dbtune

database tuning software as a service

DBtune security and privacy

For self-managed PostgreSQL

March 17, 2024

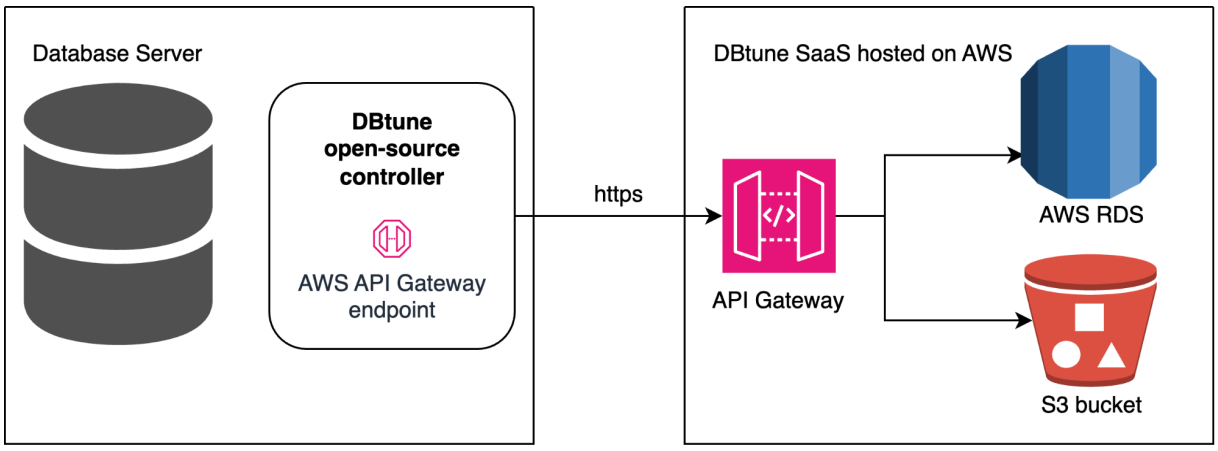
Security overview

At DBtune, we prioritize the privacy and security of our customers and their data. As a machine learning and database management company, we understand the critical importance of safeguarding sensitive information and maintaining the highest standards of data protection.

This section serves as a comprehensive guide to the information and data points that DBtune requests and leverages from the customer database server in order to achieve optimal parameter configuration and drive peak performance.

We aim to be transparent about how we handle data, empowering our customers to make informed decisions regarding their privacy and security.

DBtune architecture diagram



Data transmission security

The DBtune open-source controller, also known as the DBtune client or the DBtune agent, manages communication with the DBtune Optimizer as a Service (OaaS) hosted on AWS. It connects the database server to the OaaS in a unidirectional way, which means that the controller is able to request an operation from the OaaS. The OaaS is explicitly forbidden to issue requests to the controller, which ensures higher security because an external entity is not allowed to issue operations for the controller. To ensure secure unilateral data transmission, the following measures are implemented:

1. **HTTPS API Gateway endpoint**: The DBtune controller utilizes an AWS HTTPS API Gateway endpoint for transmitting data to the DBtune OaaS. HTTPS (Hypertext Transfer

Protocol Secure) ensures encrypted communication between the client (DBtune controller) and the server (DBtune OaaS). This safeguards against eavesdropping and data tampering during transmission.

2. **AWS data transfer security**: Upon reception of data by the OaaS hosted on AWS, AWS handles data transfer security. AWS implements robust security measures at the infrastructure level to protect data in transit and at rest. This includes encryption protocols, network firewalls, and access controls to safeguard data integrity and confidentiality.

By leveraging AWS's infrastructure and the HTTPS protocol, we ensure that data transmission between the DBtune controller and the DBtune OaaS is secure and compliant with industry best practices.

Furthermore, we utilize Amazon Cognito User Pools for managing user authentication and creation. This service securely handles user sign-up, sign-in, and account recovery processes, providing a seamless and secure experience for our users.

