

[left=1.5in,top=1in,right=1in,bottom=1in]geometry

# GIS PROCESSING ON COMPUTE CLUSTERS

by

Nathan Thomas Kerr

A Thesis Presented in Partial Fulfillment  
of the Requirements for the Degree  
Master of Science

ARIZONA STATE UNIVERSITY

August 2009

[left=1.5in,top=1in,right=1in,bottom=1in,includefoot]geometry

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam id ipsum ut ligula fringilla eleifend nec posuere dui. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Ut lobortis lobortis varius. Morbi pulvinar risus in ante imperdiet vitae pharetra enim vestibulum. Aliquam erat volutpat. Duis sed justo sed ante venenatis euismod a ac arcu. Donec et iaculis nisl. Sed ac augue massa, eget interdum sapien. Donec nec lorem mauris, ac ultricies mi. Phasellus ligula ipsum, feugiat ut malesuada nec, suscipit a turpis. Etiam mattis aliquam mattis. Phasellus lorem magna, facilisis vitae scelerisque sit amet, eleifend nec tellus. Etiam feugiat, magna at pharetra luctus, dolor odio mattis erat, eget elementum magna nunc id lacus. Quisque tempor auctor ultrices. Suspendisse potenti.

Sed sed sapien tellus, in posuere metus. Pellentesque condimentum pellentesque erat. Sed tincidunt sodales semper. Sed condimentum, nibh in tempus ornare, magna turpis tempus nisl, non tincidunt metus velit sit amet nibh. Suspendisse et augue mauris, non hendrerit quam. Sed ut lacus quis tellus congue blandit. Phasellus in neque massa. Donec posuere molestie aliquet. Donec quis nulla nibh, at mattis orci. Donec sit amet purus erat. Vivamus elementum nulla et enim iaculis quis blandit leo egestas. Maecenas lobortis, metus id ultrices facilisis, metus felis venenatis urna, nec interdum nulla justo at odio.

Sed vehicula condimentum faucibus. Pellentesque nec neque urna, et malesuada justo. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec

faucibus porta purus eget sagittis. Suspendisse lacinia nunc id lacus ultricies et vulputate nisl sollicitudin. Phasellus sed tortor velit, id gravida lorem. Fusce suscipit metus ut nisi facilisis vitae egestas neque sagittis. In nec libero nec velit ultricies cursus. Suspendisse potenti. Fusce vel velit in orci condimentum consequat. Nunc in sodales sem.

Pellentesque laoreet, velit at sollicitudin facilisis, leo orci bibendum nibh, ut pretium sem dolor quis magna. Aenean in odio dolor. In sit amet nunc libero, sit amet porta nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec dapibus mauris at odio vulputate eu feugiat ligula iaculis. Integer pulvinar imperdiet nisl vitae adipiscing. Donec viverra enim ut neque consectetur sit amet malesuada enim vehicula. Nam vel quam in quam dapibus blandit. Praesent ullamcorper urna at nibh sodales a hendrerit nisl porta. Phasellus ultrices dapibus ante, ac tincidunt enim semper et. Maecenas congue aliquet sapien ut blandit. Duis pellentesque velit non arcu lobortis at feugiat lectus vestibulum. Nulla euismod vestibulum nisl, vel sollicitudin felis scelerisque id. Sed facilisis pellentesque mi eget consectetur. Nulla in dolor sit amet magna interdum elementum. Cras non lorem in ipsum vehicula scelerisque at non tellus. Nullam accumsan posuere vulputate.

Sed vitae augue vel dui vehicula congue. Maecenas magna nibh, tincidunt sit amet molestie ut, congue ut velit. Proin ante purus, dapibus a accumsan et, vulputate ac neque. In et dolor non dolor porta ornare. Curabitur diam est, mattis et rhoncus ut, sodales mollis enim. Vivamus eget dolor arcu, et aliquam ipsum. Fusce quis pulvinar purus. Ut eu lectus nunc. Vivamus ut dolor felis, at convallis

lorem. Cras ut urna a nisi euismod lobortis. Donec id ipsum non nisl tempus luctus. Sed id enim velit. Integer urna arcu, consequat et sodales sed, luctus sed risus. Morbi eu ipsum urna. Aenean non sem velit, ac porttitor tortor. Donec at tempus nisi. Vivamus nec laoreet velit. Praesent aliquet ipsum tempus orci pharetra in eleifend nisi porta. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam consequat mollis tellus vitae interdum.

