

クラウド型教育環境での検索エンジン構築演習

伊東, 栄典

九州大学情報基盤研究開発センター | 九州大学大学院システム情報科学府

Flanagan, Brendan

殷, 成久

九州大学情報基盤研究開発センター | 九州大学大学院システム情報科学府

廣川, 佐千男

九州大学情報基盤研究開発センター | 九州大学大学院システム情報科学府

<http://hdl.handle.net/2324/26502>

出版情報 : 電気学会研究会資料. IS-13, pp.1-5, 2013-05-28. 電気学会

バージョン :

権利関係 :



クラウド型教育環境での検索エンジン構築演習

伊東栄典* Brendan Flanagan 殷成久 中藤哲也 廣川佐千男 (九州大学)

近年、大学における情報機器を援用する教育活動や、情報科学技術の教育で、仮想化システムやクラウドシステムの利用が進んでいる。筆者らが所属する九州大学情報基盤研究開発センターでは、2011年度末に CloudStack と VCL によるクラウドシステムを導入した。本稿では、導入した VCL システムの構成を説明し、かつ検索エンジン構築演習に用いた事例を述べる。また VCL システムを使った演習で得た知見を述べ、かつ問題点についての改善策を考察する。

A Study of search engine practice using VM on a private cloud

Eisuke Ito*, Brendan Flanagan, Chengye Yin, Tetsuya Nakatoh, Sachio Hirokawa (Kyushu University)

Kyushu University installed a private cloud system, named “campus cloud system”, using VCL and CloudStack. For the exercise of web search engine, the authors prepared a virtual machine, which is installed apache web server and GETA indexer, on the VCL. This paper introduces the outline of the exercise, and also reports advantages and disadvantages of cloud based education.

キーワード：クラウドコンピューティング、プライベート・クラウド、仮想マシン、VCL、情報演習、検索エンジン (Cloud computing, private cloud, virtual machine, VCL, exercise, web search engine)

1. はじめに

仮想化システムやクラウドシステムが普及している。近年では情報機器および情報技術を用いた大学の教育活動や情報科学技術の教育においてもクラウドの利用が検討されている。我々が所属する九州大学情報基盤研究開発センターでは、2011年度末に CloudStack と VCL から構成されるプライベート・クラウドシステムを大学院での教育および研究のために導入した。

CloudStack は、仮想マシンやストレージを提供する IaaS (Infrastructure as a Service) 型のクラウドサービスを実現するソフトウェアである。VMOps 社が VM Instance Manager の名称で開発したもので、2010年5月の cloud.com ドメイン取得に併せて、社名もソフト名称も CloudStack となり公開された。2012年9月以降、Citrix 社が提供している。

VCL (Virtual Computing Lab) は、ノースカロライナ州立大学 (NCSU, North Carolina State University) で開発された仮想マシン環境の運用システムである^{(1)~(3)}。VCL は大学現場での利用を目的として、NCSU の工学部 (College of Engineering) と情報技術室 (Office of Information Technology) が共同で2004年から開発を行った。その後、NCSU は Apache Software Foundation (ASF) に VCL のソースコードを寄附し、オープンソースとなっている⁽⁴⁾。

VCL は DaaS (Desktop as a Service) と呼ばれる形態のシステムである。管理者 (あるいは授業担当者) はアプリケーションを導入した仮想マシン (Virtual Machine, 以下

VM) のイメージを VCL 内に保存する。利用者は、Web インターフェイスを介して使いたい VM の起動を要求する。要求に応じて VCL は VM イメージを物理マシンに配信し、かつ VM を起動する。その後、利用者は起動した VM に RDP (Remote Desktop Protocol) や SSH を用いて接続し、計算機を利用する。

日本国内でも2009年頃から VCL が普及しつつある。名古屋大学では2010年度から2012年に実証プロジェクトを行なっている⁽⁵⁾。明治大学は VCL を用いたキャンパスクラウドを日本で初めて構築し、2010年度から学内に提供している^{(6)~(8)}。名城大学も VCL システムを導入・提供している⁽⁹⁾。我々の所属する九州大学では2010年に情報系の大学院である科学研究府で⁽¹⁰⁾、2011年に情報基盤研究開発センターで導入している^{(11)~(13)}。

