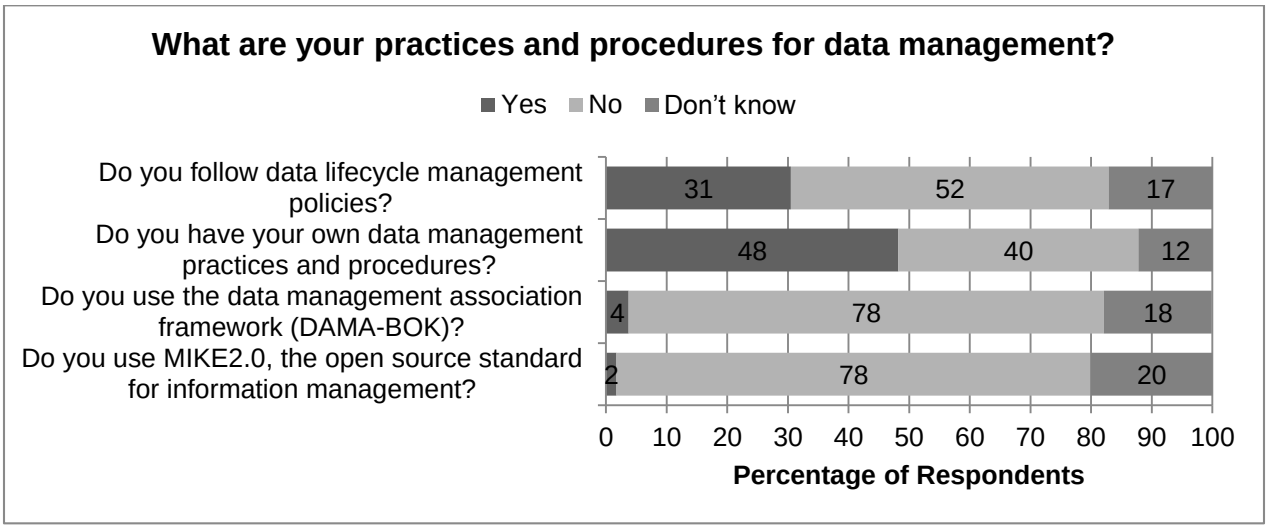

FIGURE 27 PRACTICE AND PROCEDURES FOR DATA MANAGEMENT FRAMEWORKS

Data transfer policies have not been developed in all cases: 43% of respondents had policies in place to transfer data on site, 33% had policies for offsite data transfer, and 37% had no policies (Figure 28).

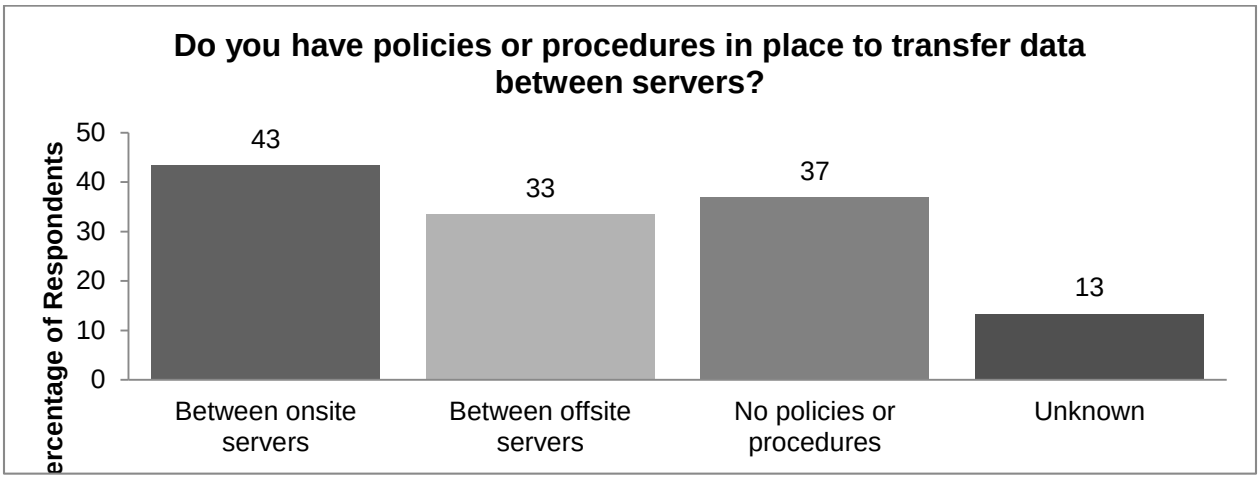

FIGURE 28 TRANSFER DATA BETWEEN SERVERS

Application Centric

Database applications provide a plethora of features to meet the demands of business. Tools are required to manage these features and to control and manage fundamental parts of the database systems.

59% of respondents stated that the type of database management that can be carried out was governed by the database software features (Figure 29). 72% said database

application scalability was a requirement, although 52% of the respondents did not have procedures to manage scalability.

62% did not have procedures to select different database engines for the task. 52% did not have procedures for reviewing new database engine changes, while 51% had procedures in place for managing virtualized databases.

