# The Data Management Survey 19

The voice of the data management community

# The Sample, KPIs and Methodology

This document provides background information to help gain a clearer understanding of The Data Management Survey 19



This document is not to be shared, distributed or reproduced in any way without prior permission from BARC

## Table of contents

Introduction
The sample4
Sample size and make-up4
Organization sizes by headcount5
Vertical markets5
Featured products
Peer groups
Overview of the key calculations in The Data Management Survey 19
Means and medians10
Understanding multiple response questions
Survey data collection
Understanding the KPIs
Reading the KPI charts
The KPIs (overview)
Developer efficiency
Time to market
Maintenance efficiency 17
Compliance
Performance17
Scalability
Data access
Price-to-value
Recommendation
Product satisfaction
Data governance
Skills availability
Product picklist used in The Data Management Survey 19
About BARC

## Introduction

The Data Management Survey 19 is the first edition of BARC's new annual study into the market for data warehousing tools. It employs the same proven methodology as The BI Survey and The Planning Survey, which have been conducted annually since 2000, setting out to analyze market trends and produce meaningful comparisons of competing products across a wide range of critical software and vendor-related criteria. The Data Management Survey also provides a detailed quantitative analysis of why customers buy data warehousing tools, what they are used for, the problems they experience with the tools and how successful they are.

The Data Management Survey 19 features 11 data warehousing products from 7 different vendors. It includes products from well-known global giants such as Microsoft, Oracle and SAP but also tools from much smaller vendors that ordinarily don't get much press but which, in many cases, offer outstanding value to customers.

After data cleansing and removing responses from participants unable to answer specific questions about their use of data warehousing products, we were left with 782 people who responded to the survey with 690 answering a series of detailed questions about their use of a named product. Participants from all over the world took part in The Data Management Survey 19.

The findings from The Data Management Survey 19 are presented in several documents, each focusing on a specific set of the survey results.

## The sample

Most surveys are conducted or sponsored by an organization based in, and focused on, one country. However, data management is a worldwide market and we wanted to capture a larger international sample.

The net result was an extraordinarily international panel. Respondents were located in 56 countries. The countries with the most respondents are Germany, the United States and Switzerland. The regions with the most respondents are Europe, North America and Asia Pacific.

The online questionnaire was published in two languages: English and German.

#### Sample size and make-up

Many thousands of people around the world were invited to participate in The Data Management Survey 19, using BARC's online research panel and the support of vendors and various websites. The questionnaire offered different sets of questions for vendors and users (or consultants answering on behalf of users).

The results of the online data collected are shown in the following chart, with the numbers of responses removed also displayed.

#### Table 1: Responses to the survey

	Resp	Responses	
Total responses	804	100%	
Filtered during data cleansing	-22	-3%	
Remaining after data cleansing (total answering questions)	782*	97%	
Non-user (did not answer questions about products)	-52	<1%	
Vendor (did not answer questions about using products)	-87	-11%	
Total answering product-related questions	690	86%	

\* Users and consultants had the option to review more than one product.

The number of responses is split between users, consultants, vendors and non-users. Vendors answered a different set of questions to those answered by end users. This document focuses on the analysis of the user results.





# BARC #DATA MANAGEMENT SURVEY 19

#### Organization sizes by headcount

Specialized data management software is most commonly found in medium and large organizations. A high percentage of the responses we received were from medium-sized companies between 101 and 2,500 employees (see Figure 2).



# Figure 2: How many employees are there in your entire organization, including all of its branches, divisions and subsidiaries? (n=620)

#### **Vertical markets**

We asked all respondents which industry sector their company operates in. The chart below shows the results of this question. Most respondents work in services, followed by manufacturing and IT.



#### Figure 3: Which of the following best describes your organization's industry sector? (n=782)

## **Featured products**

When grouping and describing the products featured in The Data Management Survey, we did not strictly follow the naming conventions the vendors use. Note that the names we use in this document are our own and are not always the official product names used by the vendors.

