Big Data Comes of Age:
Shifting to a Real-time Data Platform

EMA™

*IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING*

Big Data Comes of Age:
Shifting to a Real-time Data Platform

Table of Contents

Introduction

In 1959, famed business thought leader Peter Drucker coined the term *Knowledge Worker*, defining it "as one who works primarily with information or one who develops and uses knowledge in the workplace."[1] As forward thinking as Drucker was, he might be surprised at the level of sophistication today's data-driven executives are applying to critical business challenges. Google-generation knowledge workers arrive in the enterprise with an expectation of speed, power and ease of use and they are comfortable putting these assets to work on analytic and operational data driven projects.

New data sources are fueling innovation while stretching the limitations of traditional data management strategies and structures. Data warehouses are giving way to purpose built platforms more capable of meeting the real-time needs of a more demanding end user and the opportunities presented by Big Data. Significant strategy shifts are under way to transform traditional data ecosystems by creating a unified view of the data terrain necessary to support Big Data and real-time needs of innovative enterprises companies.

Drivers of Change

It's an exciting but challenging time in the data management world. New platforms are finding their way into our ecosystems. Big Data is a disruptive force causing many companies to totally rethink their strategies for leveraging data across the enterprise, and innovative companies are powering their analytic and operational processes with real-time data, upping the game on execution and insight. Price optimization in retail, content targeting in media and fraud detection in financial services are all areas where real-time is powering new levels of execution.

> **Big Data is a disruptive force causing many companies to totally rethink their strategies for leveraging data.**

As with most paradigm shifts, there are forces at play that are unstoppable, causing acceleration toward change and adoption. This is especially true where Big Data and real-time are concerned. Each of these drivers is timely and when combined, cause significant change.

- **Maturing Users and Applications** – There are clear shifts converging on our analytic and operational environments. Great demands are being put on traditional systems to support the maturing needs of end users and to address the creativity required by a new breed of knowledge worker. A greater complexity coupled with a larger more diverse population of users is taxing the systems beyond their abilities.

- **Economics** – Commodity hardware, low-cost storage and memory are creating an opportunity to address projects that once were beyond the fiscal reach of most companies. The ability to add new purpose-built solutions to the data management landscape is affordable for many and driving a decentralization of the Enterprise Data Warehouse (EDW) in favor of solutions better suited to specific needs.

- **Technology Advancements** – Moore's law[2] is alive and well in the enterprise. The systems available today eclipse the scale and performance of those we invested in just a few short years ago. This ongoing improvement arc is fueling our ability to address Big Data and real-time to create business value.

---

[1] Knowledge Worker, Wikipedia.com, http://en.wikipedia.org/wiki/Knowledge_worker
[2] Moore's Law, Wikipedia http://en.wikipedia.org/wiki/Moore's_law

EMA™

- **Valuable Data Sources** – For years, we have been forced to ignore data sources that could prove valuable to our work processes and analytic insights. The combination of the above drivers now makes it possible to execute Big Data strategies and expand how we accomplish that task. New data sources that include social data, machine generated data and sensor data are all important to powering more complex analysis of our businesses.

Companies who embrace these drivers will realize faster Return On Investment (ROI) and competitive advantage over companies that hold fast to traditional data management practices. Embracing this opportunity will expand capabilities but at the same time cause greater complexity across the data management ecosystem.

> **Companies who embrace these drivers will realize faster Return On Investment (ROI) and competitive advantage.**

Challenges to Innovation

The need to address data integration and management/administration challenges is paramount as the ecosystem becomes more diverse. Utilizing purpose-built solutions creates flexibility but adds to the risk of creating data and compute silos eliminating the environment necessary to execute Big Data driven, real-time workloads. Additionally, as work processes become more complex and reliant on multiple data stores, it will be critical to implement well-planned solutions that unify the ecosystem and make management of the data and systems transparent to the end users.

Making data available for immediate insight requires planning and a strategy to leverage the best possible platform(s) for success. Often times an investment has been made to place data within a certain platform that makes it difficult or impractical to move (i.e., cloud or Hadoop) so decisions of where and how the computing aspects of these processes are made become more critical within a complex ecosystem, building the case for data integration and management services designed to work across the ecosystem, not just with a single element of it.

Traditional data management tools such as Extract-Transform-Load (ETL), data replication and data virtualization will play a role, but more comprehensive solutions will be required to deliver the value, agility and flexibility required within a diverse platform driven ecosystem designed to address Big Data the real-time business.

