# Modern Data Warehousing & Business Intelligence

Great strides have been made in the world of Big Data technologies over the past decade. Cloud infrastructure along with new storage paradigms like Hadoop, graph databases and message buses have provided a new dimension to the arsenal of tools available to tackle the deluge of data the world is facing.

A well-structured SQL-based data warehouse offers a technological maturity that is still lacking from these new world tech solutions. A "modern" data warehouse architecture that combines SQL and not only SQL (NO-SQL) technologies is a preferred way to store and enrich enterprise data. "Small data" is now seen as latest fad, where the focus is back on quality rather than quantity.
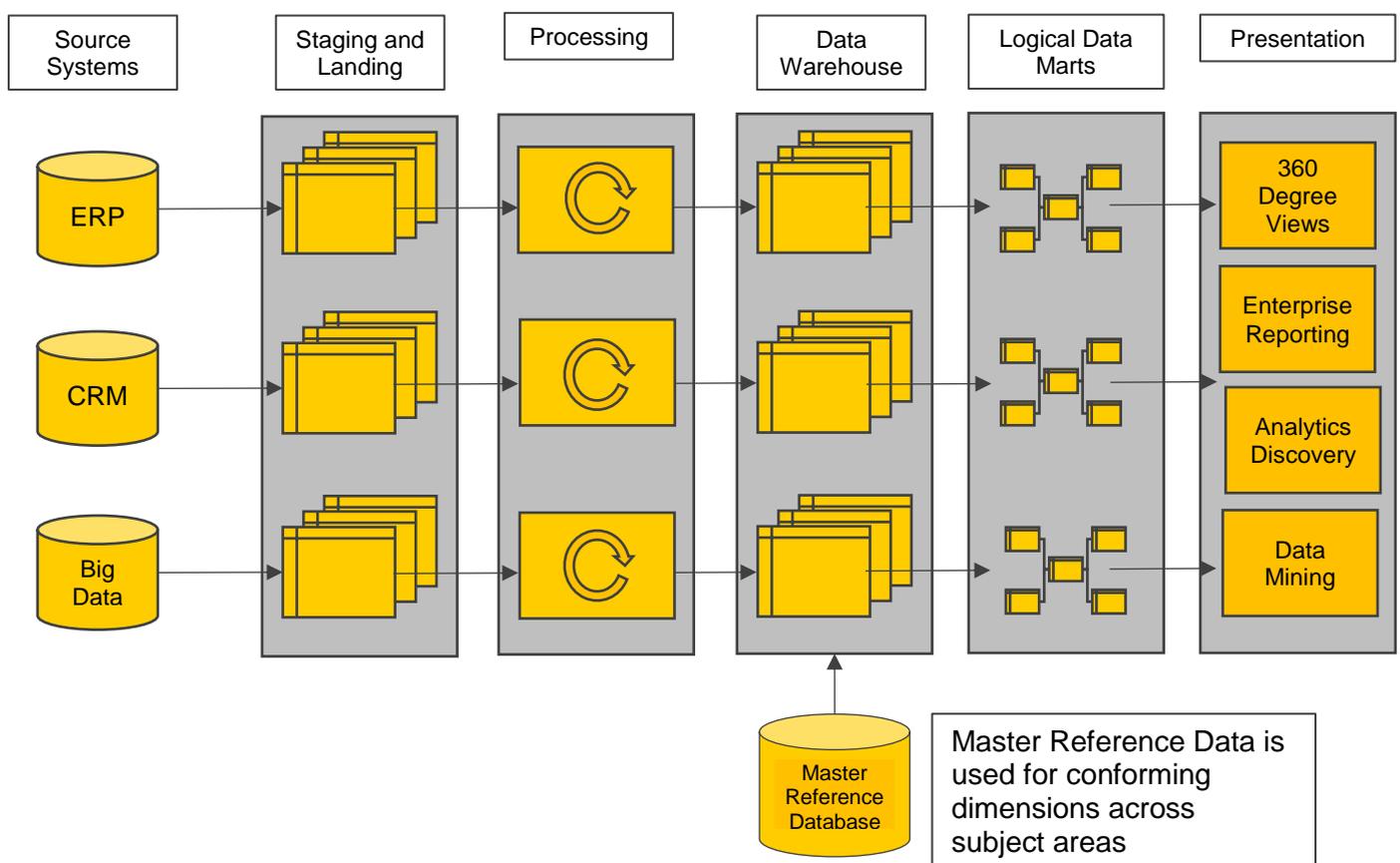
Data Warehouse High-Level Architecture

The architecture for a data warehouse hasn't changed much over the years except for the inclusion of "Big Data" sources. The figure above, outlines the layers and concerns of such an environment:

1. Source systems – Transactional systems used for day-to-day business operations.
2. Staging and Landing – Data is received from the source systems either by extracting data or having the source systems pushing the data into a landing area.
3. Processing – Applying business logic and rules in one central location
4. Data Warehouse – Location where the processed data is published and conformed to enterprise entities.
5. Logical Data Marts – Logical or virtual subject orientated data markets (i.e. sales, marketing, finance, etc.) that feeds off the data warehouse.
6. Presentation – Presentation layer where the business users will consume the information.

