Chapter 1

Introduction

1.1 In This Chapter

- Data Warehousing and Business Intelligence: What, Why, How, When, When Not?
- Open Source DW and BI: Much Ado about Everything DW and BI, When Not, and Why So Much Ado?

This chapter details the foundations and frameworks of an open source–enabled EDW/BI solution from concepts to customization, the
best-fit pragmatics in terms of contextual relevance and usability for implementing such a solution in the real world, and how the solution can elevate
the contextual customer to an intelligent customer. From its early adoption
to going mainstream, the journey of the open source Oracle has not only
helped businesses as a cost-container and information-to-application integrator, but also enabled using such methodology as an innovative business
model not just for delivery and deployment, but also as a singular elastic,
embeddable, and executable “architecture-as-a-serviceable-process-model”
that is efficient, effective, and uniquely exhaustive for anything and everything that’s fit to be termed business. When it comes to EDW/BI, this
translates to all data that’s knit to such a business, from precipitation to pervasive-in-action. Leveraging open source as an AaaS, as opposed to an
architecture-as-a-design aspect, enables business-to-business engagement
across the dimensions of information, intellect, and intelligence at any
scale whatsoever, but with a near-zero solution footprint. Thus, an open
source enabled information solution by way of an open source AaaS model
can be standardized as a “rules-of-engagement” blueprint as well as a “rules-
of-management” blueprint—all in a unified near-zero footprint. After all,
the basics of BI—business intelligence is for implementable business analytics, and business success is empirically customer success—demand the
same, and open source EDW/BI is as imperative as getting the implementable and empirical.

This is the key selection indicator in using open source techniques, technologies, and tools for an EDW/BI solution. It begins by opening up the high-fives of DW/BI in terms of What, Why, How, When, and When Not To—identifying the key aspects of each of these for the solution orientation spotlights three primary key performance indicators (KPIs) for an open source–enabled and –enabling EDW/BI pragmatics. It then dives into the Open Source arena of EDW/BI to explain each one of these three KPIs categorized across four major dimensions: the business landscape, the technology landscape, the programmer-to-implementer landscape, and the social landscape (which incorporates the evolving customer experience landscape, also known as the customer–user experience [C-UX] lifecycle, as an exclusive imperative in overall solution value).

1.2 Data Warehousing and Business Intelligence: What, Why, How, When, When Not?

Let’s revisit this callout from the Introduction:

From the combination of a people–processes–perceptions–places perspective; the evolution of business (operations) lifecycle to IT lifecycle to business–IT (solution) lifecycle to business–social lifecycle; and the customer–customization–customer experience lifecycle comes a distinctly differentiating denominator and a universally enabling factor: An open source–enabled solution is both a barometer and a benchmark-enabling bar for the end-to-end EDW/BI Oracle—from data to data-in-action, serving the business customer to the intelligent customer, and breaking through the barriers of heterogeneous data sources, time zones, platforms, technologies, methodologies, and most importantly customers with 360-degree variance in requirements-to-results!

To put these pragmatics into practice in the best possible manner to result in a “best-case” solution, the methodologies of data warehousing and BI optimally help draw the fine line between information centralization, consolidation, and decentralization—just that, nothing but that—from data-on-board to dashboard and everything in between. If one takes a snapshot of the current and future path of the information highway, from
production to provisioning to processing to protection to prediction and security, the preview shows a “data big bang” view—one that’s zoomed in and growing at lightning-fast speed. To make all of this data “work” in any and all desired fashions, it takes more than just a powerful data processing and analysis engine or tool—it requires a solution of the order of magnitude of \( n \) power-centric workhorses of data-to-information-to-knowledge-insight-full decisions enabling information, or otherwise—what can be termed an “intelligent information” churner, anytime, anywhere, and by anyone. Based on the pragmatics and practices involved in implementing such a solution, I can identify three primary business–technology–customer–social headliners that can be standardized as a necessary KPIs for the high-fives of DW and BI:

- **Taking IT Intelligence to Its Apex**: an innovative information model by design for a DBMS-based EDW/BI solution
- **Taking Business Intelligence to Its Apex**: intelligent content for insightful intent—the ability to derive intelligent decisions from the information-in-sight, i.e., information visualization for actionable business insight
- **Business** as the key driver of such a solution, as opposed to IT from its inception to implementation and beyond—Self-serviceability by way of context-aware self-adaptability in terms of its relevance, importance, and significance for continuous business efficiency and effectiveness.

