# Komprise Deep Analytics
## A new approach to managing unstructured data

Finding just the right data across billions of files can be a time-consuming challenge. Komprise Deep Analytics enables you to quickly search and find data that fits your specific criteria across all your storage silos. Use the search results to create a company-wide metadata lake for data management planning and for uses like Big Data Analytics.

## What is Komprise Deep Analytics?

Businesses want to leverage data for uses, such as Big Data Analytics, or to run applications in the cloud. But to get to data science, you first need the right data. Studies show that 80% of the time in Big Data is spent finding the right data and getting it out of data centers.[1] Komprise can change all that.

### Easily Find Just the Right Data

Komprise Deep Analytics not only helps customers quickly find the data that fits their criteria across all their storage silos, but it also exports this dynamic data lake to any analytics application or destination they choose, such as Hadoop or Amazon Lambda. It creates a highly efficient, searchable distributed index of files with support for both standard metadata and custom metadata (tags). Customers can quickly find data that fits their criteria, regardless of where the data actually lives. The resulting data set can be operated on as a discrete entity.

[1] InfoWorld: https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html

 *Deep Analytics - A New Approach to Managing Unstructured Data*

Komprise Deep Analytics is an add-on component that can be deployed either in the cloud or on-premises. Komprise Deep Analytics Cloud is a fully managed solution that can be turned on with the click of a button and scaled on-demand, with no additional infrastructure or management needed.

## How Komprise Deep Analytics Fits

Industry research shows that over 75% of data is "cold" (infrequently accessed) within months of creation. Because 80% of data's cost is in its management, efficiently identifying and managing cold data yields significant savings. Unfortunately, most organizations store, replicate, and backup cold data (usually multiple times) the same as hot--on expensive Tier 1 storage. Given flat or shrinking IT budgets, a more efficient approach is critical.

Komprise Intelligent Data Management quickly identifies cold data across a customer's NAS storage. Users can then move it to more cost-effective storage options without any impact to users or applications, using Komprise Transparent Move Technology™. This approach ensures that only active "hot" data is kept on expensive Tier-1 primary storage.

**Create Virtual Metadata Lakes**

Deep Analytics extends the powerful data analytics that Komprise provides with a fully searchable index of all your data with more granularity. This enables users to easily find specific data sets among billions of files across storage silos. Users can create custom queries to find the data, then tag them with custom tags to assemble virtual metadata lakes. Now users can create complex queries, find and tag specific projects, identify types of cold data, find orphan data, search for data owned by specific users, and more. Big data analytics projects and AI/ML applications become much less time-intensive.
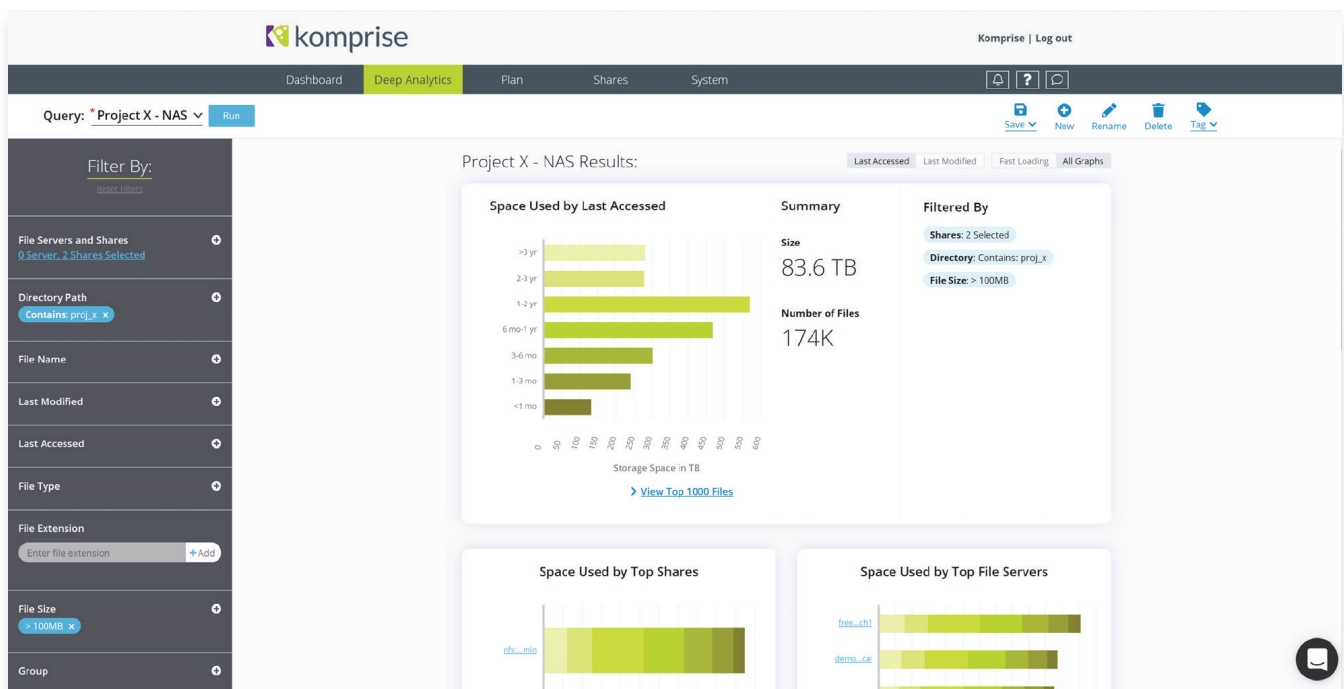


