

A SOLIX WHITEPAPER

# SOLVING THE DATA GROWTH CRISIS: SOLIX BIG DATA SUITE

JULY 2014

**Vikram D. Gaitonde**

Vice President, Products

Solix Technologies, Inc.

<http://www.solix.com>

## Contents

Introduction — Data Growth Crisis.....	2
Operational Efficiency vs. Business Intelligence.....	3
Advantages of Apache Hadoop.....	3
Solix Big Data Suite — The New ILM Continuum.....	4
ILM Best-Practice.....	5
Solix Big Data Suite — Enterprise Archiving.....	6
Data Archiving.....	7
Application Retirement.....	7
Print and Purge Archiving and Decommissioning.....	8
Benefits of Enterprise Archive.....	8
Solix Big Data Suite — Data Lake.....	8
Solix App Store.....	10
Connector Framework.....	11
The Solix Difference.....	11
Biography.....	12

## Introduction — Data Growth Crisis

Today's Chief Information Officer operates in a perfect storm of data growth. Left unchecked data growth negatively impacts application performance, compliance goals and IT costs. Yet, this very same data is the lifeblood of today's organizations.

The storm has grown to crisis level with the increased need to manage unstructured data, such as documents, images, video, machine generated data and social files, along with traditional structured data.

According to a recent survey by Gartner, data growth is now the leading data center infrastructure challenge.<sup>1</sup> "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," said April Adams, research director at Gartner.<sup>2</sup>

Adding to the crisis is the increasing business-user demand for specialized analytics to mine enterprise data for better business results. Gartner analysts said "by 2016, 75% of structured data archiving applications will incorporate support for big data analytics."<sup>3</sup>

Business Intelligence is no longer a luxury. It is a necessity. Analytics are the foundation upon which organizations are making business decisions to not only survive in the new world-economy but thrive. Gartner believes BI will become an asset in itself. "By 2016, 30% of businesses will have begun directly or indirectly monetizing their information assets via bartering or selling them outright."<sup>4</sup>

## Operational Efficiency vs. Business Intelligence

CIOs are in a difficult position. The demand for operational efficiencies and the demand for Business Intelligence seem to be odds. Traditional thinking says the processes and infrastructure required to control costs, meet compliance objectives and keep costs in check do not mesh with the processes and structures needed for BI. Duplication of effort, hardware and operational costs could cripple IT.

The CIO must change the paradigm to ensure organizations can compete in today's economic climate as structured and unstructured data continue to explode. The time has come for a new enterprise blueprint. Apache Hadoop makes it possible. Solix Big Data Suite makes it a reality.

---

<sup>1</sup> "Big Data Benefits are Hampered by 'Culture Clash,'" Gartner, Sept. 12, 2013

<sup>2</sup> "Big Data Benefits are Hampered by 'Culture Clash,'" Gartner, Sept. 12, 2013

<sup>3</sup> "Big Data Benefits are Hampered by 'Culture Clash,'" Gartner, Sept. 12, 2013

<sup>4</sup> "Big Data Benefits are Hampered by 'Culture Clash,'" Gartner, Sept. 12, 2013

