Craft Your Future State BI Reference Architecture

by Boris Evelson and Noel Yuhanna, November 1, 2012

KEY TAKEAWAYS

BI Is Different From All Other Enterprise Platforms And Applications
While some IT best practices are applicable to BI, many are not. BI is different. Specifications are often outdated by the time they are collected. Business users demand more control; therefore, IT standards often take a back seat. Enterprises that understand and embrace these differences are often the most successful with BI.

The Importance Of Agility Often Outweighs A Single Version Of The Truth In The Modern Enterprise
In today’s complex, ever-changing world, a single version of the truth is relative and contextual. While a single version of the truth is still highly important for many mission-critical applications, in numerous cases, it’s the lack of agility and inability to react to a new requirement in time that may kill otherwise successful BI initiatives.

Picking The Best BI Tool For Each Job Requires A Flexible Approach To IT Standards
An overinsistence on enterprisewide standards and a single version of the truth has traditionally ruled BI from the top down. But a purely standards-based approach to addressing most current business requirements is neither flexible nor agile enough to react and adapt to ever-changing information requirements. “Just get it done” is BI’s new mantra.
Craft Your Future State BI Reference Architecture
Road Map: The Business Intelligence Playbook
by Boris Evelson and Noel Yuhanna
with Stephen Powers, Brian Hopkins, Gene Leganza, Michele Goetz, Leslie Owens, Paul D. Hamerman, Henry Peyret, and Shannon Coyne

WHY READ THIS REPORT
In the face of rising data volume and complexity and increased need for self-service, enterprises need an effective business intelligence (BI) reference architecture to utilize BI as a key corporate asset for competitive differentiation. BI stakeholders — such as project managers, developers, data architects, enterprise architects, database administrators, and data quality specialists — may find the myriad of choices and constant influx of new business requirements overwhelming. Forrester’s BI reference architecture provides a framework with architectural patterns and building blocks to guide these BI stakeholders in managing BI strategy and architecture.

Table Of Contents
2 BI: The Last Milestone Toward Business Users In EIM
16 Forrester’s BI Reference Architecture — Detailed View
17 RECOMMENDATIONS
17 Every Organization Needs A BI Reference Architecture
18 ALTERNATIVE VIEW
18 A Sneak Peek Into An Alternate BI Reference Architecture Future
18 Supplemental Material

Notes & Resources
Forrester discussed data management landscape approaches with numerous vendors and user companies during the past six months to gather information to derive this architecture model. Companies interviewed include Composite Software, Hexaware Technologies, IBM, Infosys, and Microsoft.

Related Research Documents
Drive Business Insight With Effective BI Strategy
April 30, 2012
The Future Of BI
January 10, 2012
BI: THE LAST MILESTONE TOWARD BUSINESS USERS IN EIM

Enterprise information management (EIM) is complex — from a technical, organizational, and operational standpoint. But to business users, all that complexity is behind the scenes. What they need is BI, an interface to enterprise data — whether it’s structured, semistructured, or unstructured. Our June 2011 Global Technology Trends Online Survey showed that BI topped even mobility — the frontrunner in recent years — as the technology most likely to provide business value over the next three years.¹

Forrester often defines BI in one of two ways. Typically, we use the following broad definition:

*Business intelligence is a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision-making.*

From this definition, BI must include technologies such as data integration, data quality, data warehousing, master data management, text and content analytics, and many other topics that the market sometimes lumps into the EIM segment. Therefore, we also refer to data preparation and data usage as two separate but closely linked segments of the BI architectural stack. We define the narrower data usage BI market as:

*A set of methodologies, processes, architectures, and technologies that leverage the output of information management processes for analysis, reporting, performance management, and information delivery.*

BI architecture is never simple, and this is especially true in large, heterogeneous enterprises with global reach and multiple product and service lines. Such enterprises always have more than one enterprise data warehouse (EDW), hundreds of data marts, and several BI platforms. BI reference architecture has to account for agility enablers, such as allowing exceptions to the standards and business user self-service, due to ever-changing business and regulatory requirements. The resulting BI reference architecture may not look pretty and simple, but it’s pragmatic and practical (see Figure 1).
**Data Sources: Where The Data Comes From**

The data sources layer represents the databases, files, streams, and repositories that provide the source material for all information-related activity. These components form the persistence layer that supports all online transaction processing (OLTP) applications. These applications represent the starting point for all data in the information life cycle, whether that data is structured, semistructured, or unstructured and whether it is stored in the cloud, on a desktop, inside the data center, on a smartphone, or on a thumb drive (see Figure 2). Structured data sources include data residing on mainframe, distributed, desktop, and mobile databases as well as relational and nonrelational sources. Unstructured data sources include unformatted streams, sensor data, files, documents, images, binary, video, and audio formats. Semistructured data includes XML, formatted log file and streams, and formatted application data, such as office documents.
Figure 2 Forrester’s BI Reference Architecture — Data Sources Layer

Data Rationalization — Mapping Apples To Oranges

This layer represents the data integration, data quality, data profiling, and data services technologies (see Figure 3). It includes best practices that support the physical and virtual movement of data across the disparate data stores and applications managed within a BI architecture. In addition, this layer applies transformation logic and business rules that standardize, validate, cleanse, aggregate, enrich, match, and profile data as it moves from point of capture to consumption in the data management life cycle. Data rationalization represents a pass-through layer in which data is not persisted (although various staging areas can be used throughout the process) but directly feeds the derived data sources, including data marts (DMs) and data warehouses. Derived data sources can also act as a data source to the data rationalization layer to support additional processing, transformation, cleansing, and integration. This process can be repeated several times in order to enhance data to support complex analytical requirements. A typical set of architectural components includes:

- **Virtualized data access.** Virtualized data access is a services layer that integrates heterogeneous data and content in real time or near-real time. Terms for such layers include information fabric, data services, information-as-a-service, and data federation. These layers provide virtual views that represent information in forms that applications and users need while hiding the true complexity of data. Applications or users access virtualized data through the fabric or layer, enabled by middleware. An information fabric enables sharing of common views of derived data with improved data quality, regardless of the data’s location or its format. Applications may access the layer via structured query language (SQL), XQuery, simple object access protocol (SOAP), and/or representational state transfer (REST) calls.²
- **ETL.** Extract, transform, and load (ETL — where most of the processing occurs in a specialized server) and extract, load, and transform (ELT — where most of the processing occurs in a database management system [DBMS]) tools are used for batch data movement and transformation. Enterprise service bus (ESB) tools provide near-real-time application integration, and enterprise information integration (EII) tools enable the creation of federated or virtualized views of data. Firms leverage data quality and data profiling software to analyze, baseline, prevent, and mitigate data quality issues. A significant difference between ETL and ELT is the order of the transformation process, which is performed either before or after loading data from the staging database. Besides moving data, ETL and ELT support complex transformations such as cleansing, reformatting, aggregating, and converting large volumes of data from many sources. Both of these technologies complement change data capture (CDC) and data replication to support real-time data requirements. In addition, data virtualization delivers real-time integration of heterogeneous sources including structured, unstructured, and semistructured data.

- **Business event processing.** Business event processing applications answer the question, “What activities are happening in the business now that I need to know about?” by searching for patterns and values within several streams of actively flowing data. The streams almost always represent real-world information such as location of transportation vehicles and the goods they carry, stock prices fluctuating in real time, and machine-to-machine (M2M) communication and process control systems. These applications often use a combination of business event processing technologies, such as streaming DBMS, complex event processing (CEP), or business activity monitoring (BAM) technologies.

- **CEP.** A CEP (one of the technologies that support business event processing) platform detects patterns of events (and expected events that didn’t occur) by filtering, correlating, contextualizing, and analyzing data captured from disparate live data sources. CEP recognizes crucial events in the business, makes those events known, and initiates either automated or manual responses. CEP platforms tend to have greater capabilities in handling live data, diverse data, pattern detection, diverse business processes, and volumes of data when compared with other applications with event processing capabilities, such as BAM and business process management (BPM).

- **Text processing/natural language processing (NLP).** This technology converts, segments, tags, and derives meaning from information. Machine learning algorithms and/or linguistically based rules perform NLP. First, information retrieval technology crawls data sources and filters content and metadata from diverse file types. Then, NLP detects language and parts of speech, deduplicates files, extracts text patterns, and recognizes entities (e.g., people, places, things, and events) and the relationships between them. Then, the information
may be coded, labeled, classified, scored on sentiment, and mapped to predefined facets using a lexicon or ontology. This process transforms unstructured source materials into semantically annotated information compliant with standard models for data interchange, such as RDF and XML. Practical applications include mining call center records (text and/or digitized voice notes) for patterns in customer requests or complaints.

- **Data quality software.** Data quality management provides technologies for implementing policies and processes that ensure that enterprise data has relevance, freshness, accuracy, and integrity. The software itself is essentially a fit-for-purpose business rules engine that standardizes, validates, cleanses, aggregates, enriches, matches (e.g., deduplicates), and merges. Stakeholders can apply data quality rules in batch processes or call them as a service within other processes. Many data quality platform vendors also offer integrated data profiling capabilities to measure and monitor data conditions for data stewardship and remediation. Many enterprise-class vendors combine data integration and data quality processes for complete management of data transformations.

- **Master data management (MDM) and reference data.** MDM enables an organization to identify trusted master data and leverage it to improve business processes and decisions. An MDM solution acts as either a virtual or a physical repository to persist, share, and synchronize master views of data and data relationships (such as hierarchies) for any subscribing operational or analytical application. MDM solutions provide the primary data stewardship mechanism to manually administer and mitigate data relationship issues and views. Most MDM solutions target critical data domains such as parties (e.g., customer, supplier, patient, employee, and citizen), products, materials, securities, assets, and locations. MDM tools also support shared reference data used in lists of values or pick lists for subscribing applications. Both MDM and data quality applications can be unidirectional or analytical (data is cleansed and reconciled with MDM only on its way to derived data sources, such as data warehouses — a good place to start) or bidirectional or operational (cleansed and reconciled data is sent back to the data sources as well — usually a more complex and a costly endeavor).