Aenean dapibus pharetra neque sit amet tincidunt. Ut venenatis sem sed ipsum tempor facilisis. Phasellus nunc lectus, imperdiet quis sagittis a, molestie vitae nibh. Pellentesque quis mi sed ligula faucibus eleifend ac eu arcu. Nunc vitae lacinia nunc. Aenean felis nisi, semper sed volutpat nec, pharetra sed odio. Sed tempor est ac velit fermentum vel aliquam nunc dignissim. Sed lacinia mauris nec metus bibendum vehicula. Donec turpis purus, ultrices non lacinia eu, gravida quis risus. Donec dapibus adipiscing aliquam. Vestibulum ac nunc ante, ut volutpat mi. Pellentesque pellentesque pretium arcu, id tincidunt odio congue nec. Proin sodales nunc ut ipsum fermentum sit amet sollicitudin lectus sagittis. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In vestibulum orci et elit rhoncus vitae porta ipsum euismod. Integer id augue sem. Ut ut lectus et nisl fringilla posuere vel sit amet mauris. Praesent a accumsan diam.

Phasellus sit amet magna id urna convallis lobortis sed ut turpis. Aliquam vel tempor massa. Fusce adipiscing pulvinar consequat. Praesent vel scelerisque sem. Curabitur pharetra turpis ac nulla dictum sit amet vestibulum velit pellentesque. Sed eu ligula mi. Nullam velit nunc, laoreet ut convallis vel, mattis in lacus. Nulla tempor tincidunt consectetur. Donec vitae felis vel tortor blandit ornare ac

nec lacus. Ut risus nunc, euismod gravida pellentesque quis, ultricies sed neque. Donec ultrices lacus eu est facilisis congue vel vitae lacus. Ut consequat turpis eget justo rhoncus quis lacinia metus facilisis. Nulla rhoncus faucibus enim, sit amet euismod nunc tempor sed. Nullam dignissim gravida neque nec cursus. Morbi tellus risus, sodales in tristique sit amet, euismod in purus.

Quisque dictum scelerisque quam, nec tristique elit mattis vel. Praesent euismod aliquam eros, a consectetur tellus interdum venenatis. Etiam vel aliquam lectus. Aliquam quis leo sed neque consequat dictum. Suspendisse et pretium mauris. Suspendisse vel tellus vitae erat cursus bibendum. Proin et magna et diam tincidunt tincidunt eu et risus. Quisque mi odio, fermentum non sagittis id, pretium ac ante. Etiam ac mauris vitae justo dapibus laoreet vel nec mauris. Donec vitae feugiat ipsum. Etiam sit amet arcu ac lorem cursus pretium euismod in ante. Suspendisse a mauris a ligula ultrices luctus. Aliquam ut est neque, vel accumsan nisi. Curabitur venenatis pulvinar urna eu pretium.

Quisque quis mi id eros hendrerit mattis ac malesuada erat. Curabitur vestibulum consequat sapien, nec bibendum massa faucibus eget. In pulvinar congue tristique. Cras risus lectus, ornare nec iaculis congue, blandit vitae lorem. Sed non molestie urna. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Maecenas malesuada, felis vehicula pretium pretium, turpis magna vestibulum libero, a pharetra felis velit eget tortor. Duis tempus erat hendrerit sapien commodo congue. Aliquam turpis ipsum, hendrerit ut mollis a, tincidunt id metus. In porta aliquam turpis, sit amet posuere risus varius ac. Nullam tincidunt dui ut odio lobortis cursus. Phasellus nec felis eros. Nam tempus

vehicula feugiat. Maecenas scelerisque, urna a tincidunt venenatis, metus diam fringilla ligula, et pharetra orci est vitae urna. Ut consectetur ante at velit blandit ultrices. In hac habitasse platea dictumst. Quisque odio odio, molestie ut euismod sed, volutpat et sem.