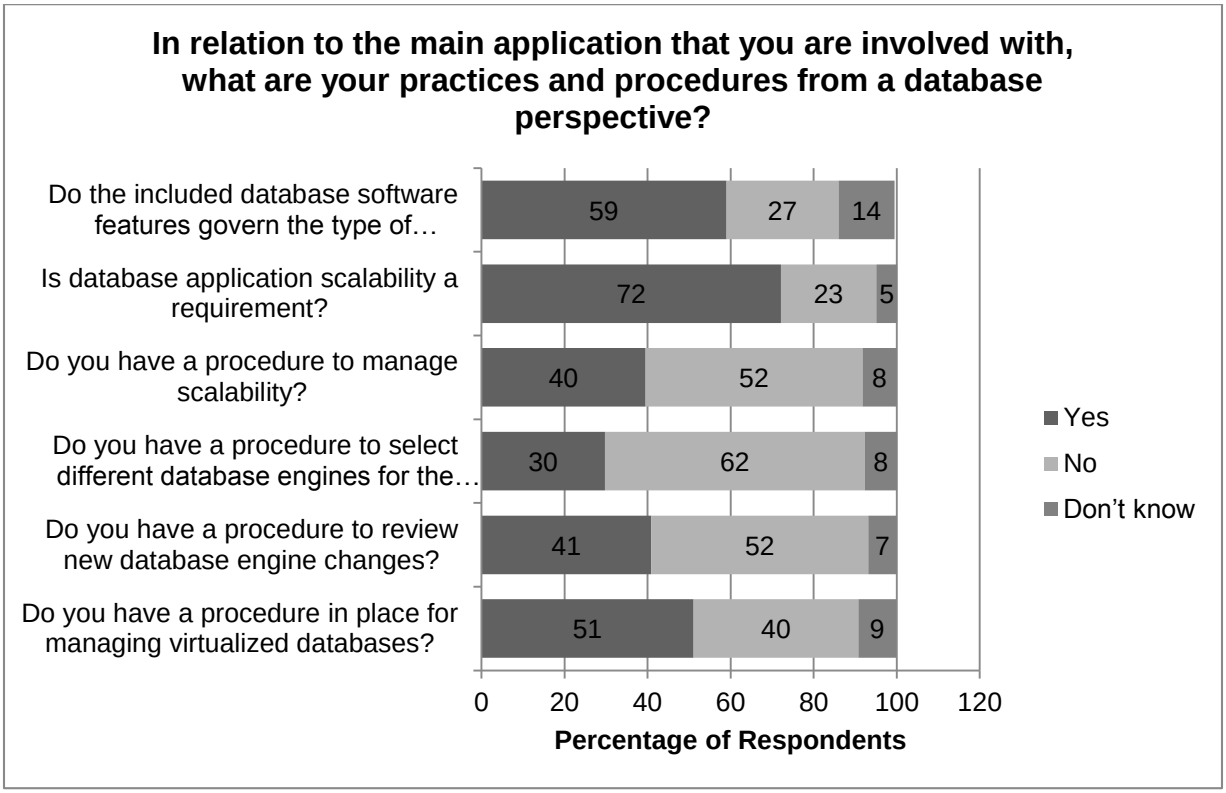

FIGURE 29 PRACTICES AND PROCEDURES FOR THE MAIN APPLICATION

45% of respondents (Figure 30) had different management practices for different database products used.

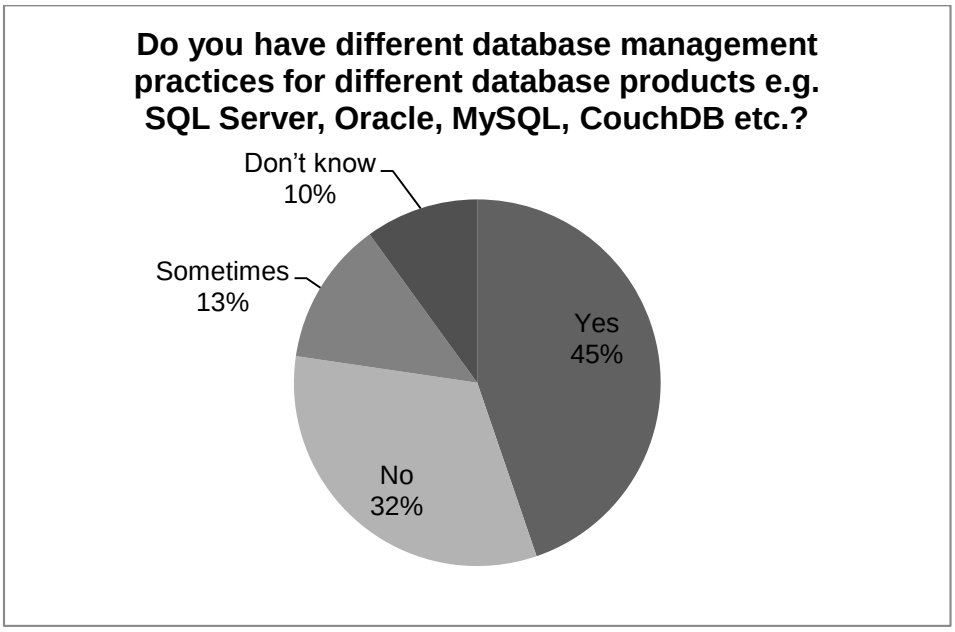

FIGURE 30 DIFFERENT MANAGEMENT PRACTICES FOR DIFFERENT DATABASE PRODUCTS

As shown in Figure 7, respondents use a large selection of database products, so it is important to know whether they apply different database management practices to different products. 63% of the respondents were not managing more unstructured data than last year, compared to 22% who *were* managing more unstructured data (Figure 31).

There were very few database administration practices and procedures for managing 'Big Data' (Figure 31). From Figure 6 only 1.3% of the respondents had a database 'over 100TB' in size. 79% of respondents did not manage Big Data (Figure 31). 81% did not have any procedures for the management of Big Data. 63% did not have different management practices for different sizes of database (Figure 31).