Big Data Value

Our ability to manage and leverage Big Data to fuel complex workloads is growing in importance as operational processes demand real-time data to deliver value. Financial instructions, for example, are leveraging multiple data systems to power next-best offers solutions to increase customer-based revenue. These workloads often leverage customer data from an EDW, real-time clickstream analysis from corporate websites, credit information from third-party vendors, and sentiment data from call centers, enabling organizations to match customers to robust product offering catalogs.
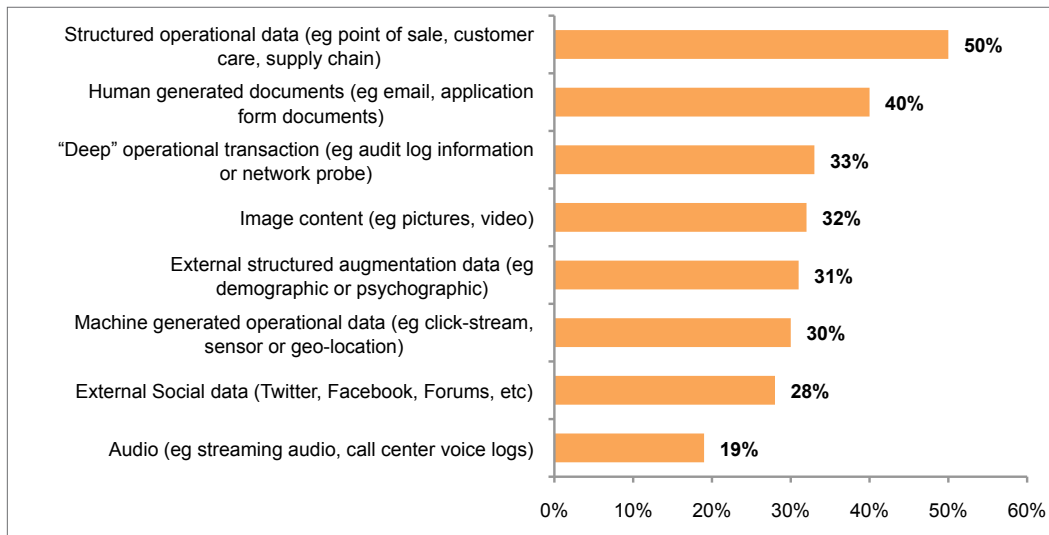
Adoption

Big Data adoption is moving at a fast rate. Recent ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) research shows that 36% of study respondents are already in operation with a Big Data project, 35% are planning to launch projects within the next six months and the remaining 28%
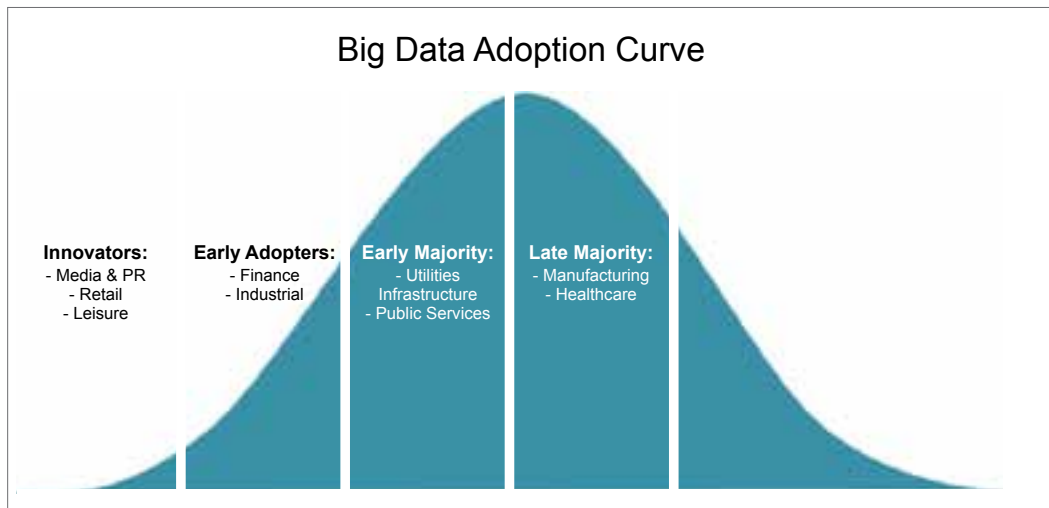
are researching and investigating the technology. The research further illustrated that analytic workloads were top of mind with end users – the top response showed that 42% of users intended to use Big Data as a primary operational platform. This speaks directly to the growing need to power operational systems with fast, real-time data.

EMA research shows that 36% of study respondents are already in operation with a Big Data project,

As discussed, today's data sources are widely dispersed and varied in content. Sophisticated applications and processes require the flexibility to tap into these sources easily and with confidence. EMA research shows a wide range of sources in use, lead by structured operational data and including social, machine generated, image and deep operational transaction data.

Data Source	Percentage
Structured operational data (eg point of sale, customer care, supply chain)	50%
Human generated documents (eg email, application form documents)	40%
"Deep" operational transaction (eg audit log information or network probe)	33%
Image content (eg pictures, video)	32%
External structured augmentation data (eg demographic or psychographic)	31%
Machine generated operational data (eg click-stream, sensor or geo-location)	30%
External Social data (Twitter, Facebook, Forums, etc)	28%
Audio (eg streaming audio, call center voice logs)	19%

EMA research showed that when segmented by industry, traditional innovators and early adopters included Media/PR, Retail, Leisure/Gaming along with finance and industrial segments.