Phasellus quis odio nec justo tincidunt placerat. Suspendisse justo magna, cursus ac facilisis a, elementum quis quam. Pellentesque a massa non odio sagittis varius eu quis arcu. Donec faucibus est et enim sagittis commodo. Morbi elementum, augue vehicula volutpat euismod, ipsum metus accumsan ante, ac porta dolor enim vitae enim. In varius sem sed nisl ullamcorper vulputate. Sed cursus, est eu malesuada consectetur, odio sapien faucibus leo, at suscipit leo urna non metus. Donec ullamcorper placerat commodo. Etiam in elit sem, eu consectetur diam. Duis ultrices lectus vel erat placerat eget consequat ligula tempor. Praesent dignissim, felis sit amet accumsan venenatis, risus nulla egestas odio, feugiat sodales tellus dolor at nulla. Ut non magna erat, facilisis cursus massa. In at velit massa, sagittis fringilla eros. Vestibulum semper ipsum eget velit sodales pretium.

body of research



Back Matter

# **Chapter 1**

## **Introduction**

GIS processing is important

### **1.1 Current methods**

Desktop GIS

Database GIS

### **1.2 Significance**

Limitations of single processor implementations

### **1.3 Attempted solutions**

Parallel DB

Parallel GRASS

Problems with desktop programs on clusters

## **1.4 Needed solution, cluster programming model**

batch

mpi

scalable architecture

spmd

Thesis statement goes here

How this document goes about supporting the thesis statement.

# **Chapter 2**

## **Related Work**

### **2.1 Desktop GIS**

ArcInfo, QuantumGIS, GRASS GIS

### **2.2 Database GIS**

PostGIS, Oracle Spatial, Client-Server Paradise

### **2.3 GIS libraries**

JTS, GEOS

## **2.4 Parallel DB**

Paradise

## **2.5 Parallel GRASS**

Other converted programs?

## **2.6 MRGIS MapReduce GIS**

# Chapter 3

## Requirements

batch mode processing

- data distribution options

- standard geospatial operations

- scalable, high performance

- dataset centric

### **3.1 easy to use to form new operations**

Sample program:

- initClusterGIS

- loadData (distributed/replicated)

- process (user code here)

- saveData (distributed/replicated)

`finalizeClusterGIS`

# **Chapter 4**

## **Design/Implementation**

How to fulfill requirements

- Architecture of solution

- Split dataset across tasks (MPI I/O)

- Each task works on their own part

- MPI to communicate if needed



# **Chapter 5**

## **Experimental Setup/Design**

How to show how well requirements were filled

### **5.1 Operation set**

These operations are representative of problem space.

Data access based: parallelization is based off data decomp, so operations are too.

OGC standard requirements - survey of methods requiring 1, 2, or more data points.

### **5.2 Dataset Description**

Full datasets; 34k employers, 1.2m parcels

sub datasets, how to generate from full

listing of datasets used

## **5.3 Hardware setup**

Saguaro

## **5.4 implementations**

### **5.4.1 PostGIS**

### **5.4.2 Hadoop Prototype**

### **5.4.3 ClusterGIS**

# Chapter 6

## Results

### 6.1 Performance Analysis

vary procs, maintain data

vary data, maintain procs

vary data, vary procs

### 6.2 Comparisons

PostGIS as baseline

PostGIS to Hadoop

Hadoop to ClusterGIS

PostGIS to ClusterGIS

# **Chapter 7**

## **Recommendations**

Synthesis

### **7.1 Applicable Problem Spaces**

Analysis of problem types that are good/bad for this approach

# Chapter 8

## Conclusion

### 8.1 Future Work

additional decomposition methods, combined with alternative MPI communicators

- addition of preprocessing methods (indexing, etc)

- chunking of replicated dataset