

Position paper

Contents

- 03 Datalake is the answer to data highest stakes
- 04 Enabling visibility and control of data
- 05 Use cases
- 06 Codex Datalake Engine
- 07 Atos & Cloudera
- 08 Towards a hybrid datalake architecture
- 08 Datalake is the answer to data highest stakes

Datalake is the answer to data highest stakes

“Though 85% of companies are trying to be data-driven, only 37% of that number say they’ve been successful”

As of this moment, only 0.5% of all accessible data is analyzed and used. Imagine the potential here.

Building trust in your data

Digital transformation is a business requirement for organizations. Data is critical in executing this transformation, enabling innovation and competitive differentiation.

The results for data-driven organizations are better decision-making abilities with an immediate positive impact.

Increasing data within companies create new needs:

- Capacity to request in real time interactive dashboards from multiple sources
- Guarantee flexibility of platform to follow business growth & data graduality
- Take real time decisions based on accurate data
- Ensure highest security & guarantee sovereignty of data

“About three out of five people in leadership roles say a failure to get on board with big data could lead to obsolescence”

These needs arise new questions and challenges in the use of datalakes:

- Who in my organization is using data and how?
- Is my data secure and can I trust it?
- How do we handle large quantities of structured and unstructured data?
- Who in my organization is using data and how?
- Is my data secure and can I trust it?
- How do we handle large quantities of structured and unstructured data?

Data-driven digital transformation requires robust and suitable data architecture with strong governance not only to maintain security but also to enable activities such as analytics, enterprise artificial intelligence and interconnected objects.

Here is where Codex Datalake Engine offers the solution. Meeting your trust and compliance needs and leaving you free to get maximum value from your data.

Codex Datalake Engine is a well-defined, new generation of data lake solution. It is an end-to-end data management and security platform, which enables organizations to build a highly scalable, easy-to-use and cost-effective hybrid or on-premise only datalake. It is certified by Cloudera.

Enabling visibility and control of data

Chief data officers, data stewards have the following concerns:

- How is data being used?
- How can data be optimized?
- How do you efficiently manage the lifecycle of your data?
- How do you overcome data silos?
- How do you ensure data regulations are met?
- How can I guarantee a high ROI?
- How can I implement seamlessly a datalake without all the skills required?
- Can I have a cloud platform on an on- premise datalake?

Compliance groups track and protect access to sensitive data. Their primary task is to always be prepared for an audit: tracking who is accessing data, what data they're accessing and what it is being used for.

Their job is to ensure that sensitive data is well governed and protected and in line with General Data Protection Regulation (GDPR) within the European Union and the European Economic Area.

Data scientists, AI developers, BI users need to find the data that matters the most for their business. They want to be able to explore data, trust what they find, and be able to visualize relationships between data sets and make the most from their data.



<p>Finance</p>	<ul style="list-style-type: none"> • Globally distributed datalake for AI-driven document analysis and content detection • Datalake platform industrialization and datascience/ML services to support various use cases, such as customer profiling and scoring for credit renewal and churn prevention (data treasure). • Consulting & data-science services for a datalake to support corporate social responsibility (prevent fines when a company goes bankrupt)
<p>Global functions</p>	<ul style="list-style-type: none"> • Datalake and analytics factory for marketing, sales operations & Finance • Datalake & analytics platform to provide a consolidated vision of the group performance • Map external data (market studies, information news, social media) with internal data (financial results,
<p>Government</p>	<p>National big data platform for French Ministry of Defence:</p> <ul style="list-style-type: none"> • Ships & aircraft monitoring • Knowledge share • Maintenance of equipment • Personnel health • Network threats analysis • Data processing and exploitation
<p>Media & events</p>	<ul style="list-style-type: none"> • Datalake industrialization, AI and data-science services for automated image and video processing • Datalake for connected stadium, customer 360° and predictive analysis initiative (ticket sales prediction through ML algorithms). • Capacity to analyze and cross data from thousands of external & internal media sources to evaluate a trend, social stake or movement
<p>Utilities. oil & gas</p>	<ul style="list-style-type: none"> • Consumption forecast • Smart meters implementation <p>On a recent wastewater treatment project, Atos and Cloudera assisted the plant team in identifying diverse opportunities for optimization, namely by:</p> <ul style="list-style-type: none"> • Building one platform to process multiple data sources such as clean historical data, batch and real-time IOT data collected at the plant, and third-party external data like weather forecasts, population variations and behaviors, and other data affecting its operations • Potential optimization impact for time, quality and cost factors were identified to assist in real-time decision-making

Codex datalake engine

Codex Datalake Engine is a preconfigured, scalable, easy-to-use and fully virtualized appliance. It is cost-effective with minimal administration needs or energy usage. As a result, organizations spend less time installing, tuning, operating, troubleshooting, patching, upgrading, and dealing with integration, adoption of technologies and scale-related issues.

