DASSAULT SYSTEMES

EXALEAD

EXALEAD CLOUDVIEW™ PLATFORM HIGHLIGHTS

3DS.COM/EXALEAD

EXALEAD CloudView™ is a unified information access platform enabling a new generation of innovative Search-Based Applications (SBAs) as well as providing superior enterprise and Web search.

IDC on EXALEAD

"EXALEAD is disruptive because the company has moved aggressively from Web search to enterprise search, and now to information access. The firm's technology makes it possible to integrate structured and unstructured content in a unique way to address mission-critical applications in areas such as extended business intelligence, customer support, compliance, and many others. In addition, EXALEAD offers its customers scalability, reaching across multiple content repositories including desktop, legacy apps, third-party outsourcing providers, and, finally, the Internet—to bring information access to new levels and decision-making intelligence to business professionals throughout the enterprise."

Susan Feldman, Stephen E. Arnold & Ryan Patterson, IDC Vendor Profile

WE WELCOME YOUR FEEDBACK

Whatever your role—IT analyst, system administrator, application end user, business manager, security expert, or simply a curious reader—your feedback is important to us. We invite you to contact us at http://www.3ds.com/products/exalead/ with your comments, suggestions or questions.

EXECUTIVE SUMMARY

Companies today are facing an information access crisis. Most of the essential information they need to thrive in a highly competitive environment is inaccessible to the people who need it most: their employees, customers and partners. Specifically, steep learning curves and heavy licensing and infrastructure costs hamper access to valuable information stored in corporate databases and enterprise applications like Enterprise Resource Planning (ERP) and Customer Relationship Management (CRM) (i.e., 'structured' information). And while database access tools are restrictive, employees often have no tools at all for locating and exploiting the 'unstructured' data that makes up the bulk of corporate information assets—information encapsulated in resources such as email, chat, blogs, forums, RSS feeds, videos, and Office documents.

Online businesses face a similar challenge. They need to provide easier, innovative access to a broader range of information to attract and build their audiences, but the cost of doing so can be exhorbitant, even when and if the technical challenges can be overcome.

Introducing EXALEAD CloudView™: Eliminating Information Access Barriers

CloudView is a one-of-a-kind search engine that collects unstructured and structured data from any source, in any format and in any volume, and automatically transforms it into a single structured information resource. This resource, which continually evolves and adapts as your data evolves, can be directly searched or used to develop innovative business applications.

CloudView for Enterprise Search

Deployed directly as a search engine, CloudView lets employees instantly locate files no matter where they are stored—on their desktop, on network servers, the company intranet or out on the World Wide Web—and to easily discover related content, automatically generating a unique menu for refining searches and exploring related material for each user query.

Beyond Search... Bringing Agility, Innovation & Lower Costs to Application Development

Rapidly deployed without altering existing information systems, CloudView is also enabling a new generation of Search-Based Applications (SBAs) that are reducing IT costs and complexity and driving innovation. CloudView reduces IT costs by:

- Providing alternative data access that is as rich as relational database querying yet 100s of times faster and far cheaper
- Scaling infinitely and on-demand by simply adding inexpensive commodity hardware
- Reducing time to market for new applications from months or years to days or weeks

CloudView drives innovation by enabling you to incorporate an unprecedented depth and variety of information in your enterprise and Web applications, including emotive and qualitative data from unstructured sources.

The pages that follow provide details about the CloudView platform and how it provides these benefits and more.
We hope you will find this information helpful as you evaluate information access options and opportunities for your organization.

TABLE OF CONTENTS

FIGURES

TABLES

1. CLOUDVIEW PLATFORM OVERVIEW

1.1 What is CloudView?

EXALEAD CloudView™ is a revolutionary search engine and unified information access platform enabling better search and innovative Search-Based Applications (SBAs).

1.1.1 Unified Information Access for Better Search

CloudView was developed simultaneously for the enterprise and for the Web, driving a 16-billion page public search engine and serving 100 million researchers a month through CloudView-powered websites. Because of this dual Web/enterprise DNA, the CloudView search engine alone combines Web simplicity, scalability and innovation with features essential for the corporate environment, including:

- The ability to efficiently access and index structured data (data stored in corporate databases and enter prise applications)
- The ability to automatically organize and classify staggering volumes of unstructured content (such as email, Web pages, RSS feeds, multimedia files and Office documents)—and to intelligently synthesize this data with structured data
- The search refinement tools essential for task-based business search
- The capacity to fully adhere to stringent data security requirements

1.1.2 Unified Information Access for Better Web and Business Applications

Beyond search, CloudView provides a unified information access platform that is revolutionizing the development of Web and business applications. CloudView can collect data in any format from any source (the Web, email servers, databases, intranets, multimedia archives, etc.), and automatically transform it into a cohesive, meaningful information resource.

This resource, the CloudView index, is continually evolving, endlessly scalable and easily accessible via standard Web-based technologies. Businesses are using it to provide alternative data access that reduces the load on over-taxed database systems, and to construct a new generation of innovative Web and enterprise applications. Far more efficient than traditional database-driven applications, these new applications harness an unheralded depth and breadth of information sources yet are quick and easy to construct.

