OKERA

SOLUTION BRIEF

Simplify Fine-grained Access Control for Amazon EMR

Enjoy the cost savings and flexibility you want from Amazon EMR and the sensitive data protection you need from Okera

Amazon EMR (Elastic MapReduce) is a cloud-based platform that provides the elasticity and engines for running Petabyte-scale analysis at a fraction of the cost of traditional on-premise clusters. Customers use EMR to reliably and securely handle big data use cases like machine learning, deep learning, bioinformatics, financial and scientific stimulation, log analysis, and data transformations (ETL). It's easy to set up, operate, and scale your big data environments by automating time-consuming tasks like provisioning capacity and tuning clusters.

Using Okera for EMR can make it practical to run one multi-tenant EMR cluster to support a wide variety of users and tools. Okera can deliver significant cost reduction in AWS storage and processing fees, while also significantly reducing your attack surface, through the use of Okera nScale - a very high-performance, distributed data access layer that runs co-located on the EMR cluster in a secure, isolated process.

With nScale process isolation, user code does not access S3, and thus does not have access to full fidelity data. This allows organizations to implement a zero trust approach to data access, where the cluster does not have IAM access to S3. Instead, Okera provisions data dynamically and in the approved format according to current policies. Dynamically filter, hide, mask, tokenize, and anonymize sensitive data as users work with Amazon S3 data.

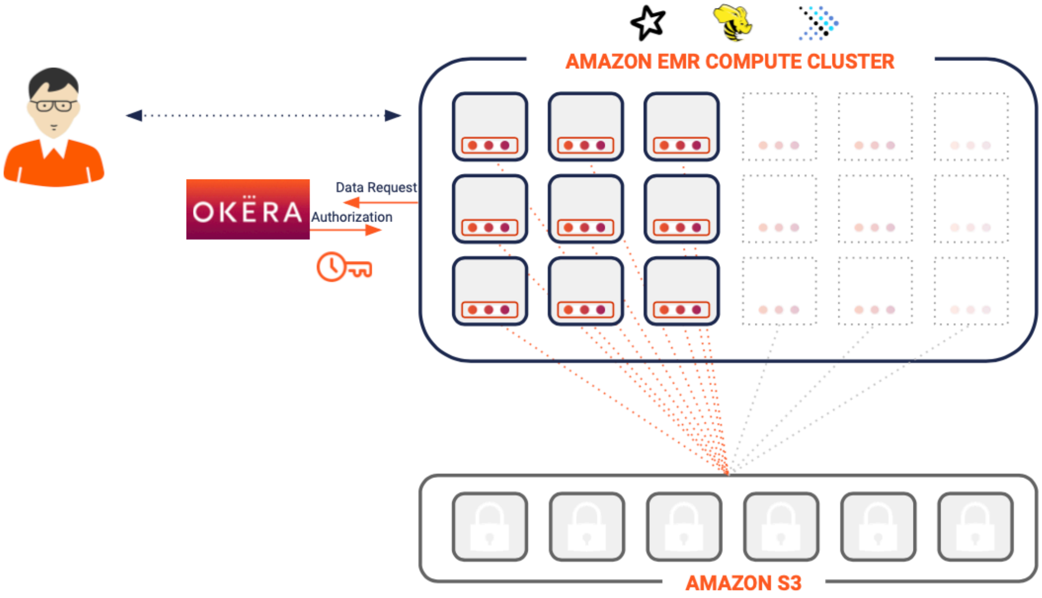
Okera's universal data authorization is complete, clear, and consistent across big data frameworks such as Spark, Hive, and Presto. With Okera for EMR, you can maximize the value of Amazon EMR while protecting confidential, personally identifiable, and regulated data.

Complexity is the Enemy of Security

If managing data authorization gets harder over time, you have a security gap, not a working solution.

Alternate approaches are unnecessarily complicated to set up, and become increasingly unwieldy when adding more data, users, use cases, and tools.

Okera's approach makes data authorization easier to do over time. Registering new data, onboarding new users, and updating policies can be automated and validated within minutes or even seconds. Choose Okera for EMR to get complete, consistent, and clear universal data authorization that can scale as demand increases and business requirements evolve.



1. Okera intercepts data requests sent to Spark, Hive, or Presto and authorizes queries off-cluster using metadata stored in AWS Glue or a Hive Metastore.

2. Authorized data access requests are then delegated to Okera nScale.

3. Okera nScale receives temporary privileges to directly access data in S3 for the exclusive purpose of applying data authorization policies.

4. Okera nScale streams the authorized data on-cluster to the Spark, Hive, or Presto framework for compute processing.

CONTROL

Dynamically apply user-appropriate row-level filters, data masking, and de-identification techniques at the point of query to comply with data privacy regulations and security mandates.

CENTRALIZATION

Okera's data authorization policies are managed, enforced, and audited on a unified platform. There are no gaps or inconsistencies, no complex policy synchronization to manage, and no plugin inconsistencies to work around.

SIMPLICITY

Okera's data authorization policies are easy to write and manage because they are agnostic to the data platform. Define no-code policies through an intuitive web UI or programmatically via API. Once you have a policy in place, simply register the data to be governed.

VISIBILITY

All data access requests are automatically logged for every individual user, down to the exact query, timestamp, access method, sensitive data attribute, and whether requests are approved or denied. Every administrative task is also audited, so you know when policy and platform changes are made, and by whom.

ABOUT OKERA

Okera provides the first universal data authorization platform which empowers companies to accelerate business agility, minimize data security risks and demonstrate regulatory compliance. The Okera Dynamic Access Platform automatically authorizes and audits all data requests, dynamically enforcing data security and compliance policies across all data platforms in hybrid and multi-cloud environments. Okera can be deployed into production in days, adapts to any common data authorization framework, and seamlessly integrates into existing data governance ecosystems.

The company is headquartered in San Francisco and backed by Bessemer Venture Partners, ClearSky Security, and Felicis Ventures. For more information, contact us at info@okera.com or connect with the team on Facebook, Linkedin, or Twitter.