As organizations aim to differentiate their offerings, they increasingly turn to data-driven strategies to increase revenues, better target consumers, and discover new means to lower operational costs. These data-driven organizations are making data the core input or “fuel” of their business models, as opposed to the exhaust from internal and external transactions. They are accomplishing this by using data from multiple internal sources such as operational support systems like billing, customer care, and supply chain management, and reaching out to external sources such as third-party data brokers, geolocation mapping services, and external payment platforms. With new opportunities to create value with information, companies want to implement environments that take advantage of a broad range of data, such as enterprise, social, and sensor data. However, managing information across database platforms presents significant hurdles to organizations that are trying to achieve these goals. The problem is not that the individual technical tasks for digital transformation, new business models, and data-driven exploration present an unachievable challenge. It is the overall complexity of the endeavor that is difficult.

The Enterprise Management Associates® (EMA™) Hybrid Data Ecosystem (HDE) provides a logical framework to link the business requirements of data landscapes with various underlying data platforms that support enterprise and multi-structured data. EMA developed the concept of the Hybrid Data Ecosystem to provide an architecture for next-generation data management and the effective implementation of those data landscapes toward business ends.

The HDE framework helps enterprises break down the complexity associated with these big data implementations in an organization’s overall data landscape. One of the best ways to break down the barriers of complexity is using technologies and techniques that:

- Allow business stakeholders such as analysts and data scientists to access deeper functionality, such as metadata management and data integration pipelines.
- Enable implementers within the CIO’s office and the IT department to scale their efforts across multiple projects.
- Give IT architects the flexibility to add new processing capabilities or data platforms without being “locked” into an architectural environment.

The SAP Data Hub gives organizations the opportunity to make a rapid impact on their data-driven initiatives by helping simplify the data landscape. The solution orchestrates, manages, and governs information across multiple usage scenarios from multiple data management platforms.

In a single, visual design environment, organizations create data pipelines that access, blend, transform, process, and move information from a variety of sources across an organization’s data landscape.

**Executive Summary**

**IN THE KNOW**

**WHO:** Business executives, architects, and stakeholders who want to be more data-driven

**WHEN:** When the challenges of complex data landscapes and traditional techniques overwhelm implementation teams and restrict data-driven innovation

**WHAT:** Learn how data-driven organizations implement innovative strategies, recognize the data landscape complexities that will challenge them, and understand modern orchestration, management, and governance solutions that will enable their success.
Digital Transformation and Dark Data Driving Change

Data-driven organizations seek opportunities for new business models that drive additional revenue, innovative ways to target consumers and improve customer experience, and new means to lower operational costs to improve margins.

Companies implementing these strategies make data the core input, or “fuel,” of their business model. They are doing this by using data from multiple internal sources such as operational support systems like billing, customer care, and supply chain management, and reaching out to external sources such as geolocation mapping services and external payment platforms. This data can come from many sources, including internal data centers, cloud-based repositories, and third-party providers. Using data as the fuel for these strategies disrupts and ultimately validates new avenues for business models and markets to operate.

As enterprises take on the process of digital transformation they uncover new, untapped sources. Many of these new data sources already existed within the organization, but were not utilized. This untapped and hidden information is called dark data. Often times, dark data already exists within an organization. However, because it is relatively hidden from the analysts, operations teams, and other decision makers that might make the best use of it, the data is considered a mystery. But the key for an organization is to be able to quickly and proactively leverage dark data.