The first two of these serve as the necessary enablers in realizing the third. And an enterprise-wide data warehouse solution complemented by a BI solution results in a complete solution that can deliver results, to the point of a completely satisfied customer experience.

### 1.2.1 Taking IT Intelligence to Its Apex

Business–IT efficiency is an incessant IT evolution that resembles \( n \)-dimensional phenomena—it is constantly changing across a multitude of imperatives, most essentially across the business, technology, time, location, cost, and user experience (UX) dimensions.

- No more one-size-fits-all, as there is more than one “all”
- Best-fit is the new best
- Right time is the new real-time
Context-specific is the new content-specific, and customer-centric can be one or a combination of any business-process touch-point.

Business context, not the prevalent technology, drives the solution architecture—hence, a “best-fit” solution is one that can mix, match, or merge any technology, methodology, or tool to get the optimal solution. The implication is that results are the only reality—everything else is virtual reality!

An enterprise data center can be the foundation for lightning-fast responses to market changes—but only if it can take full advantage of recent developments in servers, networking, virtualization, and cloud-computing strategies.

An enterprise solution can be an intelligent one by being context-aware, self-adaptive, and self-service-able; this can be achieved via an architectural framework that leverages the latest-and-greatest technologies, methodologies, and tools that help enable the transformation of analysis (i.e., insight from business domain, user/customer experience) into analytics, be it advanced or predictive; a framework that facilitates extreme interactivity by way of rich visualization, collaboration, and dynamic on-demand super- and sub-componentization (e.g., live on-the-fly streaming of a super- or subset of the entire end-to-end solution). The key ones in this list of technologies, methodologies, and tools are:

- Business continuity and operational efficiency end-to-end—from the desktop to the data center to the access touch point (continuous operational BI)
- Enterprise information integration (EII), enterprise application integration (EAI), e-solution-as-a-service integration on the fly—using real-time change data capture (CDC), data, and data integration services (via Web services) and service-oriented architecture (SOA)—based servicing to make it software-as-a-service (SaaS)-enabled and -enabling
- Combination of virtualization, clustering, and hybrid cloud computing for resource consolidation, compute collaboration, and business–IT SaaS
This can serve as a viable business–IT process framework for a foundational architecture for an intelligent EDW-BI solution, one that

- Delivers information intelligence using information technology: anyone/anymime/anywhere and getting the right information for the right purpose to the right people at the right time; fast, actionable, synchronized, and tested (FAST) is sharable, though distributed, and serviceable (on-demand or otherwise)
- Is reusable, replicable, retainable (archival deduplication, columnar/hybrid-compression), refactorable, and reformable (even if it is time variant and location variant [geospatial])
- Has the ability to handle any and all types of information, from legacy to legendary (i.e., data that’s insight-rich; this needn’t necessarily be legacy data); from persistent to consistent; from just-in-time to any-point-in-time; from near-real-time to right-time; from big data to better data; from content-aware to context-aware; from single source to multisource to open source(d) (i.e., extreme flexibility in terms of elasticity and hot-pluggability of data sources, application sources, and other embeddable services on-the-fly)
- Is actionable from data to dashboard; from data-in-motion to dashboards-in-mobilization; from data, data everywhere to data centers everywhere

And this intelligent IT results in the UX “where’s all that (huge chunks of) data gone?” for “I can only see the information that I need, and nothing else,” taking IT intelligence to its apex!

1.3 Open Source DW and BI: Much Ado about Anything-to-Everything DW and BI, When Not, and Why So Much Ado?

As the design of solution architectures for Very Large Databases (VLDBs) or otherwise-sized databases continually evolves, from the primary database being a relational database management system (RDBMS) to a no-database or rather an invisible-database environment, the use of innovative technologies and methodologies such as pipelining, parallelization, data grids, database virtualization, application grids, virtual federated views, columnar orientation at the core data model level, columnar compression, isolation, and interoperability of the “application” from the database layer thereby enable linear scalability (scaling-out), vertical scalability (scaling-in), cloud
2.3 What’s In, What’s Not: Available and Viable Options for Development and Deployment 19
  2.3.1 Semantic Analytics 19
  2.3.2 Testing for Optimizing Quality and Automation—Accelerated! 19
  2.3.3 Business Rules, Real-World Perspective, Social Context 19
  2.3.4 Personalization Through Customizable Measures 20
  2.3.5 Leveraging the Cloud for Deployment 21

