# 北海道大学ハイパフォーマンス インタークラウドの設計，構築，運用まで

杉木　章義，棟朝　雅晴

北海道大学情報基盤センターでは，2018 年 12 月に新しい学際大規模計算機システム（通称：北海道大学ハイパフォーマンスインタークラウド）を導入し，運用を開始した．他大学や他研究機関の高性能計算を目的としたシステムと比較した場合の最大の特徴は，スーパーコンピュータとクラウドの両者のシステムを併設し，一体的に設計から構築，運用までを行っている点にある．本稿では，ハードウェアおよびソフトウェアを含むシステムの構成，利用者に向けた各種の提供サービス，最近の活用事例などを明らかにし，将来的な展望についても言及する．

キーワード：クラウドコンピューティング，インタークラウド，高性能計算，仮想化，ストレージ，情報基盤

## 1. はじめに

北海道大学情報基盤センター [1] では，2018 年 12 月に学際大規模計算機システム [2] を更新し，運用を開始した（図 1）．以前の学際大規模計算機システムでは，後ほど増設されたペタバイト級データサイエンス統合クラウドストレージを含めて「北海道大学アカデミッククラウド」と名づけ，利用の促進を行ってきたが，新しいシステムでは「北海道大学ハイパフォーマンスインタークラウド」という名前に変更し，普及を図ることとなった．新世代のシステムでは，その名のとおり，高性能であると同時に，インタークラウドによる接続性や柔軟性を重視したシステムとなっている．

北海道大学ハイパフォーマンスインタークラウドは，前回のシステムからの大幅な飛躍となる総理論演算性能，約4.0PFLOPS を達成し，システム全体で約 1,300 台以上の計算機で構成され，SINET5 [3] で接続された複数の遠隔拠点にサーバやバックアップ装置を新たに設置し，北海道から九州に至るまでの全国規模の広域分散クラウドシステムとして構築されていることが特徴である．

本システムは，従来からの計算科学および計算機科学の研究を引き続き支援するとともに，ネットワークや IoT (Internet of Things) などの分散システムに関する研究にも活用されることが期待される．よって，本システムは，支援できる研究の幅・深さともに，大幅に強化されたシステムとなっている．

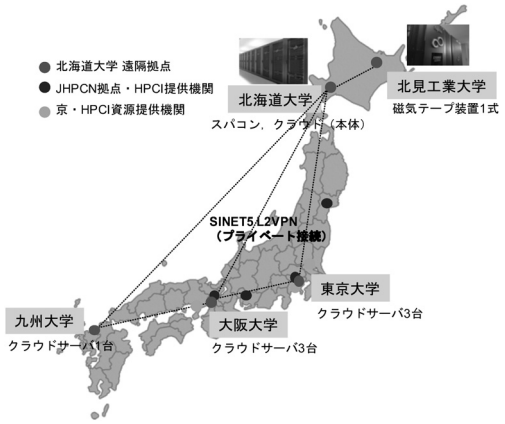


**図 1**　北海道大学ハイパフォーマンスインタークラウド

## 2. 北海道大学情報基盤センター

北海道大学情報基盤センターは，1962 年に全国共同利用施設として発足した大型計算機センターに由来し，同じく 1979 年に発足した情報処理教育センター，後の情報メディア教育総合センターと合流し，2003 年に両センターを統合して設立されたセンターである．また北海道大学内では，スラブ・ユーラシア研究センター，人獣共通感染症リサーチセンターとともに「研究センター」の位置づけにあり，研究活動を行うことが主要なミッションとして規定されている．

さらに，情報基盤センターは，学際的な共同研究・共同研究拠点の利用を推進するにあたり，8 大学で構成される学際大規模情報基盤共同研究・共同研究拠点

すぎき　あきよし，むねとも　まさはる
北海道大学情報基盤センター
〒 060–0811 北海道札幌市北区北 11 条西 5
sugiki@iic.hokudai.ac.jp
munetomo@iic.hokudai.ac.jp

