Retain, Search and Analyze more Data

elysium
ANALYTICS

Retain more data for 10% of **Amazon Elasticsearch Ultrawarm**

Elysium Analytics' disruptive ML-based log analysis solution is a true cloud-scale solution built on Snowflake with a 90% cost reduction for hot log data storage, search, analytics, and visualization compared to Amazon Elasticsearch Ultrawarm. Our solution carries close-to-zero operational overhead, gives you complete control over your data, and scales seamlessly.

Machine-generated data is growing exponentially in volume and variety. This data is mission-critical for identifying operations issues in all software applications and infrastructure and to detect first-seen security risk. The ability to collect, process, and analyze this vast amount of data in real time, a flexible and scalable cloud service like Elysium Analytics is required. All agile IT and security operators recognize the benefits from having real time access to this data and understand that the more data they have available to them, the better they are equipped to do their job. However, storing all this data in hot storage has been prohibitively expensive, at rates up to $700 per TB per month. To find the balance between how much data to keep in cold storage, warm storage, and eventually cold storage is challenging. On top of that, to configure an elaborate architecture consisting of various types of nodes and find the balance between disk storage, memory, and compute, is often a full-time job. This complexity is not just due to of the lack of scalability of legacy applications but also driven by the strong need to mitigate the ballooning cost caused by a seemingly never-ending wave of data and rapidly escalating storage cost.

Elysium Analytics, running on Snowflake, the leading cloud-scale data warehouse, has solved this problem. You can now store all your data for as long as you want for as little as 1% of the cost of hot data storage on a solution like Elasticsearch. Even more modern iterations of Elasticsearch, such as AWS Ultrawarm, will end up as much as 10 times more expensive even before taking operational overhead into consideration.

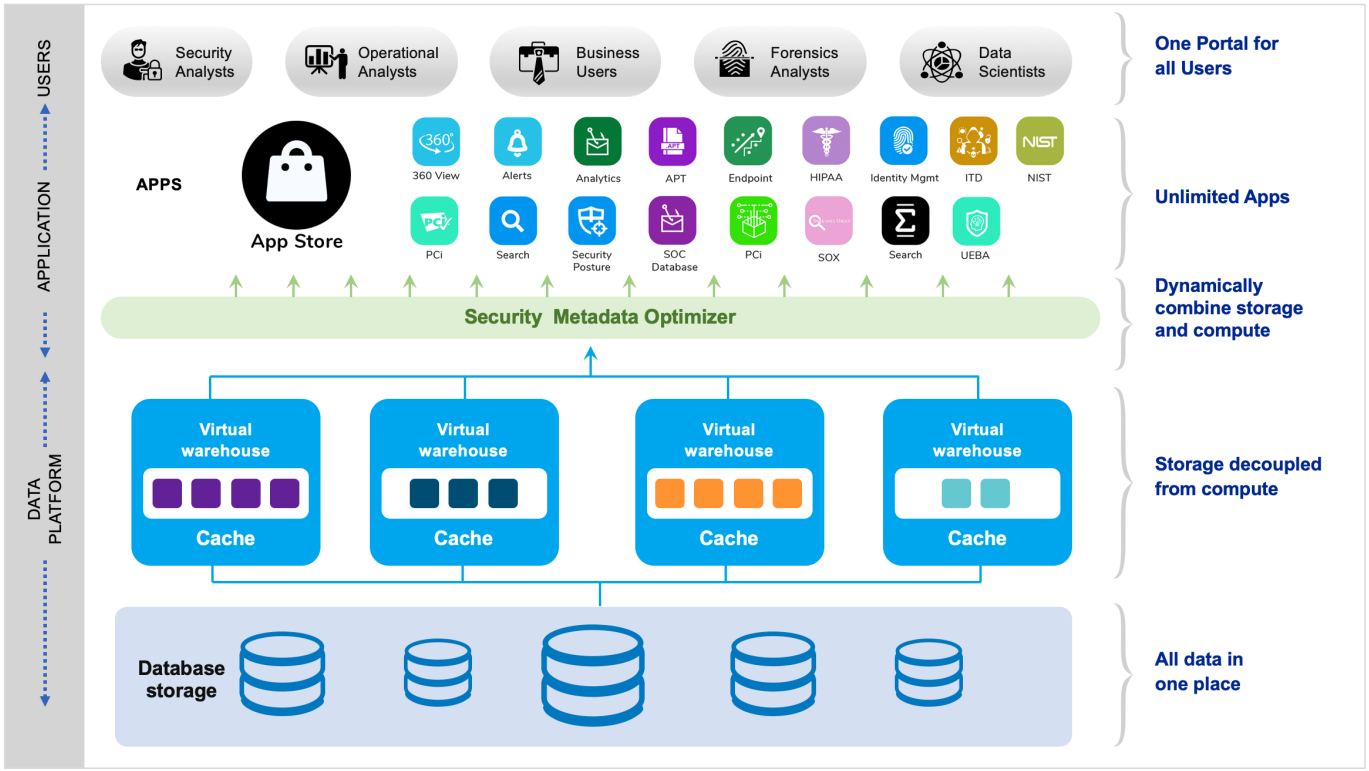
Best of all, the interactive analytics experience remains the same, or even better, with both Kibana and Looker bundled in with the service at no additional cost.

