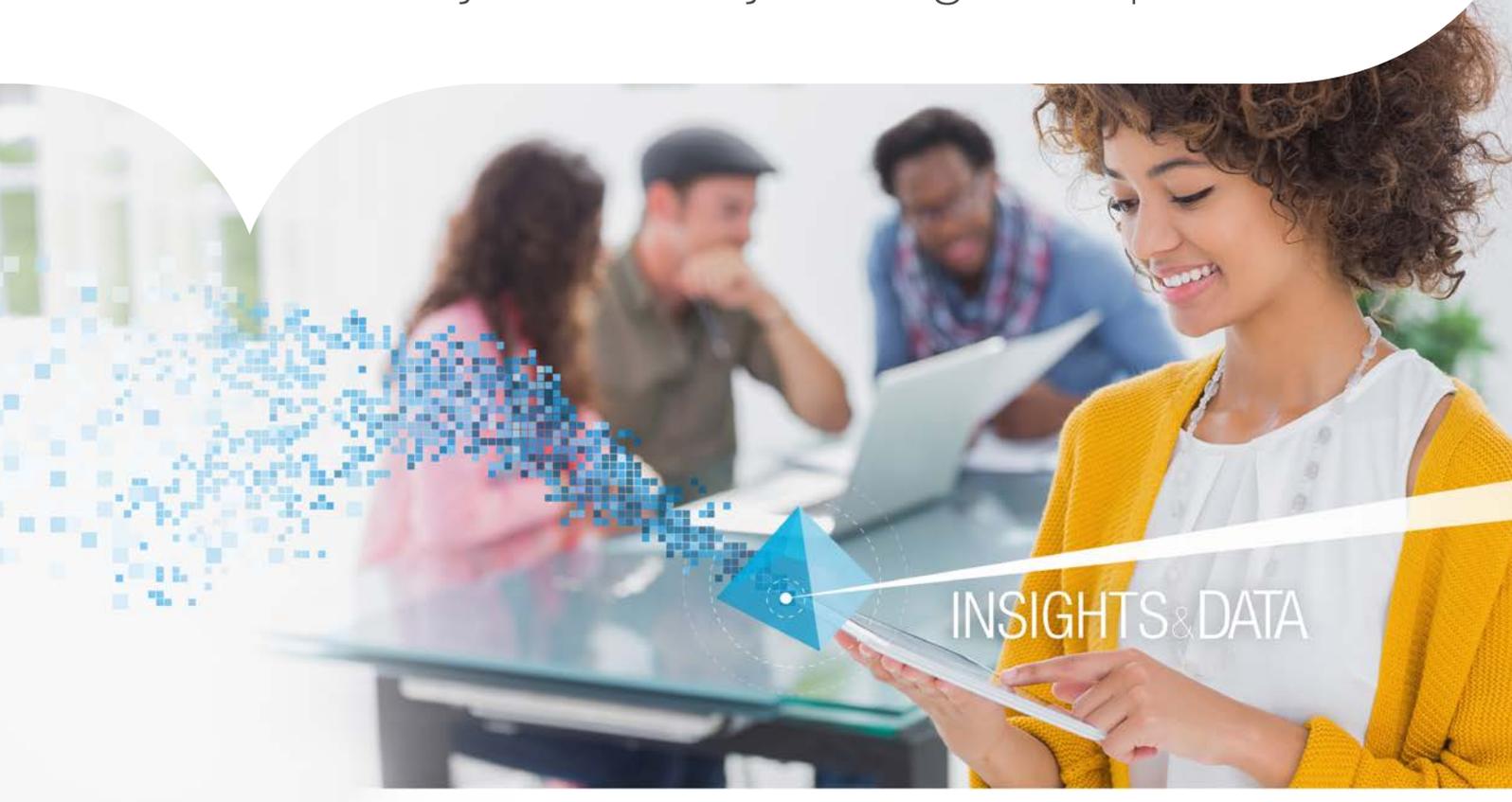


# The Business Data Lake: delivering the speed and accuracy to solve your big data problems



**Get rid of your “data swamp”. The Business Data Lake makes it easy to obtain full business value from all your enterprise information. You can straightforwardly govern and manage data, and rapidly distill all the views that the business needs, refining as required.**

## **High-quality, accessible information is the key to business transformation**

A business's success or failure in the digital age depends on its use of information – and specifically on its ability to deliver insights at the point of action based on all the available data. This means that ease of access to information, and the quality of that information, are more critical than ever.

Organizations therefore need new enterprise data management platforms that create maximum value from data. The platforms must be able to combine existing data resources with new types of data, both structured and unstructured – including data associated with social media, documents, and the Internet of Things. They must make the combined data available for both fast real-time analytics in operations and strategic predictive analytics.

With these platforms in place, your company can use all its data assets, whatever their size, to facilitate business transformation and obtain competitive advantage. You can derive the information you need to identify the new services and products that your customers really want, and hence create new revenue streams.

