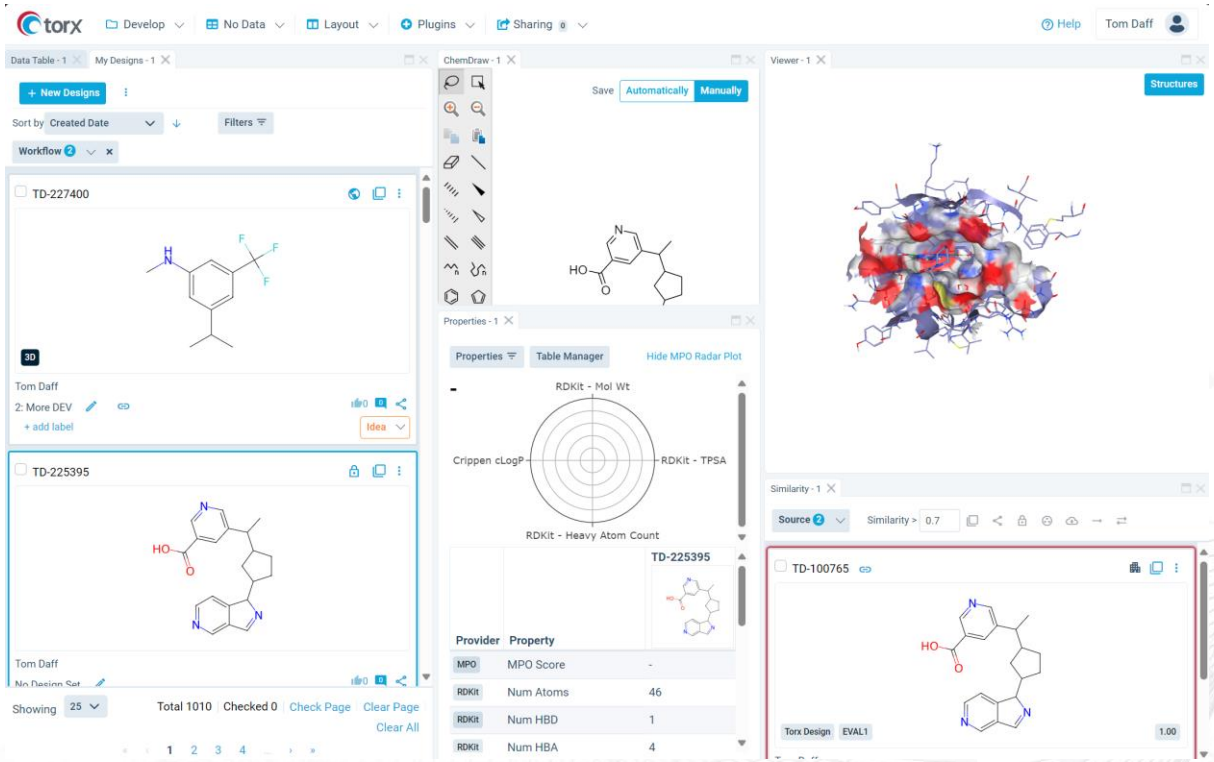
RDKit on RDS

Tom Daff
Web Development, Torx

Torx DMTA

- Torx is a small molecule DMTA platform
- RDKit is used:
  - > Structure aware database storage and searching from Django
  - > Properties
  - > Structure rendering
- Users typically work with hundreds to thousands of structures at any time
- “Cloud native” software-as-a-service



RDKit Postgres cartridge

Dockerfile 3.08 KiB | Blame | Edit | Lock | Replace | Delete

```
# Build of RDKit database cartridge into the offical Postgres Docker
# image. Modified in the Torx fork to use specific versions of Postgres
# and RDKit. Debian preferred as Alpine reduces structure search speed
# by ~30%.
ARG RDKIT_VERSION=Release_2023_09_5
ARG POSTGRES_VERSION=14.11-bookworm

FROM postgres:${POSTGRES_VERSION} AS rdkit-postgres-build-env

# hadolint ignore=DL3008
RUN apt-get update \
 && apt-get install -yq --no-install-recommends \
    build-essential \
    ca-certificates \
    cmake \
    git \
    libboost-dev \
    libboost-iostreams-dev \
    libboost-serialization-dev \
    libboost-system-dev \
    libeigen3-dev \
    postgresql-server-dev-14 \
    wget \
    zlib1g-dev \
 && apt-get clean \
 && rm -rf /var/lib/apt/lists/*

# Refresh arg from outer scope
ARG RDKIT_VERSION
RUN wget --quiet https://github.com/rdkit/rdkit/archive/${RDKIT_VERSION}.tar.gz \
 && tar xf ${RDKIT_VERSION}.tar.gz \
 && mv rdkit-${RDKIT_VERSION} rdkit \
 && rm ${RDKIT_VERSION}.tar.gz \
 && mkdir /rdkit/build
WORKDIR /rdkit/build

# Disabled Freetype as it fails to configure.
RUN cmake -Wno-dev \
    -D CMAKE_BUILD_TYPE=Release \
    -D RDK_BUILD_FREETYPE_SUPPORT=OFF \
    -D RDK_BUILD_INCHI_SUPPORT=ON \
    -D RDK_BUILD_PYTHON_WRAPPERS=OFF \
    -D RDK_BUILD_PGSQL=ON \
    -D PostgreSQL_ROOT=/usr \
    -D PostgreSQL_TYPE_INCLUDE_DIR=/usr/include/postgresql/14/server/ \
    .. \
 && make -j "$(nproc)"

# Run the tests
ENV RDBASE=/rdkit/

# Postgres tests need a working database and write access
# for the postgres user to the test directory
RUN cp /rdkit/build/Code/PgSQL/rdkit/rdkit--*.sql /usr/share/postgresql/14/extension \
 && cp /rdkit/Code/PgSQL/rdkit/rdkit.control /usr/share/postgresql/14/extension \
 && cp /rdkit/build/Code/PgSQL/rdkit/librdkit.so /usr/lib/postgresql/14/lib/rdkit.so \
 && chmod 777 /rdkit/build/Code/PgSQL/rdkit/ \
 && chmod 777 /rdkit/build/Testing/Temporary/
```