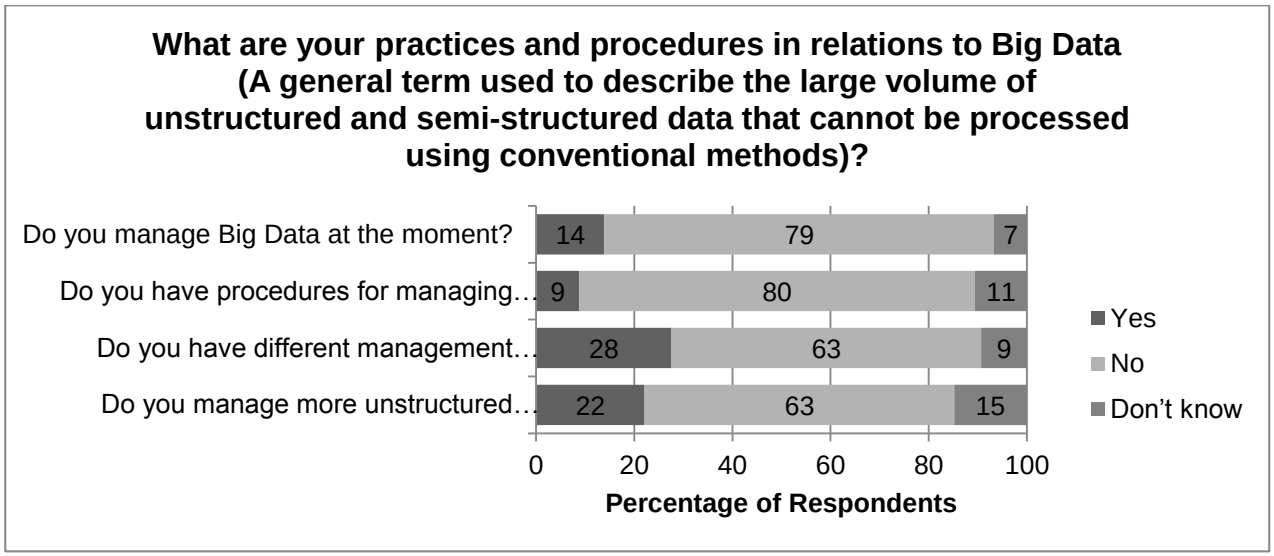

FIGURE 31 PRACTICES AND PROCEDURES FOR BIG DATA

Change Management

Most respondents reported having practices and procedures to manage changes for database servers. 57% reported that database changes required sign off by business users. 88% reported that changes to the database server could not be carried out by just anyone. 48% of respondents reported that change procedures were enforced for all database engines, while 46% did not enforce such procedures (Figure 32).

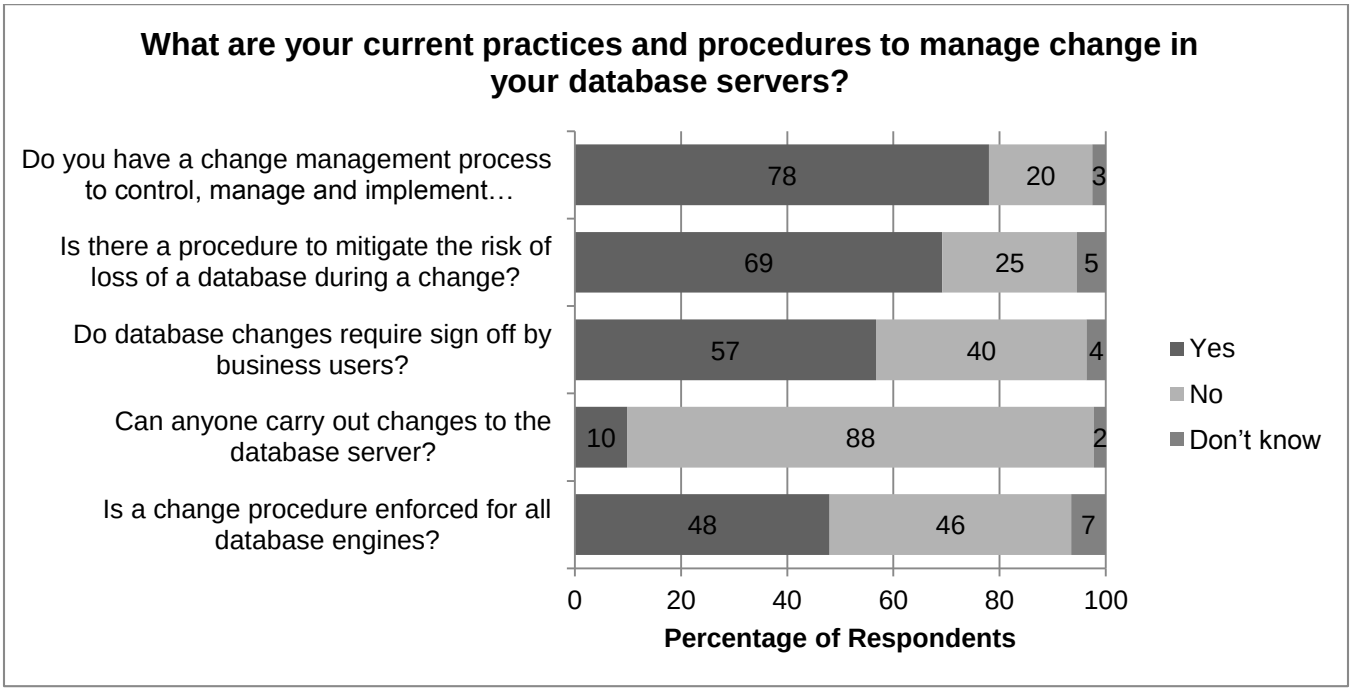

FIGURE 32 PRACTICES AND PROCEDURES FOR CHANGE MANAGEMENT

Changes were often made to the database environments in respondents' organizations (Figure 33). 27% of respondents said changes were carried out less often than weekly, while 6% made more than 50 changes per week.

