

doi:10.15199/48.2016.03.14

## Implementacja cyfrowego filtra wygładzającego w środowisku chmurowym

**Streszczenie.** Artykuł przedstawia wyniki badań eksperymentalnych implementacji cyfrowego filtra wygładzającego Savitzky'ego-Golaya w środowisku chmurowym z wykorzystaniem języka programowania R. Dzisiejsze systemy pomiarowe stosowane w urządzeniach nadprzewodnikowych są w stanie rejestrować ogromne ilości danych eksperymentalnych. Tego typu układy generują liczbę punktów pomiarowych przekraczającą możliwości obróbki w komercyjnych środowiskach eksploracji danych. Rozwiązania chmurowe dedykowane do obróbki dużych ilości danych są w takich przypadkach pewnym rozwiązaniem. W artykule dokonano porównania wyników badań dla implementacji filtra w środowisku chmurowym oraz w komercyjnym rozwiązaniu klasy Enterprise.

**Abstract.** The article presents the results of experimental implementation of a digital Savitzky-Golay smoothing filter in the cloud environment using a R programming language. Today's measurement systems used in superconducting devices are capable of recording huge amounts of experimental data. Such systems generate a number of measurement points exceeding the processing capabilities of commercial data mining environments. Cloud computing solutions dedicated to processing large amounts of data in such cases are some way out. Comparison of test results for the implementation of the filter in the cloud environment and in commercial enterprise-class data mining system was presented. (*Digital Smoothing Filter Implementation in Cloud Computing Environment*).

**Słowa kluczowe:** filtr Savitzky'ego-Golaya, eksploracja danych, chmury komputerowe, programowanie w R.

**Keywords:** Savitzky-Golay filter, data mining, cloud computing, R programming.

### Wstęp

Systemy eksploracji danych odgrywają bardzo ważną rolę w branży IT oraz nauce. Popularność tych systemów jest wynikiem szerokiego zastosowania metod eksploracji danych w różnych dziedzinach. Odkryta wiedza jest wykorzystywana w optymalizacji procesów biznesowych, wykrywania nieprawidłowości i osobliwości, przewidywania przyszłych zdarzeń i zachowań, jak również do analizy ogromnej ilości danych pomiarowych [1-8]. Aplikacje eksploracji danych w cyfrowym przetwarzaniu sygnałów można lokować w następujących dziedzinach [9-12]:

- klasyfikacja – gdzie można rozróżnić fragmenty sygnału zgodne z opracowanym modelem i zdefiniowanymi klasami,
- grupowanie – wiąże wykryte segmenty w klasy (taksony, klastry) z wykorzystaniem specjalnych algorytmów (k-mediana, k-średnia, itp),
- segmentacja – może rozdzielić sygnał na fragmenty o różnych rozmiarach, na ogół posiadających jednorodną właściwość,
- eksploracja i analiza sekwencyjna – może znaleźć logiczne prawidłowości w strukturze sygnału i relacje między różnymi zarejestrowanymi zdarzeniami, jest również możliwe znalezienie wzorców czasowych, które charakteryzują naturę sygnału,
- wizualizacja wielowymiarowa – jej zaletą jest wykreślenie n-wymiarowego zestawu danych o określonym celu końcowym mającym określić ich bliskość lub oddalenie.

Środowisko R bardzo dobrze nadaje się do obliczeń statystycznych i technik eksploracji danych. Ponieważ R przechowuje dane obliczeniowe w pamięci, jest ograniczone do dostępnej dla użytkownika pamięci operacyjnej RAM. Obliczenia w chmurze pozwalają na rozszerzenie możliwości środowiska R, poprzez uruchomienie go na znacznie lepszym sprzęcie, jak również wykorzystaniu obliczeń zrównoleglonych. Następujące projekty rozszerzają możliwości środowiska R wykorzystując ideę przetwarzania w chmurze [9]:

- pakiety RevoScaleR oraz RHadoop opracowane przez Revolution Analytics – firma oferuje swoje rozwiązania jako serwis do wykupienia w chmurze Amazon EC2 oraz przez zakup licencji [13],

- Oracle R Enterprise – zapewnia zestaw narzędzi pozwalających obsłużyć duże ilości danych w rozwiązaniach Oracle [14],
- Renjin – jest oparty na interpreterze JVM dla języka R i pozwala programistom wdrożyć swój kod do chmur typu PaaS dostawców, takich jak Google App engine, Amazon czy Heroku [15],
- RAmazon – zestaw pakietów do interakcji z chmurą Amazon [16,17].

