A data source is the location where data that is being used come from. In a database management system, the primary data source is the database, which can be located in a disk or a remote server. The data source for a computer program can be a file, a data sheet, a spreadsheet, an XML file or even hard-coded data within the program. The purpose of a data source is to gather all of the technical information needed to access the data — the driver name, network address, network software, and so on — into a single place and hide it from the user.

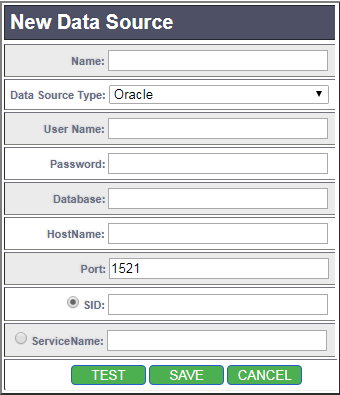
Data sources can differ according to the application or the field in question. Computer applications can have multiple data sources defined, depending on their purpose or function. Applications such as relational database management systems and even websites use databases as primary data sources. Hardware such as input devices and sensors use the environment as the primary data source.

 In basic terms, a data source is a facility for storing data. It can be as sophisticated as a complex database for a large corporation or as simple as a file with rows and columns. A data source can reside on a remote server, or it can be on a local desktop machine. Applications access a data source using a connection, and a DataSource object can be thought of as a factory for connections to the particular data source that the DataSource instance represents. The DataSource interface provides two methods for establishing a connection with a data source.

A good example is a temperature and pressure control system for a fluid circulation system such as the ones used in factories and oil refineries, which take all related data from the environment or whatever they are monitoring; so the data source here is the environment. Data such as temperature and pressure of the fluid are taken by sensors regularly and then stored in a database, which then becomes the primary data source for another computer application that manipulates and presents this data.

A data source is most commonly used in context with databases and database management systems or any system that primarily deals with data, and is referred to as a data source name (DSN), which is defined in the application so that it can find the location of the data. It simply means what the words mean: where data is coming from.

To add a new data source, user needs to click on “Add New…” hyperlink display on top of the page. There is a new web page to create new data source entry, before doing any testing user has to create new data sources.



**Name:** This is data source name; this name should be much relevant to the data so anyone can understand before using this. For ex. PRODUCT\_DW, this tells underline it connects to Product level information.  This is a name to identify the data source configuration.

**Data Source Type:** This can be multiple data source types, for ex Oracle, SQL Server, Sybase, XML file type, CSV file type, XLS sheet etc. This tells which type of data source user want to use.

**User Name & Password:** To connect with any DB, user needs to give user id and password. User Name is the name of a valid user for the database and password for the user name that user provided to connect to the database.

**Database:** While creating data source, user needs to define which data base it has to select. It is the name of the database to connect to by default, which is assigned by the database administrator.

**Host Name & Port:** This define server information which user wants to connect. Host name is the name or IP address of the machine where the database is running. Port is the number on which the database is running.

**SID or Service Name:** If the name of the Oracle driver that was selected contains the words "for Instance connections" enter the SID.

If the name of the Oracle driver contains the words "for Service connections" enter the service name.

References:

“Enterprise Data Warehouse Initiative EDW Planning Project.” Phase I Final Report 23 Feb. 07. University of California – Berkeley. Web. 7 Mar 2011.

Li, Jian; Xu, Bihua; "ETL tool research and implementation based on drilling data warehouse," Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on , vol.6, no., pp.2567-2569, 10-12 Aug. 2010 doi: 10.1109/FSKD.2010.5569836

McKnight , William. “Data Quality for the Next Decade.” Information Week. Nov 2010. Web. 20 Oct 2011. <http://www.information-management.com/issues/20_7/data-quality-for-thenext-decade-10019112-1.html>.

“Reporting Tool Selection in Data Warehouses.” 1keydata.com. n.d. Web. 20 Oct 2011. <http://www.1keydata.com/datawarehousing/toolreporting.html>.

Rogers, Denise. “Data Warehousing Architecture - Designing the Data Staging Area.” Database Journal. 22 June 2010. Web. 20 Oct 2011. <http://www.databasejournal.com/sqletc/article.php/3888696/Data-Warehousing-Architecture--Designing-the-Data-Staging-Area.htm>.

