Big & Fast Data: The Democratization of Information

Moving from the Enterprise Data Warehouse to the Business Data Lake
# Table of Contents

Forewords 04  
Steve Jones, Vice President for Global Big Data, Capgemini Insights & Data 04  
Paul Gittins, Global Alliance Executive, Capgemini Global Channels & Partners 05

Executive summary 06

What organizations want to achieve with data-derived insights 07

Setting the strategic direction but delivering incrementally 07

The need for the democratization of data and analytics 08

Yesterday’s technology falls short of today’s aspirations 09

Emerging approaches enable the insight-led organization 11

Benefits of the BDL 11

How it works 12

Speed of adoption is of the essence 12

Architectural view of the Business Data Lake 13

Use case 1: Internet of Things 14

Use case 2: Transforming operational reporting 15

Use case 3: Anomalous behavior detection 15

IT’s new opportunity as the data broker to the enterprise 16

7 guiding principles 17

Conclusion 19
Capgemini has always taken a business rather than technically focused approach to information. With the Business Data Lake Capgemini and Pivotal started from a simple question: what does the business want to see?

From that question came recognition that traditional approaches were more about IT cost control than delivering business value.

What is clear from the report is that many companies are already starting a transformation and seeing significant new business value from breaking out of traditional information silos such as the enterprise data warehouse. The majority however are suffering frustration with their current approaches, most notably the delivery methods of IT, and are not realizing the value that they say is there.

This report outlines many of the challenges companies are facing and solutions to removing that frustration in order that they can transition to a new information landscape - one which works in harmony with the business.
Is it possible that ubiquitous analytics represents the next phase of the information age? New business models are emerging, enabled by big data that business leaders are eager to adopt in order to gain advantage and critically, to mitigate disruptive threats from startups and parallel industries. The winners are likely to be those that master a cultural shift as well as a technology evolution.

Our view is this will be realized through the alignment of a business-centric big data strategy, combined with democratization of the analytical tools, platforms and data lakes that will enable business stakeholders to create, industrialize and integrate insights into their business processes.

Innovative approaches are needed to free up data from silos whilst encouraging both the sharing and the continuous improvement of insights across the business. While it will be evolution for some, revolution for others; the risk of status quo is not just the loss of opportunity but also a widening gap between business and the internal technology functions.
Executive summary

Data-driven insights, enabled by big data and integrated into operational processes, are delivering four key types of business value:

- Efficiency and cost focus
- Growth of existing business streams
- Growth through market disruption from new revenue streams
- Monetization of data itself, with the creation of new lines of business

Underpinning these new models are new technology capabilities that allow businesses to deliver value in line with their strategic goals. In practice, this means democratizing access to analytics so that decision-makers at all levels can create and enhance the insights that they need, posing their own “what if” questions to a rich substrate of data.

Our recent research¹ confirms that current business intelligence technology such as the enterprise data warehouse (EDW) is not able to provide these capabilities. At best, only the most senior decision-makers have access to the insights that they need. More often, everyone is limited to structured data created within a fragmented set of spreadsheets, data marts and warehouses.

The research shows that the traditional lengthy development cycles for new insights from IT based on EDW architectures has led to businesses not only questioning those traditional processes but also bypassing IT entirely in their quest to deliver new business value from analytics.

The Business Data Lake (BDL) represents a new approach to the creation of analytical insights for the business, from the acceleration of traditional enterprise reporting through to new analytics driven by data science. The BDL works with high volumes of structured and unstructured data, storing them at low cost and making insights rapidly available throughout the enterprise. It can coexist with earlier investments, accelerating the evolution of the information landscape.

By introducing the BDL, and moving towards a more iterative and agile approach to analytics and insight delivery, CIOs can evolve their information landscape to reposition the IT function as the business’s “insight enabler”. As a result, decision-makers at all levels of the business will be able to seize the opportunities of big and fast data more rapidly and so gain competitive advantage.

