SOLIX®
Empowering the Data-driven Enterprise

SOLIX COMMON DATA PLATFORM:

Advanced Analytics and the Data-Driven Enterprise

EXECUTIVE SUMMARY

"You cannot manage what you cannot measure."

Those prescient words from management guru Peter Drucker more than 30 years ago encapsulated the evolution of enterprise software platforms and have paved the path to the era of the data-driven enterprise.

Data is remaking industries and reshaping the global economy. Those who embraced data found growth areas and improved earnings. Organizations that ignore the promise of data can no longer survive in the new world economy.

Today, Drucker's words are as true as ever. To continue being data-driven, organizations must be able to ingest and analyze new forms of data from constantly developing new sources. Those who are ready to mine new data streams, such as social, IoT and more, are primed to transform economies and reap the benefits with new growth opportunities. Businesses ready to leap into the fray are adopting Big Data. Big Data brings together structured, semi-structured and unstructured data. When all forms of data are brought together, it not only multiplies the value of every single piece of data, but it also presents a new set of challenges around data storage, governance and consumption.

In today's enterprise world, business users want to make real-time, data-driven decisions using the vast amount of data available. Yet, IT departments are faced with the challenge of increasing storage and Business Intelligence (BI) costs, complex governance and Information Lifecycle Management (ILM) for data, which is now in the scale of petabytes and beyond. Unfortunately, current enterprise ready technology offerings are not capable of managing this data tsunami, let alone take advantage of all the possibilities this data offers. The tension created within organizations is clear.

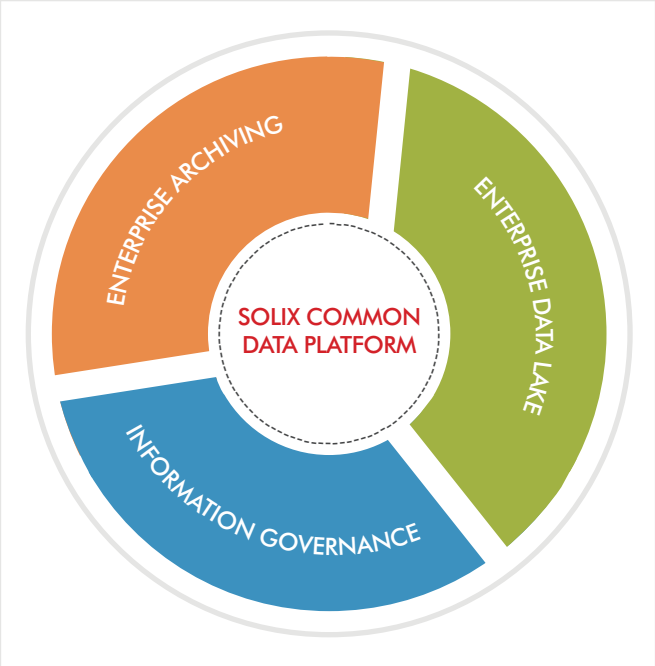
> "As Forrester also points out in its research report, in the era of Big Data, traditional EDW is failing to meet new business requirements, such as support for real-time and ad hoc customer analytics, new sources of data, and self-service capabilities.[1]

The Solix Common Data Platform (CDP) allows organizations to embrace Big Data, while keeping the challenges in check. The Solix CDP helps organizations leverage their existing infrastructure and allows them to collect, store and analyze massive amounts of data from every source without sacrificing governance, security or management. Further, with the Solix CDP all data keeps its original context and structure, allowing organizations to ask complex questions and gain deep contextual insights from data at any point. The Solix CDP creates a new paradigm fostering a meaningful, frictionless partnership between IT departments and business users. IT departments can now become the guardians of data and business users can become the owners and direct consumers of data.

Solix created the CDP to bring ILM to the Data Lake and innovation to the EDW. The Solix CDP is the next evolution in the new enterprise blueprint, offering Enterprise Data Archiving and Enterprise Data Lake to create an Advanced Analytics platform with unprecedented levels of ILM in a Big Data setting.

[1] Forrester Report on The Next-Generation EDW is the Big Data Warehouse, August 2016

INTRODUCTION — SOLIX COMMON DATA PLATFORM



Solix CDP = Enterprise Archiving + Enterprise Data Lake + Information Governance

At the core of the Solix CDP are Enterprise Archive and The Enterprise Data Lake.

The Solix CDP utilizes the Solix Big Data Suite to provide comprehensive enterprise data management and robust ILM. With the CDP, organizations can vastly expand the reach of analytics by creating an Advanced Analytics platform. For the CIO, Enterprise Archiving offers a quick ROI that will ensure budgetary support from the organization and dissolves the obstacles between Big Data and ILM.

The Solix CDP brings enterprise-grade capabilities to the Hadoop framework, addressing all shortcomings of the Data Lake. Solix CDP provides uniform data collection, metadata management, ILM and secure data access for Advanced Analytics.

The Solix CDP does this all while maximizing an organization's existing infrastructure. With no need to reboot the organization's enterprise architecture, the Solix CDP harnesses the current architecture to develop a new enterprise blueprint, capable of evolving with the business requirements of an organization.

The Solix CDP is also capable of evolution. As businesses stretch Hadoop to its limits, new Big Data technologies will emerge. The Solix CDP is primed to adapt with them.

