GETTING THE BEST FROM ENTERPRISE CLOUD SEARCH SOLUTIONS

Implementing effective enterprise cloud search demands in-depth know-how, tuning, and maintenance

WHY MOVE TO THE CLOUD?

More and more organizations recognize that moving enterprise systems to the cloud is a cost-effective, highly-scalable option. In fact, by 2020, according to IDC, 67 percent of enterprise infrastructure and software will include cloud-based offerings.

No surprise then that we're seeing such rapid growth in enterprise cloud search solutions.

These can help modern organizations stay agile, providing multiple benefits including:

- high availability
- flexibility and scalability
- security
- ease of maintenance and management
- cost reduction (no cost for an on-premise infrastructure and more transparent pricing models)
- improved relevance – accessing more data to feed into search machine-learning algorithms
- unified search – easier integration with content from other products offered by the same vendor

Benefits like these are all within reach. But how to capture them? While cloud companies provide the infrastructure and toolsets for most search use cases – from intranet search and public website search to search-based analytics applications – truly effective search capabilities demand in-depth knowledge and expert implementation, tuning, and maintenance.

With so many options, and more cloud solutions launching all the time, what's the right option for you?

This white paper provides a useful guide, exploring some of today's leading enterprise cloud search solutions, unpacking their key features, and highlighting challenges that can arise in their deployment:

- Cloud search use cases

- Cloud search solutions, pricing models, and key features:
  - Elastic Cloud
  - AWS Elasticsearch
  - AWS CloudSearch
  - SharePoint Online
  - Azure Search
  - Google Cloud Search
  - Coveo Cloud
  - SearchStax
  - Algolia

- Challenges and considerations when deploying a cloud search platform

PUTTING CLOUD SEARCH TO WORK –
LEADING USE CASES

CLASSIC DOCUMENT SEARCH

This encompasses holistic enterprise search solutions that enable users to find information easily and securely. These applications can index content from multiple, disparate document repositories (often heterogeneous and containing a variety of file formats) and serve a range of user needs (from deep, specific research to simple fact-checking and document retrieval). Classic document search applications seek to provide an information retrieval and discovery tool that can offer a "360-degree view" of the enterprise.

Examples of classic document search use cases include:

- **Intranet portals**
  For enterprises with multiple documents in various repositories (perhaps billions of documents amounting to petabytes of data), these can give internal users quick access to relevant enterprise-wide documents.
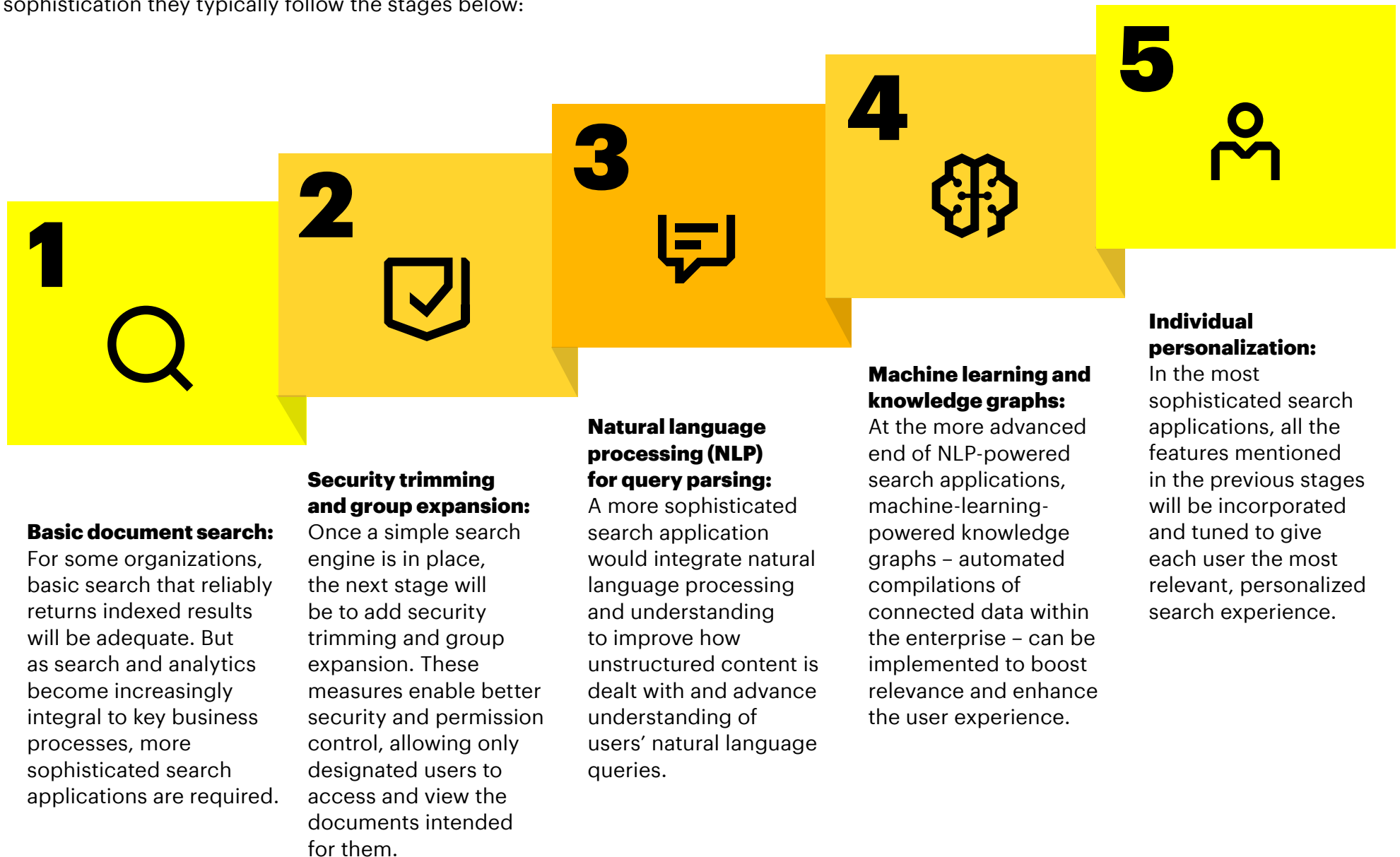
- **Public-facing website search**
  Encountered by most people daily, this is one of the most common search use cases: the search box on a website helps users find the information they're looking for quickly, without having to navigate through all the site's various sections.

- **Document repositories search**
  Helps users retrieve document repositories: these are groups of documents that share common features (defined by a set of fields, with field values captured when documents are indexed).

While all these applications are at different levels of maturity, as they grow in sophistication they typically follow the stages below:

1

Basic document search:
For some organizations, basic search that reliably returns indexed results will be adequate. But as search and analytics become increasingly integral to key business processes, more sophisticated search applications are required.

2

Security trimming and group expansion:
Once a simple search engine is in place, the next stage will be to add security trimming and group expansion. These measures enable better security and permission control, allowing only designated users to access and view the documents intended for them.

3

Natural language processing (NLP) for query parsing:
A more sophisticated search application would integrate natural language processing and understanding to improve how unstructured content is dealt with and advance understanding of users' natural language queries.

4

Machine learning and knowledge graphs:
At the more advanced end of NLP-powered search applications, machine-learning-powered knowledge graphs – automated compilations of connected data within the enterprise – can be implemented to boost relevance and enhance the user experience.

5

Individual personalization:
In the most sophisticated search applications, all the features mentioned in the previous stages will be incorporated and tuned to give each user the most relevant, personalized search experience.