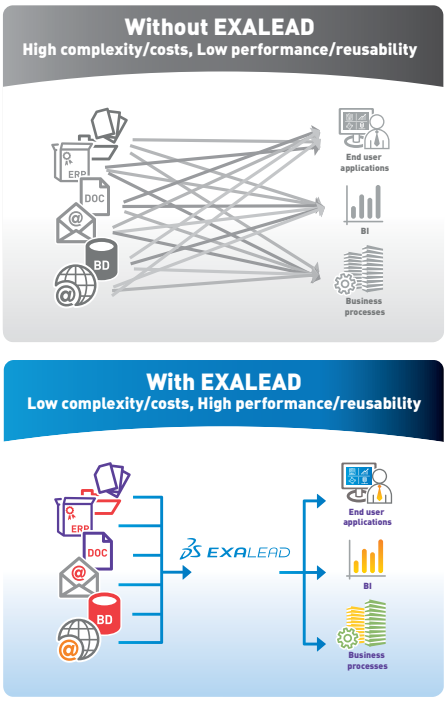

Figure 1: CloudView Simplifies Information Access and Reduces IT Costs

1.2 HOW IS CLOUDVIEW BEING USED?

1.2.1 CloudView Solutions

CloudView solutions include 1) superior search for the enterprise and the Web, 2) embedded search and information access for Original Equipment Manufacturers (OEMs) and Independent Software Vendors (ISVs), and 3) innovative Search-Based Applications (SBAs) that leverage search engine-derived technologies to more effectively exploit database assets, expand the scope and improve the performance of business applications, and bring innovation and depth to Web applications.

Enterprise and Web Search

CloudView helps you find information—fast. Looking for a PowerPoint presentation given to Company X? From a single text box, you can easily find that file no matter where it is stored—on your desktop, network servers, or out on the Web—and at the same time explore related information: email feedback from Company X about the presentation, the Company X profile on the intranet, transcripts of the presentation debriefing call... the options are limited only by your security rights.

This winning combination of effective search and rich content discovery is ideal for website search as well, helping users locate information or products quickly while enticing them to explore related content and services.

Embedded Search & Information Access

CloudView is also being used by OEMs and ISVs to embed seamless, scalable search and information access functionality in their own commercial products. CloudView's flexible architecture, unlimited data source connectivity, and open Application Programming Interface (API) framework enable the platform to be embedded in virtually any type of application, including Storage & Archiving, Messaging, Enterprise Content Management, Information Lifecycle Management, Compliance and eDiscovery.

Search-Based Applications (SBAs)

- **Improved Database Applications**
  SBAs provide access to the information contained in database systems via a search engine index and complementary Web technologies rather than through direct database queries. This index-based database offloading strategy enables IT to cut costs for licensing, infrastructure and development while simultaneously boosting performance and expanding access. Index-based query processing is 100s of times faster than traditional database querying, and deeply structured queries, mathematical operations, Web-style fuzzy natural language search, and faceted navigation are all natively supported (CloudView's semantic processors preserve the extensive classification information contained in relational database tables. CloudView also enables real-time operational reporting to be generated on-the-fly for any or all data characteristics maintained in the database.

Figure 2: Unified Intranet Search for CEA; Better Web Portal Search for Rightmove
France's Atomic Energy Agency, CEA, unified search across 150 intranets with CloudView. Rightmove improved search features and performance on their real estate classifieds portal while reducing the cost of search from £0.06 to £0.01 per 100 queries.

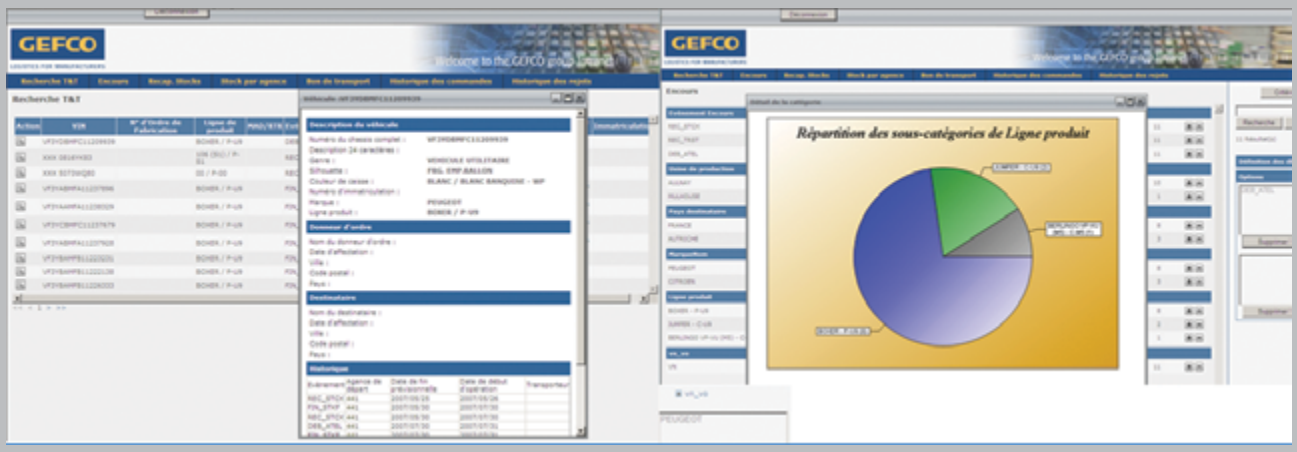

Figure 3: Database Offloading for GEFCO: Reduced Costs, Improved Performance

GEFCO customers use this CloudView-powered extranet to track and optimize vehicle transport across 80 countries. Deployed in only 60 days, CloudView reduced the load on GEFCO's Oracle databases while improving performance, with data latency cut from 24 hours to 30 seconds. Users can also now drill down on an endless number of characteristics for reporting and research, a breadth and depth impossible to achieve using standard pre-determined SQL queries.