Leveraging the power of the ultimate BullSequana S and SA20 server

Codex Datalake Engine is delivered as an appliance that features BullSequana S and SA20, two of the most agile, scalable and open servers. With their dynamic reconfiguration capabilities, BullSequana S, SA20 and SA20G server combine exceptional performance with unprecedented levels of agility and efficiency.

BullSequana S

The BullSequana S server range features from 2 to 32 Intel® Xeon® Scalable processors in a single server.

- Codex Datalake Engine for development

For development and test purposes Codex Datalake Engine is bundled into a single compact server, scaling from BullSequana S200 (2 CPUs) hosting 5 TB of user data, up to BullSequana S800 (8 CPUs) hosting 20 TB to 50 TB of user data.

- Codex Datalake Engine for production

For production configuration starts with 3 BullSequana S200 and can be extended up to 12 BullSequana S200, with 50TB to 350 TB of user data. For hefty workloads, Codex Datalake Engine leverages the BullSequana S400, starts with 6 BullSequana S400 and can support several PB of user data with 60 TB upgrade increments.



BullSequana SA20 & SA20G

With its dynamic reconfiguration capabilities, BullSequana SA20 & SA20G server combines exceptional performance with a limited carbon footprint.

BullSequana SA20 and SA20G rack servers powered by 2nd generation AMD EPYC™ processors bring a cost-effective balance of performance and virtualization. By offering up to 100TB in a 2U form factor, and best-in class SAS/SATA and ultra-fast NVMe drives, they provide optimal performance for on-premise or hybrid datalake architecture. In addition, this outstanding density makes it possible to minimize the footprint and improve the energy efficiency.

BullSequana SA20 best-in-class server for stockage

- Storage on multiple hard disks allows data redundancy and performance improvement. RAID technology provides the right balance between reliability, availability, performance, capacity - and cost!
- Enabled up to 2 TB RAM and to 100 TB storage powered by 2 CPUs

BullSequana SA20G best-in-class server for Ai inference

- Offering a powerful AMD CPU's and up 3xA100 GPUs combining ML and deep learning capabilities in a single platform.
- Enables Spark 3.0 applications to run on the serve thanks to GPU enablement, accelerated Spark components and end-to-end integrated data pipeline for CPU/GPU.
- Offering flexible provision of GPU resources for data scientists to integrate deep learning capabilities into Cloudera Datalake

An appliance certified by Cloudera

Cloudera provides a scalable, flexible, integrated platform that makes it easy to manage rapidly increasing volumes and varieties of data. Cloudera products and solutions enable to deploy and manage Apache Hadoop and related projects, manipulate and analyze data, and keep that data secure and protected.

Codex Datalake Engine is Cloudera certified. It is the result of a joint effort between Atos and Cloudera to achieve a solid and trustworthy data architecture and deliver the most complete, secure, industrial and qualified data lake solution on the market.



Atos is Cloudera's
EMEA Partner of
the Year 2020

Cloudera CDP

Cloudera evolved its offering by merging the best of Cloudera Enterprise Data Hub and Hortonworks Data Platform Enterprise along with new features and enhancements across the stack.

With the all-new Cloudera Data Platform, the field of possibilities of analytics on any Cloud is widen - enabling the speed and agility the business wants with the security and governance the enterprise demands. Cloudera Data Platform (CDP) is an integrated data platform that is easy to deploy, manage, and use. It delivers powerful, self-service analytics across hybrid and multi-cloud environments with the granular security and governance policies that IT leaders demand.

CDP is a new approach to enterprise data that delivers a full range of analytic capabilities from the Edge to AI.

CDP is different from other data platforms and analytics services in four important ways:

Any cloud - CDP provides maximum choice and flexibility with the option to manage, analyze, and experiment with data on-premises, in hybrid, private cloud, and multiple public cloud environments.

