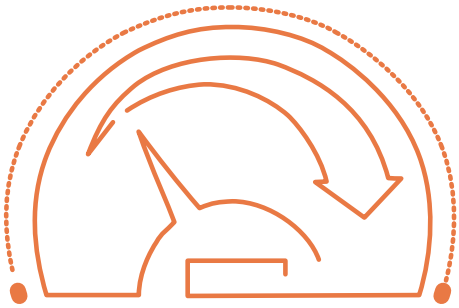
Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Elastic Analytics

When big business meets big data, a dynamic approach to analytics is essential

People matter, results count.



BUSINESSES NEED TO MAKE DECISIONS AT AN EVER-INCREASING PACE, BASED ON MORE RELIABLE ANALYTICS THAT MUST HANDLE GROWING VOLUMES OF DATA.

Turning analytics into a business expense

Businesses need to make decisions at an ever-increasing pace, based on more reliable analytics that must handle growing volumes of data. That means they can't afford to wait for the normal "procure and install" cycle that comes with traditional IT-delivered business intelligence (BI). It isn't reasonable to undertake large-scale capital expenditure just to find out that using a new big data source wouldn't improve forecasting. And if there is a clear need for analytics to be run every hour rather than every day, then waiting three months to make that change means losing competitive advantage.

An associated challenge is that businesses need their analytics to be scalable. This isn't simply about scaling up to meet a peak: critically, it's also about the ability to scale down when that peak doesn't exist. A business that makes large capital expenditures on infrastructure to cover a monthly or seasonal peak is effectively like a store that always has every checkout manned, no matter how many customers are in the store. It's simply a waste of money.

In addition, businesses need an analytics environment whose cost is reflected in the value it delivers *at that point in time* by moving analytics away from IT CapEx into business OpEx. Businesses in other words need analytics which stretch and contract with demand – Elastic Analytics.

The CapEx cliff

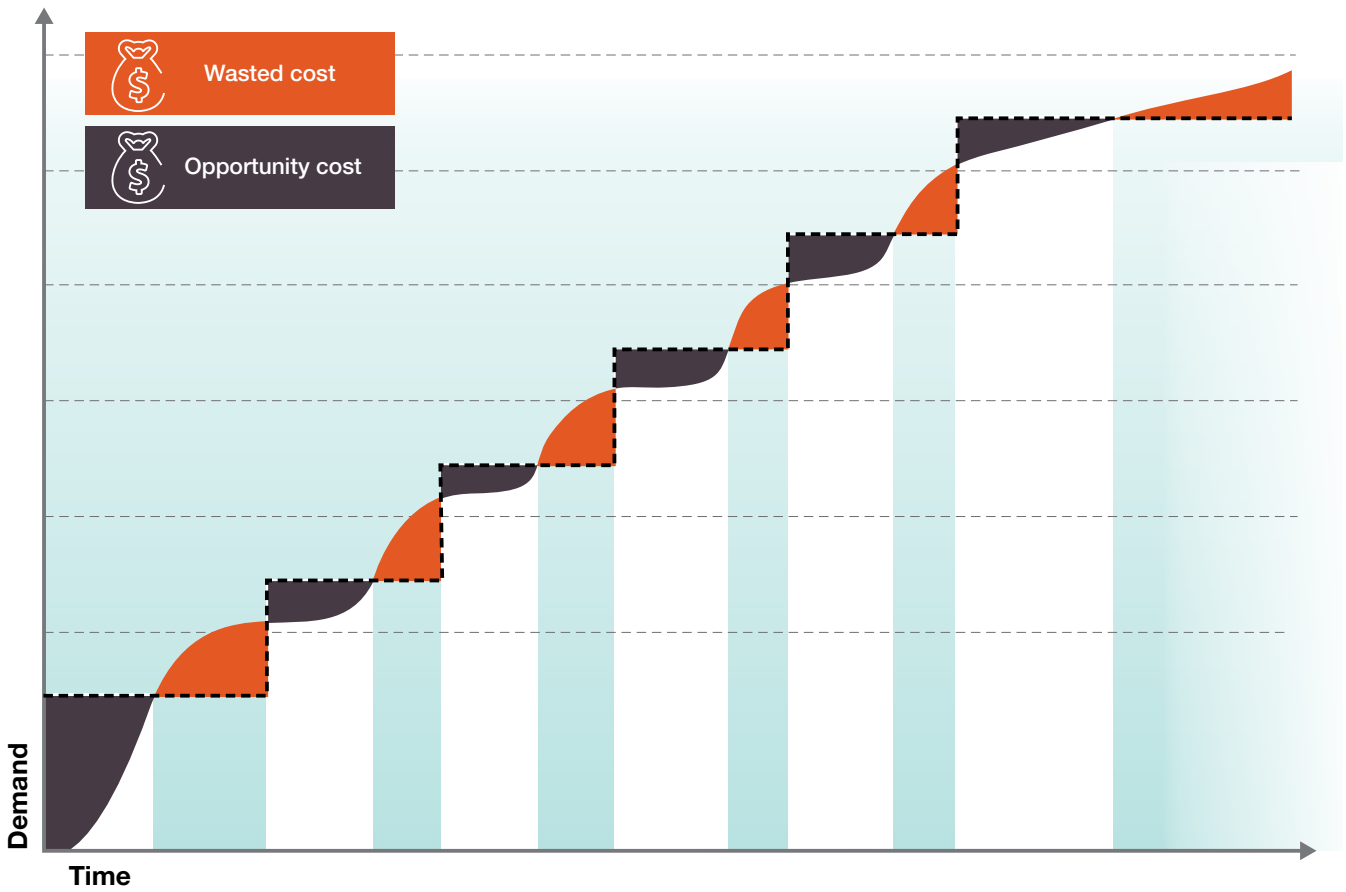
Good analytics involves hard, intensive mathematics. This isn't about the simple summation of a column: it's about applying complex algorithms across massive data sets to get a real view of what has happened and a clearer view of what could happen in future. This complexity means longer processing time, and therefore cost. The volume of data means more storage, which also translates into cost.