- **Smarter Business Applications**

  CloudView is also being used to bring new agility and expanded scope to enterprise applications like Business and Competitive Intelligence (BI and CI), Customer Relationship Management (CRM), and Supply Chain Management (SCM). CloudView can infuse these applications with important emotive and qualitative data from 'unstructured' sources like email, blogs, chat, telephone transcripts, Web pages and more, boosting application relevancy and improving decision making. Because CloudView was designed to manage real-time data updates, it also improves the timeliness of information, enhancing business agility and competitiveness. More intuitive, Web-style search also boosts end user adoption rates and system usage, increasing application ROI.

- **Innovative Web Applications**

  CloudView enables online business to add instant depth and stickiness to their sites through innovative mashup applications, that is to say, applications that merge content and functionality from diverse sources such as databases, mapping services, business applications and the Web. In contrast to traditional mashups that awkwardly juxtapose multi-source content, EXALEAD mashups use CloudView's fully unified information platform for seamless presentations that are as deeply engaging as they are easy to manage.

> ***"Businesses now need to keep tabs on thousands of blogs and billions of other Web pages to understand what people are saying about their products. The real insight from this content comes when it's aggregated and summarized in some meaningful way for deeper analysis."***
>
> Forrester Research, *Search + BI = Unified Information Access*

CloudView unifies and structures diverse content from unlimited sources, enabling dynamic mash-ups that are seamlessly presented and easy to manage

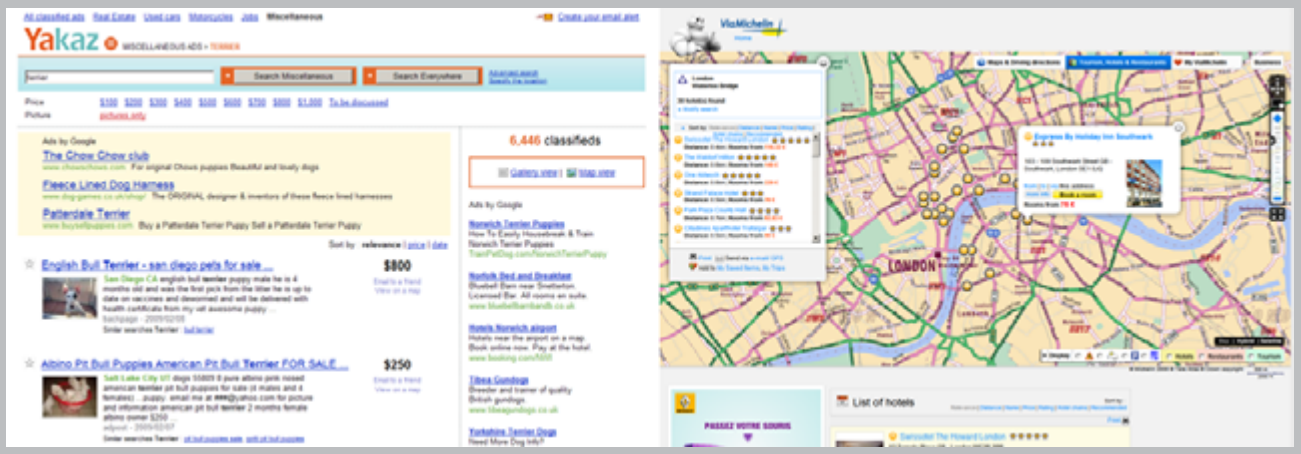

Figure 4: Online Innovation for Yakaz and ViaMichelin

For the classified ad portal Yakaz.com, CloudView provides seamless structured search of content culled from nearly 7,000 websites. Developed in only 4 weeks, ViaMichelin, Michelin's CloudView-powered travel portal, is an engaging mashup of database information, Web content and dynamic mapping for 15 million points of interest (hotels, restaurants, attractions, etc.).

1.2.2 CloudView Products

The CloudView platform is available in three editions:

- CloudView Search
- CloudView OEM
- CloudView 360°

CloudView Search provides a feature-rich, endlessly scalable and quickly deployed solution for enterprise and Web search. Both CloudView Search and CloudView 360 provide a unified information access platform for developing Web and business applications. CloudView 360 features unique semantic tools (i.e., computer-based interpretive tools) for analyzing data and optimizing it for business use. CloudView OEM is used exclusively by OEMs and ISVs to embed search and related information access functions within their own commercial products.

1.3 PLATFORM DIFFERENTIATORS

1.3.1 Automatic Structuring of High Volumes of Unstructured Data

CloudView is especially adept at deep text processing of massive volumes of unstructured data. Beyond simply identifying keywords in a document, deep text processing means transforming unstructured content into a fully classified resource that can be synthesized with existing structured data, such as that from corporate databases and business applications.

For example, CloudView can instantly contextualize a product sales report in a BI system by incorporating Web 'notoriety' statistics for that product (i.e., the number and emotive quality of product mentions on the Web), as well as qualitative data drawn from support forums, email messages, phone transcripts, and much, much more.

1.3.2 Infinite, Cost-Effective Scaling

Engineered for Web-scale processing, CloudView is the only enterprise search engine designed from inception for multibillion document scalability. More importantly, CloudView scales effortlessly and cost effectively. The system is extremely resource efficient, supporting real-time indexing of 100 million documents and processing up to 20 queries per second on a single dual-processor server. And, thanks to its distributed architecture, you can scale CloudView on demand by simply adding inexpensive commodity hardware.

