The Trusted CMDB:
Data Quality and Governance

EMA™
®

IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING

The Trusted CMDB: Data Quality and Governance

Table of Contents

EMA

Executive Summary

The Configuration Management Database (CMDB) is a central capability for large scale IT management. It integrates the most critical IT data for operations and planning purposes, and provides context for understanding the broader landscape of IT management information. However, like any data-intensive production system, it suffers greatly if its data is perceived to be inaccurate.

In turn, shortcomings in this information can translate to higher costs and risk for the enterprise. Blazent offers a unique quality assurance capability for the CMDB to increase trust in and value of this critical IT system.

Introduction

Historically, the data repositories for IT management have been some of the most fragmented systems in the enterprise. The Configuration Management Database[1], or CMDB, emerged as the best known alternative of various "repository" proposals throughout the 1980s and 1990s. Given visibility through the IT Infrastructure Library (ITIL), it has achieved widespread mindshare and a corresponding degree of industry skepticism.

Enterprise Management Associates (EMA) has covered the concept and evolution of CMDB since 2003. This research has shown that CMDB/CMS investments have brought significant benefits across a wide range of use cases including IT financial and asset management, change management, incident and problem management, operations management, and even IT portfolio management.

However, while very real benefits have been achieved, the CMDB is a complex, integrated system presenting various challenges and there are a number of ways it can fail:

- Lack of management support
- Not understanding requirements
- Poor data quality
- Insufficient perceived value

This EMA white paper will focus on the issue of CMDB data quality, as this is a challenge that must be addressed head-on by the CMDB team. Failures in CMDB data quality can have significant impacts on an enterprise's IT cost and risk profile. Large enterprises increasingly outsource some or all server or desktop maintenance (patching for example). Outsourcers charge on a per-device basis, and if the server count is inaccurate, the service provider may bill incorrectly. Change management relies on correct CMDB information to ensure operational stability. Network security is utterly dependent on a full understanding of the perimeter.

The consequences of inaccurate IT management data are therefore problems with significant operational consequences for IT and ultimately IT-business relationships. Ensuring the accuracy of this data becomes a governance function that cannot itself be outsourced.

---

[1] While ITIL version 3 replaced the term Configuration Management Database (CMDB) with Configuration Management System (CMS), and ITIL 2011 now is advocating the Service Knowledge Management System (SKMS), this paper opts to remain with the term CMDB which in EMA's experience still has the greater industry recognition.

The CMDB Challenge

The CMDB's architecture and role is consistent with what Bill Inmon termed the "Operational Data Store" [1., 2.]. It is an integrated, current repository that bears some resemblance to a data warehouse, but does not keep extensive history, and can be updated transactionally. Inmon warned of the complexity of these kinds of systems, and the CMDB experience demonstrates he was correct to do so.

After some earlier attempts at stand-alone CMDBs, these repositories increasingly are found at the heart of integrated IT Service Management (ITSM) suites. The combination of CMDB with Incident, Problem, Change, IT Asset, and Service Request Management is becoming a de facto functional core for many ITSM systems. However, the CMDB still has many integrations with external tools (Figure 1).
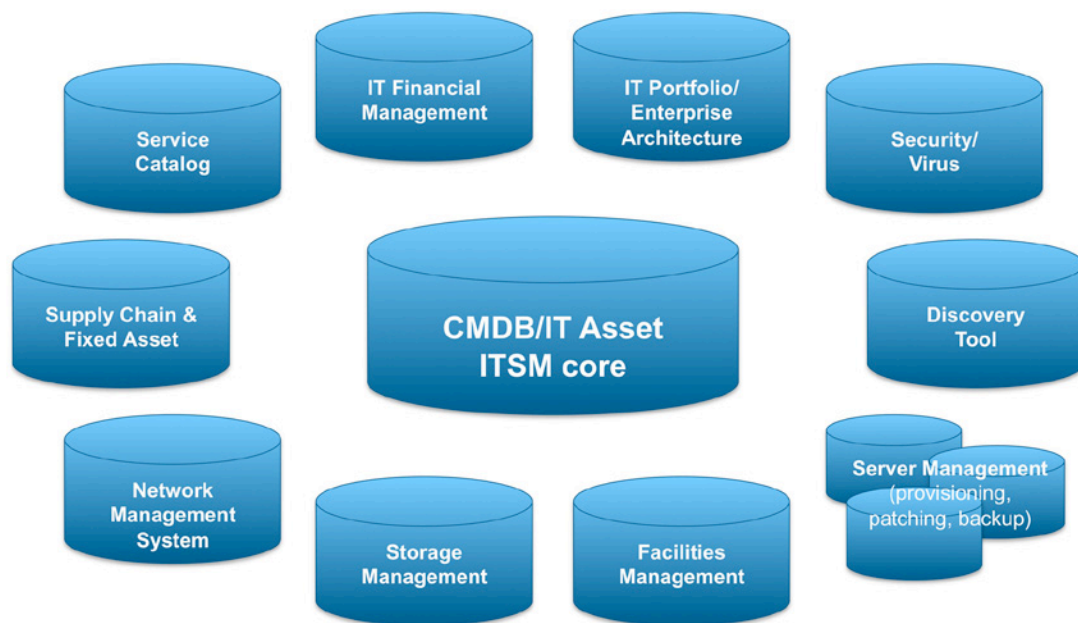

