



Insights-Driven Operations with SAP HANA[®] and Cloudera Enterprise

Unleash your business with pervasive Big Data Analytics with SAP HANA[®] and Cloudera Enterprise



INSIGHTS & DATA

The missing link to operations

As big data solutions have matured, and are preparing for the imminent wave of the Internet of Things, organizations across verticals are looking to collect and transform data into business value.

For example, consumer product companies are trying to understand better the end users of their brands, to better anticipate the changing market trends and design future products. Telecommunications companies are prioritizing their network optimization efforts to serve their high-value customers more effectively, thus gaining a competitive edge in this dynamic market.

Traditional business intelligence architectures, which rely only on well-known relational systems, excel at processing standard business-generated data. However, today's environment increasingly demands insights derived from larger and more complex datasets. As a result, traditional architectures need to adapt to meet the modern needs.

New big data architectures take aim at the aforementioned challenges of volume and variety. However, they can lack key components and the familiarity required to truly bridge the gap between business operations on one hand, and insights generated from big data solutions on the other.

To address these issues, organizations have moved to modernize their composite ERP application landscapes, by deploying the innovation of SAP HANA®. In parallel, enterprises have also started Hadoop-based big data initiatives, which can sometimes be disjointed from operations.

The biggest value for your big data initiatives actually lies in connecting the dots between the two.

A new solution for Insights-driven Operations

Collaborating with SAP®, Cloudera and Intel®, Capgemini has created a reference architecture, integrating all the components needed to put "Insights at the point of action".

We have designed an innovative reference architecture that can:

- Enable real-time integration with business processes and operational systems through SAP HANA's link with SAP Business Suite
- Benefit from high-performance computing, combined with optimized total cost of ownership (TCO) and scalability
- Provide an integrated data governance and security management framework
- Provide an implementation framework for Intel® IOT Platform reference architecture for secure end to end Internet of Things infrastructure
- Ingest (real-time, micro-batch, or batch), distill and process heterogeneous types of data, internal or external, SAP or non-SAP, structured or unstructured, integrating the lightning-fast in memory SAP HANA platform with the advanced data management capabilities of Cloudera Enterprise
- Support a wide variety of analytical workloads, real-time event detection, patterns detection, agile data exploration, visualization and search
- Continuously update with upcoming additional features brought by SAP, Cloudera and Intel (Spark use extension, model self-optimization with Cloudera Navigator, performance optimization).

Capgemini has brought together the engineering and innovation skills of SAP, Cloudera and Intel to integrate game-changing features into this ready-to-use reference architecture, as illustrated in Figure 1.