¹ Big & Fast Data: The Rise of Insight-Driven Business, 2015
What organizations want to achieve with data-derived insights

In our recent report Big & Fast Data: The Rise of Insight-Driven Business, we discussed how data-driven insights are changing businesses, and identified the four key ways businesses are looking to drive new value from analytical insights:

1. **Efficiency and cost reduction**
   By leveraging both real-time data and historical information predictive analytics and adaptive insight-driven processes can reduce bottom line costs and drive operational improvements. These can provide increased margins or reduced cost to market. An example would be the use of predictive maintenance and intelligent workforce scheduling to reduce down-times and minimize service visits.

2. **Growth of existing business streams**
   New analytics is changing the way businesses engage with customers. Top-line growth and increased margins can be achieved by more targeted insights that select, target and support key customers. For instance, better insight in retail marketing and couponing can result in not only better direct impact but also secondary impacts through the use of social media distribution.

3. **Growth through market disruption from new revenue streams**
   Analytics and insight can deliver new ways to sell existing products. Here the service model is changed and so the economics of addressing the market. The shift of the aero engine industry away from selling engines to providing engine hours is an insight and data driven shift.

4. **Monetization of data itself, with the creation of new lines of business**
   As data volumes increase and analytics becomes the differentiation, so the data itself becomes a valuable business commodity. Companies like Facebook, Twitter and Google are fundamentally driven by their data and it is this that underpins their financial valuations. The ability to find value from data and monetize it provides dramatically different economic leverage in an insight driven world.

To realize these new opportunities, the business needs technology capabilities that can deliver insights to support both the organization’s overall strategy and the demands of individual lines of business. The majority of businesses are adopting an incremental approach, as described in the panel.

---

**Setting the strategic direction but delivering incrementally**

A new mindset is emerging that involves quickly building proofs of concept, “failing fast”, and then, where value is found, scaling rapidly – in weeks or months – rather than launching multi-year programs. Execution takes place in three phases:

1. **Proving value**: We find the most successful programs start by addressing a few real business use-case opportunities. This creates a much stronger dialogue between IT and business. Even at this point it is important to understand the path to scaling out the capability for the business, otherwise these become dead end initiatives. (Some programs try to shortcut this phase by performing a technology evaluation. In our experience this approach actually increases time to value, as the business is less engaged and cannot attach real business value to the outcome.)

2. **Expansion to pilot**: Companies often choose an entire line of business for migration to the new environment, with a focus on ensuring scalability, performance and adoption.

3. **Enterprise adoption and uptake**: Some companies start to migrate specific business units or functional areas, with a focus on expanding use cases and enhancing platform capabilities.

What sets the successful programs apart is that they take this business-centric incremental approach within the framework of a clear strategy, ensuring a long term focus on moving towards a converged information landscape which is aligned with long term business objectives.

Source: Big & Fast Data: The Rise of Insight-Driven Business, p15
The need for the democratization of data and analytics

In determining the right technology strategy to realize the types of opportunities identified in the preceding section, business users are asking key questions:

- How do I align the pace of insight creation – the “time to insight” – to the pace of my business opportunity?
- How do I integrate the insight into my decisions to drive an appropriate action?
- How can I ensure I am not constrained by the insights that other areas of the business are trying to obtain?
- How do I industrialize and deploy these insight services across the business and minimize the time to value?

It is not usually realistic to expect IT alone to be able to respond to these needs individually. Historically, requests to IT for mechanisms to analyze data in new ways just take too long to fulfill. The constraints of single-schema EDW models mean that by the time the insight is available the opportunity has passed.

To overcome these limitations, instead of just delivering insights to business functions on request, IT needs to provide business users with first an increasingly rich data substrate that they can access to create insights of their own, and second analytical tools and services that vastly expand the analytical capacity of the organization.

By providing these two things IT enables and supports the evolution of the way the business uses data and insight:

- Business areas can focus on the distillation they want to match their needs, rather than all parties needing to agree on a single enterprise wide schema before getting any value. These business area focused views match their challenges and perspectives, without having a multi-year program to gain alignment on the meaning of each data type being used in the business.
- The business can move beyond viewing history to adopting new capabilities: real-time reporting, predictive analytics, and machine learning. These new capabilities can move far beyond the traditional EDW, with in-transaction response times or complex mathematics that require fundamentally different approaches from those that simply report the facts of the past.