Figure 1: CMDB Integration Architecture

CMDBs often have a data processing layer, perhaps called a "reconciliation engine," which checks incoming data against current records and either allows or disallows CMDB additions, changes, and deletions. However, in CMDB products this reconciliation engine is limited in functionality. The focus is on supporting the CMDB as a presumed authoritative source. Towards this end, CMDB reconciliation practices may actually discard some data that is important to understanding and improving the IT environment. This can be especially problematic when first establishing the CMDB, but remains an ongoing challenge in federated environments.

> In CMDB products this reconciliation engine is limited in functionality.

CMDBs are also limited in the use cases they support. They don't provide sufficient tools for working with **sets of data**. Many service providers are interested in being able to identify a set of servers meeting some exception criteria, and then being able to dispatch and track some effort to address those exceptions. CMDBs too often require such use cases to extract a static spreadsheet, which then quickly becomes stale.

EMA™

"Somebody once asked me to make a guess about how many CIs are in our CMDB across the whole estate and my rough estimate was 20%.... It's really low. And there's also a lack of confidence that it's accurate. What's in there might not even be real, might not even be right." ~ Senior IT Executive, Large Financial Services Firm

Continuous Improvement and Data Quality

Data quality is a concern for any recordkeeping system, whether manual or automated. Issues with poor data quality can build to the point where an IT system's capabilities and even value become degraded. In response, the field of data quality management has developed. Based on well-established principles of quality management and continuous improvement [3., 8.], data quality management seeks to identify information management issues and their root causes. Master Data Management is another, related topic of importance for CMDB and IT Asset Management practitioners [9.].

For example, a data quality effort might start by examining ("profiling") data in a given system and examining it for obvious exceptions. But exception reporting can only address some aspects of data quality. Data may appear to be fine – it may satisfy all known business rules – and yet simply be wrong. At scale, statistical sampling to infer the level of data quality may be needed.

> Data quality management seeks to identify information management issues and their root causes.

It's critical to understand the cause of data errors. It is not sufficient to report on exceptions and merely go and fix them. Quality management requires digging into the reasons for the errors. For example, Lean practices as originated at Toyota call for the "Five Whys": asking "Why" five times.

Without such a continuous improvement perspective, any data quality effort is at risk, and if data quality is at risk, the integrity of a systems architecture actually can be compromised. EMA analysts have observed what they call the "Data Quality Vicious Cycle" (Figure 2).
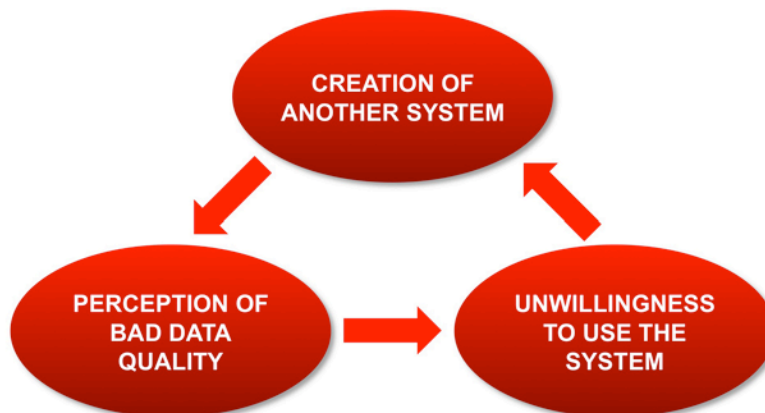

Figure 2: The Data Quality Vicious Cycle

In the Data Quality Vicious Cycle, the perception of poor data quality actually leads to the creation of competing, redundant systems (sometimes spreadsheets or desktop databases), which in turn have their detractors and own problems of data quality – a self-reinforcing, downward spiral of increasing cost and decreasing effectiveness. There is only one way to reverse the vicious cycle: **treat data quality as a continuous improvement problem and measure and show progress over time.**

Data quality has benefits beyond reversing the vicious cycle. Quantifying them depends on the business problem being solved, but they have been demonstrated on both the top and bottom line. Sometimes, risk management analyses may show clear value at risk if poor data quality led to operational failures. In other cases, poor data quality may inhibit cross-selling, or interfere with a clear view of enterprise profitability. EMA sees more and more examples of such quantified benefits, and Master Data Management (MDM) and Data Quality (DQ) vendors are doing very well as part of the broader Big Data marketplace.

