CRUNCHY Enterprise PostgreSQL

Crunchy Certified PostgreSQL Technical Overview

This document introduces Crunchy Certified PostgreSQL and details various features and configurations for deploying Crunchy Certified PostgreSQL in your enterprise.

What is PostgreSQL?

PostgreSQL is a powerful, open source object-relational database system with more than 20 years of ongoing development. The active community, proven architecture, and reputation for reliability, data integrity, feature robustness, and ease-of-use makes PostgreSQL the prime candidate for enterprises looking to move from expensive proprietary technologies.

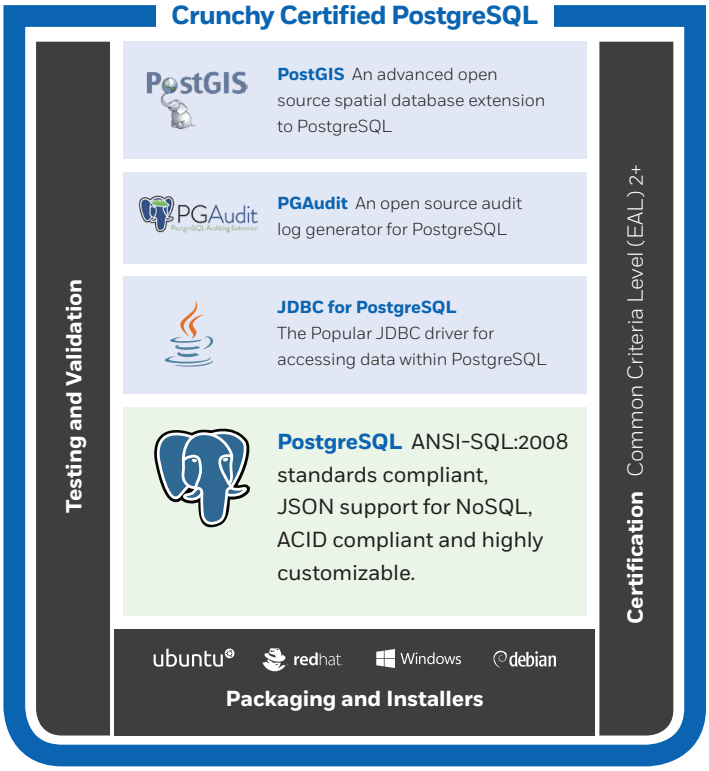
The origins of PostgreSQL date back to 1986 where it was part of the "POSTGRES" project at the University of California at Berkeley. Since then, much of the development focus of PostgreSQL has been on enabling developers to build robust applications, allowing administrators to build reliable, fault-tolerant systems, and helping decision makers eliminate expensive, proprietary licensing fees. PostgreSQL has been fully ACID-compliant since 2001 and is highly complaint with the SQL standard, conforming to 160 of the 179 mandatory SQL:2011 Core conformance, more than any other relational database.

PostgreSQL is also designed to be very extensible. For example, one of the world's most popular geospatial data management tools, PostGIS, is built as an extension to PostgreSQL and takes advantages of the database's data type and indexing infrastructure.

What is Crunchy Certified PostgreSQL?

Crunchy Certified PostgreSQL is certified at Common Criteria Evaluation Assurance Level (EAL) 2+ to ensure it is the most trusted open source enterprise PostgreSQL distribution. Crunchy Certified PostgreSQL is the first commercially available open source relational database management system to receive Common Criteria certification.

Crunchy Certified PostgreSQL is not a fork of PostgreSQL: all of the code that is certified is uses only upstream project components, all of which are actively support by the open source community.



Find out more at www.crunchydata.com

PostgreSQL Feature Overview

Below are highlights of features available in PosgreSQL

Data Types

- Primitives: Integer, Numeric, String, Boolean
- Structured: Date/Time, Array, Range, UUID
- Document: JSON/JSONB, XML, Key-value (Hstore)
- Geometry: Point, Line, Circle, Polygon
- Customizations: Composite, Custom Types

Concurrency, Performance

- Indexing: B-tree, Multicolumn, Expressions, Partial
- Advanced Indexing: GiST, SP-Gist, KNN Gist, GIN, BRIN, Hash, Bloom filters
- Sophisticated query planner / optimizer, index-only scans, multicolumn statistics
- Transactions, Nested Transactions (via savepoints)
- Multi-Version concurrency Control (MVCC)
- Parallelization of read queries
- Table partitioning
- All transaction isolation levels defined in the SQL standard, including Serializable

Extensibility

- Stored procedures
- Procedural Languages: PL/PGSQL, Perl, Python (and many more)
- Foreign data wrappers: connect to other databases or streams with a standard SQL interface
- Many extensions that provide additional functionality, including PostGIS

Data Integrity

- UNIQUE, NOT NULL
- Primary Keys
- Foreign Keys
- Exclusion Constraints
- Explicit Locks, Advisory Locks