## Mastering, Distilling and Refining information

Most companies are well aware of new big data technologies that offer innovative capabilities at a fraction of the cost of traditional technologies. However, until now there have been some obstacles to adoption.

The first is that to gain value from information you must be able to govern it. This means putting in place master data management (MDM) and data governance to enable key pieces of information (products, customers, locations, etc) to be uniquely identified across all systems so that, for instance, the customer prospect list from one source can be combined with the invoice history and Twitter stream.

The second challenge is the lack of a business-friendly distillation capability – that is, of an easy way to extract what is needed from shared data for a specific business challenge or channel and present it in the required form. Because of this lack, it has been difficult for companies to make integrated data solutions and consistent information available to all without sacrificing the focus, and the flexibility, associated with line-of-business solutions such as spreadsheets.

The final challenge is that big data solutions need to provide an option to refine information as part of the distillation process. This involves leveraging MDM, and adding in further quality, security and data lineage capability. These cannot be adjuncts to a big data environment: Instead, they must be integral parts of the distillation process. Otherwise, the solutions can exacerbate existing problems instead of resolving them.

These limitations of first-generation data lakes have been summed up by Gartner: “By its definition, a data lake accepts any data, without oversight or governance. Without descriptive metadata and a mechanism to maintain it, the data lake risks turning into a data swamp. Without metadata, every subsequent use of data means analysts start from scratch, like a form of data amnesia.”<sup>1</sup>

To overcome these challenges and limitations, Capgemini, Informatica and Pivotal have brought the disciplines called for by Gartner together within a platform that is flexible and business-focused: the Business Data Lake (BDL)<sup>2</sup>.

### The BDL solution: overview

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. As the first data lake architecture to combine data integration, distillation and refinement for a fully governed big data platform solution it stops data lakes from becoming data swamps and makes them instead into assured solutions for businesses.

The BDL represents an evolution away from the “one size fits all” attempts of the previous EDW generation and towards a

## 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.

more business-centric approach to data. Its key principles are that information must be governed, but governed in line with the business, and that it is crucial to enable the local view to be distilled.

This unified approach to information management, reporting and integration helps businesses leverage all their organizational data in a way that makes sense to individuals and business units rather than constraining it within a single enterprise view. It addresses the requirement for consistency and collaboration across the enterprise, but also recognizes that insight has the most impact when it is specific to an action or decision.

Users therefore get a business-centric view of information, while all the data, structured and unstructured, is governed consistently organization-wide. This means you can realize the power of all your information while also retaining the flexibility of traditional desk-based spreadsheets. You can deliver the specific information that is required for a specific challenge, rather than trying to force everyone to use the same perspective.

### How the solution streamlines refinement and distillation

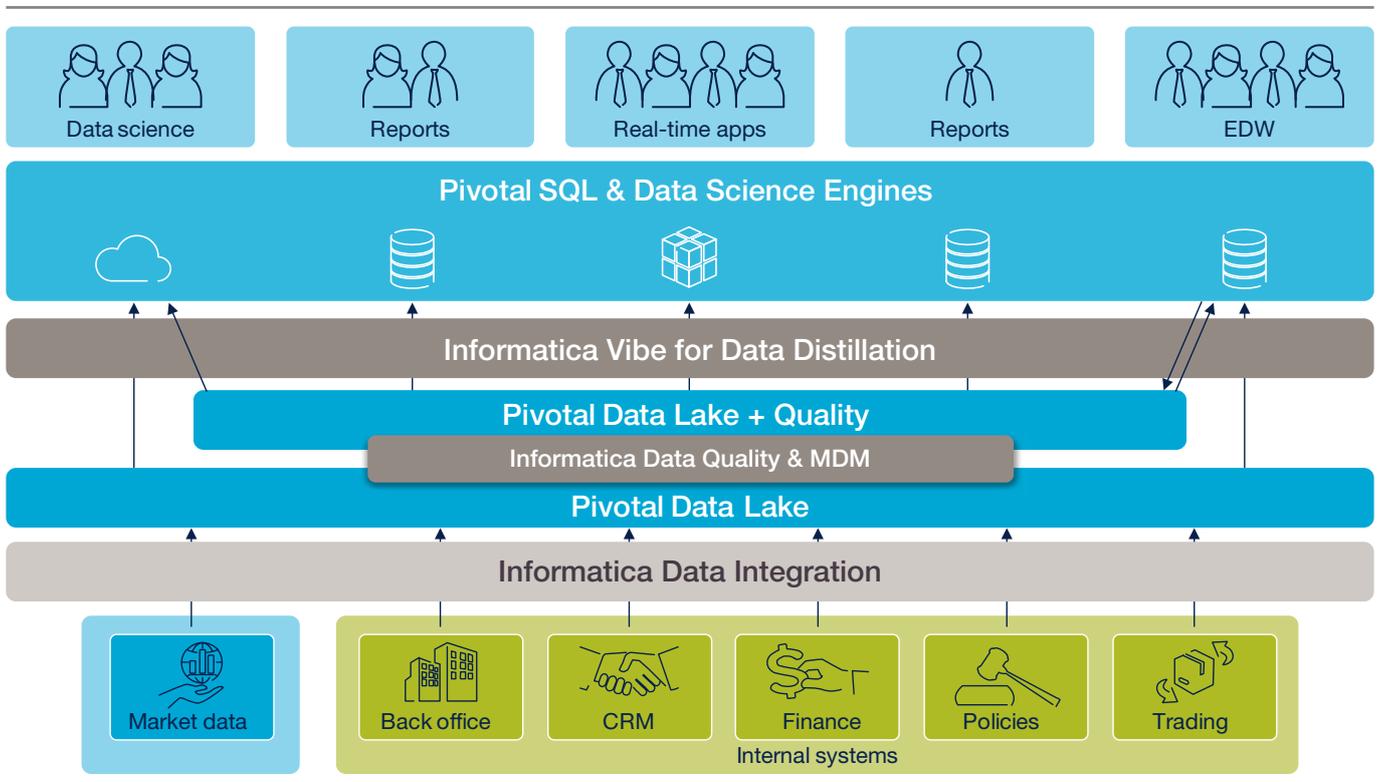
Complementing Pivotal's advanced information storage and horizontally scalable analytical engines, Informatica's capabilities in the area of MDM, distillation and refinement make it easier than ever before for companies to build the platform they need to transform to a data-driven enterprise.