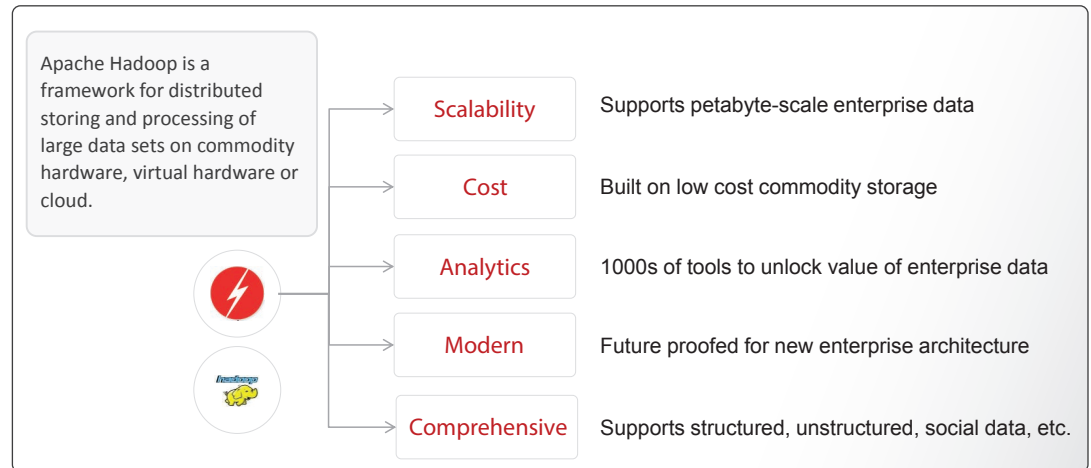


## Advantages of Apache Hadoop

Apache Hadoop is a framework for distributed storing and processing of large data sets on commodity hardware, virtual hardware or the cloud. Hadoop has been designed to provide analytics over petabyte-scale data sets, and the infrastructure can be leveraged for long-term storage and analytics.

The Hadoop Distributed File System (HDFS) has rapidly emerged as the leading storage platform because it provides secure, stable storage for structured and unstructured enterprise data with enhanced access. Moreover, Apache Hadoop represents the lowest cost alternative for highly scalable, bulk storage of enterprise data. Hadoop's MapReduce framework can process large data sets across distributed compute nodes in parallel, allowing "commodity" hardware to be used as the most efficient and cost-effective bulk data storage solution available.

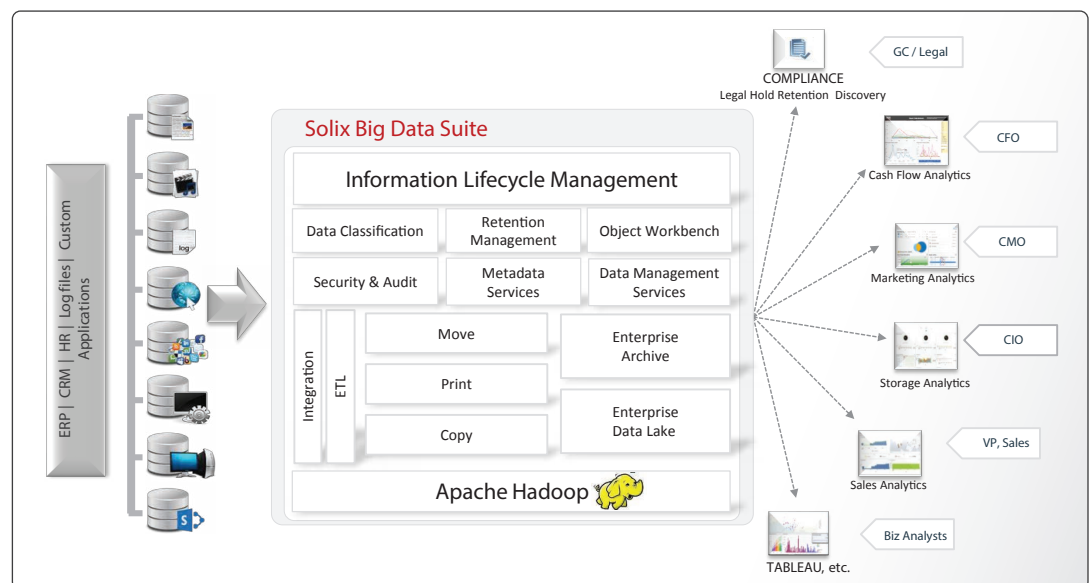
The Solix Big Data Suite harnesses the capabilities of Hadoop to create a comprehensive and efficient platform that creates unified ILM and BI infrastructures for all data, requiring smaller teams with fewer IT skills, while allowing quicker rollouts and faster results.



## Solix Big Data Suite — The New ILM Continuum

### Capture. Organize. Analyze.

Solix Big Data Suite offers an opportunity to create a new enterprise blueprint for the way data is handled. Solix Big Data suite provides the framework for an ILM continuum that ensures CIO's don't have to choose between operational efficiencies and BI. It's the only Big Data management platform that ensures a data lake does not turn into a data swamp.