For example, in the wireless telecommunications industry, much of the information about the mobile network and how customers interacted with it used to sit isolated with the network operation (NOC) teams. As the wireless phone industry matured and markets expanded, customer experience became an important component of customer retention. Through customer feedback, it was determined that poor connections to a provider’s data network and dropped calls in areas of poor network quality had a significant and direct impact on customer churn. For the most part, all the information was available within the NOC team repositories, but no one in marketing and customer care had the interest or the ability to ask about network quality data. With the prominence of customer churn in a highly penetrated marketplace, suddenly, the dark data of network quality information became important. It could directly impact how a mobile carrier managed customer complaints, marketed to new customers, and predicted customer churn better than their rivals. Just like those mobile carriers, companies today are focusing on shining a light on more of the data within their organization in order to increase their competitiveness.
Becoming Data-Driven: From Exploration to Operational Analytics

Although there are significant opportunities to create value from information hidden across the landscape, organizations can be challenged to implement environments that can truly take advantage of enterprise data, high volume and high velocity data (such as social and sensor data), and third-party data. The problem is not that the individual technical tasks for digital transformation, new business models, and data-driven exploration present an unachievable challenge. It is the overall complexity of the endeavor that is difficult.

Exploration and Discovery

The initial starting point for most companies is to explore, discover, and validate data, whether it exists within a new dataset or part of the dark component of an existing one. Effective exploration involves placing information in a location where stakeholders can discover the information, profile it, and test different possible uses.

Operational Improvements

After a dataset is validated via the exploration process, the various lines of business work to make that dataset available to create revenue, lower costs, or improve customer experience from an operations perspective. Again, this requires that data is prepared and sent to the right location. Only then can it be most effective for revenue, customer experience, or cost reduction. Perfecting this process means that a single set of data may be used in multiple locations, or a single task might invoke multiple sets of data to make a strong operational case.

Analytics

Enterprises use analytics to learn how those operational use cases are working. Across multiple dimensions from customer to product to process, companies need to evaluate and assess how the assumptions developed in the exploration and operations stages work in the real world. To accurately perform this monitoring and management, the data needs to be linked with contextual information on the dimensions of customer, product, and process, among others.

Operational Analytics

Finally comes the ability to perform real-time analytical algorithms on specific operational situations. EMA calls these situations “operational analytics,” which include an operational usage scenario where, at particular points, a real-time analytical result set needs to be invoked. For example, analyzing orders in real time is key for online merchants.

Determining if an order(s) fits an individual customer’s purchase profile lowers operational risks and prevents losses. This mixes the operational and analytical usage scenario workloads to minimize exposure and vulnerability. However, since operational analytics use a mixed set of workloads and data constraints, it can be the most complex usage scenario to implement and execute.

<table>
<thead>
<tr>
<th>Analytical processing (e.g., data science, data mining, advanced analytics)</th>
<th>18.8%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operational processing (e.g., rating, billing)</td>
<td>18.0%</td>
</tr>
<tr>
<td>Operational visibility (e.g., process status and reporting)</td>
<td>17.8%</td>
</tr>
<tr>
<td>Online archiving (e.g., replacement for tape archives and offline storage)</td>
<td>15.5%</td>
</tr>
<tr>
<td>Exploratory analysis (e.g., data discovery)</td>
<td>15.1%</td>
</tr>
<tr>
<td>Utilization of streaming data (e.g., IoT, streaming analytics)</td>
<td>14.8%</td>
</tr>
<tr>
<td>Other (Please specify)</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Figure 1: Usage scenarios implemented in big data initiatives
It used to be that the CIO and IT departments would support a few core initiatives over a long-time scale, but the emergence of data-driven project initiatives and big data implementations have changed that standard. They’ve increased the pressure of delivering across multiple projects and accelerated the tempo of deployment and updates.

More Projects, More Pressure

As the number of projects grows, the effort to manage and maintain data projects needs to expand to match. However, there are limits on the ability of traditional processes and practices to manage the sharp increase in the number of projects associated with data-driven initiatives.

EMA research has shown that big data implementations often include five or more projects that IT architects need to support. Over 6 of 10 organizations implementing a big data initiative have three or more different big data projects.

Greater Change, Greater Complexity

With the increasing number of projects comes an augmented tempo of the adjustments to projects driven by the continuous exploration of the data, and opportunities to improve on previous data components.