- **Data-governance-enabling tools.** Despite some vendor claims (“data governance in a box!”), data governance is not a technology; it’s a functional business capability, often supported by tools in the data rationalization category. Technology, however, can enable data governance responsibilities and workflow management. Tools that add early value to a data governance initiative include: 1) data profiling software, which provides visibility into the current state and trends of data quality and completeness, and 2) business glossary tools, which help define and capture standard definitions and maintain business rules for the combination and use of shared data. Other policy definition, workflow management, and collaboration tools can also add significant value to operationalize a data governance program. Enterprise-level MDM and data quality solutions also offer collaboration capabilities for establishing and managing business rules and data remediation.
**Direct data source access.** Reporting and analyzing data directly from operational, transactional data sources is an option, and an agile one. It could be used as part of a natural progression from: 1) a quick and dirty win such as pulling reports directly from an ERP application; to 2) putting the data virtualization layer between the report and ERP as the next step toward “productionalizing” that report; and finally, 3) using ETL tools to move data to a data warehouse and updating data virtualization views to point to the data warehouse. It’s a win-win: Business users get a report ASAP, while IT works in the background to move the report to a stable and secure production environment. However, the downside of direct data source access for BI is a lack of a strong data cleansing and data integration process, which can affect data quality.

**Figure 3** Forrester’s BI Reference Architecture — Data Rationalization Layer

---

**Derived Data Sources — A Single EDW May Not Be A Practical Option**

A derived data source serves as an intermediary node in the delivery of intelligence from one or more originating sources to one or more target repositories, applications, or views (see Figure 4). Derived data sources such as data warehouses, operational data stores, data marts, and MDM hubs reconcile, consolidate, synchronize, and persist shared, trusted views of information and make it available to applications via batch, real-time, or virtual access. A typical set of architectural components includes:
- **Staging area.** While there are many definitions and possible use cases for a staging area, most often it’s just something as simple as a physical copy of OLTP tables. It is often used to minimize the load on OLTP apps for data extraction and replicate their tables for further BI processing.

- **Operational data store (ODS).** An ODS combines most data sources and optimizes data mostly for operational, near-real-time reporting. An ODS can serve some other purposes, such as acting as a staging area, consolidation area, or a staging area that supports specific parts of the information journey such as automated or manual data quality processes. It may even serve as a consolidated transactional database for multiple OLTP applications. ODS generally only stores recent history at most.

- **Data warehouse (DW).** In contrast to an ODS, a DW is an integrated analytical data store. It combines most data sources and subject areas across the enterprise and optimizes data mostly for historical reporting. For that purpose, DW data is usually denormalized and organized into star schemas. A DW is the main historical source where organizations perform most time-series analysis.

- **Data marts.** Data marts are a subject-matter-specific data store similar to the DW but serving a single segment, such as finance, HR, sales, or marketing. If a DW contains all details, then data marts may only contain aggregate information, which prevents tasking the DW with processing for aggregate reporting. Companies may physically separate data marts from the DW or may create them as virtual data marts within a DW as a set of subject-specific aggregate tables. You can “spin off” data marts from your DW or ODS or generate them in your ETL/ELT process. In any case, make sure that you only use one set of transformation logic and reconcile with the same reference/master data to generate all of your derived data sources to ensure consistency. If you do, you can call your data marts “conformed.”

- **OLAP cubes.** These typically further specialize data and analysis on top of data warehouses or data marts. For sales data marts, you may want to build a customer profitability cube that performs very specific profitability calculations by allocating revenues and expenses to the individual customer level. Cubes can be physically separate (Multidimensional Online Analytical Processing [MOLAP] as in Hyperion Essbase or IBM Cognos PowerCubes) or virtual within your DW (such as cubing engines in the relational database management system [RDBMS]) or in your BI platform (as in MicroStrategy or Oracle OBIEE ROLAP engines).

It’s important to mention that very few enterprises have a need for and actually deploy all of these five derived data sources components (staging, ODS, DW, data marts, and cubes). Also, the lines of demarcation between where one component ends and the other one starts vary greatly. Rather than seeking academically correct definitions, it’s much more important to reach a consensus in your organization on which types of data (e.g., transactional versus analytical, aggregate versus detail, operational versus historical, strategic versus tactical, etc.) belong to which component. Use
“architecture in zones” to tolerate different architecture patterns for different zone of the business. Then create a set of policies and processes to enforce and monitor compliance with these standards across the enterprise.

Other architectural components and considerations when crafting your BI reference architecture include:

- **BI-specific DBMS.** Mainstream RDBMSes are an awkward fit for BI; they require tuning, customization, and constant optimization. In order to tune such a RDBMS for BI usage, specifically data warehousing, architects usually: 1) denormalize data models to optimize reporting and analysis; 2) build indexes and aggregate tables to optimize queries; and 3) build OLAP cubes to further optimize analytic queries. RDBMS will never be a perfect fit to handle unstructured content and complex data structures with uneven, unbalanced hierarchies. To address these challenges, Forrester recommends supplementing or even replacing a BI DBMS with new technologies including columnar, in-memory, inverted, and associative indexes and graph DBMS (aka Resource Description Framework [RDF], triple stores).