2.4 The Foundations Underneath: Architecture, Technologies, and Methodologies 21

2.5 Open Source versus Proprietary DW and BI Solutions: Key Differentiators and Integrators 27

2.6 Open Source DW and BI: Uses and Abuses 28
  2.6.1 An Intelligent Query Accelerator Using an Open Cache In, Cache Out Design 30

2.7 Summary 31

Chapter 3 Open Source DW & BI: Successful Players and Products 33

3.1 In This Chapter 33

3.2 Open Source Data Warehousing and Business Intelligence Technology 35
  3.2.1 Licensing Models Followed 35
  3.2.2 Community versus Commercial Open Source 37

3.3 The Primary Vendors: Inventors and Presenters 38
  3.3.1 Oracle: MySQL Vendor 38
  3.3.2 PostgreSQL Vendor 39
  3.3.3 Infobright 41
  3.3.4 Pentaho: Mondrian Vendor 41
  3.3.5 Jedox: Palo Vendor 42
  3.3.6 EnterpriseDB Vendor 42
  3.3.7 Dynamo BI and Eigenbase: LucidDB Vendor 42
  3.3.8 GreenPlum Vendor 43
  3.3.9 Hadoop Project 43
  3.3.10 HadoopDB 44
  3.3.11 Talend 44
3.4 The Primary Products and Tools Set: Inclusions and Exclusions
   3.4.1 Open Source Databases 45
   3.4.2 Open Source Data Integration 65
   3.4.3 Open Source Business Intelligence 70
   3.4.4 Open Source Business Analytics 81

3.5 The Primary Users: User, End-User, Customer and Intelligent Customer
   3.5.1 MySQL 89
   3.5.2 PostgreSQL 89
   3.5.3 Mondrian Customers 91
   3.5.4 Palo Customers 91
   3.5.5 EnterpriseDB Customers 91
   3.5.6 LucidDB Customers 91
   3.5.7 Greenplum Customers 92
   3.5.8 Talend Customers 92

3.6 Summary 92
3.7 References 93

Chapter 4 Analysis, Evaluation and Selection

4.1 In This Chapter 99
4.2 Essential Criteria for Requirements Analysis of an Open Source DW and BI solution 100
4.3 Key and Critical Deciding Factors in Selecting a Solution
   4.3.1 The Selection-Action Preview 103
   4.3.2 Raising your BIQ: Five Things Your Company Can Do Now 107
4.4 Evaluation Criteria for Choosing a Vendor-Specific Platform and Solution 110
4.5 The Final Pick: An Information-Driven, Customer-Centric Solution, and a Best-of-Breed Product/Platform and Solution Convergence Key Indicator Checklist 115
4.6 Summary 116
4.7 References 118

Chapter 5 Design and Architecture: Technologies and Methodologies by Dissection

5.1 In This Chapter 119
5.2 The Primary Aspects of DW and BI from a Usability Perspective: Strategic BI, Pervasive BI, Operational BI, and BI On-Demand 120
5.3 Design and Architecture Considerations for the Primary BI Perspectives 121
  5.3.1 The Case for Architecture as a Precedence Factor 122
5.4 Information-Centric, Business-Centric, and Customer-Centric Architecture: A Three-in-One Convergence, for Better or Worse 123
5.5 Open Source DW and BI Architecture 125
  5.5.1 Pragmatics and Design Patterns 126
  5.5.2 Components 127
5.6 Why and How an Open Source Architecture Delivers a Better Enterprisewide Solution 128
5.7 Open Source Data Architecture: Under the Hood 131
5.8 Open Source Data Warehouse Architecture: Under the Hood 133
5.9 Open Source BI Architecture: Under the Hood 136
5.10 The Vendor/Platform/Product/Tools That Fit into the Open DW and BI Architecture 139
  5.10.1 Information Integration, Usability and Management (Across Data Sources, Applications and Business Domains) 141
  5.10.2 EDW: Models to Management 143
  5.10.3 BI: Models to Interaction to Management to Strategic Business Decision Support (via Analytics and Visualization) 144
5.11 Best Practices: Use and Reuse 146
5.12 Summary 147

Chapter 6 Operational BI and Open Source 149

6.1 In This Chapter 149
6.2 Why a Separate Chapter on Operational BI and Open Source? 150
6.3 Operational BI by Dissection 151
6.4 Design and Architecture Considerations for Operational BI 156
6.5 Operational BI Data Architecture: Under the Hood 157
Chapter 6: Operational Business Intelligence