(JHPCN) [4] の構成拠点の一つであり，先日運用を停止した「京」を頂点とする革新的ハイパフォーマンス・コンピューティング・インフラ (HPCI) [5] への資源提供機関の一つでもある．現在，ポスト京計算機「富岳」の運用開始までの受け皿となる第二階層システムの一つとして，日本国内外のさまざまな研究活動を支えている．

情報基盤センターでは，以上のような経緯の中で，高性能計算，情報ネットワーク，デジタルコンテンツおよびメディア教育，システムデザイン，サイバーセキュリティなどに関する研究を幅広く実施している．

## 3. システムの概要

図 2 に北海道大学ハイパフォーマンスインタークラウドの全体システム構成を示す．これまでにも述べてきたとおり，本システムはスーパーコンピュータとインタークラウドの両者を併設したシステムとなっている．インタークラウド部分に限らず，スーパーコンピュータ部分を含めて，「北海道大学ハイパフォーマンスインタークラウド」と総称している点に注意が必要である．

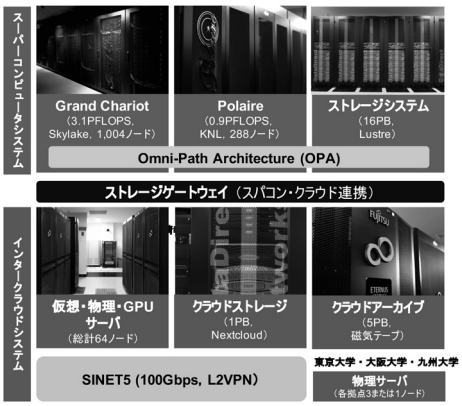


**図 2**　システム全体の概要

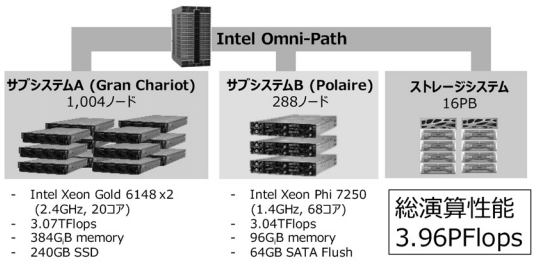


**図 3**　スーパーコンピュータシステムの概要

スーパーコンピュータシステムは，実際には「Grand Chariot」と「Polaire」の二つの異なるサブシステムで構成されており，インターコネクトを経由して，両サブシステムを支える高速かつ大容量な共用ストレージシステムが接続されている．

一方のインタークラウドシステムは，仮想マシンおよびベアメタルマシンを提供するための複数のサーバが導入されており，クラウドストレージのサービスを提供するための磁気ディスク装置，これらのバックアップを行うための磁気テープライブラリ装置が導入されている．

スーパーコンピュータとクラウドの両システムの連携はこれまでも行われていたが，さらなる緊密な連携を目的として，高速なデータ転送を可能にするストレージゲートウェイ装置が 2 台導入されている．

**表 1**　ハードウェア構成 (Grand Chariot)

| 項目 | 性能・諸元 |
|---|---|
| 製品名 | FUJITSU PRIMAGY CX400 M4, CX2550 M4 |
| プロセッサ | Intel Xeon Gold 6148 （Skylake, 2.4 GHz, 20 コア）×2 |
| メモリ | 384 GB |
| ストレージ | 240 GB SSD × 1 |
| インターコネクト | Omni-Path（2 ポート） |
| 総理論演算性能 | 3.08 PFLOPS（倍精度） |
| 総メモリ容量 | 376.5 TB |
| 総ノード数 | 1,004 ノード |

**表 2**　ハードウェア構成 (Polaire)

| 項目 | 性能・諸元 |
|---|---|
| 製品名 | FUJITSU PRIMAGY CX600 M1, CX1640 M1 |
| プロセッサ | Intel Xeon Phi 7250 （KNL, 1.4 GHz, 68 コア） |
| メモリ | 96 GB (+ 16 GB MCDRAM) |
| ストレージ | 64 GB SATA Flush Drive × 1 |
| インターコネクト | Omni-Path（1 ポート） |
| 総理論演算性能 | 0.87 PFLOPS（倍精度） |
| 総メモリ容量 | 31.5 TB |
| 総ノード数 | 288 ノード |