With a data substrate and matched analytical services in place, lines of business can interrogate wide data sets using a broad choice of analytical tools, asking “what if” type questions, predictive analytics, machine learning as well as traditional linear reports. They can leverage existing insight models that have been enhanced with previous insights or create new ones, both at a pace that enables them to seize opportunities as they emerge, without being constrained or rushed by the pace of other units’ needs.

This approach is democratic in two senses. First, data is directly available to the business with minimal IT intervention. Second, decision-makers at any level of the business can access all the information they have permission to receive. This moves away from a central, planned, committee-based approach that dictates the way data is organized and limits the variety of new analyses that are possible.

The result is a more agile business that is better poised to build competitive advantage.
Yesterday’s technology falls short of today’s aspirations

Democratizing insight is not a totally new idea. Ideas like self-service business intelligence have been around for a long time, and there were high hopes that the EDW would make business users less dependent on the IT function for their insights and more able to interrogate the data for themselves.

Traditional solutions provide various levels and types of analyses of structured data, but are not designed to handle either the volumes or varieties of data that are currently available to the business.

Fundamentally, these older solutions were designed to report on data after the fact: to show what has already happened. Today’s challenges are to report on what is happening now and what will happen in the future.

As new data sources, especially unstructured data, become important to the business, the restrictions lead to breakaway business solutions that can better represent what the business wants to see.

One provocative view is that it is not surprising that the EDW has failed to achieve democratization. Its goal was to be the single place for the end-to-end enterprise information capability, empowering individuals across the organization.

The reality, however, is that EDW-based approaches are biased towards committee-based decision-making. In an EDW, there is a single schema defining the data; experience shows that this definition usually ends up serving the needs of a narrow subset of the organization, while the rest generally have to make do with views of the data that represent a significant compromise of various conflicting or misaligned requirements.

The compromise of data views in enterprise environments

The reality in the business is that analytics to date is built from data that has a fundamental compromise: it is of uncertain veracity; the real-time data that is increasingly required for decision-making is not available; and insights are not connected to business actions.

The 4 key Dimensions of data use in the enterprise

- **Fit**: How fit for purpose is my view of data?
- **Detail**: How much granularity is available in my view?
- **Freshness**: How recent is the data? +24hrs = 20% less useful
- **Fidelity**: How much compliance & quality has been applied?
Our study found that current approaches are not achieving what the business wants and are not delivering to the level that today’s information-centric business requires. The responses clearly indicate that traditional approaches cannot provide the complete answer in a world increasingly dominated by a high diversity of structured and unstructured data, much of which is not directly under the control of the organization.

There is a clear divide between senior and operational management: Of those organizations that have not yet implemented big data technology, 62% consider the data analytics provision to be either good or very good for senior management, but only 43% feel the same way about the provision to non-management employees. Supplementary solutions are widespread, with 72% using local business intelligence tools and 63% using local data marts.

This does not just suggest that effort is being duplicated and data used in inconsistent ways; it also implies that people are often working with different and out-of-date snapshots of data. This inconsistency often drives contradictory decision-making and poor operational efficiency. The use of delayed insight also represents a challenge, with 77% of our respondents saying that decision-makers increasingly require data in real time.

The need for advantage through insight, operational efficiency, fast data, and business agility is beginning to fundamentally transform how businesses approach information. A batch-fed EDW cannot be the whole story when insight needs to be delivered to the point of action.

FIGURE 1: Traditional approaches are constrained in a data world

1. The volumes of data are exploding
2. The ability to control and dictate in an “outside-in” world is minimal
3. More and more business value is beyond the core transactions
4. The old approach of “a single view” is impossible in a world of federated internal and external data
Emerging approaches enable the insight-led organization

The BDL can provide the democratization that businesses need using the latest in big data technology. It is designed to overcome the limitations of traditional business intelligence (BI) systems and provide an evolutionary approach that helps organizations augment and transition to a data centric business.

The BDL provides a way forward with a complete rethink of the way data is ingested, stored and organized for analytical purposes. It democratizes insight by giving the business the capabilities it needs to seize opportunities rapidly, providing a common analytical engine for the business.