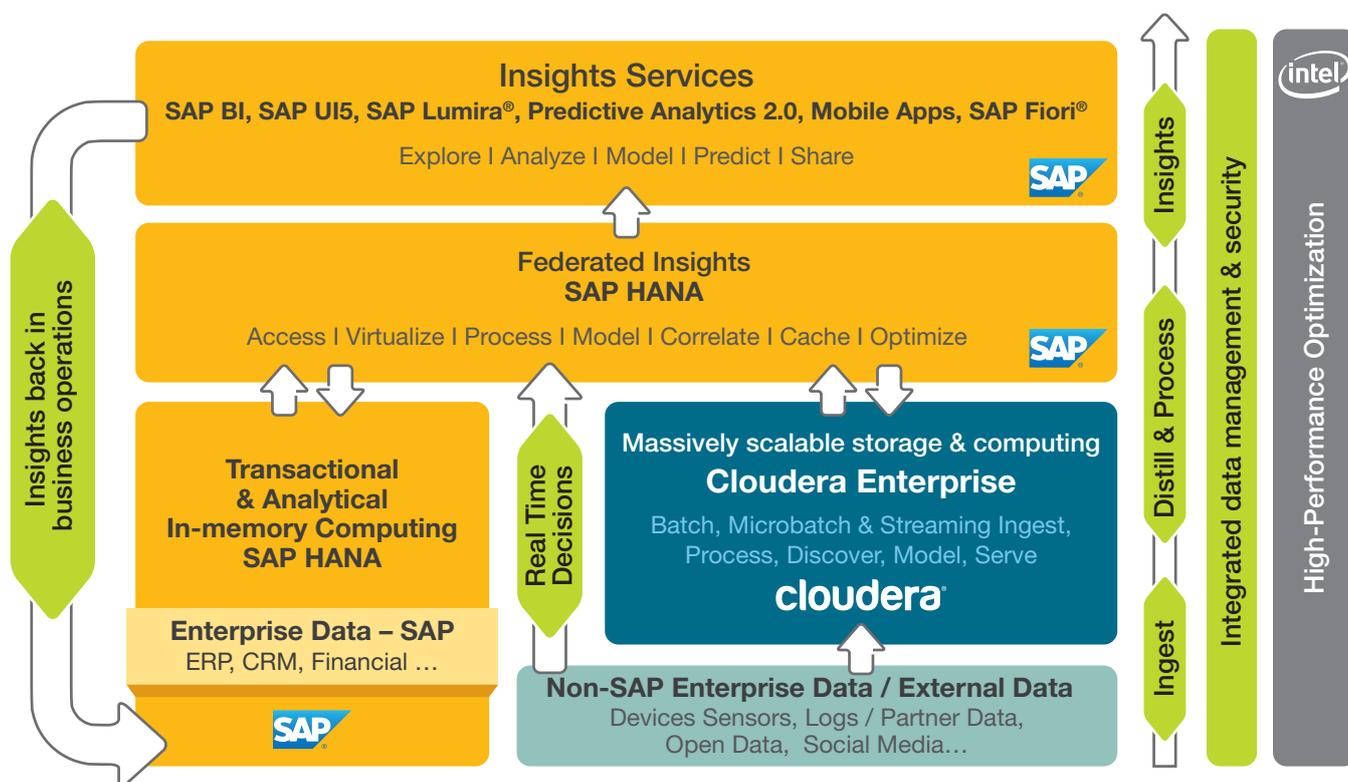
Tangible impacts for your business

With the integration of SAP HANA and Cloudera Enterprise, it is now possible to look at the net new value that can be brought to business operations, by bridging the gap between transactions and analytics. This solution is designed to both support the new Internet of Things and connected people applications, and to keep improving your **existing landscape**.

In addition to extending the scope of data, this solution gives you access to new analytical capabilities (machine learning, long-term signal detection on the longest history of data). Training these algorithms on more data means more accurate predictions, and more accurate preventative actions helps to retain your most **valuable customers**.

However the field of application of Insight-driven Operation goes beyond customer-centric use cases. We highlight below a few examples of transformative use cases that rely on this innovative **reference architecture**.

Figure 1: High-level building blocks of Capgemini's Insights-driven Operations solution



Insights-driven Operations in Consumer Products and Retail

To stay competitive in their markets, retailers need to be able to react quickly to new trends and events across all channels. Optimized sales, demand and inventory planning are key to a retailer's bottom line. To enable this, a more detailed knowledge of their customer base is critical, how it is changing over time (and over seasons), and what are the singular events that are driving this change. For consumer products companies, getting to know the end customer better is the major driver of a sustainable digital transformation strategy.

To effectively produce and execute on these new insights, staying focused on only enterprise transactions is just not enough. Getting a deeper view of each customer's touchpoints, on direct as well as indirect channels, and being able to run wide trend analysis can help you determine the right offer for any particular customer.

Moreover, how well you execute on that offer, how and when you deliver it, how you optimize your assortment planning, pricing and promotion strategy to make it more

profitable, will make the final impact on a customer's basket, their overall loyalty, and your overall sales (see Figure 2).

Insights-driven Operations in Manufacturing

Sensors and connected devices are facets of the Internet of Things wave that organizations are integrating into their near-term strategies. In particular, manufacturers are looking at how they can optimize their end-to-end operations with new datasets that would enrich the detailed view of a production line, the real-time view of stocks for parts that are needed for maintenance, and the overall impact on the quality chain.

By capturing diagnostic events and detecting the fault sequence patterns, there's an opportunity to optimize the entire value chain of your manufactured assets (eg automotive, medical equipment, etc), through a complete view of its lifecycle, from manufacturing to operations to services provider.

By closing the loop and propagating these insights back to product engineering at the appropriate manufacturing stages (ie R&D, design and production), you can optimize the maintenance, upgrade or decommissioning plans for these devices; see Figure 3 .

Leverage the best of both worlds

Our Insights-driven Operations solution delivers a seamless environment, using the latest integration features available in SAP HANA through SAP Smart Data Access, as well as the advanced data management and analytics features of Cloudera Enterprise; see Figure 4.

