FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN

Master's Thesis

Your name – `your@email.haps`

This is the report title

Very nice subtitle

Supervisor: Your Supervisor's Name

March 5th 2017

Abstract

This is an abstract abstract. Too abstract for my taste, really.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam ac sagittis mauris, vitae faucibus felis. Donec posuere massa massa, non tristique nunc vulputate ornare. Donec maximus ipsum purus, id posuere velit volutpat eu. Nunc quis tellus sem. Nulla scelerisque bibendum luctus. Proin et risus vel elit consectetur volutpat. Sed id malesuada ipsum. Donec mollis cursus massa, nec finibus augue maximus at. Ut sagittis, mauris et feugiat eleifend, ex dolor vulputate sapien, vitae venenatis nibh elit vestibulum nisi.
Maecenas varius mattis porta. Curabitur egestas erat eget dictum scelerisque. Morbi id sem quis magna dictum lobortis. Nulla dictum vel risus sit amet molestie. Aliquam fermentum lobortis semper. Aenean rhoncus viverra mi ut tincidunt. Ut fringilla risus varius nisi dictum, eu accumsan ex elementum. Sed porttitor odio sit amet iaculis tristique. Integer finibus accumsan ante non dapibus.
Nullam eros enim, feugiat eget orci non, rhoncus posuere mauris. Curabitur est elit, lacinia et ultrices vitae, ultricies et augue. Sed massa lacus, egestas nec aliquam non, interdum id odio. Quisque sapien augue, pellentesque a rutrum ac, consequat non urna. Praesent vulputate, mi et sollicitudin tincidunt, turpis nisi hendrerit quam, fringilla sodales ligula elit in diam. Integer tincidunt dolor ac purus placerat sollicitudin. Proin sollicitudin mauris quis semper pulvinar. Mauris a varius sem, tempus vestibulum tellus.
Nam nunc justo, sagittis at leo ut, molestie aliquet libero. Praesent tristique dictum vestibulum. Nunc elementum risus sed leo rhoncus venenatis. Praesent mollis, elit vitae auctor fermentum, ante nisl ultricies sapien, at volutpat orci ex vitae nibh. Quisque dictum accumsan felis, vitae condimentum nulla pharetra nec. Mauris luctus elit quis odio accumsan semper. Nullam lacus nisl, imperdiet nec lacus vel, gravida lobortis metus. Nullam venenatis est a risus iaculis pulvinar. Cras suscipit, nulla sed gravida ornare, turpis leo pulvinar ipsum, vitae luctus elit enim vel purus.

Contents

Chapter 1

Introduction

Motivation

This is the movivation!

Problem

Uhhhh, big problems.

Structure and contributions

The report is structured in this and that way.
The three main contributions of our work, listed in the order of appearance, are:

1. Provide example of thesis written in markdown
2. Provide build script
3. More contributions

Chapter 2

Figures and stuff

In this section, we introduce the background knowledge required to understand the rest of this report. We begin with the introduction of global selections. We conclude the chapter with an introduction of the system level performance techniques that we hope can improve the performance of global selections, indexing, and parallelism.

2.1 Figures

These are some side-by-side figures.

Listing

This is a listing

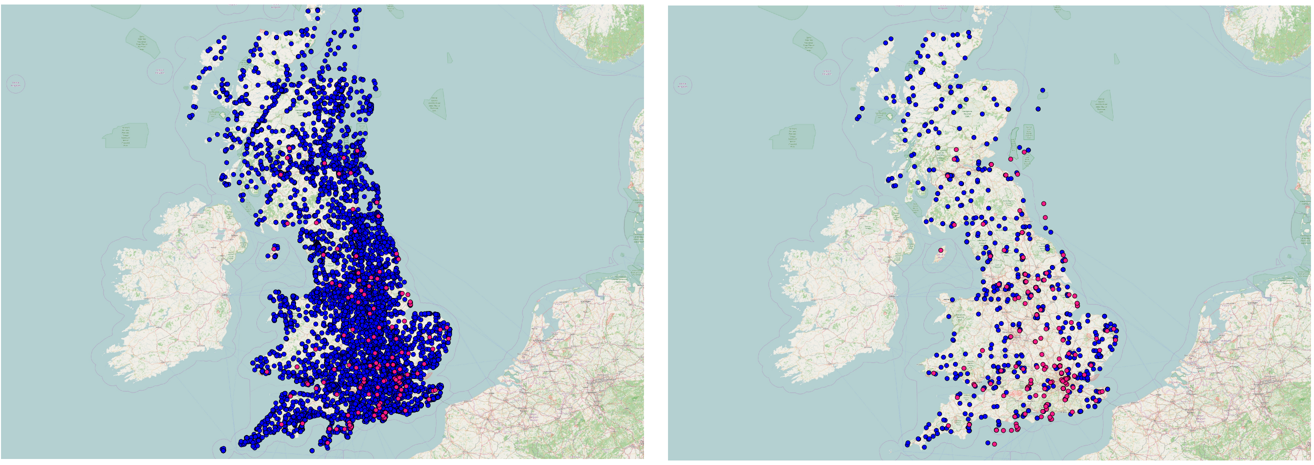
Table

This is a simple table.
This is a more fancy table.

