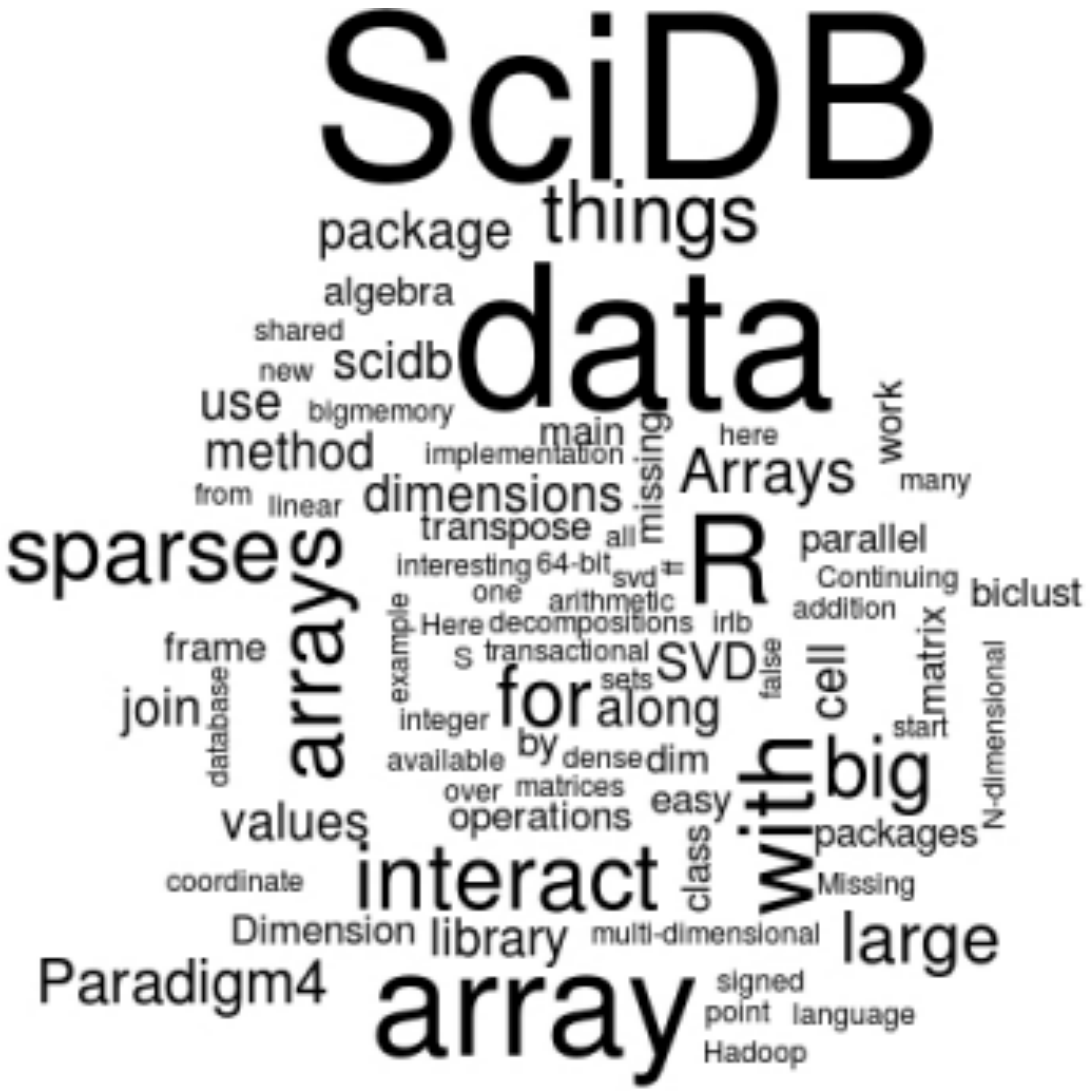

http://goo.gl/Hn0Ro

Bryan Lewis, Paradigm4
blewis@paradigm4.com

Free and open-source array database

Sparse/dense, multi-dimensional arrays

Distributed storage, parallel processing

Excels at parallel sparse/dense linear algebra

ACID, data replication, versioned data

The package defines two main ways to interact with SciDB:

1. Iterable data frame interface using SciDB query language directly

2. **N-dimensional sparse/dense array class for R backed by SciDB arrays**

```
library("scidb")
scidbconnect(host="localhost")


# An example reference to a SciDB matrix:
A <- scidb("A")
dim(A)
[1] 50000 50000
```

Subarrays return new SciDB array objects

```
A[c(0,49000,171), 5:8]
```

```
Reference to a 3x4 SciDB array
```

Use `[]` to materialize data to R

```
A[c(0,49000,171), 5:8][]
```

```
           [,1]       [,2]       [,3]       [,4]
[1,]  0.9820799 -0.4563357 -1.2947495 -0.8085465
[2,] -1.5090126  0.1547963 -0.2435732 -0.1836875
[3,]  1.3296710 -1.5006536 -0.5980172  0.3752186
```

Arithmetic

```
X <- A %*% A[,1:5]
dim(X)
```

```
[1] 50000     5
```

Mixed **SciDB** and **R** object arithmetic

```
Z <- A[c(0,49000,171), 5:7]

(0.5*(Z + t(Z)) %*% rnorm(3))[, drop=FALSE]

          [,1]
[1,]  3.707263
[2,] -2.833560
[3,]  3.518370
```

Basic aggregation (scidbdf class)

```
A <- as.scidb(iris)
Warning message:
In df2scidb :Attribute names have been changed

aggregate(A, Petal_Length ~ Species, "avg
(Petal_Length) as mean")

     Species  mean
1     setosa 1.462
2 versicolor 4.260
3  virginica 5.552
```

It is sometimes possible to use SciDB arrays in R packages with little modification.

```
library("biclust")
library("s4vd")
data(lung)
A <- lung
x <- biclust(A, method=BCssvd, K=1)

# Now with SciDB arrays:
library("s4vdp4")
X <- as.scidb(A)
x1 <- biclust(X, method=BCssvd, K=1)

# Compare the results:
sqrt( x@info$res[[1]]$u - x1@info$res[[1]]$u) )
              [,1]
[1,] 5.202109e-16
```

SVD and principal components

```
S <- svd(A, nu=3, nv=3)
dim(S)
[1]     4 50000 50000

# Result is a 3-D array containing U,
  S (sparse), and V
```

paradigm4

data-driven discovery

Virtual machines and EC2 images ready to roll (including Rstudio) available from:

www.scidb.org

R package on CRAN and development version at:

github.com/Paradigm4