The Solix CDP brings enterprise-grade capabilities to the Hadoop framework, addressing all shortcomings of the Data Lake.

WHY SOLIX COMMON DATA PLATFORM?

The need to become data-driven is clear. Transformation has hit every major industry and disruptors have become powerhouses in the global economy based on their capabilities to mine data. Any organization wanting to compete must become data-driven or it is destined to fail.

The current enterprise architecture offering, EDW, provides a canonical, top-down view of enterprise data to meet end user requirements, but those views rarely satisfy the function-specific requirements of data-driven applications.

> As per Forrester research, Big Data platforms such as Hadoop have made Big Data architectures more affordable, allowing companies to pursue new business insights for increased data-driven competitive advantage.[2]

The Solix CDP, built on top of Hadoop distributions, enables data-driven organizations to gain more value from their data because now data can be visualized in more specific ways.

The cost of relying on the EDW to collect and analyze all of this data would also exceed the budget of most organizations. The Solix CDP is a uniform data collection system for structured, unstructured and semi-structured data featuring low-cost data storage and Advanced Analytics. Solix CDP stores data "as-is" to reduce costly Extract, Transform and Load (ETL) operations, as well as transforms data to feed downstream NoSQL and analytics applications. Solix CDP enables organizations to create a true enterprise Data Lake with full access to the data, rather than a data swamp where the data gets lost. This enables the CIO to find a better solution than trying to collect and store all of the enterprise data in the expensive Tier 1 storage and existing EDW architectural offerings.

The Solix CDP does not require costly infrastructure and offers the scalability and flexibility the Big Data platform architecture provides, along with enterprise-grade governance and security. The Solix CDP lays the foundation for information governance, efficient infrastructure utilization and Advanced Analytics at petabyte scale.

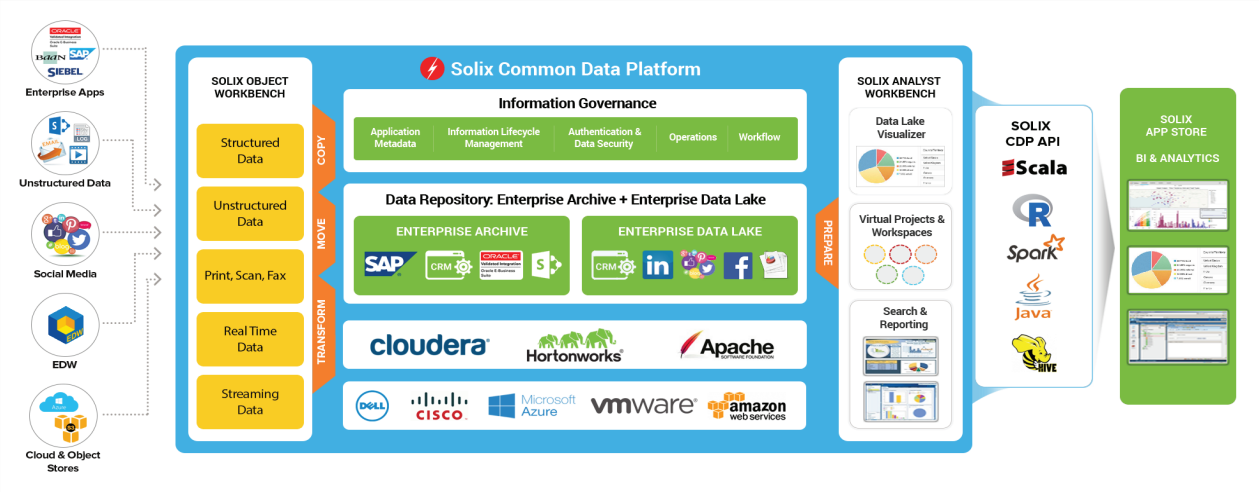
The Solix CDP lays the foundation for information governance, efficient infrastructure utilization and Advanced Analytics at petabyte scale.

[2] Forrester Report on Big Data Fabric Drives Innovation and Growth, March 2016

Here is the comparison on how Solix CDP differs from a traditional Data Warehouse and a Data Lake:

	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

PRODUCT/ SOLUTION OVERVIEW OF THE SOLIX COMMON DATA PLATFORM



Built on top of Hadoop distributions, such as Cloudera CDH or Hortonworks HDP, the Solix CDP provides an integrated suite of enterprise connectors through its Object Workbench to build a consolidated repository of enterprise data and metadata. The Solix Analyst Workbench allows multiple teams and users to collaborate, create virtual workspaces and projects to access the data without compromising on compliance and security. Because it runs on top of both of the most popular Hadoop distributions, it eliminates one of the basic questions behind the creation of a Hadoop stack, and it can bridge the two in environments where both are being used.

The data-driven enterprise does not wait for the business question to develop and then use the data to answer it. The data-driven organization uses Advanced Analytics and Business Intelligence to mine the data for the questions and then the answers. With Solix CDP, business users can create data models and derive the insights needed to move the organization forward. The self-service model takes IT out of the equation, freeing it to focus on its work, while ensuring security and governance measures are also met. The Solix CDP brings robust ILM to all data.

The Solix CDP ensures all data retains context by retaining its metadata, meaning its value is never lost. This ensures the business questions being raised by the data are truly valid and the answers analysts find are relevant.