**表 3**　ストレージシステム

| 項目 | 性能・諸元 |
|---|---|
| 製品名 | DDN ES14KX |
| ファイルシステム | DDN EXAScaler (Lustre) |
| 総物理ストレージ容量 | 16 PB |

  

表 4　ソフトウェア構成

| カテゴリ | ソフトウェア |
|---|---|
| 言語処理系 | インテル Fortran/C/C++ コンパイラ，GCC コンパイラ，Java，Python |
| ライブラリ | インテル MKL（数値計算），MPI（通信），DAAL（データ解析・機械学習） |
| アプリケーション | OpenFOAM，PHASE，FrontFlow/red，Chainer，TensorFlow，Caffe |

## 4.　スーパーコンピュータシステム

### 4.1　ハードウェア構成

スーパーコンピュータシステムのシステム構成を図 3 に示す．詳細な計算ノードのハードウェア構成を表 1 および表 2，ストレージシステムの構成を表 3 に示す．

二つのサブシステムともに，共通の x86 アーキテクチャを採用しながら，両者が採用するプロセッサは異なっている．サブシステム A の Grand Chariot は，Skylake 世代の Intel Xeon Gold 6148 を採用したマルチコア型の計算機システムである．一方のサブシステム B の Polaire は，Knights Landing (KNL) 世代の MIC アーキテクチャに基づく，メニーコア型の計算機システムである．

### 4.2　ソフトウェア構成

両サブシステムとも同じインテル製のプロセッサを採用するため，表 4 に示すほぼ同じソフトウェア環境が共通で導入されている．Fortran，C/C++，Java，Python の各言語処理系に対応し，従来からの数値計算用途に限らず，データ解析および機械学習に対応したソフトウェアも導入されている．

また，以前のシステムとの継続性を考慮して，2 台のアプリケーションサーバも導入されており，ライセンス対象者およびライセンス契約の範囲内で，Gaussian，Mathematica，MATLAB，COMSOL Multiphysics，AVS/Express Developer，Field View，Pointwise などの商用ソフトウェアが利用可能となっている．なお，Gaussian と V-FaSTAR については，Grand Chariot でも利用可能である．

### 4.3　提供サービス

北海道大学情報基盤センターのスーパーコンピュータシステムでは，まずは「基本サービス」の契約を締結し，利用者と支払予算の組に紐づけられた「利用者番号」を取得した後，必要に応じて，演算時間やストレージ容量などの「付加サービス」を契約する負担金体系となっている．

- **基本サービス（インタークラウドと共通）**

試用・デバッグを目的とした共用ノード利用，home 領域の基本容量 (100 GB)，アプリケーションサーバの利用，クラウドストレージの基本容量 (100 GB) の提供．

- **付加サービス（サブシステム A・B ごと）**

**共用ノード利用**

演算可能時間（トークン）を消費することで，ジョブを実行する利用形態．一度に多数の計算ノードを使用する大規模並列演算に向く．

**占有ノード利用**

年間を通じて計算ノードを占有する利用形態．演算可能なノード数は限られるが，ほかのグループの利用者によるジョブ実行待ちが発生しない．

**ストレージ追加**

home 領域または work 領域のそれぞれに対して，ストレージ容量の追加が可能．

基本サービスの申し込みはスーパーコンピュータシステムにログインするユーザごとに必要であるが，学生については負担金を大幅に低減する措置が新たに導入されている．また，付加サービスについては，演算時間およびストレージ容量ともに利用者間でグループを構成し，共用利用することが可能である．なお，原則として電気代相当額の実費を負担金として利用者に求めているが，大学からの研究支援により，軽減措置が図られている．

