

## Cool Vendors in Data Management and Integration, 2010

Eric Thoo, Donald Feinberg, Ted Friedman, Andreas Bitterer

This research analyzes five vendors with innovative approaches and technologies to help organizations improve the scalability of databases, introducing new optimization techniques, integrating data for delivery to software as a service (SaaS)-based applications and ensuring the quality of data across applications and data structures.

### Key Findings

- Data integration requirements are expanding to encompass data beyond the boundaries of individual enterprises, involving both internal and external data sources and targets.
- Innovation in data quality tools offers the use of expert system-based techniques for enabling greater efficiency and reusability for data quality functions deployed using improved management of business rules in data quality improvement projects.
- New database management system (DBMS) technologies continue to emerge through the use of new optimization techniques to manage mixed data types.

### Recommendations

- Look for ways in which data integration deployments can improve reuse and scalability. Explore the use of semantically based approaches to resolve integration challenges.
- Investigate data integration technologies that can unify data residing in applications both inside and outside an enterprise's firewall, to address alternative models for adopting business applications (such as SaaS) with flexible, yet cost-effective approaches to data management.
- Explore alternative DBMS technologies and optimization techniques for addressing workload scalability challenges in data persistence activity, such as data warehouses for mission-critical front-line operations.
- Seek benefit from data quality technologies using adaptable approaches, such as expert systems, focusing on increased business orientation.

## TABLE OF CONTENTS

---

Analysis .....	3
1.0 What You Need to Know .....	3
2.0 Adeptia .....	3
3.0 Algebraix.....	4
4.0 Clavis Technology.....	5
5.0 Cloudera.....	6
6.0 Mastersoft Research .....	7
Recommended Reading.....	7

## ANALYSIS

---

*This research does not constitute an exhaustive list of vendors in any given technology area, but rather is designed to highlight interesting, new and innovative vendors, products and services. Gartner disclaims all warranties, express or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.*

### 1.0 What You Need to Know

Data management tools are evolving, with innovative approaches and technologies for enhancing the scalability of database management systems introducing new optimization techniques, innovative approaches to improve data quality and integrating data for delivery to SaaS-based applications and ensuring the integrity of data across applications and data structures.

An increase in the uptake of hosted SaaS models for application delivery is driving demand for data integration capabilities that involve both internal and external data sources and targets. Despite the long dominance of relational technologies, new DBMS technologies continue to emerge through the use of new optimization techniques. Explore the use of inference-based approaches for data quality improvements to meet diverse business requirements. Look for ways in which data management tools have advanced to solve data related challenges and understand opportunities to deploy data management capabilities more efficiently and cost-effectively.

### 2.0 Adeptia

Chicago, Illinois ([www.adeptia.com](http://www.adeptia.com) )

*Analysis by Andreas Bitterer*

**Why Cool:** The single Adeptia platform is able to support batch and bulk data integration in extraction, transformation and loading (ETL) environments, business process and enterprise application integration and long-running transactions that include human workflow and approvals and errors and exceptions handling. Adeptia is a single install product and its components are switched on by entering a corresponding license key. The data integration platform is built on a service-oriented architecture, so all reusable components can be integrated into other solution hosts.

Adeptia is a privately held (\$12 million), venture-backed company founded in 2000, with its headquarters in Chicago, where the vendor also maintains its sales, marketing and technology architecture function. The development and testing group is a wholly-owned subsidiary located in New Delhi, India. Adeptia focuses on mid-market (organizations up to \$1 billion in revenue) solutions for a variety of departmental data integration requirements and business-to-business (B2B) data exchange. The company has approximately 200 deployments and 150 customers. Adeptia has about a dozen system integration partners, in the U.S. and in Europe, most of which are specialized vendors, such as Netherlands-based Enable-U.

