

IT@Intel: Granulate™ Optimizes Memory and CPU Utilization in Intel IT's Cloudera Platform

Improving big data platform efficiency is an ongoing effort for Intel IT. In initial testing, Granulate™ reduced both memory and CPU utilization in our production environment via a light-weight agent, with no code changes or developer resources.

In April 2023, Intel IT performed a proof-of-concept (POC) with Granulate software in our Cloudera platform. During a week of testing in our production environment, Granulate optimized more than 1000 unique jobs – equivalent to approximately 42,000 executions of Apache Spark jobs/applications. We saw an average 25% reduction in memory utilization and an average 44% reduction in CPU utilization.¹ This provides us with **two primary benefits**:

1

We can run more jobs concurrently without adding memory or compute infrastructure.

2

As our platform grows, we can avoid capital and operational expenses in the future.

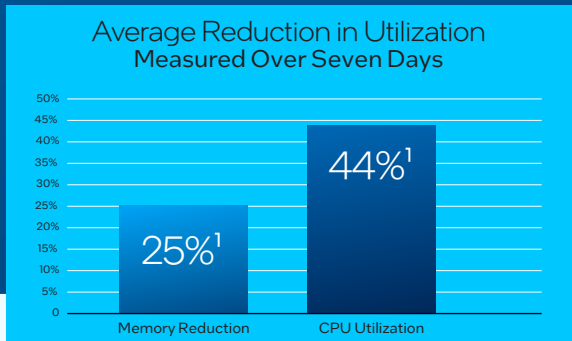


Figure 1: Granulate substantially reduced both memory and CPU utilization.

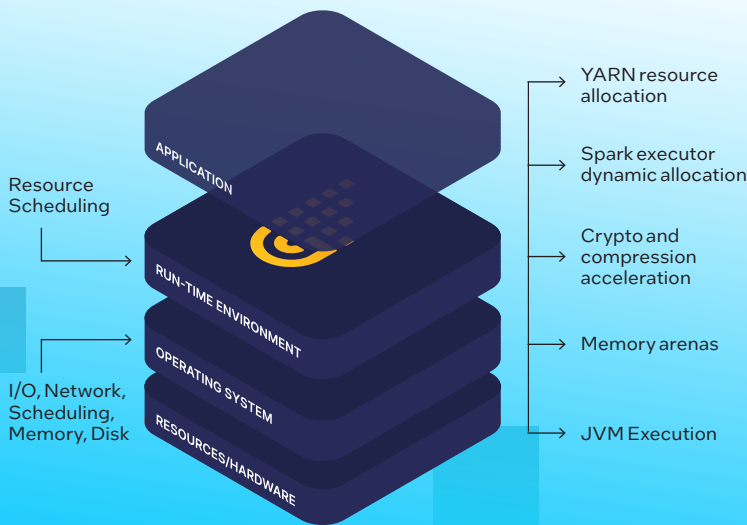


Figure 2: Five key Granulate optimizations improve platform efficiency.

How Granulate Works:

Granulate is an autonomous workload optimization solution. It continuously profiles and learns the specific characteristics of each job and then optimizes resource scheduling to reduce inefficiencies and improve throughput of big data workloads – both on-premises and in the cloud. Granulate optimizes resources for many run-time environments, including Java, Python, Scala, Go, Node, Kotlin, and Clojure. The optimizations are enabled at both the run-time and operating system levels. Figure 2 shows the five key optimizations for big data platforms.

Intel IT Enterprise Data Platform

Intel IT’s Enterprise Data Platform supports real-time and batch ingestion of structured, semi-structured, and unstructured data. Our data warehouses, together with our data lake, provide data to our visualization and data science platforms. These platforms are used by a variety of dashboards, application frameworks, and custom applications.