The following high-level stages are generally prescribed when implementing a data warehouse that will provide a return on investment to the business:

High-Level Stages and Steps for Establishing A Modern Data Warehouse

Stages	Step	Tasks
Analysis & Design	Identify the top burning questions business needs to answer to create a competitive edge	• Hold brainstorming workshops with business stakeholders to identify the issues that keep the executive awake at night. • Identify and analyze business unit's key concerns and underlying business processes
	Develop a business case for performing the analytics	• Quantify the economic impact of the identified concerns. • Determine the value of "**perfect information**" to the organization. • Document a business case that defines the return on investment to the organization
Develop	Establish a foundational data warehouse Aligned to industry best-practice and internal organisation governance standards.	• Provision server infrastructure for development, user acceptance testing, pre-production and production • Profile the source data systems to identify missing gaps • Undertake a data modelling of the fact and dimension tables. • Establish an operational data store (ODS) making use of new big data technologies, like Hadoop • Identify the high-level entities that need to be modelled in the business (i.e. customer, accounts, products, transactions, etc.). This feeds from the analytical mastering exercise. • Develop the extract transform and load (ETL) packages required for staging, processing and publishing tasks. • All business rules and logic need to be centralised for ease of maintenance.
	Testing	• Develop automated testing scripts for matching source to target systems
	DevOps	• Make use of DevOps processes to move the developed ETL packages and supporting scripts to a production environment. • Maintain a master build that is moved through development to production
Transition	Hand over operations	• Handover key technical documents, such as high-level architectures, metadata and data dictionaries

		• Hand over the running and maintenance of the warehouse system to client • Provide training workshops

Many analytical tools and insights are enabled as the data warehouse starts to be populated with data emerge as a reliable source of information. The list of these can be exhaustive, however below are the ones that we consider relevant.

Analytical Insight Benefits

Analytical Insights		High Level Description	Typical Attributes
360 Degree Views	Relational	These types of reports address a subject area in the enterprise (i.e. the customer, a supplier, asset, product, etc.), typically the customer and their behavior. These types of reports are useful for downstream advanced analytical discovery using advanced visualization, statistical and or machine learning.	• Reports at the level of the subject • Many attributes (sometimes 100s of columns)
	Graph	True 360-degree view of a subject that allows for lateral exploration of connected attributes. Due to the unstructured nature of this type of view, entity information is stored in graph format requiring a semantic database engine. Allows transversal pattern discovery, specifically useful for forensics and root-cause analysis.	• Requires a semantic graph database • Delivers on the promise of a single view
Enterprise Reporting	Operational	Reports that address the real-time operational needs of the enterprise. This includes human and systems consumers. These reports generally retrieve information from an operational data store (ODS) or directly from the operational system	• First order reporting • Build for purpose
	Strategic / Tactical	Reports that address the higher-level requirements of the enterprise. These are tactical and strategic in nature and typically report off aggregate level data. Typically, the sources of these reports are derived from data warehouses having cleansed and conformed data. These reports generally source data from OLAP cubes.	• Single-Version-of-the-Truth reporting • Typically, OLAP cube based • Drilldown capabilities

	OLAP	An analytical tool used for slice-and-dice operations typically deriving information from OLAP cubes. Has the ability to process aggregate level calculations through MDX calculations. Movement towards in-memory processing which offers speed advantages.	• Power Users reporting • Allows Drill-through
Analytical Discovery	Visualization	Visualization platforms that dynamically and visually represent a subject area. The goal of these types of tools is to expose high dimensionality to the user through visual means as the optical system of a human user has the highest interpretation throughput capability. High dimensionality refers to the ability to represent data series, category information, values on multiple axes, time, animation and data point attribute representation (i.e. colors, point size and shapes).	• Discovery tools • Requires agile provisioning of data to be effective
Data Mining	Predictive Modelling	Business users can use of software to describe and predict outcomes based by making use of statistical techniques.	• Statistics • Machine Learning

Proposal / Call to action

Ilion are looking to help financial services organisations optimise their return of investment in data centred around financial products and services. Our executive team have extensive experience in financial analysis for banks, credit risk and stress testing and data management. We see a overlap in these areas that provide a sweet spot for a return on investment made.

When you are working on your next financial analytical undertaking, please don't hesitate to call us. We would be happy to help to find a solution that is optimal for your organisation.

Contact Details:

Peter Gross
Director: Data Management
Email: peter@ilion.co.za
Mobile: +27 (0)83 388 1674