Value-based data storage balancing

A classic way to describe how data and workloads are shared between those two key components is the easy “data temperature” breakdown. In terms of data temperature management, the “hot” (most critical) data and workloads reside within the SAP HANA platform, to provide near real-time in-memory analytics, using business applications data (ERP).

Cloudera Enterprise is used as the flexible, massively scalable data hub keeping the entire history of data (all data, including unstructured data like logs or social media, “cold” data that is less critical at a given point of time but also hot data), allowing complete trend analytics of the largest possible history of data.

In addition, if data that was considered “cold” in terms of date for example, but is suddenly critical to solve a business issue

(retrospective search for fraud patterns, incident diagnostic), it is then possible to access that data from SAP HANA and utilize its real-time capabilities to answer any business question in an extremely agile way.

This combination allows you to optimize the TCO of the overall platform, and use the lightning-fast capabilities of SAP HANA for mission-critical workloads.

Augmented data management and analytics

But considering SAP HANA as only the “insights oven” and Hadoop as the “data storage” would be extremely limiting. Cloudera Enterprise supports advanced analytics such as machine learning, with algorithms trained on extremely detailed datasets, thus providing more accurate predictions. Thanks to the richness and variety of the tools brought by Cloudera’s ecosystem, near real-time analytics are also enabled on non-SAP or more complex data, using the Spark¹ framework when relevant.

Cloudera Enterprise allows ad-hoc and discovery usage on detailed data, thanks to Cloudera Search for schema-less data, or Cloudera Impala for direct SQL querying. In terms of data integration and Extract Transform Load (ETL) pipelines, Cloudera Enterprise, including Spark Streaming, is able to accelerate drastically the data integration pipeline of complex data within the platform. It is also the best way to store, with minimum latency, datasets with schemas that can change over time (typically external data from providers or partners).

Figure 2: Illustrative Consumer Products and Retail use cases

Insights	Actions
Smart inventory planning	
Real-time prediction of inventory positions, utilizing internal data such as demand plans and forecasts, sales history and clickstream data, as well as external contextual factors like geo-based social media trends and local open data (weather, demographics, local events)	Accelerate optimization cycles for demand planning and forecasts, to reduce out-of-stocks as well as reduce over-stocks
Audience-optimized store operations	
Fine prediction and analysis of store visits thanks to demand signal variables, (store location and segmentation, local contextual data – demographics, weather, events), as well as data acquired from partners (eg store public demographics from telcos)	Enable store managers and sales representatives with a real-time updated heat map of population categories susceptible to visit their stores during the day / time of day. Specific alerts triggered when a known high-value customer enters the store (eg through the customer’s mobile app, I Beacons). Optimize the availability of employees in the store
Just-in-time next best offer	
Clickstream data analysis of the retailer’s online channels to get a more detailed view of the current search/purchase activity of a customer or customer segment	Push automatically personalized offers or promotions to a customer’s smartphone during his next in-store visit or after a defined period of time

¹ Spark, the open-source, “lightning-fast cluster computing” framework has been widely endorsed by the Big Data ecosystem, especially since the partnership announced by Cloudera on October 2013, where Cloudera first integrated Spark into its Hadoop distribution.

Putting insights back into operations

On top of this Hadoop-based data hub, SAP HANA provides an easy-to-use, in-memory accelerated analytical engine to access critical insights in real-time.

Moreover, SAP HANA enables the final “missing link” in today’s classic big data architectures to business operations within an ERP system.

In this reference architecture, SAP HANA has the power to “verticalize” the impacts of the insights, all the way to business operations, at the point of action.

Enabling a seamless, real-time “insights fabric”

SAP applications data (ERP, CRM, BW) will easily be accessible through SAP HANA, either because these applications are running directly on SAP HANA, or by replication using SAP Data Services or Sybase Replication Server, or virtualized through SAP Smart Data Access. The entire history of SAP data will be stored in Cloudera Enterprise, allowing the SAP HANA system to focus its real-time capabilities on mission-critical data and workloads.

Data coming from non-SAP applications or external data is ingested following the value-based data storage principle. Either way, this data will be stored in Hadoop with its history, using data ingestion modules from the core Hadoop platform (depending on the nature of the data source and its velocity) or Spark Streaming for near real-time ingestion.

Cloudera Enterprise will be able to easily handle complex ETL processes with its advanced data ingestion and processing capabilities, and also complex or changing structures, thanks to its schema-less architecture.