Multi-function - CDP reduces the time and effort to deploy common application types with five new self-service experiences: flow & streaming, data engineering, data warehouse, operational database, and machine learning.

Secure and Governed - CDP simplifies security, privacy, and compliance for diverse enterprise data on any cloud through shared data experience (SDX) technologies. SDX makes it easy to create a secure data lake in hours instead of weeks and replaces tedious scripting with "set it and forget it" convenience.

Open - CDP is 100% open source, open compute, open storage and open for integration - enabling rapid innovation and protecting organizations from vendor lock-in.

Migration to CDP

As Codex Datalake Engine is Cloudera CDP certified, Atos offers a migration accelerator towards CDP for a seamless implementation. Repurposing existing hardware by gradually migrating data and workloads to the new cluster.

“While enterprises are rapidly adopting data platform technologies and are eager to explore the cloud for driving analytical workloads, many face challenges in trying to leverage all of their data to achieve better business outcomes, solutions such as the forthcoming Cloudera Data Platform help enterprises navigate complex data processes across multiple clouds, manage data governance, and enable multi-function analytics, regardless of where the data resides.”

James Curtis, Senior Analyst for the Data, AI & Analytics Channel at 451 Research

Towards a hybrid datalake architecture

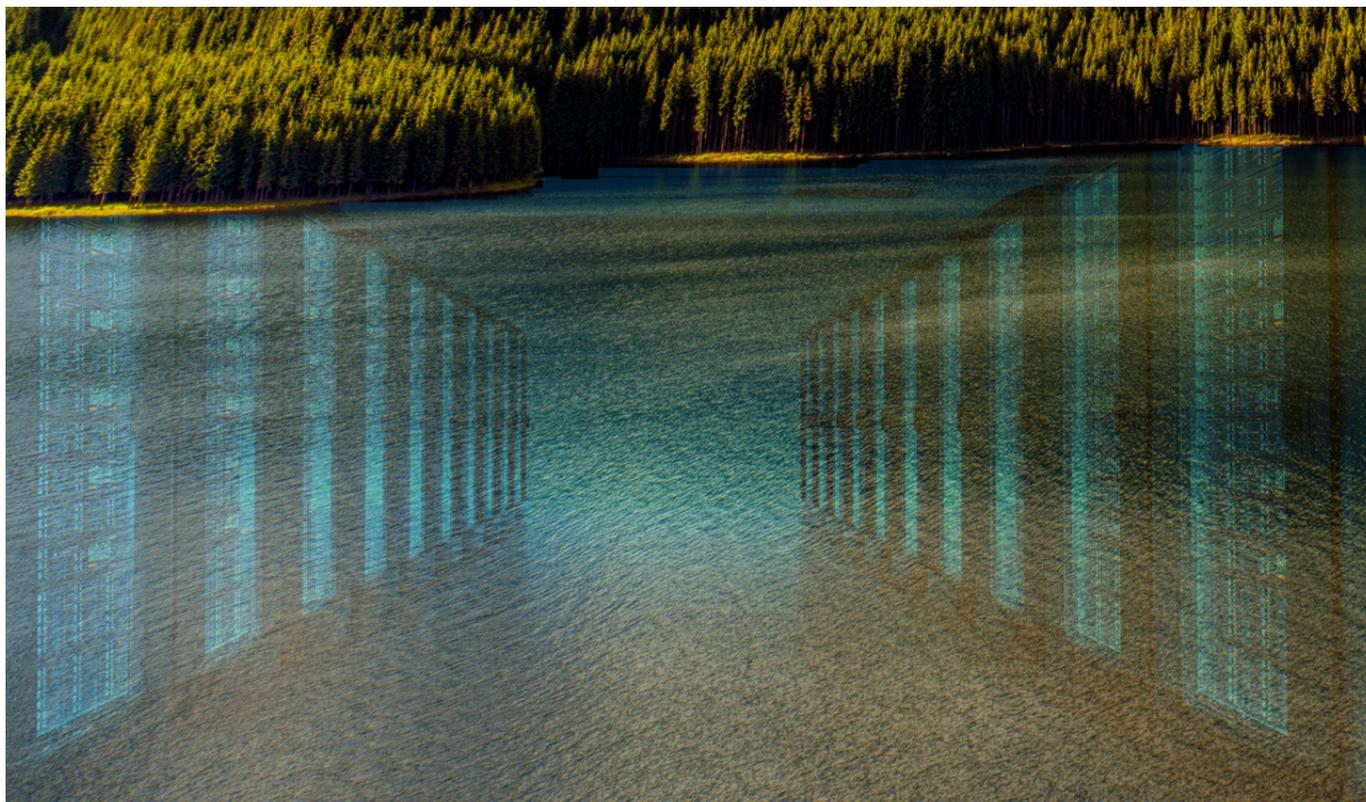
Atos has developed a platform approach that consists of horizontal hybrid cloud enablement, providing technology and cloud enterprise agnostic data approach, so that functionality can run anywhere in cloud, in DC or on the Edge, connecting these for the realization of pervasive business end-to-end use cases, that bring insights, guidance and assistance to where the business end user needs it.

