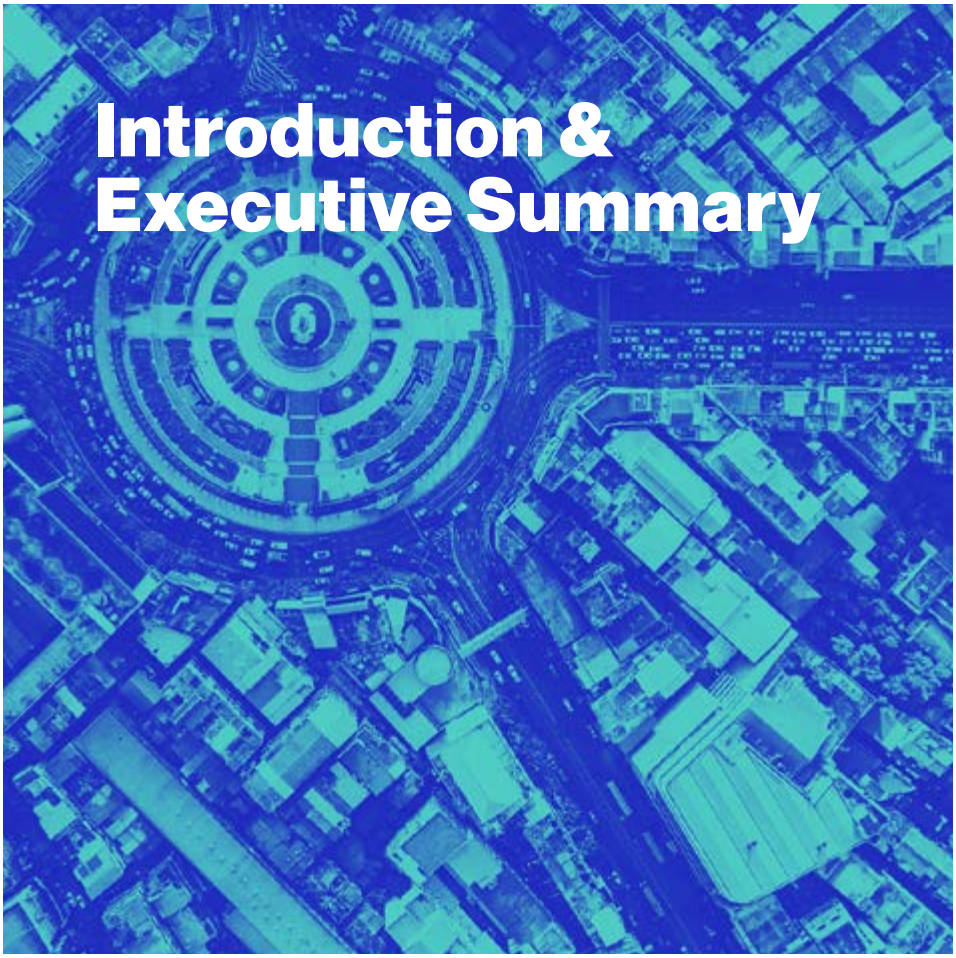
Integrate Big Data Analytics with Content

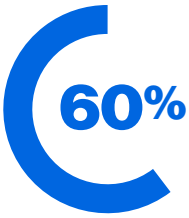
Using Systems of Insight to Optimize Business Processes and Content Performance



Introduction & Executive Summary

Big Data Projects Never Finished/ Ultimately Abandoned

Infochimps Survey 2013[1]



Gartner forecast through 2017[1]

Organizations use big data analytics to identify potential opportunities or problems that may otherwise go unnoticed. Launching a successful big data project however, is much easier said than done, as **more than half of big data projects are unsuccessful.**

Successful big data initiatives enable organizations to effectively integrate, manage, and understand all relevant data sources, **to optimize business processes**[2]

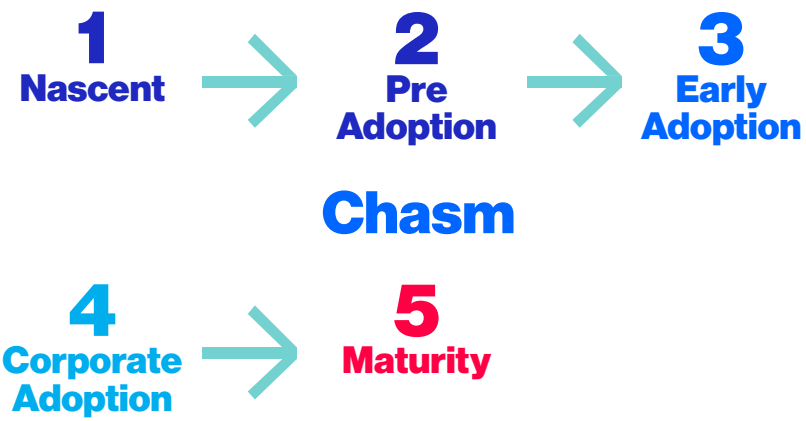
Big data engines provide the computing power to process clickstreams, sensor data and other machine-generated data. They calculate scores for one or more entities using machine learning libraries for predictive analytics, auto-classification, sentiment analysis and other advanced analytic models. An integrated content management system can leverage this information to provide further insight into the application performance and user experience.

