

The FlyMine Project

University of Cambridge, UK
<http://www.flymine.org/>
<http://www.intermine.org/>

Data warehouse for bioinformatics

Uses the InterMine Data Warehousing System

Written in Java

Full object-oriented database and query system

Data integration toolset

Web application

I will be talking on two parts that may be useful to you:

BatchWriter

Bioseq

BatchWriter

High performance Java database writing API

Asynchronous, writes in background, but has well-defined behaviour

Uses dirty tricks™ and binary COPY to maximise performance

`addRow()`, `deleteRow()`, `clear()`

`flush()`, `backgroundFlush()`, `batchCommit()`,
`close()`

Bioseg

An R-Tree GiST index implementation copied from seg and altered to handle two integers (rather than two floating point numbers), using an improved algorithm

Much better than using the geometric shapes index for two integers (index size and performance)

Even so, a built-in R-tree implementation would be very welcome

<http://www.bioinformatics.org/bioseg/wiki/>