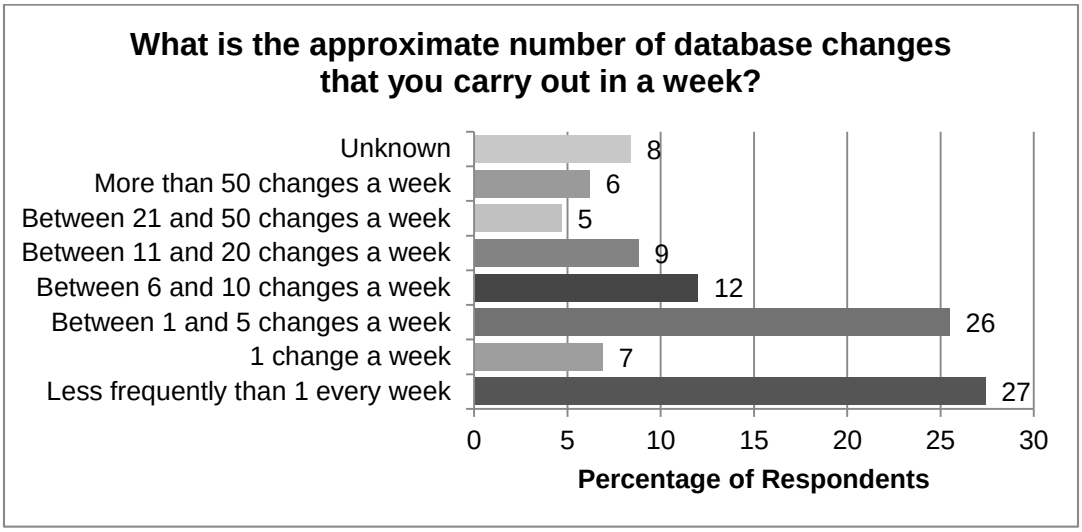

FIGURE 33 APPROXIMATE DATABASE CHANGES A WEEK

Formal change processes were not always used. 40% of respondents reported that sometimes changes occurred without following policies and procedures, while 5% reported that this happened very often (Figure 34).

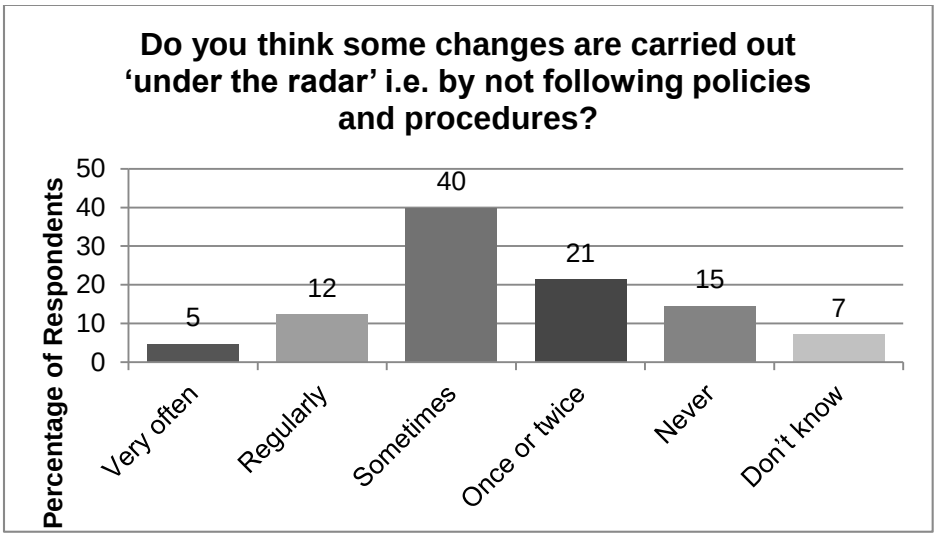

FIGURE 34 CHANGES NOT FOLLOWING POLICIES AND PROCEDURES

Organizational Culture

The responses regarding working team practices give insight into the working conditions, communication, control, strategy and budget within the organizations (Figure 35).

Communication between management and database team members, as well as cross-team communication, was often or always good. Within-team communication was seen as always good for 41% of respondents, whereas only 11% of respondents stated that cross-boundary communication was always good.

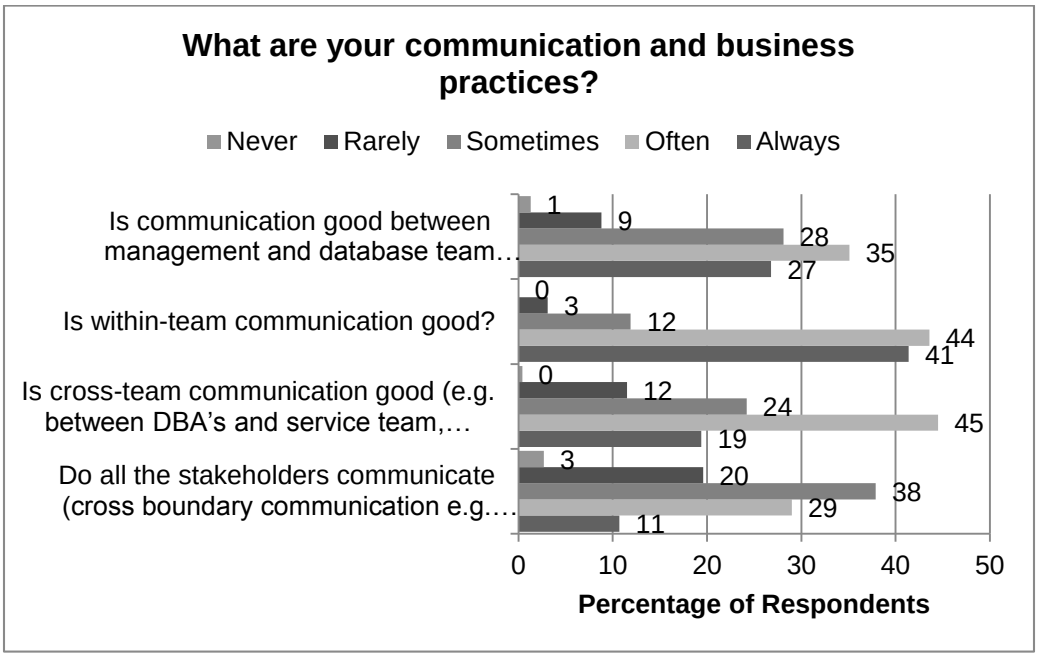

FIGURE 35 COMMUNICATION AND BUSINESS PRACTICES

5% of respondents stated that database management decisions were based solely on customer requirements (Figure 36). Nearly 50% reported that customer requirements often changed in projects and only 10% of respondents said customer requirements were always clearly identified at the outset.