The challenge for businesses today is to know how much cost – in other words how much demand – there will be, and when. The traditional IT approach is to buy and install the necessary technology. This means buying hardware and software licenses to meet peak demand in order to ensure there is enough capacity *ahead of time.*

Scaling in anticipation of a peak has a number of challenges of its own. Firstly, it requires estimation of what the peak could be. IT departments tend to overestimate this peak owing to the second problem: if there isn't enough capacity, then the business can't do the analytics it needs and the IT department gets the blame.

This approach drives two poor behaviors: over-procurement (wasted spend), and under-procurement (opportunity cost). A traditional view sees this as an ever-increasing cycle of procurement versus demand. This view is based on an assumption that peak demand keeps increasing and that CapEx is linked to that peak demand.

Figure. 1: Traditional capital expenditure vs demand



Traditional approaches require an estimation of the future, which means wasted money and lost opportunities.

The reality of analytics demand is that it can be much spikier in nature, driven by short-term business challenges as well as by seasonal or regional variations. With big data, this spiky behavior is exaggerated as the business looks to see what value new information sources can deliver, leading to exceptional peaks. In many cases, the conclusion may be that the new data source doesn't bring value, and therefore that additional provisioning of technology is not required after all.

As analytics grows in importance, this elasticity of demand will increase, meaning that the traditional CapEx-based scaling model will constrain business, and do so at a significant level of wasted cost.