Sallam, Rita. Et al. “Magic Quadrant for Business Intelligence Platforms.” Gartner Research. 27 January 2011. Gartner, Inc. 25 April 2011.

G. Sanders, S. Shin, "Denormalization Effects on Performance of RDBMS," hicss, vol. 3, pp.3013, 34th Annual Hawaii International Conference on System Sciences ( HICSS-34)-Volume 3, 2001

Schaffhauser, Diane. “Florida State U Transforms Reporting with Business Intelligence.” Campus Technology. April ,2010.

Usman, M.; Asghar, S.; Fong, S.; , "Data mining and automatic OLAP schema generation," Digital Information Management (ICDIM), 2010 Fifth International Conference on , vol., no., pp.35-43,

Wai, T.T.; Aung, S.S.; , "Metadata Based Student Data Extraction from Universities Data Warehouse," 2009 International Conference on Signal Processing Systems , vol., no., pp.670- 673

Whitehorn, Mark. Et al. “Best Practices for Data Warehousing with SQL Server 2008 R2.” Microsoft Technet. December 2010. Microsoft Corporation. 24 Jun. 2011.

A. Simitsis, P. Vassiliadis, and T. Sellis, “Optimizing etl processes in data warehouses,” Data Engineering, International Conference on, vol. 0, pp. 564–575, 2005. 8

M. I. Hwang and H. Xul, “The effect of implementation factors on data warehousing success : An exploratory study,” Journal of Information, Information Technology, and Organizations, vol. 2, 2007. 13

M. Golfarelli and S. Rizzi, “Designing the data warehouse: Key steps and crucial issues,” Journal of Computer Science and Information Management, vol. 2, 1999. 14, 15

T. Ariyachandra and H. Watson, “Key organizational factors in data warehouse architecture selection,” Decision Support Systems, vol. 49, pp. 200–212, May 2010. 17, 18, 19, 20, 21

H. Watson and T. Ariyachandra, “Data warehouse architectures: Factors in the selection decision and the success of the architectures,” tech. rep., Terry College of Business, University of Georgia, July 2005. 21

O. Corporation, “Best practice for real-time data warehousing.” Oracle Corporation, World Head Quaters 500 Oracle Parkway, 2010. 32, 33, 34

Trevor Atkins, COTS Tools vs. Scripting Languages for Test Automation, Thinking Through Testing, http://thinktesting.com/articles/cots-tools-vs-scriptinglanguages-for-test-automation (2002)

George Candea, Stefan Bucur, Cristian Zamfir, Automated Software Testing as a Service, School of Computer and Communication Services, EPFL, Lausanne, Switzerland (2010)

Berenguer, G., Romero, R., Trujillo, J., Serrano, M., and Piattini, M. A Set of Quality Indicators and Their Corresponding Metrics for Conceptual Models of Data Warehouses. In A. Tjoa and J. Trujillo, eds., Data Warehousing and Knowledge Discovery. Springer Berlin / Heidelberg, 2005, 95-104.

Bhatti, K. and Ghazi, A.N. Effectiveness of Exploratory Testing An empirical scrutiny of the challenges and factors. 2010.

Brereton, P., Kitchenham, B., Budgen, D., Turner, M., and Khalil, M. Lessons from applying the systematic literature review process within the software engineering domain. Journal of Systems and Software 80, 4 (2007), 571-583.

Chays, D., Dan, S., Frankl, P.G., Vokolos, F.I., and Weber, E.J. A framework for testing database applications. SIGSOFT Softw. Eng. Notes 25, 5 (2000), 147-157.

Ciszak, L. Application of clustering and association methods in data cleaning. International Multiconference on Computer Science and Information Technology, 2008. IMCSIT 2008., IEEE (2008), 97–103.

Dell’Aquila, C., Tria, F. Di, Lefons, E., and Tangorra, F. Logic programming for data warehouse conceptual schema validation. DaWaK’10 Proceedings of the 12th international conference on Data warehousing and knowledge discovery, Springer-Verlag Berlin, Heidelberg (2010), 1–12.