- **Reference data model.** The data model acts as a common communication medium between IT and business. Unlike data models tied to various data sources that serve the underlying transactional or operational applications, BI requires its own reference data model focusing on logical and conceptual data models for analytical purposes. This includes modeling for fact tables, dimension tables, star schema, and temporal data and providing the linkage to various data sources and derived data sources. Industry-specific data models can help jump-start a BI platform, defining logical-to-physical linkage to data sources and modeling the fact, dimension, and multidimensional tables to support vertical industries such as retail, manufacturing, financial services, and telecom.
**Analytical Data Virtualization — Hiding Data Complexities From The Business Users**

Analytical data virtualization, AKA the semantic layer, hides data complexities such as cryptic table and column names from business users and report developers. BI pros use this layer to create user-friendly business definitions of data, subject-matter-specific logical views (which may span several physical tables via joins), standard and popular metrics, and predefined queries (see Figure 5). As most enterprises use more than one BI platform, when architecting this layer, Forrester recommends:

- **Prioritizing and emphasizing common metadata efforts.** Organizations should select a metadata repository (those that come with ETL tools offer some of the best options) as a hub to synchronize metadata among all BI platforms and applications. To ensure that these metadata repositories integrate across competing BI applications, BI pros should look for the OEM standard metadata adapter vendors from Meta Integration Technology or Silwood Technology for ERP packages. They could also adhere to metadata exchange standards such as XML Metadata Interchange (XMI) or Common Warehouse Metamodel (CWM). BI pros should also prioritize key shared entities such as metadata for key reference data, key master data, and industry-standard canonical models.

- **Pushing rules and definitions down a layer for reuse and leverage.** Organizations can use data virtualization technology for a central metadata repository and can push metadata, models, and objects (joins, queries, metrics, and aggregates) into that layer so that BI tools can just plug into it. This eliminates some of the metadata duplication in BI semantic layers.

- **Leveraging the common metadata management layer in the data governance process.** Data quality and MDM is often the best place to start your data governance programs. Metadata management should be a key component to define and govern data quality standards and master data structures.

*Figure 5 Forrester’s BI Reference Architecture — Analytical Data Virtualization Layer*
Data Usage — What Business Users Touch And Feel

The next layer of the BI architecture reference model shows the general usage scenarios that drive information access in an enterprise. We’ve included different types of business use cases for data access and usage that have different implications for the people, process, and technology involved in the usage scenarios. The model shows that some usage scenarios (BI, analytics, and dashboards, for example) typically access derived data sources such as data warehouses, while others (open database connectivity [ODBC/JDBC] data access calls or online transaction processing [OLTP] workloads) typically access original data sources (see Figure 6). When planning your BI architecture, consider the following usage scenarios:

- **Reports and ad hoc queries.** This includes analytical reporting (based on data warehouse/data mart data organized in a star schema including time-period snapshots) and operational reporting (based on operational DBMS organized in the third normal form [3NF] or otherwise optimized for OLTP). Reporting tools often include pixel-perfect positioning of data and graphics, a scripting language equal in power to a full programming language, and the ability to handle complex headers, footers, nested subtotals, and multiple report bands on a single page. Ad hoc query tools provide quick answers to business questions.7

- **OLAP.** Otherwise known as “slicing and dicing” analysis, OLAP tools allow business users to see any facts (numerical, typically additive numbers such as transaction amounts and account balances) almost instantaneously regrouped, re-aggregated, and resorted by any dimension (e.g., descriptive elements such as time, region, organizational unit, or product line).

- **Dashboards.** Dashboards mash up different historical, current, and/or predictive information into one efficient user interface. Typically, dashboards display key performance indicators (KPIs), using visual cues to focus user attention on important conditions, trends, and exceptions.8

- **Exploration and discovery.** In the days before big data, BI requirements often came from business analysts, who based them, among other things, on transactional data. Back then, analysts loaded transactional data into spreadsheets or queried it with SQL as part of the process of figuring out what the final reports and dashboards should look like. But in the world of big data, one can’t just load all relevant transactional data into a spreadsheet (it won’t fit) or query it with SQL (it’ll take too long). So how does one understand what the data looks like and come up with BI requirements? Also, what if the requirement is “I want to see all data cut and cross-referenced in every possible dimension and attribute”? Building such a model using relational or multidimensional technology would be close to impossible. Data exploration and discovery tools allow business users to look at data without a data model. A special case of exploration and discovery is when it is done in memory: In-memory data, the data model, and a report are one and the same. By changing a report, the user is changing the model (AKA micromodeling or in-line modeling) or exploring and discovering previously uncovered insights.9
- **Advanced predictive analytics.** This refers to platforms and applications that support the identification of meaningful patterns and correlations among variables in complex, structured and unstructured, and historical and potential future data sets for the purposes of predicting future events and assessing the attractiveness of various courses of action. Advanced analytics initiatives typically incorporate such functionality as data mining, descriptive modeling, econometrics, forecasting, operations research, optimization, predictive modeling, simulation, statistics, and text analytics.10

- **Process analytics.** Process analytics involve reporting on and analyzing data and state of processes as part of BPM solutions. This includes real-time information about process execution, such as throughput, service-level agreements, and backlog or patterns of incomplete documentation from any business process.11

- **Analytical performance management.** Analytical performance management is a key part of business performance solutions (BPSes), which includes a collection of components to apply strategic frameworks, track business initiatives, set goals and measure progress, develop plans and forecasts, and monitor KPIs. These solutions are based on a BI foundation to support the analytical performance measures and dashboards and also include a set of tools to construct complex planning models, manage collaborative workflows, and track progress against measures and plans.12

---

**Figure 6** Forrester’s BI Reference Architecture — Data Usage Layer

- Reporting and querying
- OLAP
- Visualization dashboards
- Exploration and discovery
- Advanced analytics
- Process analytics
- Performance management

---

**Data Delivery — Where It All Goes**

The final, top layer of the BI reference architecture refers to how BI users and other enterprise applications consume BI output and includes (see Figure 7):

- **BI services.** BI services offer the ability to embed actions, alerts, and reports directly in another application, tool, user interface (UI), or process. BI services hide the complexity of the data sources, data rationalization, and semantic and integration layer, delivering a powerful capability that can be leveraged by any application or business users. They can automatically
refresh the embedded action, alert, or report based on a predefined schedule or delivered real time. With BI services, the embedded BI solution renders its own experience on the target device within a container provided by the host application, delivering a seamless BI experience to the user.