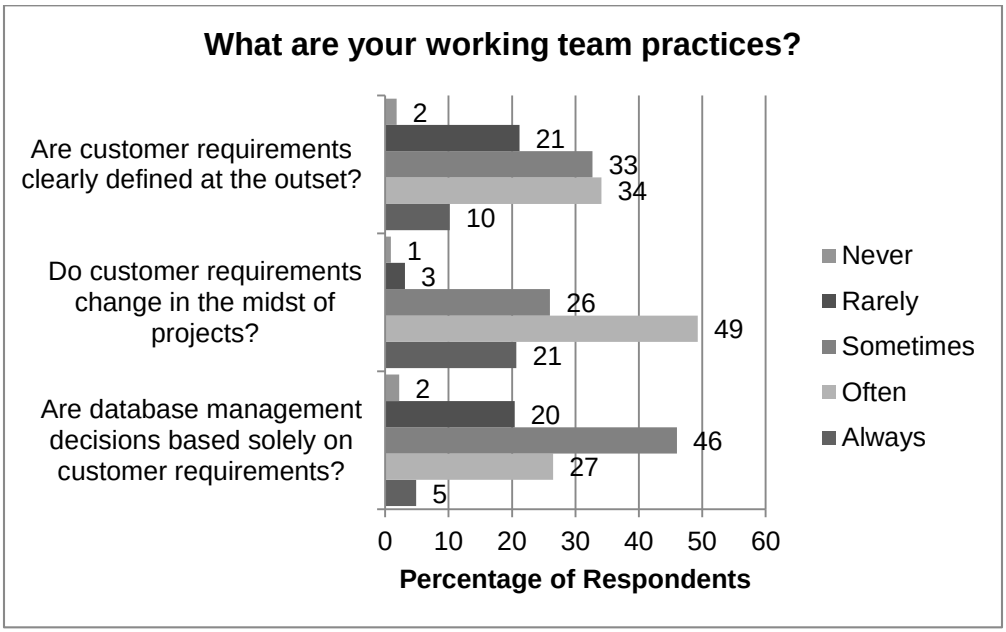

FIGURE 36 WORKING TEAM PRACTICES

It was not clear from the data whether or not the database software product selection is constrained due to the employee in-house skill set: for 48% of respondents it was a constraint, while an equal number said it was not (Figure 37). 55% of respondents stated that the budget did not determine what database *platform* was used, although 57% of respondents stated that financial reasons influenced the *version* of the database software.

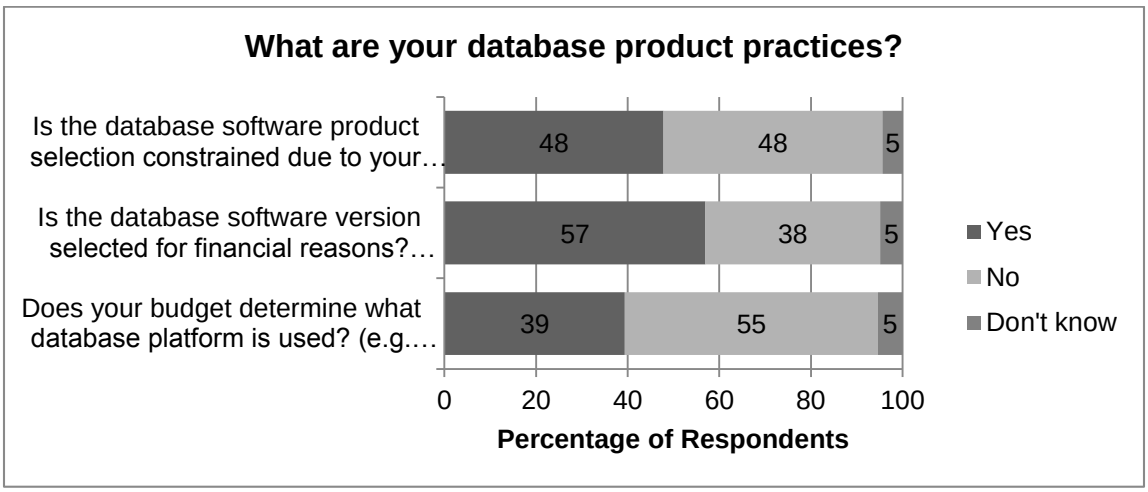

FIGURE 37 DATABASE PRODUCT PRACTICES

DISCUSSION OF FINDINGS

The questions in the survey captured the real world situation of the respondents, covering all aspects of the database lifecycle. The survey explored: architecture, development, operational management, security, cloud, cross engines, database management, data management and organizational culture, including training for database employees. In total, 493 database professionals from across the world responded to the survey, and respondents' represented a wide variety of database technologies, organizational settings and management practices (see Tables 1-5 and Figures 1, 6-8). Overall the majority of respondents were shown to have over 5 years experience in the database field. This section highlights key findings from the survey.

Best Practice

It is apparent that 'best practice' has many different meanings and best practices are always changing. Respondents took best practice guidelines from a diverse range of sources, and many respondents' organizations had created their own 'best practices' (Figures 9, 10 and 11). The way in which organizations maintain control of best practices varies considerably and relates to the nature of the database architecture (Figure 13). It was striking that 94% of respondents thought it was important to have best practices, despite the drawbacks (Figure 12).