References

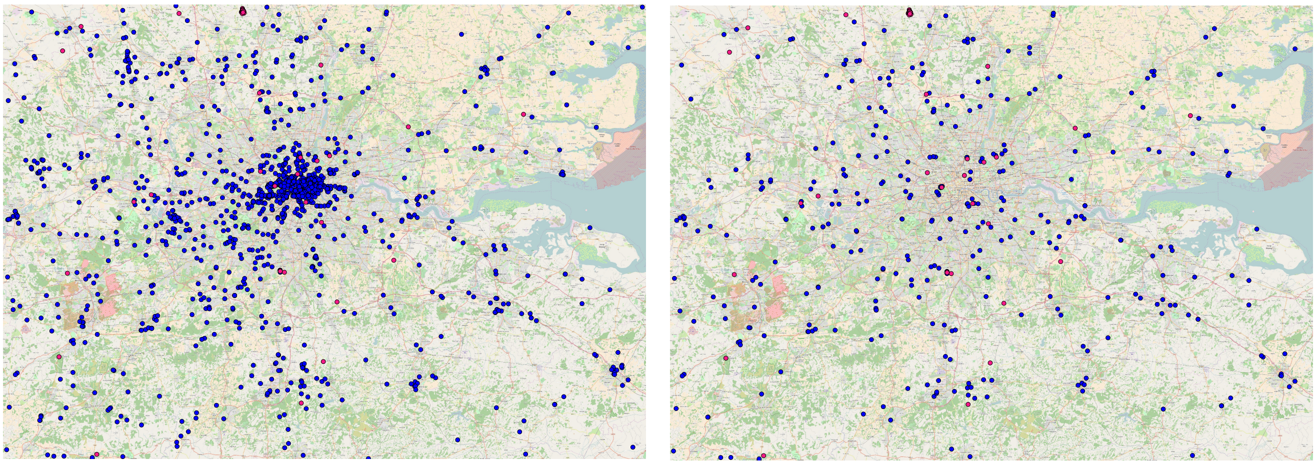
We reference the literature [1].
Reference literature by page [2, pp. 729–730].
By page and multiple refs[1], [2, pp. 729–730].

We make a reference to lines 7-14 in Listing 2.1.



(a) (b)

Figure 2.1: **(a)** Raw hotel (blue) and animal park (pink) data from OpenStreetMap project. **(b)** Result of global selection (shown on Listing 2.1) on the same data.



(a) (b)

Figure 2.2: **(a)** Raw hotel (blue) and animal park (pink) data from OpenStreetMap project, zoomed on the London area. **(b)** Result of global selection (shown on Listing 2.1) on the same data.

This is a reference to Figure 2.3.