PUTTING CLOUD SEARCH TO WORK –
LEADING USE CASES

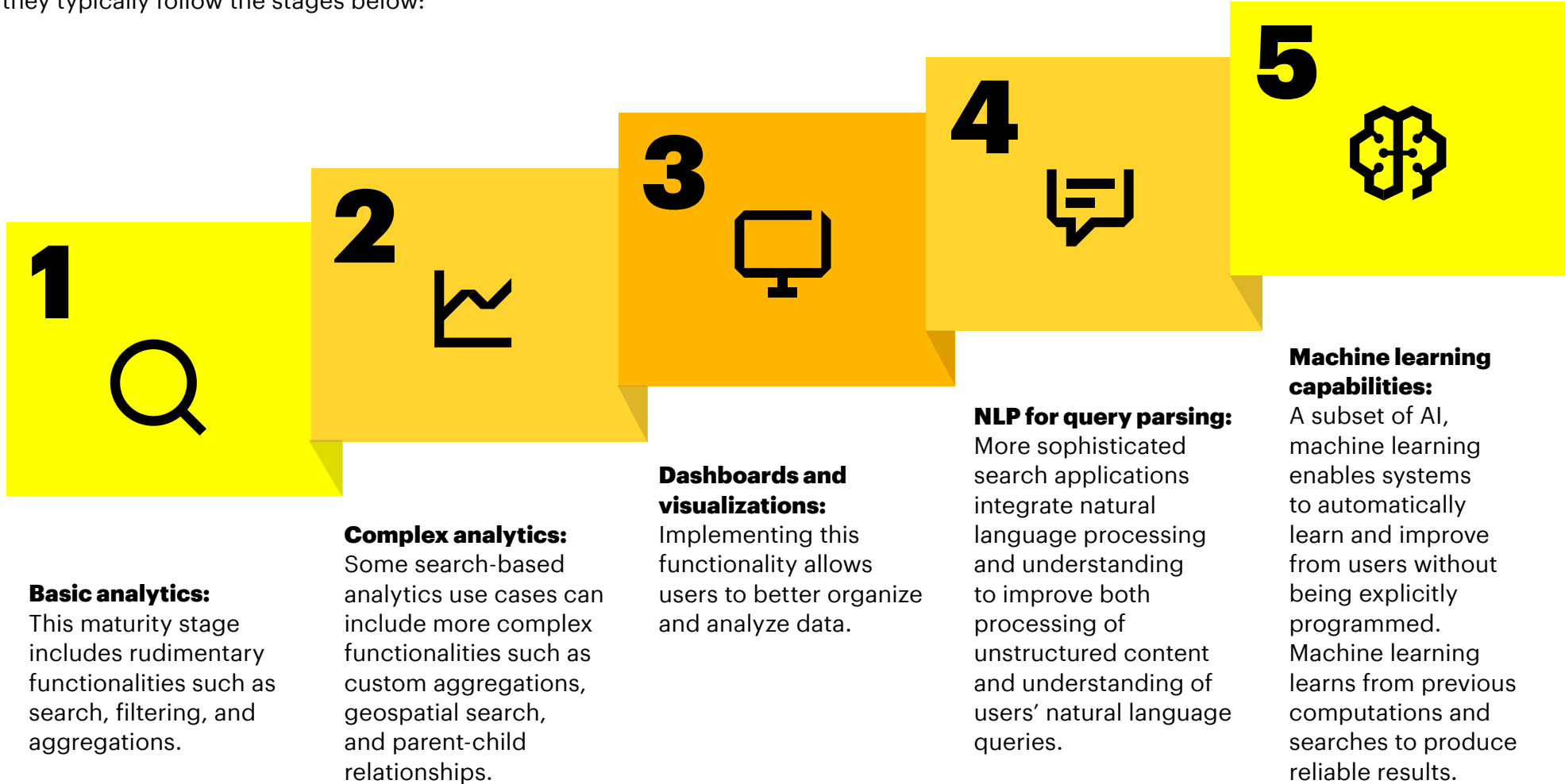## SEARCH-BASED ANALYTICS APPLICATIONS

Unlike classic document search, search-based analytics applications allow users to visualize and analyze their own data questions. This gives them better insights into the data (structured and unstructured). By drawing on text search input and results, it guides users to the information they're looking for.