## 5.　インタークラウドシステム

### 5.1　ハードウェア構成

インタークラウドシステムはスーパーコンピュータや以前のシステムと比較して，小規模ながら，インタークラウドによる接続する機能を大幅に強化したシステムとなっている（図 4）．東京大学，大阪大学，九州大学に本学の遠隔サイトの位置づけでサーバを設置し，北見工業大学に磁気テープライブラリ装置を設置し．遠隔バックアップを行っている．

これらの遠隔拠点との接続には，国立情報学研究所 (NII) が提供する学術情報ネットワーク SINET5 [3] を利用しており，SINET5 の高速性および低遅延性，フルメッシュ性を最大限活用している．

ハードウェアは運用性を考慮して，可能な限り，統

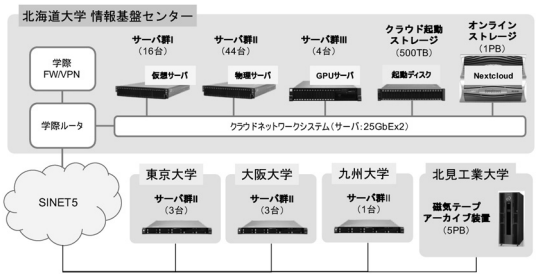


**図 4**　インタークラウドシステムの概要

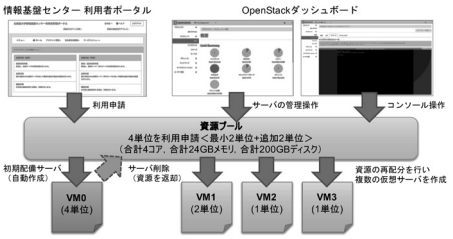


**図 5**　仮想サーバの提供

**表 5**　ハードウェア構成（仮想・物理サーバ）

| 項目 | 性能・諸元 |
|---|---|
| 製品 | FUJITSU PRIMAGY CX400 M4, CX2550 M4 |
| プロセッサ | Intel Xeon Gold 6138（Skylake, 2.0 GHz, 20 コア）× 2 |
| メモリ | 256 GB |
| ストレージ | 240 GB SSD × 2 (RAID1) |
| ネットワーク | 25 GbE × 2 |

**表 6**　ハードウェア構成（GPU サーバ）

| 項目 | 性能・諸元 |
|---|---|
| 製品 | FUJITSU PRIMAGY RX2540 M4 |
| プロセッサ | Intel Xeon Gold 6138（Skylake, 2.0 GHz, 20 コア）× 2 |
| メモリ | 256 GB |
| ストレージ | 3.84 TB SSD × 2 (RAID1) |
| GPU | NVIDIA Tesla V100（PCIe 接続，16 GB）× 2 |
| ネットワーク | 25 GbE × 2 |

一が図られている．表 5 と表 6 にサーバの詳細なハードウェア構成を示す．

記憶装置としては，各種サーバの起動用に合計 500 TB の物理容量を有する FUJITSU ETERNUS DX200，クラウドストレージの提供用に 1 PB の物理容量を有する DDN GS7K と，遠隔バックアップ用に総計 5 PB の物理容量を有するテープを搭載した FUJITSU ETERNUS LT270 (LTO) を導入している．

### 5.2　ソフトウェア構成

ソフトウェア構成としては，大きく OpenStack [6] と Nextcloud [7] の二つが導入されている．以前のシステムでは，仮想マシンの提供用に CloudStack，クラウドストレージの提供用に ownCloud が導入されていたが，それぞれ一新された．

OpenStack 環境の構築や運用は煩雑であることが知られているが，ミランティス社の Mirantis Cloud Platform [8] を活用し，導入が行われている．仮想マシンの提供も前回の XenServer から変更し，KVM [9] で行われている．OpenStack の仮想マシン提供機能は成熟しており，統一的な GUI からマシンのインストール，起動や停止，仮想コンソールが利用可能である．一方のベアメタルマシンの提供では，OpenStack Ironic を活用し，仮想マシンとの一体的な運用が可能であるが，導入バージョンの Pike であっても，未だ発展途上であり，仮想ネットワークのテナント間分離や仮想コンソールの機能が弱く，外部のソフトウェアやカスタマイズ，手動による運用に頼っているのが現状である．運用期間中に OpenStack のバージョンアップが予定されているため，改善されることを期待している．