6.6 A Reusable Information Integration Model: From Real-Time to Right Time 160

6.7 Operational BI Architecture: Under the Hood 161

6.8 Fitting Open Source Vendor/Platform Product(s)/Tools(s) into the Operational BI Architecture 164
   6.8.1 Talend Data Integration 164
   6.8.2 expressor 3.0 Community Edition 164
   6.8.3 Advanced Analytics Engines for Operational BI 165
   6.8.4 Astera’s Centerprise Data Integration Platform 165
   6.8.5 Actuate BIRT BI Platform 165
   6.8.6 JasperSoft Enterprise 166
   6.8.7 Pentaho Enterprise BI Suite 166
   6.8.8 KNIME (Konstanz Information Miner) 167
   6.8.9 Pervasive DataRush 167
   6.8.10 Pervasive DataCloud2 167

6.9 Best Practices: Use and Reuse 167

6.10 Summary 169

Chapter 7: Development and Deployment

7.1 In this Chapter 171

7.2 Introduction 171

7.3 Development Options, Dissected 172

7.4 Deployment Options, Dissected 179

7.5 Integration Options, Dissected 182

7.6 Multiple Sources, Multiple Dimensions 185

7.7 DW and BI Usability and Deployment: Best Solution versus Best-Fit Solution 186

7.8 Leveraging the Best-Fit Solution: Primary Considerations 188

7.9 Better, Faster, Easier as the Hitchhiker's Rule 189
   7.9.1 Dynamism and Flash—Real Output in Real Time in the Real World 190
   7.9.2 Interactivity 190

7.10 Better Responsiveness, User Adoptability, and Transparency 191

7.11 Fitting the Vendor/Platform Product(s)/Tools(s): A Development and Deployment Standpoint 195

7.12 Best Practices: Use and Reuse 202

7.13 Summary 204
9.4 Accelerating Application Development: Choice, Design, and Suitability Aspects
  9.4.1 Visualization of Content: For Better or Best Fit
  9.4.2 Best Practices for Autogenerating Code: A Codeless Alternative to Information Presentation
  9.4.3 Automating Querying: Why and When
  9.4.4 How Fine Is Fine-Grained? Drawing the Line between Representation of Data at the Lowest Level and a Best-Fit Metadata Design and Presentation

9.5 Best Practices for Application Integrity
  9.5.1 Sharing Data between EDW and the BI Tiers: Isolation or a Tightrope Methodology
  9.5.2 Breakthrough BI: Self-Serviceable BI via a Self-Adaptable Solution
  9.5.3 Data-In, Data-Out Considerations: Data-to-Information I/O
  9.5.4 Security Inside and Outside Enterprise Parameters: Best Practices for Security beyond User Authentication

9.6 Best Practices for Intra- and Interapplication Integration and Interaction
  9.6.1 Continuous Activity Monitoring and Event Processing
  9.6.2 Best Practices to Leverage Cloud-Based Methodologies

9.7 Best Practices for Creative BI Reporting

9.8 Summary

Chapter 10 Best Practices Beyond Reporting: Driving Business Value

10.1 In This Chapter
10.2 Introduction
10.3 Advanced Analytics: The Foundation for a Beyond-Reporting Approach (Dynamic KPI, Scorecards, Dynamic Dashboarding, and Adaptive Analytics)
10.4 Large Scale Analytics: Business-centric and Technology-centric Requirements and Solution Options
  10.4.1 Business-centric Requirements 310
  10.4.2 Technology-centric Requirements 313
10.5 Accelerating Business Analytics: What to Look for, Look at, and Look Beyond 320
10.6 Delivering Information on Demand and Thereby Performance on Demand
  10.6.1 Design Pragmatics 326
  10.6.2 Demo Pragmatics 328
10.7 Summary 329

Chapter 11 EDW/BI Development Frameworks 331
  11.1 In This Chapter 331
  11.2 Introduction 332
  11.3 From the Big Bang to the Big Data Bang: The Past, Present, and Future 332
  11.4 A Framework for BI Beyond Intelligence
    11.4.1 Raising the Bar on BI Using Embeddable BI and BI in the Cloud 335
    11.4.2 Raising the Bar on BI: Good to Great to Intelligent 335
    11.4.3 Raising the Bar on the Social Intelligence Quotient (SIQ) 338
    11.4.4 Raising the Bar on BI by Mobilizing BI: BI on the Go 341
  11.5 A Pragmatic Framework for a Customer-Centric EDW/BI Solution 343
  11.6 A Next-Generation BI Framework
    11.6.1 Taking EDW/BI to the Next Level: An Open Source Model for EDW/BI–EPM 352
    11.6.2 Open Source Model for an Open Source DW–BI/EPM Solution Delivering Business Value 353
    11.6.3 Open Source Architectural Framework for a Best-Fit Open Source BI/EPM Solution 355
    11.6.4 Value Proposition 356
    11.6.5 The Road Ahead . . . 357
11.7 A BI Framework for a Reusable Predictive Analytics Model 357
11.9 Summary 360

