



*White Paper presented by*

# Analytic Data Solutions, LLC

**Title:** Examining Top Performing Best Practices of BI Solutions

**Author:** By Gary Hanson

**Date Written:** 1/7/2013

**Date Posted (to website):** 1/31/2013

**Brief Description:** This paper is second in a series addressing best practice of developing BI solutions. The objective of this paper is to highlight and describe 10 common, top performing BI best practices.



# Examining Top Performing Best Practices of BI Solutions

By Gary Hanson, Principal Consultant, Analytic Data Solutions, LLC

January 7, 2013

## Table of Contents

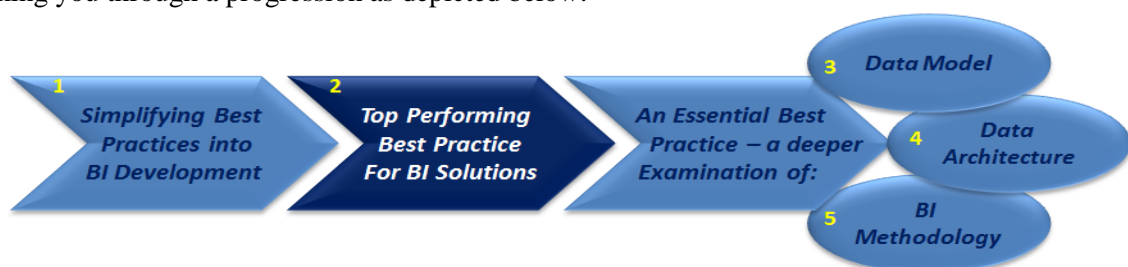
<b>OVERVIEW.....</b>	<b>2</b>
<b>TOP PERFORMING BEST PRACTICES FOR BI SOLUTIONS.....</b>	<b>2</b>
1. ESTABLISH A DATA DESIGN DRIVEN BY BUSINESS REQUIREMENTS AND FOSTERED BY STEWARDSHIP .....	3
2. ADOPT AN ITERATIVE, FLEXIBLE DEVELOPMENT METHODOLOGY ADDRESSING DW AND BI SOLUTIONS .....	4
3. ADHERE TO THE DESIGN RULES OF DIMENSIONAL MODELING FOR ALL ANALYTIC DATA MODELS .....	4
4. ALLOW ONLY DATA THAT HAS BUSINESS ANALYTIC VALUE TO BE STORED IN AN ANALYTIC DATA MART LAYER .....	5
5. UTILIZE A THREE TIER DATA ARCHITECTURE TO MANAGE ENTERPRISE DATA .....	5
6. ACQUIRE DATA FROM THE TRUE DATA SOURCE OR SYSTEM OF RECORD .....	6
7. LEVERAGE THE CORPORATE INVESTMENT AND FUNCTIONALITY OF STRATEGIC TOOLS .....	6
8. APPLY APPROPRIATE DATA RECONCILIATION AND DATA QUALITY TO THE DATA ARCHITECTURE .....	7
9. APPLY APPROPRIATE ACCESS CONTROL AND SECURITY CONSIDERATIONS TO THE DATA ARCHITECTURE .....	7
10. KEEP PERFORMANCE IN MIND WHEN DESIGNING SOLUTIONS RATHER THAN WAITING .....	7
<b>CONCLUSION .....</b>	<b>7</b>

### Overview

There are many proven development best practices and the purpose of this paper is to examine ten top performing best practices regarding Business Intelligence (BI) development. It is fair to say, a common goal of all development projects is to implement the solution on-time and on-budget. Unfortunately, there are too many times that we miss that goal. A common reason for a significant project delay or stoppage is when the project team either bypasses a necessary step or applies an inferior design technique. More often than not, following a proven **best practice** can help to avoid costly mistakes, keeping the project on-schedule and on-budget.

As stated in the first paper, ‘*Simplifying Best Practices*’, a cause of project delays may come from the tendency to rush the design activities by either cutting corners or not properly applying a technique. Subsequently, a significant benefit of **best practices** is that it contributes to a form of mentoring and educating of your staff to the proper techniques, ensuring consistent results. Your investment in maintaining and evolving **best practices** will pay dividends in delivering timely BI solutions.