One of the key reasons for this is that the products we analyze are not necessarily the latest version of the tool. Vendors often change the product name between versions, making it difficult to have a single official name for several versions of the same product. The point is not to challenge the naming conventions of the vendor, but simply to reduce the complexity of the survey findings for the convenience of the reader. In some cases, we also shorten the names of the products to improve the formatting of the charts.

We asked respondents explicitly about their experiences with products from a predefined list, with the option to nominate other products. Our predefined list can be found at the end of this document. In cases where respondents said they were using an 'other' product, but from the context it was clear that they were actually using one of the listed products, we reclassified their data accordingly.

The following table shows the products included in the detailed analysis. In this, the first edition of The Data Management Survey, a minimum of around 25 responses is required for a product to be included in the detailed analysis. The number of responses about 'other' products is not included in the following table.

Vendor	Product name	Respondents
Datavault	2150 Datavault Builder	33
AnalyticsCreator	AnalyticsCreator	38
Informatica	Informatica Intelligent Data Integration incl. PowerCenter	32
Microsoft	Microsoft Azure	28
Microsoft	Microsoft SQL Server	110
Microsoft	Microsoft SQL Server Integration Services	49
Oracle	Oracle Database	32
Pentaho	Pentaho Data Integration	30
SAP	SAP BW	38
SAP	SAP BW on HANA	44
SAP	SAP BW/4HANA	34

In this first edition of The Data Management Survey, eleven products (or bundles of products) are featured. The products in the sample vary in their market focus and origin.

## Peer groups

The Data Management Survey 19 features a wide range of data management tools. Therefore, we use peer groups to help readers identify and compare competing products. The peer groups are defined using the criteria outlined in Table 3.

The peer groups are designed to help readers compare similar tools in terms the scenarios the products are used in. See Table 4 for an overview of the products in each peer group. These functional peer groups are mainly data-driven and based on how customers say they use the product.

#### Peer group Description Analytical database products prepare, store and provide data Analytical database products for analytical purposes. Data warehousing automation products cover data-driven or Data warehousing automation requirements-driven data warehouse design and products implementation. They mainly focus on the simplification and automation of data integration and data modelling tasks. ETL products connect, extract, transform and load data from **ETL products** various source systems to a target system for analytical purposes. Global vendors have a sales and marketing reach through subsidiaries and/or partners which gives them a truly global **Global vendors** presence. They are present worldwide and their products are used all around the world.

#### Table 3: Peer group descriptions

#### Table 4: Products by peer group matrix

	Analytical database products	Data warehousing automation products	ETL products	Global vendors
2150 Datavault Builder		х		
AnalyticsCreator		х		
Informatica PowerCenter		х	х	х
Microsoft Azure	Х			х
Microsoft SQL Server	х			Х
Microsoft SSIS		х	Х	Х
Oracle Database	х	х		Х
Pentaho Data Integration		х	х	Х
SAP BW	х	х		Х
SAP BW on HANA	х	х	х	х
SAP BW/4HANA	Х	х	Х	Х

## **Overview of the key calculations in The Data Management Survey 19**

#### Means and medians

This survey makes frequent references to different forms of averages — means and medians. Just in case your statistical knowledge is a little rusty, here's a quick reminder of the definition of the terms:

The mean is the usual arithmetic average. Its value is affected by every value in the sample, so a single large outlier can materially affect the mean, particularly with small samples.

The median is the value in the middle of the sample; that is, half of the sample is larger than the median, and the other half is smaller. It could be regarded as the 'typical value', and is affected by the number, but not the value, of outliers. One or two large or small outliers therefore do not affect the median.

#### Understanding multiple response questions

Several questions in The Data Management Survey 19 allow the user to make multiple responses. For example, we asked users what problems (if any) they encountered in their projects. Because many users had more than one problem, the number of responses is larger than the number of respondents.

This means that there are two ways to calculate the percentage of a given response: based on the total number of responses or based on the total number of respondents. We present The Data Management Survey results based on the number of respondents.