The Solix CDP ensures all data retains context by retaining its metadata, meaning its value is never lost.

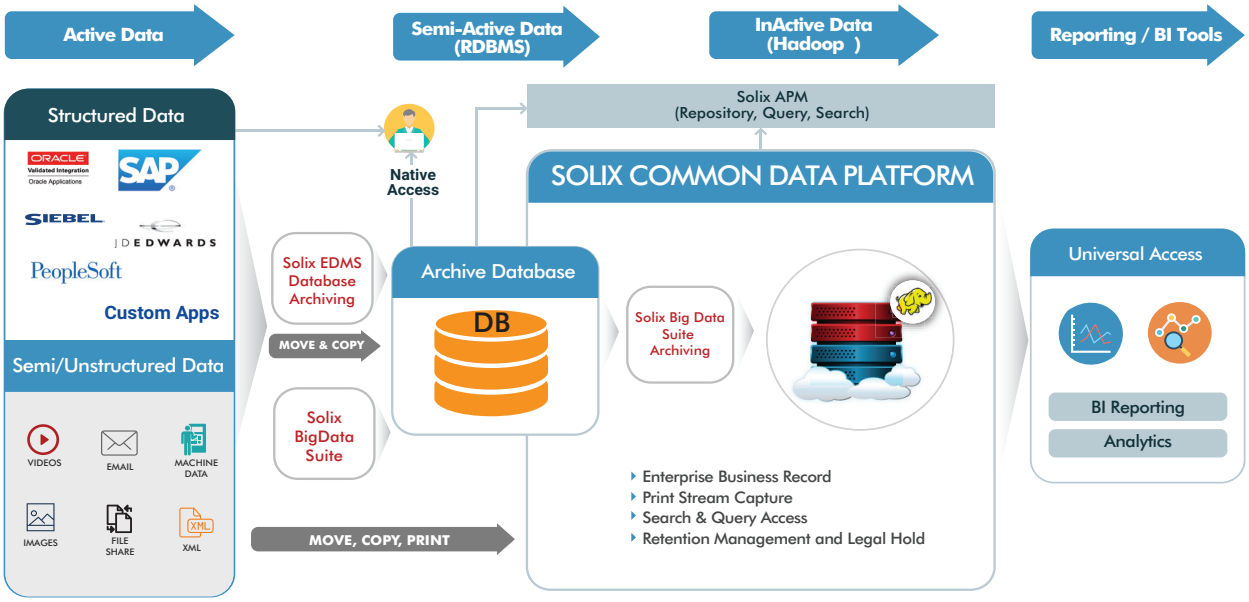
Enterprise Archiving

> In the era of Big Data, Archiving is a No-Brainer Investment.[3]

Up to 80 percent of production data used by core applications is inactive. Data archiving has emerged as an ILM best practice to meet data growth challenges. Solix CDP ensures that Enterprise Archiving improves production application performance, reduces infrastructure costs and meets the regulatory and compliance needs.

As part of Enterprise Archiving, application data running online is first moved into Tier 2 or Hadoop infrastructure, and then purged from its source location, according to ILM policies. Data archiving best practice requires that MOVE and PURGE processes be coordinated and validated. Enterprise Archiving on Solix CDP ensures proper data governance since enterprise data is ingested and stored based on ILM retention policies and business rules.

Archive data is classified for security and compliance requirements, such as legal hold, and universal access is provided for business users through structured reports and full text search for business objects.



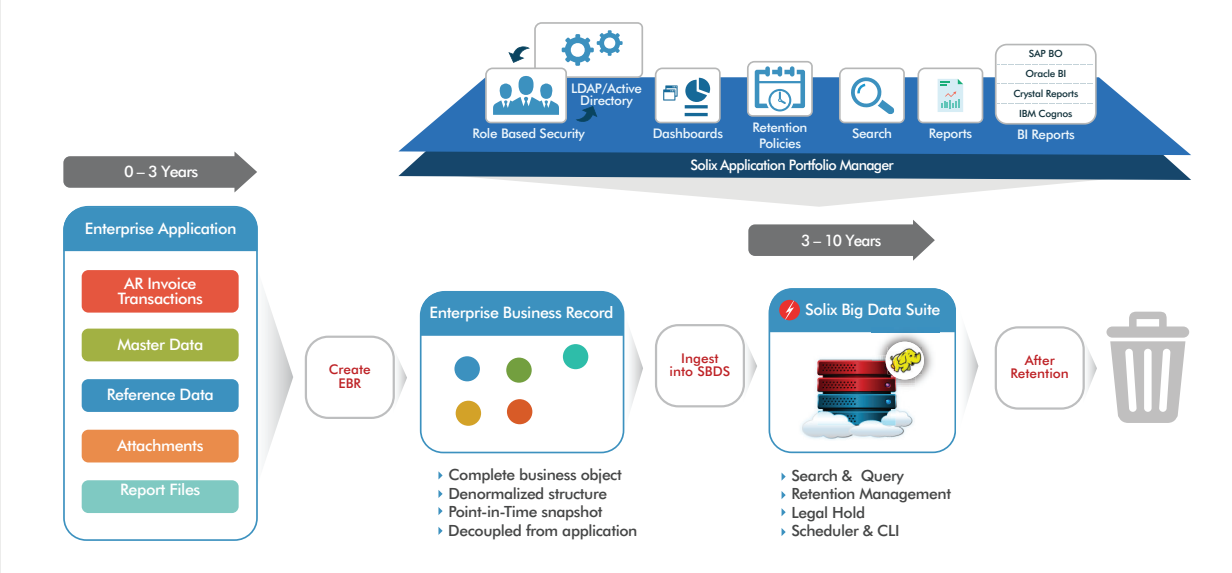
Solix CDP ensures that Enterprise Archiving improves production application performance, reduces infrastructure costs and meets the regulatory and compliance needs.