Architecture

Database physical architecture design can significantly affect the performance and availability of the database servers. Customer requirements and the design architecture are key areas in which changes will inevitably cause service downtime, once systems have moved into a production environment. Data structure within the databases, if not optimal, can affect the quality and performance of the data. (Agrawal et al., 2009) state the importance of this, referring to "architectural shifts in computing" – in their view, fundamental software changes are being triggered by advances in hardware, data management variety and cloud computing.

The present survey findings showed the following points:

- The use of mandated processes for requirements gathering, at the architectural stage, was low (Figure 15). Customer requirements were rarely defined at the outset and customer requirements could be changed in the middle of projects (Figure 36). Decisions about database management were not always based on customer requirements.

- There was a lack of use of core architecture frameworks for database design, although documented design patterns were used (Figure 14). Some of this lack of use occurred when an application came from a software vendor, including a wholly or partially complete database design, where only on-going support was required.
- Processes at the design stage were rarely used for NoSQL, NewSQL, Cloud, and In Memory databases and for database sharding (horizontally scaling of a database). About half of all respondents did not follow data structure and hardware manageability processes at the design stage (Figure 16).
- Although database scalability was a requirement for organizations there were few procedures to manage this (Figure 29). Elastically scalable database systems have increased due to global business (Abadi, 2012) making procedures to manage scalability increasingly important.

Development

Agile database management techniques (using shorter sprints with a set time limit for repeatable work patterns) can be used to help improve the effectiveness of database management. Agile can be used for database development – (VersionONE, 2012) found that 37% of respondents were using Agile for "76-100% of projects". However, the present survey showed that Agile is not without its own problems, such as "lack of upfront planning" (34% of respondents) and "a loss of management control" (31% of respondents). This survey also found that:

- The usage of agile development methodologies was high; indicating that an interactive and adaptive approach is seen as effective (Figure 17).
- Half of the respondents had no standard testing processes (Figure 18).
- About half of the respondents' organizations had no defined database development lifecycle (Figure 18).

Operational Management

The operational state of databases is probably the most well managed area of the database system due to the potential impact to the business if databases become unavailable. Challenges reported in an Independent Oracle Users Group survey (Mckendrick, 2011b) for operations included: an increase in the number of databases, databases of larger size, a reduction in the number of older systems being retired as they are kept in operation for longer, as well as more features and functionality being

included in newer systems. Stonebraker et al. (2013) also discussed the operational challenges of new features which add to the complexity. Specific issues from the present survey include:

- The availability of data and database servers was a very important factor. However, recovery time objectives (RTO) or recovery point objectives (RPO) are not defined in many cases (Figure 19), so when the database becomes corrupt or unavailable, recovery defaults to an individual's best endeavors. This could result in lengthy outages.
- About a third of respondents did not use an IT Service Management framework (Figure 20), which can help ensure best practices are followed.
- Approximately half the respondents did not use problem management methods (Figure 21). Frequent malfunctions were often dealt with in a reactive way (Figure 22). Dealing with malfunctions in a reactive way could indicate that improvement within the system was not taking place.
- The state of a database continually changes as changes in the real world are reflected within it, through database structure, data, design or architecture. It is important to understand the rate of change carried out on databases, and this may suggest a greater need for processes. Although practices and procedures were set out for changes and were regularly adopted (Figure 32), respondents reported that they did not always follow such practices and procedures (Figure 34).

Cloud

Cloud database services are becoming a popular choice, as they offer a cheaper service, have automated high availability and are an easier and quicker route to market than in-house databases. They bring many advantages and disadvantages. When deploying databases in the cloud, computation, storage and availability are no longer the concern of the business and the corresponding technical skills are not required. Issues associated with cloud databases include the potential trust and security of data, loss of critical data, multiple customer databases sharing the same infrastructure with different access patterns and longevity of the companies supplying these services. The survey showed that:

- Cloud databases have only become available for usage in recent years, and adoption is currently fairly low (Figure 8).

- Practices and procedures to manage these cloud databases were in use by only a small percentage of respondents (Figure 25).
- In-house database management skills are not usually required by users of cloud databases. The introduction of cloud databases shifts part of the database administration practices and procedures from the organization to external suppliers. Thus it is possible that a different set of best practices are required to manage cloud databases.

Cross engines

Many database engines exist (Maslett, 2012) which offer different features and in many organizations multiple database engines are used (Figure 7). The newer NoSQL engines are increasingly used in addition to traditional engines (Atzeni et al., 2013). There are various architectural models for transactional, analytical or scale-out architectures. To achieve service availability, it is important to select the correct system for the organizational requirements; and engines from different suppliers may need to interact with each other. Key points from the survey included:

- In most organizations, there were no procedures to select different database engines (Figure 29). Without a method for choosing database engines / applications, the wrong type of system could be selected, causing problems with database management.
- There were different management practices for different database products (Figure 30) which could mean added complexity when managing multiple systems. An example of this could be the use of different management tools. An Independent Oracle Users Group survey (Mckendrick, 2011b) reported that 77% of respondents use different tools for each database management (DBMS) platform.