Eno, J. and Thompson, C.W. Generating Synthetic Data to Match Data Mining Patterns. Internet Computing, IEEE 12, 3 (2008), 78-82.

Farré, C., Rull, G., Teniente, E., and Urp’\i, T. SVTe: a tool to validate database schemas giving explanations. Proceedings of the 1st international workshop on Testing database systems, ACM (2008), 1–6.

Golfarelli, M. and Rizzi, S. A comprehensive approach to data warehouse testing. International Conference on Information and Knowledge Management, Proceedings, (2009), 17-24

Haraty, R.A., Mansour, N., and Daou, B. Regression testing of database applications. Proceedings of the 2001 ACM symposium on Applied computing, ACM (2001), 285-289.

Huynh, T., Nguyen, B., Schiefer, J., and Tjoa, A. BEDAWA-A Tool for Generating Sample Data for Data Warehouses. Data Warehousing and Knowledge Discovery, (2000), 83–93.

Jarke, M. Architecture and quality in data warehouses: An extended repository approach. Information Systems 24, 3 (1999), 229-253.

Jeske, D., Lin, P., Rendon, C., Xiao, R., and Samadi, B. Synthetic data generation capabilties for testing data mining tools. MILCOM’06 Proceedings of the 2006 IEEE conference on Military communications, IEEE Press (2006), 1-6.

Kimball, R. and Ross, M. The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Wiley Publishing, 2010.

Mathen, M.P. Data Warehouse Testing. www.Infosys.com, 2010. <http://www.infosys.com/offerings/IT-services/independent-validation-testing-services/whitepapers/Documents/data-warehouse-testing.pdf>.

Mookerjea, A. and Malisetty, P. Data Warehouse / ETL Testing: Best Practices. Test India, (2008).

Muñoz, L., Mazón, J.-N., and Trujillo, J. A family of experiments to validate measures for UML activity diagrams of ETL processes in data warehouses. Information and Software Technology 52, 11 (2010), 1188-1203.

Muñoz, L., Mazón, J.N., and Trujillo, J. Measures for ETL processes models in data warehouses. MoSE+DQS ’09; Proceeding of the first international workshop on Model driven service engineering and data quality and security, ACM Hong Kong, China (2009), 33–36.

Singh, J. and Singh, K. Designing a Customized Test Data Generator for Effective Testing of a Large Database. International Conference on Advanced Computer Theory and Engineering, 2008. ICACTE’08, IEEE Computer Society (2008), 84 - 88.

Singh, J. and Singh, K. Statistically Analyzing the Impact of Automated ETL Testing on the Data Quality of a Data Warehouse. International Journal of Computer and Electrical Engineering (IJCEE) 1, 4 (2009), 488-495.

Tanuška, P., Moravčík, O., Važan, P., and Miksa, F. The Proposal of Data Warehouse Testing Activities. Proceedings of the 20th Central European Conference on Information and Intelligent Systems, (2009), 7-11.

Willmor, D. and Embury, S.M. A safe regression test selection technique for database-driven applications. Software Maintenance, 2005. ICSM’05. Proceedings of the 21st IEEE International Conference on, (2005), 421-430.

Manoj Philip Mathen. Data warehouse testing. Infosys DeveloperIQ Magazine, pages 1–8, 2010.

Jing Han, E Haihong, Guan Le, and Jian Du. Survey on NoSQL Database. In 6th International Conference on Pervasive Computing and Applications, pages 363–366, Oct 2011.

QuerySurge: big Data Testing, ETL Testing & Data Warehouse Testing. http://www. querysurge.com/ (Accessed 2017-09-20).

Hajar Homayouni, Sudipto Ghosh, and Indrakshi Ray. Data Warehouse Testing. Accepted for Publication in Advances in Computers, 2017.

Mladen Varga. On the Differences of Relational and Dimensional Data Model. In 12th International Conference on Information and Intelligent Systems, pages 245–251, 2001.

Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, and Bob Becker. The Data Warehouse Lifecycle Toolkit. Wiley, 2nd edition, 2008.

