Rockset Security Design

Abstract

Rockset is a serverless search and analytics engine. This document describes Rockset's security design.

October 2022
security@rockset.com

Rockset Security Design

Security is a top priority at Rockset and is taken very seriously. This document describes the current state of security of Rockset. This information can change rapidly, as we are constantly improving our security and as we continue to build our product. Rockset is SOC2 Type II compliant, GDPR & CCPA compliant, HIPAA ready, but not yet compliant with SOX or PCI.

Rockset is a data processor. All data loaded into Rockset is owned by the customer. Rockset indexes and stores all the data to allow SQL queries, but the data is always owned by the customer. The protection of this data is a shared responsibility. Rockset employs a split-responsibility model. Rockset is responsible for ensuring the protection of the underlying data, and the customer is responsible for the protection of their data within our platform. This document describes the steps we take to protect your data, and the tools we provide customers to help ensure they protect their data within our platform.

Use of Cloud Infrastructure

Rockset was born in the cloud. Rockset uses cloud native best practices and leverages the underlying security policies of the public cloud it is hosted on. All of our infrastructure is managed as code, with controls within our CI pipeline to ensure compliance to best practices, and to ensure simple configuration changes won't expose your data to the world. Architecturally, Rockset is designed to be cloud agnostic,however, currently, all of Rockset's services are run and hosted in Amazon Web Services (AWS). Hence our security policies follow AWS best practices at this time. Any time this document mentions Rockset servers, it means EC2 instances in AWS. Rockset does not operate any physical hosting facilities or physical computer hardware of its own. You can view the AWS security processes document here: https://d1.awsstatic.com/whitepapers/Security/AWS_Security_Whitepaper.pdf

Security Features

Data masking

For sensitive data like Personal Identifiable Information (PII), Rockset supports data masking at the time of ingest utilizing field mappings. A field mapping allows you to specify transformations to be applied on all documents inserted into a collection. This can be used for type coercion, anonymization, tokenization, etc. When a particular field is masked using a hashing function like SHA256, only the hashed information is stored in Rockset. You can see how to do field mappings in Rockset here: https://docs.rockset.com/field-mappings/

Views

Rockset allows you to create a view, which is a virtual collection defined by a SQL query. It can be used to limit which fields of a collection can be viewed by the users who are authorized to query the workspace according to RBAC (see below). You can read how to create views here: https://rockset.com/docs/views/

Role based access control

Role-based access control (RBAC) helps you manage user privileges in your organization by enforcing least privileged access to all resources within Rockset. Specify roles and privileges to ensure that you limit access to your data to the individuals that need it.

Rockset grants access to data and actions through role-based authorization and provides built-in roles that provide the different levels of access commonly needed in a database system. You can additionally create custom roles. A role grants privileges to perform sets of actions on defined resources. A role can grant access down to a workspace level of granularity. Custom roles are extended to API keys as well so developers can expose API keys with only limited privileges. You can learn more about RBAC Custom Roles here: https://rockset.com/docs/iam/

Security Architecture

Data in flight

Data in flight from customers to Rockset and from Rockset back to customers is encrypted using TLS 1.2, with certificates created and managed by AWS Certificate Manager.

The only entry-points to the Rockset VPC are through the Rockset API and VPN for Rockset operations staff. Access to these is described in the following sections.

Data at rest

Data is persisted in three places within Rockset:

1. In a log buffer service on encrypted AWS EBS volumes. Rockset uses this log buffer as transient storage to independently scale data indexing (writes) and data serving (reads).
    a. To learn more, visit http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSEncryption.html
2. On our servers, which have local solid state drives which are encrypted using an XTS-AWS-256 block cipher implemented in a hardware module on the instance.
    a. To learn more, visit https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/data-protection.html
3. In AWS S3, where all stored objects are encrypted
    a. To learn more, visit http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingKMSEncryption.html

In all cases, the encryption keys are managed by AWS Key Management Service (KMS). The master keys are never exposed to anyone (not even Rockset), as they never leave the KMS hardware. For evaluation accounts, the master keys used are created in Rockset's AWS account. For commercial installations, Rockset will provide a way for customers to provide their own KMS master key. All customer API keys and integration credentials are stored in a secure admin database that is encrypted at rest.

Access controls

The only ways to access any stored data, servers, or services running inside Rockset's VPC are through VPN servers for the Rockset operations staff and the Rockset API endpoints.

