

Empowering the Data-driven Enterprise with Solix Common Data Platform



Index

- 2** Executive Summary
- 3** The Future of Enterprise: Data
- 5** Becoming A Data-driven Organization
- 6** Data Warehouse and Its Limitations
- 7** The Hadoop Alternative: The Data Lake
- 10** The Solution: Solix Common Data Platform
- 13** About Solix Technologies

According to Gartner¹ every business is facing unprecedented change at a speed never seen before. The age of “smart” devices and the rise of machine intelligence is here²

Executive Summary

Data powers the most successful enterprises – and it is everywhere, ready to be mined for new insights. Business leaders recognize that new data horizons are constantly emerging and need to be explored for business opportunities.

Organizations such as Amazon, Airbnb and Uber have understood the power of data and have been pioneers in the push towards the digital economy. They are mining data to disrupt traditional industries and to create new parallel digital economies.

This trend of disruption driven by the power of data is spreading rapidly across banking, insurance, healthcare, energy and even government sectors. To seize the opportunities being presented by the digital economy, organizations must be prepared for the onslaught and variety of data available. A new paradigm in information management is required.

Current enterprise architecture offerings are not sufficient. Business stakeholders require data – in all its forms – to make real-time decisions. CIO’s, on the other hand, must keep infrastructure costs down, while dealing with complex governance and Information Lifecycle Management (ILM) for data on the petabyte scale and beyond. The

tension is clear. Forrester urges enterprises not just to ease the tension, but erase it and forge new partnerships to ensure business stakeholders have the infrastructure they need to make data-driven decisions and evolve in real-time.³ Enterprise architecture must be ready to respond.

The Solix Common Data Platform unites the best of Hadoop and the Enterprise Data Warehouse (EDW) into a new vision, one that will allow organizations to put the promise of visionary data at their fingertips. It leverages an organization’s existing infrastructure and allows organizations to collect, store and analyze massive amounts of data from every source without sacrificing governance, security or management. This new paradigm allows for a meaningful, frictionless partnership between IT departments and business users. Further, with the Solix Common Data Platform – all data keeps its context, meaning decisions are based on facts and not interpretations, helping fuel transformation to a data driven enterprise.

Source: Solix

Solix Common Data Platform leverages an organization’s existing infrastructure and allows organizations to collect, store and analyze massive amounts of data from every source without sacrificing governance, security or management.

¹ Gartner, “Top Strategic Predictions for 2016 and Beyond: The Future Is a Digital Thing”

² Gartner, “Top Strategic Predictions for 2016 and Beyond: The Future Is a Digital Thing”

³ Forrester, “Put The Business Back in Your Data Management Business Case”

The Future of Enterprise: Data

As the digital economy gains ground, the factors driving a business shift. According to Harvard Business Review, “Decision-making processes are becoming more standardized, with data as the foundation and starting point for discussions.”⁴ Data will continue to empower people, technology and processes. Data allows organizations to understand their customers, predict their needs and interest, and respond faster. Data has also become the backbone for everything to do with improvements in product quality and capabilities. It even allows organizations to understand themselves and operate more efficiently through analysis of business cycles, employee habits and customer needs.

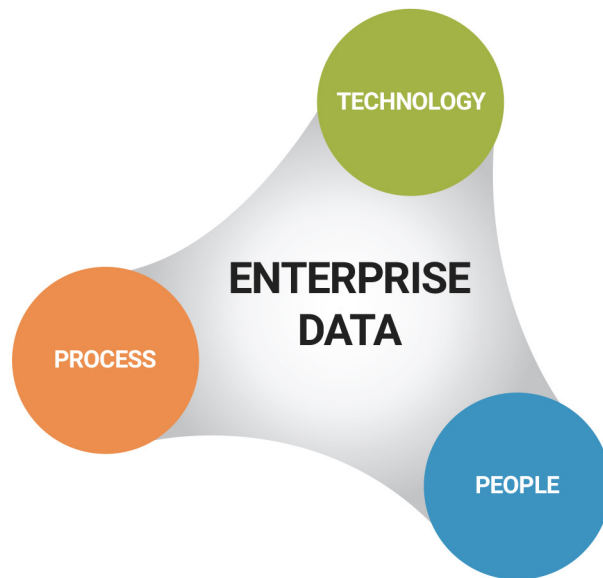
Let us take the example of one of the world’s largest aerospace company. Boeing’s jet engine

create 20 terabytes of data per hour. That data has allowed Boeing to improve its engine by analyzing the information created by its use.⁵

