ascent
Thought leadership from Atos

white paper

# Data Analytics as a Service: unleashing the power of Cloud and Big Data

Your business technologists. Powering progress

Atos

Big Data and Cloud, two of the trends that are defining the emerging Enterprise Computing, show a lot of potential for a new era of combined applications. The provision of Big Data analytical capabilities using cloud delivery models could ease adoption for many companies, and in addition to important cost savings, it could simplify useful insights that could provide them with different kinds of competitive advantage.

Along these lines, **Data Analytics as a Service (DAaaS)** represents the approach to an extensible platform that can provide cloud-based analytical capabilities over a variety of industries and use cases. From a functional perspective, the platform covers the end-to-end capabilities of an analytical solution, from data acquisition to end-user visualization, reporting and interaction. Beyond this traditional functionality, it extends the usual approach with innovative concepts, like Analytical Apps and a related Analytical Appstore. In addition, the platform supports the needs of the different users who interact with it, including those of the emerging 'Data Scientist' role.

Architecturally, and due to the intrinsic complexities of analytical processes, the implementation of DAaaS represents an important set of challenges, as it is more similar to a flexible Platform as a Service (PaaS) solution than a more fixed Software as a Service (SaaS) application. Aspects like the PaaS internal architecture, the distinction between real-time vs. non real-time processing, the specific characteristics of the Analytic Services, the needs for data storage and modeling, the delivery over hybrid cloud models and several others, make its design a complex challenge.

In order to validate the base concepts of a DAaaS platform, the Atos Scientific Community conducted research that resulted in a Proof of Concept (PoC), based on a concrete scenario for the Oil and Gas industries. This scenario explores some of the most important points that need to be solved in a DAaaS.

Published in March 2013

# Data Analytics as a Services

## Contents

### About the Authors

Edited by Celestino Güemes (celestino.guemes@atos.net) SmartUtilities R&D Manager at Atos Worldgrid in Spain, with contributions from Jordan Janeczko, Thierry Caminel and Matthew Roberts

### About the Atos Scientific Community

The Atos Scientific Community is a network of some 100 top scientists, representing a mix of all skills and backgrounds, and coming from all geographies where Atos operates. Publicly launched by Thierry Breton, Chairman and CEO of Atos, the establishment of this community highlights the importance of innovation in the dynamic IT services market and the need for a proactive approach to identify and anticipate game changing technologies.

# The Scenario for Big Data and Cloud Convergence

It's really difficult to miss Big Data in the media. Major news outlets like the New York Times or the Guardian have made everyone aware of the exponential growth of the information that businesses need to cope with. The importance of Big Data is however not the volume of data, neither its heterogeneity, but instead it is about the insights that businesses can get from understanding this entire "data deluge" using the appropriate analytical methods.

"Doing" Big Data in a proper way is not easy. In addition to the complexity of tools and infrastructures that are required to manage huge volumes of data, we need to identify and resolve the scarcity of talent that can properly take advantage of these tools to make data "talk". Some of the techniques associated with Big Data are not really new (statistics and machine learning have a long history) but these techniques need people with a deep understanding; in a new role that now is often called the 'Data Scientist'. So, Big Data can be a tough proposition for many companies as conventional tools and on-premise techniques could be the wrong approach.

To ease many of these "pain points" another transformative trend has entered the market place - Cloud Computing and the continuous movement towards a Utility ICT model. So instead of deploying complex solutions in-house, companies can take advantage of services provided by third parties, using economic models that offer them flexibility and adaptability to the changing needs of their environment.

It's easy to understand that the combination of Big Data with Cloud can ease the adoption of advanced analytic capabilities over the bigger and more heterogeneous data sources that companies need to handle, letting companies benefit of the insights derived from it. Value is in the data itself, and not in the technology that is used to process it. What companies need is not the deployment of a complex Big Data infrastructure and the associated capital investment, but the access to the services that provide advanced Data Analytics on their data now and in the future. Providing this service in a flexible and scalable format is the main purpose of **Data Analytics as a Service**.

> It's easy to understand that the combination of Big Data with Cloud can ease the adoption of advanced analytic capabilities over the bigger and more heterogeneous data sources that companies need to handle, letting companies benefit of the insights derived from it.

# What is Data Analytics as a Service (DAaaS)?

Data Analytics as a Service (DAaaS) is an extensible analytical platform provided using a cloud-based delivery model, where various tools for data analytics are available and can be configured by the user to efficiently process and analyze huge quantities of heterogeneous data.

Customers will feed their enterprise data into the platform, and get back concrete and more useful analytic insights. These analytic insights are generated by Analytical Apps, which orchestrate concrete data analytic workflows. These workflows are built using an extensible collection of services that implement analytical algorithms; many of them based on Machine Learning concepts. The data provided by the user can be enhanced by external, 'curated' data sources.

The DAaaS platform is designed to be extensible, in order to handle various potential use cases. One concrete case of this is the collection of Analytical Services, but it is not the only one. For example, the system can support the integration of very different external data sources. To enable DAaaS to be extensibility and easily configured, the platform includes a series of tools to support the complete lifecycle of its analytics capabilities.