## ILM Best-Practice

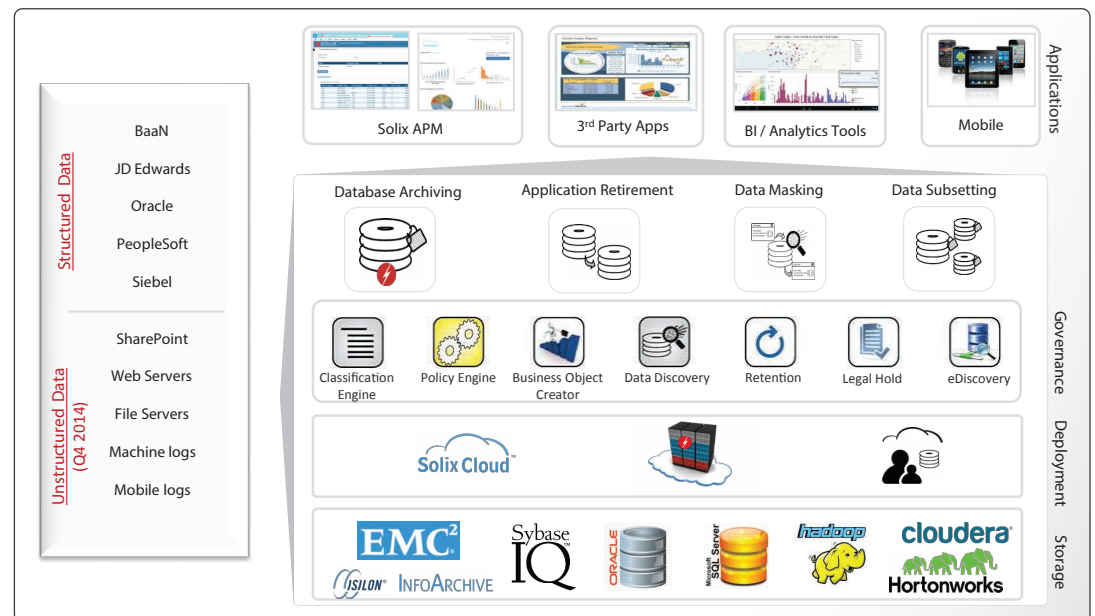
Gartner says “any organization thinking of simply applying existing information governance practices to big data will likely fail — not least because much data is ungoverned; or governed by others according to a different set of objectives.”<sup>5</sup>

Solix Big Data Suite provides the first true ILM continuum that addresses the complexity of governance in the Big Data world while ensuring governance for core enterprise applications is not sacrificed.

Solix Big Data Suite’s ILM framework manages the data within HDFS and HBASE. The Solix ILM framework also provides an integrated retention-management and legal-hold capability for data within Apache Hadoop.

Structured and unstructured data from other data sources are migrated into HDFS/H-BASE with full data-validation and audit reports. These reports provide the necessary defensibility and chain of custody for compliance and data governance.

