

Legacy Application Retirement Guide

A comprehensive overview for SAP and non-SAP environments

ABSTRACT

This white paper examines technology solutions and approaches that your enterprise can use to retain and access business-critical data when migrating legacy applications and data to new systems. The paper explains how you can accelerate the retirement of your legacy applications while maintaining compliance with your corporate data retention standards.

November 2017

TABLE OF CONTENTS

INTRODUCTION.	1
CONSIDERATIONS FOR A LEGACY APPLICATION RETIREMENT SOLUTION	1
Scalable architecture	1
Cost-effective architecture	2
Ability to extract and store data from a variety of systems	2
Enterprise-level security	2
Ease of use and flexible reporting	2
Data governance and retention compliance	2
STRATEGIES FOR LEGACY APPLICATION RETIREMENT	3
LEGACY APPLICATION RETIREMENT BY CAPTURING DATA IN A LOW-COST DATABASE.	4
Extract, transform and load	5
Security	5
Data governance and retention compliance	5
LEGACY APPLICATION RETIREMENT PROCESS	5
Discovery Phase	5
Design Phase	5
Build Phase	5
Go-live Phase	5
INFRASTRUCTURE DEPLOYMENT OPTIONS	6
Extract	6
Transform	7
Store and report	7
On-premises options	8
Hybrid and cloud options	8
ACCELERATING YOUR JOURNEY WITH DELL EMC AND AURITAS.	8
ABOUT THE AUTHORS	9

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2017 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Intel, Xeon and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries. Other trademarks may be the property of their respective owners. Published in the USA November 2017, White Paper.

Dell EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

INTRODUCTION

Imagine you're the CIO of a large multinational company. You and your forward-thinking team have implemented a shiny new ERP application that replaces many of your older applications and offers great benefits in terms of efficiency and cost savings. Suddenly, a new problem arises before the cleaning crew even has a chance to clear out the remnants of last night's celebration party. What do you do with all those old applications that have been replaced?

Though you migrated your master data and open transactions to the new applications, you still have historical data in the old applications. These old, or legacy, applications contain data that your legal department insists on keeping available for audits. So, should you keep the legacy applications running?

As you grapple with this problem, your company makes a few acquisitions and informs you that the new ERP system must be rolled out to them as well. You now have numerous legacy applications that need managing. Before you know it, maintenance costs have ballooned, data governance has weakened, and your IT team must support a variety of applications that no one really knows much about.

Your challenge is that once the master data and open transactions have been migrated to newer applications, the value of the data remaining in your older applications doesn't justify the costs of maintaining them. What then, are your options? How do you comply with legal requirements and demands for historical data without incurring the huge licensing, maintenance and personnel costs?

In such cases, instead of pulling their hair out, smart managers like you pull the plug on these old applications by migrating essential data to a cost-effective repository with reporting capabilities. This paper will help you understand how to design and build cost-effective solutions to overcome your legacy applications retirement problem.

CONSIDERATIONS FOR A LEGACY APPLICATION RETIREMENT SOLUTION

Once you've diagnosed the need to retain legacy data for future retrieval, it's time to think about the ROI of available options. In many organizations, the retrieval of legacy data is purely necessary for legal purposes. Practicality is key in selecting a retirement solution for retaining data. Why pay premium costs for access to data with little operational value? In most cases, if the cost of the application retirement solution is less than that of maintaining a legacy application, then a retirement solution wins out.

When the value of retaining legacy data is low, organizations pursue cost-effective solutions that provide the following features:

SCALABLE ARCHITECTURE

Whether an organization chooses to host its legacy data on premises or in the cloud, the architecture should be highly scalable. Data growth and accessibility are the primary factors in scalable architecture. If more legacy applications are to be retired, the implemented infrastructure solution should scale to meet future requirements.

COST-EFFECTIVE ARCHITECTURE

One of the main drivers for legacy application retirement is their high cost of maintenance. Also, the data from legacy applications is not perceived as very high in value as it's mostly retained for legal and compliance reasons. While the cost of non-compliance and not adhering to legal requirements can be high, the value for the day-to-day operations of the organizations is minimal. This suggests that the replacement should have a low cost of maintenance.