Database Management and Data Management

The management of databases requires diverse knowledge and skills, which are continually changing, and there are numerous core technical practices. It is important to understand current practices and procedures in data management, in relation to database management, due to the close connection between these disciplines. Aiken et al. (2011) highlighted trends which link together components of data management. There were several important findings from the present survey in this area:

- A few respondents followed data lifecycle management policies (Figure 27), although just under half the respondents' organization had their own data management practices and procedures. Management of data was regarded by Aiken et al. (2007, p. 49) as a maintenance cost rather than an asset.
- Little time was spent solely on managing database servers (Table 4), which could indicate there are many other tasks required in addition to the management of the database server. Due to the time split between tasks this could affect the number of practices and procedures created or adopted.
- Database management decisions were mostly based on customer requirements (Figure 36), which indicates that decisions on the key factors in database management follow the customers' requests and not what the product manufacturer or industry necessarily recommend. This could potentially affect which best practices are adopted. In a survey by (Mckendrick, 2013), 47% of respondents reported that the leading database administrative challenges for Oracle databases was the diagnosis of performance problems.
- There was a lack of choice of management processes to match the different sizes of data (Figure 31).
- There are a large number of database software providers available (Maslett, 2012). Software features were fairly important in product choice (Figure 29) and this could govern the type of database management available, through the tools which come as part of the database products.
- Documentation usually contains the architecture design (Figure 15), the development design, the configuration and specific practice and procedures. Most respondents considered documentation important to improve practices and procedures.
- Data governance and master data management were frequently lacking, and data quality procedures were only partially used (Figure 26). Data quality procedures are key to ensuring that the database provides good quality information.
- The volume of unstructured data is rapidly growing (Gantz & Reinsel, 2010). A small group of respondents reported growth in the management of unstructured data over the previous 12 months (Figure 35); however the majority of respondents did not currently manage such data.
- Best practices and procedures appear to be widely adopted for data security, with the exception of procedures to transfer data between servers (Figure 28).

There have been a number of major security issues where data has been lost in transit between different geographic locations and within offices.

Culture

The culture in an organization affects how certain tasks are carried out. The management of the organization dictates resourcing levels, which tools and internal systems are available for use and whether time is available for proactive work. With the evolution of database technology and the surrounding hardware technologies, more teams of people are involved in bringing about the successful operation of the entire database platform. The survey showed the following issues relating to organizational culture:

- Certification was rarely encouraged by respondents' organizations (Figure 3), and neither was formal training forthcoming (Figure 4). Yet the involvement by respondents in database community associations was high (Figure 5). Attendance at community events provides free or inexpensive training and an environment where problems can be discussed.
- Improved communication at all levels was required (Figure 35). Many respondents stated that communication deteriorated when communicating across multiple teams and that the worst communication was cross boundary communication with the stakeholders. Poor communication could lead to incidents or problems occurring.
- It was striking that up to date training was not always considered vital in respondents' organizations: 23% of respondents were never given the opportunity to undertake formal training (Figure 4), 24% were never given the opportunity to attend database conferences workshops or seminars (Figure 5), and 28% were never encouraged to obtain certification (Figure 3). 12% reported that no database training was provided at all. A low in-house skillset could result in problems occurring in database selection and operation.

CONCLUSION

The analysis of the findings from the survey suggests there are many and varied practices and procedures throughout the database lifecycle. The complexities discussed in the Claremont Report (Agrawal et al., 2009) have entered the everyday working environment for the management of database systems. There are many

interconnected components and stakeholders in database management, and best practices and procedures may be affected by these. Best practices are continually changing and many organizations have their own custom best practices.

The survey reported in this paper shows that there are a variety of adoption levels for best practices and procedures. The management of the servers was only a part of managing database systems and the increased growth of data was not being managed by the respondents. Documented design patterns were used rather than established frameworks. Operational work was often reactive but the only framework significantly adopted by the respondents was ITIL for service management. Documentation was important to the respondents. Cloud practices and procedures were not well established and it is unclear what skills will be required in future. The method for selecting database engines was often unclear, although software features were an important factor in the choice. Financial budgets had some effect on the version of database platforms selected.

A lack of control of best practice could be a contributing factor as to why best practices are not always followed. Sometimes enforcement is required to ensure conformity. Cross boundary communications with stakeholders require improvement. Formal training on existing systems, and keeping up to date with new systems, were acknowledged as important, but were sometimes lacking. The results of the survey highlight that there are a vast array of technical skills and technical knowledge required for the management of database systems. There are various areas within the database lifecycle where the survey suggests that there are current gaps in practices and procedures for database management.

The next stage of this research is to examine in more depth the complexities and interactions between the key components of the database system. The research will look at recommended practices and the reasons behind usage, with an indication of a suggested route towards general improvement in the operation and management of database systems.

REFERENCES

Abadi, D. J. (2012). Consistency Tradeoffs in Modern Distributed Database System Design. *IEEE Computer Society*, *2*(February), 37–42.

Abiteboul, S., Agrawal, R., Bernstein, P., Carey, M. J., Ceri, S., Croft, B., … Naughton, J. (2005). The Lowell Database Research Self-Assessment. *Communications of the ACM*, *48*(5), 111–118. doi:10.1145/1060710.1060718

Agrawal, R., Ailamaki, A., Bernstein, P., Brewer, E. ic A., Carey, M. J., Chaudhuri, S., … Weikum, G. (2009). The Claremont Report on Database Research. *Communications of the ACM*, *52*(6), 56–65. doi:10.1145/1516046.1516062

Aiken, P., Allen, M. D., Parker, B., & Mattia, A. (2007). Measuring Data Management Practice Maturity : A Community's Self-Assessment. *IEEE Computer Society*, *40*(4), 42–50. doi:10.1109/MC.2007.139

Aiken, P., Gillenson, M. L., Zhang, X., & Rafner, D. (2011). Data Management and Data Administration: Assessing 25 Years of Practice. *Journal of Database Management*, *22*(3), 24–45. doi:10.4018/jdm.2011070102

Atzeni, P., Jensen, C. S., Orsi, G., Ram, S., Tanca, L., & Torlone, R. (2013). The relational model is dead, SQL is dead, and I don't feel so good myself. *ACM SIGMOD Record*, *42*(2), 64–68. doi:10.1145/2503792.2503808

Ballew, R. M., Sabelko, J., Gruebele, M., Natl, P., Oas, T. G., Biol, N. S., … Auffray, J. C. (1998). A Physical Map of 30,000 Human Genes. *Science*, *282*(5389), 744–746. Retrieved from http://74.220.219.81/~mrflipco/teach/writing/GeomProj/images/PhysicalMapOf30000HumanGenes.pdf

Blasis, J. De. (1977). Database Administration as a Team Function: An Analysis from Survey Data. In *Proceeding SIGCPR '77 Proceedings of the fifteenth annual SIGCPR conference ACM* (pp. 227–240). doi:10.1145/800100.803248

Bradley, N. (1999). Sampling for Internet Surveys. An examination of respondent selection for Internet research. *Journal of the Market Research Society*, *41*(4), 387–395. Retrieved from http://www.wmin.ac.uk/marketingresearch/Marketing/sam06.html

Buckingham, A., & Saunders, P. (2009). *The Survey Methods Workbook* (Reprint.). Cambridge: Polity Press Ltd.