Search-based analytics also allows organizations to understand what users are thinking or how they might search within an application (whether that's an e-commerce website or an intranet).

Some examples of search-based analytics use cases include:

- **Transactional and business records analytics**
  Collecting data from in-scope transactions and correlating that data to get insights into IT operations, customer experience, business outcomes, patient clinical records, etc.

- **Dashboards-based interactions**
  Compiling and presenting results from multiple data sources, graphically in real-time, for immediate trend spotting.

- **Log analytics**
  Assessing a recorded set of information from one or more events taken from a computer, network, or other IT ecosystem component. Organizations can use log analytics to identify patterns in user behavior, recognize problems, ensure compliance, etc.

As search-based analytics applications become more refined and sophisticated, they typically follow the stages below:

1

Basic analytics:
This maturity stage includes rudimentary functionalities such as search, filtering, and aggregations.

2

Complex analytics:
Some search-based analytics use cases can include more complex functionalities such as custom aggregations, geospatial search, and parent-child relationships.

3

Dashboards and visualizations:
Implementing this functionality allows users to better organize and analyze data.

4

NLP for query parsing:
More sophisticated search applications integrate natural language processing and understanding to improve both processing of unstructured content and understanding of users' natural language queries.

5

Machine learning capabilities:
A subset of AI, machine learning enables systems to automatically learn and improve from users without being explicitly programmed. Machine learning learns from previous computations and searches to produce reliable results.

CLOUD SEARCH PLATFORMS

PRICING MODELS AND KEY FEATURES

OPEN-SOURCE-BASED TECHNOLOGY

Elastic Cloud (Elasticsearch)

AWS Elasticsearch (Elasticsearch)

Azure Search (Lucene)

SearchStax (Solr and Elasticsearch)

PROPRIETARY TECHNOLOGY

AWS CloudSearch

SharePoint Online

Google Cloud Search

Coveo Cloud

Algolia

CLOUD SEARCH PLATFORM FEATURES

Below are some of the key cloud search features that we will be discussing throughout this white paper.

Search UI
An end-user search interface, provided out of the box, with full source code that can be customized for specific requirements.

Natural Language Understanding (NLU)
Adds a notion of semantic interpretation to the natural language text, in addition to processing tokens and patterns extraction.

Interactions API
The ability to ingest interactions from webpages even when they occur outside the search application (e.g. number of user views of a specific page directly in the content management system). This information can be used to improve the relevance of search results.

User personalization
Personalizing search for each user, displaying the most customized and relevant results.

Parent-child relationships functionality
Allows users to associate one document type with another in a "one-parent-to-many-children" relationship. The parent document can be updated without re-indexing the children. Child documents can be added, changed, or deleted without affecting either the parent or other children. Deletion of a parent can trigger the automatic deletion of all children underneath it.

Dashboard and visualization
Enables users to filter and analyze data via graphical dashboards, charts, graphs, and other visual elements.

Geospatial search
Ability to return search results based on a certain geographical location and/or cluster (aggregate) search results by map regions.

Complex aggregation and faceting
With an aggregate framework in place, a user can aggregate data based on a search query to build complex summaries of the data. By combining faceting, a search technique, users can narrow down search results by applying filters based on faceted classification of search results.
Examples: custom aggregations, significant terms aggregations or extends stats (standard deviation, variance, sum of squares) aggregations.

Index API
Adds or updates a typed JSON (or XML) document in a specific index, making it searchable.

Query API
For querying a data source for information about entities in that data source.

Security trimming
The ability to set document security policies that ensure users can only view documents that they have permission to access.