- **Alerts.** Unless a BI environment has built-in alerts, business users may not know what’s happening with data unless they actually look at a report or a dashboard. By setting various conditions, rules, and triggers, BI pros can program the software to automatically notify relevant stakeholders if a certain event occurs or does not occur. They can set these triggers in any derived data source (data warehouse, data marts, or cubes) or in the BI applications. The application may deliver alerts via reports, dashboards, email, or mobile device.

- **Actions.** While alerts notify you of a certain condition, actions allow you to do something about it. This includes the ability to update a value in a cube, as in creating budgeting and planning scenarios (AKA write back). Actions could take you directly into your operational application and make an update (removing a duplicate record, for example, or approving a purchase order). Or actions can kick off a complex workflow process (usually via integration with BPM applications), such as approving a credit application. While most BI applications are analytical and even predictive, when actions are automated (based on a certain rule or a condition), BI can become prescriptive.

- **Portal.** Portals may serve as a single access point to catalog and index, classify, and search for BI objects such as dashboards and reports. Most BI platforms come with their own portals, but to organize multiple BI applications based on different platforms under one portal umbrella requires legwork. Organizations can apply enterprise portal standards to all BI applications and expose all BI metadata to the portal search engine to allow users to search for all BI content from a single point.

- **Collaboration.** Most leading BI platforms offer limited collaboration features such as report annotations and discussion threads. For richer collaboration functionality, organizations can integrate BI tools with enterprise social platforms to provide blogs, wikis, profiles, microblogging, activity feeds, community capabilities, tagging, tag clouds, and Really Simple Syndication (RSS).

- **Mobile.** Enterprises should extend portal and collaboration capabilities to mobile devices so that business users can access the same content in the same manner while on the road. Additionally, they should strive for consistency by creating the same look and feel (optimized for specific BI platforms) for all mobile BI applications.
Office applications. While all leading BI platforms integrate with Microsoft Word, PowerPoint (for documents and presentations that use “live” data), and Excel (so that one can use Excel as a graphical user interface [GUI] to secure and trusted corporate data), very few provide tight integration with email beyond emailing reports or report URLs. Organizations can embed BI dashboards into an email application (such as another panel — similar to mail, calendar, and tasks — in Outlook). Further customizations make the data more actionable. For example, an email from a certain customer could expose a dashboard about that customer.

![Figure 7 Forrester's BI Reference Architecture — Data Delivery Layer](source)

Other Components Span All Layers Of The BI Reference Architecture

Lastly, the following components span all of the layers of the BI reference architecture:

- **Integrated metadata.** Metadata describes and provides context to data and is an integral part of data governance and data usage. Metadata has grown beyond just data models and business definitions and now includes storage, life cycle, flow, confidentiality, quality criteria, stewardship, and semantic meaning. For multimedia, it can also include format and tags. Integrated metadata (ETL, DW, and BI tools metadata) is vital to BI application development and debugging, as it can support data lineage (tracing report content to its origin) and impact analysis (understanding the impact of changing any BI component or an object on all other components/objects). Taken one step further, ETL/DW/BI metadata that is integrated with ERP metadata can enable and support truly plug-and-play BI applications.

- **Information life-cycle management (ILM).** ILM comprises policies, processes, practices, and technologies used to manage data from its conception through its disposal. It involves all aspects of dealing with data to support a BI platform, including security, archiving, retention, and availability. ILM polices define the security controls, retention period, availability requirements, backup schedule, and type of media used.
**Enterprise content management (ECM).** In large enterprises, especially the ones in highly regulated industries, report generation is just the beginning of the BI cycle. Enterprises can use ECM platforms to store, index, classify, and archive reports.

**BI out of the box.** When complex and expensive ETL, data modeling, and BI tools are too much (too many moving pieces) and analytics-only tools are not enough (a certain degree of data integration and cleansing may be required), power users have other choices. Some vendors offer so-called BI suites where ETL, data modeling, and BI (sometimes including predictive analytics) are all part of one integrated — usually proprietary — suite of tools. Some other vendors offer a single design platform to automatically generate ETL, logical and physical data models, and BI metadata for multiple leading ETL, DW, and BI platforms.

**BI on BI.** These are BI metrics and reports that provide insights into efficiencies and effectiveness. For example, efficiency metrics could include the usage of various BI resources (servers, applications, tools, data sets, specific reports and queries, etc.). Firms can then proactively — rather than reactively — take steps to improve the performance of popular resources and eliminate waste by removing or deprioritizing seldom-used resources. Metrics that measure effectiveness of BI usage are trickier to quantify, but they are critical to a BI program’s overall success.