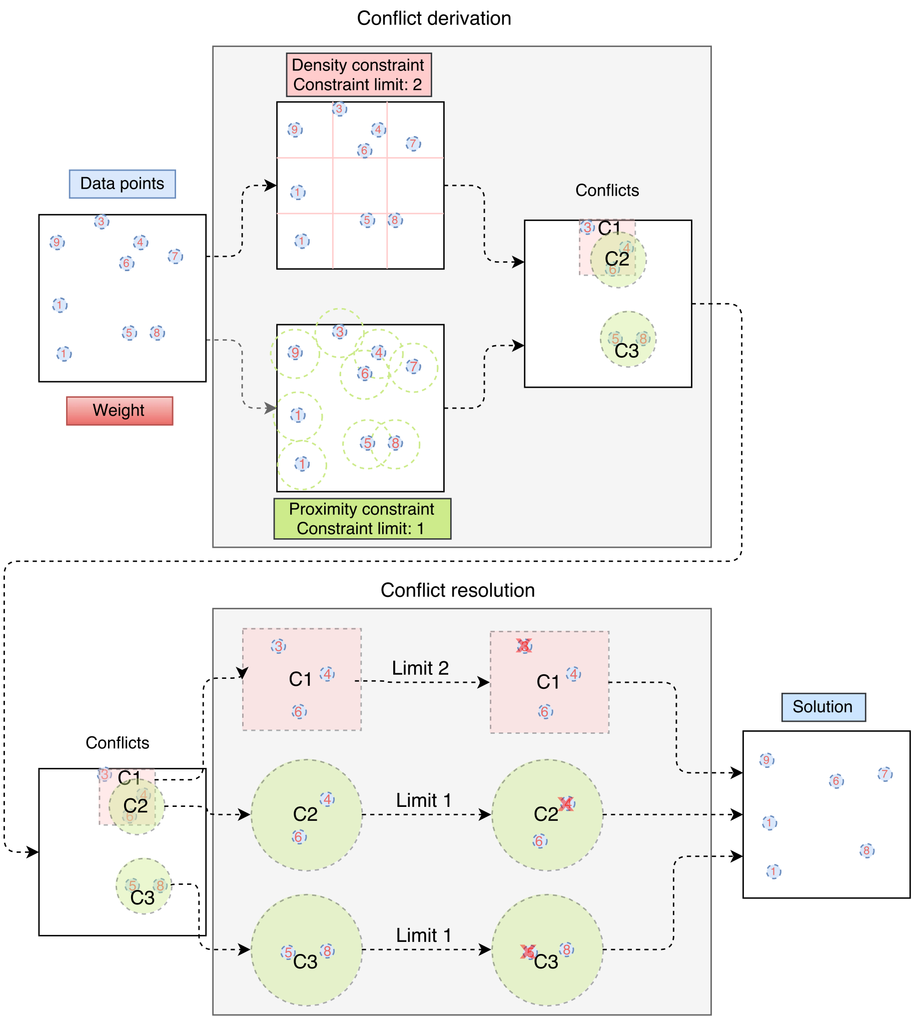


Figure 2.3: This is a single figure

Listing 2.1 Global selection formulation of query for finding hotels that are within 1 km of public transportation, preferring hotels that are closer to animal parks, tiling hotels according to zoom level, allowing at most three hotels per zoom level-appropriate tile.

```
SELECT osmid, name, lon, lat
FROM uk_hotels WITH KEY osmid
ITERATE OVER RANGE FROM 18 TO 1 AS zoom
          -- Prioritize hotels with animal parks nearby
WEIGH BY (999999 - distance_to_closest_animal_park(geom))
SUBJECT TO
  -- Remove hotels not close to public transport
  MAP BY (SELECT array_agg(osmid) ids
          FROM uk_hotels
          WHERE osmid NOT IN (
              SELECT h.osmid
              FROM uk_hotels h, uk_transport p
              WHERE ST_DWithin(h.geom, p.geom, 1000))
  )
  LIMIT 0
AND
  -- Ensure uncluttered map, at most 3 hotels per tile
  MAP BY (SELECT array_agg(osmid) ids
          FROM uk_hotels
          GROUP BY TileOnZoomLevel(geom, zoom)
  )
  LIMIT 3
```

Reference section Section 2.1.

Reference chapter Chapter 2.

Reference appendix Section A

Formatting and stuff

Math $\frac{10^6}{10^5} = 10$
bold italic `code`

1. list

Listing 2.2 A Python implementation of the Haversine formula, for calculating the distance between two geospatial points.

```
def haversine_distance(p1, p2):
    diff_lat = p2.lat - p1.lat
    diff_lon = p2.lon - p1.lon

    a = (sin(diff_lat/2)^2 +
         cos(p1.lat) * cos(p2.lat) +
         sin(diff_lon/2)^2)
    c = 2 * atan2(sqrt(a, sqrt(1-a))
    return 6371000 * c
```

Table 2.1: The factor of difference between the running time of the density constraint using Python and Scala

Tuples	Scala	Python	Factor of difference
488k	1.25	3.32	2.65
976k	1.13	3.58	3.17
1.9m	1.29	4.28	3.32
3.9m	2.14	8.44	3.95
7.8m	3.37	13.84	4.10
15.6m	5.95	25.31	4.25
31.2m	12.01	49.25	4.10
62.4m	22.30	96.71	4.34

2. item2
3. item3

This is a line.
This is one kind of line break.

This is another kind of line break.

Table 2.2: Summary of traits implemented by each of the systems considered

System	Partitioning	Parallelism	Indexing
Greenplum	Non-spatial	Distributed	Spatial/R-tree
Hadoop-GIS	Spatial	Distributed	Spatial/R-tree
Oracle	Non-spatial	Distributed	Spatial/R-tree/Quadtree
PostgreSQL	Non-spatial	Restricted single-machine	Spatial/R-tree
Simba	Spatial	Distributed	Spatial/R-tree

Appendix A

This be appendix A

A.1 Single-machine experiments

PostgreSQL

- `shared_buffers` 30GB
- `temp_buffers` 4GB
- `work_mem` 50GB
- `max_wal_size` 5GB
- `autovacuum` off

Greenplum

- `max_wal_size` 5GB
- `gp_vmem_protect_limit` 6GB
- `autovacuum` off

We ensure that the data is evenly distributed by using hash-partitioning on the primary key.

Simba

- `spark.driver.memory` 30g
- `spark.executor.memory` 80g

- `spark.python.worker.memory` 1g
- `spark.driver.maxResultSize` 0
- `SPARK_WORKER_INSTANCES=1`
- `SPARK_WORKER_CORES=16`

A.2 Distributed experiments

PostgreSQL

- `shared_buffers` 15GB
- `temp_buffers` 4GB
- `work_mem` 40GB
- `max_wal_size` 5GB
- `autovacuum` off

Simba

- `spark.driver.memory` 15g
- `spark.executor.memory` 40g
- `spark.python.worker.memory` 1g
- `spark.driver.maxResultSize` 0
- `SPARK_WORKER_INSTANCES=1`
- `SPARK_WORKER_CORES=8`

Appendix B

This be appendix B

B.1 Some section

Some subsection

TEXT!!!

References

[1] Wikipedia, "Set cover problem." 2016.
[2] Q.-S. Hua, D. Yu, F. C. M. Lau, and Y. Wang, "Exact algorithms for set multicover and multiset multicover problems." in *ISAAC*, 2009, vol. 5878, pp. 34–44.