[3] Forrester Report on Vendor Landscape: Big Data Archiving, August 2015

Enterprise Business Record (EBR)

An EBR is a de-normalized, point-in-time snapshot of a business transaction, which may include structured or unstructured elements. The Solix CDP helps model, ingest and manage EBR data into a Hadoop optimized file format that is fully accessible for text search or structured query.

Data can be ingested to build both a long-term Enterprise Archive and a transient Enterprise Data Lake. For the archiving use case, older inactive data is moved from the source application to the Solix CDP. For the Data Lake use case, current data can be transformed and then copied from the source application to the Solix CDP.



EDW Augmentation

Currently enterprises are struggling to maintain costs associated with both storage and processing capabilities around traditional EDW implementations. Offloading storage as well as costly ETL functions to a commodity hardware such as Hadoop enables enterprises to focus on utilizing the existing Data Warehouse infrastructure to its best ability in doing BI and Advanced Analytics.

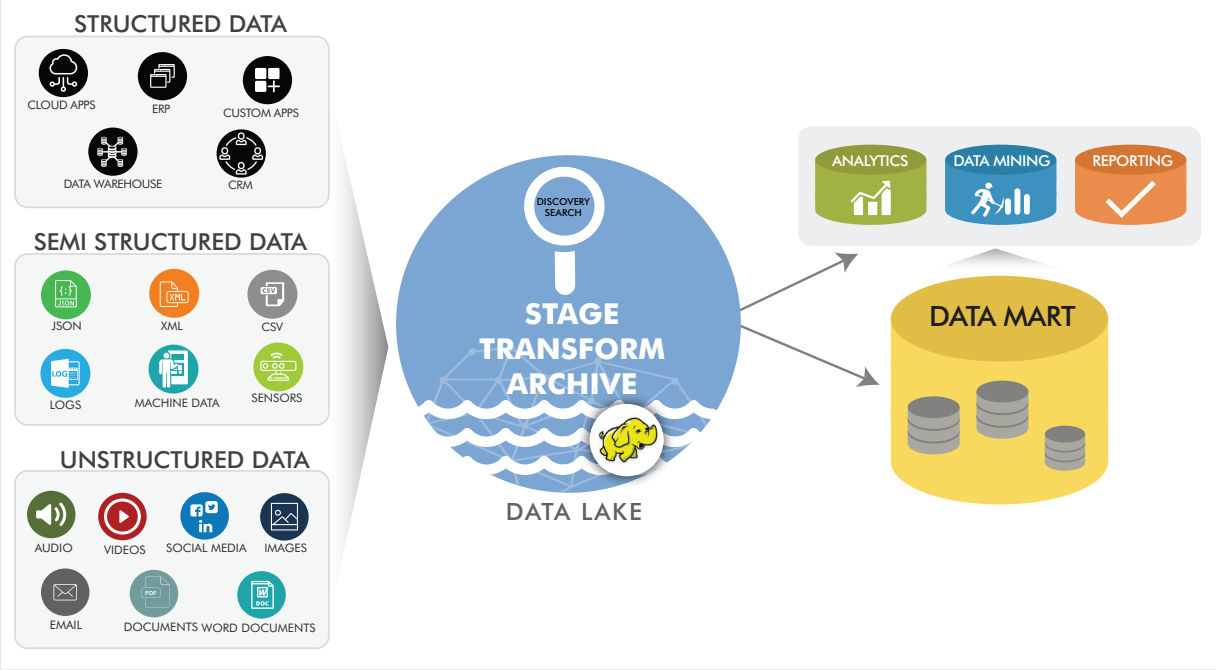
Migrating warm or cold data from the EDW via archive onto low cost bulk storage system such as Hadoop enables organizations to save millions on storage costs and significantly speeds up the processing power to get more value from the data warehouse by extracting valuable insights at a quicker pace from the collected data.

Currently enterprises are struggling to maintain costs associated with both storage and processing capabilities around traditional EDW implementations.

Enterprise Data Lake and Advanced Analytics

The Solix CDP based on Apache Hadoop establishes new capabilities for Advanced Analytics applications. It stores data "as-is" eliminating the need for demanding ETL processes during ingestion. It captures and maintains the metadata connected to each byte of data, which is half or more of the value of the data itself. The Enterprise Data Lake may then be mined for critical business insights using text search, structured query or further processing by downstream analytical applications. The Solix CDP utilizes either Hive or Spark query frameworks dependent on the user requirements.

The Solix Enterprise Data Lake reduces the complexity and processing burden of staging EDW and analytics applications and provides highly efficient, bulk storage of enterprise data for later use. Once resident within HDFS, enterprise data may be more easily distilled and better described at petabyte-scale by business analytics applications. This allows organizations to develop an enterprise architectural strategy that is responsive to the business stakeholders without driving up the investment in hardware and software.