**Embedded BI.** Software developers often need to embed BI capabilities within applications, products, or online services. The degree and nature of the capabilities they require shapes the architecture of the solution, which can range from a complete, contained solution to one or more layers that developers embed within their application architecture “stack,” which delivers analytics through an application programming interface (API) or service or as a data set intended for subsequent read-only processing. Embedded APIs, services, or data integrate with the host app at a lower level in the stack and may be embedded within a software product as an OEM solution, within an application via licensing, or as a cloud service.

**Big data.** Forrester defines big data as “techniques and technologies that make getting value from data at extremes of scale economical.” There are three parts to this definition: techniques and technology; affordability; and extremes of scale (a combination of volume, velocity, variety, and variability that makes handling data issues with existing technology expensive or awkward). Big data technology generally has two or more of the following: 1) massively parallel processing; 2) the ability to do analytics against schemaless or flexible schema in a lightweight, agile fashion; 3) horizontal scale-out across hundreds or thousands of nodes in a cluster; and 4) use of advanced computing technology such as in-memory to improve real-time analytic responsiveness.
FORRESTER’S BI REFERENCE ARCHITECTURE — DETAILED VIEW

The final, complete detailed view of Forrester’s BI reference architecture includes all the layers and components described in this report (see Figure 8).

Figure 8 Forrester’s BI Reference Architecture — Detailed View
**RECOMMENDATIONS**

**EVERY ORGANIZATION NEEDS A BI REFERENCE ARCHITECTURE**

Without a reference architecture, BI pros, data architects, and analysts find it more challenging to use and manage information. A reference architecture provides the clarity necessary to develop a BI strategy that provides the capabilities that the business needs and the resources that are required to execute that strategy. In addition to all of the best practices already covered in this report, Forrester recommends that organizations:

- **Build BI architecture to support any application and platform.** Keep your understanding of the BI layer independent of the particular application, business process, platform, or tools. The architecture should be flexible enough to support any business requirements and also extend to integrate with newer technologies such as Apache Hadoop, in-memory, and cloud computing platforms.

- **Plan for and allow exceptions.** Use Forrester’s BI reference architecture as your strategic BI future state. However, recognize that situations, such as high-priority, urgent business requirements, will come up that will require deviations. The ability to deal with exceptions rather than fighting them is often what separates a successful BI environment from a failed one.

- **Start and proceed small but think big.** If the way you plan to reach the target state of your BI reference architecture appears to be a “boil the ocean” approach (first building an EDW, then addressing individual business requirements), then you are not doing it right. Rather, Forrester recommends the following five-step approach: 1) list your top BI business (not IT) priorities that you plan to address with this target state reference architecture; 2) rank them by easy/medium/hard, remembering that high-priority, low-effort initiatives are your sweet spot as a starting point; 3) pick as many as you can do simultaneously given your budget and resource constraints; 4) start executing on these tactical initiatives with rapid deliverables and a short cycle; 5) tack on, say, 20% of the resources and time on top of what you estimated to dedicate to the strategic initiatives such as building a new EDW (this approach is sometimes referred to as “EDW tax”).

- **Don’t expect mass migrations to cloud-based services to obviate the need for expertise.** A good deal of processing will move to cloud-based platforms, and even cloud-based data integration may be relevant to some organizations. But cloud-based services and platforms will not be a replacement for traditional BI, especially as it relates to data management practices for the foreseeable future. Rather, cloud-based capabilities will provide complementary techniques for targeted use cases. Think of the cloud less as a competing approach and more as an additional set of tools, techniques, and components to consider when building out your BI architecture.
ALTERNATIVE VIEW

A SNEAK PEEK INTO AN ALTERNATE BI REFERENCE ARCHITECTURE FUTURE

The need for business to access BI data in real time coupled with technology advancements around in-memory distributed cache, massively parallel processing (MPP), and data virtualization drives the need to create a single, real-time, scalable data management platform for transactional and analytical systems. Some ETL jobs move data from transactional and operational systems to data warehouses, shrinking their batch cycles to accommodate faster BI queries and reports, but the physical movement often requires considerable system resources and effort beyond the transformation, cleansing, and integration that is still required after the data movement. With the new distributed in-memory, MPP, and data virtualization technologies such as Apache Hadoop, IBM InfoSphere Information Server, IBM PureSystems integrated systems, Informatica Data Services, Microsoft xVelocity, Oracle Exadata, and SAP HANA, we are a step closer toward a common data management platform for transactional and analytical systems. However, additional efforts are still needed in the areas of real-time data quality, governance, transformation, aggregation, and MDM as well as tighter integration with the distributed in-memory, Apache Hadoop, MPP, and data services layer to deliver this real-time, scalable data management platform. If and when such technologies and architectures deliver this real-time platform, organizations will need to adjust their BI reference architectures to accommodate this change.

SUPPLEMENTAL MATERIAL

Companies Interviewed For This Report

Composite Software
Hexaware Technologies
IBM
Infosys
Microsoft

ENDNOTES

1 In 2011 we asked 208 IT executives to rate the top five technologies in terms of business value and expected change. They rated BI tops in both categories. See the October 7, 2011, “The Top 10 Business Technology Trends EA Should Watch: 2012 To 2014” report.

