Universität Hamburg

Fachbereich Informatik

Cluster-Handbook

Projekt "Parallelrechnerevaluation" im SoSe 2016

vorgelegt von

Olga Perevalova, Kamila Mazur

bei Dr. Julian Kunkel, Dr. Michael Kuhn,

30-09-2016

Inhaltsverzeichnis

Themeneinführung 3
Die Veröffentlichung von Cluster-Handbook 4
Medienauswahl 4
Umwandlung von LaTeX - Format in Mediawiki - Format 6
Die Erweiterung von Cluster-Handbook: Spack 10
Die Erweiterung von Cluster-Handbook: Debugging 11
Fazit 12
Zeitaufwand 13
Quellen 14

Themeneinführung

Unsere Projektarbeit befasst sich mit dem Cluster-Handbook.
Mit Hilfe von Cluster-Handbook können folgende Fragen erklärt werden:

- Welche Bedeutung hat Cluster?
- Wie wird ein effektiver Cluster gebaut?
- Welche Management Optionen bietet ein Cluster?

Cluster-Handbook wurde von Studenten der Universität Hamburg geschrieben und als LaTeX - Datei bereitgestellt. Obwohl das Buch den großen Themenumfang bietet, bleiben noch und bestehen immer neue Erweiterungsmöglichkeiten von Cluster-Handbook.
Unser Projekt wird in zwei Phasen unterteilt:

- Die Veröffentlichung von Cluster-Handbook.
- Die Erweiterung von Cluster-Handbook.

Weiter wird näher auf die oben genannte Phasen des Projekts eingegangen. Beginnend mit der Veröffentlichung von Cluster-Handbook, die besteht aus der Bestimmung von Anforderungen an Veröffentlichung, Medienauswahl, Umwandlung Cluster-Handbook in geeignetes Format, darauf die Erweiterung von Cluster-Handbook um Kapitel "Spack" und "Debugging".

Die Veröffentlichung von Cluster-Handbook

Medienauswahl

Anforderungen

Eine Internet - Ressource muss bestimmte Kriterien erfüllen und Optionen anbieten, die für Veröffentlichung von Cluster-Handbook geeignet sind:

- Cluster-Handbook muss gut auffindbar und langfristig verfügbar sein.
- Cluster-Handbook muss schneller, kostenloser Zugang besitzen.
- Cluster-Handbook muss Erweiterungsmöglichkeiten für andere Nutzer anbieten.

Unsere Suche nach Medien wird durch oben genannte Anforderungen unterstützt und alle Alternative, die diese Kriterien nicht erfüllen, werden abgelehnt**.**

Wikibooks

Am 10 Juli 2003 wurde Wikibooks als Lehrbuchabteilung der Wikipedia eingerichtet.
Wikibooks ist eine Sammlung freier Lehr-, Sach- und Fachbücher unter der Share-Alike Lizenz 3.0 und GNU-Lizenz für freie Dokumentation.
Wikibooks wurde von uns als geeignetes Internet - Ressource für Cluster-Handbook ausgewählt. Unsere Entscheidung basiert sich auf folgenden Vorteilen von Wikibooks:

- Jeder kann und darf alle Bücher frei nutzen.
- Materialien werden Urheberrechtlich geschützt.
- Alle Wikimedia-Seiten können (auch ohne Einrichtung eines Benutzerkontos) gelesen, erweitert oder verbessert werden. Dies ist sehr wichtig, da jeder die Möglichkeit bekommt das Cluster-Handbook zu erweitern.
- Statistik zeigt eine große Anzahl der Wikibooks-Nutzer. Englische Version : 18.711.176 Views pro Monat[1]

[1] https://stats.wikimedia.org/wikibooks

Ein Nachteil von Wikibooks besteht darin, dass LaTeX - Format nicht automatisch in Mediawiki - Format umgewandelt wird. Es kostet Zeit, mit Hilfe von Parser ein LaTeX Datei in HTML oder Mediawiki umzuwandeln.