Chapter 12  Best Practices for Optimization 363
12.1 In This Chapter 363
12.2 Accelerating Application Testing: Choice, Design, and Suitability 364
12.3 Best Practices for Performance Testing: Online and On Demand Scenarios 366
12.4 A Fine Tuning Framework for Optimality 369
12.5 Looking Down the Customer Experience Trail, Leaving the Customer Alone: Customer Feedback Management (CFM)-Driven and APM-Oriented Tuning 372
12.6 Codeful and Codeless Design Patterns for Business-Savvy and IT-Friendly QOS Measurements and In-Depth Impact Analysis 373
12.7 Summary 375

Chapter 13  Open Standards for Open Source: An EDW/BI Outlook 377
13.1 Introduction 377
13.2 Summary 384
13.3 References 385

Index 387
Foreword

It was nearly three decades ago when a young precocious computer geek serving the U.S. Air Force and National Security Agency (NSA) was first trained to process and analyze intelligence. Armed with what was then the latest technology and tools, I was catapulted into the world of military intelligence. I diligently pursued this field for nearly a decade and after serving under President Reagan and President Bush (the 1st one that is) and the Gulf War (again, that is, the 1st one) I was ready for something new in the “outside” world. But as Life would have it, I found myself gravitating toward what I knew best – perhaps summed up as Software, Technology, and Data—gained from all my years in military intelligence. This began my career of Business Intelligence (BI) which greatly involved all three components. After building more than a dozen BI solutions from the ground up and consulting to nearly fifty of the world’s most prestigious and prolific organizations on the matter, I can now surely say that that young precocious computer geek has turned into an old precocious computer geek.

With the same love of the craft and a passion for intelligence and analytics, I traded the world of military intelligence for business intelligence but my end mission and goal has basically remained the same now as it was then—that is, to get the right information to the right person at the right time.

It sounds simple enough when put in these very generic points but as every experienced BI architect or developer would attest, this is absolutely easier said than done. Thus we, as BI practitioners, have all tried to deal with the current technologies and tools in hope that when properly applied, it would lead to a successful delivery of a BI solution. Open Source BI is among those that have contributed to the plethora of offerings in the current BI landscape today. With the demand for better BI and analytics, it will continue to be in the forefront of IT for years to come as different technologies, tools, and strategies are explored.
Having been a long-time practitioner with a multitude of implementations in most of the major BI software packages I tended to stay close to my expertise of the “big players”. So when Mr. Bulusu approached me about the subject of Open Source BI I have to admit that I was quite surprised that he would even consider it knowing that he too had worked with the big vendors, and in fact, had authored a handful of books on using their technologies and tools. Surely, I thought, he would share an elitist view similar to the one that I harbored. It is an elitist view that in fact came to surface recently just last year when I was asked to speak on providing a small budget solution for DW and BI. For this, I admittedly shied away from even addressing at length an open source package and chose rather to address it with other alternatives.

Nevertheless, I continued this discussion with Mr. Bulusu; however, for me, his book has helped to legitimize and validate the use of Open Source BI. It just might also be the catalyst needed to help bring the subject into serious consideration for other experienced practitioners as well. As a practitioner himself, Mr. Bulusu is able to pinpoint the more critical aspects for consideration that a BI expert would truly appreciate. He does a great job of covering the subject thoroughly by first working on the high-level view of BI and then breaking down the components to the most granular detail applying Open Source BI’s feasibility and validity as a viable solution.

As a leader of a consulting firm dedicated and focused on the BI space and touting being tool-agnostic, I now have it ready as an offering and will indeed utilize and apply Mr. Bulusu’s writings when asked for an alternative. In other words, for me, Open Source BI is now a viable solution in light of the ever-changing corporate world of IT and consequently BI. As Mr. Bulusu so eloquently puts it in succinct bullets:

- No more one-size-fits-all, as there is more than one “all”
- Best-fit is the new best
- Right time is the new real-time

With Mr. Bulusu’s book, I have surprisingly started on a new path of acceptance. I knew that Bulusu’s thorough, meticulous approach and in-depth examination of the subject; taken from a standpoint of an experienced author, educator, and technologist has finally provided justice to this subject.