Reliability, Disaster Recovery

- Write-ahead Logging (WAL)
- Replication: Asynchronous, Synchronous, Logical
- Point-in-time-recovery (PITR), active standbys
- Tablespaces

Security

- Authentication: GSSAPI, SSPI, LDAP, SCRAM-SHA-256, Certificate, and more
- Robust access-control system
- Column and row-level security

Internalization, Text Search

- Support for international character sets, e.g. through ICU collations
- Full-text search

Source: https://www.postgresql.org/about/

Running PostgreSQL in Enterprise

Crunchy Certified PostgreSQL packages together the essential open source tools for running PostgreSQL at scale and fully compliant with enterprise requirements, including:

High Availability

- Primary / Replica with asynchronous/synchronous replication
- Transaction log shipping and archiving
- Hot/cold primary with shared disk failover
- File system (block device) replication

Backup, Restore, and Disaster Recovery

- WAL log archiving enables point-in-time-recovery (PITR)
- pgBackRest: an open source backup/restore systems designed for terabyte-scale databases
  - Full, incremental, and differential backup
  - Parallel backup and restore
  - Parallel, asynchronous transaction log streaming
  - Encrypted backups
  - Checksumming and backup integrity checks and validations

Foreign Data Wrappers (FDW)

- Foreign data wrappers allow you to connect from within PostgreSQL to a remote file system that is defined by the data wrapper. You can query these systems from directly within PostgreSQL
- Implemented based on the SQL/MED standard
- Support for relational databases such as PostgreSQL, MySQL, Oracle, Sybase, and more
- Support for NoSQL databases like MongoDB, Cassandra, Redis, and more
- Support for messaging platforms like Kafka, RabbitMQ
- Foreign data wrappers enable the creation of federated databases and to aggregate data from multiple sources, specially for reporting/BI uses cases

Auditing

- PostgreSQL has many logging capabilities as well as the ability to generate custom logging parameters
- Crunchy Certified PostgreSQL comes with the pgaudit extension, that generates the information required to evaluate audits performed by government, financial, or ISO certifications

Parallelism and Distributed Query Execution

- Query parallelism introduced in PostgreSQL 9.6 with improvements in all subsequent releases
- Query distribution across different nodes can be achieve with the Pgpool-II and PL/Proxy are open source projects

Logical Replication

- Change-data-capture (CDC) introduced into PostgreSQL in version 9.4
- Logical replication introduced in PostgreSQL 10

JDBC

- Crunchy Certified PostgreSQL comes with the PostgreSQL JDBC driver, which provides an implementation of the standard JDBC specification as well as several PostgreSQL-specific extensions

NoSQL Support

- PostgreSQL supports JSON, XML, key-value data, and more with full transactional support
- PostgreSQL has two data types for JSON: JSON and JSONB
  - JSON is a text-based implementation
  - JSONB is a binary implementation with advanced indexing capabilities

Geospatial Data Management

- PostgreSQL comes with built-in support for simple geometric data types
- Crunchy Certified PostgreSQL comes with the PostGIS geospatial database extender, which adds robust GIS functionality
- PostGIS itself has extensions that can be used, such as pgRouting, which provides geospatial routing functionality

Full Text Search

- Native support for document-based full text search across different languages
- Full text search support within JSON and JSONB data types since PostgreSQL 9.6

PostgreSQL Architectures for Different Workloads

Heavy Write Workload

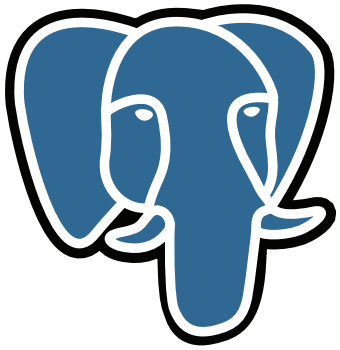
- I/O spreading via tablespaces
- Write spreading using WAL relocation
- Application-based sharding
- Custom database sharding
- Table partitioning
- Unlogged tables
- PostgreSQL 11 introduced advanced partitioning support that, in conjunction with foreign data wrappers, enables you to distribute and shard data across different notes. Prior to PostgreSQL 11, this can be accomplished using PL/Proxy

Analytical Workload

- Declarative table partitioning (PostgreSQL 10+)
  - Range partitioning
  - List partitioning
  - Hash partitioning (PostgreSQL 11+)
- Table Inheritance
- Federating database content via the postgres_fdw

Heavy Read Workload

- Load-balanced read-only, replicas & streaming replication is a proven way to scale your cluster to accommodate large number of read requests.
- Read scaling using Pgpool & streaming replication
- Cascading replication
- Scaling using Foreign Data Wrappers
- Application-based Sharding
- Sharding using PL/Proxy or, in PostgreSQL 11 a combination of partitioning + foreign data wrappers.
- Read reduction via increased memory
- Read parallelism using effective_io_concurrency



162 Seven Farms Drive Suite 220 Charleston, SC 29492

+1-843-737-6045 • info@crunchydata.com • crunchydata.com