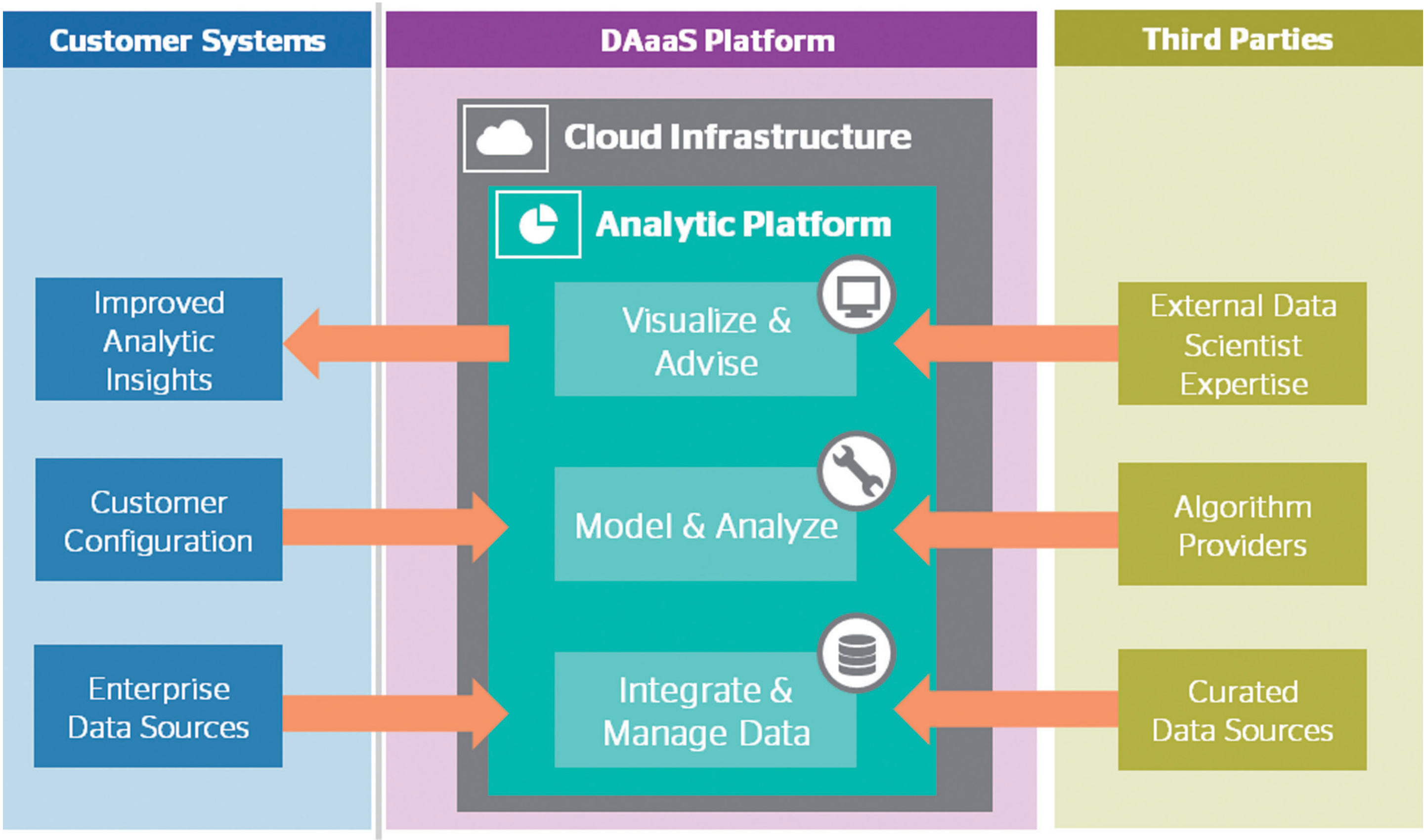


**Figure 1: DAaaS Concept**

# Functional Elements of a DAaaS Solution

In order to deliver all the capabilities of a DAaaS solution, a complete platform needs to be implemented, as shown in the following figure:

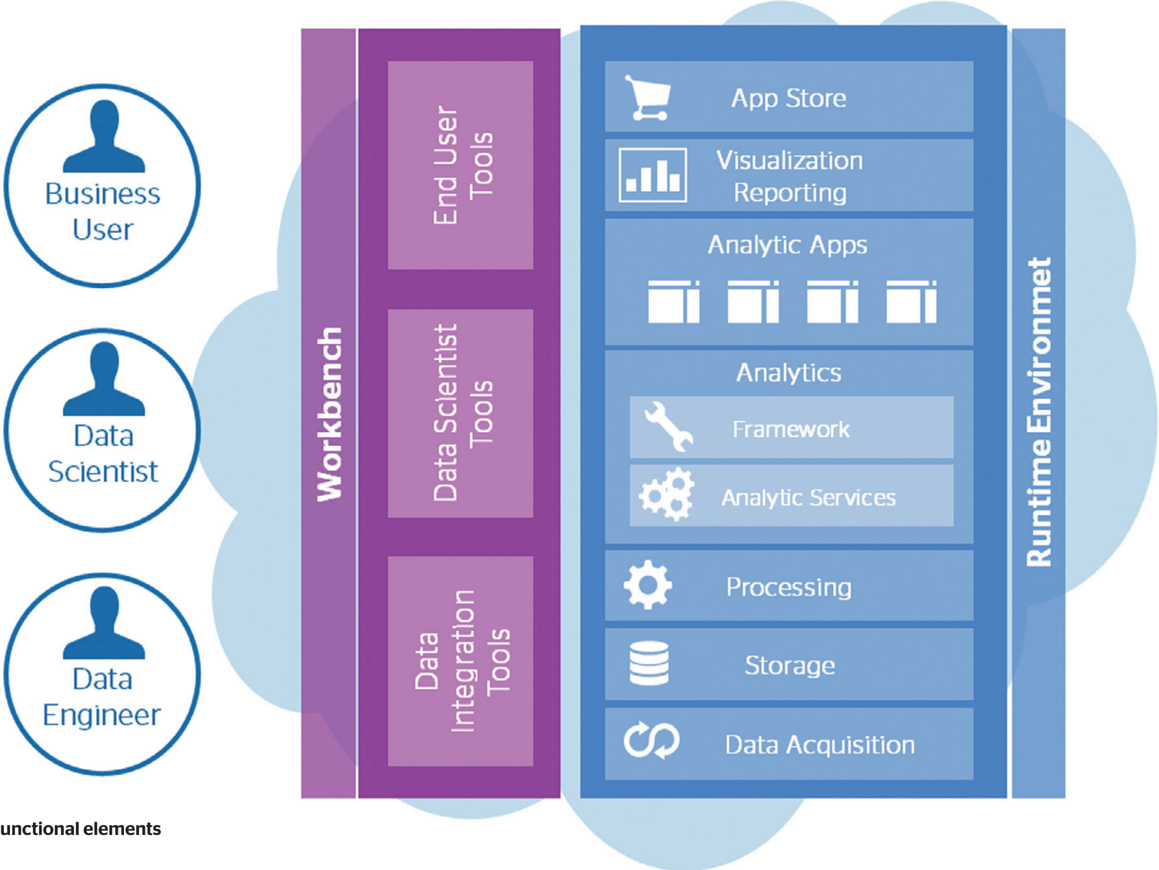


**Figure 2: DAaaS functional elements**

First, we need to differentiate the two groups of elements:

- Those related with the runtime aspects of the solution, that is, the platform that processes the data. We've identified this as **Runtime Environment**.
- Those that control the interaction with the user, mainly for the configuration of the system, using a set of tools that we have called **Workbench Environment**.