*Figure 1: Komprise Deep Analytics enables you to search across all of your unstructured data.*

 *Deep Analytics - A New Approach to Managing Unstructured Data*

## Use Cases for Deep Analytics

**Plan data management strategies**
Even with petabytes of data, within minutes of setting up Komprise you can see how data is growing, how it's being used, how much cold data you have, and your estimated ROI based on different scenarios that help you plan your data management strategies.

**Greater insight and control over data**
Deep Analytics enables users to perform deeper analysis at the file level, using custom queries and filters for specific data sets within or across shares. This enables storage IT administrators to gain deep insights into, and greater control over their enterprise's data.

**Create a searchable index**
Deep Analytics creates a searchable index of all the standard metadata as well as extended metadata (or tags) of data across storage. All data sources analyzed by Komprise are also indexed by Deep Analytics when it's enabled.

**Generate reports**
With Deep Analytics, users can download summaries and detailed reports or summarize results by file metadata parameters.

## Key Features

- Search and find specific data and files easily across storage

- Create custom queries and tags both via an intuitive UI and via API

- Visualize query results with interactive graphs and summaries with drill down

- Export graphs and download detailed reports

- Generate dynamic data lakes of data that fit specific criteria

- Export dynamic data lakes to other applications

*Deep Analytics - A New Approach to Managing Unstructured Data*

| Usage | Benefits |
|---|---|
| Identify all data belonging to a project, across file servers, shares, and directories | Enables project-based archiving |
| Identify the types of "cold" data, e.g., email, documents, videos, custom data files | Determine just how much cold data of various types you have to better plan your archiving strategy |
| Identify which owners are the biggest users of storage | Generate reports that help IT work with end users to get the needed approval to archive or confine data after project completion |
| Identify which departments are the biggest storage users | Apprise departments of their storage impact or chargeback based on usage |
| Identify data owned by specific users | Enables special handling of data if required |
| Identify the largest "cold" files | Get the most cost savings from archiving as quickly as possible |
| Identify "cold" orphan (zombie) data, by department or group | Shows how much storage they would free up if all that data were archived |
| Identify data with specific tags | Create dynamic data lakes based on specific criteria; this data set can be copied to applications like Big Data Analytics |

*Figure 2: Uses and benefits of Deep Analytics.*

**Typical Data Query Examples**

- Find top users in the engineering department who have the largest amount of data on file server "NAS92"

- Find out which departments are creating large video and archive files across all shares

- Find out which users in R&D have not accessed most of their data in the last two years

- Find data of users who are no longer with the company *(zombie data)*

 *Deep Analytics - A New Approach to Managing Unstructured Data*

## How Deep Analytics Works

Deep Analytics is a licensed feature that must be specified in your license. A setting easily enables or disables Deep Analytics.

When shares are added to Komprise and enabled, Komprise starts to rapidly aggregate analytics information across these shares producing results in the Plan page. If Deep Analytics is enabled, Komprise builds the Deep Analytics index in the background. Because it examines and indexes every file's metadata, Deep Analytics runs take longer run than the regular faster analysis.

After a Deep Analytics run has begun on a share, queries can be made against data (files) on that share. Query results, however, will be partial until the Deep Analytics run has completed on the share. Subsequent Deep Analytics runs will occur on each enabled share after a default delay interval of 30 days.

Deep Analytics keeps an index of all the standard and extended metadata; no file contents are stored. Indexed metadata includes:

- File name
- File parent directory
- File size
- File extension
- File type *(directory, file, symbolic link, Komprise file link)*
- File creation date
- File last modified date
- File last accessed date
- Owner ID *(uid/sid)*
- Owner name
- Group ID *(uid/gid)*
- Group name
- Tags *(custom Komprise metadata)*

**NOTE:** *No file content is ever read or stored by Komprise.*

Query: *Project X - NAS ∨    Run

Filter By:

Reset Filters

**File Servers and Shares**    ⊕
0 Server, 2 Shares Selected

**Directory Path**    ⊕
Contains: proj_x ×

**File Name**    ⊕

**Last Modified**    ⊕

**Last Accessed**    ⊕

**File Type**    ⊕

**File Extension**

Enter file extension    +Add

**File Size**    ⊕
> 100MB ×

**Group**    ⊕

*Figure 3: Run a search in Deep Analytics for files with the Project X tag and operate on the data set.*

*Deep Analytics - A New Approach to Managing Unstructured Data*

## Deep Analytics Tagging

Komprise Deep Analytics enables data to be tagged, which can be used in queries.