Group expansion
Identity service – allows for items to be properly filtered in the search results by comparing the item's ACLs to the user's ID and its groups. Identity service enables the search engine to identify the group(s) to which the user belongs.

Ingest transform API
A framework that allows the setup of processing pipelines for data transformations and enrichment of every document that gets ingested.

Connector and connector framework
Helps bring content from external sources into the search engine. Some search engines provide pre-built connectors to common business applications while others do not. In some cases, out-of-the-box connectors may be sufficient. In others, custom-built connectors are needed to ensure availability, data quality, and security. Custom connectors are typically built on a common connector framework that can be provided with the search engine.

ELASTIC CLOUD

OVERVIEW

Elastic Cloud is the official cloud search solution offered by Elastic – the company behind Elasticsearch. Fully hosted and managed, it offers a single console for provisioning, management, and monitoring of Elasticsearch applications, at any scale. Elastic Cloud provides the most functionality and comes the closest, from a features-parity perspective, to an on-premise ELK deployment.

PRICING MODEL

As a pay-as-you-go solution, Elastic Cloud's standard subscription pricing is determined by several factors: deployment size (RAM), choice of cloud provider (Amazon Web Services or Google Cloud Platform), number of availability zones, and deployment type, which affects the underlying infrastructure.

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Security trimming

- Group expansion – identity service

- View the complete feature list at: **www.elastic.co/cloud/elasticsearch-service/subscriptions**

> **The features available with this offering make it a strong candidate for search-based analytics use cases.**

AMAZON ELASTICSEARCH SERVICE

OVERVIEW

Amazon Elasticsearch Service is a fully managed service that makes it easy for you to deploy and operate the open source Elasticsearch version securely, at scale, and with zero down time. The service offers open source Elasticsearch APIs, managed Kibana, and integrations with Logstash and other AWS Services, enabling you to securely ingest data from any source and search, analyze, and visualize it in real time.

PRICING MODEL

Amazon Elasticsearch Service's pricing model is pay-as-you-use. There is no minimum fee or usage requirement. Factors impacting pricing include instance hours, Amazon EBS storage (optional), and data transfer.

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Integrations with open source tools, such as Kibana and Logstash for data ingestion and visualization

- Integrations with other AWS services such as Amazon Virtual Private Cloud (VPC), AWS Key Management Service (KMS), Amazon Kinesis Data Firehose, AWS Lambda, AWS Identity and Access Management Service (IAM), Amazon Cognito, and Amazon CloudWatch

- View the complete feature list at: **aws.amazon.com/elasticsearch-service/**

> **The features available with this offering make it a strong candidate for search-based analytics use cases.**

AMAZON CLOUDSEARCH

OVERVIEW

Amazon CloudSearch, available in the AWS Cloud and based on the Solr search engine, is a managed service that makes it simple and cost-effective to set up, manage, and scale search solutions for an enterprise website or application.

Supporting 34 languages, Amazon CloudSearch enables search across large collections of data, such as web pages, document files, forum posts, or product information. Basic search capabilities can be quickly added to the application without search expertise or concerns about hardware provisioning, setup, and maintenance. The solution can scale to meet enterprise requirements for data volumes and traffic fluctuations.

PRICING MODEL

Offered as a pay-as-you-use solution, there are no set-up fees or upfront commitments to begin using Amazon CloudSearch. The major portion of a typical domain's costs come from search instance usage. Customers are billed according to their monthly usage across these dimensions: search instances, document batch uploads, IndexDocuments requests, and data transfer.

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Free text, Boolean, and faceted search

- Autocomplete suggestions

- Customizable relevance ranking and query-time rank expressions

- Geospatial search

- View the complete feature list at: **aws.amazon.com/cloudsearch/**

> **The features available with this offering make it a strong candidate for both search-based analytics and classic document search use cases.**

SHAREPOINT ONLINE

OVERVIEW

The built-in search functionality in SharePoint Online helps users to find enterprise information quickly and easily. Providing both a classic and a modern search experience, SharePoint Online allows administrators to customize the search experiences for users. They can define searchable managed properties in the search schema, identify high-quality pages to improve relevance, manage query rules and result sources, remove individual results, and evaluate any changes by viewing reports about usage and search.