Now imagine the potential of advanced real-time analytics based on emerging data streams from smart devices. The opportunity of collecting data insights from streaming devices still remains a largely untapped frontier. However, it is poised for a change. In fact, Gartner predicts spending on new IoT hardware will exceed \$2.5 million a minute this year.⁶ New varieties of data sources are beginning to emerge, and more and more organizations are equipping themselves to tap into this opportunity. The question is — “is your organization ready for a future driven by data insights?”

The Harvard Business Review notes in a survey it conducted that the window for decision-making has shrunk in recent years as businesses become more and more digital⁴

FIGURE 1 Enterprise Data



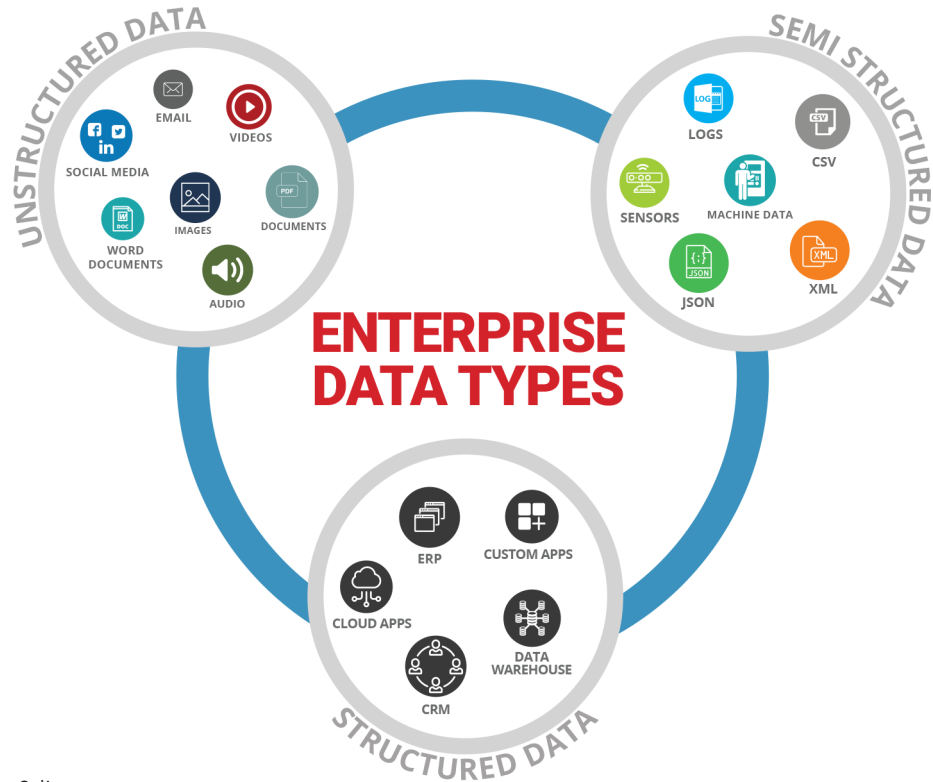
Source: Solix

⁴ Harvard Business Review, “The Evolution of Decision Making: How Leading Organizations Are Adopting a Data-Driven Culture”

⁵ Forrester, “Big Opportunities in Big Data”

⁶ Gartner, “Top Strategic Predictions for 2016 and Beyond: The Future Is a Digital Thing”

FIGURE 2 Enterprise Data Types



Source: Solix

Enterprises must now assume competitors are capturing data in all its varieties and analyzing it in real-time to turn it onto actionable insights.