Są też inne prace związane z cyfrowym przetwarzaniem sygnałów, systemami eksploracji danych i technikami przetwarzania w chmurze [9-12,18-20]. Najbardziej zbliżone badania do tematyki prezentowanego artykułu są też prowadzone przez Adama Laiacano w [10]. Jednak te badania wykonane zostały w środowisku Scalding, które jest rozszerzeniem biblioteki Java Cascading obsługującej niskopoziomą abstrakcję środowiska Hadoop.

W tym artykule zaprezentowano implementację wygładzającego filtra Savitzky'ego-Golaya opracowaną w środowisku Hadoop z wykorzystaniem środowiska R. Realizacja filtra została wykonana w prywatnym systemie eksploracji danych, ale może być bezpośrednio przeniesiona do chmury publicznej, w której zainstalowano platformę Hadoop. Filtr zastosowano do danych pomiarowych pochodzących z systemu akwizycji rejestrującego napięcia i prądu w układzie zawierającym taśmę nadprzewodzącą wykonaną z nadprzewodnika wysokotemperaturowego. Systemy pomiarowe urządzeń nadprzewodnikowych prądu przemiennego często generują ogromne ilości danych. Zwykle ważne jest, aby przetwarzać uzyskane dane niemal w czasie rzeczywistym, w celu utrzymania stabilności pracy urządzenia i procedur związanych z quenchem. Wyniki otrzymane z realizacji filtra w środowisku RHadoop zostały porównane z wynikami pochodzącymi z komercyjnego środowiska eksploracji danych NI Diadem. Zaletą opracowanego filtra jest to, iż posiada bezpośrednie wsparcie w technologiach chmurowych. Technologie chmurowe korzystają z rozwiązań wirtualizacji w celu osiągnięcia korzyści zwielokrotnienia przetwarzania oraz przechowywania danych. Wirtualizacja jest narzędziem do budowy wydajnych oraz efektywnych infrastruktur działających w modelu przetwarzania w chmurze.

Chmury komputerowe oferują szerokie możliwości przetwarzania dużej ilości danych, jak również nowoczesne implementacje algorytmów obliczeniowych [5,9]. Przykładem takiego rozwiązania jest implementacja paradygmatu MapReduce w środowisku chmurowym Cloudera. Dzięki wykorzystaniu implementacji tego paradygmatu możliwe są efektywne obliczenia dużej liczby danych pomiarowych (tzw. Big data).

### Konfiguracja platformy testowej

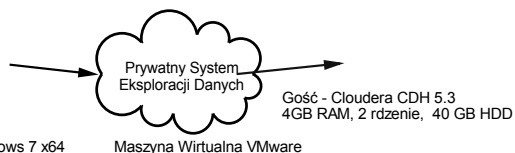
W celu realizacji cyfrowego filtra Savitzky'ego-Golaya w prywatnym systemie eksploracji danych wykorzystano dystrybucję chmurowego systemu Cloudera CDH 5.3. System Cloudera pozwala zarówno na budowę klastrów obliczeniowych jak i prywatnych chmur. Dzięki temu, użytkownik systemu wykorzystuje zasoby w taki sam sposób jak w chmurze publicznej oferowanej na przykład przez Amazon. Pozwala to na łatwą integrację tego rozwiązania w projektach chmur hybrydowych [21-23]. Hadoop, który jest częścią dystrybucji systemu Cloudera składa się z następujących elementów [24,25]:

- HDFS – Hadoop Distributed File System jest rozproszonym systemem plików, który został zaprojektowany do funkcjonowania na powszechnie dostępnym sprzęcie, jak również zawiera biblioteki zapewniające mechanizmy dostępu do zasobów rozproszonych,
- Hadoop YARN – szkielet aplikacji wykonany do zarządzania rozproszonymi zasobami i obliczeniami (równoważenie obciążenia, mechanizmy odzyskiwania systemu po awarii, planowanie procesów obliczeniowych),
- Hadoop MapReduce – implementacja przetwarzania danych zgodnie z paradygmatem MapReduce przy użyciu szkieletu YARN,
- Hadoop Common – zestaw narzędzi i udogodnień do obsługi innych modułów.