Calculating percentages based on the number of respondents tells us how likely a given respondent is to have the problem but results in percentages higher than 100 percent when all the problems are added together (e.g., 47 percent of all respondents reported that they have no significant problems). Conversely, calculating percentages based on the total number of responses would result in a total of 100 percent.

## Survey data collection

The survey was conducted by BARC, with data captured from January to May 2019. All data was captured online from a total of 782 respondents.

Respondents were solicited individually via BARC's own research panel and from dozens of vendor and independent lists, as well as websites from many different countries, with emailed invitations being sent to the lists in a staggered fashion.

At our request, most of the vendors notified their customers about The Data Management Survey using either their regular newsletters or websites. We also asked some bloggers to mention it. Each list and website had a different survey URL, though in all cases, the same questionnaire (in English or German) was used.

# BARC #DATA MANAGEMENT SURVEY 19

## **Understanding the KPIs**

The goal of this section is to help the reader spot winners and losers in The Data Management Survey 19 using well-designed dashboards packed with concise information. The survey includes a set of 12 normalized KPIs for each of the 11 products.

We have calculated a set of KPIs for each of the four peer groups. The values are normalized on the whole sample. Peer groups are used to enable fair and useful comparisons of products that are likely to compete.

The KPIs all follow these simple rules:

- Only measures that have a clear good/bad trend are used as the basis for KPIs.
- KPIs may be based on one or more measures from The Data Management Survey.

• Only products with samples of at least 15 - 30 (depending on the KPI) for each of the questions that feed into the KPI are included.

• For quantitative data, KPIs are converted to a scale of 1 to 10 (worst to best). A linear min-max transformation is applied, which preserves the order of, and the relative distance between, products' scores.

KPIs are only calculated if the samples have at least 15 - 30 data points (this varies from KPI to KPI) and if the KPI in question is applicable to a product. Therefore, some products do not have a full set of root KPIs. It is important to exclude KPIs based on small (and therefore not representative) samples to ensure that the graph scales are not distorted by outlier KPIs. In such cases, the product is still shown in the tables, but with a blank KPI value and no bar in the bullet graph or bar chart.

SARC ≝DATA MANAGEMENT SURVEY 19

#### Table 5: The KPIs

KPIs
Product satisfaction
Recommendation
Developer efficiency
Time to market
Data access
Price-to-value
Performance
Scalability
Maintenance efficiency
Compliance
Data governance
Skills availability

## **Reading the KPI charts**

We provide two different types of dashboards for viewing the KPIs. The first type is the Product Dashboard. A Product Dashboard displays all the KPIs for a single product. The second type is the KPI Dashboard, which displays the KPI values for each product in a peer group using simple bar charts. The products are sorted by value in descending order.



Figure 4: KPI dashboard used for displaying KPIs

In the KPI Dashboards (see Figure 4), the peer group average is indicated by a light blue bar.

# ISARC ≝DATA MANAGEMENT SURVEY 19

				Rank	in pee	er gro	up			Value		
Produ	ct satisfaction	5	5						6.5			
Recor	nmendation	7	,						2.7		I	
Devel	oper efficiency	3	}						5.7			
Time t	to market	5	5						7.4			-
Data a	access	5	5						7.2			•
Price-	to-value	4	ŀ						6.9		1	
Perfor	mance	3	}						7.8			-
Scala	bility	3	3						8.1			
Mainte	enance efficiency	3	3						7.2			
Comp	liance	1							10.0			1
Data g	governance	2	2						9.5			-
Skills	availability	e	3						7.0			0

Figure 5: Product dashboard used for displaying all KPIs for a product in a specific peer group

In Figure 5, the first column shows the KPI name and the second column indicates the product rank in the specific peer group. As previously mentioned, not every product is represented by the complete set of KPIs. The gray squares show how many products in the peer group have an adequate sample to be classified in each KPI. The next column shows the KPI values for the product in question in each KPI and the blue bars in the final column represent those KPI values against the peer group average, which is indicated by a vertical gray line.