This paper explains how to integrate big data analytics with a content repository and the benefits of using big data calculations for operational improvements. Implementing an integrated solution helps organizations make big data operational, pervasive, and actionable. Learn how a content management platform can use big data analytics to expedite this benefits.

1. Infochimps, CIOs & Big Data (2013) & Gartner, Business Intelligence and Analytics Leaders Must Focus on Mindsets and Culture to Kick Start Advanced Analytics (Press release, Sept. 15, 2015).
2. Fern Halper and Krish Krishnan, TDWI Big Data Maturity Model Guide (2013-2014; emphasis added).

Make Big Data Operational

TDWI offers a Big Data Maturity Model (see chart below) to help organizations assess where they currently stand with big data initiatives and what must be done to maximize business value:



Let's "begin with the end in mind" and look first at the desired end state of TDWI's big data maturity model. What benefits will your organization see as a result of successfully formulating and executing a big data project?

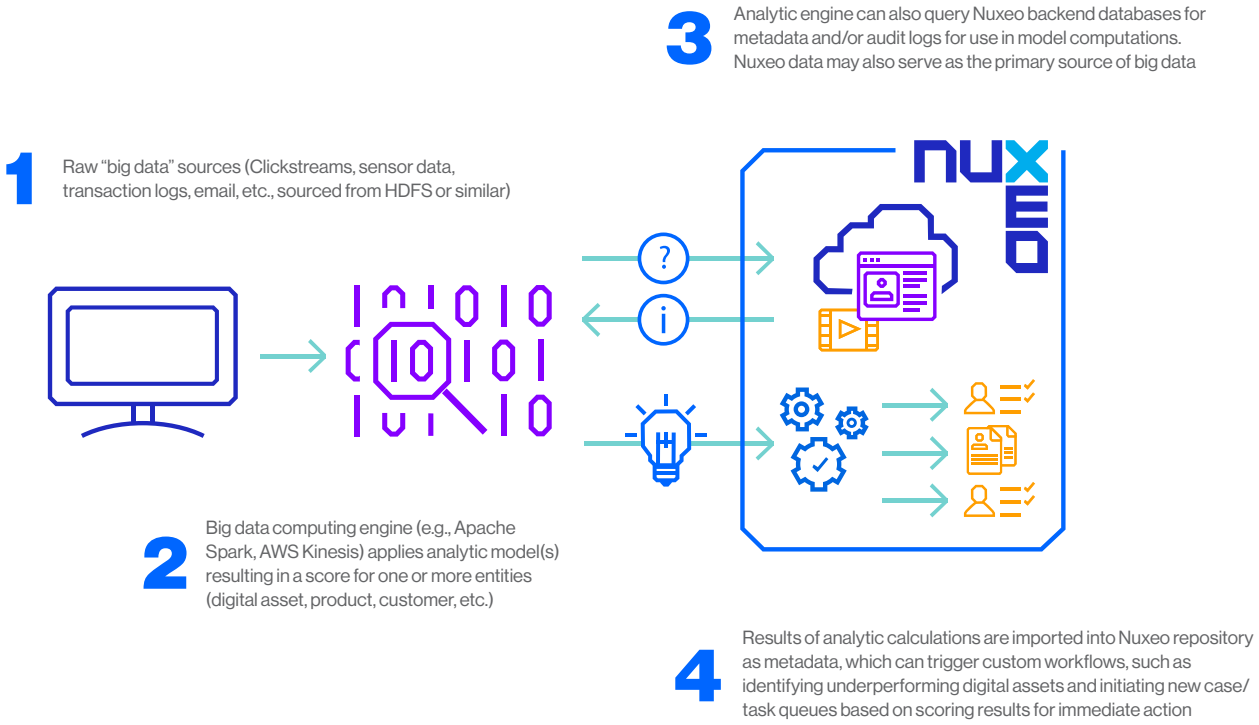
TDWI answers this question by describing what big data analytics looks like in organizations achieving its Visionary level of big data maturity. These organizations share key big data capabilities **at the operational level—**including:

- **New data coming into the organization can be quickly analyzed and made part of the logical infrastructure**
- **Improved ability to "connect the dots" between new data and existing assets**
- **Increased integration of big data analytics within business processes**
- **Collaborative decision making processes are directly supported by big data analytics**

A content repository integrated with big data analytics, makes these benefits possible.