A recognized leader in data management and integration, Informatica provides a business-proven solution for ensuring data isn't dragged into a data swamp. Incorporating Informatica's leading capabilities for data quality, complex data parsing, and MDM, the BDL delivers accurate information and a framework for building lightweight governance processes.

Informatica's data integration solution industrializes the distillation process required to turn data from the lake into insight for the business. A business-user-friendly graphical user interface (GUI) provides the crucial data ingestion and extraction capabilities, and real-time stream collection helps capture all the data, whatever the type and source.

<sup>1</sup> Gartner “The Data Lake Fallacy: All Water and Little Substance,” Nick Heudecker and Andrew White, July 23, 2014

<sup>2</sup> Capgemini, “You have to manage your Data Lake – the fallacy of technology being magic,” Steve Jones, August 5, 2014 <http://www.capgemini.com/blog/capping-it-off/2014/08/you-have-to-manage-your-data-lake-the-fallacy-of-technology-being-magic>



The unified Informatica big data suite provides the data discovery, data lineage, and metadata management to allow business users to build and refine their own custom views for their specific requirements. The platform allows separate groups to share insight, driving re-use and productivity.

### Benefits of the solution

Capgemini, Pivotal and Informatica provide managed proven big data infrastructure that has MDM and data quality built in, and that enables any part of the business to distill the view it requires from the shared data lake. The result is an industrialized solution providing a level of quality, flexibility and efficiency that was impossible with previous solutions.

Specific benefits include:

- Rapid delivery of value – reducing the time taken to get from challenge to solution – and then the ability to evolve applications and add new, more specific, solutions, avoiding both “rip and replace” and long change cycles.
- Reduction of project risk by using proven technologies on a proven platform with proven skills.
- Dramatic reduction of the cost of supporting data and analytics, with self-service enabled through the use of easy-to-use distillation engine and visual tools. Business users can create their own views in HAWQ or Greenplum.
- Massively parallel processing (MPP) performance, allowing business intelligence (BI) and data science applications previously used with small-scale databases to maintain their performance when applied to the BDL.
- Progressive improvement of data quality.

- Governance and cross-referencing to enable information from multiple sources to be combined into a single database. Data warehouses and line-of-business databases as well as analytics can be drawn from the BDL so that all available data can be used in a consistent and traceable way.
- Leveraging of existing ETL skills not only to populate the BDL from source systems but also to provide business-centric distillation. This reduces the adoption time and re-training costs traditionally associated with big data programs.

In short, we make it easy for your organization to manage its data lake, derive maximum business value from it, and prevent the lake from becoming a swamp.

### Why work with Capgemini?

We are the market leader in both Pivotal and Informatica technologies, with a highly regarded Pivotal Center of Excellence (CoE) and a large number of trained Informatica professionals, backed by a global group focused on data management and MDM.

By working with us you can safely adopt a best-of-breed approach, instead of being locked into a single vendor. We take care of integration issues for you.

If required, we can provide a managed service for big data and analytics. Whatever approach you choose, our focus is not on technology but on business outcomes – we ensure you have the right governance mechanisms to enable flexibility plus control.

Find out more at  
[www.capgemini.com/bdl](http://www.capgemini.com/bdl) and  
<http://www.pivotal.io/big-data/businessdatalake>



## 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/bim](http://www.capgemini.com/bim)

or contact us at  
[bim@capgemini.com](mailto:bim@capgemini.com)