

The Opportunity of the Business Data Lake



Information is advantage

Capgemini's work with MIT Sloan has shown how companies can drive more profitable growth by becoming digital organizations and how this has a significant, up to 24% impact, on a company's stock market valuation¹. Information therefore is no longer just something that is good to have; it's a fundamental basis for a successful organization. When looking at what good means the standard bearers are those organizations that have been able to grow up as digital companies from the start. This view of the 'consumer grade company' has been behind the establishment of Pivotal.

In order for companies to undertake digital transformation and become that consumer grade company, delivering the revenue, profit and market valuation increase that goes with it, those companies need to have an approach to information management and consumption which is driven inline with the business culture. This is what Capgemini and Pivotal are working on with the Business Data Lake – a solution that provides companies with the foundations of digital transformation and the quality of the consumer grade enterprise.



¹ The Digital Advantage: How digital leaders outperform their peers in every industry. <http://ebooks.capgemini-consulting.com/The-Digital-Advantage/index.html>

Technology must stop fighting Culture

The main reason why most organizations struggle with making this transformation is that their technology actively fights against the business culture. This is because the mantra in IT has been to create a single Enterprise Data Warehouse (EDW) for all information. Within information the whole benefit, however, is to have the information you want, when you want, in the way that you want.

IT's answer to this has historically been to create a single EDW with a single schema that aims to be all things to all people and if the information isn't there, then it can take many months to have it added. To speed up the process the business then creates their own data sources and solutions which is why companies have multiple data marts and Excel spreadsheets – the single view is not what the business wants.

Fragmentation is reality

For many years, with its Greenplum, Gemfire, Spring and Hadoop technologies, the team at Pivotal has been concentrating on delivering local views over single enterprise views because this is how businesses want to work. Capgemini has long espoused the business centric view of information, and in particular the value of governing information at the level that allows business collaboration rather than requiring all information to be governed before progress can be made. These two views are what are being brought together in the Pivotal Business Data Lake, the technology assets of Pivotal and the business experience of Capgemini.

Figure 1: Why the single view fails



Delivering Information based on Business Culture

The change with the Pivotal Business Data Lake is based on four key tenets:

Figure 2: Four key tenets



The purpose here is to leverage the business culture and deliver to the business what is required. At the heart of this is the recognition that information is often suited only to a specific context and is not required for collaboration across multiple business areas. This has been the finding from Capgemini's work on master data management and information governance over the last several years.

Store everything

The first challenge is to store everything. This provides the business with the first thing that it needs: access. Traditional approaches extract information based on a pre-defined schema and require new integration for new information requirements. By using Pivotal HD, a low-cost Hadoop based data storage system, it becomes cost effective to store all information, including its history. By providing a data substrate in which all information is made available and where its access can be tracked and traced it becomes possible to provide a centrally managed solution, but not require the extensive and expensive investment in predefining an all encompassing single canonical schema.

Encourage local

The next piece is to stop trying to shut down local data initiatives and instead actively encourage them to leverage the single data substrate. Here Pivotal's technologies such as Pivotal Data Dispatch and HAWQ enable business users

to create their own local views on the data and interrogate that with traditional data manipulation and reporting tools. By encouraging multiple local views on the information it becomes possible to see where the real value is, and to allow the business to choose the sort of performance model that they want.

Pivotal's Business Data Lake provides a series of choices on the type of analytics that can be done by the business, from predictive analytics and complex statistical models, through traditional disk based structured SQL analytics to real-time in-memory analytics. This ability for the business to choose the type of interaction model that they want and to change that model gives the business a dynamic analytics environment that enables them to modify their local solutions without having to undergo major technical modifications. The Pivotal Business Data Lake can also supplement existing Data Mart and EDW technologies by providing a data offload and distillation process that frees up capacity and extends the usable life of existing IT investments. Thus the Pivotal Business Data Lake provides a single managed infrastructure.

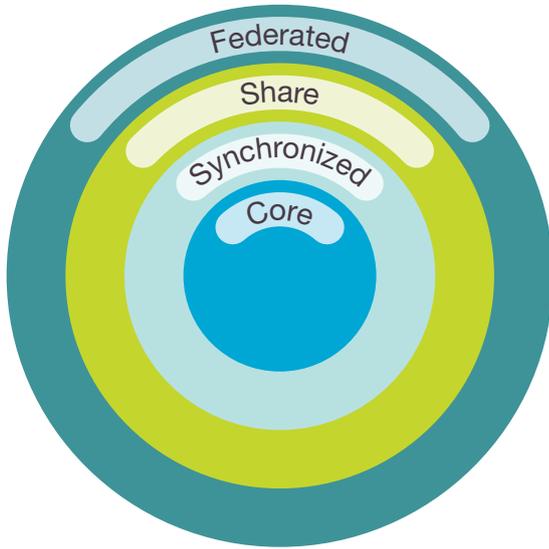
Govern only the common

Governance is good, it gives rigor and accuracy to information but that rigor always comes at a price. The key switch for information governance is something that Capgemini has instilled for several years into the delivery programs with clients: govern for collaboration, don't govern everything. By concentrating on master data management and reference data management it becomes possible to focus governance and thus investment and enable businesses to concentrate their efforts on delivering impact where it drives real value. Capgemini's view on this is shown with the information radar that splits information into four groups:

- **Federated** – transactional information that is of local use
- **Share** – information governed and published by one area of the business
- **Synchronized** – information with multiple editors
- **Core** – information required for the unique identification of a piece of master data.