How Nuxeo Helps Make Big Data Operational

The Nuxeo Platform enables organizations to rapidly and systematically apply big data analytics as a driver of optimized business processes, as shown in the conceptual chart below.



Big data analytic engines such as Apache Spark and Amazon Kinesis provide the computing power to process clickstreams, sensor data and other high volume, high velocity machine-generated data. They calculate scores for one or more entities using machine learning libraries for predictive analytics, auto-classification, sentiment analysis and other advanced analytic models—all at massive scale.

Working in tandem with big data analytics, the Nuxeo Platform provides early-warning indicators of new opportunities or problems to the right managers and teams at the right time, as well as empowering them to take timely and decisive action:

- **Improve analytic calculations by integrating metadata from the content repository.** Analytic engines may achieve more precise and useful analytic results by integrating other external data sources—**including the Nuxeo Platform itself.**

  The Nuxeo repository offers a highly flexible source of structured master data for customers, users, digital assets, products and other entities referenced within big data sources.

Nuxeo utilizes a schema-flexible metadata model designed to adapt to rapidly changing business needs, as well as enable the creation of **intelligent content objects.**

While typical document management tools often rigidly define a document as a single file, a Nuxeo document can be a highly complex content object. Each content object can contain multiple documents and binary types, along with all related attributes (metadata and hierarchy), which you are free to define and modify to match your business needs. Unlike other tools that store content as "files" in "folders," a **Nuxeo content object** can include multiple files and file types with advanced metadata including nested properties, which in turn can include complex types and lists of complex types.

- **Utilize audit logs as a big data source.** The Nuxeo Platform provides an advanced audit log module to capture and record any and all system or application events. You can configure custom events to be logged and store anything you need alongside each audit log entry. Nuxeo provides the massive scalability necessary to manage the most demanding audit logging requirements, with support for NoSQL (MongoDB) as a content backend, as well as Elasticsearch for storing and searching of audit logs.

  Nuxeo's audit logs can also serve as a primary source of big data to analyze enduser productivity or perform in-depth application security audits. For example, audit logs for a video-on-demand application could enable advanced analytics of customers' viewing preferences, popularity of video assets and much more.

- **Add analytic results to the content repository as new metadata.** Results from big data analytic models can also be imported within select Nuxeo content objects (see page 3 chart, item 4) as new metadata associated with related digital assets, customers, products and more, enabling immediate action as well as valuable insights from new trend-based analyses.

- **Automatically initiate new case queues.** The Nuxeo Platform includes an event bus that allows the registration of event listeners, which can trigger one or more workflows upon a given event taking place.

  For example, if a certain digital asset, such as a marketing campaign video, is considered to be underperforming based on its big data-generated score, Nuxeo workflows can dynamically create new cases and prioritized lists of tasks assigned to select managers. Nuxeo can also include links to related content in the Nuxeo repository, equipping managers with essential contextual information to take informed action.

  It is also possible to apply business logic that initiates actions taken by the Nuxeo Platform without requiring human interaction, such as automatically replacing an underperforming digital asset with the next most similar asset.

Nuxeo Platform In Action:

Improving ad campaign engagement through digital clickstream analysis and targeted digital assets

A global marketing agency uses the Nuxeo Platform for digital asset management, along with big data analytics to help maximize the performance of multi-channel advertising campaigns in near real-time.

Global advertising campaigns require multiple versions of videos, display ads, images and ad copy, which must be localized to match the culture of its particular target audience. All of these assets are stored in the Nuxeo repository, linked and associated together with extensive metadata, easily accommodated by Nuxeo's schema flexible data model. This includes in-depth tracking of all talent and usage rights (e.g., signed releases from actors and other performers defining how and when they may use their likeness and work) to avoid costly contract breaches.

The agency recently added clickstream analysis to identify which online videos, web content and other online materials were most frequently accessed by geographic location, age groups and other data points. Results are imported into the Nuxeo repository as new metadata for each asset. Content not performing at expected levels is identified quickly for review and possible modification or replacement.

This initiative has already improved consumer engagement, and identified new opportunities for combining print and web-based assets in campaigns.

Benchmarked to the tune of over one billion documents, the Nuxeo Platform provides massive scalability using either NoSQL (MongoDB) or a relational database (PostgreSQL, Oracle, MySQL, etc.).