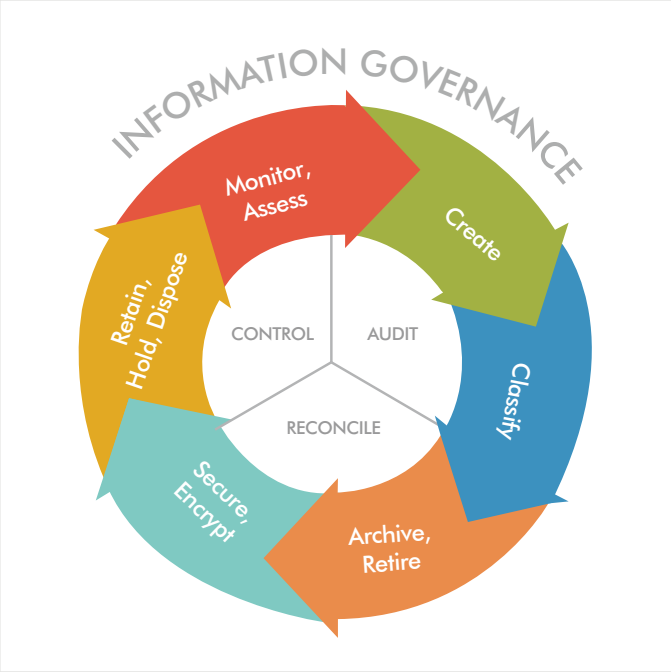
Information Governance

Analysts have warned that applying existing information governance practices to Big Data will result in failure. Comprehensive information governance provided by Solix CDP establishes the control framework necessary for proper data access control, data assessment, data discovery, data classification, data validation, retention management, legal hold and privilege management.

> Forrester estimates that the average Hadoop repository doubles in size every year; some implementations double in volume every month. More Hadoop silos are creating data challenges around security, integration, governance and delivery.[4]

To achieve robust ILM, new security and governance measures must be put into place to match the variety and complexity of the new data assets. The Solix CDP provides a true ILM continuum that addresses the complexity of governance in the Big Data world, while ensuring governance for core enterprise applications is not sacrificed. The Solix ILM framework manages the data within HDFS and provides an integrated retention-management and legal-hold capabilities.

Structured and unstructured data from various data sources are migrated into HDFS with full data-validation and audit reports. These reports provide the necessary defensibility and chain of custody for compliance and data governance. ILM policies and business rules may be pre-configured to meet industry standard compliance objectives, such as COBIT, or custom designed to meet more specific requirements.



Additionally, ILM also helps to solve the data growth problem by moving less frequently accessed data from high-cost Tier 1 infrastructure to Hadoop, leveraging cheap commodity infrastructure. Relocating inactive data to low-cost bulk data storage creates enormous infrastructure cost savings. Because governance, risk and compliance concerns grow by the terabyte, the Solix CDP ensures ILM for data throughout its lifecycle.