The CMDB and Data Quality

As a large, complex Operational Data Store, the CMDB suffers from all the same issues of data quality that a customer database might. Internal IT systems are just as prone to the Data Quality Vicious Cycle as business-facing systems.[2] And the value of data quality for the CMDB can be quantified in terms of risk reduction, increased agility, better IT decision making, and, especially for Managed Service Providers (MSPS), even top line revenue (Figure 3). It's possible to quantify risk and value of this kind; see [10.] and consider using tools such as Monte Carlo analyses.
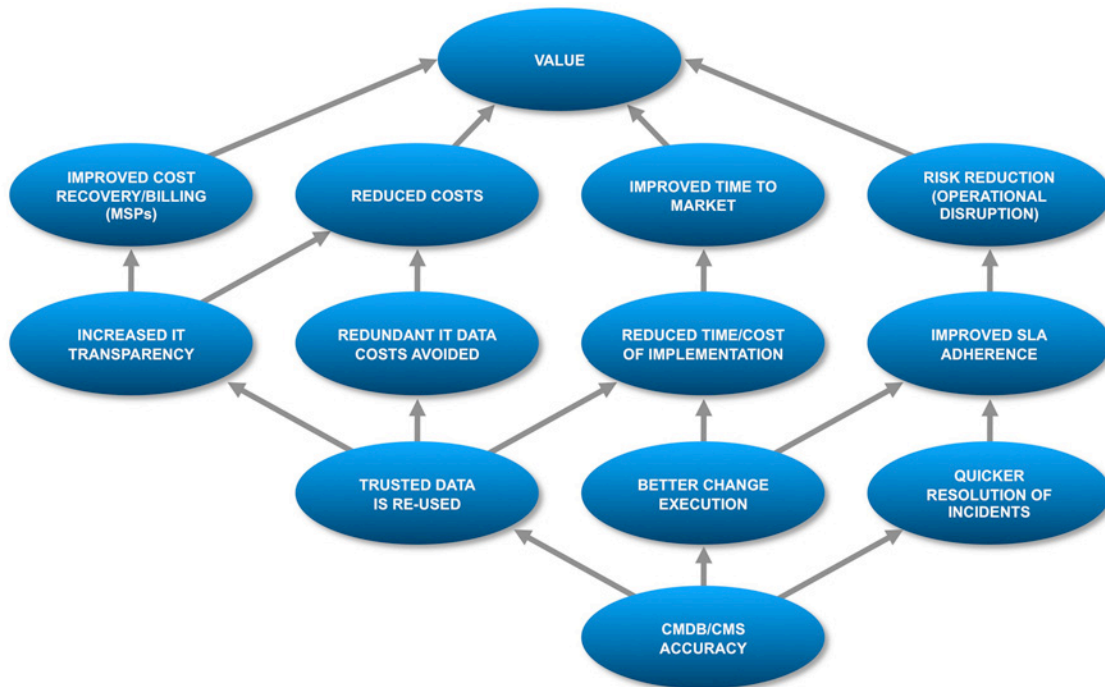

Figure 3: CMDB Data Quality Value Proposition

[2] Perhaps more so, since IT staff have the tools and skills to create new systems if they are dissatisfied with present ones. IT staff also can be resistant to process and unfamiliar with quality management principles.

"Dealing with bad data is troublesome when you're trying to focus on quality, service excellence, and service availability. In the financial industry, if we are down, people can't trade, people can't pay for dinner. If we have bad data, if I can't connect this infrastructure device to that infrastructure device, I've got people spinning and I just can't afford to do that." ~ Senior IT Executive, Large Financial Services Firm

---

And data quality practices apply just as well to the CMDB as to any business-facing data store. Business rules can be defined for expected CMDB data characteristics, exception reports can be generated and worked, root cause investigations can and should be performed, and progress on data quality should be tracked and reported on over time.

Exception reporting in the CMDB often is performed when multiple sources are brought together. If server records from three sources are being integrated, there are multiple possibilities (Table 1).