Thank you Mr. Bulusu. You have opened my eyes to Open Source BI.

Rosendo Abellera
President and CEO, BIS3
Introduction

As I begin the introduction to this book, a plethora of thoughts randomly fill my mind in regard to open source. From the open skies since the beginning of time to the NEW and NEXT-GEN open-source paradigm involving people–processes–pragmatics–practices NOW, open has become infinite—and yet it remains as finite in its ability as getting the globe in the palm of my hand while at the same time enabling me to travel back and forth, in varying degrees of function, form, and fastness. The sometimes-fast, sometimes-just-fast-enough, sometimes-too-fast, and sometimes-faster-than-I-can-think skewing and shearing dynamics of the currently trending and future-enabled open-source technologies are the game changers in information delivery—what I term as the OPEN SOURCE PERFECT: from OPEN-NESS to OPEN-SENSE and BEYOND, yielding a “What-I-get-is-what-I-want” solution, nothing more and nothing less, at any time and at any point in time, thus producing, preserving, and providing currency of the desired information.

The key benefits of such a solution from a customer-centric point of view are as follows:

- Anytime, anywhere, by anyone (A-A-A) accessibility and availability, all authentic and authorized
- The right information for the right purpose to the right people at the right time (R-R-R-R)
- Uncompromising on the business basics of affordability, flexibility, reliability, stability, scalability, and efficiency
- Business assurance in terms of a single frame of reference for operational continuity, agility, and customer assurance
An open source solution is the secret sauce behind the appraisal, analyses, architecture, adoption, and adaption of such a solution, one that is at least sufficiently best-fit for meeting the demanding needs of current-day superior customer experience in the unpredictably complex web of business revolution and competitiveness. The key differentiator is the loosely coupled architecture that open source methodologies support, which can be easily and transparently extended to support a loosely coupled business model if needed.

Open source gains precedence when it comes to the simplicity and convenience involved in terms of being a unified and universally available, flexible, and scalable on-demand solution that is at the same time compliance-ready, implementation-ready, deployment-ready, and customer-awareness-ready. The business performance indicators from such a solution are as clearly visible from the preview of the solution implementation as the need for such indicators to drive business performance:

- Content and context convergence on instance, constancy, and consistency on-premise to online to on-the-go (mobile) to out-of-premise (in the cloud)
- Dynamic analytics by way of metrics generation through automation
- As open source moves mainstream, business moves to business-on-the-go (or business-on-the-fly, to be precise). This seemingly slight shift in the business model has enormous potential to transform the business–IT–customer–social landscape of such a solution point north of the center so that:
  - The business accelerates in terms of improvement to involvement.
  - The business–IT forecast looks all CLOUD–y (irrespective of climate change!), starting from cloud bursting to crowd sourcing.
  - The business–social relationship gets wiser than business-as-usual, as data becomes intelligent to information-in-action.
  - The business network communicates and collaborates to become business competition, from inside-the-net in the real world through the Internet to the social net, resulting in competitive business processes.
  - Business services (including business solutions and support and operations solutions and support) evolve into business–social
services, thereby granting them tenure—and business assurance—through the customer lifecycle of the business.

- Businesses can leverage the so-called “gold” mines of intelligent information (as more and more data mining becomes practical) in unimaginably fast, actionable, synchronized, transparent and social ways, wherein users across all touchpoints can become involved interactively in a 360-degree variance to feed and feedback their (business) input. This self-serviceable input can then be auto-rewired as “intelligent business rules” via the tenor of self-adaptability of the open source BI solution in near real time.

With all these pros come the cons, too. Business security tends to become increasingly insecure. The streams of “intelligent information” transcend the enterprise boundaries into multiple endpoints and clouds, in all their sensitivity and identity, and converge into the social enterprise; their security gets social too—and ensuring the social security of the same demands opens multiple channels for what is termed social engineering, from ethical hackers to ethical blockers and beyond—wherein the acts of hacking, phishing, VISHING, and threats become alarmingly ethical!

An open source solution can drastically minimize the impact of the risk while it also improves the remediation of the same through its ability to automate, accelerate, and reduce the time to recovery in easy, elastic, duplication-less, and dynamic architectural virtualization and execution capabilities. The faster the time to insight, the faster the time to recovery—and the lesser the risk of disaster and the minimal the time spent in disaster recovery mode.

