

# An Extensible, Scalable, Interoperable Pilot-Abstractions for Iterative MapReduce Applications on Clouds and Grids

Melissa Romanus, Andre Luckow, Pradeep Mantha, Shantenu Jha\*

*Radical Research Group, Rutgers University*

## SUMMARY

Copyright © 2012 John Wiley & Sons, Ltd.

Received ...

**KEY WORDS:** MapReduce; Grid Computing; Cloud Computing; K-Means; Data-Intensive; Compute-Intensive

## 1. INTRODUCTION

Motivation of problem in scaling data intensive applications- Interoperability, Scalability, Extensibility/Flexibility/usability. - Why is this a problem? Any real application requires this problem to be solved?

- CMS, Atlas generates PBs of data/ day.

Why Iterative Mapreduce?

What Application? ( k-means?)

Why k-means?

- twister mapreduce used k-means?

- k-means implemented using windows azure

What Infrastructure?

- FutureGrid/ XSEDE ( Sierra, Kraken )

- OSG ( Need Bliss Condor adaptor - Not yet developed )

- Eucalyptus Cloud ( Ashley's contributions ) - Get widely distributed instances.

- OpenStack Cloud ( Melissa contributions ) - Get widely distributed instances.

Experiments -

- Scale data 1GB, 10GB, 100GB, 1000 GB - Scale Resources 1000 cores, 10000 cores, 50,000 cores

Some Research Questions?

Can we say something when to use cloud or When Grid ? Does minimizing queue wait time, distributed nature of Pilot-abstractions motivate Domain Scientists to use freely available Grid resources? Waiting time and cost increases as Number of instances required increase? HOW is it beneficial than Grids?

Why Domain scientists are moving to cloud? Hype? Due to non-availability of necessary simple abstractions to scale applications on Grid?

---

\*Correspondence to: Journals Production Department, John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK.

†Please ensure that you use the most up to date class file, available from the CPE Home Page at <http://www3.interscience.wiley.com/journal/117946197/group/home/home.html>

```

\documentclass[times]{cpeauth}
%\documentclass[times,doublespace]{cpeauth}%For paper submission

\begin{document}

\runningheads{<Initials and Surnames>}{<Short title>}

\title{<Initial cap, lower case>}

\author{<An Author\affil{1},
Someone Else\affil{2}\corrauth\ and Perhaps Another\affil{1}>}

\address{<\affilnum{1}First author's address
(in this example it is the same as the third author)\break
\affilnum{2}Second author's address>}

\corraddr{<Corresponding author's address (the second author in
this example)>. E-mail: <corresponding author's email address>}

%\cgs{<Contract/grant sponsor name (no number)>}
%\cgsn{<Contract/grant sponsor name>}{<number>}

\begin{abstract}
<Text>
\end{abstract}

\keywords{<List keywords>}

\maketitle

\section{Introduction}
.
.
.

```

Figure 1. Example header text.

## 2. GETTING STARTED

### 3. THE ARTICLE HEADER INFORMATION

The heading for any file using `cpeauth.cls` is shown in Figure 1.

#### 3.1. Remarks

- (i) In `\runningheads` use '*et al.*' if there are three or more authors.
- (ii) Note the use of `\affil` and `\affilnum` to link names and addresses. The author for correspondence is marked by `\corrauth` and `\corraddr` is used to give that author's address, which will be printed as a footnote, prefaced by 'Correspondence to:'.
- (iii) For submitting a double-spaced manuscript, add `doublespace` as an option to the `documentclass` line.

```

\begin{table}
\caption{<Table caption>}
\centering
\tabsize
\begin{tabular}{<table alignment>}
\toprule
<column headings>\\
\midrule
<table entries
(separated by & as usual)>\\
<table entries>\\
.
.
.\\
\bottomrule
\end{tabular}
\end{table}

```

Figure 2. Example table layout.