Prywatny system eksploracji danych został zaimplementowany przy użyciu maszyny wirtualnej zawierającej dystrybucję systemu Cloudera Linux z Apache Hadoop (CDH) w wersji CDH 5.3 [26].

Architektura platformy testowej została przedstawiona na rysunku 1. Elementy sprzętu oraz oprogramowania zostały przypisane w następujący sposób:

- Komputer gospodarza (PC – CPU - AMD Athlon X2 240, 2 rdzenie, 6 GB RAM, 250 GB HDD SATA), system operacyjny gospodarza Windows Professional x64, hipernadzorca VMware Player v.7.0 z zainstalowanymi VMware Tools,
- Gość – maszyna wirtualna z systemem operacyjnym Cloudera CDH 5.3.0.0 (Centos 6.4 x64) oraz 4 GB RAM, 2 rdzenie, 40 GB HDD.



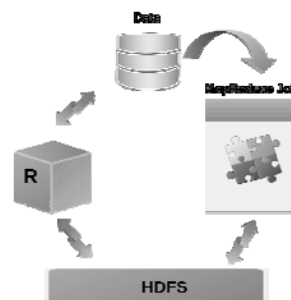
Rys.1. Architektura platformy testowej

Filtr Savitzky'ego-Golaya został zaimplementowany za pomocą języka programowania R, który to nie jest włączony standardowo do dystrybucji Cloudera. Z tego powodu dodatkowe kroki konfiguracyjne były wymagane aby móc wykorzystać zalety tego rozwiązania. Firma Revolution Analysis opracowała zestaw aplikacji łączących dla środowiska Hadoop zwanych RHadoop. Zestaw ten pozwala na manipulowanie magazynem danych Hadoop bezpośrednio z HDFS i HBase przy użyciu języka R. Dało

to programistom języka R możliwości pisania prac zgodnych z paradygmatem MapReduce z wykorzystaniem strumieni Hadoop [27,28]. RHadoop składa się z następujących pakietów środowiska R pozwalających użytkownikom na zarządzanie i analizowanie danych w platformie Hadoop, w tym tworzeniu nowych zadań MapReduce [29]:

- rhdfs – pakiet dla podstawowej łączności w systemie plików HDFS. Pakiet ten umożliwia przeglądanie, czytanie, zapisywanie i modyfikowanie plików zapisanych w HDFS wprost ze środowiska R. Powinien być zainstalowany tylko na węźle, który będzie uruchamiał klienta R.
- rhbase – zapewnia podstawową łączność HBase, przy użyciu serwera Thrift. Tabele zapisane w HBase można przeglądać, czytać, tworzyć i modyfikować z poziomu języka R. Moduł instalowany tylko na węźle, na którym będzie uruchamiany klient R.
- plymr – umożliwia wykonywanie typowych operacji manipulacji danych, jakie można znaleźć w popularnych pakietach, takich jak *plyr* i *reshape2*, na bardzo dużych zbiorach danych zapisanych w systemie Hadoop. W trakcie wykonywania zadań opiera się on na paradygmacie MapReduce. Powinien być zainstalowany na każdym węźle w klastrze.
- rmr2 – pakiet, który umożliwia wykonywanie analiz statystycznych poprzez funkcjonalności MapReduce w klastrze Hadoop. Powinien być zainstalowany na każdym węźle klastra.
- ravro – dodaje możliwość odczytu i zapisu plików w standardzie avro zarówno lokalnie jak też z systemu plików HDFS. Dodaje wsparcie avro dla rmr2 i moduł powinien być instalowany tylko na węźle gdzie będzie uruchamiany klient R.