To the last point, we use here the concept of a user quite broadly, including not only the end business users, but also all other individuals that interact with the platform in more "technical" roles, like programmers in Data Modeling and Integration roles, or Data Scientists that configure Analytical Services and Data Flows.

## Runtime Environment

The Runtime Environment is the execution platform of the DAaaS solution. We describe the various components from the bottom to the top as shown in the figure, following the logical data flow from ingested data (input) to generated insights (output).

**Data Acquisition**: it provides a web services / messaging interface to the external world, for the acquisition of data, both from the end customer but also for the alternative 'curated' data sources. To provide the flexibility to cope with very diverse data sources and protocols, the solution needs to be modular with components that implement widely recognized EIP (Enterprise Integration Patterns).

**Storage**: the core data repository is used to store the information in the system, both from the customer but also other data. The solution should be able to cope with up to petabyte-size quantities of data, but also needs to be flexible enough so very different data models can be implemented and supporting strict multi-tenant capabilities. There are several potential NoSQL databases that allow these kinds of capabilities, each having strong and weak points for different scenarios.

**Processing**: In order to be able to process huge volumes of data, it is essential that some kind of distributed processing capability exists, so that different processing algorithms can be implemented and executed in parallel. This processing layer acts as an interface between the Data Storage and the analytical services. Nowadays the more popular solution for this kind of Distributed Processes is the Hadoop Map-Reduce solution. It is widely supported, both by different NoSQL data sources, but also by programming languages and analytic tools in upper layers of the stack. However other emerging alternatives like Spark, or Storm, are appearing with some specific advantages, for example in real-time scenarios.

**Analytics**: in a sense, this is the most crucial part of the platform. This is where all the analytic processes reside. We can deconstruct it further in two main elements:

- The **Analytic Services**, which are well-defined components that implement concrete data analysis algorithms. These can be quite varied and could be implemented using different base technologies. But they have a defined, concrete scope related to a specific Data Analysis technique over specific dataset classes. Many of them will be based in Machine Learning techniques, both using Supervised and Unsupervised approaches.
- An **Analytic Framework**, that acts like a glue that brings together different Analytic Services in order to achieve a concrete business outcome. So a specialized programmer could use this framework to implement a complete analytic functionality.

**Analytic Apps**: as we've just seen we can combine different Analytic Services using the capabilities of the Analytic Framework. We call these bundles of complete analytic functionality an Analytic App. These Analytic Apps are the elements that expose the final, business-oriented capability to the end-user.

**Visualization / Reporting**: although a great part of the functionality of the DAaaS platform could be accessed using a web services interface, a complete solution will integrate some form of end-user oriented visualization / reporting to simplify access to information. An ample set of tools that provide these kind of capabilities exist, some commercial (Tableau, QlikView) and some others open source (Pentaho, Jasper Reports).

**App Store**: The solution will provide high-level functional bundles or Apps to end-users. An enterprise App-Store front-end will provide the mechanisms to control the lifecycle of an app in an organization, from acquisition to retirement.

## Workbench Environment

As we will justify later, the DAaaS solution has some intrinsic characteristics of a Platform as a Service (PaaS). So in addition to the runtime environment, some specific tooling is needed to customize the solutions (the Analytic Apps) to the specific needs of the end-user. We call this set of tools the **Workbench Environment**. It includes tooling for different kind of roles:

**Data Integrator**. This role takes responsibility for interfacing the existing internal enterprise systems with the DAaaS system. Specific tooling for the Data Acquisition subsystem (and, to a lesser degree, for the data storage and processing layers) is needed, that enables ETL (extraction, transformation and load) of this information, so it is properly modeled and incorporated into the DAaaS environment.

Once the data is in the DAaaS Storage subsystem, the next role takes care of a correct implementation of the analytical capabilities required by the enterprise.

**Data Scientist**. A substantial part of the work of a Data Scientistis to correctly model, test and check analytic workflows exposed in Analytic Apps, for validity for the specific datasets used. Going even further, we can assume that specific Analytic Apps (or even, Analytic Services) could be built by them. Data Scientists can be employees of the company, but they could also work for a specialized service provider.

The final step is by the **Business Users**. One important point to bear in mind is that from the DAaaS perspective, there exist different kinds of business users. Obviously, we have the non-technical, business end-user, who only needs to have the results from the tool, either integrated in their existing enterprise apps or using simple to use visualization tools.

But we have other, more specialized types of users, that we could call **Subject-Matter Experts** (SMEs for short). These users, without having the deep mathematical knowledge of a Data Scientist are essential as they provide the business context to the statistical findings. In addition, they can have an active role in DAaaS. For example, they can help in the determination of Business Pattern Rules based on the feedback from analytic patterns, that may later be implemented as Real-Time Analytics capabilities.

DAaaS has some intrinsic characteristics of a Platform as a Service (PaaS). So in addition to the runtime environment, some specific tooling is needed to customize it to the specific needs of the end-user. We call this set of tools the Workbench environment.

# The Challenges of Analytics in the Cloud

Analytic solutions that need to support Big Data services present additional challenges. This is even more the case if these services are intended to be delivered through a cloud environment.