[4] Forrester Report on Big Data Fabric Drives Innovation and Growth, March 2016

COMPONENTS OF THE SOLIX COMMON DATA PLATFORM

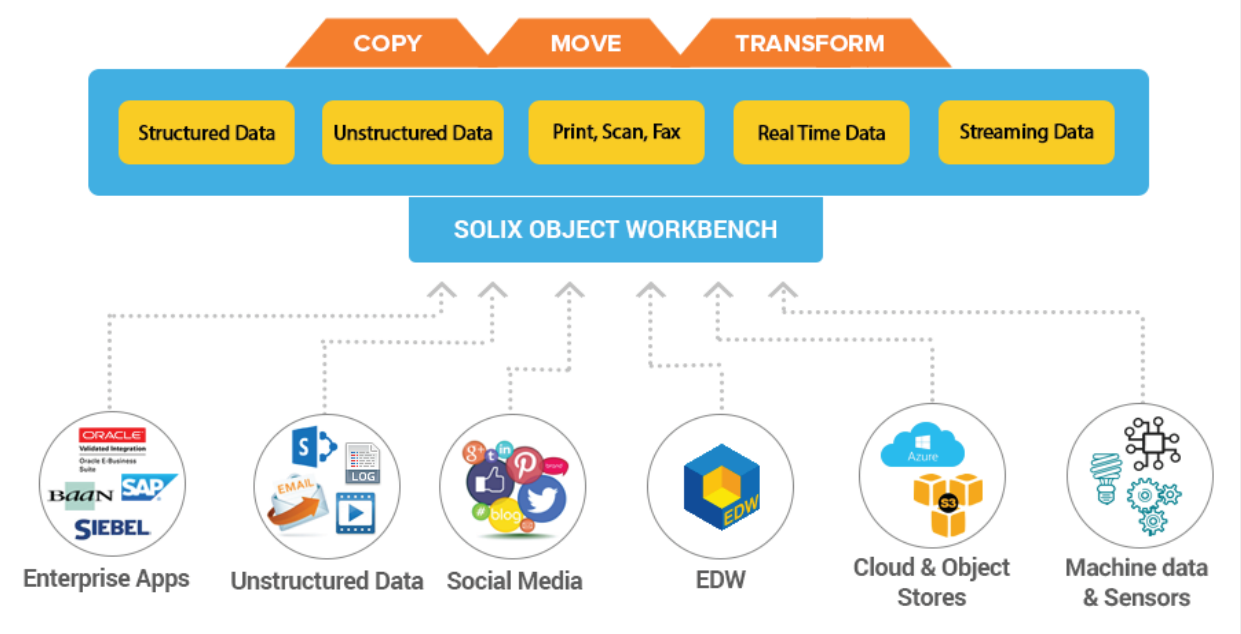
Solix Object Workbench

Integrated Connectors

Solix Object Workbench provides integrated connectors that can extract and ingest vast amounts of data "as-is" from an extensive set of enterprise data sources, including structured, semi-structured, unstructured and streaming data sources. The Object Workbench provides functionality to copy, move, and transform data from various data sources into the Solix CDP.

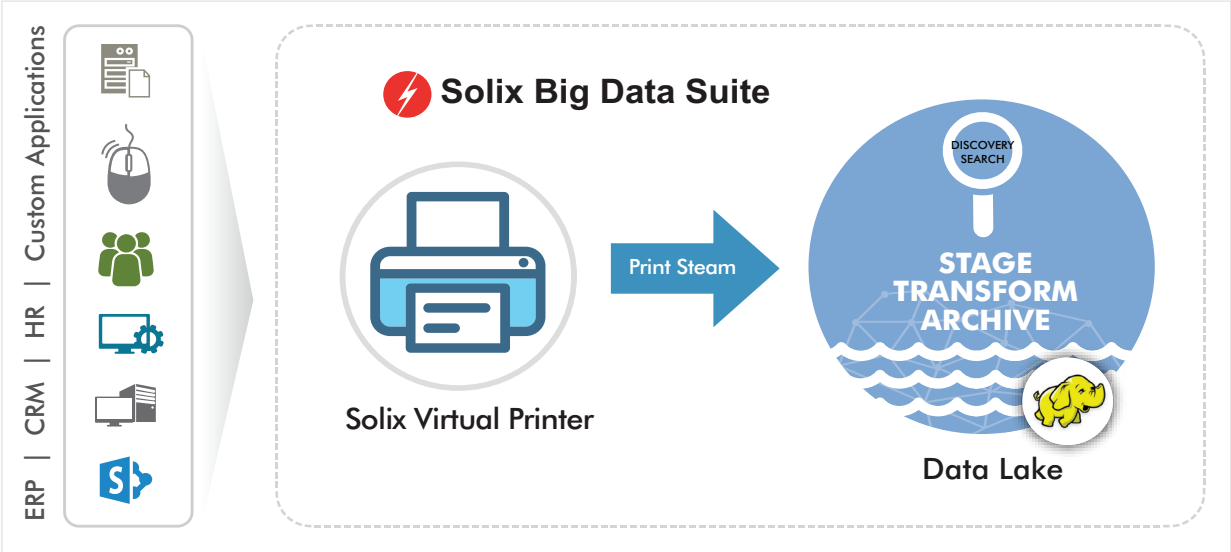
Extract, Transform and Load (ETL)

The Solix CDP Object Workbench also enables the ETL process to be undertaken as data is moved into the Enterprise Data Lake. This provides the ability to transform complex application data into meaningful data in a ready-to-use format from which the business user can gain immediate insight, with the use of BI tools.



Solix Virtual Printer

The Solix Virtual Printer provides functionality to capture print stream output from any application, transform it into a PDF document, automatically ingest it into Hadoop, index it and make it available for search access with full role-based security.



The virtual printer can be used to supplement an archiving project by capturing key report output — including all formatting — from the source application and storing it alongside the structured data.

The virtual printer can also be used to support a streamlined archiving approach called "print-and-purge." Using this approach, key documents, such as invoices or customer documents, are first "printed" by the Solix Virtual printer and ingested into Hadoop, after which the underlying data from the source application can be purged.

Real-Time and Streaming Data

> "Business users want data that's integrated in real time from multiple sources, including legacy data, social media, sensor data and weblogs, so they can make better decisions and increase their company's competitiveness.[5]