Our platform is securely hosted within an Amazon Virtual Private Cloud (VPC), which serves as the cornerstone of our network infrastructure. By leveraging VPC, we ensure that our resources are logically isolated from other tenants within the AWS cloud, enhancing the security and privacy of our data. Overall, our reliance on Amazon VPC underscores our commitment to providing a secure and resilient platform for our users.

Report a security issue

If you have discovered a potential security issue in DBtune, please let us know immediately by sending your report directly to support@dbtune.com. We will respond to your concerns as soon as possible, usually within 24 hours.

How we investigate security issues

When we receive a report we will:

- Acknowledge that we received your report (usually within 24 hours).
- Investigate your report by attempting to reproduce the issue and determine its impact on our customers. We will work with you to make sure we fully understand the issue during our investigation.
- Once our investigation is complete, we will determine how best to resolve the security issue. This might include mitigation, patches, or configuration changes. We will contact all impacted parties with the resolution.

Data storage and processing

DBtune only collects performance telemetric data. DBtune doesn't access data in the database tables and doesn't access table meta-data. The only data accessed by DBtune is typically considered irrelevant for privacy concerns by enterprise Infosec teams irrespective of the geography and the vertical, including for example finance, telecom, automobile, retail.

The performance data collected by DBtune is highlighted below. Once collected this information is stored in the DBtune backend in an Amazon RDS database as shown in the architecture diagram above. The backend is based on AWS Well-Architected Framework (WAFR) standards.

SOC 2 Type II and ISO27001 compliance

Building a culture of security: SOC 2 and ISO 27001

At DBtune, security is paramount. We understand the importance of safeguarding your data and are committed to building a robust security posture. We are actively working towards achieving SOC 2 Type II and ISO 27001 compliance.

Understanding SOC 2 and ISO 27001

- SOC 2 (service organization control 2): Developed by the American Institute of CPAs (AICPA), SOC 2 defines industry best practices for managing customer data security, availability, processing integrity, confidentiality, and privacy. There are two types of SOC 2 reports:
    - Type I: Provides a point-in-time snapshot of your security controls.
    - Type II: Offers a more in-depth assessment by verifying the effectiveness of controls over a period (typically 3-12 months).
- ISO 27001: This international standard outlines a framework for an Information Security Management System (ISMS). It helps organizations systematically manage information security risks and implement best practices.

Our commitment to security and continuous improvement

While we haven't yet achieved formal compliance and certification, we are actively implementing the controls and processes outlined in these frameworks. These include:

- Developing and documenting security policies and procedures.
- Regularly conducting security risk assessments and vulnerability testing.
- Implementing access controls to safeguard your data.
- Training our team on security best practices.

We plan to achieve SOC 2 Type II compliance and ISO 27001 certification within 12 months. We will keep you updated on our progress towards these important security milestones.

In the meantime, you can be confident that:

- We take data security seriously.
- We have implemented strong security measures to protect your information.
- We are continuously working to improve our security posture.

Top 10 OWASP security

At DBtune, we take a proactive approach to application security by adhering to many of the Open Web Application Security Project (OWASP) Top 10 principles. These principles provide a robust framework for identifying, preventing, and mitigating common web application vulnerabilities.

Here's a breakdown of how we address some of these principles:

- **A03:2021** — Injection: We implement secure coding practices and input validation techniques to prevent injection attacks, such as SQL injection and cross-site scripting (XSS).
- **A07:2021** — Identification and authentication failures: We leverage strong authentication mechanisms like AWS Cognito to ensure secure user access and prevent unauthorized activity.
- **A09:2021** — Security logging and monitoring (partially implemented): We actively monitor our systems for suspicious activity and are continuously working to improve our logging and monitoring capabilities for comprehensive security coverage.
- **A3:2021** — Cross-Site Scripting (XSS): We leverage the built-in security features of React, a JavaScript library known for its focus on preventing XSS vulnerabilities.

Security is an ongoing process. We are committed to continuously improving our security posture by:

- Regularly reviewing and updating our security practices based on the latest threats and vulnerabilities.
- Investing in security training for our team members.
- Conducting security audits and penetration testing to identify and address potential weaknesses.