To properly address the breadth of this theme, **best practices**, this paper is the second in a series of white papers taking you through a progression as depicted below:



The purpose of “*Examining Top Performing BI Best Practices*” is to highlight the significance or resolve of each **best practice** by providing a summary outline as an important factor to both Data Warehouse (DW) and Business Intelligence (BI) development. A key component is the discipline of capturing your experiences or lessons learned back into the **best practice** itself. This enhances the value proposition of **best practices** and contributes to achieving the highest level of business capabilities via BI solutions.

### Top Performing Best Practices for BI Solutions

BI solution is a broad term typically used to describe some sort of business capability involving technology and analytics. So before we get into the ‘how to incorporate **best practices** into your development processes’, let’s talk a bit in terms of what is a BI solution and why create a **best practice**. I think it is safe to say that most BI solutions have a fair amount of design complexity and business complexity associated with them. When appropriate, these complexities will benefit the most by applying repeatable and consistent **best practices** to them.

*BI solutions represent many types of deliverables that come from either DW or Analytic Reporting development projects. As examples, a BI solution could range from creating a new data mart to delivering a new analytic reporting capability to deploying a new metadata repository.*

A way to determine if creating a **best practice** merits consideration is to weigh its importance related to BI solutions as well as the level of risk associated to the effort. Here are a few questions to ask:

Is it important to design certain components using a consistent and repeatable technique? Is it important to follow a certain process using a consistent and repeatable method? Do you have a rather inexperienced team? Are the implications or consequences of not conforming to **best practices** considered high risk? Is BI solution development occurring across several organizations and/or geographically separated groups?

**Best practices** are meant to augment your development process and not replace it, so look for areas to complement work activities and avoid creating conflicting instructions. The following outline summarizes a set of common **best practices** that I have used and found very valuable in developing DW and BI solutions:

1. Establish a data design that is driven by business requirements and fostered by business stewardship
2. Adopt an iterative, flexible development methodology addressing DW and BI solutions
3. Adhere to the design rules of dimensional modeling for analytic data models
4. Allow only data that has business analytic value to be stored in the analytic data mart layer
5. Utilize a three tier data architecture to manage Enterprise data
6. Acquire data from the true data source or system of record
7. Leverage the corporate investment and functionality of strategic tools
8. Apply appropriate data reconciliation and data quality considerations to the data architecture
9. Apply appropriate access control and security considerations to the data architecture
10. Keep performance in mind when designing solutions rather than address it as an afterthought

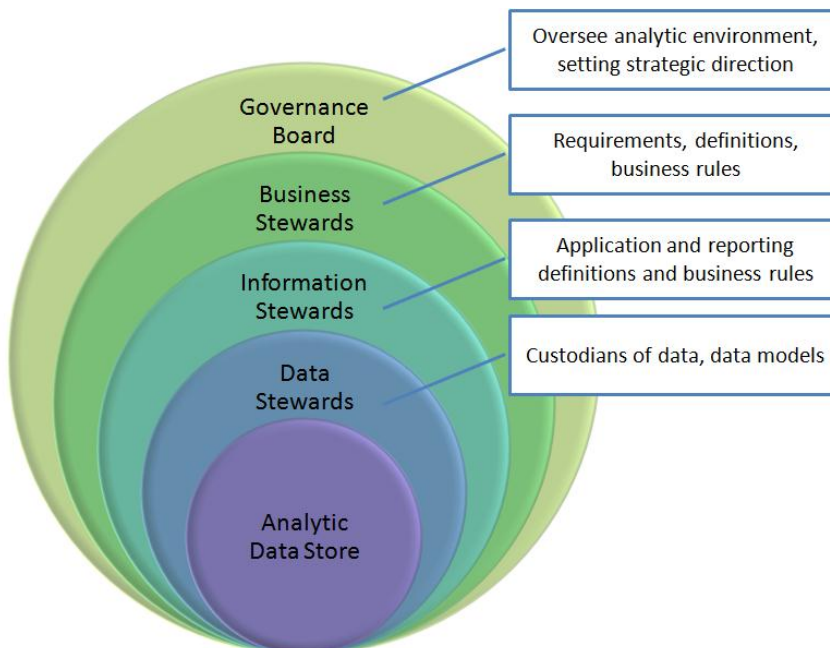
A key take-away is each **best practice** is branded to the way your organization performs it and over time, it is enhanced by applying your experiences to each one. The following outlines represent a summary view of each **best practice**; subsequently, there is much more information available on each topic. A third series of white papers, addressing this important theme, will take a deeper examination of selective **best practices**.

