PostgreSQL DBA

Step by step guide to becoming a modern PostgreSQL DB Administrator in 2022

The intent of this guide is to give you an idea about the DBA landscape and to help guide your learning if you are confused. The roadmap is highly opinionated — neither, knowing everything listed in the roadmap, nor the order of items given in the roadmap is required to be followed in order to be a DBA.

Learn basic RDBMS terms and concepts

Get basic understanding of Postgres key terms and basic RDBMS concepts.

- **Object model**: data types, columns, rows, tables, schemas, databases, queries.
- **Relational model**: domains, attributes, tuples, relations, constraints, NULL.
- **Databases high-level concepts**: ACID, MVCC, transactions, write-ahead log, query processing.
- **Links**:
  - **Postgres Glossary**
  - SQL and Relational Theory - Christopher J. Date, 2009
  - Database Design and Relational Theory - Christopher J. Date, 2012

Learn how to install and run PostgreSQL

Get practical skills of how to set up and run Postgres to get a working environment for further learning.

- Using package managers (APT, YUM, etc.)
- Using `docker` .
- Managing Postgres service using `systemd` (start, stop, restart, reload).
- Managing Postgres service using `pg_ctl` , or OS-specific tools (like `pg_ctlcluster`).
- Connect to Postgres using `psql` .
- Deploy database service in cloud environment (AWS, GCE, Azure, Heroku, DigitalOcean, etc...).
- **Links**:
  - **Official download and install instructions**

  - **Official Docker images**

Learn SQL concepts

Get practical skills of how to create and manipulate database objects and how to execute queries using `psql` client.

- Understand basic data types.
- **DML queries**: querying data, modifying data, filtering data, joining tables.
  - **Advanced topics**: transactions, CTE, subqueries, lateral join, grouping, set operations.
- **DDL queries**: managing tables and schemas (create, alter, drop).
- Import and export data using `COPY` .
- **Links**:
  - **DB Fiddle**
  - **PostgreSQL Tutorial**
  - **PostgreSQL SQL Getting Started**
  - **The SQL Language**

Learn how to configure Postgres

Get understanding of the main aspects of how Postgres could be configured. Deep understanding of Postgres internals is not yet necessary here.

- postgresql.conf:
  - Resources usage
  - Write-ahead Log
  - Checkpoints and Background Writer
  - Cost-based vacuum and auto-vacuum
  - Replication
  - Query planner
  - Reporting, logging and statistics
  - Adding extra extensions
  - ...keep exploring other configuration options
- **Links**:

  - **Postgresqlco.nf**

Learn Postgres security concepts

Get understanding about basic security concepts and common ways of how to deploy secure configurations.

- Authentication models, roles, pg_hba.conf, SSL settings.
- **Objects privileges**: grant/revoke, default privileges.
- **Advanced topics** - row-level security, selinux.
- **Links**:
  - **Client authentication**
  - **Roles and users managements**

Develop infrastructure DBA skills

Get practical skills of how to deploy, extend, maintain and support Postgres installations and 3rd-party Postgres ecosystem software.

- **Replication**: streaming replication, logical replication
- **Backup/recovery tools**:
  - Built-in: `pg_dump` , `pg_dumpall` , `pg_restore` , `pg_basebackup`
  - 3rd-party: `barman` , `pgbackrest` , `pg_probackup` , `WAL-G`
  - Backup validation procedures
- **Upgrading procedures**
  - Minor and major upgrades using `pg_upgrade`
  - Upgrades using logical replication
- **Connection pooling**:
  - `Pgbouncer`
  - Alternatives: `Pgpool-II` , `Odyssey` , `Pgagroal`
- **Infrastructure monitoring**: `Prometheus` , `Zabbix` , other favourite monitoring solution
- **High availability and cluster management tools**:
  - `Patroni`
  - **Alternatives**: `Repmgr` , `Stolon` , `pg_auto_failover` , `PAF`

- **Applications Load Balancing and Service Discovery**: `Haproxy` , `Keepalived` , `Consul` , `Etcd`
- **Deploy Postgres on `Kubernetes`** : Simple `StatefulSet` setup, `HELM` , operators
- Resource usage and provisioning, capacity planning

Learn how to automate routines

Get practical skills, learn automation tools and automate existing routine tasks.