CloudView can index 100 million documents on a single server

1.3.3 Agile, Open Architecture

CloudView further provides the most flexible, extensible platform on the market. Its service-oriented architecture (SOA) and extensive Application Programming Interfaces (APIs) ensure:

- Maximum data flexibility, with the ability to connect to any internal or external source
- Agile, low-cost application development, with an independent data layer and support for standard Web formats and protocols
- Maximum scalability, performance and availability, with built-in service distribution and data replication capabilities for easy, low-cost scaling

1.3.4 Distinctly Easy to Install and Use

Though all core CloudView functions are openly accessible and configurable, CloudView is also a smoothly packaged solution designed for rapid deployment. In fact, CloudView typically deploys in days or weeks, not months or years as is common for other solutions.

Ongoing maintenance is also fast and easy with a Web interface for platform administration, and the patented user interface is highly user-friendly, having been refined for immediate "zero-training" use by millions of Web users. Indeed, CloudView is so easy to deploy, administer and use, it has earned EXALEAD an exceptional 100% customer loyalty rate.

CloudView's ease of deployment, administration and use has earned EXALEAD a 100% customer loyalty rate

2. PLATFORM ARCHITECTURE

2.1 Core Services: Collect, Process, Access, Interact

The CloudView platform is composed of four core services operating within a secure framework:

- **Service 1: COLLECT**
  Collects unstructured and structured data from internal and external sources
- **Service 2: PROCESS**
  Transforms the data collected into a single structured resource
- **Service 3: ACCESS**
  Updates the enhanced data and processes user and application queries
- **Service 4: INTERACT**
  Provides interaction via a customizable Web interface or visual dashboards

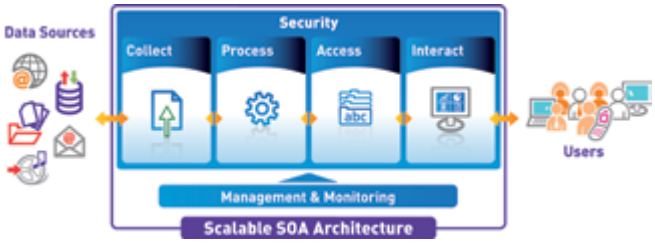

Figure 5: CloudView Platform at a Glance

2.1.1 Service 1: COLLECT

First, CloudView gathers data from designated internal or external sources across the enterprise Cloud. The platform provides native support for more than 120 languages and 300+ data formats, with built-in connectors to enterprise information sources ranging from groupware applications to intranets, Content Management Systems, file servers, email systems and over 50 types of data sources. Furthermore, special Web connectors make it possible to build thematic Internet crawls around specific subjects (competitors, industries, products, etc.).

CloudView provides built-in support for 300+data formats and 50+ types of data sources

In addition to this extensive array of built-in connectors, CloudView offers an open, standards-based data collection API ('Push API', or PAPI) to extend connectivity to virtually any data resource, even legacy and non-standard systems.

An open Push API extends data source connectivity infinitely

2.1.2 Service 2: PROCESS

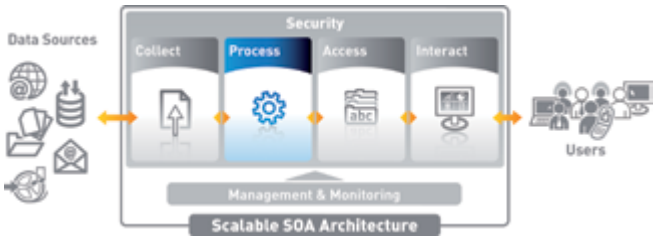

Figure 6: Service 2: PROCESS

The platform next transforms all the heterogeneous data collected into a single exploitable resource. First, CloudView automatically and independently analyzes, classifies and categorizes all unstructured and structured data, identifying attributes such as document keywords and keyword variants, proper nouns (for example, the names of people, places and organizations), and metadata like document location, file type, author and creation date.

It then identifies embedded meanings and relationships within and across these resources, meanings and relationships that can be used to extend business applications or create powerful content mashups.

This data, together with information like unique document identifiers, security rights, and ranking and relevancy indicators, constitutes the CloudView index. This rich index can be made available to other applications in virtually any format desired, including the versatile default format, XML, today's leading standard for data encoding and exchange.

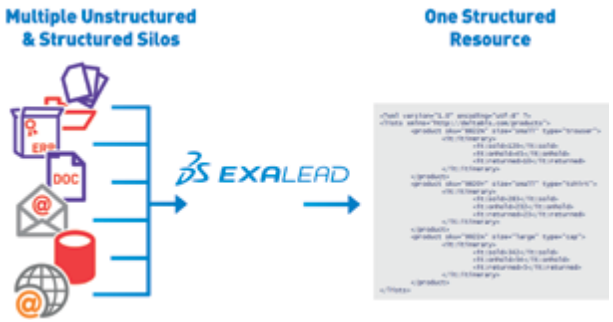

Figure 7: Unstructured & Structured Data Becomes a Single Structured Resource

TECH NOTE

CloudView Natural Language Processing

Natural language processing refers to the techniques computers use to 'read' and analyze text the way we human beings do. The goal is to find a way to analyze, classify and tag volumes of data that would be impossible to process manually, and to help computers better understand and respond to queries formulated as 'natural' language questions as opposed to complex programming queries. Accordingly, CloudView uses numerous natural language processors during both indexing and query processing. These processors, which are individually accessible and configurable, provide functions such as:

- Language detection
- Tokenization and normalization, with sentence boundary recognition (parsing text into individual words and sentences, applying language-specific rules regarding separators like white space and punctuation)
- Stemming (identification of words sharing the same stem, for example, "engine" and "engines")
- Lemmatization, morphological and syntactic processing (identification of not only basic stems but of more complex variants, like "good" and "better," and applying language-specific knowledge of word and sentence construction patterns)
- Part of speech tagging

In addition to these basic modules, activation of certain linguistic resources during indexing automatically triggers deployment of additional processing modules.

