Become our client

We provide Data– driven insights to a diverse client base that spread across various industry sectors. Be confident to receive the Expertise, the Skills, the Knowledge, the Resourcefulness, and the Assurance that a seasoned Expert will be there every step of the way.

Email: enquiries@expertskills.com.au

Upcoming Events

• Mastering Business Analytics with SAP Solutions

Fantastic event for companies that use SAP BusinessObjects BI as an enterprise platform to enhance decision making. Practitioners from global organizations come together to share expedience, knowledge and innovation tips. October 23-26, Melbourne. Contact The Eventful Group for further information

• Tableau Test Drive - Melbourne

Tableau Test Drive session is designed to demonstrate ow to connect to data and visualize queries without writing a single line of code. Self-serve analytics for everyone.
February 16, 2016 , Contact Saxon's Training Facilities for further information.

Expert Skills Pty Ltd
PO BOX 1164
BLACKBURN NORTH
VIC 3130
P: 8819 0000
M: 0403 575244
W: www.expertskills.com.au

ISSUE
01
Q1
2016

QUARTERLY
JOURNAL OF
BIG DATA
DATA WAREHOUSING
BI & ANALYTICS
DATA ANALYTICS
DATA MANAGEMENT

DataInsights

this issue

Data Warehouse Features

Subject Oriented -Data is selected based on a particular subject of interest or business functions and stored in an effective format as a single set of data (Inmon & Kimball 1996).

Integrated - Whilst underlying source systems may store data in a particular format unique to that system, data that is stored in a data warehouse is defined and stored according to specific set of guidelines including naming standards and physical attributes (Inmon & Kimball 1996).

Non volatile - Data in a data warehouse is static and is not instantly affected by updates that may be occurring at a source systems level (Inmon & Kimball 1996).

Time variant - Data in a data warehouse is both historical and current. Historical data can be used for trending for up to 10 years where as operational is limited up to 90 day periods (Inmon & Kimball 1996).

Data Warehouse Concepts

With numerous industry definitions available in reference materials, in general terms, data warehouse is a database, different to transactional processing, that can combine data from multiple disparate sources designed to enable business intelligence activities through the use of data analysis and reporting by organizations.

Ponniah (2010) defines the data warehouse as an informational environment that: (a) provides an integrated and total view of the enterprise; (b) Makes the enterprise's current and historical information easily available for strategic decision making; (c) makes decision-support transactions possible without hindering operational systems; (d) renders the organization's information consistent; (e) presents a flexible and interactive source of strategic information.
Similarly, Inmon and Kimball (1996) define Data Warehouse as a set of Subject-Oriented, Integrated, Time-Variant, Non-Volatile data used in support of management decisions.

Furthermore, Imhoff, Galemmo and Geiger (2003) suggest that the data warehouse has been a part of the Business Intelligence (BI) system's architecture from the very beginning.

The Data Warehousing Institute (cited in Andrews & Guerra, 2013, p. 2) defines business intelligence as the process, technologies, and tools needed to turn data into information, information into knowledge, and knowledge into plans that drive profitable business action.

Business Intelligence is not just about glossy reports and ad-hoc analysis; it's a system in its own right that is supported by data warehousing, theories, methodologies, architectures, and analytical and content management tools (Andrews & Guerra 2013). These systems support organizations by turning data into meaningful information to aid in better decision-making (Rud 2009).

Demand for BI applications is strong and continues to grow (Soejarto 2003). To succeed and survive in today's competitive business environment, reliance on effective and timely business information is essential (Lönnqvist & Pirttimäki 2006).

Data Warehouse Benefits

Numerous data warehouse benefits are well documented and circulated in both the academic and IT professional community.

Meyer and Cannon (1998) explain that a data warehouse equips its users with effective decision support tools by integrating corporate wide data into a single repository from which end users can run reports and perform ad hoc data analysis.