- Information Lifecycle Management: the complete analytical workflow can get very complex with lots of important steps: data acquisition (data access, setting parameters, transformation, data cleansing, data quality...) data modeling (definition of logical model, linking with other data...), data mining (variable identification, algorithm selection, validation...) and visualization (customized reporting, advanced graphics...). In contrast with transactional solutions, which are more fixed in nature, Analytics need a flexible approach to adapt to all this potential variability.
- Data model diversity: a diversity of potential types of data models exists for specific business needs - and these data models are tightly coupled to specific types of analytics. For example, time series data is modeled quite differently from social network data and also the potential algorithms to be used are distinct.
- Analytic knowledge: although not really new, many of the advanced techniques related to advanced analytics (like Machine Learning) are quite complex and demand people with very specific knowledge, such as the previously mentioned Data Scientists.
- Data volume: even when technology exists for processing huge volumes of data, it is not easy. Moving big volumes of data to a cloud solution can be difficult, and sometimes, it is much easier to bring computation to where the data is.
- Real-time analytics: more and more, the value of analytics demands quicker insights, progressing towards the concept of real-time analytics. Even if we are talking here of soft real-time, this has a definite impact on how a cloud solution is designed to handle these loads.
- Security: like in any other cloud solution, security is a very complex issue. Some companies, due to the data criticality or regulatory constraints, may be reluctant to move data to the cloud, but could benefit of the analytical capabilities offered in a private cloud.
- Privacy: for some specific types of data, privacy considerations may impact the potential of cloud analytics - not only due to the data in itself, but also due to the potential that data will not remain anonymous after analysis.

In addition, we should note that in addition to issues that are specific to Analytics, there are others that are of a more technical nature, related to how existing Analytics and Big Data platforms need to adapt to a cloud infrastructure, as many of them started out in more classical 'on-premises' infrastructure.

# Architectural Aspects of a DAaaS Platform

As the previous point has shown us, a DAaaS platform can be an important venture, as it implies some interesting but complex engineering challenges. Here we analyze some potential approaches to the previous challenges. Due to space restrictions, we can't cover them all in this White Paper. For example, security and privacy would by themselves demand a complete paper each[1].

## DAaaS is not SaaS, but a Specialized PaaS

Analyzing the DAaaS problem as if it were another SaaS solution would lead to some very thorny and critical issues. As we saw previously, many of the steps in an analytical workflow can demand considerable customization due to the characteristics of the data sources but also of the analytic outcomes that are intended. If we take a basic analytical workflow, we will find at least the following steps:

- Adaptation of the ETL (Extraction-Transformation-Load) processes from enterprise data sources.
- Adaptation to the DAaaS Data Models and Metadata standards.
- Modeling and configuration parameters of Analytical Apps, by defining workflows using Analytical Services.
- Validation of results from end-users.
- Integration of results back into Enterprise systems.
- Configuration of GUIs for reporting and visualization.

Even for simple cases, it seems obvious that the typical functionally-fixed SaaS solution doesn´t provide enough flexibility. One potential approach to these difficulties could be to see the DAaaS as a form of specialized PaaS[2] (Platform as a Service). This is a lower level view, more oriented towards customized applications, which the customers (or integrators) would develop to their specific needs.

In this scenario, end-developers for the external application will have access to the elements in the DAaaS in a programmatic way. They could develop solutions, which would run in the DAaaS cloud, adapted to the specific needs of their system. This will imply the existence of the PaaS solution and some specific tooling to be used by end-developers.

So they will:

- Develop the acquisition of data to the interface standards in the Data Acquisition Layer for the data that they need to analyze.
- Select and configure the specific analytic services for that specific data, to their custom specification. They could additionally arrange different analytic workflows, which could combine different kind of algorithms.
- Access the results from the execution of the analytical services and integrate them into their system.
- Customize visualizations running in the DAaaS to their specific needs.

A possible scenario for the interaction of the client with this PaaS could follow the following script:

- Client has a specific Analytic need that demands lots of customization and integration with their own systems.
- Client registers with the Analytics PaaS which provides access to the platform and to the development tools that may be needed.
- Client looks for specific analytic services and external data sources in the Catalog, simulates the potential cost, and purchases the ones that could interest him.
- Client customizes specific modules for his needs:
  - Module to export data from his system to the DAaaS,
  - Module to get results back into his systems,
  - Customized Analytics module, which uses the purchased analytical services and data sources, according to specification,
  - (Optional) Creates a Dashboard (that runs in the DAaaS) to visualize information.
- Client deploys modules in the DAaaS platform.
- Custom analytic module is executed according to client's schedule.
- Client supervises the system (usage, billing, service level agreements) using monitoring tools provided by the DAaaS.

Even in this PaaS approach the important characteristics of Cloud are preserved. For example, the billing of the system could combine considerations about Data Storage, Computing Time used and specific analytic services used. And the concept of Analytical Services catalog remains valid - the customer could purchase specific analytic services that implement concrete algorithms using the Analytic App Store, which would be enabled for their developers.

One very important point: the PaaS approach does not invalidate a higher level, less "customizable" SaaS approach. In some sense the PaaS approach could be the platform where non-customer-specific analytic applications could be developed, for well-defined use-cases and for some specific, well documented data formats.

## Data Storage and Modeling Issues

The DAaaS platform is intended as a horizontal platform that can be used in quite different vertical sectors. So it is designed to be a generic, extensible platform. One point where this generic approach raises important issues is regarding Data Storage and Modeling. Basically, the base storage technology needs to:

- Be able to represent and store different kinds of data.
- Be capable of storing huge volumes of data in a distributed manner.

[1] Atos Scientific Community has written a White Paper ("Privacy and Social Networks") about the main issues related to privacy

[2] For a detailed explanation of Platform as a Service (PaaS), see Atos Scientific Community White Paper, "PaaS - Making the Most of Clouds".

- Integrate those data models with the processing layer so analytical services can process information with the required performance.
- Provide strong multi-tenancy capacities.
- Satisfy strict security requirements.

There are several potential data storage engines or databases in existence that could satisfy this requirement. They belong to what has been called NoSQL databases[3], a varied set of solutions that provide very different approaches to data representation and storage: document-oriented, key-value, column oriented, network-oriented, etc. Each solution has its strong points and weakness, mostly because they fit some specific kind of data models better.

So the DAaaS platform has to choose between two options:

- Select one data storage option and adapt it to be used in all the potential cases.
- Open the possibility of using distinct data storage technologies.