ABILITY TO EXTRACT AND STORE DATA FROM A VARIETY OF SYSTEMS

Legacy systems come in a variety of shapes and sizes. Some use databases and open relational tables for their back ends, while others use flat files and proprietary formats to store their data. The proposed legacy application retirement solution should be capable of extracting data from a variety of sources, such as:

- Relational database management systems (RDBMSs)
- Flat files
- Applications (SAP, Oracle E Business Suite, PeopleSoft, etc.)
- REST APIs
- JSON streams

The system should allow data to load seamlessly. Extract, transform and load (ETL) features, such as auto-detection of source schemas and automatic transformation to the target schema, make the ETL process simple and time-efficient. Such features are essential in a legacy application retirement solution.

ENTERPRISE-LEVEL SECURITY

Typically, data that is available in the legacy system is controlled by the application's security features. Because these restrictions still apply when the data is moved into a new system, the legacy application retirement solution should be security enabled to integrate with enterprise security systems. In cases where the legacy application retirement solution will be actively used by many users, it should support interfacing with Active Directory (AD) and Lightweight Directory Access Portal (LDAP), and support security features such as single sign-on.

EASE OF USE AND FLEXIBLE REPORTING

Legacy data will frequently be retrieved for legal requirements and analytic purposes. To support these tasks, the system should provide strong visualization and reporting capabilities. Developers and end users should have access to tools that allow them to quickly create custom reports.

DATA GOVERNANCE AND RETENTION COMPLIANCE

Like security, data governance applies to all data retained by the organization. While organizations are legally forced to retain data for certain periods of time, they are also legally forced to remove and delete data that is older than the specified retention period. Therefore, the legacy retirement solution should offer robust lineage tracking, monitoring and retention compliance. To support auditing, the system should allow easy tracking and lineage reporting for the data stored in the legacy system. To enable retention compliance, the system should incorporate automatic rules to delete data subject to retention requirements and integrity rules.

STRATEGIES FOR LEGACY APPLICATION RETIREMENT

Given the above considerations, there are many strategies that organizations can adopt depending on the individual organization's needs. The table below lists some of the strategies and their pros and cons:

Strategy	Pros	Cons
Migrate data from legacy environments into active systems	<ul style="list-style-type: none"> • Single solution for accessing all information • Having one platform for managing compliance, security, etc. 	<ul style="list-style-type: none"> • Expensive • ROI of retaining legacy data does not justify expense
Capture data in magnetic formats and shut down servers	<ul style="list-style-type: none"> • Savings on hardware and software maintenance • Reduction in data maintenance requirements 	<ul style="list-style-type: none"> • Potential issues with hardware compatibility • Longer lead time for information access • Inflexible
Capture data in a low-cost database	<ul style="list-style-type: none"> • Easier maintenance of database • Reporting on data is fast and easy • Using database queries for reporting or ODBC connection for external reporting 	<ul style="list-style-type: none"> • Migrating data into a database is a difficult process • Effort required to build relationships to tables in new database • Effort required to create new reports • Potential issues with data integrity and format
Capture all the information in a report format	<ul style="list-style-type: none"> • Static report output • Details accessible at a short notice • Very little or no management needs 	<ul style="list-style-type: none"> • Capturing reports does not guarantee that all historical data requirements will be met • Issues with reporting performance and display format • Little flexibility on drill-down, mining and analysis
Capture data in a business warehouse solution	<ul style="list-style-type: none"> • Dynamic report outputs • Details of data are quickly accessible 	<ul style="list-style-type: none"> • Capturing reports does not guarantee that all historical data requirements will be met • User training required and purchase of software and hardware needed • Cost of migration and maintenance could be high
Capture data using tools like DART	<ul style="list-style-type: none"> • Minimum user training required • Single application for accessing legacy and active data • Data is still mutually exclusive 	<ul style="list-style-type: none"> • Only used for retiring SAP landscape • Purchase of ILM software and hardware required • Understanding of fields and tables knowledge is required for migration and reporting
Migrate data into third-party tools (Informatica, OpenText, HPE, IBM, Business Objects etc.)	<ul style="list-style-type: none"> • Built-in extraction process • Single application for retiring multiple legacy environments • Data is still mutually exclusive • Retention management and data tagging possible 	<ul style="list-style-type: none"> • Software and hardware investment • Knowledge transfer and training • Proprietary solution

