

# TDDE31 Lab1

Kristian Sikiric (krisi211) & Pontus Svensson (ponsv690)

1)

Highest:

(u'1975', 36.1)

(u'1992', 35.4)

(u'1994', 34.7)

...

(u'1960', 29.4)

(u'1950', 29.4)

(u'1998', 29.2)

(u'1951', 28.5)

(u'1965', 28.5)

(u'1962', 27.4)

Lowest:

(u'1990', -35.0)

(u'1952', -35.5)

(u'1974', -35.6)

(u'1954', -36.0)

...

(u'1967', -45.4)

(u'1987', -47.3)

(u'1978', -47.7)

(u'1999', -49.0)

(u'1966', -49.4)

a)

Highest (with station number):

(u'1975', (36.1, u'86200'))

(u'1992', (35.4, u'63600'))

(u'1994', (34.7, u'117160'))

...

(u'1950', (29.4, u'75040'))

(u'1960', (29.4, u'173810'))

(u'1998', (29.2, u'63600'))

(u'1951', (28.5, u'75040'))

(u'1965', (28.5, u'116500'))

(u'1962', (27.4, u'76380'))

Lowest (with station number):

(u'1990', (-35.0, u'166870'))

(u'1952', (-35.5, u'192830'))

(u'1974', (-35.6, u'166870'))

(u'1954', (-36.0, u'113410'))

...

(u'1967', (-45.4, u'166870'))

(u'1987', (-47.3, u'123480'))

(u'1978', (-47.7, u'155940'))

(u'1999', (-49.0, u'192830'))

(u'1966', (-49.4, u'179950'))

b) Non-parallelized: For small file; 267 seconds, for big file; 2120 seconds.

The spark version was considerably faster, 86 seconds for the small file and a couple of minutes for the big file. The spark version uses parallel computation which probably is the reason it is faster than the non-parallelized program.

2)

Number of readings for each month:

(u'2008', u'11', 1663)

(u'1970', u'01', 1)

(u'1971', u'06', 47448)

(u'1989', u'12', 32)

(u'1966', u'11', 213)

...

Number of distinct readings:

(u'2004', u'06', 153)

(u'2009', u'04', 16)

(u'1973', u'10', 24)

(u'1984', u'04', 14)

(u'1991', u'09', 69)

...

3)

(u'1971', u'10', u'71470', 8.901612903225805)

(u'1992', u'07', u'178850', 9.80483870967742)

(u'1990', u'03', u'76160', 6.104838709677419)

(u'1963', u'12', u'97200', -2.693548387096774)

(u'2010', u'04', u'93520', 3.961290322580644)

...

4)

The result for this assignment was empty.

5)

(u'2008', u'05', 4.082352941176464)

(u'1998', u'07', 15.029411764705932)

(u'1996', u'12', 6.979411764705853)

(u'1993', u'04', 0.0)

(u'1996', u'07', 14.829411764705902)

...

6)

(u'02', u'1974', 4.413315683772325)

(u'02', u'1955', -2.0995875420341283)

(u'02', u'2002', 5.0617027805465185)

(u'02', u'1991', 1.0407350386110341)

(u'02', u'1986', -4.857652058163159)

...

The following plot shows the difference in temperature between the average temperature for a month and the long-term average for that month.