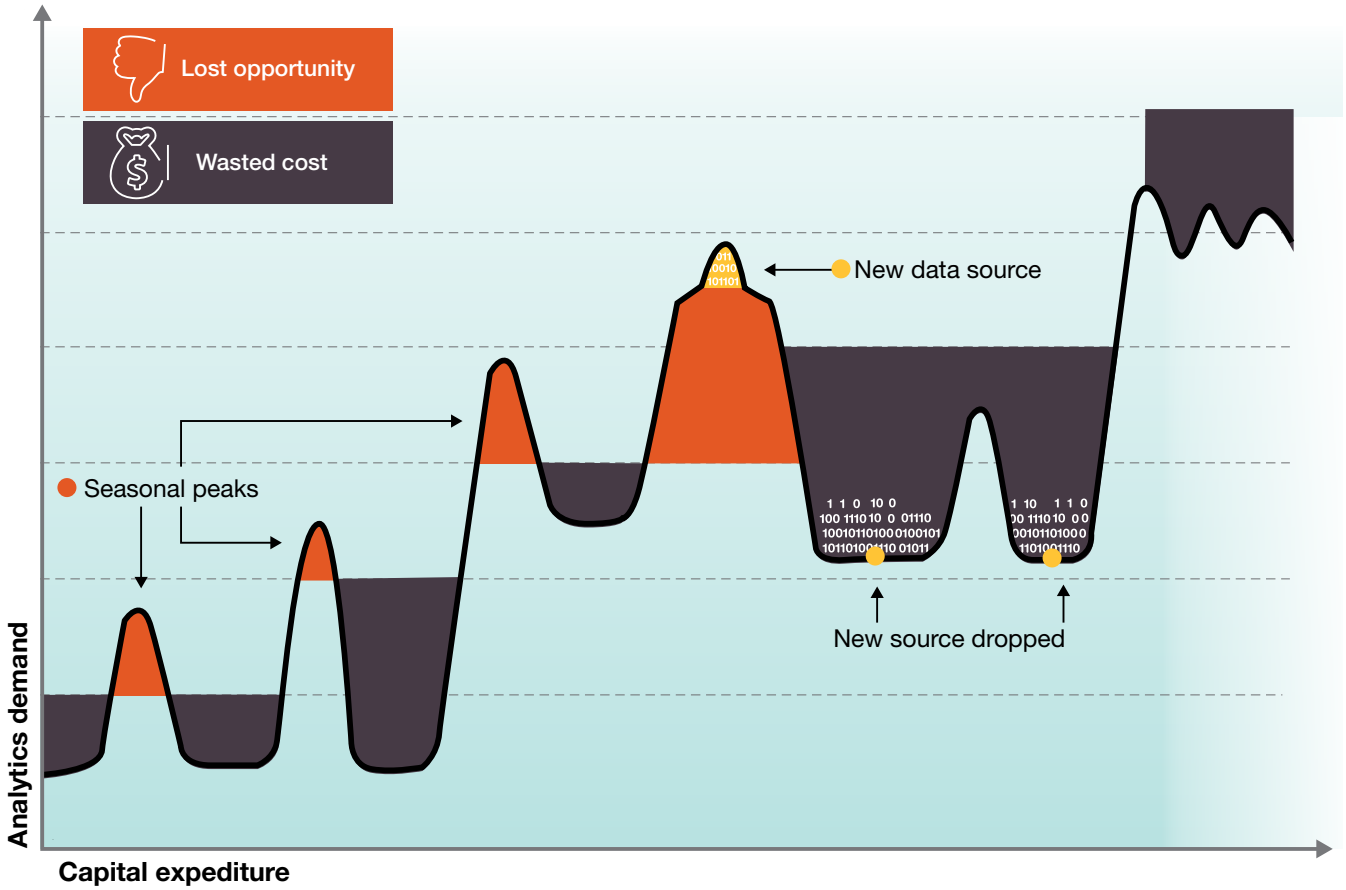
Insight into demand

Instead, a business should be able to flex its analytics demand based on the current value that analytics delivers. This means separating the question of scale, both up and down, from procurement cycles: in other words, having an analytics service where the cost is driven by current demand, and where the choice to scale is practically instantaneous. This is where cloud-based analytics solutions deliver, but they must meet a few key criteria:

- Costs must scale as a single entity based on volume – no separation of software versus hardware costs
- Capacity must be able to be added or removed quickly (e.g. within four hours)
- New analytics models must be deployable by end-users
- Costs must be attributable to the model or analytics being run
- Security must meet organizational standards.

The goal here is to have an analytics environment where the business can see the direct cost of running an analytics scenario and decide on that basis how to run it. For example, a stock forecast could normally be produced by a four-hour overnight run, but during critical periods extra capacity could be commissioned to complete a new forecast every 30 minutes.