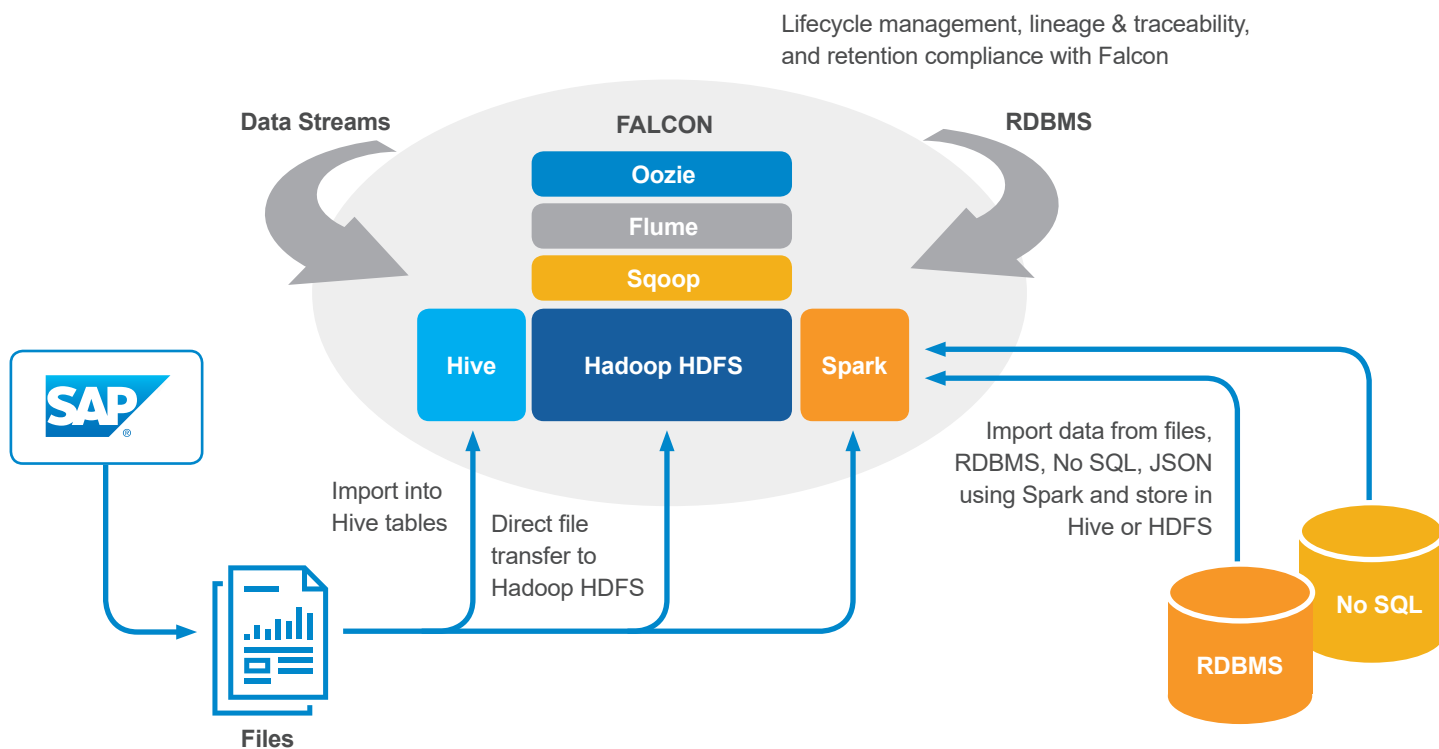
LEGACY APPLICATION RETIREMENT BY CAPTURING DATA IN A LOW-COST DATABASE

Capturing legacy data in a low-cost database is a proven, cost-effective and flexible solution adopted by various organizations. In this section, we highlight how Apache™ Hadoop® can be used as a low-cost database to capture legacy data.

Apache Hadoop is an open-source, distributed computing platform that supports scalable data storage, processing and analytics. One of the key features of Hadoop is that, to achieve scalability, the system is designed to handle failure of its nodes. The upside of this is that less expensive hardware can be used to build the system. This leads to a cost-effective and highly scalable system.

Since its early development stages, Hadoop's ecosystem of components has grown above and beyond primary components such as HDFS (Hadoop Distributed File System), MapReduce and YARN. The ecosystem now provides various components that help Hadoop deliver many of the legacy application retirement solution requirements discussed earlier. These features, in addition to low-cost and scalability benefits, make Hadoop an ideal solution for legacy application retirement.

The following sections provide details on how Hadoop proves to be a scalable, secure and cost-effective legacy application retirement solution.