Depending on the “real time intensity” that is required for a specific use case, complex event processing (CEP) can be executed in the SAP HANA platform, while writing both event and decision in Hadoop to keep an entire history and traceability.

Building and crossing insights

Each component of this architecture will provide analytics capabilities for the data that they are storing, according to the principles outlined above.

SAP HANA will offer its real-time analytics capabilities especially for SAP data (or non-SAP data necessitating CEP), while Cloudera Enterprise allows the use of its data management and analytics component (Spark Core for fast transformations, Hive for SQL querying and Impala for interactive query, or Search for unstructured data) on all data.

In the same manner, Machine Learning techniques (eg Mahout or Spark MLlib) can be applied to the entire data sets in Hadoop, including history, to deliver the best possible predictions (eg churn or fraud detection). R analytics and visualizations will preferably be executed on Hadoop, to benefit from better scalability and a broader scope of data.

Figure 3: Illustrative Manufacturing use cases that can be implemented on our reference architecture

Insights	Actions
Predictive asset maintenance	
Identification of symptoms that lead to a particular part failure thanks to the correlation of diagnostic events with device usage data	Plan and perform preventative repairs before they become critical, reducing down time and optimizing parts back-order
Parts inventory optimization for maintenance planning	
Real-time alerts of potential failures to field teams, parts inventories and stock organization	Alert field engineers in real-time that a part is about to fail Ship replacement parts preventatively to a regional warehouse to enable ‘Fix-it Right First Time’ = FRFT Allow quick access to the right parts in the right zone/warehouse in the right quantities, to reduce cost of storage and maintenance time
Closed-loop product quality monitoring	
Repeated failure trends	Propagate repeated failure trends to Product Engineering and Quality, especially on newly-commercialized products. Corrective actions can be taken to fix the problem and protect the manufacturer’s brand reputation

To federate those insights, and create new ones based on the crossing of data stored or produced in SAP HANA or Cloudera Enterprise, our solution makes use of SAP Smart Data Access (SDA). SDA enables data structures to be exposed from Hadoop to SAP HANA in a transparent way to the final user, giving access to all data with the strength of Hadoop.

Through this federation layer, Machine Learning models trained on the entire history and scope of data can be applied in real-time to new data coming in. Data exploration scenarios can also be enabled in this layer, by pulling and caching data from Hadoop to SAP HANA, allowing real-time exploration and prototyping on top of selected datasets.

Analytics produced, and models prepared within SAP HANA using SDA, can be archived or stored back in Hadoop, to keep the history of all analytical scenarios applied to the data, for traceability or future comparison purposes.

Exposing insights to analytical tools and integration back to operations

The insights produced above will be accessible through the variety of analytical tools and API developed by SAP, especially SAP UI5 for advanced visualization (including geo-based), SAP Lumira® for data exploration, or Predictive

Analytics 2.0. SAP Fiori® applications will be able to access data shared through SDA.

Through SDA or by write-back, computed insights will also be made accessible to SAP business applications (Business Suite, CRM, BW), as “enhanced attributes” (eg customer attributes like churn scores back in BW, sentiment temperature in SAP CRM, or cost and profitability attributes in Business Suite).

End-to-end security framework

Our solution leverages both SAP’s and Cloudera Enterprise’s robust security features. Centralized user provisioning is done through SAP HANA. Thanks to SAP Smart Data Access and the federated insights it provides, information access privileges are defined in SAP HANA. Authentication is done through LDAP and Active Directory, enabling security policies for both SAP HANA as well as Cloudera Enterprise, illustrated in Figure 5. Our reference architecture takes advantage of the Intel-developed features such as encryption-decryption optimization at chip level and information access optimization, which are part of the latest versions of Cloudera Distribution of Apache Hadoop (CDH).