Cuddeford-Jones, M. (2013). Predicting the Future. *Marketing Week (01419285)*, 31–33.

Data, Data everywhere. (2010). *The Economist*, *394*(8671), 3–5.

Dembowski, F. L. (2013). The Roles of Benchmarking , Best Practices & Innovation in Organizational Effectiveness. *International Journal of Organizational Innovation*, *5*(3), 6–20.

Denscombe, M. (2008). *The good research guide: for small-scale social research projects* (3rd ed.). Maidenhead: Open University Press.

Gantz, J., & Reinsel, D. (2010). *The Digital Universe Decade – Are You Ready?* Retrieved from http://idcdocserv.com/925

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *IEEE Intelligent Systems*, *26*(3), 10–14. doi:10.1109/MIS.2011.52

Gillenson, M. L. (1982). The State of Practice of Data Administration-1981. *Communications of the ACM*, *25*(10), 699–706. doi:10.1145/358656.358664

Gillenson, M. L. (1985). Trends in Data Administration. *MIS Quarterly*, (December), 317–326. doi:10.2307/249232

Gillenson, M. L. (1991). Database Administration at the Crossroads: The Era of End-User-Oriented, Decentralized Data Processing. *Journal of Database Administration*, *2 (4)*, 1–11. doi:10.4018/jdm.1991100101

Guion, L. A. (2002). *Triangulation: Establishing the Validity of Qualitative Studies*. *Qualitative Research* (pp. 1–3). Retrieved from http://www.rayman-bacchus.net/uploads/documents/Triangulation.pdf

Haigh, T. (2012). Fifty Years of Databases. ACM Sigmod special interest group on Management of Data. Retrieved from http://wp.sigmod.org/?p=688

Hensley, Z., Sanyal, J., & New, J. (2014). Provenance in sensor data management. *Communications of the ACM*, *57*(2), 55–62. doi:10.1145/2556647.2556657

Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data Intensive Scientific Discovery*. (K. T. Tony Hey, Stewart Tansley, Ed.) (pp. 1–252). Redmond Washington: Microsoft Research. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf

Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Sceince Quarterly*, *24*(4), 602–611. Retrieved from www.jstor.org/stable/2392366

Kahn, B. K. (1983). Some Realities of Data Administration. *Communications of the ACM*, *26*(10), 794–799.

Maslett, M. (2012). Database Landscape Map. *451 Group*. Retrieved May 12, 2013, from http://blogs.the451group.com/information_management/2013/02/04/updated-database-lanscape-map-february-2013

McCririck, I.B., & Goldstein, R. C. (1980). What do data administrators really do? *Datamation*, *26*(8), 131–134.

Mckendrick, J. (2011a). *Managing the Rapid Rise in Data Growth: 2011 IOUG Survey on Database Manageability*. *Unisphere Research, a division of Information Today, Inc*. Retrieved from http://www.oracle.com/us/products/database/ioug-managing-db-growth-355001.pdf

Mckendrick, J. (2011b). *Data Cross-Curents : 2011 Survey on Cross-Platform Database Adminsitration*. *Unisphere Research, a division of Information Today, Inc*. Retrieved from http://www.dbta.com/DBTA-Downloads/ResearchReports

Mckendrick, J. (2013). *From Database Clouds to Big Data: 2013 IOUG Survey on Database Manageability*. *Unisphere Research, a division of Information Today, Inc* (pp. 1–28). Retrieved from http://www.oracle.com/webapps/dialogue/ns/dlgwelcome.jsp?p_ext=Y&p_dlg_id=13789408&src=7897724&Act=33

Mullins, C. S. (2012). *Database Administration: The Complete Guide to DBA Practices and Procedures* (Second.). Upper Saddle River, New Jersey: Addison-Wesley.

Oxford Dictionaries. (2014). *Oxford University Press*. Retrieved February 24, 2014, from http://www.oxforddictionaries.com/

Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*, *23*(5), 528–543. doi:10.1108/IntR-06-2013-0115

Segal, B., Robertson, L., Gagliardi, F., Carminati, F., & Cern, G. (2000). Grid computing: The European data grid project. *IEEE Nuclear Science Symposium and Medical Imaging Conference*, *1*, 2. Retrieved from http://www.researchgate.net/publication/3914275_Grid_computing_the_European_Data_Grid_Project/file/72e7e52864de518172.pdf

Stonebraker, M., & Robertson, J. (2013). Big Data is "Buzzword du Jour;" CS Academics "Have the Best Job." *Communications of the ACM*, *56*(9), 10–11. doi:10.1145/2500468.2500471

Thurmond, V. A. (2001). The Point of Triangulation. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau*, *33*(3), 253–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11552552

Verner, J., & Babar, M. a. (2004). Software quality and agile methods. *Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004.*, 520–525. doi:10.1109/CMPSAC.2004.1342889

VersionONE. (2012). *7th Annual State of Agile Development Survey* (pp. 1–14). Retrieved from http://www.versionone.com/state-of-agile-survey-results/?mkt_tok=3RkMMJWWfF9wsRouuq3JZKXonjHpfsXx6OUpUaeg38431UFwdcjKPmjr1YsAScR0dvycMRAVFZl5nRpdCPOcc45P9PA=