### 5.3　提供サービス

上述のハードウェアおよびソフトウェアを活用し，下記のサービスを利用者に対して提供している．

なお，下記は 4.3 節のスーパーコンピュータの付加サービスに相当し，インタークラウドのサービスの利用にあたっても，基本サービスの契約締結が必要である．

**● 仮想サーバ**

物理サーバと同一構成のハードウェアを仮想化機能により分割し，2 コア以上，1 コア単位で仮想マシンとして提供する（図 5）．物理サーバと比較して，負担金が抑えられるため，研究予算に応じて，柔軟に利用することが可能である．また，GUI によるダッシュボードのみならず，OpenStack API も開放しているため，短期間での自動的な仮想サーバの生成や破棄も可能である．実際に Terraform や Rancher などの OpenStack 連携ソフトウェアとの接続も確認している．

標準的な OpenStack 環境と同様に「資源プール」での利用が可能であり，本学では 1 CPU コア，6 GB メモリ，50 GB ディスクの組を 1 単位として提供している．図 5 に示すように，たとえば 4 単位を契約した場合に，4 コアの仮想マシン 1 台として利用することも可能であるし，2 コアの仮想マシンを 1 台と 1 コアのマシンを 2 台などのように，複数の仮想マシンに分割

  

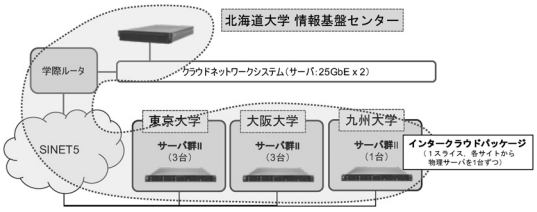


**図 6**　インタークラウドパッケージの提供

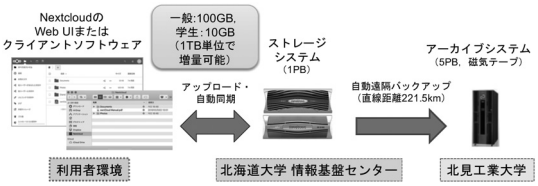


**図 7**　クラウドストレージ (Nextcloud) の提供

して利用することも可能である．

**• 物理サーバ**

物理サーバをベアメタルマシンとして提供する．サーバ性能のみならず，ネットワーク性能にも配慮されており，CentOS および Ubuntu の標準提供 OS では，2 本の 25 Gbps Ethernet をアクティブ・アクティブ接続，総計 50 Gbps 接続で利用可能である．

**• GPU サーバ**

上記の物理サーバと同様にベアメタルマシンのサービスであるが，GPU が搭載されている点が異なる．

**• インタークラウドパッケージ**

本システムの特徴的なサービスであり，北海道大学と東京大学，大阪大学，九州大学の各遠隔サイトから物理サーバを 1 台ずつ取り出し，SINET5 の L2VPN でプライベート接続して，スライスとして一括提供するサービスである（図 6）．ネットワークに関する研究や広域分散システムの研究に活用されることが期待される．

インタークラウドパッケージを含む SINET L2VPN 接続時には，インタークラウドシステムから学内ファイアウォール装置を迂回して，SINET5 に直接，物理リンク 100 Gbps で接続するバイパス線を使用する．また，ほかの遠隔サイトも各大学のキャンパスネットワークを経由するが，少なくとも上流スイッチには，それぞれ東京大学では 40 Gbps，大阪大学では総計 20 Gbps，九州大学では 10 Gbps で接続されており，非常に高いネットワーク転送性能が期待される．

**• クラウドストレージ**

Nextcloud によるクラウドストレージサービスを提供する（図 7）．基本サービス経費の負担により，一般は 100 GB，学生は 10 GB まで標準で利用可能であり，1 TB 単位でのストレージ容量の追加も可能である．Nextcloud は Web UI または利用者端末にインストールされたデスクトップクライアントから利用可能であり，Nextcloud に格納されたデータは情報基盤センターに設置されたストレージに格納された後，北見工業大学の磁気テープライブラリ装置に定期的かつ自動的にバックアップされる．