Data Empowers Disruptors To Create New Economies

- Amazon understood the power of data years ago and used it to not only change e-commerce, but change the very idea of what an online business could be. Amazon built itself upon its knowledge of its customers — tracking every click, not just purchases — and, in the process, changed how the world shops.
- Uber uses Big Data to gather customer insights and became a world-wide phenomenon by doing it. Data literally drives Uber. It allows the organization to choose areas to expand into and where and when to position cars to meet customer need and satisfaction

⁷ Gartner, "Top Strategic Predictions for 2016 and Beyond: The Future Is a Digital Thing"

Gartner has identified a "Nexus of Forces" — social networking, cloud computing, mobile communications and information — that will continue to evolve and change the business landscape.⁷

Becoming A Data-driven Organization

Data has become the biggest and most valuable asset an enterprise has. For enterprises to succeed and evolve they must become data driven. It is important that decisions on investment into data are driven by business needs and not infrastructure costs.

Business Intelligence and Advanced Analytics are the foundation upon which organizations are making decisions not only to survive, but to thrive in the new world economy. For the CIO, this means developing an enterprise architecture strategy that will be responsive to the business stakeholders without driving up the investment in hardware and software.

Organizations across industries are looking to Advanced Analytics to solve significant challenges, provide better service and to discover new opportunities. Two industries – healthcare and finance – are using analytics to create unprecedented levels of transformation for these very reasons.

Advanced Analytics are allowing Healthcare organizations to survive an unprecedented shift in their business models. Analytics are paving the way to determine new service models, coordinate a continuum of care and help focus staffing and operational efficiencies.

The financial industry finds itself at the forefront of the digital economy and the emergence of cashless society. Advanced analytics make it possible for institutions to ensure security and governance, while allowing customers to pay for goods and services with a variety of emerging digital options. This more inclusive and cashless economy seemed impossible just a few years ago. Data and Advanced Analytics are at the center of this transformation.

“Gartner predicts that by 2018, more than half of large organizations globally will compete using advanced analytics and proprietary algorithms, causing the disruption of entire industries.⁸”

– Jim Hare, Gartner Analyst

FIGURE 3 Data-driven Organization



Source: Solix

⁸ Gartner, “Top Strategic Predictions for 2016 and Beyond: The Future Is a Digital Thing”

CIO's are tasked with the challenge of capturing, organizing and managing data in a way that supports thorough and accurate analysis and ensures governance, security and ILM are supported

- Data has become the biggest and most valuable asset of an enterprise
- Business Intelligence and Advanced Analytics are the foundation upon which organizations are making decisions not only to survive, but thrive in the new world economy

Data Warehouse and Its Limitations

The Enterprise Data Warehouse offers the opportunity to mine structured data with analytics for improved business results. The EDW captures the relational data from various transactional systems and supplies access to multiple important record systems. All data retains its metadata and the relationships between pieces of information are kept clear. Traditional EDW platforms deliver highly specific data views based on corporate strategy. The EDW data must follow the time-

consuming extract, transform and load (ETL) process to be of use, creating limitations on the data use.

This canonical, top-down enterprise approach is deductive and has merits, but it alone will not allow organizations to stay competitive in the emerging business climate.

The EDW limitations stifle progress toward becoming data-driven:

- Only allows schema-on-write — requiring organizations to define the business questions they want answered ahead of time
- Highly dependent on the platform it is built on, making enterprises very reliant on the vendors
- Highly complex, requiring specialized technologists both to manage it and run queries to generate reports. Business users in most cases are not able to query the database directly, further stifling iterative research
- Costly, since it must be built on Tier 1 storage

Additionally, a data warehouse cannot capture all the forms of data business users clamor for. The capital investment of building a data warehouse to deal with exabytes of data is far too costly. It also fosters a top-down use of the data where data is used to answer pre-defined business questions often a month or longer after the events occur. Ad hoc, interactive querying where each answer leads to another question as the user achieves valuable insights is impossible since the data schema that

defines – and limits – the questions that can be asked is imposed during data ingestion itself.

This does not mean that the data warehouse is not valuable or that it cannot answer important business questions. It does mean the EDW cannot answer new questions and therefore cannot deliver new insights. What if the business does not yet know the questions it seeks answers for?

“What if the business does not yet know the questions it seeks answers for?”

This is where Big Data technologies such as Hadoop take over from the data warehouse. Because the data schema is imposed as part of the query process rather than during data loading, users can ask any question. This allows business

users to work directly with the data in an iterative fashion where one query can lead to the next and so on. This often produces deeper and more actionable insights.