Elysium Analytics provides full text search and a number of operational and security dashboards leveraging machine-learning analytics out of the box. Additionally, have the flexibility to customize all included dashboards and the ability for you to create your own dashboards and machine learning-based models.

It is well known that recent log data is valuable and paying for hot storage, which offers the best search performance, makes perfect sense. Log data from the past few months or years is valuable as well but the decision on where to draw the line on how much you are willing to pay for warm storage is often directed by budget. As a result, most corporations have a policy of aging out data from hot to warm after one week to three months, then move it to warm storage for up to a year to finally move it to cold archive storage for as long as the data is required from a compliance perspective. Only if there is an audit or an urgent threat hunting exercise requirement will the data be brought back to hot storage, reindexed, and eventually be ready for search and analysis. The process of transferring and indexing the data from cold storage alone can take days, if not weeks, and results in additional expenses typically not budgeted for. There is no longer a need to make this compromise with hot data stored at less than the cost of frozen archive data.

With all your data stored on AWS, Azure, or GCP, you retain full control and ownership of your data which is protected by Snowflake's security framework. If you for some reason choose to discontinue our service, you merely remove our access privileges and stop using our application. Since Elysium Analytics is licensed on a usage bases, we will only charge you what you use and there are no long-term commitments required. The other, and often more important, aspect of our billing model is the ability to scale up compute capacity to whatever level you may require getting an urgent task done. There is no concurrency limit and you have an almost unlimited pool of compute power available at seconds' notice.

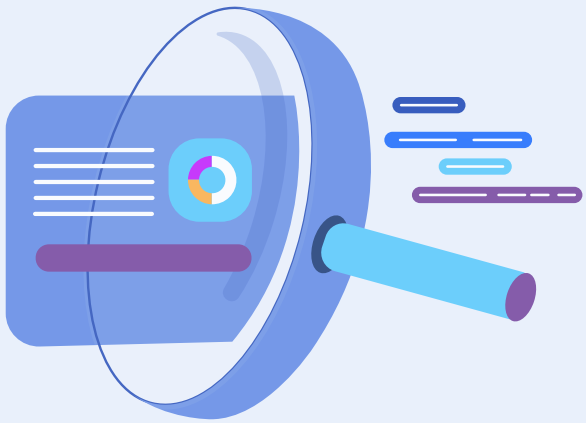
Search and ML-based analytics **on Snowflake**



How it **works**

Traditional storage has limitations in that an Elasticsearch cluster is scaled out by adding dense-storage instances which require operating system overhead, disk watermarks, and index replicas. You also end up paying for the storage and compute you provision, not what you actually use.

Elysium Analytics, on the other hand, is a radically different solution. Running on Snowflake, we load from AWS, Azure, or GCP storage containers or load directly from cloud apps and on-premises sources to Snowflake is a radically different solution. Running on Snowflake, we load your data either directly in AWS, Azure, GCP, or directly to Snowflake. We enrich your data and apply our open data model with lazy normalization for full text search and analytics across all sources without changing the raw data, providing you with a single source of truth. We set up virtual warehouses on Snowflake where you have full access to all the compute you require for all our apps on a pay-as-you-go basis. Elysium Analytics achieves significantly higher data compression rates than any other solution, typically from 10x to 15x, with no need for replicas and no overhead leading to an effective storage cost for hot data of less than $2/TB per month for ingested data. Additionally, Elysium Analytics perfor mance is generally similar to or better than traditional warm storage with the added benefit of retaining the ability to load and query with unlimited scale and zero operational overhead from adding nodes, migrating indexes, and re-adjusting shards.