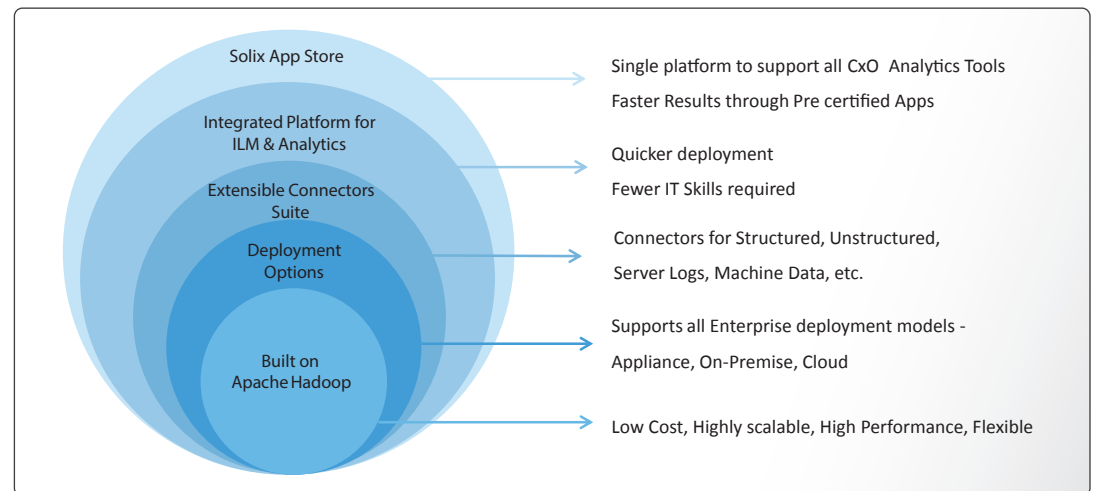
Solix’s ingestion engine captures application metadata and integrates with enterprise authentication systems, such as Active Directory, LDAP and others, to create role-based security for all data. Solix provides a single user interface to create a seamless user experience for administrators and end-users. Through role-based-security, the Solix interface allows users to access different facets of the product. For example, a compliance officer could be given access to create retention policies, apply legal-holds or generate reports on cases. End-users, however, might only be allowed to perform keyword-searches on projects assigned to them.



<sup>5</sup> “Big Data Benefits are Hampered by ‘Culture Clash,’” Gartner, Sept. 12, 2013

This extensive ILM framework allows Solix Big Data Suite to create a unified repository to **capture** all enterprise data and optimally **organize** it for **analytics tools** offered through the Solix App Store.

The **suite** is highly scalable with an extensible connector framework to ingest all the enterprise data. The integrated suite allows seamless archiving, retirement, and flexible extract transform load (ETL) capabilities to improve the speed of deployment, lower the cost, and optimize infrastructure. Solix also supports on-premise and cloud-based deployment, on a variety of Hadoop distributions.



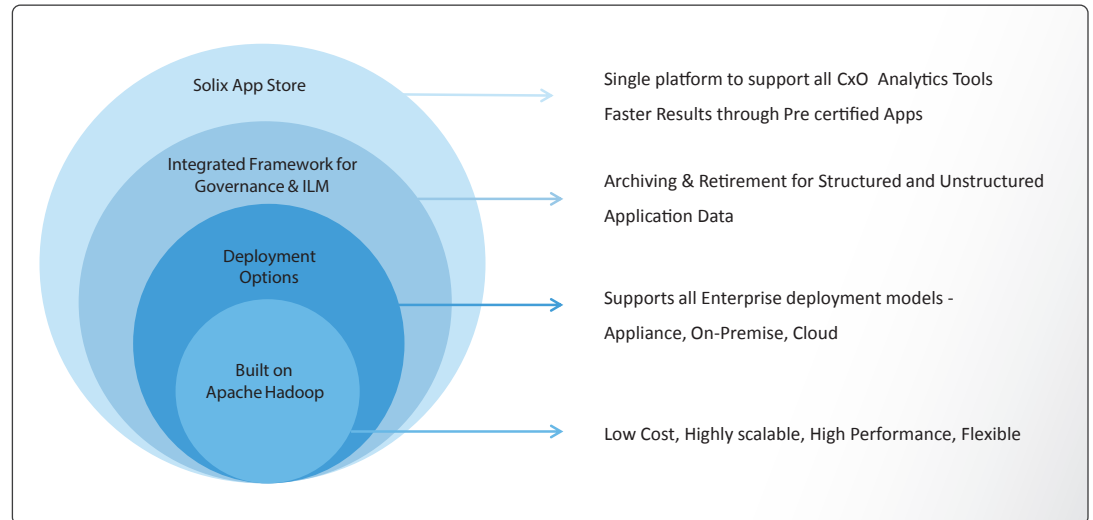
### The Solix Big Data Suite includes:

- Solix Enterprise Archiving to improve enterprise application performance and reduce infrastructure costs. Enterprise application data running online is first moved, and then purged from its source location according to ILM policies to ensure governance, risk and compliance objectives are met.
- Solix Enterprise Data Lake reduces the complexity and processing burden to stage enterprise data warehouse (EDW) and analytics applications, and provides highly efficient, bulk storage of enterprise data for later use when it is needed. Solix Data Lake provides a copy of production data and stores it “as is” in bulk for later use.
- Solix App Store offers pre-integrated analytics tools for data within the Enterprise Archiving and Enterprise Data Lake.