Tagging makes it easy to organize and find data based on extended metadata attributes beyond the standard file metadata.
This can be useful in many ways:

1. Grouping together data that easily satisfies multiple criteria
2. Creating tags outside of Komprise (e.g., tagging data at the source when you know more about the data), and leveraging the tags to search and find associated data easily
3. Managing and finding data by these tags rather than relying on just standard file metadata

Example: Say you want to run an operation on data belonging to either Project X or Project Y. You'd first run a search in Komprise for any files that belong to Project X, and tag them with Project X. Then you'd do the same for files related to Project Y. Then you'd run another search in Deep Analytics for files with either Project X or Project Y tags and operate on that data set.

Tags are key value pairs. You can specify a key and a value of that key. For instance, you could have a key = "Movie" and values can be "Shrek", "Toy Story" and so on. You can then search for all files with the key "Movie" or find files specific to the movie Shrek.

Tags can be set via the UI or via Komprise APIs for Deep Analytics. Tags can also be set via API outside of Komprise.

## Deep Analytics API

All Deep Analytics functionality is accessible through an API. The API enables capabilities, including:

- Creating, saving, renaming, deleting, and running queries
- Setting query filters on:
  - File servers and shares
  - Directory path
  - File name
  - Last modified time
  - Last access time
  - File type
  - File extension
  - File size
  - Group
  - Owner
  - Tag
- Creating tags *(keys and values)*
- Applying and removing tags from files and query results
- Monitoring tagging tasks
- Retrieving the set of all tags *(keys and values)* created
- Summary of query results
- Top 5,000 files of a query result
- Retrieving summary of query result by:
  - Top shares
  - Top file servers
  - Top owners
  - Top groups
  - File types
  - File extensions
  - File sizes
  - And retrieving the top 5,000 files for any of these

*Deep Analytics - A New Approach to Managing Unstructured Data*

## Deployment Architecture

Komprise Deep Analytics can be run in the cloud or on-premises. Deep Analytics uses secure, cloud-based services, including metadata indexes, and a powerful, open-source analytics and search engine. No Deep Analytics components need to be deployed on-premises.

The Komprise Observers used in the Komprise Intelligent Data Management solution automatically send file metadata into the secure cloud index service when Deep Analytics is enabled.

When deployed in the cloud, Komprise manages all the analytics components in the cloud and the customer needs to only deploy on-premises Observers. This deployment is shown in Figure 4 below.

## Advantages of a Cloud-based Deployment

• Fast, easy deployment, SaaS model: only standard Komprise Observers need to be deployed on-premises

• Enables use of thin Observer resources: simpler and cost-effective provisioning, maintenance, and growth accommodation

• Handles elastic growth and shrinking data sizes. Add more data to analyze or remove some; Komprise automatically adjusts