Figure. 2: Analytics demand against capital expenditure



The challenge of scaling

It's important to be able to estimate the amount of capacity required to make that type of change in the right timescale. Business users will also need feedback on the impact of volume, complexity and other factors on the time that a run might take. There are mathematical limits to how accurate this estimate can be, but guidance needs to be visible.

Certain types of analytics scale well: if twice the capacity is added, then the analytics take half the time. Other analytics scale much more slowly, and on occasion only faster processors will make the difference – not a greater number. To get the full benefit of the Elastic Analytics approach, therefore, it's important to design your analytics to take advantage of dynamic scale.

Scaling and costing has a number of dimensions, as shown in figure 3.

These scaling dimensions each have different costs associated with them. In the case of storage and memory, there is also the option of moving from traditional disk-based analytics towards in-memory analytics. This move can bring a dramatic improvement in performance, but may also result in a significant increase in cost. Determining the most time- and cost-effective way to balance cost against performance is an important part of Elastic Analytics. It's an interactive process, and one that enables the business, over time, to tune both performance and cost in line with demand.

Figure. 3: Dimensions of scaling

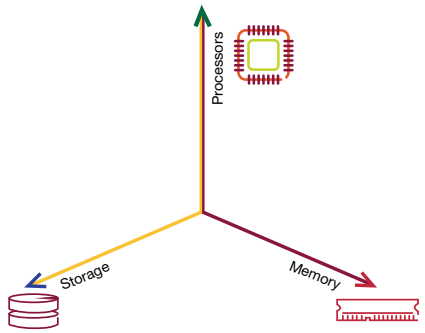
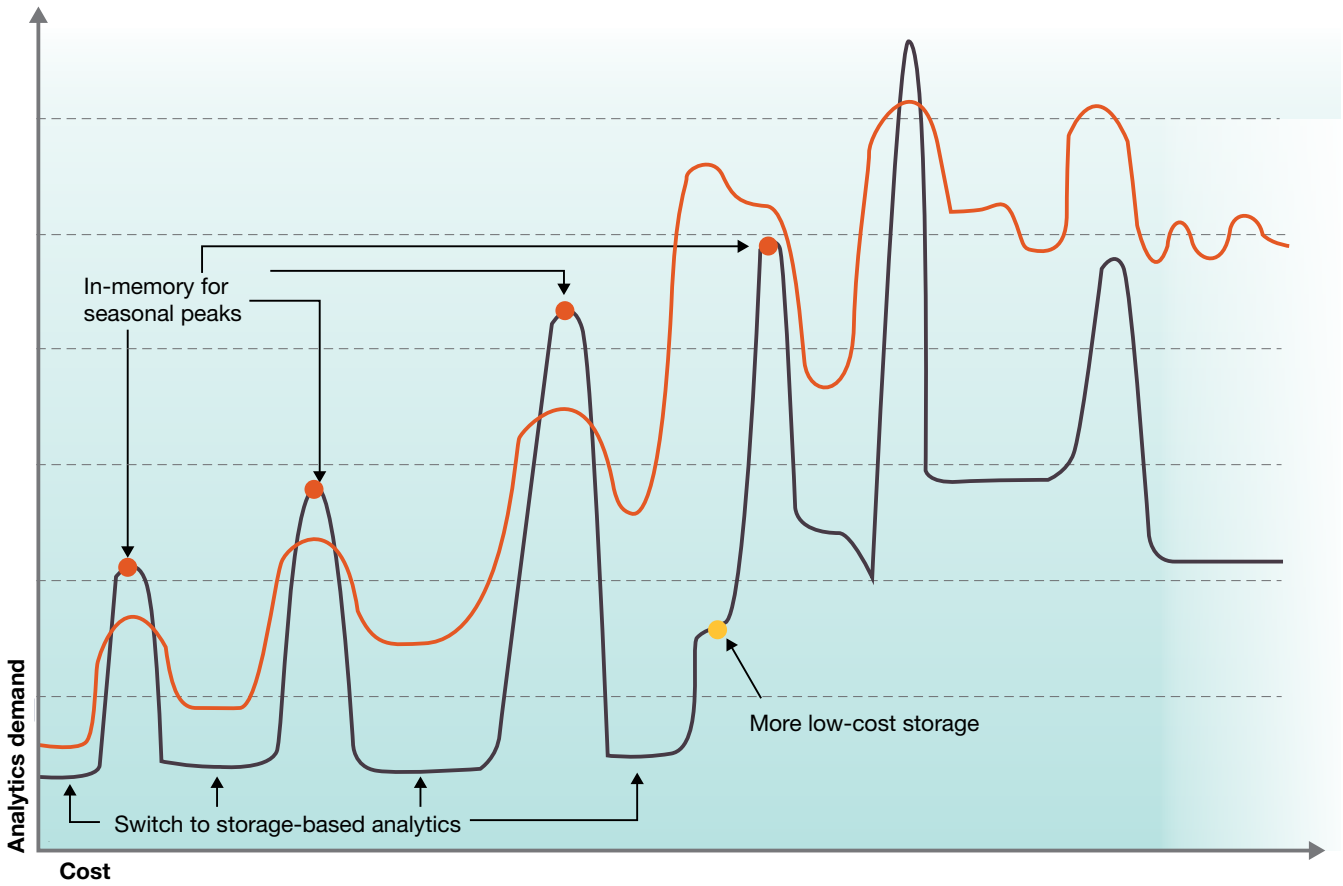