### Solix Big Data Suite — Enterprise Archiving

Solix Enterprise Archiving offers a unified ILM framework to archive and retire all enterprise data — structured and unstructured — to optimize application performance, reduce cost and achieve compliance goals. Solix’s ILM framework meets all the

necessary aspects of governance, including legal hold, retention management, eDiscovery, and auditing. Solix takes governance the next step by including validation of all archiving and retirement actions taken within the platform.



The Enterprise Archive has two components, Data Archiving and Application Retirement.

## Data Archiving

Data growth can drive infrastructure costs up while diminishing application performance and productivity. As much as 80 percent of data in production databases used by core applications is inactive. Database archiving has emerged as a key component to an ILM best-practice framework to meet data growth challenges.

Data archiving best-practice requires that move and purge processes be coordinated and validated. All data deletions are done in a defensible manner with audit records and chain of custody. Solix Enterprise Archiving ensures proper data governance since enterprise data is ingested and stored based on retention management policies with support for custom 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.

## Application Retirement

Experts estimate that up to 40 percent of applications are candidates for retirement, migration or rationalization. Retiring legacy applications creates instant cost savings by freeing staff for more important assignments, eliminating licensing and maintenance fees, and enabling IT to shut down unnecessary hardware saving power, floor space and air conditioning load.



Data — both structured and unstructured — is compressed by as much as 90 percent and stored at the lowest possible cost in an immutable format which is easily accessed by users. Yet, active data can be migrated to the production database and accessed by new core applications. Hundreds of legacy applications, in a variety of configurations and on different platforms, can be decommissioned.

## Print and Purge Archiving and Decommissioning

Solix Print Report Archiving offers the capability to generate printed reports from native applications, in readable formats such as PDF or Text, for archiving or application decommissioning. The tool allows import of previously created reports from folders. Users can access these reports for compliance, eDiscovery and other regulatory requirements through Solix's search and reporting interfaces.

## Benefits of Enterprise Archive

Together the features of Solix Big Data Suite Enterprise Archive ensure best-practice ILM strategies can be implemented for both structured and unstructured data. All move and purge processes can be automated based on organization policies and schedules. Only Solix Big Data Suite ensures all move and purge actions are validated to ensure the proper data has been processed.

User access to the data occurs through reporting and search tools utilizing metadata tags and keywords. However, archived data is classified for security and compliance requirements and access is role-based and predicated on clearance.

### **Solix Enterprise Archive will also benefit organizations through:**

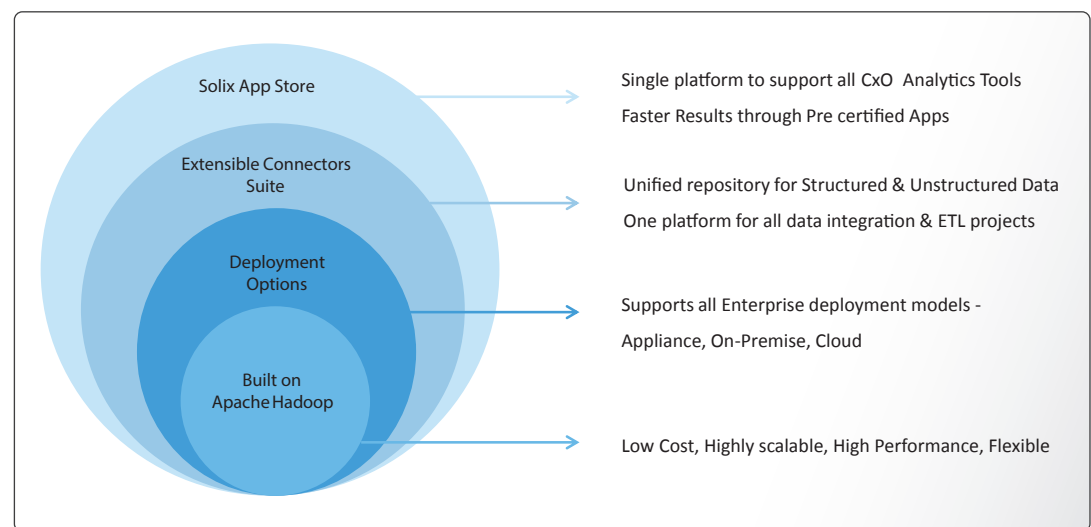
- Improved application performance by reducing the amount of production data core applications must process.
- Reduced application downtime. Smaller pools of active data means faster backups, upgrades and disaster recovery. Cloning time will also be reduced.
- Reduced operational costs including utilities and the ability to reallocate staff time to more mission-critical operations.
- Decommissioning legacy applications, which allows for the elimination of infrastructure, maintenance, and support costs