For example, in previous eras of predictive analytics and machine learning, changes to advanced analytical models were made on a quarterly basis. Data did not change as much. Business stakeholders had lower expectations. Now, EMA research is showing that companies are seeing much greater change to their advanced analytics projects. Instead of quarterly change, teams are now pressured to provide updates on a weekly or daily basis. EMA research shows that over 4 of 10 projects require updates at least weekly, and 7 of 10 require monthly change.
When companies bring together the disparate data platforms with a range of data sources to meet multiple usage scenarios, they require a coordinated data management approach to optimize the investment in data and information. This coordinated approach needs to place the business requirements of data landscape at its core, and include the “right” technical elements around them.

The Enterprise Management Associates (EMA) Hybrid Data Ecosystem (HDE) provides a conceptual framework to link the business requirements of a data landscape with disparate data management platforms. These platforms support enterprise, sensor, and multi-structured data. EMA developed the concept of a Hybrid Data Ecosystem to provide an architecture for next-generation data management and the effective use of increasingly complex landscapes.

The HDE enables organizations to coordinate between the range of platforms that now populate their big data landscapes. Validated since 2012 with end-user research relating to the implementation of big data initiatives and the associated environments, the HDE includes platforms that reside within the corporate data center, the cloud, and third-party providers, including:

- Hadoop
- NoSQL
- EDW and data marts
- Streaming systems
- Operational applications and support systems
- Analytical platforms
- Discovery platforms
- External and cloud-based data sources

Each of the platforms presents its own unique technical and operational challenges. This is over and above the difficulties with managing the different data types and workloads they were architected to support. Managing information between the platforms presents hurdles to data-driven organizations.

At the core of the HDE’s ability to coordinate between these many platforms is the idea that effective HDE implementations share operational, technical, and business metadata. With this metadata sharing component in place, enterprises implementing an HDE architecture can optimize the data residing on each of the platforms without attempting to “reinvent the wheel” with information duplicated across multiple platforms.
Finding an Easier Path Through the Complexity

There are some core concepts to consider when using the EMA HDE architecture to guide the implementation of a data landscape.

Companies must:

- Allow business stakeholders such as analysts and data scientists to access technical functionality like metadata and data integration pipelines between platforms, without having to understand the underlying technologies.
- Enable technologists within the CIO’s office and the IT department to scale their efforts across multiple projects and initiatives to maintain and improve the time to implementation across multiple projects.
- Give IT architects and directors the flexibility to add new data types and data management platforms as they emerge and not be “locked” into a technology decision.

One of the best ways to meet these objectives is to use configurable application implementations, as opposed to a custom coding approach that requires adjustments at each component of a data landscape. By using configurable implementation methodologies, organizations meet the objectives of encapsulating complexity, using technology to scale, and achieving implementation flexibility.

One of the areas where companies can have the highest impact on data landscape simplification is the area where Orchestration, Data Integration, and Security reside. Within this area, being able to easily view metadata from multiple platforms allows an organization to coordinate how information is managed across their platforms. This coordination optimizes each technical platform to the best business use.

A new offering that addresses this critical area is the SAP Data Hub. SAP Data Hub helps simplify the data landscape by orchestrating, managing, and governing the information that enables data-driven initiatives across multiple usage scenarios, leveraging data from a wide variety of data management platforms.

In a visual design environment, organizations can create data pipelines that access, blend, transform, process, and move information from a variety of sources across an organization’s data landscape. These pipelines can utilize powerful libraries for computation or machine learning and make it easy to incorporate the latest technology developments. SAP Data Hub pipelines also connect data of a wide variety of types, such as social media, customer, and product information to allow the organization to combine information in new ways and make new discoveries. SAP Data Hub leverages the existing processing investments already in the landscape, such as capabilities in SAP HANA, Apache Hadoop, SAP Vora, or Apache Spark. Finally, these pipeline models can be easily copied, modified, and reused to accelerate pipeline deployment. They can also be used to leverage best practices, and maximize reuse as the organization’s data landscape continues to expand and evolve with new initiatives, as well additional data sources and data types.
About SAP

SAP Data Hub

SAP Data Hub provides visibility, orchestration, and control across the changing data environment of an organization. By providing a broad, detailed, and easily understood view of the entire data landscape from sources like Hadoop and Amazon S3 to SAP HANA and ERP, SAP Data Hub helps organizations understand data sources, uses, interconnections, quality, and impacts. This enables enterprises to see new opportunities from data, resolve emerging data issues, and ensure that data is flowing to where it needs to go.