The server appears in systems A, B, and C	The server appears in system A, but neither B nor C
The server appears in systems A and B but not C	The server appears in system B, but neither A nor C
The server appears in systems B and C but not A	The server appears in system C, but neither A nor B
The server appears in systems A and C but not B	The server does not appear in system A, B, or C

Table 1

The scenarios above all need to be explained. What if the server is seen in System A but neither B or C? Does it exist? What further confirmation would be advisable? Is System A notably inaccurate, to the point where its data is discounted? Or is this indicative of a serious gap in processing? Perhaps even a stolen asset? And all of these scenarios are **independent of whether the server actually exists**. Appearance in multiple systems may be strong evidence for the device's existence, but perhaps the systems are all wrong.

Complicating matters is the fact that the server's existence may be represented by a number of different data points (Figure 4). It may surprise people who are not professional IT asset managers that a device may have four different serial number representations, and exceptions may be found comparing them. The plate is generally seen as authoritative, but inspecting a device's often-illegible and inaccessible plate is both expensive and error-prone. Similarly, the host name and the primary DNS name by convention should be identical, but this paper's author has seen situations where they differed. Virtualization causes further difficulties with identifiers like IP and MAC address. And Cloud options may simplify things on the physical side, but present new problems in tracking the logical assets and what is being done with them. Cloud sprawl is costly, and Cloud servers represent liabilities. Software asset management and IT services represent further data subjects a comprehensive solution needs to cover.
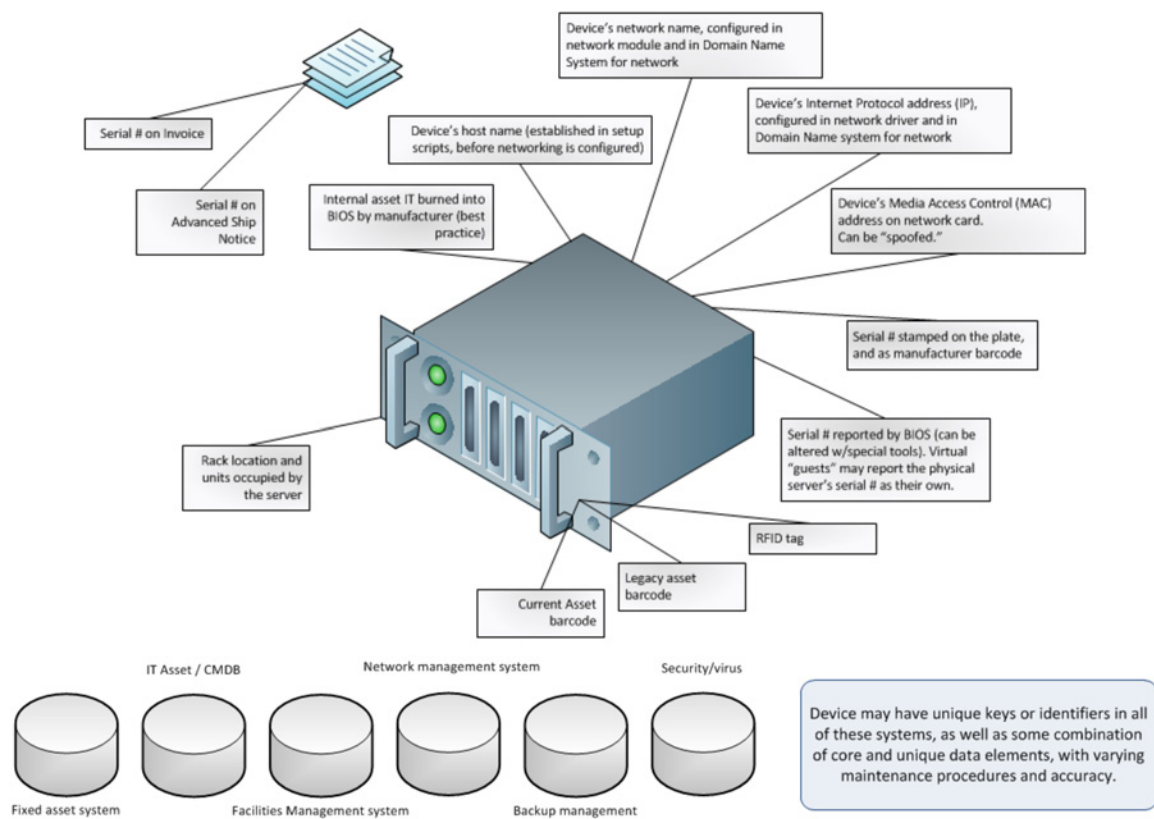

Figure 4: The Multiple Identifier Problem

All in all, the business rules for reconciling the basic physical stock of IT assets are challenging and complex. They are not easily built from scratch, and CMDB vendors have been hesitant to take a strong stand on best practices, preferring to leave that up to customer preference. Furthermore, as noted above, many CMDBs' reconciliation practices are simplistic. For example, it may be difficult to routinely determine WHY a given server record was discarded. Probabilistic approaches (we think there is an 80% chance this record is accurate) are typically not seen. Nor does the typical CMDB reconciliation engine have the intelligence to identify recurring issues with data quality.