Benefits of the BDL

This new approach to enterprise reporting and information management has fundamental advantages. It:

• Enables a focus on local requirements, whether from the board or the warehouse manager
• Provides access to everything by storing it in a single connected substrate
• Provides key insights from data where and when it’s needed
• Enables the appropriate level of governance to accelerate business collaboration

Essentially, the BDL approach gives the CIO and IT function the ability to align and accelerate the digital strategy for the business. As we found in *Leading Digital*

How it works

The BDL approach is illustrated in the diagram on the next page. Its basic principles are:

- **Ingest** any kind of data at scale, structured or unstructured data sets – from any internal, partner, IoT or open data
- **Store** for both near real-time and long-term analysis at the lowest possible cost
- **Analyze** in batch or real time; complement with data science tooling
- **Surface insight**, i.e. bring the insight arising from analysis into management tools, together with its context
- **Take action** and drive action back into the system for future improvements

The BDL supplements and incorporates data from existing systems such as the EDW, rather than requiring them to be replaced. If an EDW is satisfying its current data capture and base reporting requirements, a BDL can complement to provide to more powerful predictive analytics engine.

Insights from the BDL can supplement traditional, more linear, reporting from the EDW. Including the EDW as one of the corporate views within a BDL positions IT to help the business get what it needs in any situation.

The BDL draws information from existing systems and legacy data solutions and provides a platform on which new insight driven services can be developed and integrated back into operations or turned into entire new data monetization business lines. It demonstrates a lower risk, augmentation driven approach to the next generation of insights.

**Speed of adoption is of the essence**

The BDL enables organizations to progress much faster towards their goals than conventional approaches. To gain competitive advantage, they need to start getting acquainted with it sooner rather than later. Respondents in our study quickly spotted the potential: 61% state that the BDL architecture would be relevant to their organization, with 29% strongly agreeing this is the case.
Architectural View of the Business Data Lake

The Business Data Lake is the first unified architecture for big and fast data and the only one designed for driving third-generation data platforms.

The figure shows the key tiers of a Business Data Lake. Data flows from left to right. The tiers on the left depict the data sources, while the tiers on the right depict the integration points where insights from the system are consumed. Key tiers are:

**Storage**: Ability to store all (structured and unstructured) data cost efficiently in the Business Data Lake – Hadoop (HDFS)

**Ingestion**: Ability to bring data from multiple sources across all timelines with varying quality of service (QoS)

**Distillation**: Ability to take data from the storage tier and convert it to structured data for easier analysis by downstream applications

**Processing**: Ability to run analytical algorithms and user queries with varying QoS (real-time, interactive, batch) to generate structured data for easier analysis by downstream applications

**Insights**: Ability to analyze all the data with varying QoS (real-time, interactive, batch) to generate insights for business decisioning

**Action**: Ability to integrate insights with business decisioning systems to build data-driven applications.

**Unified data management**: Ability to manage the data lifecycle, access policy definition, and master data management and reference data management services

**Unified operations**: Ability to monitor, configure, and manage the whole data lake from a single operations environment
Use case 1
Internet of Things

The internet of things (IoT) is about connecting smart devices to the internet so that they can stream data, improving support and enabling better decision-making. This raises two key challenges. First, the speed at which these devices need to react and be updated – it’s no good telling a car about traffic on its route only after it has reached its destination, or tuning a wind turbine after it has broken. The second challenge is that of volume. With potentially millions of devices in circulation, each generating information “beeps” every second, the volume of data can quickly become prohibitively costly to handle within traditional databases.

The BDL provides key services needed to deliver business value from this data “firehose”, as it has been called. The BDL can ingest large volumes of information at a fraction (a tenth or less) of the cost of traditional technologies. It can handle high volume streams of information from devices. It can carry out predictive analytics based on historical data. And it can use machine learning to drive adaptive analytics in real time.

With these foundations in place, you can operate at the pace of the “thing” itself to create the right actionable insight – whether that is focused on cost efficiencies in a wide grid of sensors deployed across a country, or building new lines of business based on health data from a wearable device.
Use case 2
Transforming operational reporting

Operational reporting has long posed a challenge for enterprises: how to meet local reporting needs while focusing on an enterprise context. The BDL resolves this conundrum, because there is no longer any need to choose between operational and enterprise views. Instead, you just need to find the right view of data to meet a given business need.