Nuxeo also tightly embeds Elasticsearch within the platform, providing horizontally scalable enterprise search, data aggregations, real-time analytics and rapid integration of Nuxeo applications with big data analytics. Native Elasticsearch connectors already exist for Apache Spark and Hadoop.

Make Big Data Pervasive

Before organizations achieve TDWI's Visionary stage of big data maturity, where the most compelling benefits of big data analytics are found, organizations must first "cross the chasm."

The chasm in the TDWI maturity model represents challenges which must be overcome for big data analytics to be utilized within cross-department and crossfunctional business operations. Without widespread acceptance, big data initiatives will remain stuck in a niche, limiting its potential for transforming business processes.

While TDWI identifies a number of key challenges that make up this chasm, concerns over data management and data governance stand out as particularly critical issues to overcome:

In order to move forward to corporate adoption and sharing of data, a solid data management and governance plan needs to be in place. Some organizations believe this is the single most important key to big data maturity... Departments [might] begin to fight for control over who owns the data or who may want their particular vision in place [regarding infrastructure and] issues such as data security or management. [3]

How Nuxeo Helps Make Big Data Pervasive

By leveraging key functionality already provided by the Nuxeo Platform, organizations can quickly address concerns over big data management and governance, to achieve widespread big data adoption:

- **Always on, military-grade security model.** Every user request goes through a central authentication filter. Any user request must be authenticated before the Nuxeo Platform will perform anything.

  Nuxeo also provides **two** concurrent layers of security that users **cannot** bypass:

  Access control lists (ACLs), used by Nuxeo to manage security at the data level. Nuxeo provides the ability to allow administrators to define ACLs via the Nuxeo UI. ACLs can also be applied programmatically. Because ACLs can be erroneously set by an administrator, organizations can also choose to use some automatically set ACLs, or enforce company security rules by leveraging custom security policies (see next bullet).

3. Ibid.

Custom security policies consist of dynamic code that will be executed each time the Nuxeo repository needs to determine whether or not to allow a user access request or other user action. Custom security policies enable Nuxeo to enforce mandatory access controls (MACs) that may override or supplement any applicable ACLs—a common requirement for military systems.

- **Location-agnostic management of content.** The Nuxeo Platform does not require content to be stored in a centralized file system. Nuxeo stores the actual content (binary streams) in a BLOBstore, while storing its content attributes in a separate SQL or NoSQL database, along with references and links to the content, as well as versioning, security access controls, content lifecycle states and workflow details. The binary files can be stored locally or in the cloud; content delivery networks (CDNs) such as CloudFront or Azure CDN; or even residing within cloud file sharing services including Google Drive, Dropbox, Microsoft OneDrive and Box.

- **Complete audit logging at massive scale.** Nuxeo provides a comprehensive audit logging module to record, monitor and track any system or application event. In addition to its multiple proactive layers of security, an audit trail exists to provide a review and analysis of any historical user or system activity which may be questionable in hindsight.

Nuxeo Platform In Action: Electronic Arts securely delivers and tracks game builds worldwide

Electronic Arts (EA) is a global leader in digital interactive entertainment, providing blockbuster video games, content and online services for game consoles, PCs and mobile devices. The company leverages the flexible metadata and security capabilities of the Nuxeo Platform to provide global developers with reliable, secure access and delivery of game builds throughout the entire game development lifecycle.

New game builds are submitted automatically into EA's Nuxeo Platform-based application, with custom metadata fields automatically populated, including information relating to access controls and the locations of the build content: source code, artwork and audio assets. Upon validation, the application retains the information provided in a custom document type (content object) for that build, with a lifecycle status indicating it is available for access. Notifications are then automatically sent to everyone involved in the process.

Developers from approved sites submit online requests for build content, which is transferred to the destination and delivered directly to the requestor, typically via download to a console or PC.

A complete audit history for each game build is maintained, logging all user activity, such as when a build was checked in or out and by whom, and any modification of permissions.

EA fully utilizes Nuxeo's out-of-the-box access control resources, as well as some custom security policies; for example, only QA team members can update the lifecycle state of a game build from alpha to beta, pre-production and, eventually, final release of the game.

Nuxeo enables fine-grained security, by business role, down to the metadata level for safe, reliable delivery of game builds—EA's most critical asset.

Make Big Data Actionable

Finally, let's look at organizations currently within TDWI's early "pre-chasm" stages of big data maturity. What steps should these organizations take to move their big data initiatives forward?