## Solix Big Data Suite — Enterprise Data Lake

Solix Enterprise Data Lake is a repository to integrate all formats of data from structured and unstructured sources. The Data Lake offers data integration and ETL capabilities on a low-cost, highly scalable platform.

Traditional EDW platforms deliver highly specific data views based on corporate strategy. This canonical top-down enterprise approach is deductive and has merits, yet it does not allow for the inductive analytical style that drives many BI proponents. The fact is, the volume, velocity and variety of data in organizations is simply too much for the human mind to grasp. Yet, Gartner says these two approaches are locked in a ‘culture clash’ that must be resolved.<sup>6</sup>

The Solix Enterprise Data Lake reduces the complexity and processing burden to stage EDW and analytics applications, and it provides highly efficient, bulk storage of enterprise data for later use when it is needed. Proponents of both deductive and inductive analytics can be satisfied.



With the Solix Enterprise Data Lake, the data is stored on “commodity” servers that are far less expensive than the Tier 1 infrastructure enterprise data traditionally utilizes. These commodity servers can save organizations more than 80 percent on infrastructure costs alone. Add in the reduced expense for licenses and utility costs to keep Tier 1 infrastructures running and the Data Lake pays for itself quickly.

The Solix Enterprise Data Lake provides a copy of production data and stores it “as is” in bulk to be better described and distilled later. This simple copy process eliminates the need for heavy ETL processing during ingestion. Once resident within the Hadoop file system (HDFS), enterprise data can be better described or transformed for use with business analytics applications based on user need, which allows organizations to get the most out of the data.

Utilizing the Solix Enterprise Data Lake, organizations can have access to larger pools of data to drive business decisions at far lower cost. Simple tools could leverage server and

<sup>6</sup>“Big Data Benefits are Hampered by ‘Culture Clash,’” Gartner, Sept. 12, 2013

machine logs to predict system failures, alerts and the like. However, organizations need to mine data at a far more strategic level. Sophisticated BI tools can bring significant value to positively impact the revenue and growth of the organization and have become key decision drivers.

The Solix Enterprise Data Lake also offers unprecedented access to an extensive ILM framework to ensure data meets governance, risk and compliance objectives and best practices for data retention and classification are deployed. Gartner analysts say “today's information governance practices focus on controlling various information quality attributes, typically for information governed behind the firewall. Organizations are either following the same approach with big data (which is proving impractical) or looking for new ways to govern that data to reduce risks.”<sup>7</sup>

Solix Big Data Suite creates the new standard for governance in a Big Data world. 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.

Solix Big Data Suite creates the new standard for governance in a Big Data world. 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.

**Solix Enterprise Data Lake’s extensive features include:**

- An easy-to-use wizard-based user interface
- Manual or scheduled mode to update data
- Role-based security
- Instant access to all information within the data lake

## 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 offers the opportunity to utilize third-party apps. Organizations can also build custom and mobile apps through the App Store.