- **Management of zones, systems, and connections**
- **Configurable adapters**
- **Service discovery**
- **Data quality scorecard**

**Access and security center**
- **Identity control (user, groups, roles)**
- **Policy management**
- **Security logging**

**Data hub operations**
- **Data landscape management**
- **System and metadata discovery**
  - Browse connected systems
  - Profile data for data quality
  - Export data sets for modeling
  - Perform lineage and impact analysis
- **Data model lifecycle**
  - Export and import of data models
  - Configuration repository integration
  - Figuration management
- **Data pipeline creation**
  - Data model authoring
  - Data sets: SAP Vora, SAP HANA, files
  - Tasks: Flowgraph and vFlow
  - Predefined tasks: Copy, BW, PA

**Data studio and refineries**
- **Orchestrate execution**
  - Scheduling of execution
  - Workflow definition (TaskChain)
  - Monitoring and analysis

**Pipeline orchestrator**
- **Data model authoring**
- **Data sets: SAP Vora, SAP HANA, files**
- **Tasks: Flowgraph and vFlow**
- **Predefined tasks: Copy, BW, PA**

**The SAP Data Hub helps organizations:**

- **Understand and manage systems and data across the complex data landscape** – SAP enables organizations to browse connected systems to profile and expose an organization’s information for data discovery, processing, and analysis. Since it is an open architecture, systems and data can be located within existing SAP applications and platform or be located in non-SAP systems.

**Support and manage a complete data model lifecycle** – SAP enables architects to manage the lifecycle of their data model, from development, configuration, and version control to repository integration from a single environment.

**Create data pipelines that go across cloud, on-premises, and hybrid data management environments** – SAP empowers organizations to create complex, multistep pipelines that incorporate a variety of data types (big data, enterprise data, usage scenarios), operational support applications and analytical environments, and platforms such as Hadoop, NoSQL, and relational databases.

**Orchestrate and monitor the execution of data pipelines** – SAP Data Hub helps organizations schedule and monitor workflows associated with data pipelines between data and applications.

**Data landscape governance** – SAP enables organizations to manage zones, systems, and connections, as well as understand the state of system data with a data quality scorecard to ensure application of data governance policies.

**Provide enterprise data access and security controls** – SAP Data Hub safeguards data with enterprise security and data access controls implemented across all orchestrated pipelines.

**SAP Background**

SAP develops and markets enterprise application and database software. It serves various lines of business, including asset management, commerce, finance, human resources, manufacturing, marketing, research, development and engineering, sales, service, sourcing and procurement, supply chain, and sustainability, as well as the consumer, discrete manufacturing, energy and natural resources, financial services, public services, and services sectors. SAP SE was founded in 1972 and is headquartered in Walldorf, Germany.
About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help EMA’s clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise line of business users, IT professionals, and IT vendors at www.enterprisemanagement.com or blogs.enterprisemanagement.com. You can also follow EMA on Twitter, Facebook, or LinkedIn.

This report in whole or in part may not be duplicated, reproduced, stored in a retrieval system or retransmitted without prior written permission of Enterprise Management Associates, Inc. All opinions and estimates herein constitute our judgement as of this date and are subject to change without notice. Product names mentioned herein may be trademarks and/or registered trademarks of their respective companies. “EMA” and “Enterprise Management Associates” are trademarks of Enterprise Management Associates, Inc. in the United States and other countries.

©2017 Enterprise Management Associates, Inc. All Rights Reserved. EMA™, ENTERPRISE MANAGEMENT ASSOCIATES®, and the mobius symbol are registered trademarks or common-law trademarks of Enterprise Management Associates, Inc.