Figure 4: Insight-driven Operations reference architecture

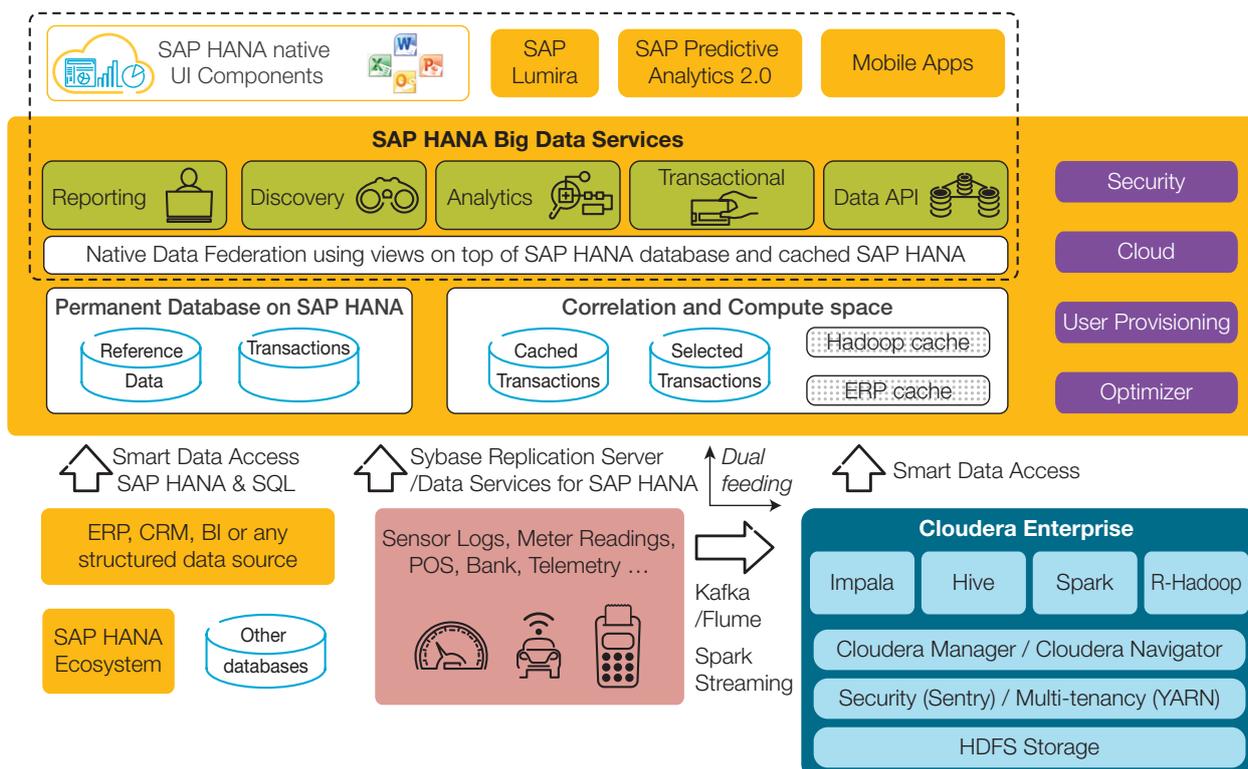
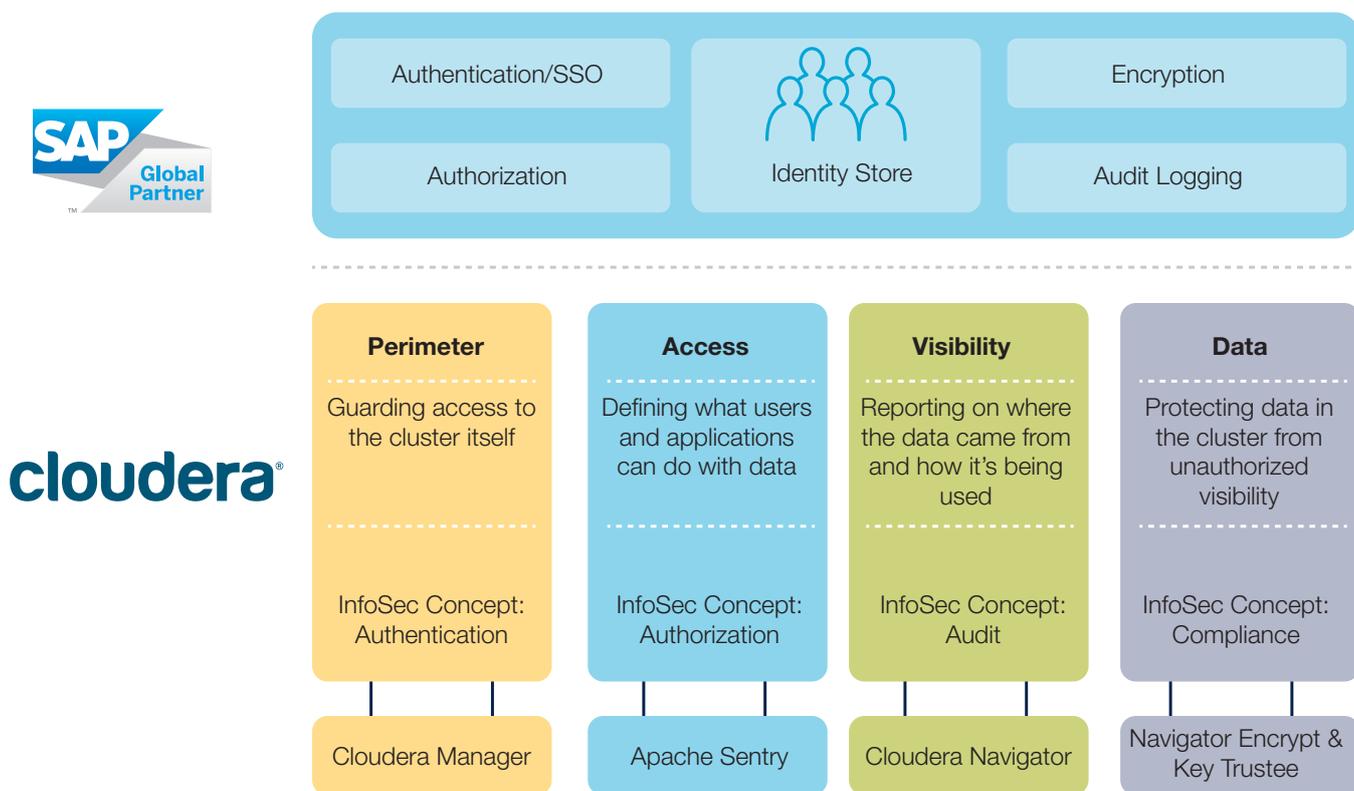