The Trusted CMDB: Data Quality and Governance

Furthermore, assuming that some business rules have been adopted for CMDB/asset reconciliation, there still is the question of what to do with exceptions once they are identified. IT process improvement, like any continuous improvement effort, requires persistent analysis.

Ultimately, EMA proposes that IT management adopt an enterprise data architecture strategy that transcends the CMDB. A robust data services layer is required, independent of any functional IT system, supporting data quality, master data management, analytics, and related use cases (Figure 5).
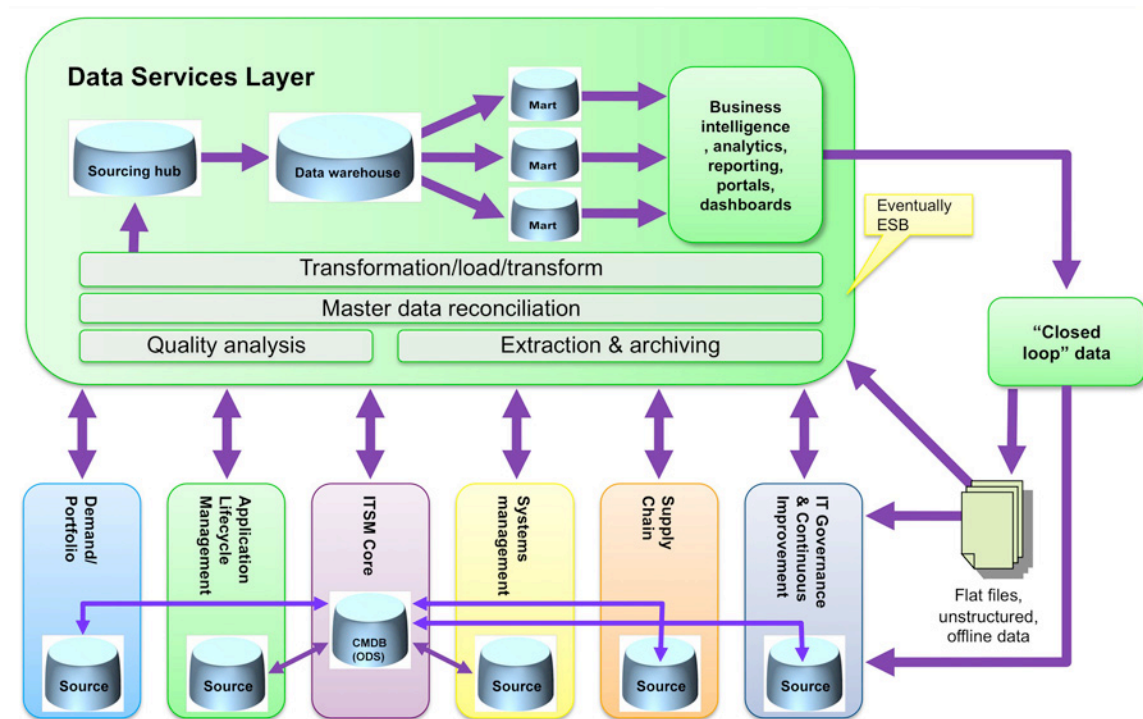

Figure 5: IT Management With Mature Data Services

In summary:

- The CMDB is a data-intensive, complex, federated system.
- Multiple sources of data must be brought together and this always requires some form of reconciliation processing; i.e., complex business rules.
- The value of data quality can be quantified.
- The CMDB should be seen as part of a broader IT data management architecture.
- A continuous improvement approach is a must.

The CMDB should be seen as part of a broader IT data management architecture.

> *"The CMDB is way over hyped - data goes in, doesn't come out. It's like an 18 wheel truck – doesn't occur to them they have a blind spot. There are too many data sources the CMDB doesn't care about. Blazent doesn't have that blind spot. Blazent is not limited to someone's vision of ITIL processes and procedures."* ~ Senior IT Executive, Managed Services Provider

Blazent's Approach

These problems of CMDB and IT Asset quality and reliability are exactly where Blazent positions its solution. As shown in Figure 6, Blazent's SaaS-delivered product starts with a flexible reconciliation engine and high capacity data store, front-ended by a commercial grade business intelligence interface.