Wykony przez firmę Revolution Analytics szkielet programowy (zwany RHadoop) umożliwia określenie specyfikacji operacji, które mają być zastosowane w eksploracji dużej ilości danych. Algorytm i dane mogą zostać podzielone i zrównoległone (Rys. 2.).



Rys.2. Główne elementy systemu RHadoop

W tradycyjnym rozwiązaniu w środowisku Hadoop, te rodzaje działalności są zapisywane w postaci pracy MapReduce w języku Java lub języku wyższego poziomu, jak Hive czy też Pig. Wykonywanie kodu R w kontekście pracy MapReduce pozwala na dodatkowe funkcjonalności takie jak filtrowanie cyfrowe, wizualizacja danych na wykresach oraz mapach [27]. W celu uruchomienia środowiska RHadoop zarówno język R jak i pakiety rozwojowe powinny zostać zainstalowane przy użyciu polecenia:

```
# yum install R R-devel
```

Kolejne elementy powinny zostać zainstalowane już z uruchomionego środowiska R. Dodatkowe pakiety należy zainstalować z repozytorium Cran przy użyciu polecenia:

```
> install.packages(c("rJava", "Rcpp",
"RJSONIO", "bitops", "digest", "functional",
"stringr", "plyr", "reshape2", "caTools"))
```

Następnym krokiem jest pobranie z serwera GitHub firmy Revolution Analytics pakietów *rhdfs* oraz *rmr2*, które będą odpowiedzialne za obsługę komunikacji z serwerem Hadoop. Cyfrowe filtry można implementować w kodzie języka R albo skorzystać z gotowych rozwiązań dostępnych w pakiecie *signal* (do pobrania z repozytorium Cran). Wyżej wspomniane moduły należy zainstalować z uruchomionego środowiska R wykorzystując instrukcje podane w [29].

### Implementacja filtru w środowisku chmurowym

Opracowana platforma testowa pozwala na szybką i kompleksową realizację cyfrowego filtru Savitzky'ego-Golaya. Po uruchomieniu środowiska R pierwszym krokiem jest inicjalizacja obsługi chmurowego systemu plików *hdfs* za pomocą polecenia:

```
> hdfs.init()
```

Dane pomiarowe zapisane do pliku w formacie CSV mogą zostać zaimportowane w postaci struktury pomiar.dane:

```
> pomiar.dane=read.csv("input.csv")
```

Obróbce można poddać dowolną liczbę kolumn, dla przykładu kolumnę zawierającą wartości chwilowe prądu i w następnym kroku przesłać dane do systemu plików HDFS

```
> I=pomiar.dane[, "I0"]
> I.index=to.dfs(I)
```

Powyższe polecenie powoduje przekazanie danych do mapera systemu Hadoop rekord po rekordzie. Następnie można te dane wygładzić za pomocą filtru cyfrowego w środowisku RHadoop używając polecenia:

```
> sg= values(from.dfs(mapreduce(input=I.index,
map=function(k,v) sgolayfilt(v))))
```

W tym przypadku wartością wejściową jest zmienna *I.index*, która zawiera obiekt typu BigData. W przeprowadzonych badaniach użyta została funkcja *map* pochodząca z paradygmatu MapReduce, a w obecnej implementacji filtru nie używano funkcji *reduce*. W procesie mapowania użyto funkcji *sgolayfilt*. Składnia tej funkcji wygląda następująco:

```
sgolayfilt(x, p = 3, n = p + 3 - p%%2, m = 0,
ts = 1)
```

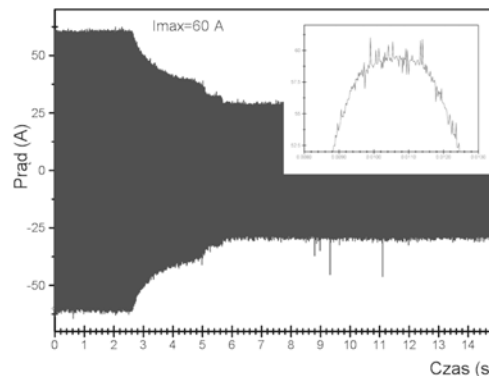
gdzie: *x* – sygnał do przefiltrowania, *p* – rząd filtru, *n* – szerokość okna (musi być nieparzysta), *m* – zwraca *m*-tą pochodną współczynników filtru, *ts* – współczynnik skalowania czasu.