### 1. Establish a data design that is driven by business requirements and fostered by stewardship

Data design should be triggered and driven by business requirements that are sponsored and overseen by a data governance or stewardship program. These business requirements are captured and maintained in a corporate metadata repository and managed by the assigned data steward. Stewards play different roles depending upon their relationship to the information or data. They may make the final decision regarding data in a specific design, they may need to influence or have input on decisions surrounding it, or they may just need to be informed of any changes or enhancements.

**Justification** – ensures that the key business drivers and requirements are captured allowing the appropriate business capabilities to be developed and implemented.

**Implications** – makes design decisions on either incorrect assumptions or incomplete requirements will require costly rework and re-engineering as well as project delays.



**Exceptions** – there may be technical reasons, such as driving a new feature of a BI solution, where stewardship does not need to be involved.

## 2. Adopt an iterative, flexible development methodology addressing DW and BI solutions

It is typical for DW and BI solution initiatives to only have high level business drivers as business requirements prior to the development effort. Subsequently, an iterative approach promotes using a more agile format that allows for different ways of capturing a full set of detailed business requirements during the early stages of development. This could include establishing a pilot team to construct a prototype process and application, for example.

**Justification** – introducing an iterative development process may help to avoid costly rework by addressing upfront creative ways of capturing business requirements as well as designing and implementing business solutions.

**Implications** – using a Standard Development Life Cycle (SDLC) may not bring out data centric designing, coding, & testing complexities found in BI solutions. Also, using a standard SDLC approach may fall short in meeting the business needs and expectations. This is because most SDLC's are tailored for transactional systems application development rather than decision support development.

**Exceptions** – when full business requirements have been captured and completely accepted by the business sponsor and/or the business solution is straight forward (simple and natural).

## 3. Adhere to the design rules of dimensional modeling for all analytic data models

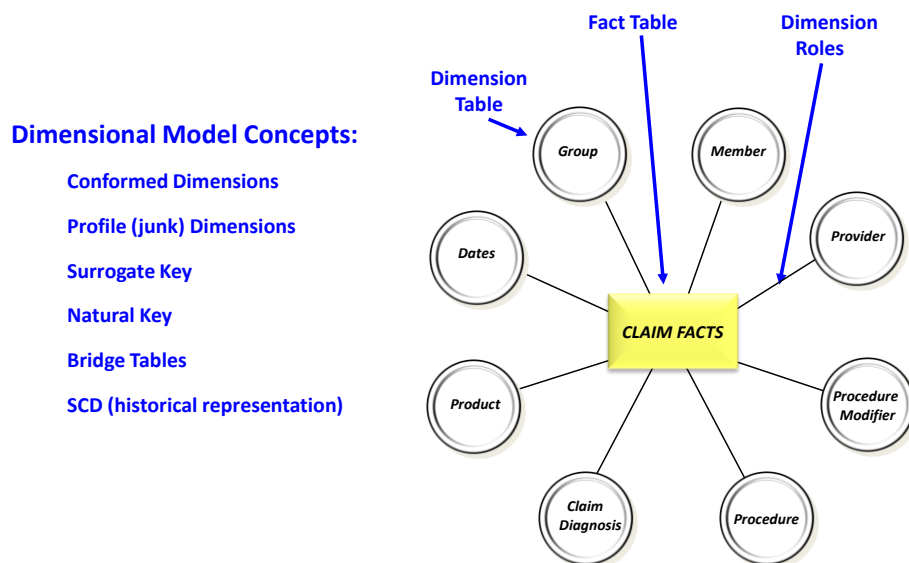
Adhere to the design rules of dimensional modeling for the format and storage of all analytic data. This data would be represented in the 3<sup>rd</sup> tier of an analytic data architecture diagram, usually called the Analytic Data Mart (ADM) layer. The design of the data model will adhere to the techniques of star-join-schema and associated population strategies.