By following these practices, we strive to provide a secure environment for your data. For any further questions about our security practices, please don't hesitate to contact us.

Privacy policy

This privacy policy explains how information about you is collected, used, stored and disclosed by DBT Solutions AB (“DBtune”, “we” or “us”) when you visit our website app.dbtune.com (the “Website”).

Please read this Privacy Policy before using the Website or submitting any personal information through registration. Through such conduct, you agree that your personal data is used in the manner described in this Privacy Policy.

Personal data

“Personal data” means any information relating to an identified or identifiable natural person (“data subject”); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Collection and use of personal information

While using our service, we may ask you to provide us with certain personal data, such as:

- Your name
- Your email addresses
- Your postal address
- Cookies and usage data

We may use your personal data to contact you with newsletters, marketing or promotional materials and other information that may be of interest to you. You may opt out of receiving any, or all, of these communications from us by following the unsubscribe link or instructions provided in any email we send or by contacting us. We may also use your personal data to gather analysis or valuable information so that we can improve our website and associated services.

Third party service providers

We may employ third party companies to monitor and analyze the use of our website. These third parties have access to your personal data only to perform these tasks on our behalf and are obligated not to disclose or use it for any other purpose. We currently use Google Analytics, which is a web analytics service offered by Google Inc. that tracks and reports website traffic. Google uses the data collected to track and monitor the use of our website. You can opt-out of having made your activity on the service available to Google Analytics by installing the Google Analytics opt-out browser add-on. More information regarding Google Analytics may be found on Google’s website: https://policies.google.com/privacy.

Retention of personal data

Personal data that we process for any purpose will not be kept for longer than is necessary for that purpose. We will however retain your personal data for a longer period to the extent that we are required to do so by law or if necessary in order to establish, exercise or defend our legal rights.

Security of personal data

DBtune is committed to keeping your personal data secure against unauthorized access or use, alteration, unlawful or accidental destruction and accidental loss. Only authorized employees, agents and contractors (who have agreed to keep information secure and confidential) have access to your personal data. However, you should be aware that there is always some risk involved in transmitting information over the internet.

Legal rights

You have the right to request information about the personal data that DBtune holds on you, its origin and recipients as well as the purpose for which it is being stored.
If your personal data is incorrect, incomplete or irrelevant, you can ask to have such personal data information corrected or removed. We will notify you within 30 days of receiving the request as to whether and, if so, to what extent, we will comply with your request. If for any reason we don’t comply with your request, we will provide you with the reasons.
You can withdraw your consent to DBtune using the personal data for such purposes that needs your consent at any time.
You may contact us with your privacy concerns or questions by sending an email to info@dbtune.com.

Amendments to this policy

This policy may be changed from time to time in order to keep pace with new developments and opportunities relating to the internet and to stay in line with prevailing legislations. Significant changes to the policy will be publicized on our website along with an updated version of the policy.

Technical specifications

Infosec whitelisting

These are the two endpoints that need to be whitelisted for the OaaS connection.

- AWS endpoints that is used in the DBtune client: https://bwqh2n66kg.execute-api.eu-north-1.amazonaws.com/prod
- AWS s3 bucket link used for downloading the client: https://dbtune-eu-client-package-prod.s3.amazonaws.com

Port

Default port used: 443

Data collection

The data in this section represent the data fetched from the server machine and PostgreSQL instance.

System information

We retrieve the following data once:

- Hardware:
    - Number of CPUs (NUMOFCPU)
    - Total memory (TOTALMEMORY)
    - Available memory (AVAILABLEMEMORY)
    - Cloud provider (CLOUDPROVIDER)
    - Instance type (INSTANCETYPE)
    - Hard drive type (HDTYPE)
    -
- Software:
    - Database version (DBVERSION)
    - Operating system type (OSTYPE)
    - Maximum connections (MAXCONNECTIONS)
    - Database size

To fetch the hardware information such as *NUMOFCPU*, *TOTALMEMORY*, and *AVAILABLEMEMORY*, we utilize Python's *psutil* library. Custom methods were developed to retrieve *CLOUDPROVIDER*, *INSTANCETYPE*, and *HDTYPE*. The software information *OSTYPE* is retrieved using a Python library named *platform*. *DB_VERSION* and *MAXCONNECTIONS* are obtained by querying the database system directly, i.e., "*SHOW server_version*" and "*SHOW max_connections*". A custom method is implemented to fetch disk size.