我々は九州大学の情報系大学院であるシステム情報科学府で Web 検索エンジンに関する講義をしてきた。2011年まで講義形式での授業を行なっていた。座学講義だけでは不足するため、ペーパーテストの導入や、学生による調査および発表形式を導入していた。しかしながら実例を伴わないため理解度が低く、それにより受講生の満足度も低いものであった。クラウドシステムの導入に伴い、実際に検索エンジンを作る演習を2012年の講義で導入した。

本稿では、九州大学で導入したキャンパスクラウドシステムについて、システム設計の理念、システム構成を紹介する。次に検索エンジン構築演習について、講義の目的、講義内容、受講生から得たアンケート結果と講義の改善策を紹介する。最後にクラウドシステムを教育で用いる場合

の利点や課題を述べる。

2. 九州大学キャンパスクラウドシステム

ここでは、九州大学キャンパスクラウドシステムについて説明する。本システムは、九州大学情報基盤研究開発センターが2012年3月に導入したプライベート・クラウド型システムである。

〈2・1〉 導入の経緯 本学の大学院システム情報科学府は、大学院向けの教育研究用計算機システムを従来から整備してきた。2011年度まで、大学院のサーバ室に機器を設置するオンプレミス (on-premises) 型のシステムを導入してきた。2012年度から研究院で新たに所有することをやめてクラウド型システムを利用する方針を定めた。そこで、学内の情報基盤研究開発センターが整備するプライベート・クラウドと、外部企業が提供する商用パブリック・クラウドを併用することにした。そこで、我々が所属する九州大学情報基盤研究開発センターでは大学院の教育および研究に適したプライベート・クラウドを導入した。

導入時には、最初にシステム情報科学府と情報基盤研究開発センターの教員で約半年間検討した。その結果、表1に示す4つの部分システムから構成されるシステムを導入することとした。

表1 キャンパスクラウドの部分クラウドシステム

Table 1. Sub cloud systems in Qcloud

名称	内容
サーバ用	常時利用するサーバのための IaaS 型クラウド
開発用	情報サービスやソフトウェア開発に使う IaaS 型クラウド
データ処理用	データ処理に使う計算機システム
教育用	大学院の演習・教育の際に用いる計算機のための IaaS 型クラウド

〈2・2〉 システム構成

図1に、導入した九州大学キャンパスクラウドシステムの構成図を示す。また、表2、および表3にシステムの性能諸元の一部を示す。

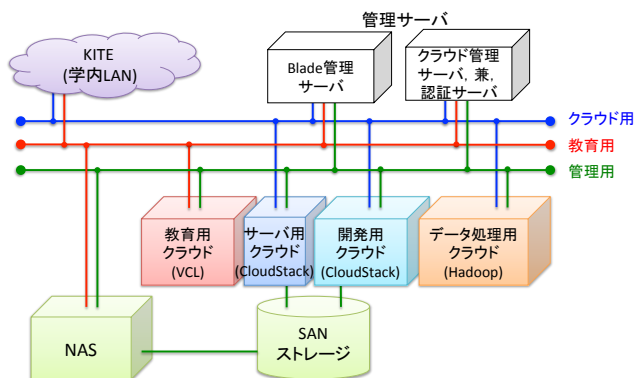


図1 キャンパスクラウド・システム構成
Fig. 1 Qcloud System Architecture

表2 キャンパスクラウドの部分クラウドシステム

Table 2. Sub cloud systems in Qcloud

システム	管理システム	Hyper visor	実コア数	メモリ量
教育用	VCL	VMware	108 (12*9 枚)	432GB (4GB*108)
サーバ用	CloudStack (ver.2)	VMware (ESXi 5)	60 (12*5 枚)	240GB (4GB*60)
開発用	CloudStack (ver.2)	Xen	108 (12*9 枚)	432GB (4GB*108)
データ処理用	CloudStack (ver.2)	なし	160 (2*80 台)	160GB (2GB*80)

表3 キャンパスクラウドの部分クラウドシステム

Table 3. Sub cloud systems in Qcloud

システム	VM Template	利用者	ログイン・アカウント	通信制限
教育用 (VCL2)	Win7, XP CentOS 5.7	教員 学生	全学共通 ID (学生 ID, SSO-KID)	なし
サーバ用	〃	教員		学内のみ (変更可)
開発用	〃	教員 学生		学内のみ (変更可)
データ処理用	CentOS 6 Hadoop 0.20	申請者	指定 ID/PW	なし