The Hadoop Alternative: The Data Lake

Big Data and Hadoop offer the capability to capture, store, organize and analyze massive amounts of structured, semi-structured and unstructured data within the Data Lake.

Apache Hadoop provides a free and open source computing framework designed to operate powerful, low-cost infrastructure in a less expensive tier, while still delivering massive

scalability and performance. Using the MapReduce or Spark programming models, the Data Lake can process large data sets across distributed compute nodes in parallel. All this means Hadoop offers dramatic cost savings over traditional Tier 1 infrastructure required by EDW. It also enables storage and processing of massive amounts of data.

Hadoop provides a foundation for The New Enterprise Blueprint and allows organizations to become data-driven.

While all the top data center hardware infrastructure challenges impact cost to some degree, data growth is particularly associated with increased costs relative to hardware, software, associated maintenance, administration and services.⁹

– April Adams,
Research Director,
Gartner

⁹ “Big Data Benefits Are Hampered by ‘Culture Clash,’” Gartner, Sept. 12, 2013

By 2020, 10% of organizations will have a highly profitable business unit specifically for productizing and commercializing their information assets.¹⁰

– Douglas Laney,
Alan Duncan,
Gartner Analysts

With Data Lakes built on Hadoop, organizations can create repositories holding a vast amount of raw data in its native format until it is needed without imposing schema or requirements. When a business question arises, a Data Lake can be queried for relevant data and a schema that is tailored to the question can be applied on demand. Storing data “as-is” in an Enterprise Data Lake enables downstream NOSQL and SQL applications to describe the data better at a later time when it is needed. Text search and structured query are also available enabling universal access to all data. Organizations no longer need to know the business questions they seek answers to, rather the data itself can be used to determine the questions.

While the benefits of Hadoop are clear, so are the limitations. Security and governance capabilities are immature and the Data Lake can easily become full of data that may not represent the best and most accurate information. Additionally, Hadoop does not capture any metadata. Thus, no one knows the context of the data, where a particular data set originated, how trustworthy the source might be, when it was captured, etc. The lack of metadata negatively impacts the ability to locate the right data to answer the question.

The lack of metadata negatively impacts the ability to locate the right data to answer the question.

Challenges with the Data Lake include:

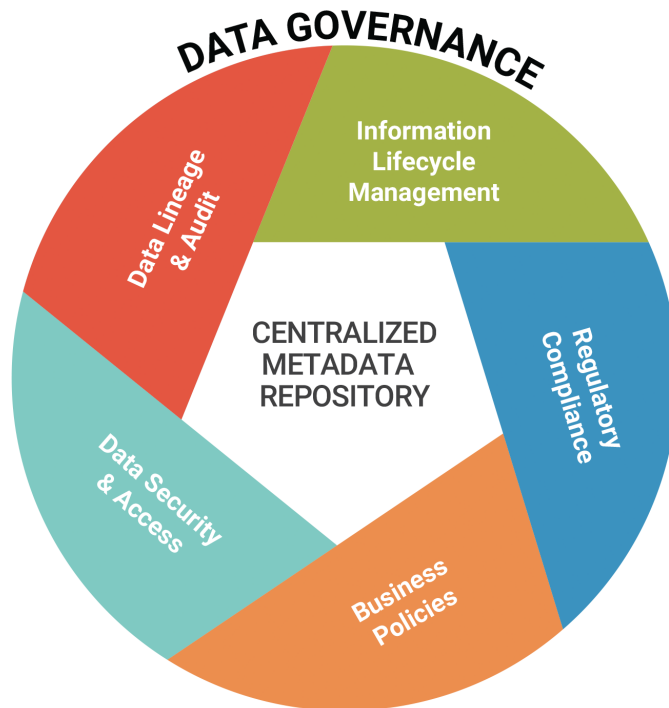
- Lacks the ability to capture and maintain metadata, so the data loses all context
- Cannot maintain the relevancy of data over time
- Security is immature without metadata
- Lacks support for Information Governance

¹⁰ Gartner, Predicts 2016: Information Strategy

Additionally, Data Governance has never been more critical, but without metadata it is also impossible. Security and privacy become impossible to ensure. With more data, risk

increases exponentially, and exposing an enterprise to such hazards is not acceptable.