Figure. 4: Elastic demand



Delivering insight at speed

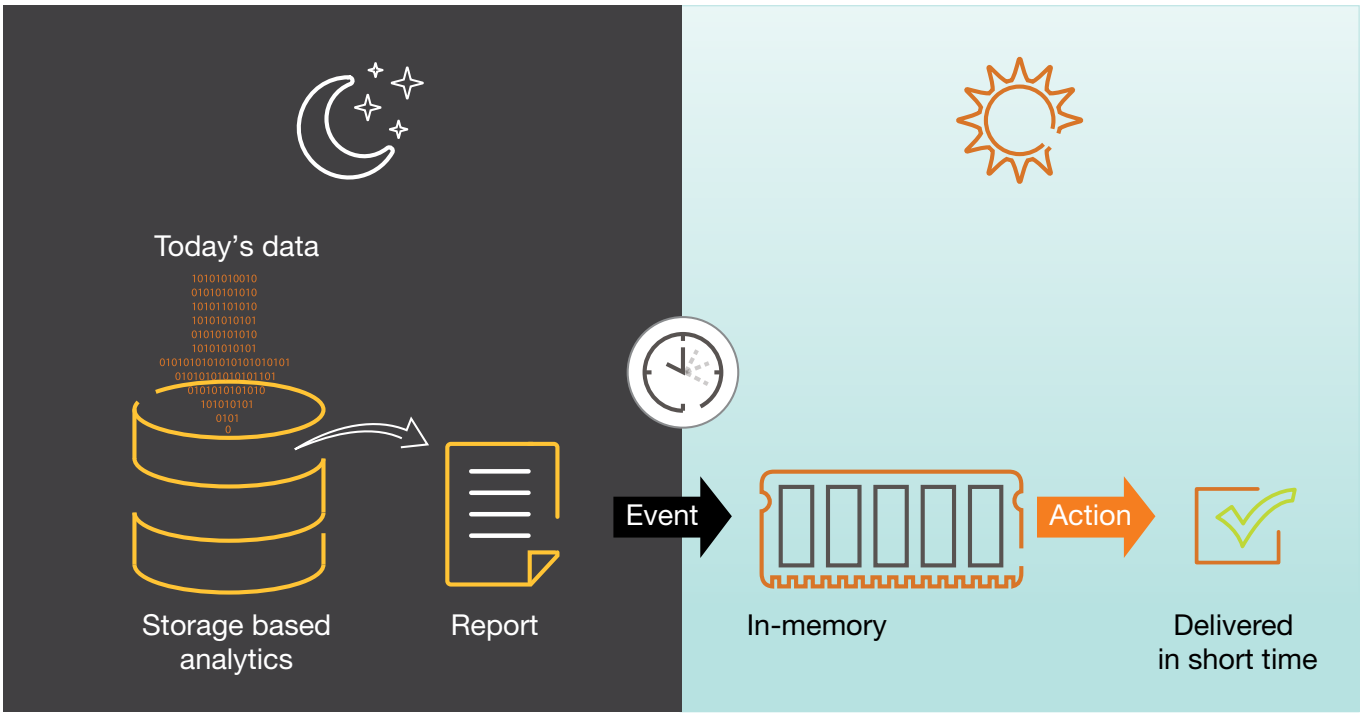
The ability to move from storage-constrained to in-memory and scale-out processing also gives the business a new opportunity: the ability to deliver insight in “business time”, that is, to undertake analytics in a timeframe that means it can be integrated within operational processes.