Furthermore, data warehouse can help reduce costs, increase value added activities, and improves efficiency (Chiang et al. 2003). According to BI-Insider (2011), there are a number of benefits for implementing a data warehouse that include:

- Enhanced BI through the use of improved information access
- Increased query and system performance through the use of designated database servers and OLAP technologies
- BI from multiple data sources through the use of data mashing
- Timely access to data through the use of ETL frameworks
- Enhanced data quality and consistency through the use of single source of truth
- Historical intelligence through the use of data warehouse's ability to store historical data
- High return on investment through the use of data warehousing technologies to make better decisions, which in turn translates to competitive advantage.

Typical Data Warehouse architecture(s)

We look at leading data warehouse architectural types that are commonly used by organizations.

Data warehousing has gained acceptance during the 1980's and 1990s (Ponniah 2010, p. 17) Since then, the typical data warehouse architectures have been evolving to range from Independent Data Marts to Centralized enterprise data warehouses, and different approaches including top-down, bottom-up, and hybrid methodologies.

The leading data warehouse architectural types are: (a) Centralized EDW; (b) Decentralized Independent Data Marts; (c) Federated; (d) Hub-and-Spoke architecture; and (e) Data-Mart Bus architecture (Ponniah, 2010). Figure 1 compares different data warehouse architectural types.

Centralized Data Warehouse

- This architecture type uses 3NF (third normal form) where data is normalized and stored at the lowest granularity level. Whilst this architecture type takes into account enterprise-level information requirements to provide a single source of truth, there are no separate data marts and this architecture is the most expensive to implement (Ponniah 2010).

Furthermore, this architecture is based on Top-Down approach noted by Inmon and Kimball (1996), who defined data warehouse as a centralized repository for the entire enterprise.

Independent Data Marts - With this architecture, independent data marts are developed by business units for specific business functions or purpose. Although being the cheapest option to implement, these data marts independently evolve over time and fail to provide a single version of truth in the long run (Ponniah 2010).

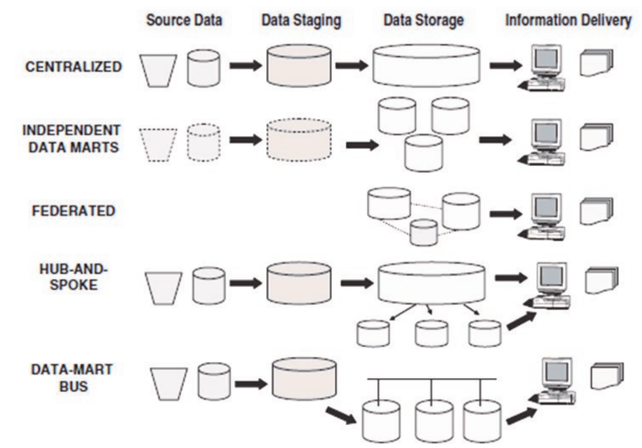
This particular architecture is based on Bottom-Up Approach and noted by Inmon and Kimball (1996) who envisions the corporate data warehouse as a collection of conformed data marts.

Federated - Noted as the practical solution, federated architecture aims to integrate data logically or physically using shared key business fields or global metadata.

Distributed queries can be executed against the data but there is no overall data warehouse (Ponniah 2010).

Hub-and-Spoke - Much like Centralized EDW architecture, overall EDW exists in the hub-and-spoke architecture and stores data in the 3NF. However, additional Decentralized Independent Data Marts are also included allowing for rapid development to onboard additional business subject (Ponniah 2010).

Data-Mart Bus - With this architecture, first super-mart is built using dimensions and metrics and focuses on a particular business subject area. Additional data marts that are brought on-line will share/re-use the same dimensions (Ponniah 2010).



Data Warehouse Components

A data warehouse can be categorized into 3x major layers. With each vendor depicting their own, typically, data warehouse architecture consists of the following: (a) Data Integration Layer; (b) Data Warehouse Layer; (c) Presentation Layer.

In addition, beneath each layer, there are a number of major components that make data warehouse tick. Kimball and Ross (2002) define the following components that make up the complete warehousing environment

- Operational Source Systems where original transactions take place
- Data Staging Area where data from various disparate operational systems is combined
- Data Presentation where data is transformed for end users to query
- Data Access Tools where software is used to interrogate the data

Operational Source Systems are designed to be highly scalable and available (up time), focused on processing performance and are used to capture business transactions (Kimball & Ross 2002).