PRICING MODEL

The search functionality is included in SharePoint Online subscriptions, which are available on a per-user-per-month basis or annual commitment payment basis (for Office 365 Enterprise plans).

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Security trimming

- Group expansion – identity service

- Connectors (out-of-the-box connectors and a Software Development Kit for building custom ones)

- Search user interface (UI)

- View the complete feature list at: **docs.microsoft.com/en-us/office365/servicedescriptions/sharepoint-online-service-description/search**

> **The features available with this offering make it a strong candidate for classic document search use cases.**

AZURE SEARCH

OVERVIEW

Azure Search is a search-as-a-service cloud solution that gives developers APIs and tools for adding a rich search experience over private, heterogeneous content in web, mobile, and enterprise applications. Query execution is over a user-defined index.

PRICING MODEL

In addition to the free plan, there are additional plans offered as a pay-as-you-use solution in combinable units. As the volume of data or throughput needs of an application change, Azure Search can scale to meet these needs, and then scale back down to reduce costs. To get more performance, a customer can combine units to gain more queries per second, or a higher document count, or both. Units can also be combined to enable high availability or faster data ingestion.

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Ingest transform API

- Cognitive search

- Geo-search

- Data integration

- View the complete feature list at:
  docs.microsoft.com/en-us/azure/search/search-what-is-azure-search

The features available with this offering make it a strong candidate for search-based analytics use cases.

GOOGLE CLOUD SEARCH

OVERVIEW

Google Search Appliance is being replaced with a new cloud offering, Google Cloud Search (renewals of Google Search Appliance end in Spring 2019).

Google Cloud Search was initially rolled-out as a G Suite extension with the ability to search content from a number of Google services: Gmail, Drive, Calendar, Sites, People, and Groups. The general availability (GA) includes support for indexing and search over third-party enterprise content, such as:

- Internal websites and portals

- Content management systems

- File systems

- Relational databases

- Content hosted in enterprise applications

While Google Cloud Search seamlessly integrates third-party and G Suite content, it's important to note that enterprises do not need to have a G Suite license to use Cloud Search. Google Cloud Search customers can choose to index only third-party content and make it available within their enterprise search solutions.

PRICING MODEL

Details for the pricing model are not yet available.

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Security trimming

- Group expansion – identity service

- Connector (out-of-the-box connectors and a Software Development Kit for building custom ones)

- Interactions API

- View the complete feature list at:
  cloud.google.com/products/search/

> **The features with this offering make it a strong candidate for classic search use cases.**

COVEO CLOUD

OVERVIEW

The Coveo Cloud is a fully scalable native enterprise-class cloud service that can meet enterprise requirements for intelligent search. Coveo Cloud supports both microservices and scalability, allowing users to easily expand search solutions at the speed of their business.

Coveo Cloud enables organizations to:

- Easily and securely unify data from disparate cloud and on-premise sources across the IT ecosystem

- Create ubiquitous search experiences wherever people work

- Recommend the most relevant insights from everywhere directly into the context of customers and employees

PRICING MODEL

There are two plans available with a monthly subscription fee. One option is called ProCloud, a quick and easy way to incorporate search to a user's Sitecore site. The other is an enterprise cloud option, which offers access to the full range of features from Coveo Cloud Platform.

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- Security trimming

- Connector (out-of-the-box connectors and a Software Development Kit for building custom ones)

- View the complete feature list at: **www.coveo.com/en/technology**

The features with this offering make it a strong candidate for both classic search and search-based analytics use cases.