EXTRACT, TRANSFORM AND LOAD

The Hadoop eco-system has numerous components that support extraction and data loading from multiple sources. For example, Hadoop's HDFS allows direct file transfer of text files to the HDFS. Hadoop also features additional tools to extract data from RDBMSs, NoSQL databases and JSON files.

SECURITY

To meet regulatory requirements for data protection, Hadoop provides secure authentication for shared clusters. Hadoop in conjunction with Kerberos, a mature open-source network authentication protocol, authenticates users and provides access to data and functions within Hadoop.

DATA GOVERNANCE AND RETENTION COMPLIANCE

While Hadoop in itself does not provide many governance features, through its components it provides an extensible set of core foundational governance services. These enable enterprises to effectively and efficiently meet their compliance requirements within Hadoop by allowing integration with the whole enterprise data ecosystem. Some of the features provided by the Hadoop ecosystem include data lifecycle management; defining, scheduling and monitoring data management policies; data lineage and traceability; auditing; and data classification.

The above features make Hadoop an exceptional legacy system retirement solution.

LEGACY APPLICATION RETIREMENT PROCESS

Process makes perfect! A strong process helps ensure the success of the legacy application retirement initiative. Below is an outline of a customizable process to meet the requirements of various organizations:

DISCOVERY

DESIGN

BUILD & TEST

GO-LIVE

DISCOVERY PHASE

1. Identify all stakeholders within and outside the organization. Stakeholders outside the organization would be regulatory bodies and similar entities.
2. Determine business value of applications and data in the legacy system.
3. Identify data to be retained.
4. Document system read-only and retirement dates.
5. Select a system retirement strategy.

DESIGN PHASE

1. Review infrastructure considerations and requirements.
2. Design the replacement landscape.
3. Size the replacement landscape.
4. Plan and design data migration and data archival.

BUILD & TEST PHASE

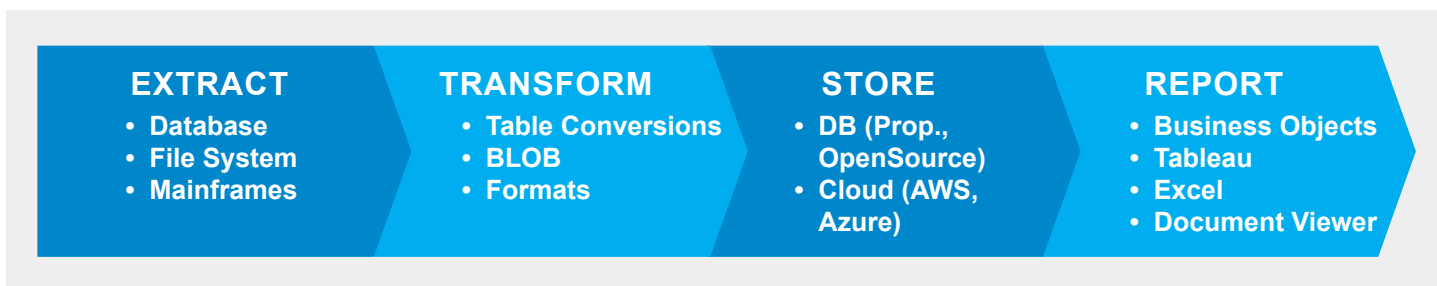
1. Install landscape.
2. Migrate data.
3. Archive data.
4. Validate data.
5. Test functionality.
6. Obtain sign-off.

GO-LIVE PHASE

1. Hand over to operations team.
2. Train staff.
3. Support the solution.

INFRASTRUCTURE DEPLOYMENT OPTIONS

Dell EMC infrastructure deployment options cover all of the key steps in the data management workflow for legacy system retirement — extract, transform, store and report.



EXTRACT

Dell EMC offers multiple options for extracting data from legacy databases, file systems and mainframe systems and moving it into lower-cost archival storage. These options include solutions that incorporate the Apache Hadoop platform for distributed data storage and management and the Dell EMC Isilon platform for dedicated network attached storage for large datasets.

DELL EMC READY BUNDLES FOR HADOOP

Dell EMC Ready Bundles for Hadoop are developed jointly with leading Hadoop distributions, and are based on extensive customer experience with real-world Hadoop production installations. These solutions include the hardware, software, resources and services needed to deploy and manage Hadoop in production environments and offer compelling total cost of ownership (TCO) benefits by using cost-optimized, industry-standard, Intel-based Dell EMC servers and storage to decrease the cost to store and process large data sets. These Ready Bundles are an ideal solution for storing small and midsize datasets from legacy environments.