Blazent is source-agnostic and accepts data from a wide variety of IT management sources, including spreadsheets. The product embodies a mature set of reconciliation techniques for IT management problems, tailored very specifically to the business of IT. Blazent demonstrates deep intellectual property in its approach to reconciling and rationalizing the data elements shown in Figure 4, when seen in multiple converged sources, using detailed sets of source and field level precedence rules derived across dozens of engagements for some of the largest IT service operations.

Time to value is a key Blazent proposition. The tool's orientation as a data workbench enables fast assimilation and analysis of various data sets. Blazent customers consistently indicate this as a strength of the product, where CMDBs implemented as part of heavy ITSM suites can take months to show any progress at all.
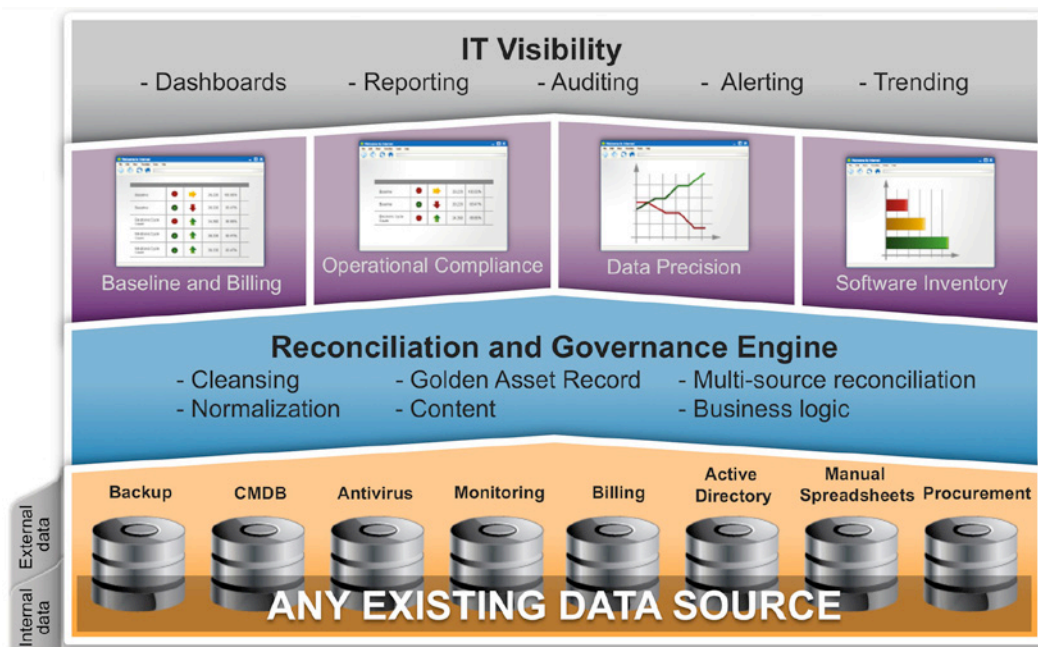

Figure 6: Blazent Architecture

> *Beyond data accuracy, Blazent has capabilities to benefit operations. Blazent makes it easy to audit if an agent is present and signature file up to date. For us to bill a customer, the asset has to be active, with anti-virus, backed up, and managed in Microsoft SCCM. The asset repository does not give us all this.* ~ Senior IT Executive, Managed Services Provider

Visually appealing dashboards are used to communicate the status of the various data subjects. Typical areas of concern are reflected in Figure 7: core Asset information and level of consistency with partner systems such as anti-virus, backup, and monitoring.