- (iv) Use `\cgs` for giving details of financial sponsors; alternatively use `\cgscn` if the grant number is also to be included. These details will be printed as a footnote, with ‘Contract/grant sponsor:’ and ‘contract/grant number:’ inserted in the appropriate places.
- (v) The abstract should be capable of standing by itself, in the absence of the body of the article and of the bibliography. Therefore, it must not contain any reference citations.
- (vi) Keywords are separated by semicolons.

## 4. THE BODY OF THE ARTICLE

### 4.1. Mathematics

`cpeauth.cls` makes the full functionality of  $\mathcal{A}\mathcal{M}\mathcal{S}\mathcal{T}\mathcal{E}\mathcal{X}$  available. We encourage the use of the `align`, `gather` and `multline` environments for displayed mathematics. `amsthm` is used for setting theorem-like and proof environments. The usual `\newtheorem` command needs to be used to set up the environments for your particular document.

### 4.2. Figures and Tables

`cpeauth.cls` includes the `graphicx` package for handling figures.

Figures are called in as follows:

```

\begin{figure}
\centering
\includegraphics{<figure name>}
\caption{<Figure caption>}
\end{figure}

```

For further details on how to size figures, etc., with the `graphicx` package see, for example, [1] or [3]. If figures are available in an acceptable format (for example, .eps, .ps) they will be used but a printed version should always be provided.

The standard coding for a table is shown in Figure 2.

### 4.3. Cross-referencing

The use of the L<sup>A</sup>T<sub>E</sub>X cross-reference system for figures, tables, equations, etc., is encouraged (using `\ref{<name>}` and `\label{<name>}`).

### 4.4. Acknowledgements

An Acknowledgements section is started with `\ack` or `\acks` for *Acknowledgement* or *Acknowledgements*, respectively. It must be placed just before the References.

### 4.5. Bibliography

The normal commands for producing the reference list are:

```
\begin{thebibliography}{99}
\bibitem{<x-ref label>}
    <Reference details>
.
.
.
\end{thebibliography}
```

where `\bibitem{x-ref label}` corresponds to `\cite{x-ref label}` in the body of the article and `{99}` is the widest such number expected and determines the width of the number column in the reference list.

Please note that the file `wileyj.bst` is available from the same download page for those authors using BIB<sub>T</sub>E<sub>X</sub>.

### 4.6. Double Spacing

If you need to double space your document for submission please use the `doublespace` option as shown in the sample layout in Figure 1.

## 5. SUPPORT FOR CPEAUTH.CLS

We offer on-line support to participating authors. Please contact us via e-mail at [cpeauth-clis@wiley.co.uk](mailto:cpeauth-clis@wiley.co.uk).

We would welcome any feedback, positive or otherwise, on your experiences of using `cpeauth.cls`.

## 6. COPYRIGHT STATEMENT

Please be aware that the use of this L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub> class file is governed by the following conditions.

### 6.1. Copyright

Copyright © 2012 John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK. All rights reserved.

### 6.2. Rules of Use

This class file is made available for use by authors who wish to prepare an article for publication in *Concurrency and Computation: Practice and Experience* published by John Wiley & Sons, Ltd. The user may not exploit any part of the class file commercially.

This class file is provided on an *as is* basis, without warranties of any kind, either express or implied, including but not limited to warranties of title, or implied warranties of merchantability

or fitness for a particular purpose. There will be no duty on the author[s] of the software or John Wiley & Sons, Ltd to correct any errors or defects in the software. Any statutory rights you may have remain unaffected by your acceptance of these rules of use.

#### ACKNOWLEDGEMENT

This class file was developed by Sunrise Setting Ltd, Torquay, Devon, UK. Website:  
[www.sunrise-setting.co.uk](http://www.sunrise-setting.co.uk)

#### REFERENCES

1. Kopka H, Daly PW. 2003. *A Guide to L<sup>A</sup>T<sub>E</sub>X* (4th edn). Addison-Wesley.
2. Lamport L. 1994. *L<sup>A</sup>T<sub>E</sub>X: a Document Preparation System* (2nd edn). Addison-Wesley.
3. Mittelbach F, Goossens M. 2004. *The L<sup>A</sup>T<sub>E</sub>X Companion* (2nd edn). Addison-Wesley.