The source systems can be a single instance or a combination of ERP or CRM systems, other operational or legacy data stores/systems and possibly excel and csv files that support business function(s). Traditionally, operational source systems are On-line Transaction Processing (OLTP) systems.

Data Staging Area in a data warehouse brings together all the data from various source systems. Data flow is performed using Extract, Transform, Load (ETL) steps (Kimball and Ross 2002, p. 33).

ETL can be seen as a critical component in overall integration architecture. It loads the data into a data warehouse and is accountable for data transformation (Bouman, Casters & Doongen 2010).

The Extract step manages connections to various source systems and allows for the source data to be available in one common Staging area, a copy of source data, for further processing steps (Bouman, Casters & Doongen 2010).

Once the data is available in the Staging area, Transformation step can be performed before the data is ready for loading to target destination (Bouman, Casters & Doongen 2010, p. 5). Transformation usually involves changes to original state data including summations, aggregations, and addition of calculated measures.

Once transformation step is complete, Load step can proceed to insert transformed data to target data warehouse (Bauman, Casters & Doongen 2010). Most often, the Load is scheduled to refresh the data warehouse on nightly basis.

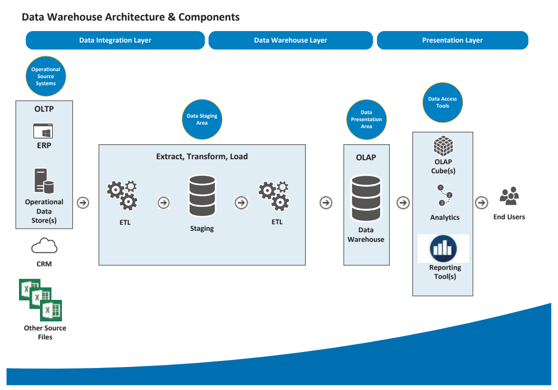
Transformed data is organized and stored at the **Data Presentation Layer.** From there, the data can be queried by business users, analyzed by data analysts, and used for reports by report writers (Kimball & Ross, 2002).

Kimball and Ross (2002) define **Data Access Tools** as the final major component of the data warehouse environment. A data access tool can be as simple as an ad hoc query tool or as complex as a sophisticated data mining or modelling application (Kimball & Ross 2002, p. 39).

Major BI software vendors compete furiously for the Leader's spot in the Gartner BI Magic Quadrant by offering agile analytical BI software that can access and query data warehouses.

These tools make it easier for business users to interrogate data and remove the need for direct SQL queries that can sometimes be erroneous (Armstrong 1997). These tools are often offered free of charge as standalone installations with an aim to penetrate business user community.

To complement this, business users are showing more drive to become more tech savvy as vendors roll out these innovative, easy to use, and scalable product offerings.



Data warehouse database uses On-line Analytical Processing (OLAP) technology that is delegated with storing transformed data to provide a scalable system that offers high performance query handling, analysis and calculations. The database stores aggregated, historical data, stored in multi-dimensional schemas.

For organizational users, data warehouse means access to the data presentation via Data Access Tools where all of the querying, analysis and report development is performed.

KEEP AN EYE ON Current Trends

2016 Gartner's Magic Quadrant

The 2016 Magic Quadrant for BI and Analytics sees SAP, IBM, Oracle, MicroStrategy and SAS drop from the Leadership Quadrant. The three companies remain as leaders are Tableau, Qlik and Microsoft. Microsoft Power BI is the focus for this Magic Quadrant and is on its second major release, offering cloud-based BI with a new desktop interface — Power BI Desktop.

SOFTWARE Quarterly Picks

Microsoft Power BI

Microsoft Power BI transforms company's data into rich visuals . It offers data preparation, data discovery and interactive dashboards via a single design tool .

Connect and transform data, create advanced calculations, and build stunning reports in minutes.

Power BI