This information is available in the open-source code of the DBtune controller and can be easily audited.

PostgreSQL performance metrics retrieved

Performance metrics are submitted every second.

- Throughput measured as transactions per second (TPS)
- Average query runtime measured in milliseconds (ms)

To measure throughput information, DBtune queries the *xact_commit* value from the *pg_stat_database* view, while for average query runtime, we retrieve relevant data from the *pg_stats_statements* view. The data is then processed to calculate the average query runtime and TPS.

System monitoring data

Every second, the client fetches and submits server metrics. We utilize Python's *psutil* library to retrieve system metrics, employing *disk_io_*counters method for the resource utilization and IOPS. The resource utilization is defined as the percentage of time during which a disk is actively performing read or write operations relative to its total available time (%util below). DBtune uses the *virtual_memory* method for memory metrics, and the *cpu_percent* method to compute CPU utilization percentage.

The following server metrics are fetched:

- cpu_stats
    - cpu_util
- memory_stats
    - free
    - slab
    - used
    - total
    - active
    - cached
    - shared
    - buffers
    - percent
    - inactive
    - Available
- io_stats
    - iops
    - %util

PostgreSQL parameters tuned

The parameter set is divided into two subsets. The first subset requires a restart of the database instance while the second subset does not require a restart. In the case the end user triggers the restart option in the DBtune web front end then both sets of parameters are tuned simultaneously. The non-restart parameters can be tuned at runtime, hence they are referred as dynamic parameters.

To customize PostgreSQL configuration, we create a file named *99_dbtune_postgres_config.conf* in the *conf.d* directory. We then edit the *postgresql.conf* file to include the config files from *conf.d*, enabling the database to adopt the proposed configuration by DBtune. Whenever updates are made to *99_dbtune_postgres_config.conf*, the database needs to be restarted (in the case the user allows restart) or reloaded (in case the user doesn't allow restart) for the changes to take effect. To revert to the default state, the user can simply remove *99_dbtune_postgres_config.conf* from the *conf.d* directory and restart the database. Reverting back to the default state can also be automated from the DBtune web front end through the configuration management panel.

Restart parameters (static)

- shared_buffers
- max_parallel_workers
- max_worker_processes
- checkpoint_completion_target
- max_parallel_workers_per_gather

Non-restart parameters (dynamic)

- work_mem
- max_wal_size
- seq_page_cost
- random_page_cost
- bgwriter_lru_maxpages
- effective_io_concurrency

In general, these parameters affect the PostgreSQL query planner and improve the way data flows from disk to the processor. A key factor for DBtune is to set the parameters in a way that they achieve good data locality both temporal and spatial locality.

Frequently asked questions

Do I really need to enable the restarting of the database?

No. It is up to you whether you enable the restart of the database during the tuning process. However, enabling database restart allows us to tune parameters that can only be changed with a restart thus making the optimization space that can be explored by DBtune to be more comprehensive. This usually leads to a higher performance improvement. If your application is resilient enough, we recommend enabling the restart.

Could you give an example of database parameters that you adjust to improve performance?

The database parameters control aspects of its runtime behavior, such as buffer pool sizes, caching policies, and other features. An example in PostgreSQL is: *shared_buffers* and *work_mem*. These are part of the tens of parameters that are shipped with PostgreSQL and need to be tuned (traditionally manually) by a DBA to achieve high performance.

Apart from access to database parameters, what type of data does DBtune access? And how is it typically collected?

The data that DBtune needs is the performance metrics of the database, i.e., throughput measured in transactions per second (TPS) and average query runtime measured in milliseconds (ms). In addition, DBtune needs access to the database instance configuration file which will allow DBtune to change and optimize the database parameters.