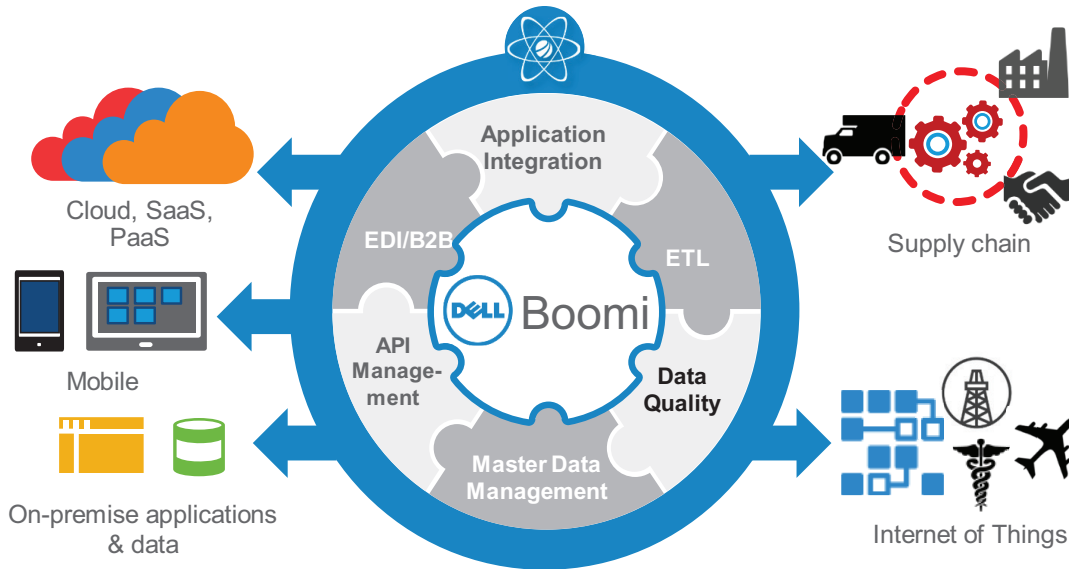
SHARED STORAGE HADOOP VS. DISTRIBUTED STORAGE HADOOP

It's a testament to Hadoop's flexibility that it supports multiple deployment models accounting for varying budget, performance, capacity and density requirements. The Dell EMC Isilon solution is a shared storage model where the persistent filesystem data for Hadoop is stored in an Isilon NAS cluster versus the distributed model where data is spread across the local storage of the Hadoop nodes themselves. These two approaches offer varying advantages:

Shared storage Hadoop	Distributed storage Hadoop
Single copy of data for IT workloads and analytics	Massive (hundreds of petabytes)
Reduced data center footprint (storage density)	Ability to use commodity platforms
Enterprise file management: data protection, security, storage tiering, etc.	Linear scaling
Independent scaling of storage and compute	Flexible replica model

TRANSFORM

Dell Boomi offer powerful data conversion capabilities that enable the migration of data from legacy systems to modern systems. The Boomi AtomSphere® integration platform as a service (iPaaS) supports all your application integration processes — among cloud platforms, software-as-a-service applications and on-premises systems. To help accelerate your data integration processes, Dell Boomi offers connectors for hundreds of applications, including connectors that enable you to move data from SaaS applications, databases and web services into an SAP environment.



STORE AND REPORT

Dell EMC offers a wide range of modern platform choices for data storage and reporting, including options for running SAP HANA in on-premises, off-premises cloud and hybrid environments.

BUY

Turnkey,
Maximum Agility

Ready Solutions Continuum | Solutions for every customer & workload

Build

Maximum
Flexibility

Hybrid Cloud Platforms

The ultimate turnkey, pre-integrated hybrid cloud system with lifecycle management and single SKU support experience



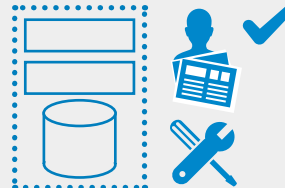
Ready Systems

Documentation, certification and/or tools to optimize workloads on a given engineered system



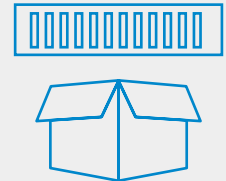
Ready Bundles

Validated and optimized complex bundle level RA + additional engineering value-add (benchmarks, deployment guides, quoting tools, etc.)