2.1.3 Service 3: ACCESS

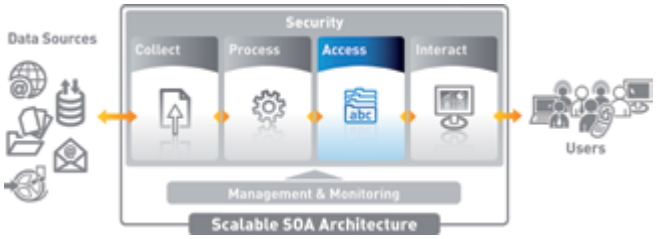

Figure 8: Service 3: ACCESS

The platform's third core function is to keep the index up to date and process user and application queries—all in real time and with outstanding performance.

Index Updates

Updates can be executed 1) in real time, 2) at specified intervals (hourly, daily, etc.), 3) or provided "just in time." More efficient than real-time updates, "just in time" updates are made on the fly only when queries are received from specified applications. This strategy satisfies both your users' need for timely data and your need to optimize system resources.

Whichever strategy you choose, references used during indexing, such as dictionaries and thesauri, are automatically updated as the index is updated.

Query Processing

For business applications, CloudView provides query processing that is 100s of times faster than relational database querying, enabling the kind of sub-second responsiveness Web-savvy users expect while reducing the operational load on expensive database systems—without compromising the integrity of the database.

On the human side, CloudView deploys a host of tools to ensure that even if an end user's input isn't perfect, their search results can be. To this end, CloudView features:

- Spell checking (a mistyped 'hammr' will return 'hammer')
- Checks for word variants (e.g., 'hammer' will also match 'hammering')
- Phonetic matching ('exaleed' will match 'exalead')
- Approximate matching ('exalaed' will match 'exalead')
- Presentation of related terms and concepts and other search refinement aids, including options based on language, data location, file type, author, creation date, and more.

> Search engine users expect search results to be accurate with respect to what they are looking for, not what they typed in the search box. And what they *type* is decidedly not always the same as *what they are seeking.*

TECH NOTE

CloudView Query Processing

Whether end user or application-generated, incoming queries may contain textual, numerical, and symbolical constraints, with extensive Boolean operator support. The queries pass through a processing pipeline which is fully configurable, supporting, for example, the expansion of the search terms using specific semantic rules and dictionaries, application of additional security restrictions, or enforcement of application-specific ranking rules.

2.1.4 Service 4: INTERACT

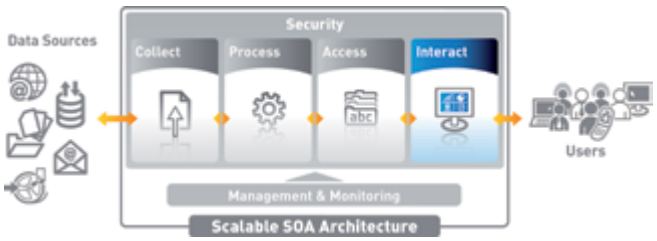

Figure 9:Service 4: INTERACT

CloudView's interaction framework provides endless flexibility in searching and exploiting your data. You can deploy EXALEAD's award-winning search interface as is (or customize it using CSS, JavaServer Faces or JavaServer Portlet technologies), or use the Search API to create custom applications.

Built-In Interface

This interface features a patented navigation system that dynamically generates a unique menu for each user query. This menu offers options for narrowing or broadening searches as well as links to related content (related terms and categories, links to other materials by the same author or from the same source, etc.).

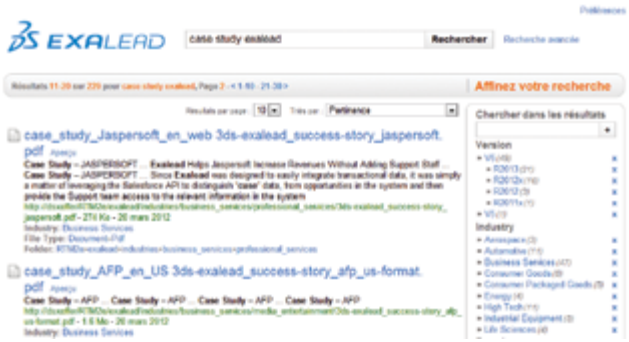

Figure 10: CloudView's Built-In Web Interface

Custom Applications

Supporting standard interfaces (see Section 2.3.2), CloudView's Search API enables you to create custom search interfaces, generate information dashboards, develop unique content mashups, and enhance or create information-rich business applications.