The BDL provides rapid analytics and mobile BI support as well as predictive analytics capabilities. These give operational reporting the power of modern BI and analytics in a platform that can be leveraged across the business.

For example, in a large consumer goods company, the head of marketing in Malaysia may need to look at just that part of the market. The local distribution center will need different insight from the same data to optimize stock levels. All of this information will also be used at the Asia and global levels for other purposes. With the BDL in place, all of them will be deriving insights from the same source information sets within a single data substrate. This means they are aligned, but not constrained.

Use case 3
Anomalous behavior detection

The BDL approach lends itself to anomalous behavior detection, addressing IT security issues that conventional methods can’t. It can help to detect and prevent theft of data or intellectual property (IP), for instance at the behest of nation states or organized crime, or by a disenchanted employee.

The solution can quickly identify when a user is behaving in a way that is abnormal for them and take appropriate action to limit what they can do, or flag up the situation for managerial attention. It can also predict when anomalous behavior is likely to occur, flagging events of interest for further investigation for potential security breach.
IT’s new opportunity as the data broker to the enterprise

Over a third (36%) of the most senior (C-level or director-level) non-IT decision-makers taking part in the study say that their business unit has circumvented IT in order to carry out the data analytics it requires. This internal and external “shadow IT” underlines the point that the IT department is already becoming marginalized.

In the early 90s, IT managers who focused on minicomputers were ousted because they didn’t see the PC revolution coming; this revolution was led by business instead. The same thing is happening again. Software as a Service (SaaS) and managed platform solutions for big data are providing business leaders with ways to directly procure insights without involving their IT department.

Why are so many business leaders bypassing IT and going outside for solutions? Our study identifies two key reasons. First, nearly half (47%) of respondents agree that their organizations’ IT systems are not optimized to enable business decision-makers in all departments to effectively do their jobs.

Second, there is a general feeling that everything is taking too long. About 45% of respondents complain that the current development cycle for new analytics is too long and does not match their business requirements. Over half (53%) consider the speed of their organization’s insight generation to be constrained by its IT development process.

All these are clear signs that IT runs the risk of being seen as a barrier to business change instead of an enabler. CIOs must act now to reverse this perception as history has shown us that the CIO as blocker of business innovation tends not to survive very long.

Instead of struggling to meet requests from business users, IT has an opportunity to become a broker of data and insights. Data comes from inside the traditional internal transactional systems but also increasingly from external sources (e.g., social media, open data, internet of things) and from implicit ones (e.g., log files, internal security, email, documents).

The key is for IT to put in place technologies and delivery methods that enable effective democratization of insights in a way that last-generation approaches such as the EDW have failed to do, meeting the increasingly pressing need for fast time to insight. In fact, more than half (54%) of respondents stated that they consider leveraging fast data to be more important than leveraging big data. In this way IT can help the business to realize the available opportunities.

For the IT function, the opportunity is to reposition itself as a data broker at the same time as it democratizes the creation of insights. In this new role, it will provide end-users with business-centric platforms and platform-based services including:

- An enterprise-wide data substrate that ingests and manages the widest variety of data.
- Big data foundations (Hadoop, MPP SQL, in-memory, streaming, etc)
- Higher-order tooling such as data quality, management services (master data management/reference data management. The focus becomes governing data where it needs to be shared, a shift away from a single-schema approach with its inevitable compromises, towards governing the data that needs to be shared.
- Platform consumption by the end users “as a service”, which includes infrastructure, capacity, identity management, etc.
- Self-provisioned business intelligence, data science tools, and access to as many previously created insights as possible.
Our guiding principles towards democratized analytics

<table>
<thead>
<tr>
<th>Guiding Principles</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Embark on the journey to insights within your business and technology context</strong></td>
<td><strong>Enable your data landscape for the flood coming from connected people and things</strong></td>
<td><strong>Master governance, security and privacy of your data assets</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Develop an enterprise data science culture</strong></td>
<td><strong>Unleash data and Insights-as-a-Service</strong></td>
<td><strong>Make insights-driven value a crucial business KPI</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Empower your people with insights at the point of action</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**The journey towards an insight-led enterprise**

**We see the following seven principles as central to making the necessary transformation:**

**Principle 1: Embark on an insights journey, respecting your business and technology context**

The starting point must be your business objectives. Design your roadmap to harness new data sources based on how they will help achieve these objectives. Equally importantly, your journey must be dictated by where you start, in terms of data maturity but also technology.