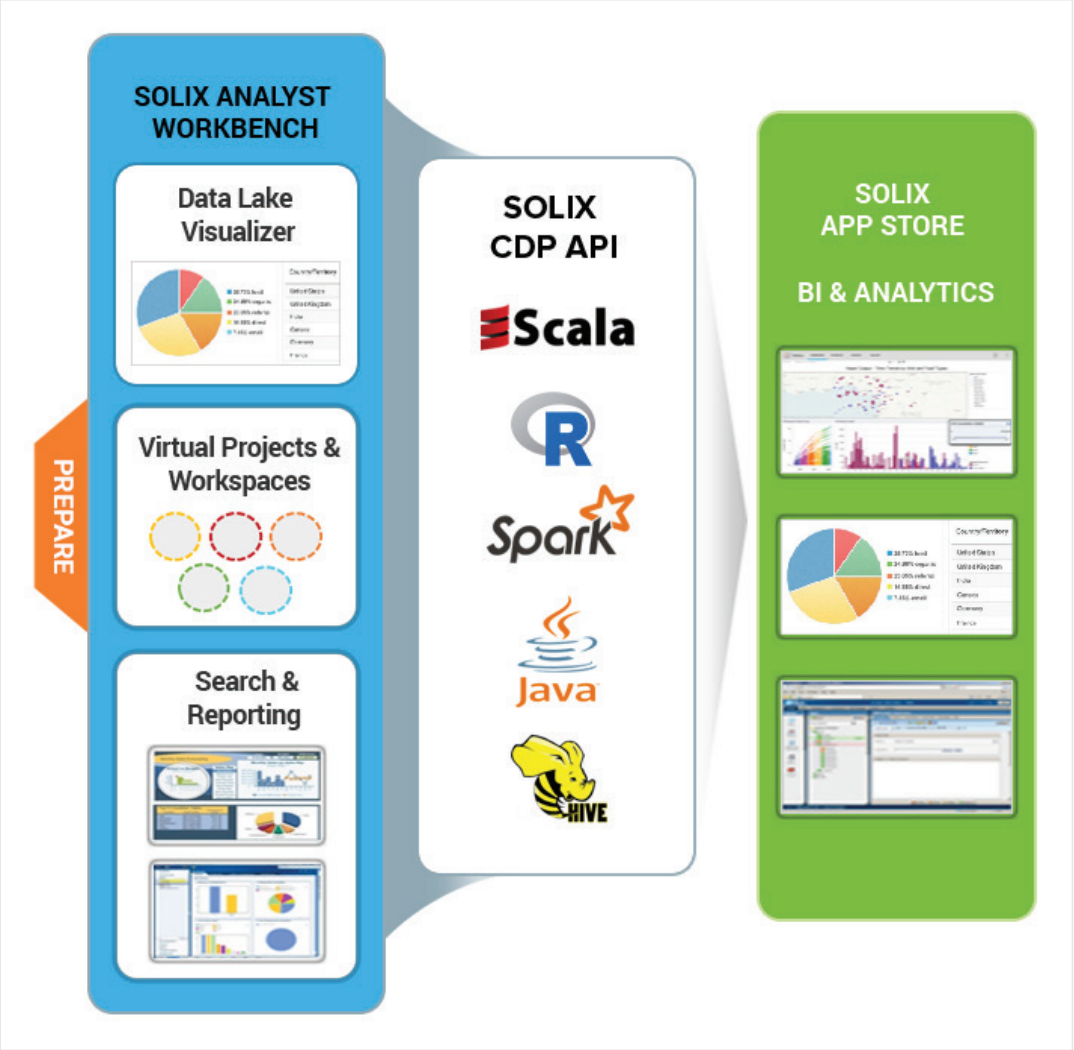
Insights from enterprise data is now not restricted to formulated data repositories, which only contain data at the end of its operational life. Huge amounts of data are now collected from both internet enabled devices and also from terminals in real-time and streaming formats.

This data can also be captured via the Solix CDP Object Workbench, enabling teams to create views of data to analyze and deliver actionable intelligence. This will enable enterprises across industries to become truly data-driven.

[5] Forrester Report on Big Data Fabric Drives Innovation and Growth, March 2016

Solix Analyst Workbench

The Analyst Workbench is designed for business analysts, data scientists, and DBAs to securely access the data within the Solix CDP and build virtual workspaces to manage analytics projects. All data within the platform is automatically made searchable and reportable in a secure and governed manner.



Functionality included with the analyst workbench includes:

Data Lake Visualizer

The Data Lake Visualizer is a graphical inventory of the data contained in the lake. Using the visualizer the data analyst can quickly find the data sets needed to complete their analytics assignment. Once the data sets are identified they can be selected for inclusion in the analytics project.

Virtual Projects and Workspaces

For each analytics assignment a virtual project can be created by the analyst. Within each project one or more virtual workspaces can be created. The objects identified in the visualizer can then be virtually copied into the workspace, eliminating the need to make physical copies of data.

Once the virtual workspace has been created, the data analyst can do data mashups by creating new composite objects to support the analytics assignment.

Data Preparation

Data preparation is the foundation for becoming a data-driven organization. To properly use data it must not only be collected from its ever-increasing variety of sources, it must also be put into a repository where those varied forms can be used by the analysts.

> *As more organizations utilize Data Scientists and Advanced Business Analysts to wrangle their data to enable digital transformation, the lack of proper tools hinders this progression. In fact, research shows that Data Scientists spend 80 percent of their time just cleaning data.*

Every analytics project requires the proper preparation of the data set. The nature of the data — from semi-structured data (such as log files), unstructured data (such as social, IoT) and structured data (such as relational databases) – must be understood, organized and transformed quickly and efficiently. The Solix CDP offers powerful, easy to use self-serve data preparation capabilities, including the ability to parse, clean, join and enrich data, as well as populate missing information and calculate new metrics.

The Solix CDP utilizes the Spark framework. Spark runs in-memory within the cluster and provides machine learning capabilities for faster and more advanced data preparation.

Search and Reporting Functionality

Solix CDP supports universal access to all enterprise data on a petabyte scale via text search, structured query or further processing by downstream analytical applications. End users gain improved data-driven results because their data is better able to be described.

Information Governance

The Solix CDP provides the ability to govern all of the data within the Hadoop repository for compliance and security. For example, automatically purge data based on a time-horizon, apply legal holds on files and transactions, enforce Kerberos/LDAP authorization for user access and more. This level of security and governance is able to be maintained because the CDP ensures all data retains its metadata.