Ready Nodes

Pre-configured server validated and/or certified for a specific workload. Optional factory software installation.



Deliver better outcomes | Deploy faster and more easily | Mitigate risk

ON-PREMISES OPTIONS

With Dell EMC infrastructure options for on-premises deployment of the SAP HANA platform, you can choose any one or a combination of data center-ready solutions to achieve your goals. With each of these options, Dell EMC offers data protection and high availability for protecting SAP HANA data and meeting your recovery-time and recovery-point objectives.

These options include:

- **Ready Nodes** — Dell EMC offers Ready Nodes dedicated to SAP HANA. Built on Dell EMC™ PowerEdge™ servers with Intel processors, these Ready Nodes are pre-sized, pre-built and delivered with SAP HANA software pre-loaded.
- **Ready Bundles** — Dell EMC Ready Bundles bring together SAP-certified Intel-based Dell EMC servers, storage and networking, including support for SAP HANA Tailored Data Center Integration (TDI).
- **Ready Systems** — These pre-built systems deliver the convenience of an appliance and the flexibility of TDI, including options that incorporate Dell EMC V[x]Block, VxRail and VxRack.

HYBRID AND CLOUD OPTIONS

In addition to on-premises deployments, Dell EMC gives you the choice of hybrid cloud and off-premises platforms for your SAP HANA environment.

- **Hybrid cloud** — The Dell EMC Enterprise Hybrid Cloud solution offers the ultimate turnkey, pre-integrated hybrid cloud system with lifecycle management. It combines hardware, software and services from Dell EMC and VMware into a platform built on converged systems to deliver the foundation for infrastructure-as-a-service. Dell EMC and VMware have spent thousands of hours of engineering time designing, testing and proving the platform in our labs, so you don't have to expend that effort yourself.
- **Off-premises managed cloud for SAP** — Virtustream is unparalleled in running SAP in the cloud. The platform was purpose-built for handling complex, mission-critical enterprise applications with performance, security, resilience and services to support your entire IT estate. Our experts have extensive experience working directly with SAP applications, so we can help you seamlessly migrate your SAP deployments to the cloud.

ACCELERATING YOUR JOURNEY WITH DELL EMC AND AURITAS

Dell EMC and Auritas are ideally positioned to help your organization accelerate the retirement of your legacy applications and at the same time remain compliant to your retention standards.

Not only does Dell EMC have a distinguished history in engineering projects, benchmarking and best practices — Dell IT has created reference architectures for various editions of Hadoop. This combination enables Dell EMC to solve today's efficiency challenges and at the same time pave the ground for a risk free transition to Hadoop and the cloud.

Auritas, in turn, has developed a very successful "LSR Approach" for a headache-free retirement of your legacy applications. This highly structured approach in checking and verifying your specific situation from both functional and technical perspectives uses a "service" approach. We deploy a set of tools that provide a clear identification of the data in your legacy system. Once we identify the data and its characteristics, we deploy our tools that help you achieve a rapid and successful migration of the data into the Hadoop platform. After migration, our process validates the data that has been moved and triggers the decommissioning of the legacy application. Following this structured approach, you can rapidly decommission your legacy applications and save on costs of licenses and maintenance.

Together, Dell EMC and Auritas are uniquely positioned to provide answers to the constant challenge of doing more with less, while at the same time helping you reduce risk on the path to digital transformation and modernization.

ABOUT THE AUTHORS



Feroze Arif is the director of analytics and innovation for Auritas, LLC. He has more than 25 years of experience spanning data analytics and SAP master data management, data quality and governance, as well as SAP data migration. He is passionate about using data science for business and regularly uses Python, R and other technologies to build machine learning and predictive analytics solutions.



Divyesh Vaidya is an enterprise solution consultant who is responsible for driving business analytics and big data solutions in support of Dell EMC's enterprise solution strategy, incorporating Dell EMC products with partner solutions. An experienced product manager with proven ability in driving cross-organizational teams and interaction with enterprise customers, Divyesh has a breadth of experience in solutions product planning and strategy, including product roadmaps based on market requirements and product messaging.

To learn more, visit: Dell.com/sap or Auritas.com