Chapel I/O: Getting to Your Data Wherever It Is

Tim Zakian
IU – Bloomington

Michael Ferguson
LTS

Brad Chamberlain
Cray

What's Wrong With This Code?

```
// example-bad.chpl

param const path = "test.txt";
// Open a file
var fl = open(path, iomode.r);
// Get the length of the file
var len = fl.length();

// In most cases non-optimal
forall i in 0..len by len/10 {
 // get a section of the file
 var reader = fl.reader(start=i,
                        end=i+len/10,
                        locking=false);
//.. do some stuff with this section
}
```

```
// example-better.chpl

param const path = "test.txt";
// Open a file
var fl = open(path, iomode.r);
// Get the length of the file
var len = fl.length();

// Return the optimal read size
var (start, end) = fl.getchunk();

// Better! Now let's optimize
forall i in 0..len by end-start {
 // get a section of the file
 var reader = fl.reader(start=i,
                        end=i+end-start,
                        locking=false);
//.... do some stuff with this section
}
```

Let's Make It Better!

```
// example-better.chpl

param const path = "test.txt";
// Open a file
var fl = open(path, iomode.r);
// Get the length of the file
var len = fl.length();

// Return the optimal size of reads
var (start, end) = fl.getchunk();
var blen = end-start;
// Better! Now let's optimize
forall i in 0..len by blen {
 // get a section of the file
 var reader = fl.reader(start=i,
                        end=i+blen,
                        locking=false);
//.. do some stuff with this section
}
```

```
// example-best.chpl

param const path = "test.txt";
// Open a file
var fl = open(path, iomode.r);
// Get the length of the file
var len = fl.length();

// Return the optimal size of reads
var (start, end) = fl.getchunk();
var blen = end-start;
// Better! Now let's optimize
forall i in 0..len by blen {
  var locs=fl.locsForRegion(i,i+blen);
  on locs[0] {
   // get a section of the file
   var reader = fl.reader(start=i,
                          end=i+blen,
                         locking=false);
//.... do some stuff with this section
 }
}
```

What If You Want to Use HDFS?

```
// example-lustre.chpl

param const path = "test.txt";
// Open a file
var fl = open(path, iomode.r);
// Get the length of the file
var len = fl.length();

// Return the optimal size of reads
var (start, end) = fl.getchunk();
var blen = end-start;

forall i in 0..len by blen {
  var locs=fl.locsForRegion(i,i+blen);
  on locs[0] {
   var reader = fl.reader(start=i,
                  end=i+blen,
                  locking=false);
//.. do some stuff with this section
 }
}
```

```
// example-hdfs.chpl

param path = "hdfs://coal:0/test.txt";
// Open a file
var fl = open(path, iomode.r);
// Get the length of the file
var len = fl.length();

// Return the optimal size of reads
var (start, end) = fl.getchunk();
var blen = end-start;

forall i in 0..len by blen {
  var locs=fl.locsForRegion(i,i+blen);
  on locs[0] {
   var reader = fl.reader(start=i,
                  end=i+blen,
                  locking=false);
//.... do some stuff with this section
 }
}
```

What If Your Data Is Elsewhere?

```
➢ chpl example-best.chpl -o example

// Open a standard (or Lustre) file
➢ ./example --path="test.txt"

// What if your data is on HDFS?
➢ ./example --path="hdfs://coal:0/test.txt"

// What if your data is a webpage?
➢ ./example --path="http://chapel.cray.com"

// Or on an FTP server?
➢ ./example --path="ftp://..."
```

Why Is This Cool?

- Can easily read data in a way that optimizes read performance on distributed file systems (locality and stripe/block size aware).
- Easily stream live data (e.g., upload current performance and results to something like Plotly or Fusion Tables).
- Access data wherever it may reside – on the file system or elsewhere (websites, Google Fusion Tables, FTP servers etc.).

See more at:

$CHPL_HOME/doc/release/README.(auxIO, curl)