# SARC ≝DATA MANAGEMENT SURVEY 19

## The KPIs (overview)

The following section provides the entire list of KPIs calculated for The Data Management Survey 19, as well as a description of the calculations.

KPIs are only calculated if the samples have at least 15 - 30 data points (depending on the KPI), so some of the products do not have a full set of KPIs. It is important to exclude KPIs based on small (and therefore unreliable) samples to ensure that the graph scales are not distorted by outlier KPIs based on small data samples. In such cases, the product is still shown in the tables, but with a blank KPI value in the bar chart.

Different readers will have their own views on which of these KPIs are important to them. For example, some people will regard compliance as vital, while others may consider recommendation or performance to be more important.

The KPIs below provide a good selection from which readers can choose those that best fit their own organization's requirements.

#### **Developer efficiency**

#### What we measure

This KPI is based on how users rate their tool in terms of developer productivity, e.g., for testing, deployment, reusability, ease-of-coding and use of metadata.

#### Why it is important

80% of development effort is spent on preparing data before users can use it. The data preparation process is complex and time-consuming. In order to make this process as efficient as possible, experts should be able to concentrate on the task at hand. Therefore, the extent to which a tool can support experts with development and testing functions or relieve them of administrative tasks is of vital importance.

#### How we measure

We ask participants to rate the developer productivity (IT cost, IT backlog reduction, reusability, focus on development) of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### Time to market

#### What we measure

This KPI is based on how users rate their tool in terms of adaptability (agility to adapt to new requirements).

#### Why it is important

This gives an indication of the extent to which the tool supports the ability to react quickly and adequately to changes or to create new applications as efficiently and quickly as possible.

#### How we measure

We ask participants to rate the time to market of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### **Maintenance efficiency**

#### What we measure

This KPI is based on how users rate their tool in terms of the complexity of system administration and optimization.

#### Why it is important

The number of objects for data storage and processing grows over time. The administration and maintenance of these objects often becomes an organizational challenge without functional support, especially in scenarios with multiple environments, developers, database administrators etc. The functional support of the tool can help to manage these artifacts. Tools can also support tracking changes in one or more artifacts and adapting them.

#### How we measure

We ask participants to rate the maintenance efficiency (e.g., for system operation / system optimization) of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### Compliance

#### What we measure

This KPI is based on how users rate their tool in terms of its fulfillment of regulatory requirements (e.g., transparency and auditing).

#### Why it is important

Statutory regulatory requirements such as Basel II, Solvency II and GDPR demand transparency and traceability of data and processes. Failure to comply can result in financial penalties. Technologies and the processing of data must also meet legal requirements. Such obligations can be met by making the necessary information accessible or by providing functions (such as audit reports and DQ dashboards) to provide this evidence.

#### How we measure

We ask participants to rate the support for compliance requirements (e.g., GDPR) of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### Performance

#### What we measure

This KPI is based on user feedback about the reasons why the product was chosen and complaints about the system's performance.

# SARC ≝DATA MANAGEMENT SURVEY 19

#### Why it is important

Performance satisfaction is crucial when loading or querying (large) datasets or when calculating data. In some ways, complaints about performance are more important than performance measured in seconds, because acceptable delays can vary depending upon how the system is used.

#### How we measure

We ask participants to rate the performance (load performance, query performance, processing performance) of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### Scalability

#### What we measure

This KPI is based on how users rate their tool in terms of the ability to upscale/downscale automatically with data volume and various workload scenarios.

#### Why it is important

For years now, we have seen an increasing demand for flexibility in the use of data and, at the same time, increasing complexity in data. The ability of a tool with analytical requirements to scale data volumes, heterogeneity and speed can be a crucial factor in the usability of the solution.

#### How we measure

We ask participants to rate the scalability (data, applications, data environments) of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### Data access

#### What we measure

This KPI is based on how users rate their tool in terms of data accessibility (number and variety of data sources supported, quality of support/connectivity to data sources, ease of access to data, data transparency, data quality).