Alternativen für Wikibooks

- ***Eine Unterseite auf Webseite von der Fakultät Wissenschaftliches Rechnen.***
  Im Vergleich zu Wikibooks wird Cluster-Handbook sofort mit der Universität Hamburg identifiziert. Aber problematische Erweiterungsmöglichkeiten und begrenzte Anzahl der Nutzer machen diese Alternative konkurrenzunfähig.
- ***Projekt Gutenberg***[2] ist eine digitale Bibliothek, die über das Internet zugänglich ist und von Freiwilligen erstellt. Diese Alternative wird abgelehnt, da Bücher nicht weiterentwickelt werden können und hauptsächlich klassische Literatur veröffentlicht wird.
- ***E-Learning-Plattformen*** werden wegen kurzer inhaltlichen Darstellung von Materialen und Notwendigkeit der Video-Tutorials abgelehnt.

[2] http://www.gutenberg.org/

Umwandlung von LaTeX - Format in Mediawiki - Format

Parser

Parser oder Konvertierungssoftware dient zur Zerlegung und Umwandlung einer Eingabe in ein für die Weiterverarbeitung geeigneteres Format. Eine Konvertierungssoftware ist notwendig für unser Projekt, um LaTeX - Format in Mediawiki - Format zu übersetzen.

Zwei folgende Alternativen wurden von uns betrachtet:

- Wir können einen Parser selbst schreiben.
- Wir können einen bekannten Parser nutzen.

Erste Alternative zeichnet sich dadurch aus, dass selbst geschriebener Parser das Cluster-Handbook vollständig automatisch übersetzt könnte. Nach lange Überlegungen wurde diese Alternative abgelehnt, da Zeitaufwand für Lösung unsere Aufgabe mithilfe dieser Alternative viel größer würde als Übersetzung mithilfe einen bekannten Parser.

Auswahlkriterien für Parser

Folgende Anforderungen wurden für Parser bestimmt:

- Parser muss kostenfreie Nutzung anbieten.
- Einfachheit der Installation und Anwendung sind vorteilhaft.
- Direkte Übersetzung von LaTeX in MediaWiki ist notwendig.
- Parser muss Hoher Automatisierungsgrad der Übersetzung gewährleisten.

Pandoc und Alternativen

Pandoc ist ein kostenfreier Parser für Multidokumentenformate, der zur Umwandlung verschiedener dokumentbasierter Dateiformate ineinander dient.

Wegen folgender Merkmale wurde für Übersetzung der Pandoc genutzt:

- Das Programm bietet direkte Übersetzung von LaTeX in MediaWiki.
- Der Pandoc ist ganz einfach installierbar.

- Der Pandoc hat eine einfache und klare Dokumentation.
- Ungefähr 90% von Cluster-Handbook wurde automatisch umgewandelt.

Folgende Tabelle gibt eine kurze Übersicht über bekannte Konvertierungssoftware, die von uns im Prozess der Suche geprüft wurde. Aber fehlende direkte Übersetzung von LaTeX in MediaWiki und kompliziertere Installation und Anwendung machen diese Alternative konkurrenzfähig.