**Justification** – dimensional modeling is a proven industry standard for analytic reporting ensuring accessibility and extensibility as well as easy query navigation and good performance.

**Implications** – not adhering to these modeling techniques may put extra strain on future enhancements forcing unnecessary data replication and reduced data quality. Deviation from this design practice may require additional training of both architecture, ETL, and BICC teams regarding maintenance, support, and training.

**Exceptions** – any exception to the above should be formally approved and well documented as an approved exception.

### Dimensional Data Model



### 4. Allow only data that has business analytic value to be stored in an analytic data mart layer

An important principle to the design of an Analytic Data Mart (ADM) and the 3<sup>rd</sup> tier of the analytic data architecture) is to understand the key purpose behind it. The primary purpose is to serve as a single data source for analytic reporting; whereas, the integrated data store (the 2<sup>nd</sup> tier) supplements other operational reporting and downstream application needs for the Enterprise. The idea is to avoid adding unnecessary attributes that result in very fat dimensions and fact tables, which can lead to poor performance and maintenance issues. Attributes that should not be included are operational or system fields as well as free form text fields. In addition, hearing, ‘well, we may need them down the road’, isn’t a strong justification and should be investigated / questioned further.

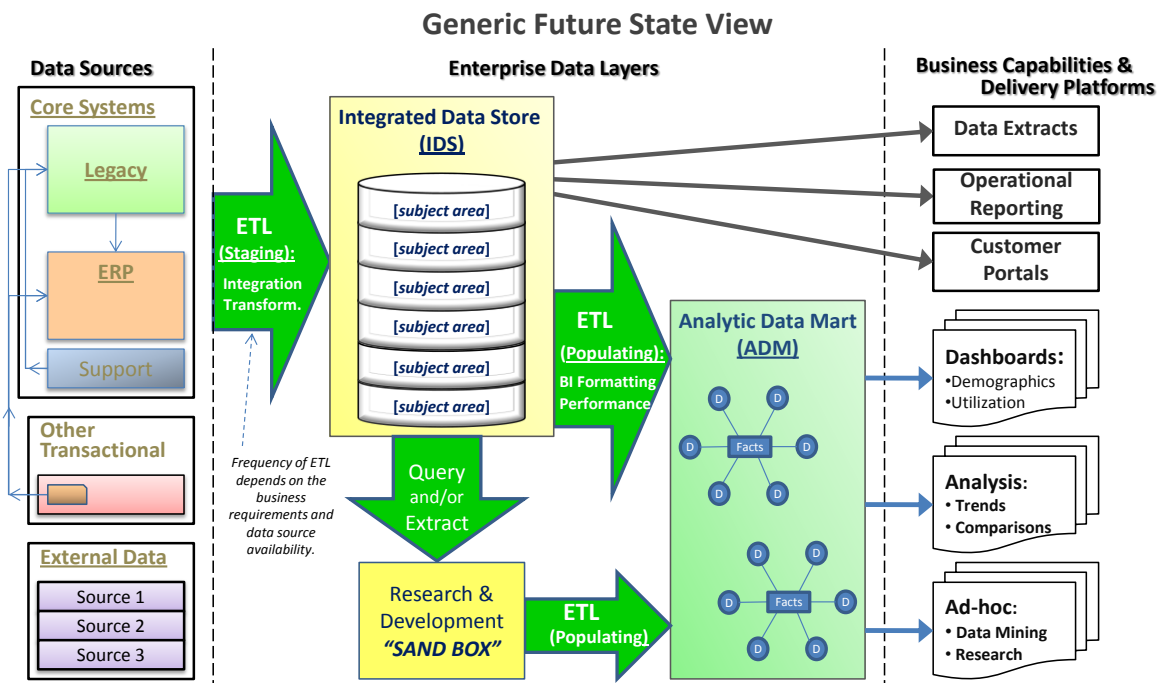
**Justification** – enforcing this design rule will keep your data marts lean and efficient ensuring the best platform for load and access performance as well as providing easier navigation of the data by business users.