2 Enterprises are facing growing challenges in using disparate sources of data managed by different applications, including problems with integration, security, availability, and quality. Forrester’s data virtualization reference architecture “information fabric” focuses on a virtualized data layer that integrates heterogeneous data and content repositories, delivering a 360-degree view of all enterprise information, not just customer or financial data. See the April 9, 2007, “Information Fabric 2.0: Enterprise Information Virtualization Gets Real” report.
Just as in 2009, technical innovation and mergers and acquisitions (M&A) activity have continued to have a significant impact on the integration space. The most common trend has been the expansion of existing products and services to the point where boundaries between specific integration market categories have blurred considerably and application, process, and data integration are no longer isolated islands of functionality. Enterprises should consider responding to this trend by consolidating their varied integration resources into a single shared services group (an integration competency center) that will be able to deal with a multitude of interconnected integration challenges in the most effective manner. See the January 12, 2010, “2010 Update: Evaluating Integration Alternatives” report.

Today, enterprises face increasing data integration challenges, largely because of growing data volumes, data protection issues, data migration complexities, the need for real-time information, integration of structured with unstructured data, and the limitations of existing data integration technologies. Business users want quick access to reliable real-time information to help them make better business decisions. As a result, information architecture and application development professionals find themselves on a never-ending quest to improve data's quality and timeliness. To provide insight into the vast array of options available to meet today’s data integration challenges, Forrester investigated 11 technologies used to support a wide variety of data integration requirements, including change data capture (CDC); data access; data discovery; data quality services; data replication; data services platforms; distributed data caching; enterprise information integration (EII); enterprise services buses (ESBs); extract, transform, and load (ETL); and master data management (MDM). We found that data quality services, data services platforms, ESBs, and MDM continue to mature with widespread adoption, while CDC, data access, data replication, and ETL have reached the Equilibrium phase. Data discovery and distributed data caching are in the Survival phase, and EII stands alone in the Decline phase. See the February 4, 2010, “Forrester TechRadar™: Enterprise Data Integration, Q1 2010” report.

Business rules platforms are a mature technology for automating decision and policy logic and for managing fast changes to that logic to keep up with business changes. Now customers are seeking more: capabilities allowing them to employ business rules to help detect and respond to business events hiding in streams of data and to automate decision life cycles. Forrester published research that reveals how well vendors are responding to these new requirements. See the July 5, 2011, “The Future Of Business Rules Platforms” report.

A typical DW dilemma is how to change history. Many financial applications allow post-closing-period adjustments, which may or may not be automatically reflected in your DW (depending on your data integration architecture and processes). Companies often make one of two choices. One is not to reflect such adjustments in the DW (because “history cannot be changed”), which keeps the DW model and reporting more straightforward. But in this case, one would always have to be aware that the DW does not reflect the current state of events, only the historical. The other option is to add an offsetting transaction to the DW and create two sets of logical views: one to show history as it was on a certain date, the other one to show history that is up to date with transactional apps and the books of the company. This is typically the best option (win-win), but it will make your data model and report creation more challenging. The last option — to actually change history — is not a recommended one, mostly for compliance reasons.
6 We know how to address the key business intelligence (BI) challenges of the past 20 years, such as stability, robustness, and rich functionality. Agility and flexibility challenges now represent BI's next big opportunity. Business process professionals realize that earlier-generation BI technologies and architecture, while still useful for more stable BI applications, fall short in the ever-faster race of changing business requirements. Forrester recommends embracing Agile BI methodology, best practices, and technologies to tackle agility and flexibility opportunities. Alternative database management system (DBMS) engines architected specifically for Agile BI will emerge as one of the compelling Agile BI technologies business process pros should closely evaluate and consider for specific use cases. See the May 27, 2011, “It's The Dawning Of The Age Of BI DBMS” report.

7 In Forrester’s 145-criteria evaluation of enterprise business intelligence (BI) platform vendors, we found that IBM Cognos, SAP BusinessObjects, Oracle, Information Builders, SAS, Microsoft, and MicroStrategy led the pack because of the completeness of not just BI but overall information management functionality. Actuate came out as a Strong Performer on the heels of the Leaders offering equal — or in some cases superior — BI functionality, but it mostly relies on partners for the rest of its information management capabilities. Tibco Spotfire also came out as a Strong Performer offering top choices for analytics, even surpassing other Strong Performers in the overall information management arena based on its traditional strength in middleware and application integration. Last but not least, QlikTech and Panorama Software moved up from Contenders and into the Strong Performers category based on the continuous improvements in their analytical capabilities. See the October 20, 2010, “The Forrester Wave™: Enterprise Business Intelligence Platforms, Q4 2010” report.

8 Enterprises find advanced data visualization (ADV) platforms to be essential tools that enable them to monitor business, find patterns, and take action to avoid threats and snatch opportunities. In Forrester's 29-criteria evaluation of ADV vendors, we found that Tableau Software, IBM, Information Builders, SAS, SAP, Tibco Software, and Oracle led the pack due to the breadth of their ADV business intelligence (BI) functionality offerings. Microsoft, MicroStrategy, Actuate, QlikTech, Panorama Software, SpagoBI, Jaspersoft, and Pentaho were close on the heels of the Leaders, also offering solid functionality to enable business users to effectively visualize and analyze their enterprise data. See the July 17, 2012, “The Forrester Wave™: Advanced Data Visualization (ADV) Platforms, Q3 2012” report.