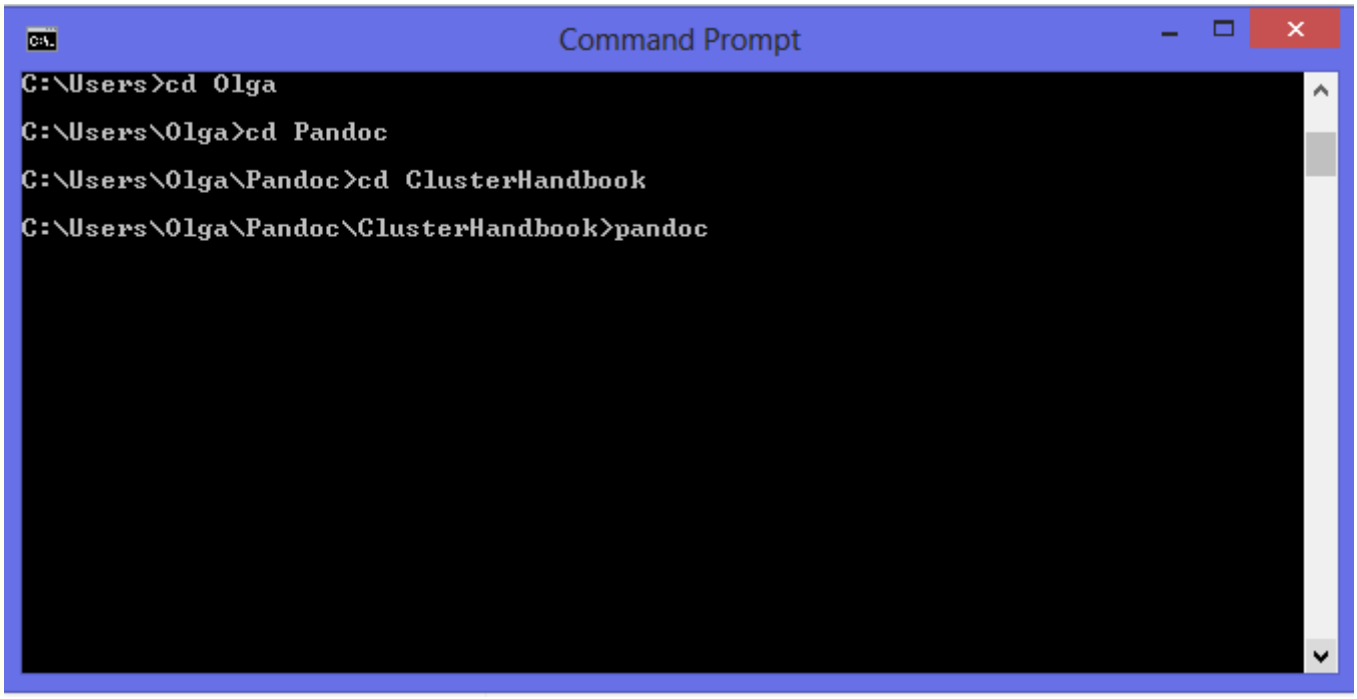
	Formate	**Betriebssystem**
LaTeX2html	LaTeX in HTML	Windows 4x, Windows NT, Linux
TeX4ht	LaTeX in HTML, XML	Windows, Mac OS X, Linux
LaTeXML	LaTeX in HTML, XML, EPUB	Windows, Mac OS X, Linux

Umwandlung mit Pandoc

Installation

Damit Pandoc auf Windows nutzen, muss man die erforderliche .msi - Datei (Link: https://github.com/jgm/pandoc/releases/tag/1.17.2) runterladen und installieren.
Damit Pandoc Dateien umwandeln kann, benötigt man die LaTeX-Bibliothek, denn die Umwandlung basiert auf LaTeX.

Pandoc bietet keine grafische Schnittstelle und nach dem Verzeichniswechsel in Kommandozeile kann die Funktionalität dieses Werkzeugs benutzt werden.



Umwandlung

Pandoc kann die Referenzen nicht automatisch übersetzten, deshalb wurde das Buch in zwei Teilen umgewandelt.

```
C:\Users\Olga\Pandoc\ClusterHandbook>pandoc clusterhandbook.tex --toc -o book1.t
xt -t mediawiki_
```

```
C:\Users\Olga\Pandoc\ClusterHandbook>pandoc software\software.tex qlistar\qlista
r.tex clustermonitoring\clustermonitoring.tex slurm\slurm.tex network\network.te
x ossec\ossec.tex munin\mu.tex snortIDS\snort.tex prelude\prelude.tex firestarte
r\firestarter.tex torque\torque.tex arm-cluster\arm.tex -o book2.txt -t mediawik
i
```

Die Erläuterung der Befehlen:

- (clusterhandbook.tex) Eingabe-Dokument
- ({toc) Generiert das Inhaltsverzeichnis in Ausgabedokument
- (-o book1.txt) Ausgabe-Dokument
- (-t mediawiki) Format des Inhalts des Dokuments

Nacharbeitung der Übersetzung

Pandoc wurde nicht speziell für Cluster-Handbook geschrieben, deshalb musste noch Zeit in die Nacharbeitung (ungefähr 10% des Dokuments) investiert werden.

Das Hauptproblem besteht darin, dass Bilder von Hand auf Wikibooks (nur als angemeldeter Nutzer) hochgeladen werden mussten.

Tabellen wurden zwar übersetzt, aber die Formatierung musste korrigiert werden.

Die Erweiterung von Cluster-Handbook: Spack[3]

Zweiter Teil der Projektarbeit befasst sich mit der Erweiterung von Cluster-Handbook um zwei Kapitel : "Spack" und "Debugging".

Kapitel "Spack" beschreibt eine Software für das Management eines Computerclusters.

Mit der zunehmenden Komplexität von HPS-Systemen (engl. High performance computing systems) steigt auch das Potenzial für Problemen und umso schwieriger wird ihre Administration. Spack ist ein Hilfsmittel. Es bietet effektive Lösungen für die Reduktion der Komplexität der Verwaltung.