Organizations in the Early Adoption stage will have developed a successful big data proof of concept that is departmental in scope and near-production ready. This requires organizations to address a variety of early data and infrastructure-related issues.

Organizational culture, however, may present a much greater challenge; for example, TDWI warns that early big data exploration may be met with skepticism among business managers.

How Nuxeo Helps Make Big Data Actionable

As noted in the introduction above, **the ultimate end goal of big data analytics is to optimize business processes.** Doing so means workers and dispersed teams must be able to freely access, share and collaborate on content on demand:

> **Enterprises differentiate themselves from competitors based on the value that individuals and teams create; it is critical for IT leaders to ensure that those workers have the content they need... Content can be the outcome of collaborative processes and tasks, and is often the evidence or substance of decision making. [4]**

The quickest way to convert big data doubters into supporters is to present new and useful big data analytic calculations along with content-driven applications that render those insights actionable.

Nuxeo serves as a critical bridge between big data analytic results and timely, focused, measurable business action.

4. Karen M. Shegda & Hanns Koehler-Kruener, Content Is a Critical Dimension to Digital Workplace Success, Gartner (May 20, 2015)

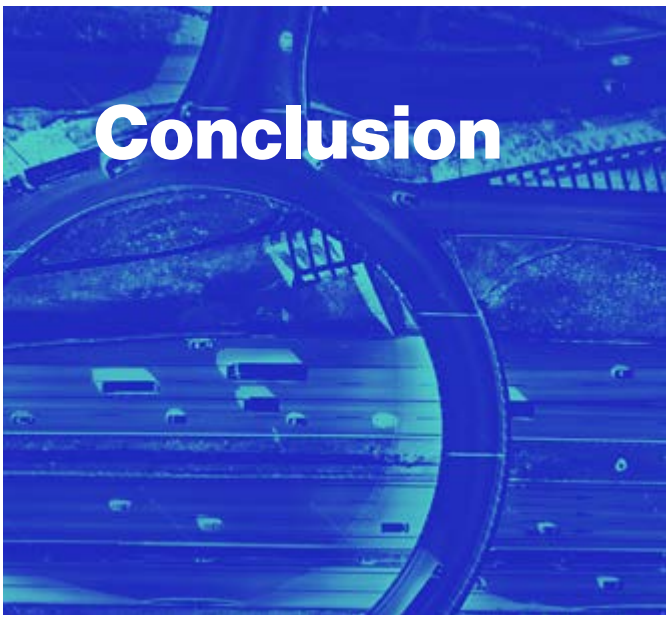
Nuxeo Platform in Action:

Identifying potentially fraudulent credit applications with big data scoring

A leading credit review agency utilizes algorithmic models to analyze data from online loan applications to identify which applications might be fraudulent and give those applications a score. Scores that signify a material fraud risk are directed to the case management application built on the Nuxeo Platform, which automatically creates a new case for each flagged application.

Credit agents are automatically assigned tasks to further investigate the applications in question and are provided the detailed scoring results from the analytic engine and other content in the Nuxeo repository associated with the credit applicant. Other Nuxeo case functionality also ensures timely action, such as automatic escalation of tasks if not completed within a set timeframe by the assigned worker.

The Nuxeo Platform helps the company manage and direct anti-fraud investigations, reducing credit fraud before it starts—at the loan application level.



Conclusion

The optimal combination of people, processes and technology enables organizations to find and act upon early indicators of business problems or opportunities in real time. Big data analytics hold tremendous potential to optimize business processes. Doing so requires big data maturity—a journey that the Nuxeo Platform will help expedite, to quickly make big data analytics:

- **Operational – freely integrating big data analytics within business processes company-wide**
- **Pervasive – resolving make-or-break concerns over big data management and governance by leveraging unique existing Nuxeo functionality that ensures security, flexibility and scalability**
- **Actionable – providing a bridge between big data analytic calculations and specific, measurable business actions**

- **Download the Nuxeo Platform.**
- **Watch the online training videos on Nuxeo University to learn how to build a Nuxeo Platform application, as well as read our online documentation center.**
- **Contact us to request a demo or find out more.**

About Nuxeo

Nuxeo enables organizations to manage complex digital content at massive scale. Software architects and developers build mission-critical applications using Nuxeo's content management platform. The Nuxeo Platform offers native integrations with leading technology providers, flexible storage options, and extensive packaging capabilities for managing media & digital assets, structured & rich content, and advanced business processes. Over 1,000 organizations rely on Nuxeo to run business-critical applications, including the IRS, Electronic Arts, Verizon, Sharp, Capital One, and the U.S. Navy. Nuxeo is headquartered in New York and Paris. More information is available at **www.nuxeo.com.**