Depending on customer preference, Adeptia offers a flexible pricing model, either based on a yearly subscription or a perpetual model. In addition, the Adeptia solution can be deployed on-site or in a "cloud" environment, including the Amazon Elastic Cloud (EC2), where small deployments run in the region of \$500 per month. For an installation of a server with four CPU cores, the Adeptia integration suite costs \$1,200 per month, while the ETL suite costs \$800 per month. For customers preferring a perpetual license, all prices simply need to be multiplied by a factor of 30. Most platform adapters come free-of-charge, while a handful of connectors (for example, SAP, EDI and NetSuite) cost an additional \$100 to \$200 per month.

**Challenges:** Due to the vendor's almost non-existing marketing, small sales force of six people, and limited partner and independent software vendor channel, Adeptia is virtually unknown among Gartner clients and the industry. While Adeptia has signed some impressive customers, its powerful competition, such as IBM, Informatica, Microsoft, Oracle and SAP is dominating the data integration domain, often with aggressive pricing, slowing Adeptia's adoption rate. Furthermore, companies looking for more enterprisewide data integration functionality do not have Adeptia on their radar. The Adeptia platform also has no data quality components. As the data integration and data quality markets continue to slowly converge, other data integration platforms, either through acquisition or through OEM agreements, already include data profiling, cleansing, matching, standardization and enrichment, for example.

**Who Should Care:** Organizations with small and medium-scale data integration needs should take a look at Adeptia and consider its broad integration technology as a component for departmental solutions or B2B integration, such as EDI. Given the relatively low price-point and the flexible pricing model, organizations investigating an open-source data integration platform, may consider Adeptia a viable alternative. Application developers may be interested to embed Adeptia routines as a service within their service-oriented architecture deployments.

**Related Research:**

"Magic Quadrant for Data Integration Tools"

## 3.0 Algebraix

San Diego, California ([www.algebraixdata.com](http://www.algebraixdata.com))

*Analysis by Donald Feinberg*

**Why Cool:** Algebraix's general availability was announced 23 February, 2010 and is the newest specialized DBMS for high-speed analytics and business intelligence (BI) applications. The DBMS, named A<sup>2</sup>DB for Advanced Analytic Database, is not based on relational algebra (as are the relational DBMSs), but on extended set algebra and does not use the standard relational techniques of a RDBMS, although fully accessible by Structured Query Language. Through a patented process called Adaptive Data Restructuring, A<sup>2</sup>DB dynamically generates data structures and access methods based on the data loaded into the DBMS and the queries performed on that data. Additionally, A<sup>2</sup>DB does not have the same restrictions on non-relational data types as the relational database management system and so can perform analytics on mixed-data types more efficiently.

Algebraix brings the first really new data model for BI and analytics to the market not based on a relational model, column-store, fast file access or other technology available in the market. It is based on rigorous mathematical models, different from those of relational. A<sup>2</sup>DB will run on x86 hardware with 64-bit Linux or Windows and has industry standard interfaces, such as Open Database Connectivity and Java Database Connectivity. These interfaces allow it to be used with standard BI tools, such as Business Objects, Cognos and MicroStrategy.

**Challenges:** A<sup>2</sup>DB is a new product that has only been available for several months and has few customers and references. The initial challenge will be finding customers with the desire to try something very new and unproven in the market. The ease of installation and set-up with automatic generation of database structure is a double-edged sword. On one side it promises simplicity, performance and scalability and on the other, it is almost like magic, causing skepticism among potential customers. Algebraix will need to get several early strong references and will be required to perform a proof of concept (POC) in every situation. This situation, coupled with initial slow adoption will put a strain on the small company.

**Who Should Care:** CTOs, BI data architects and data warehouse architects interested in analytics and BI, especially on mixed-data types and willing to try something really new, can benefit from trying a POC with A<sup>2</sup>DB from Algebraix. We would recommend starting with a small BI or analytic application, expanding from there to multiple data sources and applications.