Elysium **Analytics cost**

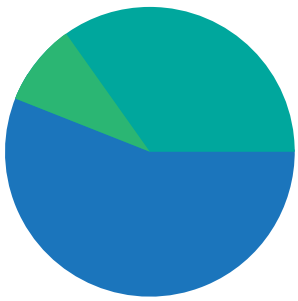
For a pricing example, consider a case where there is a need to ingest 2TB of log data daily and retain the data for one year which adds up to 720TB ingested data. With Elysium Analytics, we will provide collection, parsing and loading (ELT) as a service. The loading of your data will incur a nominal consumption-based charge and you will always only pay for compute used for the process. If you load less data one day, we will charge you less, if you have more data to load for a period of time, you will not be asked to upgrade your license but rather pay incrementally more for the increased loading time. Your data is stored in AWS, Azure, or GCP and you pay a flat rate of $23/TB/month for compressed data. When you need to search or query on the data, we will spin up the compute instance you need and automatically shut it down after use based on your timeout setting and only charge you for the compute consumed down to the second. In this example, you will have a XS instance allocated for queries 8 hours per day.

Loading compute cost:	$2,344
Storage compressed 10X:	$1,656
Query compute cost:	$2,160
Total monthly cost:	**$6,160**

AWS **Ultrawarm cost**

For comparison, consider 36 AWS Ultrawarm1.large nodes. Like with all nodes, you pay an hourly rate for each. In this case, that rate is $2.68 per hour, or $1,956 per month, per node. Together, these nodes can address up to 720TB of storage on S3 at $0.024 per GB/month. The total for a month of nodes and storage is approximately $87,696. In addition, you will need three master nodes at an annual cost of $4,231 each for a grand total of $88,753 per month.

3 master nodes r5.2xl:	$1,057
37 Ultrawarm nodes:	$70,416
720TB managed storage:	$17,280
Total monthly cost:	**$88,753**



Traditional **Elasticsearch cost**

The traditional Elasticsearch configuration would require 48 i3.16xlarge data nodes to hit just shy of 720TB of storage. But not all of those 720TB are usable. After factoring in overhead, you're left with 75% of the original space. Replicas then cut that number in half. To reach approximately 720TB of usable storage for primary shards, you actually need 132 i3.16xlarge data nodes, each of which is 45,478 annually each when paid up front for the year. Whether you use that space or not, the grand total for a month is $500,280. In addition, you will need three 5r.4xlarge master nodes at an annual cost of $8,467 for a grand total of $502,513 per month.

3 master nodes r5.2xl:	$ 2,117
132 data nodes i3.16xl:	$ 502,396
Total monthly cost:	**$ 504,513**

A word about query **compute cost**

One major difference between Elysium Analytics' solution architecture and Elasticssearch, and other log analysis vendors in general, is that Elysium Analytics is running on Snowflake's Data Cloud with separation of storage and compute. This is a major breakthrough as users can scale storage and compute completely independently and transparently as needed. If you need to load more data and store it longer, you will simply be billed accordingly at the next billing cycle and there is no need to provision additional nodes or storage capacity before you run out of space. Similarly, if you are not making any queries, you will not be billed for compute usage but when you need "all hands on deck" and you have multiple analysts accessing the data, you will have as much compute capacity as you'll need with no concurrency limitations. Contrast this to the legacy architectures that offer a fixed amount of compute where you most days are over-provisioned, and you overpay, while at other times you are grossly under-provisioned and queries are stacking up.

The question is, however; how does this translate to cost and savings? It is somewhat challenging to compare the two platforms; one that ties compute to storage and one that scales storage and compute independently. With Elasticsearch, a decision about how much compute goes with the storage is made by someone else; you want more storage, you have to procure more nodes regardless of whether you need the compute or not. However, most use cases do not see a linear demand for compute as you scale out storage. This is especially true when it comes to use cases where the end-user wants to retain their data hot for a longer period of time, often up to several years, but the number of queries remain the same as before. If you do not need the extra compute the storage increase brings with it, you still need to pay for this under-utilized resource.