FIGURE 4 Data Governance



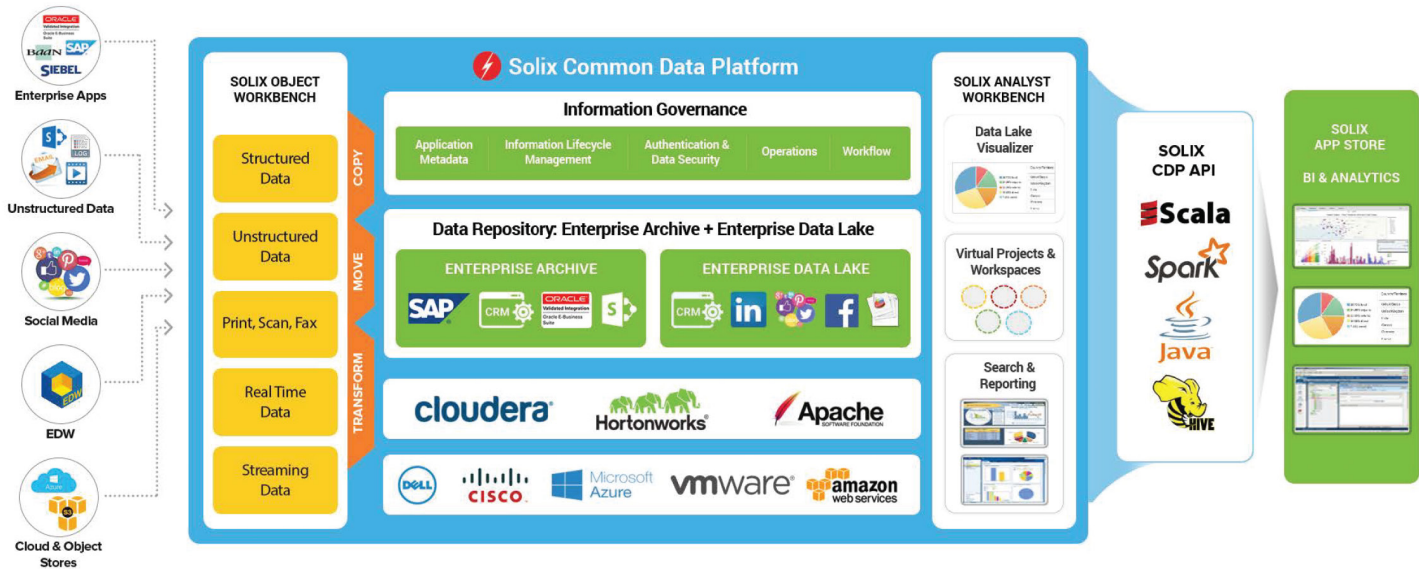
Source: Solix

“A robust ILM strategy focused on governance and security requirements must be an essential part of the evolution of an organization’s overall data strategy”

Without a well-defined data governance and a centralized metadata repository, the Data Lake can easily turn into a data swamp, rife with risk and too tangled for much reward.

Neither the EDW nor the Data Lake can manage the data needs of today’s enterprises on their own. A robust ILM strategy focused on governance and security requirements must be an essential part of the evolution of an organization’s overall data strategy.

FIGURE 5 Solix Common Data Platform



Source: Solix

The Solution: Solix Common Data Platform

Solix Common
Data Platform

=

Enterprise
Archiving

+

Enterprise Data
Lake

+

Information
Governance

The Solix Common Data Platform (CDP) brings enterprise grade capabilities to the Hadoop framework and addresses all the shortcomings of a Data Lake. Solix CDP provides uniform data collection, metadata management, information lifecycle management, governance and secure data access for advanced analytics.

The Solix information governance layer, which includes metadata management, Information lifecycle management, security and authentication coupled with Data Lake visualizer gives in-depth control of the Data Lake to the IT department, to ensure the Data Lake does not turn into a data swamp.