Until recently, storage-based and in-memory analytics represented two fundamentally different architectures, and the decision on which approach to take was dictated right at the start of the program. With Elastic Analytics, however, it becomes possible to make that decision at the right time for the business.

That means, for example, that a forecasting method can be proven in a lower-cost storage-based approach before being moved to in-memory analytics, where it can be delivered in minutes or even seconds.

An Elastic Analytics environment makes the choice of architecture into a business decision about the speed of analytics delivery. It becomes possible to move between analytics architectures based on value, not on IT restrictions.

Figure. 5: Delivering insight at speed



ELASTIC ANALYTICS IS ABOUT HAVING AN END-TO-END PLATFORM THAT CAN ACQUIRE INFORMATION, MARSHAL IT, ANALYZE IT, AND FINALLY DELIVER THE RESULTING INSIGHT BACK TO THE BUSINESS USERS WHO NEED IT.

Implementing Elastic Analytics

Analytics is the engine – it does the processing, it drives the insight – but, as with a car, the engine isn't much use on its own. The engine needs fuel: in this case, data. It also needs a way for the analytics to be delivered to the end-user: the reporting interface. And everything needs to be integrated in a manner that works without requiring continual modification by IT.

Elastic Analytics is about having an end-to-end platform that can acquire information, marshal it, analyze it, and finally deliver the resulting insight back to the business users who need it.

Elastic Analytics therefore implies Elastic Acquisition, Elastic Marshalling and Elastic Delivery. This is where some technical expertise will be required, and there are a number of ways that it can be managed. Acquiring information, for example, can be done by moving it, unprocessed, into a Hadoop environment, and from there performing the required marshalling or preliminary analytics steps if it needs to be converted into a more structured form. This sort of manipulation can be thought of as an industrialized and large-scale version of the Excel-based manipulation that business users typically do today.

THE KEY TO INTEGRATING THESE ELEMENTS SUCCESSFULLY IS TO CONSIDER ALL OF THEM AS PART OF A SINGLE OPERATIONAL COST.

The key to integrating these elements successfully is to consider all of them as part of a single operational cost. The acquisition of new information should be assigned a cost based on its volume and complexity; this information should then be available as soon as loaded for use within new analytics. Assigning a new reporting or visualization channel should be fast and simple, and down to the choice of the business user on how they wish to interact with the information. This business driven approach needs to be driven by a new financial model, one which makes clear to the business users the cost/performance choices that they have and which makes those costs a continual operating expense rather than needing up-front capital expenditure. Elastic Analytics is about both technical *and* financial flexibility.