This second option, being better adapted to the distinct potential data sources, introduces an important level of complexity in the architecture.

In the first case the choice of a specific model can be difficult. For example in the DAaaS Proof of Concept, that the Atos Scientific Community has developed, HBase (part of the Hadoop project), a column oriented database, was selected; HBase provided a good option to represent different kind of data, especially those related to Time Series information.

In any case the DAaaS platform must be designed for reusability at the Data Modeling Layer and to external layers: to the Acquisition Layer and to the interfaces to the analytic services in the Analytics layers.

## Analytic Services

The Analytic Services are the basic building blocks of the analytical capabilities of the system. The basic idea here is to implement a component architecture in which the interfaces to external systems are clearly defined so extensions (Services) can be implemented. We can distinguish several types of interfaces:

- Interfaces to the storage layer to access the information they should operate with.
- Interfaces to return results back to other components or as final results.
- Interfaces to configure or model the analytical services for execution.
- Utility interfaces for example, interfaces for monitoring the execution of the system.

As previously stated, these analytical services have a defined scope taking charge of one specific or 'atomic' analytic operation. For meaningful analytic capabilities, there will be a workflow of operations that it is coordinated by the Analytic Framework element. We have called these workflows Analytical Apps. The components (analytic services) and the Analytic Frameworks are designed to run in a distributed manner using the base capabilities of solutions like Hadoop Map-Reduce, Spark, or Storm.

Technically, these building blocks can use different technical approaches, some of them based in Statistical and Machine-Learning techniques, such as Artificial Neural Networks, Bayesian networks, Support Vector Machines, Regression Analysis, etc.

## Real-Time vs. Non Real-Time Analytics

Another important issue regarding DAaaS is related to its potential real-time capabilities. Note that here we are talking about a soft real-time; response times are in the order of seconds, not the hard real-time that is in use in embedded systems for example.

Usually, traditional analytics have been in the realm of non real-time. However, technological advances and business pressures are demanding shorter response times for analytical processes at the same time that data volumes grow exponentially. So, the architecture must provide mechanisms for this kind of near-real-time response[4].

In the case of the DAaaS platform, there is an additional problem that needs to be taken into account. Some technologies in the DAaaS stack are not very well aligned with the needs of this kind of real-time response. Perhaps the most significant one of those 'troublesome' system components is the Hadoop Platform. It is the most popular Big Data platform but at the same time it is not really real-time oriented. For example, the Map-Reduce framework is based on batch execution modes of operation. This imposes an important burden onto the system if real-time modes of operation are required.

So is it possible to have Real-Time Analytics over the DAaaS platform?

The most sensible answer is: not by itself, but it can be complemented by other systems that provide real-time response in coordination with the non-real-time capabilities of the DAaaS.

The main idea is shown in the following figure:

[3] A detailed exploration of different NoSQL options can be found in Atos Scientific Community White Paper "Open Source Solutions for Big Data Management"

[4] One example use case of this real-time needs is covered in the Scientific Community White Paper "Connected Robots"

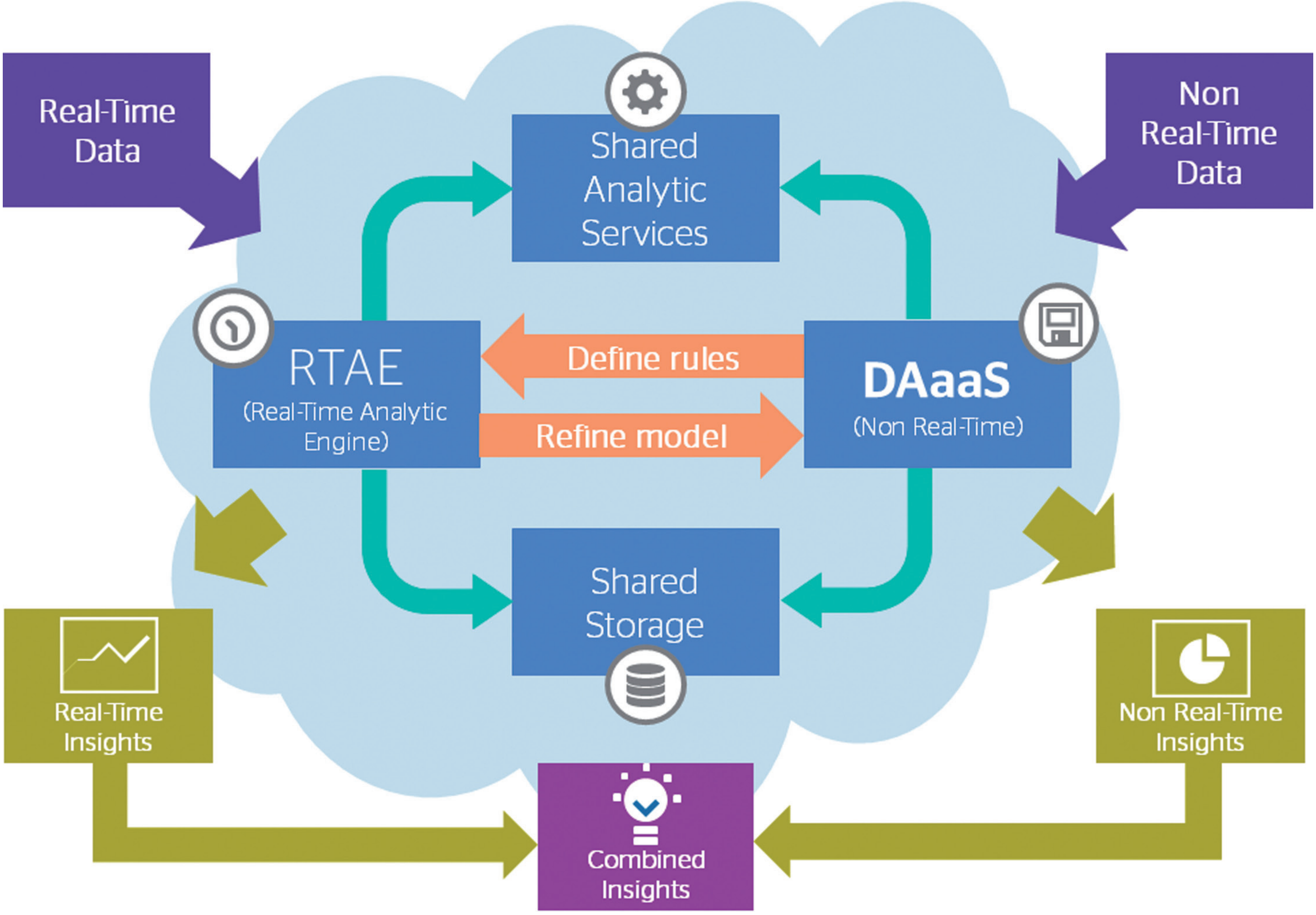


Figure 3: DAaaS and Real-Time Analytics

We can say that we have two systems running in parallel:

- The DAaaS platform.
- A Real-time Analytics Engine (RTAE).

The latter takes care of the real-time analytic capabilities, processing streaming data and looking for patterns on it. These patterns may be of different degrees of complexity and may be expressed using different mechanisms. Lower levels of complexity may be represented by a "business rules language", like the one included in Complex Event Processing (CEP) solutions.

For a higher level of complexity, complete analytic services could apply that may share common elements with those included in DAaaS.

For example, a potential approach to the RTAE part is an extension of the Atos' Context Broker Platform (CBP)[5]. The latest version of this solution will use a distributed messaging framework based in the 'actor model', called Akka implemented in the Scala Language. These models will define a 'distributed complex event processing (CEP)' platform and on top of it, the real time analytical capabilities can be implemented.

As shown in the figure, both DAaaS and RTAE can be interrelated, so that they complement each other:

- They can share some specific analytical services.
- They can share persistent storage.
- The rules at the RTAE can be configured by the outcomes of the DAaaS Platform. So, if a pattern emerges in a DAaaS App the RTAE can be configured to detect this pattern.
- Similarly, the RTAE can interact back with the DAaaS platform so real-time outcomes can alter how a DAaaS app works. For example the analytical models implemented in the DAaaS can be refined with results from the RTAE.

## DAaaS in a Hybrid Cloud Environment

A pure public cloud approach to DAaaS may have to address some important challenges for some user scenarios:

- When the volume of data is huge or rapidly growing, it may be not practical or efficient to move data from customer's premises to a DAaaS provider infrastructure due to size or communication constraints.
- In addition to this size issue there may be other limitations such as security or privacy policies, that demand on-premise or private cloud approaches.

Taking into account these restrictions is there a model that enables the benefits of the DAaaS model while still preserving data in the locations of the customer?

The answer may be an approach towards a Hybrid Cloud which would bring part of the DAaaS approach to customer premises, in effect bringing computation to the data instead of the other way round.

The approach is based in the work being done internally by Atos Scientific Community in the concept of Cloud Services Brokerage[6] and can be summarized in the following figure:

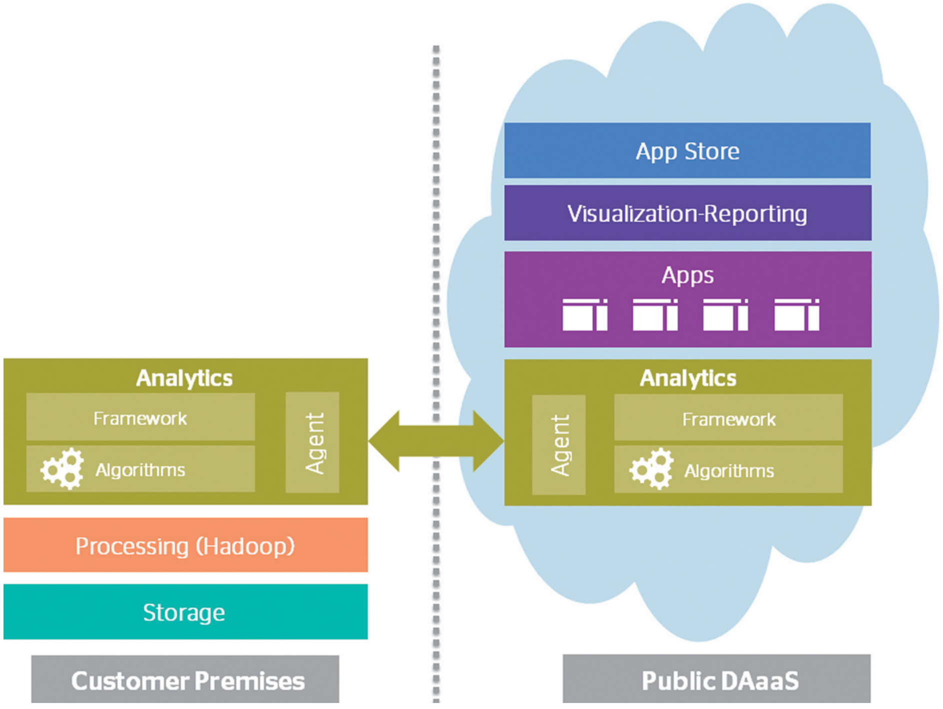


Figure 4: DAaaS working in Hybrid Cloud model

[5] Context Broker Platform (CBP) is described in detail in Atos Scientific Community "Journey 2014" document
[6] The Atos Scientific Community White Paper "Cloud Orchestration – A Real Business Need" explains the concept of Cloud Brokerage in detail.

In this model, the architecture is split with a part residing on customer premises, which comprises of the Data Storage and Processing Layers. Part (or even all) of the processing for the Analytic Services then takes place on customer's premises as Hadoop Map/Reduce processes for example, coordinated by the higher level elements of DAaaS available as external public cloud elements and using the Analytic Services present in the Public DAaaS solution. This coordination is achieved by using a series of Agents which take charge of distributing jobs between both environments.

So this Hybrid PaaS provides important advantages regarding flexibility of deployment, according to customer restrictions on data availability and minimizes the needs of data transfers as well as integrating DAaaS in a higher view of cloud services coordinated by some kinds of Cloud Orchestration services.