SEARCHSTAX

OVERVIEW

SearchStax is a fully managed offering, with a Solr-as-a-service or Elasticsearch-as-a-service option, that allows users to spend more time building search applications and less time managing, maintaining, and monitoring them. SearchStax provides a highly scalable cloud-agnostic solution that is fully managed by Solr and Elasticsearch experts.

PRICING MODEL

SearchStax offers three tiers of pricing plans with a monthly subscription fee:

- Dedicated node – ideal for prototyping, proofs of concept, or development environments

- Dedicated cluster – production-ready

- Enterprise – mission-critical, production-ready, SLA-guaranteed, customizable

SAMPLE FEATURE SNAPSHOT

- Indexing and query API

- View the complete feature list at:
  www.searchstax.com/cloud-search-service/solr/
  and
  www.searchstax.com/cloud-search-service/elasticsearch/

> **The features with this offering make it a strong candidate for search-based analytics use cases.**

ALGOLIA

OVERVIEW

Algolia provides the infrastructure, APIs, and easy-to-use documentation to help companies build, optimize, and personalize search and discovery for end-users. Algolia works with multiple platforms and devices and can scale for large enterprise deployments.

PRICING MODEL

Algolia offers monthly tiered packages, from free (qualifying open-source and non-profit projects can use Algolia at a discount or for free) to enterprise levels.

SAMPLE FEATURE SNAPSHOT

- Filters and facets

- Personalization

- View the complete feature list at:
  www.algolia.com/doc/guides/getting-started/what-is-algolia/#features

The features available with this offering make it a strong candidate for classic search use cases.

CLOUD SEARCH DEPLOYMENTS – CHALLENGES AND CONSIDERATIONS

1. NETWORK LATENCY

Systems that include components that are deployed to the cloud introduce the possibility of network latencies that will typically not apply when everything is deployed in a single data center.

Below are components applicable to search systems that could be deployed in the cloud:

- Content sources

- The search engine

- Users/groups directories

- Search application servers/APIs

The more dispersion there is (from a geographical standpoint) between these components, the higher the negative impacts on network speed.

As a mitigation option, staging repositories can be considered: data from more remote (slower) locations/content sources can initially be ingested in raw format into a staging repository that is close to the search engine. Data re-indexing into the search engine (to apply transformation rules, for example) can then be performed using the raw content from the staging repository.

Sometimes cloud providers don't clearly specify the whereabouts of the infrastructure supporting the search engine indexes. This might mean that users are only aware of the continent where search indexes are held (in which data is initially ingested or search queries directed) and nothing more. On the other hand, the providers that are less precise about their infrastructure's location will typically be the ones that sync data in data centers across the globe, in a way that is totally transparent to the user.

CLOUD SEARCH DEPLOYMENTS –
CHALLENGES AND CONSIDERATIONS

2. RATE LIMITING

Cloud services, generally, scale elastically and dynamically based on demand. However, there are situations where cloud providers need to prevent rogue or under-tested applications from overloading their infrastructure. Sometimes, there is also a need to reserve (temporarily or on a long-term basis) more hardware capacity for a customer if the number of requests increases substantially.

The most common mechanism through which cloud providers address these issues is rate limiting. In the case of cloud search services, the maximum number of documents ingested per minute (or second) and the maximum number of queries per minute (or second) are the most frequent rate limits.

Rate limiting raises some specific issues, including:

- Developers need to code around them, for example, implementing retry logic (typically exponential backoff)

- Sometimes rate limits are:
  - **non-deterministic**
    the same workloads can observe different rate-limiting behaviors, depending on the time they run

  - **not clearly or transparently stated by the cloud provider**
    this can sometimes make it difficult to plan around tests and/or significant variability in production workload volumes. Customers need to understand the cloud provider's rate-limiting policies to determine whether they will be affected. Equally important is understanding the process for increasing rate limits and how quickly it can be accomplished.