**Related Research:**

"Magic Quadrant for Data Warehouse Database Management Systems"

## 4.0 Clavis Technology

Dublin, Ireland ([www.clavistechnology.com](http://www.clavistechnology.com) )

*Analysis by Ted Friedman*

**Why Cool:** Data quality issues create pain for organizations in numerous ways. Most contemporary approaches to data quality improvement focus on cleansing data inside authoritative source systems typically managed by IT. In contrast, Clavis focuses on applying rules for data quality control at loosely-governed data collection points — such as in desktop spreadsheets, databases, and forms — which are commonly found at the edges of well-defined business processes for authoring and maintaining master data. Through Web-oriented technologies (such as representational state transfer), Clavis enables business teams to define and insert data quality rules into their data collection mechanisms via a software-as-a-service delivery model (specifically, without having to deploy software on-premises). The goal is to get the data right at the true point of initial capture.

As a recent entrant to the data quality tools market, Clavis needs to differentiate in both its positioning and its technology approach. It attempts to achieve both of these goals by focusing on an aspect of the data quality problem space, which incumbent providers do not and also by leveraging an alternative approach (SaaS) to delivering its technology capabilities. This strategy is distinctive in that it enables Clavis to engage an audience that is hungry for flexible and rapid deployment of data quality capabilities at the edges of the IT-managed systems landscape.

**Challenges:** While Clavis' clearly targets a high-demand area by offering capabilities for embedding data quality controls in common data entry points, market demand for data quality tools seeks a broader range of functionality. Clavis is lacking some of the basic operations commonly required of data quality tools vendors, such as matching/linking/merging and comprehensive data profiling. The somewhat narrow focus will pose a challenge when competing for business with organizations seeking tools suitable as an enterprise standard across all data quality requirements. In addition, as an emerging provider, Clavis' small size, limited customer and experience base and limited partner portfolio, will raise concerns with risk-averse buyers.

**Who Should Care:** Data quality teams, data governance groups and data stewards may find value in Clavis' approach to data quality rules and its focus on data collection points, which incumbent software vendors often struggle to address. In addition, organizations attempting to improve the efficiency and quality of master data definition/authorship processes could consider Clavis to focus on the initial stages of master data setup, which often happens in loosely-controlled formats and structures.

**Related Research:**

"Governance of Master Data Starts With the Master Data Life Cycle"

"Overview for an Enterprisewide Data Quality Improvement Project, 2009"

## 5.0 Cloudera

Palo Alto, California ([www.cloudera.com](http://www.cloudera.com))

*Analysis by Donald Feinberg*

**Why Cool:** Hadoop is an Apache open-source software (OSS) project implementation of MapReduce (<http://hadoop.apache.org/>) for the storage and parallel execution of analytics for both structured and mixed data types with any amount of data (not just for very large amounts of data). As with many OSS products, they are difficult to install and lack support for a production environment. Cloudera has taken the OSS product (offering it as the Cloudera Distribution for Hadoop [CDH]) and wrapped it with a toolset and API to manage a Hadoop cluster, the Cloudera Desktop. Most of its customers set up their own Hadoop cluster, but Cloudera also has advanced Amazon Machine Instances for Amazon EC2 and VMware vCloud API support, to set up a Hadoop cluster in the cloud. Cloudera has also created a suite of tools aimed at enterprise users who want to collect, store and analyze data from a variety of sources. Finally, Cloudera offers subscription support, consulting services and training for Hadoop MapReduce.

Use of Hadoop is growing in commercial organizations (not just in Web-based organizations such as Google and Yahoo). We believe there are many implementations currently in "stealth" mode, hidden within business analytic groups and with little or no support from the IT department. As the use grows within the organization and becomes more mission-critical, there is an increasing need for support and other services. As with many OSS projects, community support is not sufficient. Cloudera brings this to the market with its open-source distribution and tools. In addition, Cloudera is one of the primary contributors to the open-source Apache project.