Access to the VPN server requires a TLS auth key, a CA certificate, and the user's individual credentials. Access to the VPN server also requires a hardware token for two-factor authentication. There are no shared passwords for accessing the VPN server. Only employees that require development and administrative access will receive credentials to access the VPN server. Access to resources in the VPN is controlled on a per user basis. For example only required individuals have the ability to send SSH traffic. VPN access for each individual employee can be quickly revoked if necessary.

Access to Rockset servers are controlled by a configuration management system that distributes public SSH keys to certain employees with allowed access to certain servers, configures the UNIX group that each employee belongs to, and configures sudoers files to grant access to run commands to each employee. SSH private keys only live on the employee's laptop. SSH keys can be quickly removed and rotated on all servers in the event that the laptop is lost or the key is compromised. Only SSH key pairs are used for SSH access, not passwords. SSHing as root is also disabled. All laptops have anti-virus software with automated updates of virus and malware signature definitions, and have Mobile Device Management (MDM) enabled which enforces alphanumeric password of at least 12 characters, auto screen lock after 10 minutes, network firewall enabled, disk encryption enabled, the ability to remotely lock or wipe the computer, and periodic scans to ensure the above are in compliance.

Within the VPC, Rockset defines specific network security groups for instances depending on the services running on the network. By default, none of them are publicly reachable via the Internet. Only the VPN servers are reachable directly via the Internet, and security groups are locked down so that only the ports necessary to access the given service are allowed, and only from sources that require this access. Network access per security group is given only on an as needed basis (allowlist only the needed ports/protocols/sources as opposed to blocklist). There are Internet-facing load balancers (AWS ALBs) that service the API. These load balancers only accept inbound traffic over the HTTPS protocol and port 443, and automatically redirect HTTP traffic to HTTPS. There are also AWS CloudFront distributions that service traffic to our website and Console, and terminate TLS connections for those services. These endpoints also redirect HTTP traffic to HTTPS.

Access to Rockset's Console is based on having valid user credentials to an Auth0 user that has been granted access to the Console. Rockset relies on Auth0 for user authentication instead of

implementing our own (auth0.com/docs/overview). It is possible to configure per customer authentication using Google Authentication, Okta Single Sign On, or any identity service that can federate via SAML.

Access to Rockset's API is based on API keys, which can only be created in the Console or the Rockset API. Each API key can only access the resources granted to the user that created the key. All customer API keys and integration credentials are stored in a secure admin database that is encrypted at rest.

Access to AWS resources is provisioned through AWS Identity and Access Management (IAM) users and roles. Our EC2 instances are granted some access required for processes running on them to function (such as access to a specific set of S3 buckets) through instance roles. Developers and administrators who require AWS access are each granted access using AWS SSO which requires a company issued hardware token for two-factor authentication, and uses separate authorizations for the production accounts.

Access to our cluster manager API is done through a role based access control mechanism tied to the AWS SSO user.

CONFIDENTIALITY

Confidentiality is maintained by only allowing those users, non user accounts, and employees access to systems, services, and data when they need it. Secrets that are needed by Rockset's services during run time are stored in AWS Secret Parameter Store, with access to those secrets provisioned via AWS instance roles. Secrets needed by employees to access various third party services are stored in a commercial password management tool. Rockset user credentials are stored in Auth0.

AUDIT LOGGING

Rockset uses AWS Cloudtrail to log all actions taken on our cloud infrastructure. For each AWS account Rockset uses (including production), all logs are shipped to an S3 bucket in a completely separate security AWS account. This bucket only grants write (not read) access to the AWS Cloudtrail service. Cloudtrail log file validation is turned on so that any modification to the log files can be detected:
https://docs.aws.amazon.com/awscloudtrail/latest/userguide/cloudtrail-log-file-validation-intro.html?icmpid=docs_cloudtrail_console

Rockset uses AWS VPC flow logs for all VPCs and the logs are shipped to the same AWS account as the AWS CloudTrail logs.

DEFENSE IN DEPTH

Rockset uses a "defense in depth" strategy to protect its assets. Two-factor authentication using a company issued hardware token is required for access, and for operations which require elevated privileges. The networks are segregated so development and production systems are isolated.

We employ the principle of least privilege so the employee accounts do not have permission to perform privileged operations, and to gain elevated privileges users are required to use two-factor authentication.

We use an Intrusion Detection System to monitor all development and production systems, and immediately alert the operations staff of anomalies.