- In addition to rate limits for indexing and queries of documents, rate limits can also impact APIs used for the ingestion of users and groups with ACLs (Access Control Lists).

CLOUD SEARCH DEPLOYMENTS –
CHALLENGES AND CONSIDERATIONS

3. OTHER FACTORS

Also requiring careful evaluation before jumpstarting a cloud search implementation, typical issues include:

- Some cloud providers do not offer price breaks or discounts for non-production environments.

- As it becomes easier to spin-off resources in the cloud, careful usage monitoring is needed to avoid wasting resources.

- The pay-per-use pricing model used by cloud search providers may encourage better selection of the content to be indexed (high-quality content may ultimately increase search relevance).

- Privacy and compliance concerns regarding data storage in the cloud, especially for companies in highly-regulated sectors.

WHAT'S YOUR CLOUD SEARCH STRATEGY?

We've aimed to be fair and thorough in our evaluation of leading cloud search platforms. It's important to note that each of the solutions has its merits and could be best suited, depending on an organization's unique applications and situations. What's more? While these cloud search solutions provide the infrastructure and toolsets that can help accelerate application deployment, effective search requires expert implementation, tuning, and maintenance, whether it's an intranet search, a public website search, or a search-based analytics application.

Organizations need to ensure that their IT staff or implementation partners have the bandwidth and expertise required to conduct a thorough assessment to align each cloud search platform's features with business priorities. Because individual cloud search solutions may work better for certain use cases, it is key to identify the pros and cons for each use case in order to select the most suitable solution and plan the implementation strategy. After implementation, the search application will need to be continually evaluated and improved to achieve optimized performance and long-term ROI.

Whether organizations are at the initial phase of evaluating cloud search offerings, or have already identified their preferred solution for implementation, we can assist – from evaluating and selecting the best-suited solution to implementing, tuning, and managing the application.

OUR SERVICES INCLUDE

- Search application assessment

- Strategy recommendations

- Application design, implementation, and migration

- Content acquisition, processing, and enrichment

- Hosting, support, and managed services

Learn more at **www.searchtechnologies.com** or contact us at **info.stc@accenture.com** to discuss your search and analytics initiative.

REFERENCES

aws.amazon.com/cloudsearch/pricing/

aws.amazon.com/cloudsearch/

products.office.com/en-sg/sharepoint/compare-sharepoint-plans

www.elastic.co/cloud/elasticsearch-service/pricing

aws.amazon.com/elasticsearch-service/pricing/

aws.amazon.com/elasticsearch-service/?nc=sn&loc=1

www.algolia.com/pricing

docs.microsoft.com/en-us/azure/search/search-what-is-azure-search

www.youtube.com/watch?v=5vXygOJ6wro

cloud.google.com/products/search/

www.coveo.com/en/technology

www.searchstax.com/

www.searchstax.com/cloud-search-service/solr/

www.searchstax.com/pricing/

ABOUT SEARCH & CONTENT ANALYTICS

Search & Content Analytics, formerly Search Technologies, is part of Accenture Applied Intelligence.

We live in a data-driven world. But not everyone is making the most of their data. 80 percent of all data is unstructured – imagine the hidden insights trapped within unstructured enterprise content such as voice, images and emails.

At Search & Content Analytics, our mission is to help enterprises unlock the full value within their unstructured and structured data. We combine innovative technologies such as machine learning and natural language processing with search and big data analytics to transform the way people work.

Whether it's improving intranet and website search, monitoring internal communications to detect insider threats, helping recruiters match jobs to résumés, analyzing oil wellhead reports, or exploring molecular data, we bring comprehensive search and analytics services to clients across industries. Clients include organizations in e-commerce, media, healthcare, financial services, recruiting, manufacturing, and the government sector.

What knowledge and insights are trapped in your data? Let us help you find better answers.

ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions – underpinned by the world's largest delivery network – Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With more than 469,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives.

Visit us at **www.accenture.com**