**Implications** – not adhering to this principle has two primary consequences: 1) your data mart databases may grow to a size that requires more performance attention, prematurely and 2) analytic users may find navigating the marts more challenging as well as questioning the data content.

**Exceptions** – unless the analytic data mart layer is allowed to be used for non-analytic purposes, there are no exceptions to this best practice.

### 5. Utilize a three tier data architecture to manage Enterprise Data

Utilizing a three tier data architecture provides the ideal means for managing enterprise data in a true data integration framework as well creates the appropriate data platform for achieving a self-service reporting capability. The Integrated Data Store (IDS) contains the single source of operational and business analytic data acting as the 2<sup>nd</sup> tier and this is where full data integration for the Enterprise occurs. The Analytic Data Mart (ADM) layer contains the single source of business analytic data acting as the 3<sup>rd</sup> tier and the primary purpose here is to facilitate business analytic capabilities.





As you see in the Generic Future State View diagram (above), this platform is designed to serve all downstream application and data needs of the organization including analytics. A key premise here is to perform data integration and transformation rules once in a streamline fashion avoiding unnecessary data replication. Separating analytic data into its own data layer provides the foundation for a self-service analytic environment. To ensure consistency, accuracy, and completeness of the analytic data layer, all analytic data via dimensions and facts will be managed using the DW bus matrix. Employing this concept will eliminate or minimize the need to duplicate business descriptors (*i.e.*, using conformed dimensions) and measures (*i.e.*, fact tables), so answers will be consistent across data marts.

**Justification** – this architecture sets the stage for a foundation of containing a single source of operational and business analytic data reducing the overall maintenance costs and increasing both data integrity and quality. This form of database streamlining becomes the data foundation of achieving a self-service reporting environment.

**Implications** – not properly separating concerns between data acquisition, data integration, and data publication can create significant additional complexity, particularly in the cost of maintenance, which can severely reduce the operational lifespan of your warehouse environment in favor of a replacement solution.

**Exceptions** – a 3<sup>rd</sup> party software package may force a need to replicate data into a separate database and require different data quality controls.

### 6. Acquire data from the true data source or system of record

In selecting the data sources that feed into the DW, require the system of record (*i.e.*, the system that creates and maintains the data). Avoid using data that is stored in one-off databases, such as a replicated database, as a convenience. Sourcing data from a system of record permits a single ETL process maintaining consistent and accurate integration and transformation rules as well as minimizing data trust issues that often times surface when dealing with multiple databases.

**Justification** – the true data sources of an Integrated Data Store, IDS (also called EDW), are the transactional systems – e.g., claims adjudication (for healthcare) or an ERP system. The true data source for the Analytic Data Mart (ADM) layer is the EDW or IDS (2<sup>nd</sup> tier of the architecture).

**Implications** – a significant concern is that the chances of getting a different answer to the same question using one-off data sources is high and the overall trust of the data quality decreases. Another concern is that a non-authentic source could arbitrarily change its data support responsibilities without establishing proper notification agreements to the DW environment.

**Exceptions** – using an alternative data source such as an ODS when the true source is not accessible or when no other data store exists – for example, company descriptors that are not maintained by a system, therefore, electing to use your metadata repository.

### 7. Leverage the corporate investment and functionality of strategic tools

To make most of the corporate investment made in technologies, the BI Environment needs to leverage the functionality and capabilities of the selected strategic tools. The primary tools areas of concern are: DBMS, ETL, Reporting, and Metadata Repository. When strategic tools are available, avoid using secondary solutions even if they appear to be a faster method for a specific project. Over time, this creates both training and maintenance issues as well as questions the Enterprise's investment decision.

