

Chapter 19

Business Intelligence Architecture in Support of Data Quality

Tom Breur

XLNT Consulting, The Netherlands

ABSTRACT

Business Intelligence (BI) projects that involve substantial data integration have often proven failure-prone and difficult to plan. Data quality issues trigger rework, which makes it difficult to accurately schedule deliverables. Two things can bring improvement. Firstly, one should deliver information products in the smallest possible chunks, but without adding prohibitive overhead for breaking up the work in tiny increments. This will increase the frequency and improve timeliness of feedback on suitability of information products and hence make planning and progress more predictable. Secondly, BI teams need to provide better stewardship when they facilitate discussions between departments whose data cannot easily be integrated. Many so-called data quality errors do not stem from inaccurate source data, but rather from incorrect interpretation of data. This is mostly caused by different interpretation of essentially the same underlying source system facts across departments with misaligned performance objectives. Such problems require prudent stakeholder management and informed negotiations to resolve such differences. In this chapter, the authors suggest an innovation to data warehouse architecture to help accomplish these objectives.

DOI: 10.4018/978-1-4666-4892-0.ch019

INTRODUCTION

Companies launch business intelligence projects in pursuit of their broader information governance objectives. To leverage data assets, often some centralized business intelligence competence center (BICC) is appointed, with broad oversight of diverse groups of information consumers. For reasons of consistency and efficiency, often a centralized repository integrated data gets implemented, a data warehouse. To mitigate confusion and drive unified strategy, BICC's will aspire a "single version of the truth" to ensure all departments sing of the same hymn sheet.

Data governance programs invariably (also) want to ensure high data quality. BI projects are sometimes fraught with data quality issues. This could typically be for one of three reasons:

- The content of source system data that was supplied to the data warehouse was inadequate;
- The source system data were accurate, but the data warehouse transformations turned out to be erroneous;
- The content of source data was accurate, and data warehouse transformations were *technically* correct. However, the content of the report did not meet expectations, and further work in exploring requirements and specifying data interpretation should lead to an improved report.

The source of 'errors' in reporting can typically not immediately be known. This needs to be discovered through interaction with end-users of BI reports. This poses considerable risk to progress of data warehouse projects. To mitigate the risks associated with these various sources of data quality issues we propose choosing a hyper normalized data warehouse architecture that will better support resolving of data quality issues associated with such projects.

Business intelligence (BI) projects are risky. According to analyst firms like for instance Gartner, or the Standish Group reports, these projects have petrifying failure rates. In large part, these risks are caused by data quality problems that complicate data integration. When data from disparate business silos are confronted for the first time in the data warehouse, this often surfaces heretofore-unknown data quality issues.

BI teams merely 'hold' corporate data (as in: provide stewardship), they do not 'own' them. However, under organizational pressure to complete a project, they feel sometimes 'forced' into a role of pseudo ownership of data (and accompanying data quality errors) while they struggle to integrate fundamentally misaligned source data in the data warehouse.

This problem can either be mitigated, or exacerbated, depending on the architecture chosen for the data warehouse. An optimal architecture for data warehousing should allow interpretation of data to be left to their owners. This is *not* a responsibility of the BI team. They can facilitate discussions about its content, and shed light on source system characteristics, but the BI team should never be tasked to 'own' (as in: commit to long-term choices) interpretation of data quality problems. Interpretation of data, and decisions on how to deal with data quality errors, should be the realm of (senior) management in charge of the incumbent business silos that control source systems.

To (better) deal with quality issues that arise in the course of data integration, in this chapter I will suggest using an architecture that provides auditable and (easily) traceable data transformations. During project development, data modeling choices need to be changeable (at reasonable cost). Currently, the dominant data warehouse design paradigm (Kimball bus architecture) falls short on providing these features.

Data warehouse requirements, often carry considerable ambiguity. To cope with this ambiguity

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/business-intelligence-architecture-in-support-of-data-quality/96160

Related Content

Billing and Review Perspectives in Healthcare

Joel Fredrickson (2020). *Theory and Practice of Business Intelligence in Healthcare* (pp. 18-67).
www.irma-international.org/chapter/billing-and-review-perspectives-in-healthcare/243349

A Review of Business Intelligence and Analytics in Small and Medium-Sized Enterprises

Marilex Rea Llave (2019). *International Journal of Business Intelligence Research* (pp. 19-41).
www.irma-international.org/article/a-review-of-business-intelligence-and-analytics-in-small-and-medium-sized-enterprises/219341

Data Mining and Knowledge Discovery

Andi Baritchi (2004). *Business Intelligence in the Digital Economy: Opportunities, Limitations and Risks* (pp. 35-47).
www.irma-international.org/chapter/data-mining-knowledge-discovery/6064

Enterprise Intelligence: A Case Study and the Future of Business Intelligence

Joseph Morabito, Edward A. Stohrand Yegin Genc (2011). *International Journal of Business Intelligence Research* (pp. 1-20).
www.irma-international.org/article/enterprise-intelligence-case-study-future/55585

Machine Learning Approach and Model Performance Evaluation for Tele-Marketing Success Classification

Fatma Önay Koçoluand akir Esnaf (2022). *International Journal of Business Analytics* (pp. 1-18).
www.irma-international.org/article/machine-learning-approach-and-model-performance-evaluation-for-tele-marketing-success-classification/298014