This approach, and the cross reference between systems that it delivers enables businesses to create the horizontal views that they need, without requiring governance at the lowest levels of granularity and builds on a common practice that already exists in businesses around financial reporting and

Figure 3: Information splits by information radar



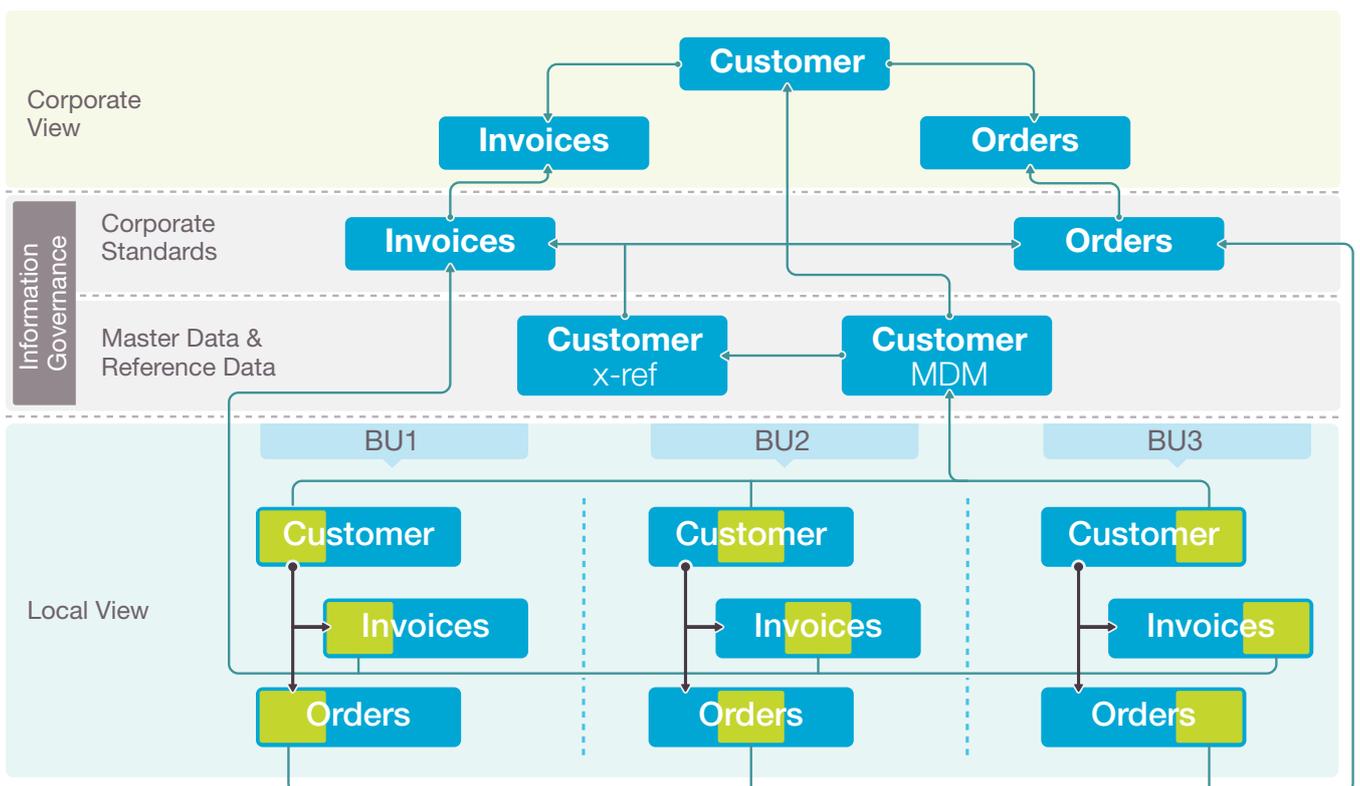
required in an invoice or transaction and it is that practice which Capgemini has successfully leveraged to deliver lighter and more effective information governance and which now forms the basis of delivering this new generation of technical solution the Pivotal Business Data Lake.

Treat global as a local view

The final part is ensuring that global views still exist but that they are focused on delivering not a view of everything but instead be the local view of global executives. With PDD and HAWQ and the new governance approach around business collaboration it becomes possible to create global views which leverage local business views rather than trying to impose the global view upon everything. In this way it becomes possible for group level executives to see divisional KPIs and measures being implemented and tracked, but not require the degree of formalism and constraint that a traditional IT driven project requires.

adjustments. With finance the goal is compliance with the key rules and measures, but not to dictate the specific fields

Figure 4: How the corporate view works



Combining Technology and Culture – The Business Data Lake

Capgemini and Pivotal are working together to deliver on the Business Data Lake view as both companies see this collaboration between technology and business as being central to the digital transformation of organizations into consumer grade enterprises. Traditional IT approaches of large centralized data warehouses have served us well in the past but are now becoming costly silos that meet few requirements and a proliferation of unmanaged or barely managed local solutions which ensures an organization remains fragmented around information. For companies to really deliver on digital transformation they need a measured and governed approach to delivering information change that supports rather than hinders companies in that transformation.

With the Pivotal Business Data Lake there is now a new way to deliver information for the enterprise, one that is based around four simple principles:

- Store everything
- Encourage local
- Govern only the common
- Treat global as a local view

Principles that match the way business works today and now principles that can be delivered efficiently in technology using the Pivotal Business Data Lake and Capgemini's information governance and delivery methods.

Capgemini and Pivotal are working together to deliver on the Business Data Lake view as both companies see this collaboration between technology and business as being central to the digital transformation of organizations into consumer grade enterprises.



Find out more at www.capgemini.com/bdl
and www.gopivotal.com/businessdatalake

Or
contact us at bim@capgemini.com



About Capgemini

With more than 130,000 people in 44 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2012 global revenues of EUR 10.3 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.