This is vital in the present fast evolving Big Data environment, in which new technologies are arising continuously in response to new business needs. Wikibon Big Data Analyst George Gilbert points out that Hadoop, which was originally

developed at Yahoo and Google to support search, is already being stretched to its limits by new business uses.¹¹ It is not impossible that an entirely new Big Data database engine technology designed to support those more demanding use cases may appear in the next few years. The Solix Common Data Platform will be able to support any new Big Data technologies and use cases that may arise.

¹¹ "Growing Complexity in Adoption Dynamics Behind Wikibon's 2016 Big Data Forecast", Wikibon.com, June 30, 2016

Benefits of The Solix Common Data Platform

- Combines the advantages of Hadoop with the ability to preserve the full metadata
- Provides advanced ILM capabilities, including the ability to copy data from the data warehouse and to archive older data
- Supports advanced data security, as well as third party analysis packages including machine learning and cognitive computing analysis of the data
- Because it preserves all data in its original format and with full metadata and supports established open standard interfaces, it future-proofs the Data Lake, ensuring the data will be usable by new technologies that are as yet unknown and for new use cases as yet undefined
- Provides a unified data governance layer from the time of data ingestion to use of data by end business users for operational insights and advanced analytics

The Solix CDP also addresses the debate around On-prem vs on Cloud deployments. There are many reasons why certain deployments are right for an organization. While most trials so far have been on premise, a growing faction in the Big Data community is advocating the advantages of cloud based data storage and analysis.

The most likely outcome is that Data Lakes will be created wherever the data originates, which means many companies will have hybrid environments for the practical reason that large volumes of data are difficult and expensive to move.

With Solix CDP, organizations can create a unified enterprise-wide Data Lake that can be hosted on prem, on cloud or in a hybrid environment.

Another major issue with Hadoop environments is the complexity of the stack and the new skills required to manage it. This has slowed progress on many Hadoop trials, delaying their evolution into production systems.

The Solix Common Data Platform provides an extensive library of APIs and a platform that can unify multiple technologies that often are not designed to work together. It uses both the MapReduce and Spark programming technology. Thus, it can help alleviate this problem as well.

The Solix Common Data Platform unites the best of Hadoop and the Enterprise Data Warehouse (EDW) into a new paradigm, one that will allow organizations to put the promise of visionary data at their fingertips

	DATA WAREHOUSE	DATA LAKE	SOLIX CDP
Data	Processed, Structured	Structured, Semi-structured/ Unstructured	Processed, Structured, Semi-structured/Unstructured
Schema	On write	On read	On read / On write
Storage Costs	High	Low	Low
Scalability	Low	High	High
Agility	Low, Fixed configuration	High, Configure & Reconfigure	High, Configure & reconfigure
Metadata Repository	Centralized MetaData Repository	No	Centralized MetaData Repository
Data Access	Query	Search	Query + Search
Query Performance	High	Medium	Medium
Security / Governance	Mature	Maturing	Mature
Users	Business Users	Data Scientists	Business Users, Data Analysts, Data Scientists
Role based Access	Yes	No	Yes
ILM	No	No	Yes
Regulatory Retention Management	No	No	Yes
Legal Hold	No	No	Yes
ROI	High	Low	High

The Solix Common Data Platform unites the best of Hadoop and the Enterprise Data Warehouse (EDW) into a new paradigm, one that will allow organizations to put the promise of visionary data at their fingertips. Overall, the Solix Common Data Platform provides a Hadoop-plus data management

environment that can create a true enterprise Data Lake with full access to the data each user needs at a particular moment, rather than a data swamp where data gets lost. To learn more about Solix, [click here](#).

Source: Solix

About Solix Technologies

Solix Technologies, Inc., a leader in empowering data-driven enterprises, helps businesses organize their Enterprise Information with optimized infrastructure, data security and advanced analytics by achieving Information Lifecycle Management (ILM) goals. Solix Big Data Suite offers an ILM framework for Enterprise Archiving and Enterprise Data Lake applications with Apache Hadoop as an enterprise data repository. The Solix Enterprise Data Management Suite (Solix EDMS) enables organizations to implement Database Archiving, Test Data Management (Data Subsetting), Data Masking and Application Retirement across all enterprise data. Solix Technologies, Inc. is headquartered in Santa Clara, California and operates worldwide through an established network of value added resellers (VARs) and systems integrators. To learn more, please visit <http://www.solix.com>