Figure 11: The Many Faces of CloudView

2.2 Service-Oriented Architecture (SOA)

CloudView features a service-oriented architecture (SOA), which means its core operations (like indexing and query processing) are made available as on-demand 'services' that can be easily tapped by other applications or services using standard Web-based technologies. This not only accelerates application development, it also enables the platform to be installed and accessed by users and applications without requiring changes to, or interfering with, existing information systems. This provides maximum business agility while preserving existing IT investments.
The platform's core services are also distributable, meaning they can be easily duplicated and/or split across an unlimited number of servers, providing maximum availability and scalability. Finally, all of the system's core services are fully configurable. System administrators can easily adapt the platform to meet any unique business need using either the Web-based management console or the standards-based Application Programming Interfaces (APIs).

TECH NOTE

CloudView's administrable components are delivered as a set of Web services communicating either locally or remotely through a secure framework (encrypted TCP/IP channels). These components are designed to be easily distributed and/or replicated, and include built-in support for Web formats and protocols: SOAP, REST, XML, RSS, RDF, OWL, etc.

2.3 Open API Framework

CloudView extends platform agility even further with three public, open interfaces (APIs) for accessing, configuring, and controlling core system functions:

- Push (Data Collection)
- Search API
- A Management API

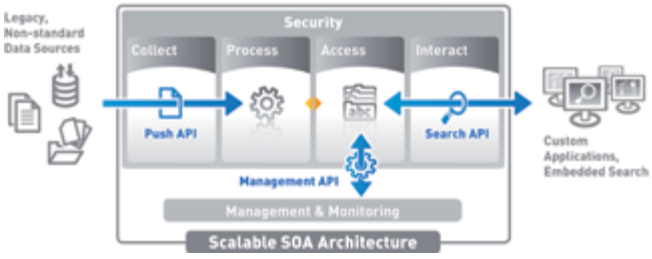

Figure 12: CloudView's Application Programming Interfaces (APIs)

2.3.1 Push API

The Push API complements the extensive library of built-in CloudView data source connectors by enabling you to create, configure and manage your own custom connectors, thus extending platform connectivity to any data repository, even legacy and non-standard systems.

TECH NOTE

Based on a simple HTTP/REST protocol, the Push API can be used by any programming language, with higher level Java, C# and Exascript wrappers provided (Exascript is EXALEAD's object-oriented XML language, blending Java and XML).

2.3.2 Search API

As noted in Section 2.1.4, you can use CloudView's built-in Web interface for standard search deployments, or you can use the Search API to construct interfaces and applications that make your Cloud data available wherever, however and to whomever you choose. The Search API is also used by OEMs and ISVs to embed information access functionality in their applications. Because it supports common leading programming languages and Web formats, development using this API is typically very fast, with an average time to market of 60 days or less (see our whitepaper, *The Hidden Costs of Scaling*).

TECH NOTE

The Search API supports multiple programming languages (Java, .NET, PHP, Ruby, Python and Perl) and Web formats and protocols (SOAP, REST, XML, RDF, OWL, etc.). To make development even easier, the system features a developer kit that includes tools like front-end code samples.

2.3.3 Management API

The Monitoring and Management Interface (MAMI) permits all stages of the indexing and search processes to be configured, managed and monitored through an API. This API can be used to build a custom administration interface, or to integrate management functions in third-party applications embedding the CloudView OEM edition.

TECH NOTE

The Management API can be used to access MAMI functions either:

- Directly using SOAP,
- Through a Java RPC client that encapsulates the SOAP, or
- Through a command line helper that exposes all operations for scripting or testing

2.4 Management & Monitoring

In addition to the access provided by the MAMI, CloudView Search and CloudView 360 feature a Web console for fast platform configuration and deployment as well as easy maintenance and management.

Administrators can use the management console for tasks such as:

- Assigning user privileges according to security access levels
- Controlling ranking criteria and the depth and freshness of data collection (crawls)

- Modifying the appearance of the results page and categories
- Monitoring usage statistics such as the top search requests, most frequently consulted documents, requests returning zero results, etc. (these statistics can be exported so administrators can use their own data-mining tools to analyze usage and performance)
- Manage index replication and data backup processes
- Control versioning, rollback and updates of sub-components (the system supports concurrent updates, enabling, for example, concurrent configuration of the search and index build processes)

The system also offers failure alarms and diagnostics that notify IT administrators upon detection of faults.

Web-based tools make configuration, deployment and management fast and easy

2.5 Security

EXALEAD's philosophy is that your information access platform should adapt to your existing corporate security infrastructure, not the other way around. EXALEAD also believes that while security should tightly enforce existing security rules, it should not interfere with the end user experience. Therefore, CloudView provides users with the convenience of single sign-on access for all resources while enforcing source-specific rules.

This means the platform behaves for an individual user as though it had only crawled and indexed the content authorized for that particular user. It not only blocks access to unauthorized documents, but also to the titles, summaries, document previews and other metadata associated with those documents.

For further protection, the CloudView platform responds in real time to changes in user permissions and rights.

CloudView combines convenient single sign-on access with deep, metadata-level security

TECH NOTE

CloudView data-level security is achieved using Access Control Lists. Internal and external network interactions are protected via secured standards (AES, HTTPS). System security can be configured and monitored via the MAMI Security Manager, which supports local system security, LDAP, OpenLDAP, Active Directory, Domino Directory, & Remote HTTP.

3 SYSTEM PERFORMANCE

3.1 Endless Scalability & High Performance

As noted in Section 1.3.2, CloudView is the only enterprise search engine designed from inception for multi-billion document scalability. This scalability extends in five directions:

- The volume of data processed
- The total number of system users
- The number of queries processed per second
- The index refresh rate
- System features and functionality

This scaling is not only unlimited, it is also resource-efficient and virtually effortless.