**Principle 2: Organize the new data landscape for a data flood from connected people and connected things (IoT)**

There are many new technologies that enable the capture and management of the data flood. Your new data landscape should be a mixture of these technologies, chosen to provide the right solution in terms of cost, flexibility and speed to suit each specific data set and to meet the insight needs of the business.
Principle 3: Create a pervasive data science skillset and attitude

Data science unlocks the insights. It needs to become part of the culture of the organization. Only by embedding it throughout the enterprise, and systematically making all decisions better informed, can organizations achieve the transformation to becoming insight-led.

Principle 4: Unlock Data- and Insights-as-a-Service

The demand from business users for information and data-driven insights is ever increasing in virtually all organizations. To harness this, business users must feel that they can rapidly access the information they need where and when they need it.

Principle 5: Introduce ROI of insights as a crucial business KPI

Measure your measurement: apply data science to your data science to see where you are adding value and where you are not. If data is becoming one of your most valuable assets then treat it as such – include it in KPIs and business reviews.

Principle 6: Master the governance, security and privacy of your data assets

Insights from unreliable data are worse than no insights at all. Equally, programs fail and businesses leave themselves exposed if data is not handled securely and with consideration of relevant privacy issues.

Principle 7: Empower your people with insights at the point of action

Ultimately all organizations are a series of decision points, both at the macro and micro level. Empowering your people with timely insights that make each of those decisions just 5-10% better will transform your business.

Through the application of these principles, we are seeing organizations start to generate real value. In our study with MIT, we demonstrated that only when technology adoption is accompanied by transformation management (including vision, business engagement, and organizational change) do businesses realize tangible benefits in terms of profitability and market valuation. This is as true for big data technology implementation as it is for any other technology.

Conclusion

Our study found business leaders saying they need better, faster and more focused access to analytical insights and that traditional solutions and methods are not delivering what they need. IT departments need to democratize the way they deliver insights to decision-makers if they are to enable the business to realize the opportunities presented by big data.

IT must provide the necessary capabilities, both in order to help the business exploit the new generation of insight-driven business opportunities and to ensure that the information landscape evolves to meet the increasing expectations of the business.

By delivering a next-generation business-centric information infrastructure, the CIO can move fast to make all this happen. Building an open, shared analytics platform for the enterprise enables rapid delivery of insights. The new solutions can be integrated with legacy platforms, or new platforms built to embrace new data monetization opportunities.

For the next generation of data-driven insights for the enterprise, the Business Data Lake represents an opportunity for the IT function to become the true data broker for the enterprise – by modernizing the data landscape and truly democratizing analytics across the enterprise, allowing the business to adopt entire new business models at pace.

About the study

This study draws on research conducted by FreshMinds on behalf of Capgemini and EMC. It combines the results from a quantitative online survey and supplementary in-depth qualitative interviews. About 1,000 senior decision-makers from across nine industries and 10 countries worldwide took part in the online survey. We would like to take this opportunity to thank all respondents for their valuable time and contributions.

FreshMinds is an award-winning insight and innovation consultancy. FreshMinds helps global organizations drive growth by discovering market opportunities, developing game-changing products and services, and creating captivating launch campaigns. Driven by a belief that businesses need to become more agile in the face of digital disruption, FreshMinds delivers agile insight that helps organizations get further, faster, with confidence. www.freshminds.net
About Capgemini

With almost 145,000 people in over 40 countries, Capgemini is one of the world’s foremost providers of consulting, technology and outsourcing services. The Group reported 2014 global revenues of EUR 10.573 billion. Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

For further information visit
www.capgemini.com/insights-data
www.pivotal.io/big-data/businessdatalake

or contact us at
insights@capgemini.com

The information contained in this document is proprietary. No part of this document may be reproduced or copied in any form or by any means without written permission from Capgemini. ©2015 Capgemini. All Rights Reserved. Rightshore® is a trademark belonging to Capgemini.