Blazent is not currently partnered with any particular CMDB/CMS vendors, but this neutrality is a good thing. While platform and other vendors may claim to fix all the problems Blazent addresses, there always remain holes and gaps—especially since real-world environments are made up of multiple brands of tools at different levels of maturity. Blazent's ability to access data is not dependent on partnerships or vendor hand holding—a simple flat file transfer is all that's needed for data access.
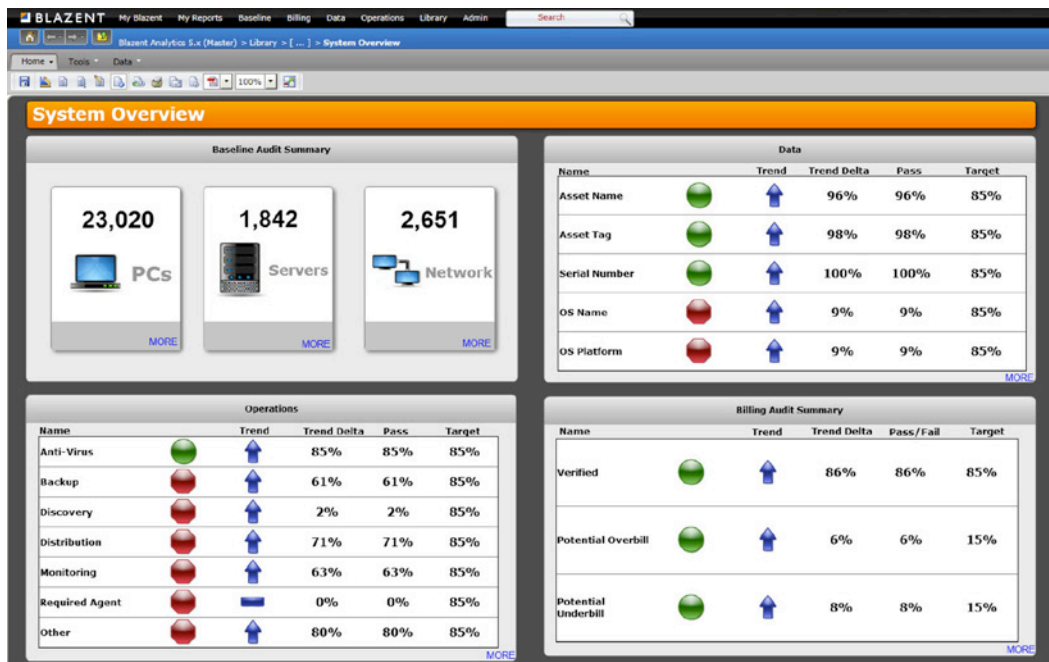

Figure 7: Blazent Dashboard

Blazent has gained great traction with large enterprises and outsourcers focusing on an essential core of data about workstations and servers. Recently, the product has expanded its focus to software inventory, and some customers are working with views of IT services.

> *"Blazent becomes an ongoing sanity check to tell us how complete we are – both in terms of the data we collect and the accuracy of our coverage. We can track this to who owns the data and then go back to the data owner and get them to correct it."* ~ Senior IT Executive, Large Financial Services Firm

EMA Perspective

Some personal experiences with CMDB data quality and reconciliation

Charlie Betz, EMA Research Director and report author

When I was working for a large firm, we encountered the exact set of problems that Blazent addresses. Auditors were questioning the integrity of the firm's Asset Management practices, and this led to a complete wall-to-wall inventory of about 100,000 IT assets across multiple data centers nationwide.

We had a CMDB, but there was also a Fixed Asset System, and the discrepancies between the two had a bright light shone on them. The painstakingly harvested wall-to-wall inventory data was yet a third source, and various infrastructure managers had their own spreadsheets.

This was not a problem that lent itself to solving entirely within the CMDB, which had a production focus. The facilities simply weren't there to load in thousands of rows from shadow spreadsheets and quickly understand the overlaps and deltas. The CMDB's reconciliation engine was too focused on pipelining information into the CMDB and did not allow for a more cautious, exploratory approach to understanding the disparate data sets. Basing the entire solution on the CMDB was therefore an approach that simply did not work either technically or politically.

We assigned a very senior data architect to the problem, a person who could have been working on other and perhaps more valuable problems for the firm. We quickly discovered that the reconciliation problem was challenging, for reasons I've outlined in this paper. While simple conceptually, the various scenarios and how to interpret them became very time-consuming, and at the end of the project all we had was a set of complex, "hand-cranked" SQL scripts that was not a production solution. (Blazent clients I interviewed also talked of having developed such interim solutions, and realizing they weren't sustainable.)

A year or so later, software asset management became a major concern, and we were again faced with a formidable reconciliation problem we just didn't have production-class data services and tools for. Ultimately, we needed a flexible data workbench for analysis and reporting, and to serve as a cross-check on the CMDB. When I encountered Blazent as an analyst, the value was immediately apparent.

EMA

It's the perspective of Enterprise Management Associates that the professionals who tend to be most directly involved in IT Service and IT Asset Management have little to no exposure to data management best practices. Data quality, data modeling, data governance, master data management, and similar disciplines are often very unfamiliar topics to the practicing ITSM professional. (See [11.] for the most comprehensive, industry leading overview of data management.) Data management tends to be better understood in business facing application areas, where data quality can have an immediate impact on customer satisfaction and the bottom line. This professional distance may be a root cause for the struggles IT has faced in executing CMDB projects.

Despite this lack of familiarity, over the past ten years, many organizations have made significant strides in IT data management. It is therefore not surprising that specialized tools like Blazent are emerging in response to widely felt pain points. Because the basic principles of data management do not change, it's also unsurprising that Blazent has significant similarities to more general purpose products found in the data management space.