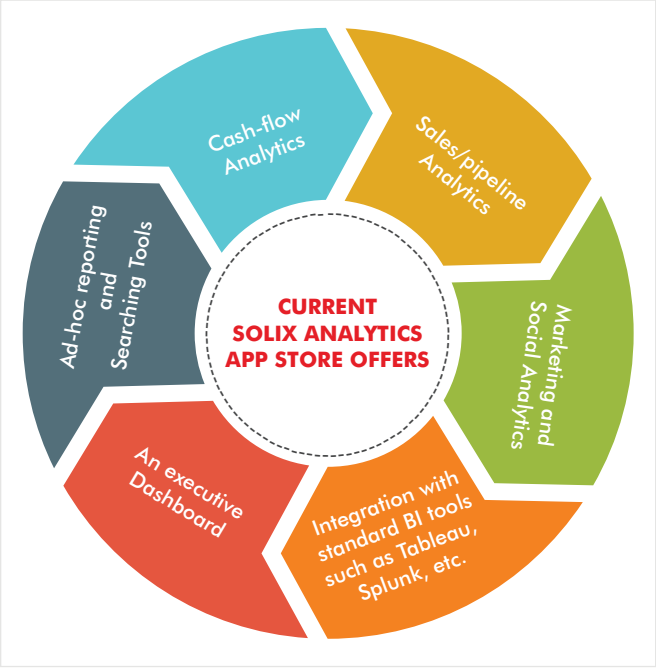
Information governance establishes the control framework necessary for proper data access control, data assessment, data discovery, data classification, data validation, retention management, legal hold and privilege management.

Solix Common Data Platform API

To enable development of custom applications and integration with existing BI and Advanced Analytics tools, the platform provides extensive APIs to access the unified repository. The API allows users to seamlessly access data from the Data Lake to enable the data-driven enterprise.

Solix App Store

The Solix App Store makes inductive BI user-friendly. The App Store offers out-of-the-box analytics through pre-integrated applications and also offers the opportunity to utilize third-party apps.



DEPLOYMENT MODELS

The Solix CDP is enabled for a number of deployment models including bare metal infrastructure, data center deployment, cloud infrastructure and also hosted multi-tenant deployment.

Solix never delivers a one-size-fits-all solution. The Solix team has experts ready to address customer needs from IT, business use and financial perspectives. The Solix team will work to understand organizational needs and then implement the best solution.

SPARK ON SOLIX CDP

The Solix CDP utilizes the Spark programing models. Spark runs in-memory within the cluster and does not depend on the two-stage Hadoop MapReduce paradigm. Therefore, repeat access to data is much faster. Spark relies on HDFS and runs on Hadoop YARN to be able to analyze the data stream.

Solix is committed to adding support for new Big Data tools as they appear in the fast-evolving Big Data ecosystem. This future-proofs Solix CDP installations, an important commitment given the speed with which the open source Big Data stack is evolving. It also simplifies the creation and evolution of an enterprise's Big Data environments.

BENEFITS OF THE SOLIX CDP

The benefits of the Solix CDP include:

- Combining the advantages of Hadoop with the ability to preserve the full metadata.
- Providing advanced ILM capabilities, including the ability to copy data from the data warehouse and to archive older data.
- Supporting advanced data security, as well as third party analysis packages, including machine learning and cognitive computing analysis of the data.
- Preserving all data in its original format and with full metadata and supporting established open standard interfaces. It future-proofs the Data Lake, ensuring the data will be usable by the new technologies and for new use cases that are as yet undefined.
- Providing a unified data governance layer from the time of data ingestion to use of data by business users for operational insights and Advanced Analytics.
- Ability to utilize either Hive or Spark query frameworks dependent on the user requirements.
- Cloud, on-premise and hybrid deployment models.
- Working with all Hadoop distributions such as Cloudera and Hortonworks.

The Solix CDP is the first solution to address all the data needs of an organization. From governance to analytics, the Solix CDP works with an organization's existing infrastructure to create a true ILM continuum that ensures the onslaught of data can be an asset and not a hindrance to business growth and development. The Solix CDP brings together the Enterprise Archiving, EDW and the Enterprise Data Lake while preserving metadata, allowing for schema on read, analytics opportunities, low cost implementation and maintenance as well as offering incredible scalability.

CONCLUSION

The era of game-changing digital disruption is here, and to thrive in this competitive environment, organizations need leadership that can effectively leverage all the data to derive actionable insights to fuel growth.

Reducing infrastructure costs, attaining operational efficiencies and deriving insights from BI and Advanced Analytics is the desire of many organizations. Solix CDP maximizes the insights that can be achieved, while reducing risk, ensuring compliance and governance to create a true ILM framework to lead organizations into the future. The Solix CDP gives organizations all the tools necessary to lower the total cost of ownership and satisfy the desire for return on investment.

Solix Technologies, Inc.

4701 Patrick Henry Dr., Bldg 20
Santa Clara, CA 95054

Toll Free:	+1.888.GO.SOLIX (+1.888.467.6549)
Telephone:	+1.408.654.6400
Fax:	+1.408.562.0048
URL:	http://www.solix.com