Figure 5: Security features for SAP applications and Cloudera Enterprise



Capgemini delivers tangible business value

Capgemini delivers tangible business value, working closely with its partners:

Business Focus

We strongly believe that business outcomes should be the key driver in big data initiatives. Our large teams of data experts, with strong domain expertise, can help you define the best business use cases to deliver both value from your data and competitive advantage in your specific industry

Technology expertise

Capgemini has been a Global SAP Partner for over 20 years with a global workforce of more than 16,000 SAP practitioners. With over 900 SAP HANA consultants in 21 countries and more than 60 SAP HANA projects delivered, Capgemini is a recognized leader in SAP HANA, winning the 2015 SAP Pinnacle award for Services Transformation and 2014 SAP Pinnacle award for SAP HANA Adoption.

As big data has become a strategic imperative for our clients, Capgemini has also built a strong team of experts to implement Cloudera Enterprise solutions. Cloudera is one

of Hadoop industry's leaders, and as the largest System Integrator partner of Cloudera, Capgemini has developed significant expertise in implementing their solutions.

Innovation

Our Insights and Data team focuses on designing and delivering innovative solutions to make sure you master your data assets and fully deliver on your digital transformation strategy, as well as prepare for the next wave of the Internet of Things. Our close partnerships with SAP, Cloudera and Intel have allowed us to bring together the most recent innovations from all three companies, and offer a 'best in class' and sustainable approach for **Insights-driven Operations**.

Seamless collaboration

Thanks to our partnerships with SAP and Cloudera, Capgemini is your 'one stop shop' for Insights-driven Operations, from provisioning development and production environments, to implementation, industrialization, and delivering 'as-a-service' models.

We believe that the first critical step is to define the key business processes your organization needs to transform short- or mid-term. The best place to start your Insights-driven Operations journey with Capgemini is an Innovation Workshop to identify the key business use cases that will create the best tangible value to your organization.



For more information on the Insights-driven Operations solution, please contact insights@capgemini.com.

Capgemini Insights & Data

In a world of connected people and connected things, organizations need a better view of what's happening on the outside and a faster view of what's happening on the inside. Data must be the foundation of every decision, but more data simply creates more questions. With over 10,000 professionals across 40 countries Capgemini's Insights & Data global practice can help you find the answers, by bringing together technology excellence, data science and business expertise. Together we leverage the new data landscape to create deep insights where it matters most - at the point of action.

To find out more visit us online at www.capgemini.com/insights-data and <https://www.linkedin.com/company/bi-big-data-&-analytics> or follow us on Twitter @capgemini.com



About Capgemini

With more than 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
or contact us at
insights@capgemini.com