Ta funkcja może być wywołana z jednym parametrem – sygnałem do przefiltrowania, wtedy pozostałe parametry przyjmują wartości domyślne (*p* = 3, *n* = 5, *m* = 0, *ts* = 1). Wyniki działania funkcji MapReduce są przechowywane w zmiennej *sg*, która jest efektem przeniesienia obiektu BigData przy użyciu funkcji *from.dfs*.

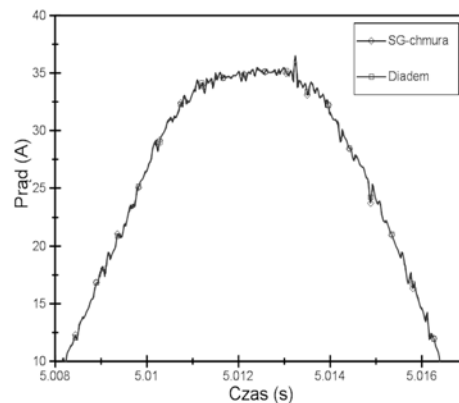
### Wyniki eksperymentu

Doświadczenie obejmowało zastosowanie opracowanego filtru Savitzky'ego-Golaya, zrealizowanego w środowisku RHadoop, do danych, które pochodziły z pomiarów a następnie porównanie wyników z wynikami pochodzącymi z komercyjnego systemu eksploracji danych pomiarowych National Instruments Diadem. W obu środowiskach zastosowano ten sam rodzaj filtru, z tymi samymi nastawami, do identycznych danych pomiarowych. Nastawy odpowiednio wynosiły: rząd filtru *p*=5, szerokość okna *n*=7. Dane pomiarowe pochodzą z układu eksperymentalnego zbudowanego i opisanego przez autorów w publikacji [30]. Idea eksperymentu pomiarowego polegała na pomiarach stabilności pracy nadprzewodnikowej taśmy drugiej generacji SF12050

wyprodukowanej przez firmę Super Power. W taśmie badano lokalny wzrost temperatury poprzez pośredni pomiar chwilowych napięć i prądu. Prąd zasilający IO o częstotliwości 50 Hz płynął przez taśmę, równoległe rejestrowano napięcia na poszczególnych segmentach taśmy. Dane rejestrowane były za pomocą karty National Instruments NI DAQ 6251 PCI z częstotliwością próbkowania równą 1000 próbek na okres.



Rys.3. Dane pomiarowe przedstawiające przebieg prądu chwilowego, wstawka – powiększenie wybranej sinusoidy prądu



Rys.4. Porównanie danych pomiarowych wygładzonych przez filtr Savitzky'ego-Golaya w środowiskach RHadoop i NI Diadem

Rejestracja danych pomiarowych z ośmiu kanałów z częstotliwością próbkowania podaną powyżej, daje 6 milionów punktów pomiarowych w czasie 15 sekund. Ponieważ eksperyment był długotrwały czasowo (do godziny) to liczba punktów pomiarowych po 12 minutach była wyższa od  $2^{31}$  (2 miliardów). Liczba ta jest limitem punktów jakie komercyjny system eksploracji danych NI Diadem jest w stanie przetworzyć [31]. Dane finalnie zapisane do pliku CSV i zaimportowane do systemu Hadoop dały plik wynikowy o rozmiarze ponad 7 GB.