**Justification** – adhering to sanctioned strategic tools and the approved features of those tools will leverage the organizations investment and keep future costs down.

**Implications** – Using non-sanctioned tools may create an additional burden to resource training and on-going maintenance as well as further costs via maintenance and support fees.

**Exceptions** – a 3<sup>rd</sup> party software package may force a need to utilize a non-standard tool and database software.

### 8. Apply appropriate data reconciliation and data quality to the data architecture

Apply appropriate data reconciliation and data quality processes to the ETL data architecture. For each data extraction development effort (i.e., source to target), there needs to be consideration to reconciliation and validation of the data being loaded into the DW and subsequent data marts. Incorporating basic row counting (source and target tables) as well as comparing amount totals to trusted sources are key reconciling processes.

**Justification** – ensuring high data quality through a data reconciliation rigor will satisfy a key business requirement of high data quality and allow business users across the organization to use and trust the data.

**Implications** – not having reconciliation checkpoints may allow either bad or incomplete data to be loaded into your data warehouse where data issues are not uncovered until weeks or months later. This may result in reporting incorrect information to both internal and external users. In some cases, the risk is loss of customers or being subject to a fine.

**Exceptions** – if the extraction process is a onetime occurrence and reconciling can be done manually.

### 9. Apply appropriate access control and security considerations to the data architecture

Apply appropriate access control and security considerations to the data architecture. It is important to ensure the right people (internal and external to the organization) have the appropriate access level to corporate data and information. Also, adhering to privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA), specifically Personal Health Information (PHI), requires additional design to the database environment.

**Justification** – ensuring protection of the Enterprise's information assets and adherence to external controls such as HIPAA and PHI as well as the Sarbanes–Oxley Act.

**Implications** – violation to HIPPA (PHI) and Sarbanes–Oxley (if applies) regulations may result in fines and government sanctions. Allowing unauthorized people access to corporate data may have negative results.

**Exceptions** – if the extraction process is a onetime occurrence and access control can be done manually.

### 10. Keep performance in mind when designing solutions rather than waiting

Keep performance in mind when designing solutions rather than addressing it as an afterthought. This would include high volume data and extraction methods (e.g., refresh vs. incremental load) as well as data design techniques. For example, anticipating and designing an aggregation (summary) layer of a high volume data mart may allow quick access to reporting analytic metrics avoiding significant response delays.

**Justification** – avoiding performance issues as well as increasing scalability and improving system resource utilization.



**Implications** – may require additional storage and processes, such as establishing a performance test environment.

**Exceptions** – low volume data.

## Conclusion

The purpose of “*Examining Top Performing BI Best Practices*” is to highlight the significance or resolve of each **best practice** providing a summary outline as an important factor to both Data Warehouse (DW) and Business Intelligence (BI) development. It is important to integrate each **best practice** into the appropriate methodology, ensuring that everyone follows a consistent and thoughtful process. This enhances the value add proposition of **best practices** that contributes to achieving the highest level of implementing business capabilities. Subsequently, your investment of maintaining and improving **best practices** will pay dividends in delivering timely BI solutions.

---

### About the Author:

**Gary Hanson** is the Owner and a Principal Consultant at Analytic Data Solutions, LLC. With over 30 years in the Information Technology field and 15 years of practicing in the data warehouse and business intelligence field, Gary has extensive hands on experience in information architecture, data integration, data management, data modeling, and business intelligence. Gary is the co-author of Quick Start Analytics for Healthcare and holds a B.A. degree in business management from Metropolitan State University

### About Analytic Data Solutions, LLC:

**Analytic Data Solutions** is a BI-data solutions consulting company with a strong business orientation specializing in creating a comprehensive enterprise analytic environment strategic roadmap or program plan. This roadmap, coupled with the foundational components of the business requirements and the architecture blueprint, provides the means to execute the appropriate sequence of projects achieving timely and business value-add analytic solutions.

To learn more, please visit us at: <http://www.analyticdatasolutions.net>