Big Data Adoption Curve

Innovators:
- Media & PR
- Retail
- Leisure

Early Adopters:
- Finance
- Industrial

Early Majority:
- Utilities Infrastructure
- Public Services

Late Majority:
- Manufacturing
- Healthcare

Hybrid Data Ecosystem

Throughout this paper many references have been made to the new or evolving data management landscape or ecosystem. Enterprise data warehouses continue to play an important role in this arena, but at the same time they have migrated away from being the center of the data management universe creating a Hybrid Data Ecosystem (HDE) that is more flexible and better prepared to meet the evolving needs of enterprise users and Big Data projects. At the center of the system are the requirements for determining which of the platforms or combination of platforms will serve the enterprise best. Big Data can be found across the entire field of solutions.

Enterprise Management Associates – Hybrid Data Ecosystem™

Response – The need for platforms to respond at new speeds and scale has opened the door for new ways to leverage data and provide insights to end users. This is especially true in the area of Big Data analytics where response rates are a key component to the value these platforms can deliver. Sub-second data delivery is not necessary for all applications and data-driven scenarios, but it's clear that real-time use cases are growing in importance and becoming more critical to many companies. Platforms such as Big Data Frameworks, analytic databases and appliances are part of this evolution and powering new solutions with improved response.

> **The need for platforms to respond at new speeds and scale has opened the door for new ways to leverage data and provide insights to end users.**

Analytics/Workload Complexity – Addressing the requirements of complexity within analytic environments is getting more challenging while running highly complex analytic models over massive data stores is becoming more commonplace. When coupling the *workload complexity* to the *response*,

selecting the best platform can create powerful tools of differentiation. The ability to introduce new data types such as social information or machine or process data can be leveraged to add even greater levels of insight and value.

Economics – The economics of technology is the great equalizer and often can attribute to an early majority adoption of the technology. This has been especially true with Big Data. Many companies have identified needs to address *response* and *workload complexity*, but the return on investment has slowed adoption. Big Data platforms are leveraging commodity hardware, and often the software is free so it breaks through the economic barrier to adoption. Companies that plan to adopt Big Data should be warned that the barrier to entry is significantly reduced, but that doesn't mean it's cheap. Special skill sets are requirements, and lack of mainstream management tools create hidden costs that need to be taken into account before adopting this type of technology.

Structure – Flexibility of structure is a growing decision point for selecting Big Data platforms. Big Data frameworks provide a level of flexibility not present in traditional data platforms. These systems can load and store data without requiring the time investment of designing and building complex data models. Analytics can be executed on these platforms without models and while running at speeds that eclipse many standard relational databases. Many users are employing "late binding" models to the data as they move it forward in the analytic process enabling a smaller set of data to be manipulated and leveraged. It is at this point the data is often moved or accessed by another system designed to execute complex analytics or to provide data to operational workflows.

Load – Data loads are growing more complex and the sources more diverse. Driven by greater complexity and demand, Big Data adoption is driven by the need to provide flexibility. The power of Big Data platforms to load a mixture of data creates an opportunity to address both analytic and operational scenarios. Without this data to fuel these workloads, it would be impossible to execute against the growing demands of enterprise applications and analytic environments.

Including these five requirements in your planning for Big Data platforms will ensure that you select the best possible solution or combination to execute your Big Data strategy. It's critical to map out the analytic process to determine which platform(s) will deliver on the *response* required, where best to execute *complex workload*, which platform(s) presents the best *economic* advantage, how best to leverage the data *structure,* and lastly, how best to support the *load* of the data.

Strategies for Managing the Ecosystem

Now that we have discussed the five requirements used to select the platforms to be included in a Hybrid Data Ecosystem (HDE), organizations begin to turn their attention to delivering the data within it and managing the overall ecosystem to create the most flexible environment possible. Most companies who embrace a HDE as a foundation for their data management strategy will utilize ETL, data replication and data virtualization as cornerstone tools to address the needs of managing data in a HDE and delivering for operational or analytic workloads. EMA research shows 63% of companies surveyed are already working with three or more platforms they need to employ a strategy that moves beyond these traditional tools and identify solutions that are designed to manage the systems and data needs of the entire ecosystem. Best-of-breed solutions will add functionality at this level to help ensure data quality, manage master data, and govern data between platforms within the HDE.

EMA

Companies who have moved beyond three platforms will also look to solutions that offer system-wide metadata management, the opportunity to parse workloads so data doesn't have to be moved or replicated, allowing the best systems to handle the compute aspects of complex workloads and then stitch them together for final delivery. Lastly, companies looking to extend their investments will utilize late-binding schemas to leverage data as a resource and avoid the time and cost involved with transforming data as it enters the ecosystem.

EMA Perspective

There is no denying that Big Data and the growing demand for real-time insights are causing disruption in the data management space. Innovative companies are aligning to new strategies and aggressively matching data, workloads and purpose-built platforms to enhance speed, agility and insight. These HDEs create opportunities along with challenges. To create an environment designed to leverage these new investments, it's critical for users to adopt sophisticated data integration and system administration strategies that will alleviate the complexity and deliver analytic and operational value.

> **There is no denying that Big Data and the growing demand for real-time insights are causing disruption in the data management space.**