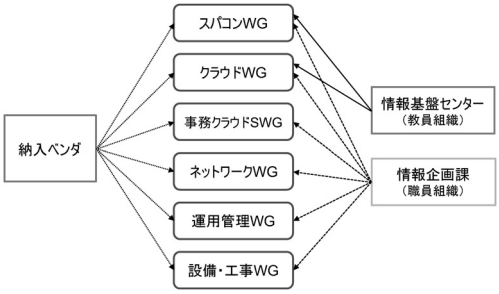


**図 8**　導入および運用体制

## 6.　設計・構築・運用体制

一般にスーパーコンピュータなどの大型計算機の調達は，非常に時間のかかるプロセスであり，本システムも設計の検討から構築，運用まで約 4 年を要している．

開札後の導入に関しては，ワークグループ (WG)，サブワークグループ (SWG) が設立され，納入ベンダ，教員および職員組織から人を出し合い，進められた（図 8）．稼働開始後も，不要となる設備・工事などを除き，ほぼ同じ WG が継続している．

途中，2018 年 9 月 6 日に発生した北海道胆振東部地震 [10] の影響を受けたものの，稼働前であったため大きな支障はなく，納入自体は順調に進んだ．インタークラウドの導入に関しては，旧システムからの仮想マシンの移行作業が大半を占めており，物理サーバの半数を移行用サーバに費やしている．今後，利用者が増加するにつれて，保守や移行作業の負荷も増加することが予想されるが，これらの軽減は今後の課題である．

利用者は順調に増加しており，特にインタークラウドシステムにおいて，2019 年 3 月時点で，すでに仮想サーバで 70%，物理サーバで 80%を超える利用率となっている．

## 7.　活用事例

導入年度となる 2018 年度は，わずか 4 カ月の期間

であったが，HPCI 採択課題で 2 件，JHPCN 採択課題で 3 件の利用があった．2019 年度はそれぞれ HPCI で 15 件，JHPCN で 6 件に大幅に増加している．

学内連携では，人獣共通感染症リサーチセンター，化学反応創成研究拠点 (WPI-ICReDD) などにより，スーパーコンピュータシステムが活用されている．

インタークラウドシステムに関しては，すでに NII が提供する SINET5 をインタークラウド実現のために最大限活用しているが，SINET5 以外の NII が提供するサービスとの連携についても模索している．

**• 学認クラウド導入支援サービス**

クラウド導入の検討や調達を進めている大学や研究機関に対して，チェックリストによる情報提供や個別相談を実施する NII のサービスである [11]．北海道大学情報基盤センターは利用者側のみならず，クラウド事業者としても参加し，情報提供を行っている．

**• 学認クラウドオンデマンド構築サービス**

クラウド上にアプリケーションの実行環境を展開する作業は，クラウドやネットワークに関する十分な理解と複雑な設定作業が必要となる．本サービスは，テンプレートベースのオンデマンド構築とネットワーク接続の支援を提供するサービスである [12]．2018 年度の 3 カ月間に実施した JHPCN 萌芽型共同研究の成果をもとに，オンデマンド構築サービスが，北海道大学ハイパフォーマンスインタークラウドに対応した．

**• SINET 広域データ収集基盤実証実験**

SINET 広域データ収集基盤は，まだ実証実験の段階ではあるが，プライベート接続された携帯電話のモバイル網を活用し，これまでの有線では到達不能であったエリアに対しても，SINET の接続性を提供するサービスである [13]．北海道大学ハイパフォーマンスインタークラウドは，大学提供のクラウドとして，実証実験に参加している．特に北海道では，広範囲な活用が期待される．

**• 研究データ管理基盤「GakuNin RDM」**

オープンサイエンスや研究データ管理計画 (DMP) の策定，研究証跡の記録などに対応し，研究データ管理機能を提供するサービスである [14]．クラウドストレージの Nextcloud は．GakuNin RDM のエクストラストレージとして接続可能であるため，長期的なトライアル実験を進めている．負担金を支払えば，学外者でも利用可能である．

