



Business Requirements Document

System that can support real-time analytics for HB data

Prepared by:

Fabien Gadet , Pierre Saint-Sorny

12/23/17

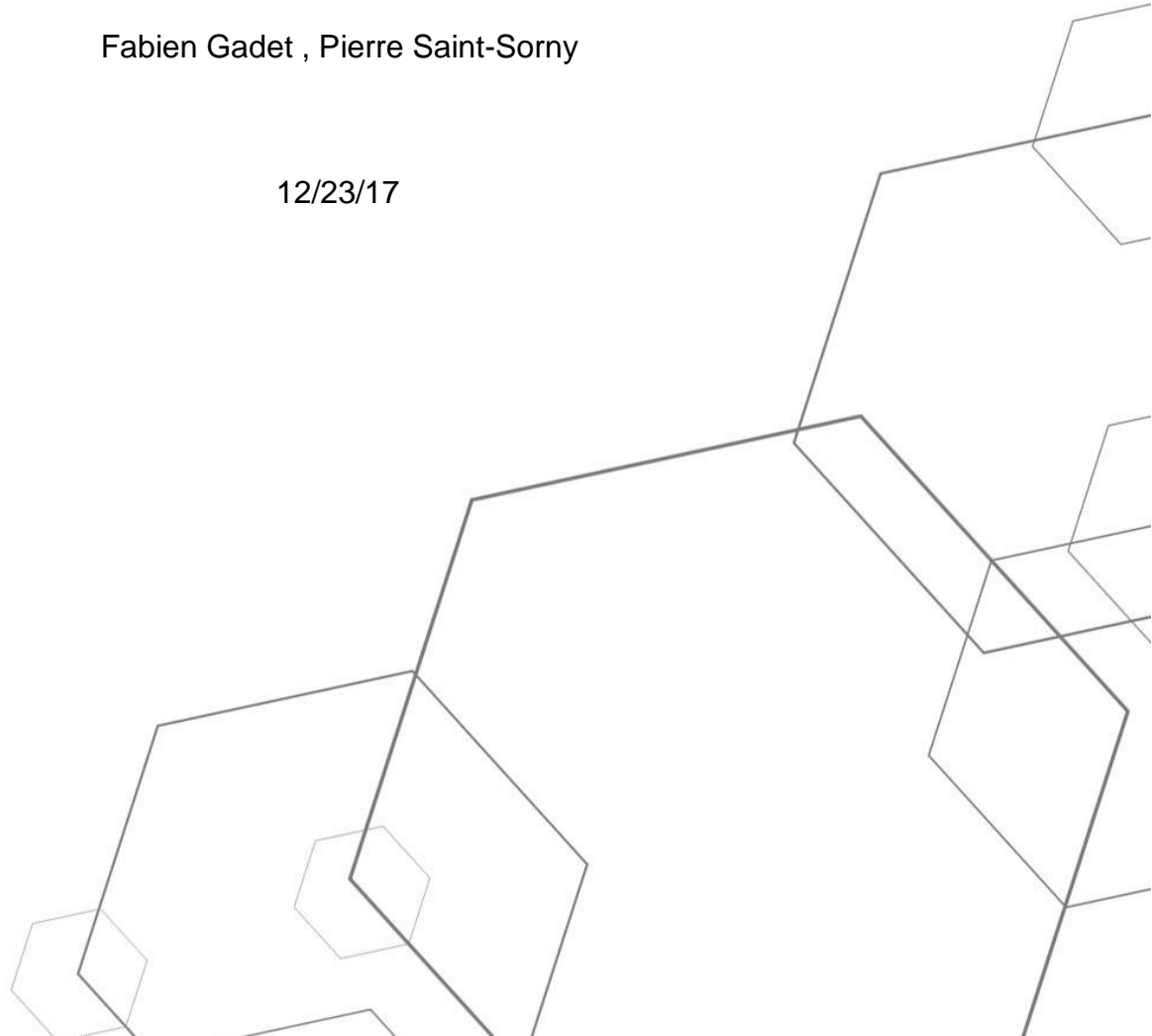


Table of Contents

Table of Contents

1. Executive Summary

2. Background

3. Features

4. Functional requirements

5. Delivery schedule

Appendix A - Glossary of Terms

1. Executive Summary

The goal of this system is to compute a large amount of data in real-time from HB data using Spark

2. Background

School Project

3. Features

Calculate SD1 and SD2 via Spark, and send out alarm is SD1 and SD2 is out of range

Read the text files of hear beat data from HDFS Can
support 100 paralleled submission of the jobs

Optionally :

Web based user interface

Use Hbase or Spark Streaming as the data input

Use the real-distributed environment(at least 2 machines)

Develop a calculation engine based on the queue system

To collect the SD1 and SD2 for different person, and set the range of SD1 and SD2 to (mean – std, mean + std) as the range

4. Functional requirements



5. Delivery schedule

6 January 2018

Appendix A - Glossary of Terms

Spark : Apache Spark is a fast and general engine for big data processing, with builtin modules for streaming, SQL, machine learning and graph processing. Spark

Streaming : same as spark but in real time (like a stream)

hbase : [open-source](#), [non-relational](#), [distributed database](#) modeled after [Google's Bigtable](#) and is written in [Java](#). It is developed as part of [Apache Software Foundation's Apache Hadoop](#) project and runs on top of [HDFS \(Hadoop Distributed File System\)](#), providing Bigtablelike capabilities for Hadoop.

HDFS : [Hadoop Distributed File System](#)