Clearly, there is a break-even point where it will be more cost effective to have dedicated resources 24x7 at a relatively lower rate than paying a higher rate on an on-demand basis. In the example above, we calculated query cost of $2,169 for 8 hours per day for a XS instance on Snowflake. The difference between $6,160 per month and $88,753 per month for AWS Elasticsearch with Ultrawarm, which is AWS' low cost option for Elasticsearch, is $82,593 per month. For $82,593 per month, you can consume 9,177 credits, including charges from both Snowflake and Elysium Analytics, which translates to 305 credits per day, or as much as 305 query-hours per day. At a query volume below this level, you will incur a significantly lower cost.

A word about query **operational cost**

Everyone who has implemented and managed an Elasticsearch implementation are acutely aware of the operational overhead burden that comes with this solution. Offerings of Elasticseach as a SaaS solution have reduced the amount of operational overhead from installing the cluster but you still need to configure the solution very carefully and have a thorough understanding of the architecture in order to optimize the nodes' compute, memory, and storage. Ongoing monitoring and performance tuning, keeping up with data growth, and overseeing security audits add a sustainable effort that can amount to a fractional fulltime employee to several employees.

Get **started**

Although the Elysium Analytics solution is radically different from Elasticsearch with a different underlying architecture and numerous innovations, the user experience remains the same with Kibana. To migrate data from Elasticsearch to Elysium Analytics is as simple as installing a plugin and adding a couple of lines to the current logstash pipeline, config.xml, which will load the data to Elysium Analytics. This way, there is no disruption to your current Elasticsearch implementation and you will be free to discontinue loading to Elasticsearch once you have settled in on the Elysium Analytics solution.

You also have the option to use our cloud-based ELT service where we handle collection directly from the sources. If your data is already is on AWS S3, Azure, or GCP, we simply load your data into your Snowflake virtual data warehouses and apply our open data model schema.

Elysium Analytics is currently available in the US, Europe and Asia https://docs.snowflake.com/en/user-guide/intro-regions.html. To get stared and see if Elysium Analytics fits your use case, contact us on sales@elysiumanalytics.ai.

About **Elysium Analytics**

Elysium Analytics, running on Snowflake, the leading cloud-scale data warehouse, has solved the problem of collecting, processing, and analyzing growing amounts of log data in real time. You can now store all your data hot for as long as you want for as little as 1% of the cost of hot data storage on a solution like Elasticsearch. Even more modern iterations of Elasticsearch, such as AWS Ultrawarm, will end up as much as 10 times more expensive, even before taking operational overhead into consideration. Best of all, the interactive analytics experience remains the same as on Elasticsearch, or even better, with both Kibana and Looker bundled in with the service at no additional cost. Elysium Analytics provides full text search, operational and security dashboards leveraging machine-learning analytics out of the box with the flexibility for you to not only customize all included dashboards but also the ability for you to create your own dashboards and machine learning-based models.

About **Snowflake**

Snowflake is a fully relational ANSI SQL data warehouse that was built from the ground up for the cloud. Its architecture separates compute from storage so that you can scale up and down on the fly, without delay or disruption, even while queries are running. You get the performance you need exactly when you need it, and you only pay for the compute you use. Snowflake currently runs on Amazon Web Services, Google Cloud Services, and Microsoft Azure.

Snowflake is a fully columnar database with vectorized execution, making it capable of addressing even the most demanding analytic workloads. Snowflake's adaptive optimization ensures queries automatically get the best performance possible, with no indexes, distribution keys, or tuning parameters to manage.

Snowflake can support unlimited concurrency with its unique multi-cluster, shared data architecture. This allows multiple compute clusters to operate simultaneously on the same data without degrading performance. Snowflake can even scale automatically to handle varying concurrency demands with its multi-cluster virtual warehouse feature, transparently adding compute resources during peak load periods and scaling down when loads subside.

elysium ANALYTICS

Running on: snowflake

Elysium Analytics, Inc. 2550 Great America Way, Santa Clara, CA 95054

elysiumanalytics.ai

Phone: +1 (669) 209-0801

info@elysiumanalytics.ai