導入したキャンパスクラウドシステムは2012年10月に試行運用を開始した。提供先は2013年4月現在、学内の大学院システム情報科学府のみに提供している。今後、学内の他学部や学府 (大学院) へ提供予定である。また学外へ提供についても検討する。

〈2・3〉 VCL システムの構成

本稿では検索エンジン構築の演習を対象としている。演習には教育用であるVCL上に構築した。図2にVCLによる教育用クラウドシステムの利用方法、VM、アカウント管理、およびファイルサーバの構成を示す。

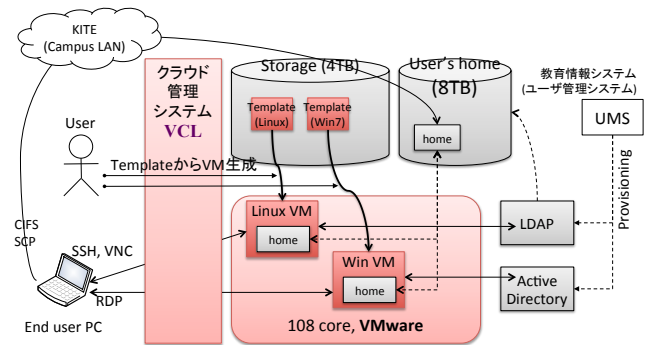


図2 VCL システム構成
Fig. 2 VCL in Qcloud

図2に有るように、教育用クラウド (VCL) ではLDAPおよびADの認証サーバ (利用者アカウント管理サーバ) を組み込んでいる。LDAPおよびADには上位の利用者アカウント管理DBから提的にデータを同期する仕組みを

実現している。そのため、学内の構成員は全学共通の ID・パスワードで仮想マシンにログインでき、新たなアカウント情報を覚える必要は無い。

また、各利用者用のホーム領域を用意しており、領域は Windows VM および Linux VM からマウントされるようにしている。講義演習終了後に VM を削除しても、講義や演習で作成したプログラム、データ、設定ファイル等を残るようにしている。ファイルサーバは学内からの CIFS アクセスを許しているため、学生は個人所有 PC から簡単にファイルのやり取りが可能である。

3. WebCGI および検索エンジン作成演習での VCL 利用

本稿の主題である、クラウドおよび仮想マシン環境を利用した検索エンジン作成演習について述べる。

〈3・1〉対象講義 本稿が対象とする講義は、九州大学大学院システム情報科学府で行なっている「分散システム特論」で、著者の廣川が主たる担当である。

「分散システム特論」は、Web に代表される広域ネットワーク環境を前提とした分散アプリケーションについて教えている。また、広域に分散したデータを収集、分析、統合、活用するための処理原理と方法を講義している。2011年度まで、Web の基本技術である HTML, URI, HTTP 等の技術を講義してきた。また具体的な実用システムとして HTTP サーバ、クライアント、Web クローラー、検索エンジンについて講義してきた。

2012年度からキャンパスクラウドシステムが稼働したため、2012年度の講義から Web 検索エンジンの構築演習を行うこととした。

〈3・2〉Web 検索エンジン構築演習 Web 検索エンジンを作成するには、前処理としての検索対象データの収集、データ整形および索引 (Index) 作成が必要である。図 3 に前処理の概要を示す。次に Web インターフェイスを介して、索引を利用する検索システムを構築する。図 4 に検索システムの概要を示す。

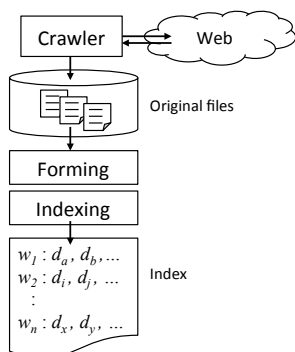


図 3 前処理 (データ収集・整形・索引作成)

Fig. 3 Preprocessing for IR system

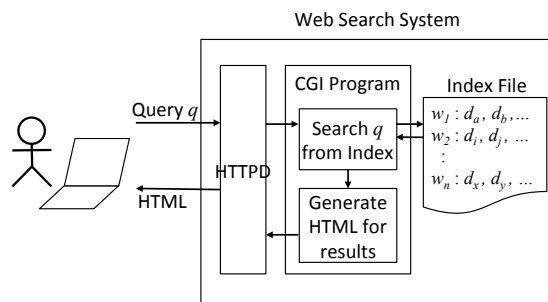


図 4 Web 検索システム

Fig. 4 Web search system

この演習を行うに当たり、Apache Web サーバ ver.2.0, 形態素解析器 chasen, 汎用連想検索エンジン GETA ver.2 を導入した VM を用意した。汎用連想計算エンジン GETA は「Generic Engine for Transposable Association」の略で、情報処理振興事業協会 (IPA) が実施した「独創的情報技術育成事業」の研究成果物である。

