

Spark + S3 Select on MinIO - A Performance Story

The world of object storage isn't just growing, it is changing at the same time - moving past the traditional use cases of disaster recovery and archiving and into more dynamic use cases that emphasize analytics and machine learning. As this transition occurs, SQL becomes increasingly critical. As the lingua franca of analytics, SQL is everywhere and has many "dialects." Object storage is no different but incorporated SQL somewhat later than other technologies - and with a potentially larger impact.

The recent extension of the S3 API to include SQL query capabilities (S3 Select) alters how machine learning, stream processing and advanced analytics are performed on large, peta-scale datasets. With S3 Select, users can execute queries directly on their objects, returning just the relevant subset, instead of having to download the whole object - significantly more efficient than the regular method of retrieving the entire object store.

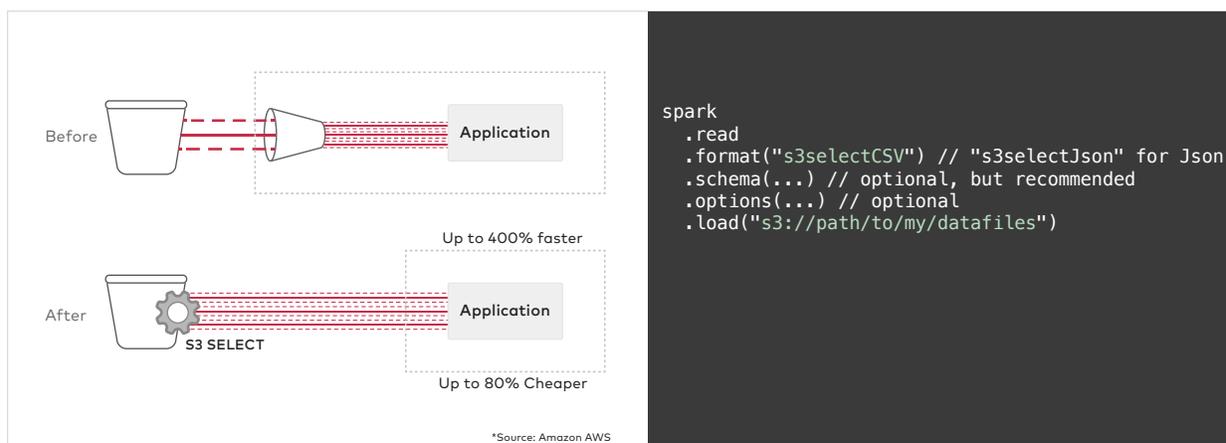
MinIO's implementation of the S3 Select API matches the native features, while offering better resource utilization when it comes to executing Spark jobs. These advancements deliver orders of magnitude performance improvements across a range of frequently used queries.

MinIO has, like all of its software, open sourced this code. It can be found here for further inspection: <https://github.com/minio/spark-select>

MinIO Spark-Select

With the MinIO S3 Select API, applications can offload query jobs to the MinIO server itself, resulting in significant speedups for the analytic workflow.

By pushing down possible queries to MinIO, and then loading only the relevant subset of the object to memory for further analysis, Spark SQL runs faster, consumes less network resources, uses less compute/memory resources and allows more Spark jobs to be run concurrently.



The Spark-Select project works as a Spark data source, implemented via DataFrame interface. At a very high level, Spark-Select works by converting incoming filters into SQL S3 Select statements. It then sends these queries to MinIO. As MinIO responds with data subset based on Select query, Spark makes it available as a DataFrame for further operations. As with any DataFrame, this data can now be consumed by any other Spark library e.g. Spark MLlib, Spark Streaming and others.

Presently, MinIO's Spark-Select implementation supports JSON, CSV and Parquet file formats for query pushdowns. Spark-Select can be integrated with Spark via spark-shell, pyspark, spark-submit etc. One can also add it as Maven dependency, sbt-spark-package or a jar import.

High-Speed Query Processing

To provide a sense of the performance, MinIO ran the TestDFSIO benchmark on 8 nodes and compared that with similar performance from AWS S3 itself. The average overall read IO was 17.5 GB/Sec for MinIO vs 10.3 GB/Sec for AWS S3. While MinIO was 70% faster (and likely even faster on a true apples to apples comparison) the biggest takeaway for the reader should be that both systems have completely redefined the performance metrics associated with object storage. Needless to say, this performance gap versus AWS S3 will increase as you scale the number of nodes available to MinIO.

This performance extends to writes as well, with both MinIO and AWS S3 posting average overall write IO of 2.92 GB/Sec and 2.94 GB/Sec respectively. Again, the differences between MinIO and AWS S3 are less material than the overall performance.

What this means for the Spark community is that object storage is now in play for Spark jobs that need performance and scalability. AWS S3 provides that in the public cloud. MinIO provides that in the private cloud. One advantage of going the private cloud route with Minio is that the private cloud offers more opportunity to tune the hardware to the specific use case. This means NVMe drives, Optane memory and 100 GbE network. This will offer at least an order of magnitude performance improvements over the public cloud numbers listed above.

About MinIO

Founded in 2014, MinIO is now the world's fastest growing object storage system. Backed by some of the smartest minds in storage and venture capital including Nexus, General Catalyst, Dell Technologies Capital, Intel Capital, AME Cloud Ventures and key angel investors, the company has raised \$23.3M through its Series A round.

Additional Information

Email: hello@min.io
MinIO Inc.
530B University Avenue,
Palo Alto, CA 94301

Resources

<https://min.io>
<https://docs.min.io/>
<https://blog.min.io/>