MolSQL Whitepaper (V3.0)

The First Chemistry Cartridge for Microsoft SQL Server

MolSQL is a SQL Server extension to enable chemistry intelligence.

Why you need a Chemistry Cartridge?

- A Chemistry Cartridge can minimize the efforts to adopt chemistry capabilities for your applications
- A Chemistry Cartridge optimizes the data architecture to maximize the structure search efficiency

What you can do with MolEngine?

- Create a molecule database
- Create a compound registration database
- Create a reaction database
- Do efficient structure search, including substructure, full_structure, similarity search
- Handle popular chemical file formats, including Molfile, Rxnfile, CML, CDX, CDXML, SMILES etc.
- Generate structure images, including PNG, GIF, JPG, TIF, EMF etc.
- Compute molecule/reaction properties, such as Molecular Weight, Formula, t_PSA etc.
- … and many more

Cartridge in Action

1. Working with Molecule

- Creating and indexing molecule table

```
CREATE TABLE moltable (id BIGINT NOT NULL PARIMARY KEY, smiles VARCHAR(4000))
GO

MolIndex 'create', 'moltable', 'smiles', 0 /* allow duplicated structures */
GO
```

NOTE:

i. Structure column can be VARCHAR or VARBINARY.

ii. The last parameter of MolIndex can be set to 1 to restrict unique structures in

iii. that table

```
CREATE TABLE moltable (id BIGINT NOT NULL PARIMARY KEY, mol VARBINARY(MAX)) GO

MolIndex 'create', 'moltable', 'mol', 1 /* restrict unique structures */
GO

/* drop structure index */
MolIndex 'drop', 'moltable', 'mol', 1 GO
```

- Searching molecule table

```
/* substructure search */
MolSelect 'SELECT id, $Formula(mol, null) formula FROM moltable
    WHERE $Search(Substructure, mol, c1ccccc1)'
GO

/* full_structure search */
MolSelect 'SELECT id, $Formula(mol, null) formula FROM moltable
    WHERE $Search(Fullstructure, mol, c1ccccc1)'
GO

/* similarity search */
MolSelect 'SELECT id, $Formula(mol, null) formula FROM moltable
    WHERE $Search(Similarity, mol, c1ccccc1, 0.8)'
GO
```

2. Working with Reaction

- Creating and indexing molecule table

  CREATE TABLE rxntable (id BIGINT NOT NULL PARIMARY KEY, smiles VARCHAR(4000))
  GO

  RxnIndex 'create', 'rxntable', 'smiles'
  GO

  NOTE:
  i. Structure column can be VARCHAR or VARBINARY.

  CREATE TABLE rxntable (id BIGINT NOT NULL PARIMARY KEY, rxn VARBINARY(MAX))
  GO

  RxnIndex 'create', 'rxntable', 'rxn'
  GO

  /* drop structure index */
  RxnIndex 'drop', 'rxntable', 'rxn'
  GO

- Searching reaction table

  /* substructure search */
  RxnSelect 'SELECT id, $Formula(rxn, null) formula FROM rxntable WHERE $Search(Substructure, rxn, >>c1ccccc1)'
  GO

  RxnSelect 'SELECT id, $Formula(rxn, null) formula FROM rxntable WHERE $Search(Substructure, rxn, c1ccccc1>>)'
  GO

  RxnSelect 'SELECT id, $Formula(rxn, null) formula FROM rxntable WHERE $Search(Substructure, rxn, c1ccccc1O>>c1ccccc1F)'
  GO

  /* full_structure search */
  RxnSelect 'SELECT id, $Formula(rxn, null) formula FROM rxntable WHERE $Search(Fullstructure, rxn, >>c1ccccc1)'
  GO

3. Utility Functions

- MolWeight/MolWeightS
- ExactMass/ExactMassS
- Psa/PsaS
- Noha/NohaS
- Nohd/NohdS
- Norb/NorbS
- Smiles/SmilesS
- Formula/FormlaS
- HtmlFormla/HtmlFormaS
- Jpg/JpgS
- Gif/GifS
- Png/PngS
- MolFile/MolfileS
- CMLFile/CMLFileS