## Running a Big Data Stack in a Cloud Environment

From a base technologies perspective, in order to support a DAaaS solution a complete Big Data stack needs to be implemented, in a way that it can run over a cloud platform and maintaining the intrinsic characteristics of clouds, like on-demand services, rapid elasticity, multi-tenancy, measured service, etc. This can be quite complex in itself.

First, a big data stack is composed by many overlaid elements[7]. There is the possibility of combining several elements from recognized projects:

- As base data storage we have the multiple options seen before when discussing about the modeling issues: NoSQL (like HBase, Cassandra, MongoDB or Neo4J), In-Memory databases like SAP HANA, 'legacy' MPP databases and even some other more 'exotic' alternatives.
- For the distributed processing capabilities needed to manage big volumes of data, Hadoop is the actual standard platform. However, potential alternatives may exist like those based in In-Memory computing (Spark) or real-time alternatives (Storm, Akka).
- On top of that analytical services can be implemented using different computing languages, some more traditional (Java), some emergent (Scala) and some specific to statistics (R). Base libraries, like Apache Mahout, or MADlib, try to provide reusable building blocks for some typical algorithm implementations.
- Finally visualization applications can be based in multiple tools, some low level (d3.js) and some linked to traditional BI solutions (Tableau, Qlikview, Pentaho…).

It is easy to see that the integration work needed for a complete solution can be quite significant. But also, making this stack a 'well-behaved' cloud platform adds to this complexity. For example, the most popular Big Data project now, Hadoop, was originally designed to run in non-virtual environments especially in what it refers to its data storage capabilities. Performance can suffer degradation in a virtualized environment and this can be important for a DAaaS solution. To overcome this cloud technology providers are adapting the Hadoop platform to run properly in a virtualized environment, like VMware's Project Serengeti, Amazon Elastic Map Reduce or Microsoft's optimized version of Hadoop for Azure.

The main benefit of the DAaaS is to lower the barrier of entry to advanced analytical capabilities, without demanding that the user commits to large internal infrastructures and human resources to the project.

[7] You can see the main elements of a Big Data Ecosystem in the Big Data chapter of Atos report "Ascent Journey 2016"

# Potential Business Use Cases

Data Analytics as a Service, as a general analytic solution, has potential use cases in very different vertical sectors.  Some concrete scenarios being studied by Atos include:

- In the Oil & Gas sector, companies could deploy predictive maintenance solutions for device fleets in remote installations, without deploying very complex solutions in-house. The solution could be rented for short-term specific analysis.
- In the Electrical Utilities sector, DAaaS is the basis of a specific solution to detect Non-Technical Losses, which cover among others, fraud detection. The customer can upload SmartMeter information into the system where it is processed by specific analytical services created and configured by experts in this kind of business analysis.
- In SmartCities solution, the DAaaS service provides analytic capabilities for the very different data sources that are provided by the city, like the sensor networks deployed in the city. As many cities are under big cost-reduction pressure, DAaaS could provide a very cost effective solution.
- In Retail, a DAaaS model can be used for campaign management and customer behavior. This could include the customer internet activities and also activity in physical stores via the customers' mobile devices.
- In Manufacturing, DAaaS can use the ever growing data coming from connected fabrication machines and when matched with demand it can allow optimal production with minimizing scrap and redundancies.

# Benefits of DAaaS

The main benefit of the DAaaS is to lower the barrier of entry to advanced analytical capabilities, without demanding that the user commits to large internal infrastructures and human resources to the project. Instead of a complex custom project the customer follows simpler steps:

- Data Scientists working for the organization explore the AppStore for an Analytical App that fits the problem.
- They rent the Analytical App for a specific time or quantity of data.
- They configure the Analytical App to its needs including, for example, the usage of external data sources provided by the DAaaS.
- Then the data is fed from the internal systems to the Analytical App.
- The SMEs in the company validate the results and even enhance them with some customization.
- Outcomes are available for all other uses.

Compare this with the typical internal Big Data project:

- Data Scientists need additional resources to design and implement the solution.
- Installs a complete Big Data infrastructure based in some complex technology like Hadoop.
- Implements complex analytical processes in low-level languages becoming in reality an expensive coder.
- Integrates the new system with your enterprise systems with more development effort.
- Examines the results and reiterates until achieving success.

Certainly using DAaaS may not be as direct as using other kinds of SaaS software. Any analytical process demands certain preparatory work: explore initial data, define analytical processes, implement and validate results using test data and optimize it as new data comes. But even so, effort is diminished. And that is without taking into account the benefits of a Cloud delivery model: no upfront costs in infrastructure and a pay per usage model allow experimentation or even temporal usage scenarios.

Technically we've seen the complex issues that the implementation of a functionally complete Big Data Analytics solution needs to overcome, if developed internally by an organization. So a DAaaS solution minimizes this technical complexity even more if it is properly designed to manage a hybrid cloud model, for those cases where information needs to be on-premises.

Also there is the issue of expertise scarcity: Data Science is hard and expert resources are not easily available. DAaaS doesn't eliminate the need for Data Scientists but alleviates some of the problems as some pre-packaged applications are provided for specific use cases. In addition to providing a DAaaS platform, analytic services companies can offer to their customers' access to Data Scientists on demand. This way combining a growing collection of Analytical Services and Data Scientist expertise, the richness of the platform and the value for customers grows.