Das Kapitel liefert die wichtige Information über Besonderheiten von Spack und verdeutlicht die Hauptverwendung des Werkzeugs an Beispielen.

Bevor es mit dem Schreiben dieses Kapitel angefangen worden ist, waren folgende Materialen studiert:

- Technisches Papier
- Software - Dokumentation für Spack
- Präsentationen der Spack-Entwickler

Der folgende Link führt zu Cluster-Handbook, wo sich die ausführliche Beschreibung von Spack befindet (und kann auch als Dokumentation zweites Teil des Projekts dienen):

https://en.wikibooks.org/wiki/Cluster-Handbook#Usage_2

[3] vorgelegt von Olga Perevalova

Die Erweiterung von Cluster-Handbook: Debugging[4]

Im zweiten Teil des Projekts habe ich mich mit dem Thema Linux und Debugging beschäftigt. Ich sollte zuerst mit der Linux-Umgebung klarzukommen - wie man das Betriebsytem installiert, wie man kleine Änderungen am System einrichten kann usw. Als ich ein bisschen Vertrauen mit Linux-Umgebung gewonnen habe, habe ich mich mit dem Thema Debugging vollständig beschäftigt. Ich habe 3 verschiedene Debugging-Tools ausprobiert, die unter Linux- System häufig verwendet werden:

- GNU Debugger - Standard-Debugger von Linux-Systemen. Ein kommandozeilenorientiertes Werkzeug für die dynamische Analyse von anderen Programmen um die darin enthaltenen Fehler zu finden und zu identifizieren,. Der Debugger ermöglicht dem Computerprogramm in der sogenannten StepModus laufen zu lassen, was die Analyse, Überwachung und Änderung von internen Variablen des Programms vor und nach Ausführung von den einzelnen Codezeilen erlaubt. (Die vollständige Dokumentation befindet sich im Cluster-Handbook).
- Valgrind Kommandozeilen-Werkzeugsammlung (mit einigen vorhandenen grafischen Frontends) zum Debuggen, Profilen und zur dynamischen Fehleranalyse für C/C++ - Programmem unter Linux. Da Valigrind ein Programm vollständig in einer Virtual Maschine ausgeführt, weiß er, welche Speicherbereiche bereits initialisiert, allokiert oder freigegeben wurden. Valgrind besteht aus verschiedenen Tools (Memcheck, Cachegrind Massif, Helgrind, Lackey, Nulgrind, Callgrind) die man bei dem Aufruf des Programms auswählen kann. Valgrind bietet auch Möglichkeit um weitere tools zu entwicken. (Die vollständige Dokumentation befindet sich im Cluster-Handbook).
- Das Allinea DDT Distributed - Debugging Tool ist ein graphischer Debugger, der speziell für die komplexen Anforderungen beim Debuggen von parallelen Codes entwickelt wurde.(Leider für das Thema habe ich keinen Kapital geschrieben,; ich habe es vollständig ausprobiert aber ich finde ganz richtige Erklärung befindet sich schon auf die Webseite Wissenschaftliches Rechnen.