Próbka danych pomiarowych zawierająca prąd chwilowy i obejmująca czas 15 sekund przedstawiona została na rysunku 3. Wstawka na rysunku pokazuje powiększenie wybranej zarejestrowanej sinusoidy. Można zauważyć, że krzywa nie jest gładka w związku z występującymi błędami cyfryzacji sygnału. Pierwszym przeprowadzonym testem było zastosowanie do danych pomiarowych filtru wygładzającego zaimplementowanego w prywatnym systemie eksploracji danych na platformie chmurowej Cludera Hadoop. Kolejnym przeprowadzonym testem było zastosowanie filtru Savitzky'ego-Golaya do tego samego zestawu danych wejściowych, ale w komercyjnym środowisku eksploracji danych NI Diadem. Filtr posiadał takie same nastawy jak w przypadku implementacji w środowisku RHadoop. Po przeprowadzeniu obliczeń wyniki zostały porównane (rysunek 4).

Analizując wartości błędu można zauważyć, że minimalna wartość wynosi około 0,5% a wartość maksymalna osiąga 1,5% (tabela 1). Wartości te są bardzo niskie i znacznie mniejsze od błędów pomiarów napięcia i prądu, które były w tym eksperymencie szacowane na 5%.

Tabela 1. Podsumowanie błędu względnego i współczynnika korelacji dla wygładzonych krzywych

Kanał	Błąd względny (%) min/max	Wsp. korelacji
I0	0,5/1,5	0,81
U1	0,4/1,6	0,8
U2	0,2/1,2	0,85
U3	0,5/1,8	0,79
U4	0,3/1,1	0,84
U5	0,35/1,2	0,83
U6	0,2/1,1	0,82
U7	0,2/1,2	0,83

## Wnioski

W artykule przedstawiono implementację filtru cyfrowego Savitzky'ego-Golaya w systemie Hadoop dystrybucji Cloudera przy użyciu zmodyfikowanego środowiska programowania R. Uzyskane wyniki porównano z wynikami pochodzącymi z systemu NI Diadem. Filtr wdrożony w środowisku testowym dał bardzo dobre rezultaty i pozwala na przetwarzanie dużych ilości danych.

Wizualizacja ogromnych zbiorów danych może dać istotne korzyści, które pomogą zrozumieć zgromadzone informacje. Dokonywanie obliczeń w środowisku R, które uruchamia zadania MapReduce, pozwala na pokonanie ograniczeń narzuconych przez komercyjnie dostępne systemy eksploracji danych. Wyniki eksperymentu wykazały, że jest możliwym, zbudowanie filtru cyfrowego w systemie eksploracji danych z wykorzystaniem języka programowania i środowiska R. Może to być bardzo odpowiednie narzędzie do analizy danych pomiarowych, zwłaszcza w układach, które wymagają zapisu dużej ilości danych (BigData) na przykład takich jak nowoczesne urządzenia nadprzewodzące.

**Autorzy:** dr hab. inż. Dariusz Czerwiński, prof. nadzw., Politechnika Lubelska, Instytut Informatyki, ul. Nadbystrzycka 36B, 20-618 Lublin, e-mail: [d.czerwinski@pollub.pl](mailto:d.czerwinski@pollub.pl)