- Automation using shell scripts or any other favourite language (`Bash` , `Python` , `Perl` , etc)
- Configuration management: `Ansible` , `Salt` , `Chef` , `Puppet`

Develop application DBA skills

Learn theory and get practical skills of how applications should work with Postgres

- **Migrations**:
  - practical patterns and antipatterns
  - tools: `liquibase` , `sqitch` , language-specific tools
- Data import/export, bulk loading and processing
- **Queues**:
  - practical patterns and anti-patterns
  - `Skytools PGQ`
- Data partitioning and sharding patterns.
- Database normalization and normal forms.
- Books:
  - The Art of PostgreSQL - Dimitri Fontaine, 2020

Learn Postgres advanced topics

Here is important to continuously extend and develop existing knowledge about Postgres.

- **Low level internals**:
  - Processes and memory architecture
  - Vacuum processing

- Buffer management
- Lock management
- **Physical storage and file layout**
- **System catalog**

- **Fine-grained tuning**:
  - Per-user, per-database settings
  - **Storage parameters**
  - Workload-dependant tuning: OLTP, OLAP, HTAP

- **Advanced SQL topics**:
  - PL/pgSQL, procedures and functions, triggers
  - Aggregate and window functions
  - Recursive CTE

- **Links**:
  - **The Internals of PostgreSQL** for database administrators and system developers
  - **PL/pgSQL Guide**

Learn Postgres troubleshooting techniques

Get basic understanding about troubleshooting tools and get practical skills of how to detect and resolve problems.

- **Operating system tools**
  - `top` (`htop` , `atop`)
  - `sysstat`
  - `iotop`

- **Postgres system views**
  - `pg_stat_activity`
  - `pg_stat_statements`

- **Postgres tools**
  - `pgcenter` - *personal recommendation*

- **Query analyzing**:

- **EXPLAIN**
- **Depesz** online EXPLAIN visualization tool
- **PEV** online EXPLAIN visualization tool
- **Tensor** online EXPLAIN visualization tool, RU language only

- **Log analyzing**:
  - `pgBadger`
  - Ad-hoc analyzing using `grep` , `awk` , `sed` , etc.
- **External tracing/profiling tools**: `gdb` , `strace` , `perf-tools` , `ebpf` , core dumps
- **Troubleshooting methods**: USE, RED, Golden signals
- **Links**:
  - **Linux Performance** by Brendan Gregg
  - **USE Method**

Learn SQL optimization technics

Get understanding and practical skills of how to optimize SQL queries.

- **Indexes, and their use cases**: B-tree, Hash, GiST, SP-GiST, GIN, BRIN
- SQL queries patterns and anti-patterns
- SQL schema design patterns and anti-patterns
- **Links**:
  - **Use the Index, Luke** - a Guide to Database Performance for Developers
- **Books**:
  - SQL Antipatterns: Avoiding the Pitfalls of Database Programming - Bill Karwin, 2010

Develop architect skills

Get deeper understanding of Postgres use cases and where Postgres is suitable and where is not.

- **Postgres forks and extensions**: `Greenplum` , `Timescaledb` , `Citus` , `Postgres-XL` , etc.
- RDBMS in general, benefits and limitations
- Differences between Postgres and other RDBMS and NoSQL databases

Develop Postgres hacker skills

Get involved to Postgres community and contribute to Postgres; be a useful member of Postgres, and the open source community; use personal experience to help other people.

- Daily reading and answering in **mailing lists**
  - pgsql-general
  - pgsql-admin
  - pgsql-performance
  - pgsql-hackers
  - pgsql-bugs
- Reviewing patches
- Writing patches, attending in **Commitfests**