# Conclusions

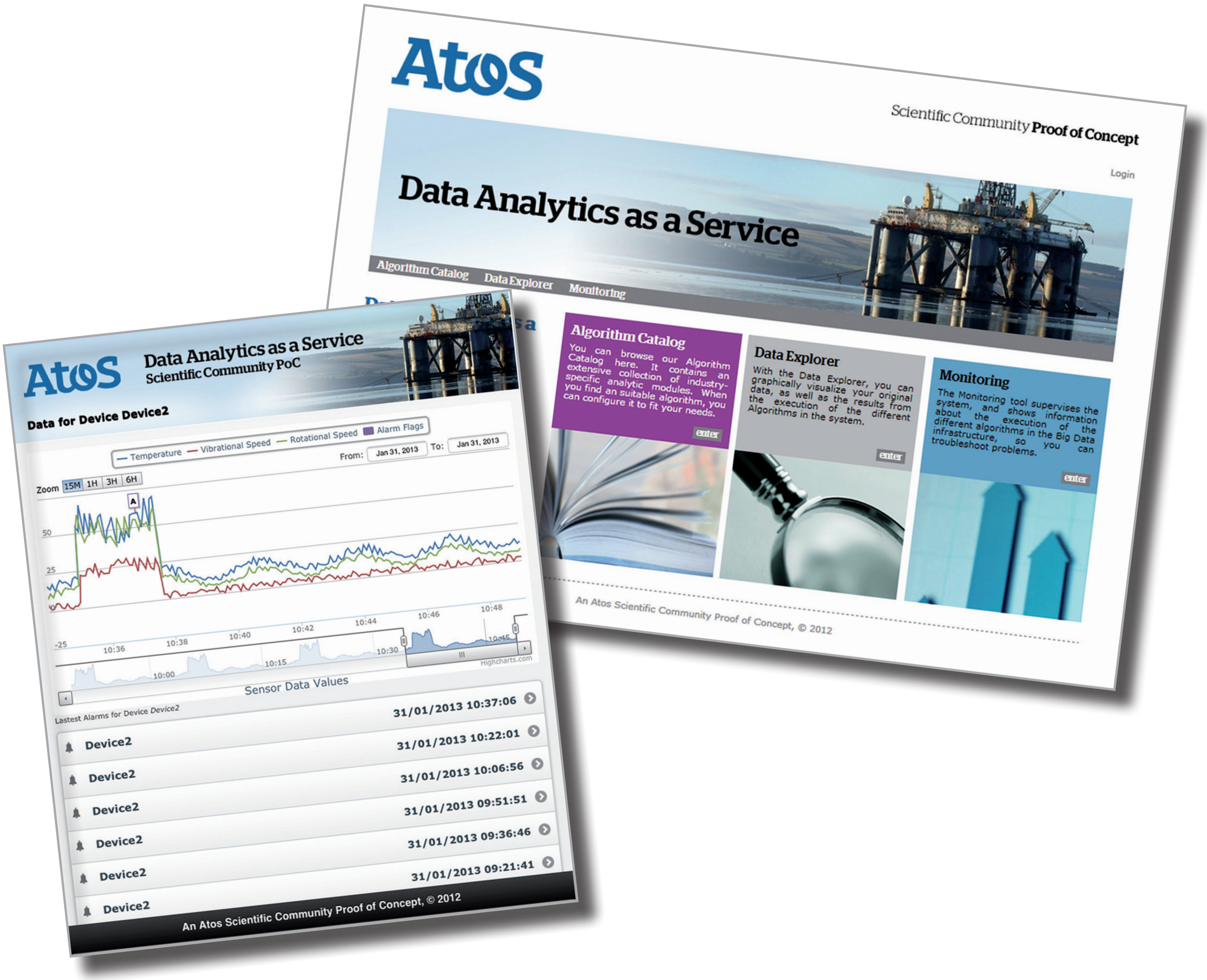
However hyped it may be currently Big Data is certainly a business changing trend, as the facts are evident: the data explosion is real and some companies have shown clear competitive advantage by creating and implementing new analytic capabilities over previously unused data. But getting this kind of capability may be not easy for some companies. Here the flexibility that Cloud delivery models bring can simplify adoption for some companies and even those that could have the resources to implement it internally can obtain significant cost advantages with DAaaS.

Data Analytics as a Service, the model we propose in this paper, can be applied to multiple use cases and industries even as the analytic approaches to different scenarios may vary considerably. Beyond that DAaaS puts analytics as a first-level element component in a new vision of Enterprise Computing which makes extensive usage of the advantages of Cloud technologies.

# Annex: Data Analytics as a Service Proof of Concept (PoC)

In 2012 the Atos Scientific Community conducted research that materialized in the development of a Proof of Concept (PoC), designed to validate the ideas of the DAaaS platform which is described in this White Paper and to serve as a starting point for the development of a commercial platform.

More information on the functionality and technologies of this Proof of Concept can be asked to Celestino Güemes (celestino.guemes@atos.net).



**Algorithm Catalog**

You can browse our Algorithm Catalog here. It contains an extensive collection of industry-specific analytic modules. When you find an suitable algorithm, you can configure it to fit your needs.

**Data Explorer**

With the Data Explorer, you can graphically visualize your original data, as well as the results from the execution of the different Algorithms in the system.

**Monitoring**

The Monitoring tool supervises the system, and shows information about the execution of the different algorithms in the Big Data infrastructure, so you can troubleshoot problems.

An Atos Scientific Community Proof of Concept, © 2012

**Data for Device Device2**

Lastest Alarms for Device *Device2*

| | |
|---|---|
| Device2 | 31/01/2013 10:37:06 |
| Device2 | 31/01/2013 10:22:01 |
| Device2 | 31/01/2013 10:06:56 |
| Device2 | 31/01/2013 09:51:51 |
| Device2 | 31/01/2013 09:36:46 |
| Device2 | 31/01/2013 09:21:41 |

An Atos Scientific Community Proof of Concept, © 2012

# About Atos

Atos SE (Societas europaea) is an international information technology services company with annual 2012 revenue of EUR 8.8 billion and 76,400 employees in 47 countries. Serving a global client base, it delivers Hi-Tech Transactional Services, Consulting & Technology Services, Systems Integration and Managed Services. With its deep technology expertise and industry knowledge, it works with clients across the following market sectors: Manufacturing, Retail & Services; Public sector, Healthcare & Transports; Financial Services; Telecoms, Media & Technology; Energy & Utilities.

Atos is focused on business technology that powers progress and helps organizations to create their firm of the future. It is the Worldwide Information Technology Partner for the Olympic and Paralympic Games and is quoted on the NYSE Euronext Paris market. Atos operates under the brands Atos, Atos Consulting & Technology Services, Atos Worldline and Atos Worldgrid.

For more information: www.atos.net.

atos.net