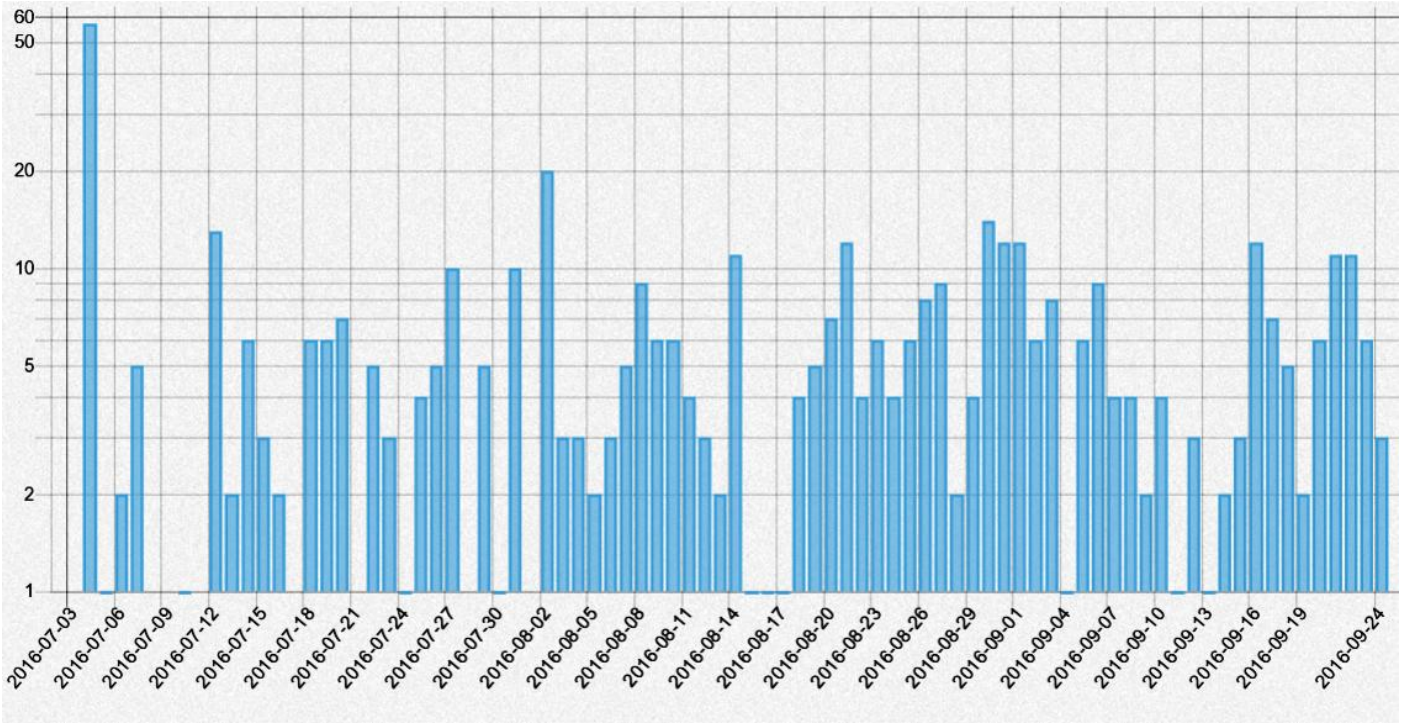
[4] vorgelegt von Kamila Mazur

Fazit

Beide Zielen des Projekts wurden erreicht. Unsere Arbeit lasst sich in folgende Phasen unterteilen:

1. Bestimmung von Anforderungen an Veröffentlichung
2. Medienauswahl
3. Auswahl von Parser
4. Umwandlung von Cluster-Handbook mit Pandoc
5. Veröffentlichung von Cluster-Handbook auf Wikibooks
6. Kapitel Spack und Debugging wurden geschrieben und übersetzt in Mediawiki-Format

Das Diagramm unten zeigt die Anzahl von Besucher, die Cluster-Handbook nutzen. Aus diesen Daten kann man schlussfolgern, dass Cluster-Handbook auf Wikibooks sich nützlich für viele Leute macht.



Zeitaufwand

	Beschreibung		Zeitaufwand
1	***Medienauswahl***	Bestimmung von Anforderungen	1 St.
2		Suche nach Medien, die entsprechende Anforderungen erfühlen	~2 Wochen
3	**Auswahl des Parsers**	Analyse der Alternativen(selbst schreiben oder einen bekannten Parser nutzen), der Bestimmung von Auswahlkriterien, Auswahlprozess	~2 Wochen
4	**Umwandlung**	Installation von Pandoc, Kennenlernen mit Dokumentation, Korrigieren des Outputs, Publizierung auf Wikibooks	**~** 2-3 Wochen
5	**Zwischenpräsentation**	Erstellung der Präsentation	~ 1 - 2 Wochen
6	**Erweiterung von Cluster-Handbook**	Das Schreiben der Kapitel: "Spack"(Olga Perevalova) "Debugging"(Kamila Mazur)	~ 5-6 Wochen
7	**Abschlusspräsentation**	Erstellung der Präsentation	~ 1 Woche

Quellen

- https://stats.wikimedia.org/wikibooks/EN/SummaryEN.htm
- https://stats.wikimedia.org/wikibooks/EN/SummaryEN.htm
- https://de.wikipedia.org/wiki/ProjectGutenberg
- http://www.gutenberg.org/
- https://de.wikibooks.org/wiki/Hauptseite
- https://github.com/LLNL/spack
- http://spack.readthedocs.io/en/latest/