#### Why it is important

Access to data is essential for BI to extract and prepare data from source systems and make it available for reporting or analysis. Access to data requires connectivity in the form of connectors, crawlers, interfaces to extract data from different source systems, files, applications and cloud services for further processing. On the other hand, data access describes the openness and interface variety of systems in order to be able to query data. Depending on the type of tool, one or the other is more important.

#### How we measure

We ask participants to rate the data accessibility (ease of access to data, data transparency, data quality) of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### **Price-to-value**

#### What we measure

This KPI is based on how users rate their tool in terms of price-to-value ratio.

#### Why it is important

Price-to-value is an important metric in today's cost-conscious age. As many a data management tool user has found, the costs of buying and supporting software quickly add up, especially when attempting to cost-justify adding new users. As more capabilities are pushed out to the business, this perception of value becomes even more critical.

#### How we measure

We ask participants to rate the price-performance ratio of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### Recommendation

#### What we measure

This KPI is based on the proportion of users that say they would recommend the product to others.

#### Why it is important

No one knows more about how a product performs in the real world than the customers already using it. All too often, they find that products don't live up to expectations, or that the vendor does not support the product properly. Therefore, if existing users say they would recommend the product, we regard this as a positive indicator of its value.

#### How we measure

Users are asked whether they would recommend the product they are most familiar with. This measure is based on the degree and proportion of positive responses.

#### **Product satisfaction**

#### What we measure

This KPI is based on the proportion of users that say they are satisfied with their product.

#### Why it is important

If a product proves unreliable at a critical time, the results can be debilitating, and can even render an application unusable. However, not all customers have the same dependency on reliability, as some applications are not mission critical or time critical.

#### How we measure

Users are asked whether they are satisfied with the product they are most familiar with. This measure is based on the degree and proportion of positive responses.

#### Data governance

#### What we measure

This KPI is based on how users rate their tool in terms of support for data governance (i.e., support of privacy/security, policy management, analysis and monitoring capabilities for data).

#### Why it is important

Data governance functions are not only used to monitor and control data through functions for data security and data protection, data transparency and traceability, measurement of data quality, access control and workflows. They can also be used as an information medium for users to find out about data and its origin and nature.

#### How we measure

We ask participants to rate the support for data governance of their chosen product. To obtain the final KPI, we calculate an average weighted score per product.

#### **Skills availability**

#### What we measure

This KPI is based on how users rate their tool in terms of availability of skilled users within and outside the company.

#### Why it is important

Without the availability of expert knowledge on the market, tools are only usable to a limited extent. Otherwise, time-consuming and costly training for employees is required until the tools can be used properly.

#### How we measure

We ask participants to rate the availability of skilled users of their chosen product within and outside their company. To obtain the final KPI, we calculate an average weighted score per product.

# ■ BARC #DATA MANAGEMENT SURVEY 19

## Product picklist used in The Data Management Survey 19

2150 Datavault Builder Ab Initio Co>Operating System Actian DataConnect Amazon DynamoDB Amazon Elastic MapReduce (EMR) Amazon RedShift Amazon Relational Database Service (RDS) Apache Hadoop Apache Spark Attunity Compose AWS Batch AWS Glue Cloudera Data Virtuality Datavirtuality **Denodo Platform** Exasol Google BigQuery Google BigTable Google Cloud Dataflow **Google Cloud Spanner** Hortonworks Data Platform **IBM Data Refinery** IBM Db2 IBM Db2 Warehouse on Cloud IBM Cloud Private for Data IBM InfoSphere Federation Server IBM InfoSphere Information Server IBM Integrated Analytics System IBM PureData System for Operational Analytics Informatica Intelligent Cloud Service Informatica Intelligent Data Integration incl. PowerCenter Information Builders iWay Data Migrator Magnitude Gateway Magnitude Kalido