VCL の機能により、講義担当教員は講義に合わせた仮想マシンイメージを更新することが出来る。CentOS 6.2 の基本機能だけをインストールした仮想マシンを起動し、Apache Web サーバをインストールした。また、マシン起動時に Apache も起動するように設定した。学外 IP アドレスからの Web アクセスを許可するように iptables の設定を変更した。また、ユーザ用 Web サイトでの CGI を許可するように設定した。Web に関する設定を終えた後、GETA を適切な場所にインストールした。

これらの Apache や GETA のインストールおよび設定は、実用的な検索システム作成には必要な機能であるものの、講義したい情報検索の本質とは関係が薄い。必要な機能を仮想マシンのイメージとして準備することで、各学生が準備する必要が無い。最初から出来上がった環境を提供するため、講義では本質部分だけの演習に専念できた。教員の側も、実マシンとは違い、仮想マシンは設定を失敗してもイメージ破棄すれば再構築が可能であるため、気軽に講義専用の環境を準備することができた。

〈3・3〉演習の手順 2012年度の演習では、以下の手順で講義を進めた。

- (1) VCL での VM 利用方法説明と、Linux 利用および Perl 言語でのテキスト処理 (1回)
- (2) Perl プログラムを用いての練習用データからの WAM ファイル作成 (1回)
- (3) 汎用連想検索エンジン GETA と、Perl 言語からの使い方の説明 (2回)
- (4) Web CGI プログラムの説明 (1回)
- (5) 実データ集合を使った検索エンジン作成 (2回)
- (6) 構築結果の発表 (1回)

上記(2)の練習用には「小説を読もう (syosetu.com)」サ

イトが提供する YAML 形式の小説メタデータを 100 ファイル提供した。これらは教員が予め集めておいたもので、当時の人気の上から 100 位までのデータを選んだ。最後の(5)の演習前に、追加データとしての 2004 年から 2011 年の間に電子情報通信学会の研究会で発表された 42921 件の論文概要データを提供した。提供にあたり、単語の形態素解析や発表年抽出などの前処理をした形式で準備した。

検索エンジン構築演習では、学生は「小説を読もう」のメタデータか、電子情報通信学会・研究会論文のデータのどちらかを選び、それを元に検索エンジンを構築した。自分でデータを集めての検索エンジン構築も許していたものの、2012 年度の演習ではデータを集めた学生は居なかった。

最後に各学生が自分の作成した検索エンジンについて発表を行った。発表にあたり、自分が自主的に工夫した点を述べさせるようにした。

4. 考察

〈4・1〉学生の評価アンケート 授業内容を評価するため、受講生に簡単なアンケート調査を行った。調査項目は以下の 7 項目で、それぞれを 5 点満点で 10 人の受講生に評価してもらった。点数評価の結果を表 4 に示す。

- (a) VCL の使いやすさ
- (b) VCL で動かす VM の利用時間
- (c) Linux VM の使い勝手
- (d) SSH 接続での Linux 利用方法の使いやすさ
- (e) CGI 用の Web サーバの使いやすさ
- (f) 汎用連絡検索エンジン GETA の使いやすさ
- (g) Perl 言語と GETA の使いやすさ

表 4 講義評価アンケート結果

Table 4. Score of enquate

		回答者										
		1	2	3	4	5	6	7	8	9	10	平均
設 問	(a)	4	4	5	3	5	4	4	5	5	4	4.3
	(b)	2	3	4	1	4	2	4	5	4	4	3.3
	(c)	4	3	5	5	4	4	3	5	5	5	4.3
	(d)	4	3	4	5	4	4	4	5	5	5	4.3
	(e)	3	4	5	3	4	4	3	2	5	4	3.7
	(f)	3	4	3	4	4	3	3	5	4	3	3.6
	(g)	3	4	4	5	4	3	4	4	4	4	3.9