## LITERATURA

- [1] Larose D.: Metody i modele eksploracji danych, PWN, Warszawa 2008
- [2] Gruca A.: Raport Obserwatorium ICT: Eksploracja Danych, Park Naukowo-Technologiczny „Technopark Gliwice”, 2013
- [3] Folkert K., Bochenek M., Huczala Ł.: The concept of using data mining methods for creating efficiency and reliability model of middleware applications. Communications in Computer and Information Science, pp 55-62, Springer 2012
- [4] Osowski S.: Metody i narzędzia eksploracji danych, Wyd. BTC 2013
- [5] Jestratjew A., Kwiecień A.: Using cloud storage in production monitoring systems. Communications in Computer and Information Science, s. 226-235, Springer 2010
- [6] Chudzian C., Granat J., Klimasara E., Sobieszek J., Wierzbicki, A. P.: Wykrywanie wiedzy w dużych zbiorach danych: przykład personalizacji inżynierii ontologicznej, Telekomunikacja i Techniki Informacyjne, pp 3-26, 2011
- [7] Ładniak M., Piórkowski A., Młynarczuk M.: Konstrukcja systemów eksploracji danych dla obrazów rastrowych, Studia Informatica, vol. 34 no. 2B, s. 7–20, 2013
- [8] Krauzowicz, L., Szostek, K., Dwornik, M., Oleksik, P., Piórkowski, A.: Numerical Calculations for Geophysics Inversion Problem Using Apache Hadoop Technology. Computer Networks 2012, CCIS vol 291, Springer, pp. 440-447, 2012
- [9] Ohri A.: R for Cloud Computing, An Approach for Data Scientists, Springer 2014
- [10] Laicano A.: Digital signal processing in Hadoop with Scalding, <https://github.com/alaiacano/dsp-scalding>, 2013
- [11] Ferzli R., Khalife I.: Mobile cloud computing educational tool for image/video processing algorithms, Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), pp 529 – 533, 2011
- [12] Sheng C., Zhao J., Leung H., Wang W.: Extended Kalman Filter Based Echo State Network for Time Series Prediction using MapReduce Framework, Mobile Ad-hoc and Sensor Networks, Ninth IEEE International Conference on. IEEE, pp. 175–180, 2013
- [13] Revolution Analytics, Enterprise Deployment, <http://www.revolutionanalytics.com/enterprise-deployment>, January 2015
- [14] R technologies from Oracle, <http://www.oracle.com/technetwork/topics/bigdata/r-offerings-1566363.html>, January 2015
- [15] Renjin: The R Programming language on JVM, <http://www.rnejin.org>, January 2015
- [16] RAmazonS3, <http://www.omegahat.org/RAmazonS3/>, January 2015
- [17] Amazon's Simple DB API, <http://aws.amazon.com/simpledb/>, January 2015
- [18] Kampf M., Kantelhardt J. W.: Hadoop.TS: Large-Scale Time-Series Processing, International Journal of Computer Applications, vol. 74, 2013
- [19] Li L., Ma Z., Liu L., Fan Y.: Hadoop-based ARIMA Algorithm and its Application in Weather Forecast, International Journal of Database Theory & Application, vol. 6, no. 5, 2013
- [20] Stokely M., Rohani F., Tassone E.: Large-scale Parallel Statistical Forecasting Computations in R, JSM Proceedings, 2011
- [21] Nurmi D, Wolski R, Grzegorzczak Ch, Obertelli G, Soman S, Youseff L, Zagorodnov D: The Eucalyptus Open-source Cloud-computing System, 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID), Vol. 0, 2009, pp. 124-131.
- [22] Donnelly, P., Bui, P., Thain, D.: Attaching Cloud Storage to a Campus Grid Using Parrot, Chirp, and Hadoop, IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom), pp. 488 – 495, 2010
- [23] Czerwiński D.: Influence of the VM Manager on Private Cluster Data Mining System, Computer Networks, Communications in Computer and Information Science, vol 4311, Springer, pp. 47-56, 2014
- [24] Strata 2012: Doug cutting, the apache hadoop ecosystem. <http://youtu.be/Ttu3ZQ58ovo>, June 2013.
- [25] Cloudera, the platform for Big Data, <http://www.cloudera.com>, January 2015.
- [26] CDH Version and Packaging Information – Cloudera Support, <https://ccp.cloudera.com/display/DOC/CDH+Version+and+Packaging+Information>, December 2014.
- [27] Revolution Analytics: Advanced 'Big Data' Analytics with R and Hadoop, <http://www.revolutionanalytics.com/whitepaper/advanced-big-data-analytics-r-and-hadoop>, 2011
- [28] Adler J.: R in a Nutshell, second edition, O'Reilly, 2012
- [29] Revolution Analytics, Packages in RHadoop Toolkit, <http://projects.revolutionanalytics.com/documents/rhadoop/rhad ooppkgs/>, January 2015
- [30] Czerwiński D., Jaroszynski L., Janowski T., Kozak J., Majka M.: Analysis of alternating overcurrent response of 2G HTS tape for SFCL, IEEE Transactions on Applied Superconductivity, No 3, vol. 24, 2014
- [31] National Instruments: What Limitations Exist With Channel and File Sizes in DIAdem?, <http://digital.ni.com/public.nsf/allkb/B391603F3CD86AE486256FAC00780122>, Knowledge Base, updated 2012