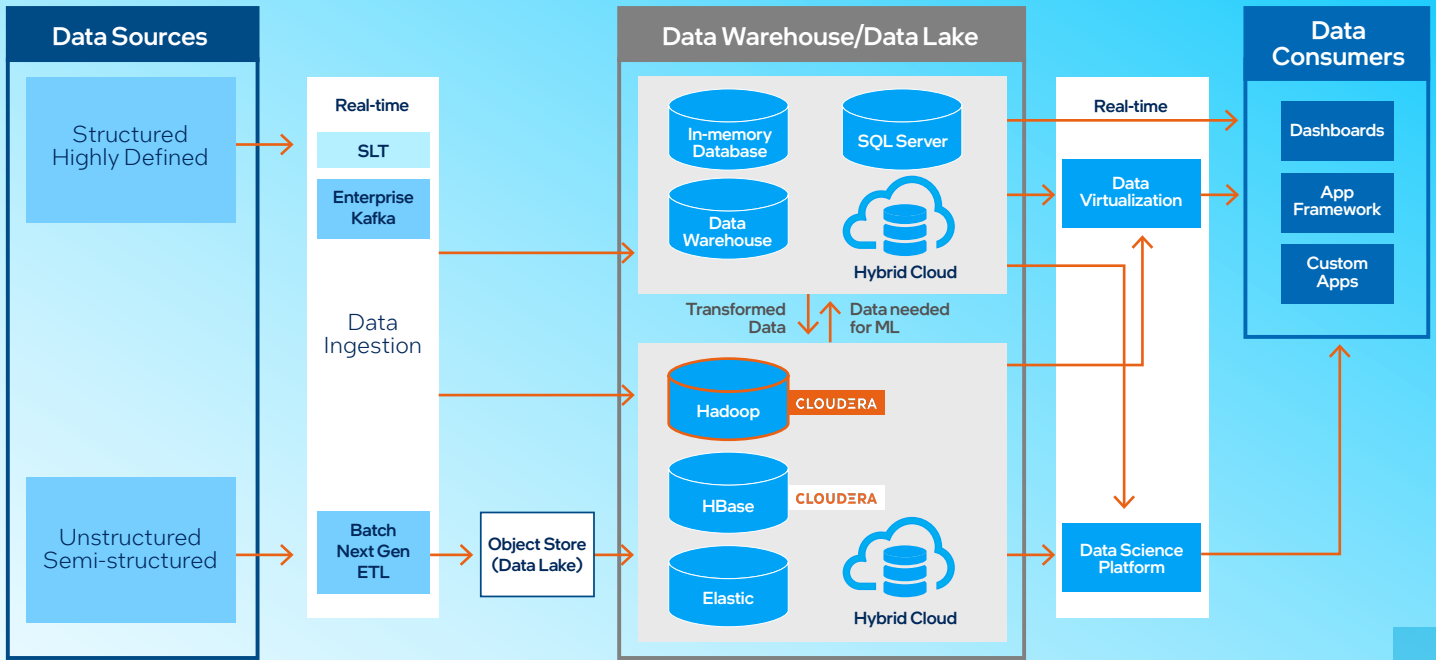


Figure 3: Intel IT’s Cloudera platform supports HPC, supply chain, finance, and advanced analytics workloads.

As a large IT shop, we continuously look for ways to reduce both operational expenses and infrastructure purchases, while simultaneously improving the user experience. Unfortunately, workload optimization can be disruptive and time consuming. Most optimization efforts require developers to take time away from business applications and product development to rewrite code – and sometimes, the tuning can lead to sub-optimal performance.

Given Granulate’s capability to optimize big data platforms with no hands-on development, we decided to test it in our Cloudera platform. First, an assessor from our Information Security team performed a risk assessment of Granulate. He reviewed several assessments by independent third parties, including Granulate’s SOC 2 Type II Report, the results of Granulate’s penetration testing and remediation actions, and SecurityScorecard’s report. In that report, Granulate earned an “A” rating of 91. The Intel Information Security team assessed that Granulate met or exceeded the pertinent standards.

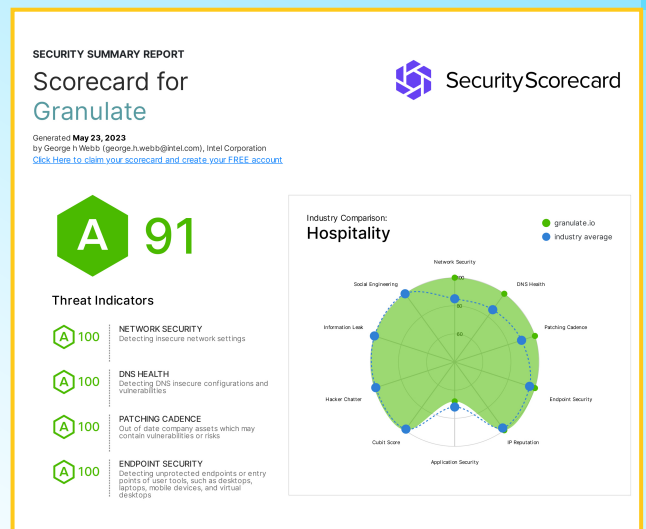


Figure 4: SecurityScorecard is a well-known third-party cybersecurity risk assessment provider.

We initially deployed the Granulate agent in our Quality Assurance environment to validate no disruption to business operations. The agent ran for about two to three weeks, learning about the platform’s Spark jobs. Then we deployed the agent to all the nodes in our production environment. Again, the agent learned for about two to three weeks. The agent identified optimization opportunities at the job level behind the scenes, with no human involvement. We then activated the Granulate agent and moved to the benchmark phase.



Figure 5: Granulate’s user interface shows the average memory and CPU overcommitment from April 21 through April 27, 2023

We gathered metrics for seven days. During this time, Granulate dynamically optimized more than 1000 jobs – equivalent to approximately 42,000 executed Spark jobs/applications. The results were compelling: an average 25% reduction in memory utilization and 44% reduction in CPU utilization. Now we can run more jobs concurrently without adding memory or compute infrastructure, which will help us avoid capital and operational expenses in the future.

Next Steps

Given the outstanding results we see with Granulate, we will continue to run it in our Cloudera platform. In addition, we started a new POC running Granulate in one of our public cloud environments. We are also exploring the use of Granulate in streaming platforms such as Apache Kafka, and log analytics platforms like Elasticsearch.

To learn more about Granulate, please contact your Intel account executive.



DISCLAIMERS:

1. Granulate identified and optimized an average of 25% reduction in memory utilization, and an average of 44% reduction in CPU utilization, when running on Intel IT’s Cloudera production platform. The test ran from 4/21/23 to 4/27/23, comparing memory and compute efficiency before and after enabling Granulate. Intel IT’s Cloudera platform configurations: 151 nodes with Intel® Xeon® processor 6420 and Xeon processor 6336Y. Total platform memory: 161 TB. Total platform storage: 14.5 PB. Software: Cloudera version 7.1.7, Java 8 and Java 11; Red Hat 7.9 and Granulate version 4.5.0. The utilization results are the average of over 1000 jobs, equivalent to approximately 42,000 executed Spark job/applications, optimized by Granulate. Testing performed by Intel.

Performance varies by use, configuration, and other factors. Learn more at www.Intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in the disclaimer above and may not reflect all publicly available updates. See disclaimer above for configuration details. No computer system can be absolutely secure.

Granulate technologies may require enabled hardware, software, or service activation.

© Intel Corporation 2023. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.