"Once we had all of the Blazent data loaded in, I had a report run to show me in the last 60 days all of the assets that have shown up that have not been contained or have not been caught by our information security tool. We gave the IP addresses to the information security team, so they could make sure that the information security tool was now capturing all of the vulnerabilities. That was a really big win." ~ Senior IT Executive, Large Financial Services Firm

---

For example, Informatica has its data quality and master data management solutions applicable to a wide variety of business information domains. However, such tools' flexibility is also their downside, as (unless the vendor or a partner has built out an accelerator) the task of defining quality and reconciliation business rules falls upon the end user. This is one of Blazent's key value propositions, that they have built this intellectual property, and EMA is not aware of any other vendor with a similar focus on data quality in the IT vertical.

Blazent has made great headway in its focus on improving the core CMDB/Asset problem. Some areas for further evolution that Blazent is currently pursuing are:

- Expanding data subjects
  - Service view
  - Increasing coverage of the software space
  - Networking
- Expanding their analytic intellectual property to correspond with new data subjects
- Continuous improvement workflow – when a data quality exception is encountered, is its root cause being investigated?
- Improving front end usability – submitting large sets of Internet Protocol addresses dynamically as a query predicate was a specific problem identified by users in interviews.

EMA encourages Blazent to continue to broaden its scope in these directions, so as to become a full service master data management and data quality service for the business of IT.

Blazent fills a unique role in the IT management ecosystem. It will shine in federated environments, and also should become a favored tool of IT auditors, looking for independent validation of automated feeds and integration logic.

Ultimately, Blazent fills a governance role, in ensuring the integrity of management data being utilized for critical billing and operational management. While some might view it as duplicative of the CMDB, that viewpoint is not held by EMA. Blazent instead serves as an essential cross-check on federated CMDB environments, providing an objective view across what can be politicized silos. In order to do this, it adopts a data workbench architecture that is source-neutral.

IT management is now too critical and too large scale to disregard well-accepted practices of fact-based data quality and data management. Blazent currently stands alone in its support for data quality management and increasing the IT's organization's trust in the integrity and accuracy of its critical information.

"Blazent come along and they look at the landscape. They look and see what you've got, what spreadsheet you might have that somebody in finance has been keeping for the last 30 years or something, the process is all about trying to open up information, all the information that you've got, feeding into the central repository, having Blazent calculate that golden record, and then working from there." ~ Senior IT Executive, Large Financial Services Firm

---

About Blazent

Blazent provides IT visibility & intelligence solutions to help companies uncover hidden opportunities to optimize IT operations, lower costs and reduce risks. By integrating data from disparate systems and applying built-in best practices, Blazent's solutions deliver complete, accurate and trusted views into a company's entire IT environment, regardless of whether it's internal or outsourced. To date, Blazent solutions are deployed at some of the world's largest, most complex IT organizations to manage more than 1.5 million end user devices, virtual and physical servers, and network assets. For more information, visit www.blazent.com or call 855.282.8571.

References

1. Betz, C. *The CMDB is an Operational Data Store*. The Data Administration Newsletter, 2009.

2. Inmon, W.H., *Building the Operational Data Store*. 1996, New York: John Wiley & Sons.

3. Brackett, M.H., *Data resource quality : turning bad habits into good practices*. Addison-Wesley information technology series. 2000, Boston: Addison-Wesley. xxviii, 354.

4. English, L.P., *Information quality applied : best practices for improving business information, processes, and systems*. 2009, Indianapolis: Wiley. xxxv, 802 p.

5. Huang, K.-T., Y.W. Lee, and R.Y. Wang, *Quality information and knowledge*. 1999, Upper Saddle River, N.J.: Prentice Hall PTR. xiv, 209.

6. Loshin, D., *Enterprise knowledge management : the data quality approach*. 2001, San Diego, Calif.; London: Morgan Kaufmann ; Academic Press. xviii, 493.

7. McGilvray, D., *Executing data quality projects : ten steps to quality data and trusted information*. 2008, Amsterdam ; Boston: Morgan Kaufmann/Elsevier. xviii, 325 p.

8. Redman, T.C., *Data quality : The Field Guide*. 2001, Boston: Digital Press. xviii, 241.

9. Dreibelbis, A., *Enterprise master data management : an SOA approach to managing core information*. 2008, Upper Saddle River, NJ: IBM Press/Pearson plc. xxxv, 617 p.

10. Hubbard, D.W., *How to measure anything : finding the value of "intangibles" in business*. 2nd ed. 2010, Hoboken, N.J.: Wiley. xv, 304 p.

11. Data Management Association, T., *The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide)*. 2009, Bradley Beach, NJ: Technics Publications, LLC.