Figure 13: CloudView Scales Endlessly in Five Directions

3.1.1 Resource-Efficient Engineering

CloudView scales in a linear manner using only a fraction of the resources of legacy search solutions. EXALEAD's founders believed that while indexing and search are intense activities, there was no reason why—with the correct approach to software engineering—a single commodity server could not support enterprise search on a massive scale. Accordingly, they engineered CloudView to deliver millisecond responsiveness from commodity hardware, providing real-time indexing of 100 million documents on a single dual-processor server. As a result, CloudView can typically handle 5 times the throughput of legacy products against 10 and 20 times the content—meaning far fewer servers to purchase, license and administer.

3.1.2 Effortless Scaling

Furthermore, scaling with CloudView is dynamic and 'pain-free': its unique distributed architecture is designed to scale on demand simply by adding commodity processors or servers—no painful migration process is required. This provides crucial business agility and continuity, enabling your business to scale serenely no matter how sharply or rapidly demand increases (see Section 3.2, page 12).

CloudView provides effortless, on-demand scaling for maximum business agility and continuity

TECH NOTE

Features enabling this high performance and scalability include CloudView's:

- Optimized C/C++ core software, with system calls optimized for each OS (Linux, Solaris, Windows), and a storage layer specifically tailored to maximize physical memory usage and disk throughput
- Unique statistical and mathematical efficiencies integrated into the indexing and query processing chains
- Optimized network communications
- An asynchronous architecture for running concurrent tasks
- A distributed architecture for the index, index-build, dictionary and query processing services, enabling rapid scaling using commodity hardware
- Index optimization tools for achieving an optimal balance between advanced features, indexing speed and search speed
- Three-level caching to boost performance: caching of user queries, the inverted list (list of terms and matching documents) and concepts (clusters of related terms similar to a thesaurus)

> ***"In my investigation of search company technology, I learned that EXALEAD's ability to scale is comparable to Google's... Most enterprise search and content processing systems cannot handle billions of documents—EXALEAD does. EXALEAD's search and content processing solutions give the company a technical advantage over vendors whose systems choke when thousands of users simultaneously want access to information."***
>
> Stephen Arnold, Industry Analyst and President of ArnoldIT

3.1.3 Performance & Scaling Benchmarks

Below are average performance statistics for CloudView indexing and query processing. For specific client benchmarking data, please see our whitepaper *The Hidden Costs of Scaling*.

Indexing Performance

Context	Record Type	Indexing Speed
Telco Log	Small	4000 records/second/server
Web Index	Medium (Web Pages)	16 billion records/week
Email	Upper Medium	200 records/second/server

TABLE 1: CloudView Average Indexing Performance

Query Processing Performance

Context	Records	Performance per Server
E Commerce	15 million	200 queries/second
Web Index	70 million	30 queries/second
Archiving	200 million	5 queries/second

TABLE 2: CloudView Average Query Processing Performance

3.2 High Availability

CloudView's index can be partitioned and duplicated across an unlimited number of servers, ensuring high performance and non-stop availability by distributing the query load, and providing back-up access in case of a hardware failure. State-of-the-art transaction logging and locking models ensure that all partitions and replicas remain fully synchronized. Clustering, load balancing and monitoring capabilities extend beyond the index to other key functions as well (like query processing) to ensure that there is no single point of failure anywhere in the architecture.

Similarly, the system allows most operational procedures to be accomplished without scheduling any downtime. Documents can be dynamically added, removed, or replaced in the index, and new index replicas can be created or removed without stopping other replicas or the core index construction process. It is also possible to update a single field of a document without modifying the content of other indexed fields (modifying, for example, the price of an indexed item without having to re-index other information such as color, size, or description).

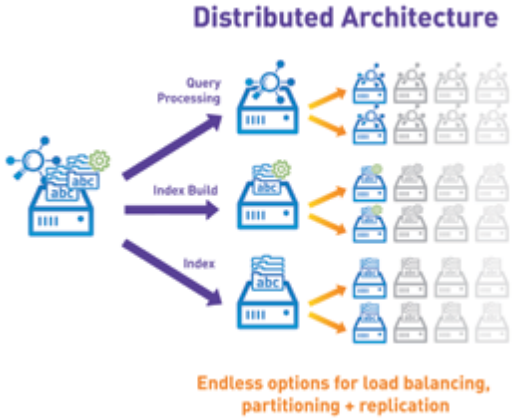

Figure 14: Maximum Availability and Performance with Load Balancing, Partitioning and Replications

3.3 Rapid Time to Market

For general search uses, CloudView can be deployed in only days with minimal professional services support (most deployments require no professional services support). Even deployments for advanced business applications and sophisticated data mashups are typically achieved within 2-8 weeks, not months or years, and likewise require minimal professional services support. This type of rapid deployment is possible because CloudView is both a fully packaged product designed for 'plug and play' use, and a standards-based system allowing open access to all core functions for fast adaptation to specific business needs. Below are deployment statistics for several recent CloudView installations that demonstrate this capacity.

3.4 Agile Development

Beyond initial deployment, CloudView provides an agile base for rapidly constructing new business applications. Application agility is assured by CloudView's fully unified data access platform, which decouples data from the underlying applications that generate it, synthesizes and structures it, then makes it available via standard Web-based standards and protocols. The API framework provides further agility by enabling adaptation of all services to evolving business needs.