分かりやすさのため、中間である 3 点を引いた値にしたものを表 5 に示す。

表 5 講義評価アンケート結果 (3 を引いた値)

Table 5. Score of enquate (-3)

		回答者										
		1	2	3	4	5	6	7	8	9	10	平均
設 問	(a)	1	1	2	0	2	1	1	2	2	1	1.3
	(b)	-1	0	1	-2	1	-1	1	2	1	1	0.3
	(c)	1	0	2	2	1	1	0	2	2	2	1.3
	(d)	1	0	1	2	1	1	1	2	2	2	1.3
	(e)	0	1	2	0	1	1	0	-1	2	1	0.7
	(f)	0	1	0	1	1	0	0	2	1	0	0.6
	(g)	0	1	1	2	1	0	1	1	1	1	0.9

〈4・2〉アンケート結果の考察 学生から集めたアンケート結果について考察する。まず全体的な評価を行い、その後に詳細について考察する。アンケートでは意見を自由に記述する部分を作っていた。その部分を読むと、受講学生の満足度は高い講義になっていることが分かる。後にも述べるが、検索エンジン構築の課題は、学生が自主的に長時間取り組もうとする課題であった。

次に表 4 および 5 で評価が低い設問(b), (e), (f)について考察する。まず、設問(b)の VCL システムの使いやすさについては、自由記述意見欄に仮想マシンの時間制限への不満が書かれていた。九州大学キャンパスクラウドの VCL では、仮想マシンの利用時間を VCL の default 設定値である 4 時間までに設定している。利用延長は出来るものの、最長 8 時間までしか使えない。検索エンジン構築の作業に長い時間が必要であるものの、VCL 側の設定制限から作業が中断されてしまう事が問題となった。VCL 側の設定は問題があるものの、学生が自主的に長時間取り組もうとする課題であった部分は評価している。実際、課題に対して興味を持ったとの意見が多くの子から出ていた。

設問(e)の CGI 用 Web サーバの使いやすさの評価が低いのは、仮想マシンの準備不足のためであろう。本演習は 2012 年度が最初であるため、演習講義の前半まで Apache の設定や permission 設定が十分ではなかった。講義後半には十分な使える環境を用意できたものの、前半の混乱が受講学生に使いにくい印象を与えたと考えている。

設問(f)の GETA の使いやすさが少し低い評価であるのは、説明の複雑さにあると思われる。まず、情報検索での文書ベクトルモデルおよび索引概念についての説明がある。その上で索引作成ツールの使い方を理解する必要がある。理解が難しい部分があるため、使いにくい印象を持ったと考えられる。

〈4・3〉VM およびクラウドの効果 仮想マシンおよびクラウドで演習を行うことは、全体的に効果が高いと感じている[16]。一つ目の利点は、計算機環境の整備のしやすさである。従来のように物理的計算機で演習する場合、計算機環境を授業毎にカスタマイズすることは、他の講義への影響、保守機構の問題、セキュリティの問題から困難であった。一方 VCL 上の仮想マシンを使う場合、気軽に本演

習講義だけの環境を用意することができた。ただし、授業担当教員側は、講義用の環境整備のための作業時間が必要であった。なれない初めての作業であるため、作業に数日を要した。

検索エンジン構築演習用の計算機環境 (VM イメージ) を整備したが、他の様々な環境を整備することは作業量的に困難である。例えば情報系の教育に限定しても、画像処理に適した VM、統計処理に適した VM、並列計算に適した VM などが必要で、これらの多様な VM イメージを用意するには多数の作業者が必要である。VM イメージには、定期的なセキュリティ対応も必要である。VM イメージの種類を増やすには、国内あるいは世界中で協力する体制が有ると良い⁽⁴⁴⁾。ライセンスの問題が無ければ、VM イメージの共有が可能であろう。あるいは、機械的に VM イメージ作成する仕組みも望ましい。

プライベート・クラウドを学生が自主的に使う VCL の利点として、計算機資源の有効活用がある。本稿では受講生も 10 名程度の大学院講義を説明したが、普通の学部では 50 名を超える講義も多い。従来の物理的な PC を使う講義では計算機資源を厳密に管理して割り当てる必要があった。VCL では必要な時に必要な量の計算機数を構築できるため、資源を有効に活用することが可能である。

