# Real-Time Large-Scale Data Analytics and Information Retrieval in Practice

Aleksandar Bradic, Igor Bogicevic

December 25, 2009

# Contents

# Chapter 1

# Introduction

## 1.1 Enter Real Time

## 1.2 Problems, Pitfalls and Challenges

# Chapter 2

# The nature of large-scale data

## 2.1 Data Archives

## 2.2 Data Streams

# Chapter 3

# The challenges of real-time information processing

## 3.1 Problem description

# Chapter 4

# Fundamental Algorithms in Data Analytics and IR

## 4.1 Statistical analysis framework

### 4.1.1 Regression analysis

### 4.1.2 Forecasting

### 4.1.3 Parameter estimation

### 4.1.4 Non-parametric methods

# Chapter 5

# Advanced Algorithms

## 5.1 Online learning algorithms

## 5.2 Kernel Methods

# Chapter 6

# Software toolkits for large-scale data analysis

## 6.1 Hadoop

## 6.2 Mahout

# Chapter 7

# Large-scale IR Cookbook

## 7.1 Building AVMs on vertical data

## 7.2 Model selection in the real world

# Chapter 8

# Moving from batch to real-time

## 8.1 Paradigm shift

# Chapter 9

# Real-world real-time applications

# Chapter 10

# Algorithms and Data Structure in support of large-scale real-time framework

## 10.1 Convolutional procedures

### 10.1.1 Example : Viterbi algorithm

## 10.2 Convolutional representation of fundamental algebraic operations

### 10.2.1 Average,Mean,Median,Variance

### 10.2.2 Matrix operations

## 10.3 Randomized Algorithms

### 10.3.1 Fast vs. Convolutional

## 10.4 Queue-based structures

# Chapter 11

# VoidBase : queue-based computing framework

## 11.1 Overview

## 11.2 Paradigms

# Chapter 12

# VoidBase cookbook

## 12.1 Zero-development dynamic resource monitoring framework

## 12.2 Automatic trend detection toolkit

## 12.3 Building automated news-based algorithmic trading app

# Chapter 13

# Future challenges in Real-Time Large-Scale analytical processing

## 13.1 Representation problem

## 13.2 Fundamental limits