Achieve a Rapid Time to Market with CloudView

Project	Time to Market	Description
Intranet Search	60 days	Intranet search on 100 million documents (Oracle db records)
Genome Knowledgebase	10 days for search component; less than 6 weeks total	Knowledgebase for genome database and related scientific articles. Initial index base of 1.2 billion documents, growing by 120 million every 2 months
Logistics Tracking Application	Prototype in 10 days; deployment in 60 days	600,000 transactions, 1TB of Oracle data; includes geolocalization and quasi-real-time data refresh
Hybrid Online Directory	60 days	Web/database mash-up & geolocalization; 40 million webpages and database records
Travel Portal	Prototype in 10 days; deployment in 60 days	Heavy-traffic portal providing information & mapping for 15 million points of interest (hotels, restaurants, attractions, etc.)

TABLE 3: Deployment Benchmarks for CloudView

For more deployment benchmarks, see the EXALEAD whitepaper *The Hidden Costs of Scaling Search*

4 PRODUCT USABILITY

4.1 "Zero-Training" End Use

CloudView's customizable interface combines the speed and simplicity of Web search with the rich output of a structured enterprise application. Users can enter their queries in a single, familiar text box, and in return are presented with deep menus for refining their search and exploring related information. The result is immediate intuitive use, more successful search, easier content exploration, and exceptionally high user adoption rates.

Features include:

- Navigation of results by dynamically-extracted categories, user ratings, related terms, file type and size, language, author and more
- At-a-glance scanning of results with extracts, file type icons, and thumbnail images
- Rich, application-independent content previews with search term highlighting
- Intuitive, 'forgiving' search, with advanced semantic processors performing the hard work of interpreting user requests and offering spelling corrections, close matches, and related content.

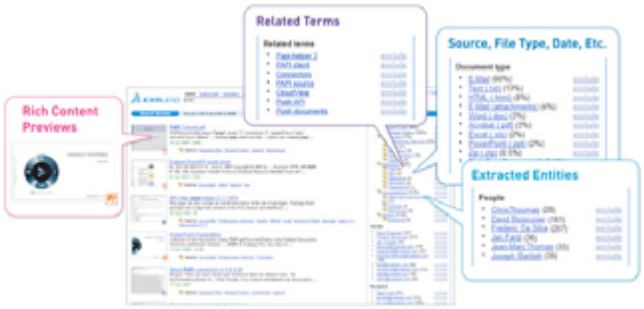

Figure 15: Intuitive, Faceted Navigation with CloudView's Web Interface

4.2 Easy Administration

This ease of use continues into the back-end, with a Web-based console for all management tasks: user interface configuration, indexing control, search performance tuning, etc.—tasks additionally exposed via APIs for maximum flexibility. CloudView's engineers have also strived to make the platform as self-maintaining as possible:

- All of the natural language processing modules (dynamic categorization, spell-checking, spelling suggestions, etc.) evolve automatically and in real-time as your data evolves.
- Index updates are also automatic, real-time and incremental. Documents can be dynamically added, removed, or replaced, and updated at the individual field level.
- Dictionaries are likewise fully automatic, incremental, and real-time. No hand-built dictionary or manual assistance is needed.
- The index build components are fully distributed and designed to run 24 hours a day, 7 days a week, without any human intervention. They also automatically perform routine maintenance tasks such as removing references to deleted documents.

These 'self-maintenance' features not only make administration easier and more pleasurable, they significantly reduce administrative labor costs, lowering TCO and augmenting ROI.

5 CONCLUSION

With a unique dual focus on both Web and enterprise search, EXALEAD has concentrated all its R&D efforts on meeting end users' need for fast, intuitive information access, and IT's need to simplify operations, gain agility, and reduce costs. The resulting CloudView platform is unique in the marketplace for its:

- Deep structuration and semantic processing of Web-scale volumes of unstructured data
- Exceptional performance and low TCO, processing on average up to 100 million documents and 20 queries per second on a single commodity server
- SOA architecture and open API framework, providing the Web-style agility essential for succeeding in a rapidly evolving Cloud environment
- Ease of use for end users, developers and administrators, with a 100% customer loyalty rate as a result

Whether you are seeking

- a better enterprise search solution,
- better performance, richer insights and a lower TCO for business applications,
- better search, enhanced content, differentiating features and reduced costs for your website, or
- an embedded information access platform for your ISV application,

CloudView provides a solution that will position you ahead of your competition, no matter what the Cloud has in store for you.

DASSAULT SYSTEMES

Delivering Best-in-Class Products

CATIA
Virtual Products

GEOVIA
Model and Simulate our Planet

SOLIDWORKS
3D Design

EXALEAD
Information Intelligence

SIMULIA
Realistic Simulation

NETVIBES
Dashboard Intelligence

DELMIA
Digital Manufacturing and Production

3DSWYM
Social Innovation

ENOVIA
Collaborative Innovation

3DVIA
3D Communication

Dassault Systèmes, the **3D**EXPERIENCE Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 150,000 customers of all sizes, in all industries, in more than 80 countries. For more information, visit www.3ds.com.

Europe/Middle East/Africa
Dassault Systèmes
10, rue Marcel Dassault
CS 40501
78946 Vélizy-Villacoublay Cedex
France

Asia-Pacific
Dassault Systèmes
Pier City Shibaura Bldg 10F
3-18-1 Kaigan, Minato-Ku
Tokyo 108-002
Japan

Americas
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

Visit us at
3DS.COM/EXALEAD