From the combination of a people–processes–perceptions–places perspective; the evolution of business (operations) lifecycle to IT lifecycle to business–IT (solution) lifecycle to business–social lifecycle; and the customer–customization–customer experience lifecycle comes a distinctly differentiating denominator and a universally enabling factor: An open source–enabled solution is both a barometer and a benchmark-enabling bar for the end-to-end EDW/BI Oracle—from data to data-in-action, serving the business customer to the intelligent customer, and breaking through the barriers of heterogeneous data sources, time zones, platforms, technologies, methodologies, and most importantly customers with 360-degree variance in requirements-to-results!
What Does This Book Cover?

This book highlights the practical aspects of using open source data warehousing (DW) and business intelligence (BI) technologies from an end-to-end lifecycle perspective, starting from concepts to implementation and customization. It follows a lively "when-to," "why-to," and "how-to" approach to explain the key differentiators and benefits of using open source from proprietary DW and BI technologies in the database and data warehouse design, real-time ETL and data integration, and presentation services and real-time reporting phases of the end-to-end DW and BI solution. These differentiators and benefits are explained in terms of very large databases (VLDBs), scalability, high performance, stability, endurance, and ease of use that save time, effort, cost, resources, and support, and maximize return on investment (ROI), whether the solution is a totally new implementation, coexistence, codeployment, or replacement. The successful players and their products in the open source DW and BI field are taken into account while describing these details.

It concentrates on the more important pragmatic aspects involved by describing the deciding factors, key design criteria and methodologies, expert techniques, and best practices by taking into account cross-functional and multi-integration scenarios.

Who Should Read This Book?

This book is targeted toward multiple audiences. First, it is meant for the team involved in deciding, choosing, architecting, developing and implementing an open source–based DW and BI solution, whether it be the CIO looking for a maximum ROI strategy, the CTO exploring the best-fit in terms of a better-faster-simpler solution, or the business analysts/solution architects trying to figure out the optimal architecture for an implement once/use forever customer-specific solution, or the design and development staff looking for a one-stop resource to uncover the under-the-hood principles, practices, and development techniques involved, or the integration experts planning an enterprise-wide one-fits-all integration solution to co-deploy with existing BI systems. All it assumes is a working knowledge of databases, data warehouses, and business intelligence.

Secondly, this book gives the foundational paradigms and patterns for research-in-action that can be followed by engineering and research and development (R&D) teams for innovation and invention of structurally unique business/solution/social/analytical models.
Additionally, this book is geared toward students, learners, and faculty in the academic field who are devoted to a course in open source data warehousing and business intelligence, who would find the book handy in problem–solution and best practices–techniques scenarios in terms of theory, pragmatics, and practice.

Why a Separate Book?

This book is unique in today’s constantly changing and evolving market, in that it serves as a single resource for comprehensive details of an open source-based DW and BI solution that is business-centric, cross-customer-viable, cross-functional, cross-technology-based and enterprise-wide, focusing on the entire lifecycle of an open source DW and BI implementation from the concepts involved all the way through to customization.

It describes the practical aspects of using an open source approach to enterprise data warehousing (EDW) and BI by way of the technologies, tools, and methodologies involved and the better business benefit obtained thereby in terms of a business-centric, cross-functional, cross-customer, cross-technology-based solution that not only enables a better-faster-easier solution model, but also gives context-aware adaptability to changing business needs, unprecedented flexibility by way of usability and deployment—on-demand by leveraging existing resources and five-nine reliability—and is pervasive, from the legacy enterprise to the current and next-gen social enterprise.