**The App Store offers:**

- Cash-flow analytics
- Sales/pipeline analytics
- Marketing and social analytics
- Integration with standard BI tools such as Tableau, Splunk, etc.
- An executive dashboard
- Ad-hoc reporting and searching tools

## Connector Framework

Solix Big Data Suite has an extensive connector framework that allows the ingestion of data from all sources. The framework can take heterogenous data streams and allow the organization to manage the data according to business policies.

### The connector framework encompasses:

- **Structured Data:** Core enterprise applications including Oracle E-Business Suite, PeopleSoft, Baan, JD Edwards, Siebel and custom applications
- **Unstructured Data:** MicroSoft Sharepoint, file systems, web servers, desktops, laptops etc.
- **Social feeds:** Including facebook, twitter, Instagram
- **Log files:** Security Logs, Network Server Logs, Mobile Logs, etc.
- **Print stream:** Native Report Archiving, Fax/Scan Data, etc.

The various streams of data are brought together through the Solix Data Integration Studio, which has comprehensive ETL capabilities to ingest enterprise data into the Solix Data Lake.

## The Solix Difference

Solix Big Data Suite manages the entire lifecycle of data from its collection to purge without sacrificing governance, compliance or performance or forcing the CIO to choose between operational efficiency or BI. Solix Data Suite is built to address the needs of each individual organization and harness the perfect storm of data growth into a cost-effect driver for business innovation.

A new enterprise blueprint is within grasp for every CIO. The Big Data Suite has a variety of deployment options, including physical or virtual appliances either on-premise or in the cloud.

With a small initial investment organizations can harness a framework to create a ILM continuum ensuring costs remain in check, core applications perform optimally and compliance objective are met while building BI capabilities to drive the organizations future in the new world economy.

Solix has designed the Big Data Suite to have flexible deployment and distribution options. Several Apache Hadoop distributions, such as Cloudera, Hortonworks and others are supported and deployment methodologies including hardware, virtual and hybrid-cloud are available to fit into all enterprises.

Solix Big Data Suite is also available as an all-inclusive appliance eliminating the hassle of installation and configuration. Lower IT cost by using inexpensive commodity hardware to scale to petabytes of data.

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.

Solix Technologies, Inc., the leading provider of Enterprise Data Management (EDM) solutions, is transforming information management with the first enterprise archiving and data lake application suite for big data: The Solix Big Data Suite. Solix is helping organizations learn more from their data with enterprise analytics and achieve Information Lifecycle Management (ILM) goals. The Solix Enterprise Data Management Suite (Solix EDMS) and Solix Enterprise Standard Edition (SE) enable organizations to improve application performance, meet compliance objectives and reduce the cost of data management across the enterprise. Solix Technologies, Inc. is headquartered in Santa Clara, California and operates worldwide through an established network of value added resellers (VARs) and systems integrators.

## Biography

Vikram Gaitonde heads the technology leadership and product strategy for Solix as Vice President of Products. Vikram is responsible for product management and engineering for Solix's next generation information management solutions.

Mr. Vikram brings 15 years of product management, engineering, and software delivery experience and has held several leadership positions within technology companies. Prior to Solix he was the Director of Enterprise Solutions at EMC, where he led the product solutions and product management teams for the information intelligence, data governance, eDiscovery, and content management products. Additionally, he has held numerous technical and product leadership positions at Silicon Valley startups and industry giants including Hotmail/Microsoft, Wipro, Kazeon (acquired by EMC), and Cast Iron Systems (acquired by IBM).

Mr. Vikram holds an MBA from The Wharton School of the University of Pennsylvania and a B.S. in Computer Science from the University of Pune, India.



**Solix Technologies, Inc.**

4701 Patrick Henry Dr.,  
Building 20  
Santa Clara, CA 95054  
Phone: 1.888.GO.SOLIX (1.888.467.6549)  
1.408.654.6400  
Fax: 1.408.562.0048  
URL: <http://www.solix.com>

Copyright ©2014, Solix Technologies and/or its affiliates. All rights reserved.

This document is provided for information purposes only and the contents hereof are subject to change without notice.

This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchant- ability or fitness for a particular purpose.

We specially disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Solix is a registered trademark of Solix Technologies and/or its affiliates. Other names may be trademarks of their respective owners.