9 In Forrester’s 31-criteria evaluation of self-service business intelligence (BI) vendors, we found that IBM, Microsoft, SAP, SAS, Tibco Software, and MicroStrategy led the pack due to the breadth of their self-service BI functionality offerings. Information Builders, Tableau Software, Actuate, Oracle, QlikTech, and Panorama Software were close on the heels of the Leaders, also offering solid functionality to enable business users to self-serve most of their BI requirements. See the June 12, 2012, “The Forrester Wave™: Self-Service Business Intelligence Platforms, Q2 2012” report.

10 In Forrester's 53-criteria evaluation of predictive analytics and data mining (PA/DM) vendors, we found that SAS Institute, SPSS (evaluated separately from new parent IBM's other PA/DM offerings), KXEN, Oracle, Portrait Software, and IBM (pre-SPSS acquisition PA/DM offerings) head the pack with mature, sophisticated, scalable, flexible, and robust solutions. SAS leads, providing the most feature-rich solution portfolio and, through its recent expansion of enterprise data warehouse (EDW) vendor partnerships,
taking the industry lead in promoting in-database analytics as an emerging best practice for high-performance analytics deployments. SPSS is rapidly integrating its already strong PA/DM solution portfolio with new parent IBM's extensive data management solution family. KXEN stands out for its focus on content analytics, sentiment analysis, and social network analysis. Oracle distinguishes itself through the depth of its PA/DM tool's integration into its enterprise database and application portfolio. Portrait provides a comprehensive set of customer analytics offerings that integrate with its core PA/DM tool. Tibco Software, FICO, and Angoss Software are Strong Performers in the PA/DM market. See the February 4, 2010, “The Forrester Wave™: Predictive Analytics And Data Mining Solutions, Q1 2010” report.

11 Companies waste millions of dollars each year on ineffective and inefficient business processes. Many companies attack this waste with process improvement exercises using approaches such as Six Sigma, Lean, Lean Six Sigma, and total quality management (TQM). However, when automating, controlling, and reporting on mission-critical business processes, many of these same firms remain stuck using rigid packaged applications or expensive custom-developed solutions. Process analysis, automation, and transformation through business process management (BPM) suites increase collaboration across business and technical teams while providing the flexibility and adaptability needed to quickly change processes based on new competitive and economic threats and opportunities. Firms can realize a return on their investment in BPM suites within three years through impressive gains in business efficiency and worker productivity. This makes BPM suites a good investment in the face of challenging economic conditions and an increasingly competitive business environment. See the August 22, 2011, “The ROI Of BPM Suites” report.

12 Business applications are mission-critical engines driving business execution in organizations, but challenges are mounting to stabilize costs of ownership, leverage new technologies effectively, and match the changing needs of the business. Application delivery leaders need a straightforward planning process to take inventory of their existing applications and determine the path forward that best aligns with business strategies. Forrester offers a five-stage guide to tuning or transforming the business application portfolio, which can be applied to specific business domains as well as to the entire portfolio. See the May 30, 2012, “Optimize Application Strategy To Generate Business Value” report.

13 At extreme scale, traditional data management and business intelligence (BI) become impractical, and your business does not get what it demands — more insight to drive greater business performance. Big data helps firms work with extremes to deliver value from data cost-effectively. However, CIOs must understand that this is not business as usual. In fact, big data will disrupt the data management landscape by changing fundamental notions about data governance and IT delivery. Take the time to understand big data as well as its implications and begin a balanced approach that considers more than just the technology hype. See the September 30, 2011, “Expand Your Digital Horizon With Big Data” report.

14 In some extreme cases, big data can be a mix, an intersection of up to nine dimensions: 1) volume; 2) variety of data sources and data times; 3) variability of data meanings; 4) velocity of data change; 5) velocity of requirements change; 6) data (model) complexity; 7) data exploration and discovery (without a structured model); 8) low cost (often using commodity open source software such as Apache Hadoop); and 9) linear scalability.
Forrester Research, Inc. (Nasdaq: FORR) is an independent research company that provides pragmatic and forward-thinking advice to global leaders in business and technology. Forrester works with professionals in 17 key roles at major companies providing proprietary research, customer insight, consulting, events, and peer-to-peer executive programs. For more than 29 years, Forrester has been making IT, marketing, and technology industry leaders successful every day. For more information, visit www.forrester.com.

About Forrester

A global research and advisory firm, Forrester inspires leaders, informs better decisions, and helps the world’s top companies turn the complexity of change into business advantage. Our research-based insight and objective advice enable IT professionals to lead more successfully within IT and extend their impact beyond the traditional IT organization. Tailored to your individual role, our resources allow you to focus on important business issues — margin, speed, growth — first, technology second.

FOR MORE INFORMATION

To find out how Forrester Research can help you be successful every day, please contact the office nearest you, or visit us at www.forrester.com. For a complete list of worldwide locations, visit www.forrester.com/about.

CLIENT SUPPORT

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or clientsupport@forrester.com. We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Focuses On

Application Development & Delivery Professionals

Responsible for leading development and delivery of applications that support your company’s business strategies, you also choose technology and architecture while managing people, skills, practices, and organization to maximize value. Forrester’s subject-matter expertise and deep understanding of your role will help you create forward-thinking strategies; weigh opportunity against risk; justify decisions; and optimize your individual, team, and corporate performance.

ANDREA DAVIES, client persona representing Market Insights Professionals