Codex Datalake Engine can be the foundation for a hybrid datalake.

- Integrated with cloud platforms
- Highly scalable to grow with your business
- A granular architecture to support heterogeneous workloads (data intensive, data & compute intensive, non compute intensive & compute intensive)
- A single point of contact

Key advantages:

- Time saving for implementation
- Combine on-premise high capacities with cloud platforms & private cloud
- High ROI for mid-long term projects
- Open platform
- Prevents vendor lock-in
- Enables to take advantage of an ecosystem of additional tools and capabilities
- Build capabilities next to legacy applications and platforms
- Maintained GPU cost



Datalake is the answer to data highest stakes



Data quality

It ensures data quality as a consistent data management and security appliance, providing a data governance solution and the management of the complete data life-cycle, from data ingestion, data cleansing, data blending, data discovery, audit, data lineage and policy enforcement. Codex Datalake Engine improves and ensures data quality and makes data trustable.



Security

Organizations are very concerned with data security and protection. With Codex Datalake Engine, sensitive data can remain on-premise in both hybrid or on-premise datalake architectures. Organizations keep full control over their data and its lifecycle as well as full control over the infrastructure, the applications and the operations, enabling more compatibility and thus minimizing the risk of failure. The complete virtualization minimizes the configuration costs and avoids extra cost when introducing new applications.



Data encryption & key management

Provides a critical layer of protection against potential threats. Encryption and key management are also required for meeting key compliance initiatives and ensuring the integrity of enterprise data. It transparently encrypts and secures data without requiring changes to applications and ensures there is minimal performance lag in the encryption or decryption process. It uses a trustee server as enterprise grade virtual safe-deposit box to store and manage cryptographic keys.



Optimized datalake

Full power in a virtualized and cost-effective design. All components of Codex Datalake Engine are integrated into a compact, cost optimized platform. For this, Codex Datalake Engine takes advantage of virtualization, which allows it to optimize the hardware sizing needed for each software component, instead of using oversized separate machines



Scalability

Codex Datalake Engine is the foundation to develop Big Data and AI applications while business growth. By leveraging all available data across organizations, it gives a reliable and comprehensive view of data. Analytics and AI developers have a full visibility and control of all data.





About Atos

Atos is a global leader in digital transformation with 110,000 employees in 73 countries and annual revenue of € 12 billion. European number one in Cloud, Cybersecurity and High-Performance Computing, the Group provides end-to-end Orchestrated Hybrid Cloud, Big Data, Business Applications and Digital Workplace solutions. The Group is the Worldwide Information Technology Partner for the Olympic & Paralympic Games and operates under the brands Atos, Atos|Syntel, and Unify. Atos is a SE (Societas Europaea), listed on the CAC40 Paris stock index.

The purpose of Atos is to help design the future of the information space. Its expertise and services support the development of knowledge, education and research in a multicultural approach and contribute to the development of scientific and technological excellence. Across the world, the Group enables its customers and employees, and members of societies at large to live, work and develop sustainably, in a safe and secure information space.

Find out more about us

atos.net

atos.net/career

Let's start a discussion together



For more information: clement.yunes@atos.net

Atos, the Atos logo, Atos|Syntel, and Unify are registered trademarks of the Atos group. July 2020. © 2020 Atos. Confidential information owned by Atos, to be used by the recipient only. This document, or any part of it, may not be reproduced, copied, circulated and/or distributed nor quoted without prior written approval from Atos.