Foster Hinshaw. Data Warehouse Appliances: Driving the Business Intelligence Revolution. DM Review Magazine, pages 30–34, September 2004.

Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. A View of Cloud Computing. Communications of the ACM, 53(4):50–58, 2010.

Jose Barateiro and Helena Galhardas. A survey of Data Quality tools. Datenbank Spektrum, 14:15–21, 2005.

Jimison Iavindrasana, Gilles Cohen, Adrien Depeursinge, Henning Muller, R. Meyer, and Antoine Geissbuhler. Clinical Data Mining: a Review. Yearbook of Medical Informatics, pages 121–133, 2009.

David Loshin. Rule-based Data Quality. In 11th ACM International Conference on Information and Knowledge Management, pages 614–616, New York, NY, USA, November 2002.

Informatica. https://www.informatica.com/ (Accessed 2017-04-14)

Neveen ElGamal, Ali ElBastawissy, and Galal Galal-Edeen. Data Warehouse Testing. In the Joint EDBT/ICDT Workshops, pages 1–8, New York, USA, 2013.

Coral Calero, Mario Piattini, Carolina Pascual, and Manuel A. Serrano. Towards Data Warehouse Quality Metrics. In the International Workshop on Design and Management of Data Warehouses, pages 1–10, Interlaken, Switzerland, June 2001.

David Chays, Yuetang Deng, Phyllis G. Frankl, Saikat Dan, Filippos I. Vokolos, and Elaine J. Weyuker. An Agenda for Testing Relational Database Applications. Software Testing, Verifi- cation & Reliability, 14(1):17–44, 2004.

Mukesh K. Mohania and A. Min Tjoa, editors. Data Warehousing and Knowledge Discovery: 12th International Conference, DaWaK. Springer, 2010.

Hajar Homayouni, Sudipto Ghosh, and Indrakshi Ray. On Generating Balancing Tests for Validating the Extract-Transform-Load Process for Data Warehouse Systems. 2018. Submitted to 11th IEEE Conference on Software Testing, Validation and Verification.

Marius Marin. A Data-Agnostic Approach to Automatic Testing of Multi-dimensional Databases. In 7th International Conference on Software Testing, Verification and Validation, pages 133–142, March 2014.

Len Wyatt, Brian Caufield, and Daniel Pol. Principles for an ETL Benchmark. In Performance Evaluation and Benchmarking, pages 183–198, Berlin, Heidelberg, August 2009.

T. N. Manjunath, Ravindra S. Hegad, H. K. Yogish, R. A. Archana, and I. M. Umesh. A Case Study on Regression Test Automation for Data Warehouse Quality Assurance. International Journal of Information Technology and Knowledge Management, 5(2):239–243, 2012.

Todd L. Graves, Mary Jean Harrold, Jung-Min Kim, Adam Porter, and Gregg Rothermel. An Empirical Study of Regression Test Selection Techniques. ACM Transactions on Software Engineering and Methodology, 10(2):184–208, 2001.

Bruno Edson Martins de Albuquerque Filho, Thiago Luis Lopes Siqueira, and Valeria Cesario Times. OBAS: An OLAP Benchmark for Analysis Services. Journal of Information and Data Management, 4(3):390, 2013.

Byung-Kwon Park, Hyoil Han, and Il-Yeol Song. XML-OLAP: A Multidimensional Analysis Framework for XML Warehouses. In Data Warehousing and Knowledge Discovery, pages 32–42, Berlin, Heidelberg, August 2005.

Earl T. Barr, Mark Harman, Phil McMinn, Muzammil Shahbaz, and Shin Yoo. The Oracle Problem in Software Testing: A Survey. IEEE Transactions on Software Engineering, 41(5):507–525, 2015.

Lawrence Corr and Jim Stagnitto. Agile Data Warehouse Design: Collaborative Dimensional Modeling, from Whiteboard to Star Schema. DecisionOne Press, 3rd edition, 2011.

Roy P. Pargas, Mary Jean Harrold, and Robert R. Peck. Test-Data Generation Using Genetic Algorithms. Software Testing, Verification And Reliability, 9:263–282, 1999.