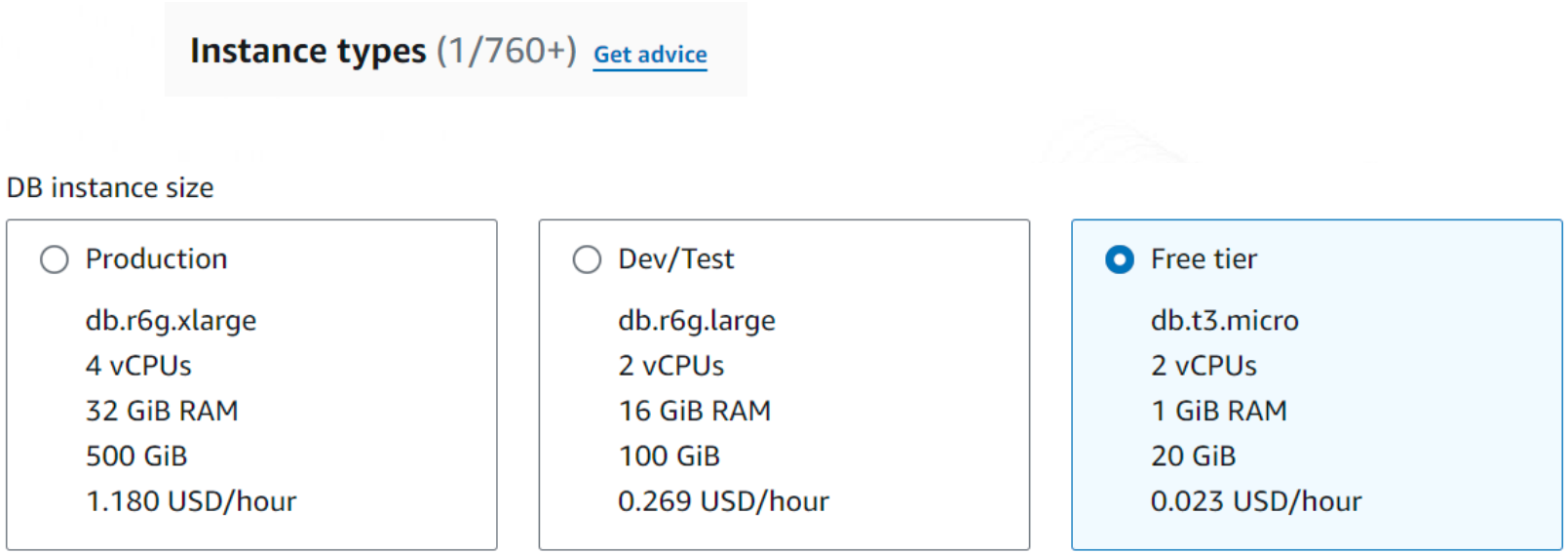
Extensions supported for RDS for PostgreSQL 16

The following table shows PostgreSQL extensions for PostgreSQL version 16 that are currently supported on Amazon RDS. For more information on PostgreSQL extensions, see Packaging related objects into an extension.

Extension	16.4	16.3-R2	16.3	16.2-R3	1(R
rdkit	4.5.0 (Release 2024_03_5)	4.4.0	4.4.0	4.4.0	4.

Tuning performance

- Noticed when testing a “security” update to move from Debian to Alpine containers there was an over 20% *decrease* in performance
- Highlighted that our current cloud setup is not well optimised
- Huge choice of options in AWS



Benchmarking

- Loosely followed the 10M compound benchmarks on the RDKit blog

	Self-managed	Managed small	Managed large	Managed xlarge
	t3a.medium	db.t3.micro	db.m5.large	db.m6g.xlarge
Cost / month	£17	**£10**	£140	£200
Create, index and fingerprint	4 hr		**2 hr**	
Similarity / s	32	29	**18**	18
Substructure / s	2.9	3.5	**2.8**	3.3

- Cheapest managed option better than untuned self-manged
- Databases scale better for concurrency than vertically for the cost
- Next up: “serverless” and specialised systems