The key standout indicator of this book is its treatment of how powerful and performing open source can be: as an enabler and benefactor of enterprise businesses without having to shake the existing business structure, and as a dynamically extensible paradigms-and-patterns expert system in form and format. As it highlights the focal points involved for such a structurally unique and dynamically intelligent business/solution/social/analytical model by way of an open source-enabled intelligence strategy as architecture-as-a-service-by-itself (AaaS) framework, as opposed to architecture-as-a-design-service, the fractal efficiencies of such an implementation grow convincingly closer to real-world customer-centric dynamics, revealed with business examples and use-case scenarios. The fine line here is that the former approach requires the AaaS as a necessary and sufficient component of the intelligence model.
The table below summarizes the primary features and their associated benefits in support of the author’s preview for a separate book for open source DW and BI; this information can at least aid in boosting the potential reader’s “learn-and-leverage” confidence. How well this preview can predict the same in terms of “true” benefits in practice is based on the readers’ and implementers’ confidence level in making a prudent decision of choice for adopting the same in desired contextual business solutions.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Corresponding Benefit</th>
</tr>
</thead>
</table>
| A single resource details the practical aspects of an open source DW and BI architecture from both technological and solution perspectives, while at the same time describing the “best practice” aspects for implementation of a “best-fit” solution. | This is a standout, in that  
- It is the only all-in-one resource across open source DW and BI technologies.  
- It focuses on architecture and design as well as live implementation aspects of the best-fit solution. |
| An exclusive section devoted to best practices from an implementation perspective for customers to adopt an open source DW and BI solution (e.g., “Customization Framework” for a best-fit solution and why “best-fit” wins over the “best” solution). | 1. It gives the reader practical advice from experience, which helps in improving productivity from the application project as well as professional standpoint.  
2. This is also helpful for winning customer confidence and getting sponsorship for an open source DW and BI implementation. Because it explains the best-case scenario to use, it helps in quick decision making for getting a go-ahead.  
3. This is indispensable for solution architecture and academic R&D projects alike, in that it provides an adoptable methodology on an “as-is” as well as on an enhancement/extensible basis. |
| Coverage of mission-critical and state-of-the-art trends from a user and business standpoint (e.g., “BI Beyond Reporting,” “Delivering Information-On-Demand,” and “Achieving Performance-On-Demand”). | These are indispensable for solution architects, business analysts, executives, and academicians alike by providing an adaptable methodology on an “as-is” as well as on an enhancement/extensible basis. |
Covers information previously unavailable or undocumented that is both technically complete and business-centric (e.g., “Accelerating Business Analytics for the Next Generation Customer—Granting A Tenure for the Customer Experience Lifecycle”) makes this book an asset for any targeted audience and a standout from other titles in the market.

The book details the contextual specifics of approaching the “open-source solution” strategy from architecture-as-a-(business process)-service model as opposed to an architecture-as-a-design aspect, giving a preview of a new baseline trend that’s both unique and innovative for a reference model for open source EDW/BI.

The last chapter (“The Prize for the Price: A Win-Win or Not?”) is a real boost for success, both for the book as well as for the business-centric customer. This makes this book a “great” book—a unique standout.

A book for a one-fit-all and implement-once/use-forever solution in the open source DW and BI space, written by a proven author-cum-architect-cum-educator-cum-researcher in the EDW/BI domain—open source and otherwise.

Increases the trust factor not only from the buying side, but also from business decision making and solution adoptability side, as the content is based on the author’s practical experience, which match real-project needs and can be treated as “benchmark” approaches for proven and successful implementations. However, zero-trust and zero-tolerance pragmatics are emphasized for information security, thereby increasing the confidence level for information assurance.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Corresponding Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covers information previously unavailable or undocumented that is both</td>
<td>1. This makes this book an asset for any targeted audience and a standout from other</td>
</tr>
<tr>
<td>technically complete and business-centric (e.g., “Accelerating Business</td>
<td>titles in the market.</td>
</tr>
<tr>
<td>Analytics for the Next Generation Customer—Granting A Tenure for the</td>
<td>2. The book details the contextual specifics of approaching the “open-source solution”</td>
</tr>
<tr>
<td>Customer Experience Lifecycle”)</td>
<td>strategy from architecture-as-a-(business process)-service model as opposed to an</td>
</tr>
<tr>
<td></td>
<td>architecture-as-a-design aspect, giving a preview of a new baseline trend that’s both</td>
</tr>
<tr>
<td></td>
<td>unique and innovative for a reference model for open source EDW/BI.</td>
</tr>
<tr>
<td>The last chapter (“The Prize for the Price: A Win-Win or Not?”)</td>
<td>1. This makes this book a “great” book—a unique standout.</td>
</tr>
<tr>
<td></td>
<td>2. Increases the trust factor not only from the buying side, but also from business</td>
</tr>
<tr>
<td></td>
<td>decision making and solution adoptability side, as the content is based on the author’s</td>
</tr>
<tr>
<td></td>
<td>practical experience, which match real-project needs and can be treated as “benchmark”</td>
</tr>
<tr>
<td></td>
<td>approaches for proven and successful implementations. However, zero-trust and zero-</td>
</tr>
<tr>
<td></td>
<td>tolerance pragmatics are emphasized for information security, thereby increasing the</td>
</tr>
<tr>
<td></td>
<td>confidence level for information assurance.</td>
</tr>
</tbody>
</table>