Figure. 6: Big data & analytics — Information landscape under the hood

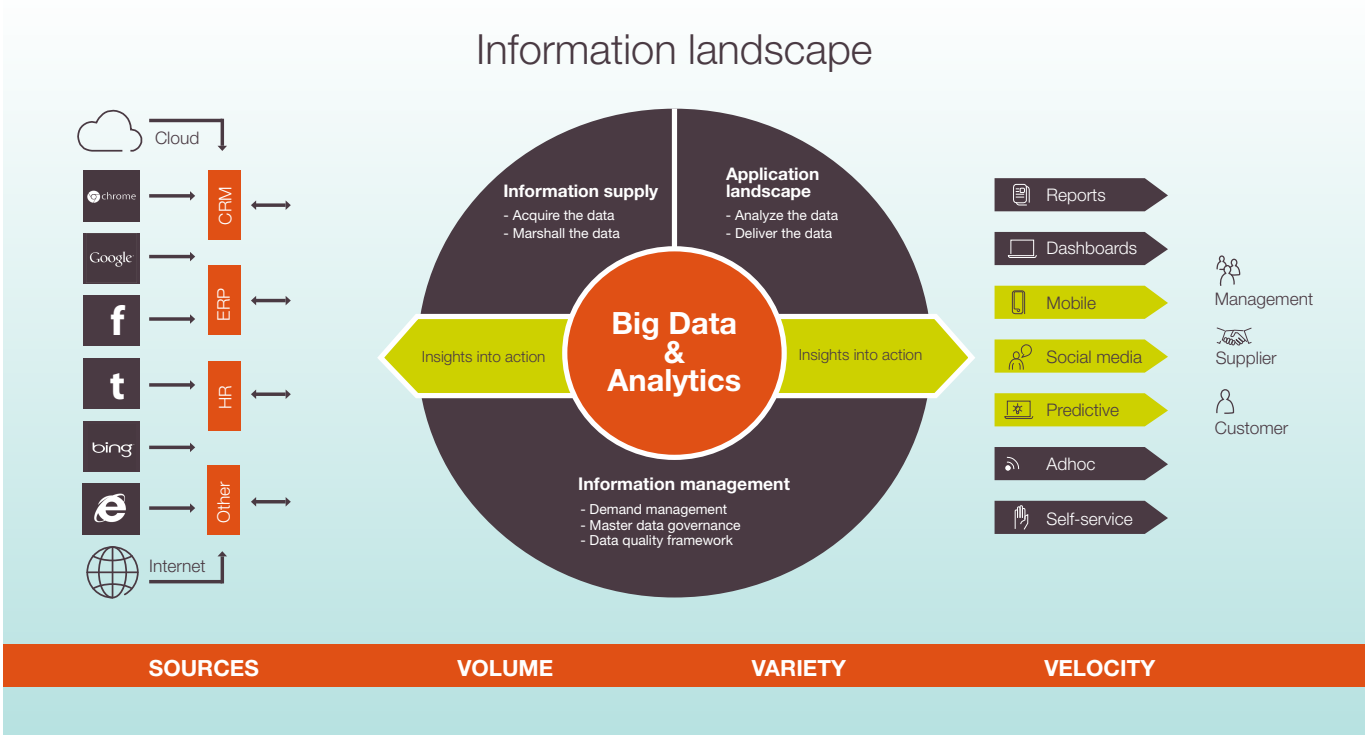
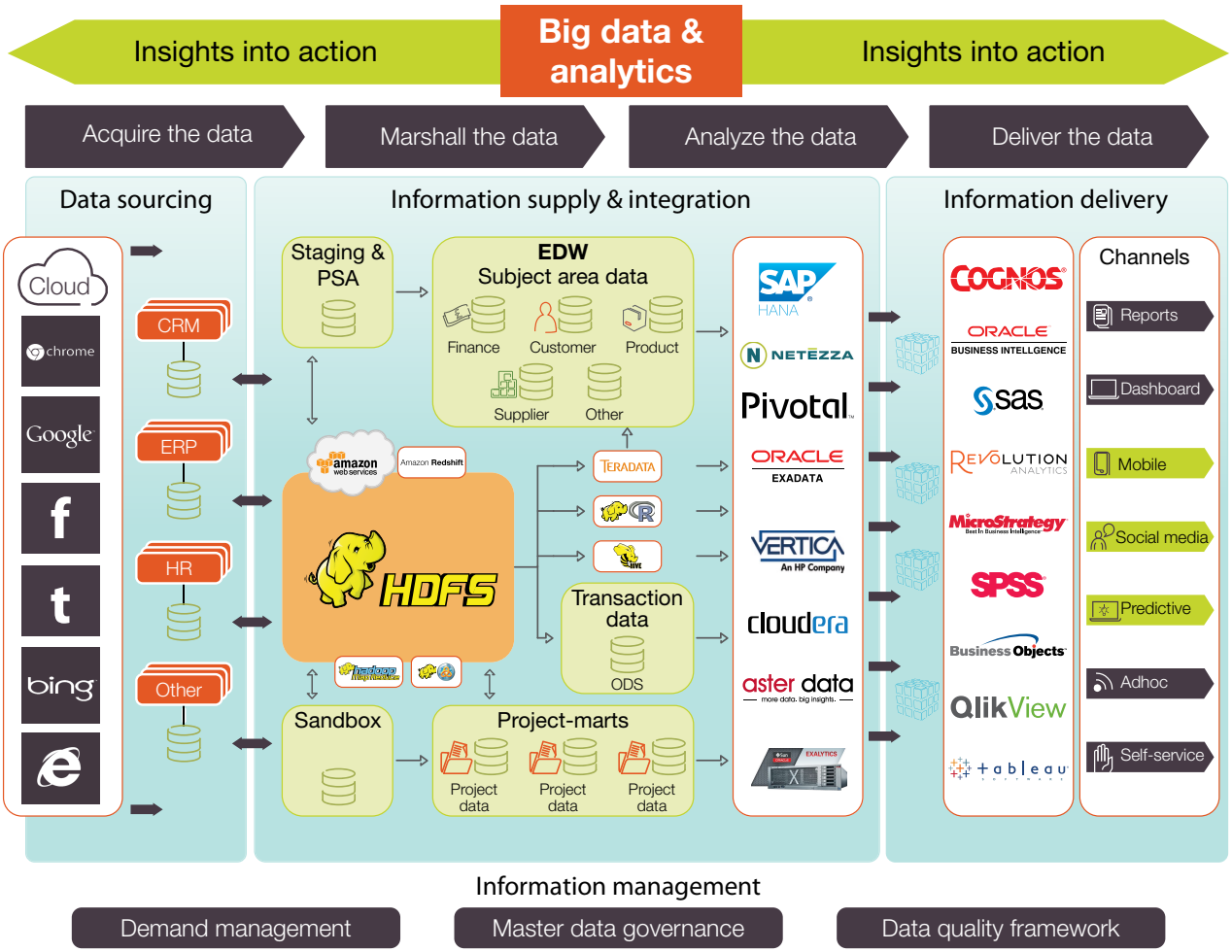