# ■ BARC #DATA MANAGEMENT SURVEY 19

MapR Data Platform Microsoft Analytics Platform System Microsoft Azure Analysis Services Microsoft Azure Cosmos DB Microsoft Azure Data Factory Microsoft Azure Data Lake Store Microsoft Azure HDInsight Microsoft Azure SQL data warehouse Microsoft Azure SQL DB Microsoft SQL Server Microsoft SQL Server Integration Services MID Innovator (for Database Architects) Oracle Autonomous Data Warehouse Cloud Oracle Data Integrator **Oracle Database** Oracle Exadata **Oracle Exalytics Oracle Warehouse Builder** Pentaho Data Integration SAP BW SAP BW on HANA SAP BW/4HANA SAP Data Hub SAP Data Services SAP HANA SAP HANA Smart Data Integration SAS Data Integration Server SAS Viya Snowflake Syncsort DMX **Talend Data Fabric** Talend Open Studio Teradata Aster Teradata Database Teradata IntelliCloud

# BARC #DATA MANAGEMENT SURVEY 19

Teradata Vantage TIBCO Data Virtualization Platform (Composite Software) TimeXtender Discovery Hub Trivadis biGENiUS Wherescape RED IBM InfoSphere Data Architect Oracle SQL Developer Data Modeler SAP PowerDesigner Erwin Data Modeler Idera ER / Studio Quest Software Toad Data Modeler Pitney Bowes Sagent Data Flow Pitney Bowes Spectrum Ataccama ONE Information Builders Omni-Gen AnalyticsCreator Acterys **IBM Watson Studio** IBM Watson Knowledge Catalog

### **About BARC**

#### **BARC** — Business Application Research Center

BARC is a leading enterprise software industry analyst and consulting firm delivering information to more than 1,000 customers each year. Major companies, government agencies and financial institutions rely on BARC's expertise in software selection, consulting and IT strategy projects.

For twenty years, BARC has specialized in core research areas including Data Management (DM), Business Intelligence (BI), Customer Relationship Management (CRM) and Enterprise Content Management (ECM). BARC's expertise is underpinned by a continuous program of market research, analysis and a series of product comparison studies to maintain a detailed and up-to-date understanding of the most important software vendors and products, as well as the latest market trends and developments.

BARC research focuses on helping companies find the right software solutions to align with their business goals. It includes evaluations of the leading vendors and products using methodologies that enable our clients to easily draw comparisons and reach a software selection decision with confidence. BARC also publishes insights into market trends and developments, and dispenses proven best practice advice.

BARC consulting can help you find the most reliable and cost-effective products to meet your specific requirements, guaranteeing a fast return on your investment. Neutrality and competency are the two cornerstones of BARC's approach to consulting. BARC also offers technical architecture reviews and coaching and advice on developing a software strategy for your organization, as well as helping software vendors with their product and market strategy.

BARC organizes regular conferences and seminars on Business Intelligence, Enterprise Content Management and Customer Relationship Management software. Vendors and IT decision-makers meet to discuss the latest product updates and market trends, and take advantage of valuable networking opportunities.

#### BARC research reports bring transparency to the market



BARC's BI Trend Monitor 2019 reflects on the trends currently driving the BI and data management market from a user perspective. We asked close to 2,700 users, consultants and vendors for their views on the most important BI trends.



The Planning Survey 19 is the latest edition of BARC's major annual study into the use of planning software. Based on a worldwide survey of over 1,350 planning users and consultants, it provides detailed user feedback about 22 planning tools. Find out more at www.bi-survey.com



The BI Survey 18 is the world's largest annual survey of BI users. Based on a sample of over 3,000 survey responses, The BI Survey 18 offers an unsurpassed level of user feedback on 36 leading BI solutions. Find out more at www.bi-survey.com Copyright © BARC GmbH 2019. All rights reserved.

#### **Business Application Research Center – BARC GmbH**

# **BARC**

Germany BARC GmbH Berliner Platz 7 D-97080 Würzburg +49 931 880651-0 www.barc.de

#### Austria

BARC GmbH Meldemannstraße 18 / 01.14 A-1200 Wien +43 1 8901203-451

Switzerland

BARC Schweiz GmbH Täfernstraße 22a CH-5405 Baden-Dättwil +41 76 3403516

Rest of the World +44 1536 772 451