二つ目の利点として、root 権限での作業を演習できたことが有る。演習用の環境が十分に整備できていなかったため、演習講義時間に Apache の起動や停止、設定ファイル `httpd.conf` を書き換えた。情報系の学生は将来計算機システム管理者としての業務を担当する可能性もあるため、管理作業の教育は有効である。実マシンで root 権限作業を行うと、設定を間違えると計算機が動かなくなる場合もあるため、受講生に作業をさせることはなかった。実マシンと異なり、VCL で作成した VM であれば、失敗しても VCL から VM を消して再度新しい VM を作れば良いため、気軽に root 権限での作業を行うことが出来た。

5. おわりに

本稿では、情報系の大学院修士課程における、クラウドシステムを用いた検索エンジン構築演習について報告した。導入したクラウドシステムを説明し、検索エンジン構築演習の講義内容を述べた。実際にクラウド内の仮想マシン上に検索エンジンを構築する演習を行った。検索対象となるデータは、予め教員が収集したデータを提供した。

受講生からのアンケートをまとめた所、受講生は非常に高い満足度を得ていることが分かった。アンケート結果の分析から、VCL システムの設定変更で講義を改善できることが分かった。またクラウドシステムを教育で用いる場合の利点を述べた。

今後はクラウド型の教育システムの規模を拡充し、情報系以外の学部や大学院における教育にも活用することを考える予定である。そのためのシステム整備内容を検討する。さらに、学内のみならず、国内他大学や世界の大学へのクラウドシステム提供を考えていきたい。

文 献

- (1) NCSU VCL, <http://vcl.ncsu.edu/>.
- (2) Henry E. Schaffer, Samuel F. Averitt, Marc I. Hoit, Aaron Peeler, Eric D. Sills, and Mladen A. Vouk: NCSU's Virtual Computing Lab: A Cloud Computing Solution, *Computer*, vol.42, issue 7, pp.94-97 (2009)
- (3) Sarah R. Stein and Henry E. Schaffer: Cloud with a Long Tail: The VCL in Support of Pedagogy, *EDUCAUSE Review*, vol. 45, no. 3 (May/June 2010), pp.10-11 (2010)
- (4) Apache VCL, <http://vcl.apache.org/>.
- (5) 梶田将司: 仮想コンピューティング実験室によるクラウド型教育学習支援環境の構築, *信学技報*, AI-110(172), pp.59-64 (2010)
- (6) 日立広報: 明治大学と日立, 米国のアプリケーションソフト提供システム「VCL」によるキャンパスクラウドを日本で初構築 (2010). <http://www.hitachi.co.jp/New/cnews/month/2010/07/0715a.html>
- (7) 齋藤孝道: 明治大学版 VCL システムの導入とその利用について, *FIT2011 学習環境のクラウド化とパーソナル化* 講演 2 (2011)
- (8) 川口謙太郎, 森皓生, 齋藤孝道: 仮想化リモートデスクトップ環境 VCL に関するパフォーマンスの計測, *情報処理学会 第 73 回全国大会講演論文集 2011(1)*, pp.405-407 (2011)
- (9) 名城大学 VCL, <http://www.wit.meijo-u.ac.jp/service/vcl.html>.
- (10) 日立広報: 九州大学大学院が授業・演習環境に「大学向けクラウド型仮想デスクトップソリューション」を導入 (2011) <http://www.hitachi.co.jp/Div/jkk/kyoiku/casestudy/kyushu/casestudy1.html>
- (11) 日立ニュースリリース: 九州大学が「キャンパスクラウド」基盤を構築し研究者・学生向けにクラウドサービスの提供を開始 (2012) <http://www.hitachi.co.jp/Div/jkk/kyoiku/casestudy/kyushu2/casestudy1.html>
- (12) 益田健, 上田将嗣, 伊東栄典: キャンパスクラウドの導入とベンチマークによる評価, 第 34 回 全国共同利用情報基盤センター 研究開発連合発表講演会 研究開発論文集, Nov.21, pp.91-97 (2012) (ISSN 0910-8769)
- (13) 伊東栄典, 堀良彰, 笠原義晃, 井上弘士: 情報系大学院での VCL 活用, *情処研報 Vol.2013-CLE-9 No.9*, pp.1-6, (2013)
- (14) 笠原義晃, 伊東栄典: 大学向けクラウド基盤における VM テンプレートの多様化, 第 1 回地域間インタークラウドワークショップ, *IIRC*, pp.65-76, (2012)