Figure. 7: Big data & analytics — Reference Architecture



One place for everyone

An Elastic Analytics environment – one where costs can be aligned to business needs – makes it easier for organizations to centralize information. The cost of on-boarding a new information source is not borne in some central pool but instead allocated based on the value it delivers to a given part of the business. This approach enables multiple areas of the business to leverage that same source in their own analytics – analytics that they pay for.

Centralization of information doesn't just help the business to share the cost more effectively between users; it also encourages people to add more information sources more quickly. This means information can be combined in ways that would not be possible in traditional ad-hoc sources such as Microsoft Excel.

To summarize, an Elastic Analytics environment can on-board information quickly at a known cost. It rapidly enables every part of the business to use that information, at its own cost, to deliver value.

Elastic Analytics is about business ROI

At the heart of cloud-based Elastic Analytics is the concept of business ROI. Analytics is about understanding the state of a business and providing guidance on its future success. This is why traditional IT-based procurement models struggle to match the requirements of analytics, and why ad-hoc analytics often dominate. Business users need to be able to commission analytics on demand, and to link the cost of a given analytic to the value they expect it to deliver. With Elastic Analytics, they can do so.

This approach creates a cloud-based solution that can scale dynamically and rapidly, and that allows business users to manipulate information in rich and complex ways. These capabilities help to reduce reliance on fragmented tactical reporting solutions. Instead, Elastic Analytics encourages the whole business to collaborate around information; each business area is able to leverage everyone else's information to supplement its own.

Elastic Analytics is about delivering what the business needs, at the point the business needs it, in the way the business needs it, and at a cost that the business decides is acceptable for that specific insight. It's about preventing procurement cycles from constraining the business and wasting CapEx.

Ultimately, it's about putting the power of information and insight into the hands of the business leaders who must use it.

About Capgemini

With more than 125,000 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.

For further information visit
www.capgemini.com/bim

or contact us at
bim@capgemini.com

MCOS_GI_AP_20130925