DBtune doesn't need anything else, the whole database instance is considered a black box for DBtune. DBtune doesn't access the tables or anything else from the user.

DBtune uses standard database commands like *pg_stats* to access the performance metrics.

Can we run DBtune in production?

Yes. Our customers successfully run DBtune in production.

Alternatively, you can use a shadow system or similar for the optimization and then use the optimized configuration provided by DBtune in the production system. While this procedure is not ideal because the database instance being optimized is a surrogate of the production instance, if the shadow database system behaves similarly to the production system the downside of this approach is negligible.

What do I do if my workload changes?

If something changes — such as your workload, your machine or the PG release for instance — you can run a new optimization session using DBtune in order to refine the configuration.

Does DBtune tune queries or indexes?

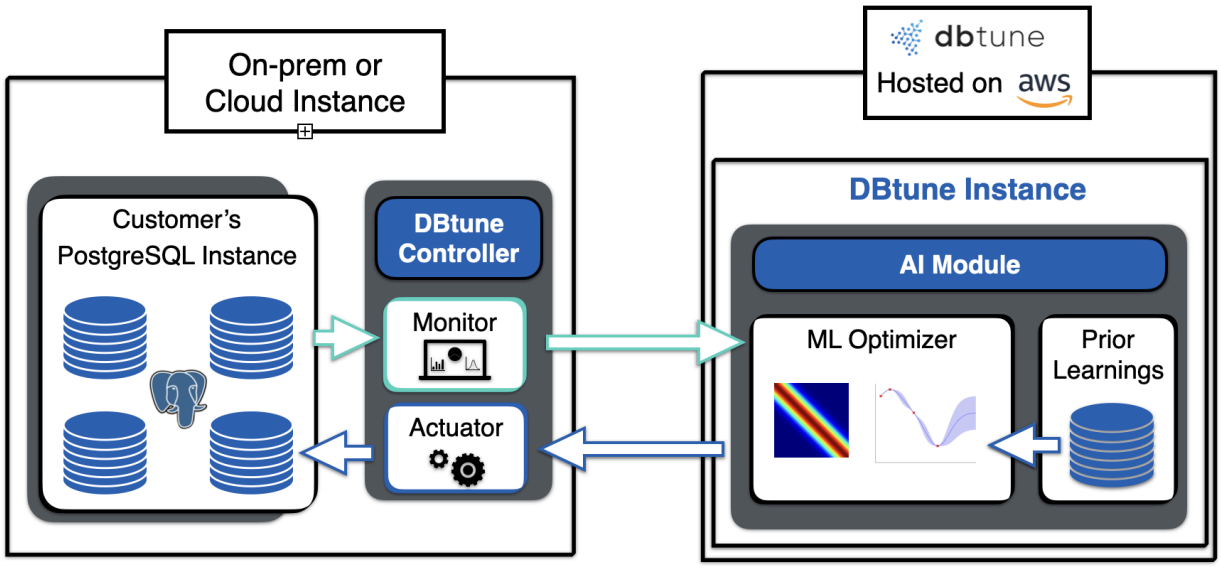
No. DBtune only analyzes and adjusts database configuration settings. It does not propose changes to queries, indexes, or database design.

What happens during the tuning process?

DBtune is an optimizer as a service (OaaS). The OaaS functions as a recommendation system. DBtune finds the best database configuration that optimizes throughput or average query runtime. It learns to optimize by observing the database system, so it generalizes to 1. a new workload, 2. a new database management system, 3. a new hardware or cloud instance type.

The DBtune controller client software is installed on the self-hosted machine by the user — The DBtune controller is open-source and written in Python. Then the DBtune service connects to the customer's database instance through the DBtune controller.

DBtune is privacy safe: The DBtune controller only monitors performance and doesn't have access to your database tables. The image below summarizes the DBtune software architecture for self-hosted database instances and the steps described above.



Can I run DBtune locally in my VPN with no connection to the internet

Yes. DBtune can be deployed offline as well with no connection to the internet. This is a separate product than the optimizer as a service (OaaS). Please contact our support team at info@dbtune.com to request this product deployment option.

dbtune