**Challenges:** Cloudera is young and faces several challenges:

- The Hadoop implementation MapReduce is OSS, which allows any organization to offer support and training and also develop products around MapReduce — causing Cloudera to lose its uniqueness in the market.
- To generate revenue, Cloudera must sell support, training, services and add-on software, as it cannot charge a license fee for the software — making revenue growth slow and more difficult.
- Some DBMS vendors have begun to offer MapReduce functionality built-in to a DBMS engine, removing the necessity of a separate Hadoop cluster. These add up to Cloudera continuing to create unique value compared to newcomers and to those DBMS vendors that integrate Hadoop in DBMS engines.

**Who Should Care:** Organizations with analytics requirements and/or large amounts of data (both structured and mixed data types) to be analyzed will be interested in exploring the potential of Hadoop for the organization. Not only the IT organization (chief technology officers [CTO], BI competency centers and database administration management) should be interested in Cloudera — business units with analysts interested in heavy use of analytics should also take a look at Hadoop and the Cloudera product line.

### **Related Research:**

"Hype Cycle for Data Management, 2009"

"Magic Quadrant for Data Warehouse Database Management Systems"

"Key Issues for Data Management and Integration Initiatives, 2010"

## 6.0 Mastersoft Research

Sydney, Australia ([www.mastersoftresearch.com](http://www.mastersoftresearch.com))

*Analysis by Eric Thoo*

**Why Cool:** Mastersoft Research offers the Harmony Data Quality Series, which uses reasoning engines based on expert systems and inference-based matching to address the challenges that enterprise organizations face with customer information quality. Harmony replicates the judgments people use when making decisions about customer information and is capable of enforcing reliable and consistent results when applying customer information across multiple business contexts. Data quality processes are dynamically constructed from Harmony modules, which can be chained together so that the output of one process is the input to another. Harmony conforms to the Customer Information Quality standard endorsed by Oasis, the leading open information standards body ([www.oasis-open.org/](http://www.oasis-open.org/)). Business processes can access any combination of data quality processes at any point within their applications. Harmony supports real-time (online) and file-based (batch) processing modes for handling of data conditions across data sets.

**Challenges:** Processes to ensure adherence to standards in data semantics and introduction of data governance-specific roles and responsibilities are also part of a holistic data quality requirements. Organizations certainly face challenges in governing the massive amounts of data they hold across a numerous applications data stores, including emergent data types in portals and content repositories that requires governance, which will be increasingly needed to present a complete solution to ensure customer data quality.

**Who Should Care:** Business leaders, CIOs and information architects will be forced to widen their focus on customer data quality improvements as the pressure increases to control and safeguard information assets. In addition, data stewards, customer campaign management teams and compliance officers, should seek technologies, which potentially, could simplify the challenge of establishing data quality controls in the diverse environments they manage.

### **Related Research:**

"Magic Quadrant for Data Quality Tools"

"2009 Survey on Data Quality Tools Highlights Broadening Deployments With Focus on Proven Functionality"

### **RECOMMENDED READING**

---

"Data Management and Integration Maturity Key Initiative Overview"

"Hype Cycle for Data Management, 2009"

## REGIONAL HEADQUARTERS

---

### **Corporate Headquarters**

56 Top Gallant Road  
Stamford, CT 06902-7700  
U.S.A.  
+1 203 964 0096

### **European Headquarters**

Tamesis  
The Glanty  
Egham  
Surrey, TW20 9AW  
UNITED KINGDOM  
+44 1784 431611

### **Asia/Pacific Headquarters**

Gartner Australasia Pty. Ltd.  
Level 9, 141 Walker Street  
North Sydney  
New South Wales 2060  
AUSTRALIA  
+61 2 9459 4600

### **Japan Headquarters**

Gartner Japan Ltd.  
Aobadai Hills, 6F  
7-7, Aobadai, 4-chome  
Meguro-ku, Tokyo 153-0042  
JAPAN  
+81 3 3481 3670

### **Latin America Headquarters**

Gartner do Brazil  
Av. das Nações Unidas, 12551  
9º andar—World Trade Center  
04578-903—São Paulo SP  
BRAZIL  
+55 11 3443 1509