• Enables transparent upgrades: only cloud-based components need to be upgraded

When Deep Analytics is deployed on-premises, the customer must deploy and scale the appropriate hardware for all components, as well as the Director, Observers, Analytics Services, and Search Cluster. See Figure 5.
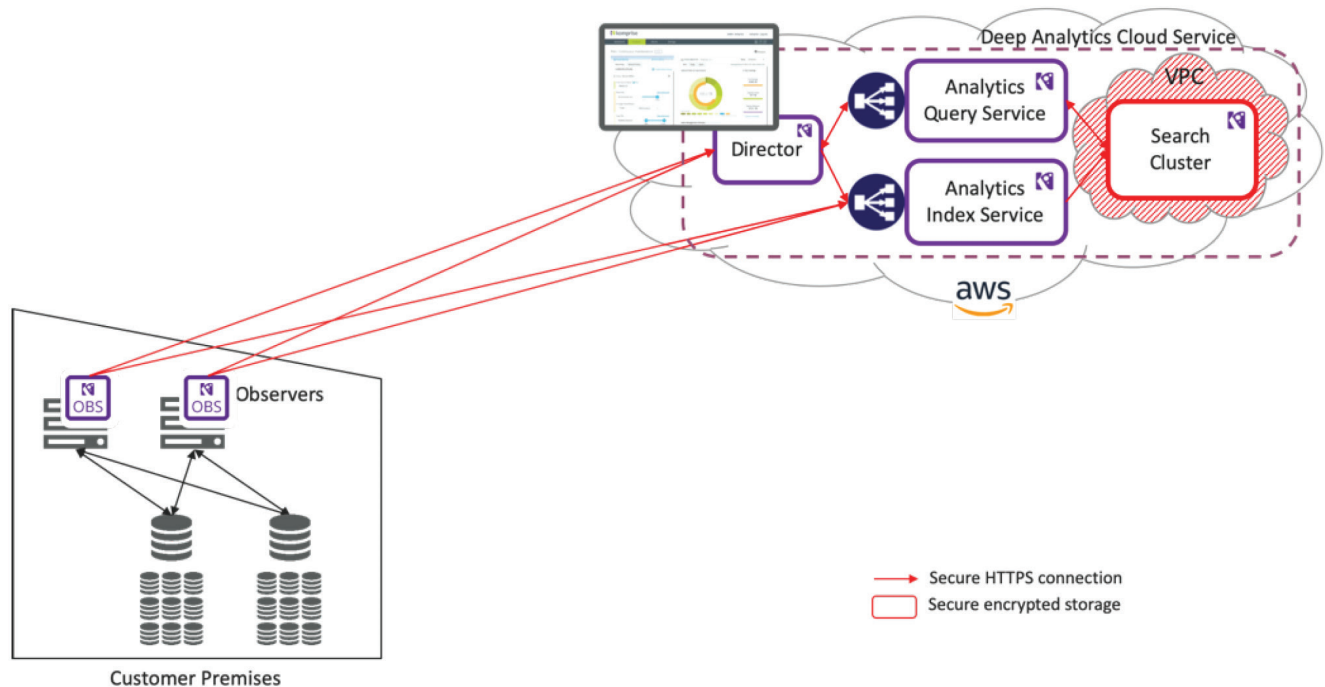


*Figure 4: When deployed in the cloud, Komprise manages all the analytics components in the cloud. The customer only needs to deploy on-premises Observers.*

 *Deep Analytics - A New Approach to Managing Unstructured Data*

Komprise provides guidelines for requirements of server sizing and software, ensuring a secure, scalable, and high-performance system deployment.
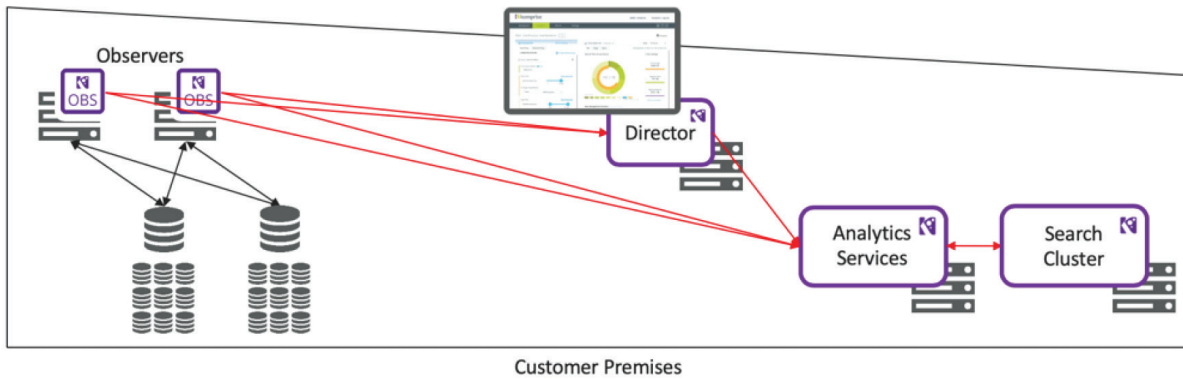


*Figure 5: Komprise Deep Analytics — On-Premises Deployment*

|  | Cloud | On-Premises |
|---|---|---|
| **Setup** | Simple | Requires VM Provisioning |
| **Scaling** | Automated | Requires additional VMs |
| **Management** | Komprise mangaged | Requires onsite management with remote Komprise support of memory, resources |
| **Security** | Encrypted, only metadata in the cloud | Fully stored on-premises |

Deep Analytics running in the cloud is convenient, flexible, worry-free, and secure. It enables the analysis of data on-demand with the click of a button without any headaches. On premises Deep Analytics is available for those with strict security requirements.

## Next Steps
To see the benefits of Deep Analytics in your environment, contact **sales@komprise.com**.



**Komprise, Inc.**
1901 S. Bascom Ave. Suite 400
Campbell, CA 95008
United States

For more information:
Call: 1-888-995-0290
Email: **info@komprise.com**
Visit: **komprise.com**

For media requests email:
**marketing@komprise.com**

WP-NewApproach-0220-a

*Deep Analytics - A New Approach to Managing Unstructured Data*