## 8. おわりに

本稿では，2018 年 12 月に運用を開始した北海道大学ハイパフォーマンスインタークラウドの概要について説明した．本システムの設計，構築，運用はオペレーションズ・リサーチ (OR) の文脈では，非常に些細な問題であるが，初歩的な手法ですら，まだ活用されていないのが現状である．むしろ，OR の研究者の先生方には，本システムの豊富な計算資源を活用して OR の研究を精力的に進める，または本システムの上で稼働する仮想マシン，OS コンテナ，アプリケーション，仮想ネットワーク，さらにはパブリッククラウドとの連携などの複雑な組み合わせ問題に対して，OR の手法を適用していくなどのほうが関心が高いのではないかと思われる．

本システムは少なくとも 2023 年 11 月まで運用予定であり，今後も継続的にさまざまな改善を図っていきたいと考えている．今回のシステムはインタークラウドを謳いながら，まだ他大学が提供する遠隔サイトの活用にとどまっており，民間パブリッククラウドの活用によるハイブリッドクラウド構成，それらを複数活用したマルチクラウド構成が期待される．実際に，本システムの利用率が非常に高い状態で推移しており，運用期間中の検討が必要であろう．また，広域データ収集基盤や研究データ管理基盤を含めた，研究データ管理も大きなキーワードとなるだろう．オープンサイエンスによる研究成果の公開が期待される一方で，研究証跡や輸出規制，人権の保護および法律の尊守への対応などが求められている．

## 参考文献


[1] 北海道大学情報基盤センター，「情報基盤センタートップページ」，https://www.iic.hokudai.ac.jp/（2019 年 6 月 7 日閲覧）

[2] 北海道大学情報基盤センター，「学際大規模計算機システム」，https://www.hucc.hokudai.ac.jp/（2019 年 6 月 7 日閲覧）

[3] 国立情報学研究所，「学術情報ネットワーク SINET5」，https://www.sinet.ad.jp/（2019 年 6 月 7 日閲覧）

[4] 学際大規模情報基盤共同利用・共同研究拠点事務局，「学際大規模情報基盤共同利用・共同研究拠点」，https://jhpcn-kyoten.itc.u-tokyo.ac.jp/（2019 年 6 月 7 日閲覧）

[5] 高度情報科学技術研究機構，「革新的ハイパフォーマンス・コンピューティング・インフラ」，http://www.hpci-office.jp/（2019 年 6 月 7 日閲覧）

[6] OpenStack Foundation, “OpenStack: Build the future of open infrastructure,” https://www.openstack.org/（2019 年 6 月 7 日閲覧）

[7] Nextcloud GmbH, “Nextcloud,” https://nextcloud.com/（2019 年 6 月 7 日閲覧）

[8] Mirantis, “Mirantis Cloud Platform,” https://www.mirantis.com/software/mcp/（2019 年 6 月 7 日閲覧）

[9] KVM contributors, “Kernel Virtual Machine,” https://www.linux-kvm.org/（2019 年 6 月 7 日閲覧）

[10] 国土交通省気象庁，「平成 30 年北海道胆振東部地震


  

の関連情報」，https://www.jma.go.jp/jma/menu/20180906_iburi_jishin_menu.html（2019 年 6 月 7 日閲覧）
[11] 国立情報学研究所，「学認クラウド導入支援サービス」，https://cloud.gakunin.jp/cas/（2019 年 6 月 7 日閲覧）
[12] 国立情報学研究所，「学認クラウドオンデマンド構築サービス」，https://cloud.gakunin.jp/ocs/（2019 年 6 月 7 日閲覧）
[13] 国立情報学研究所，「2018 年度 SINET 広域データ収集基盤実証実験」，https://www.sinet.ad.jp/wadci（2019 年 6 月 7 日閲覧）
[14] 国立情報学研究所オープンサイエンス基盤研究センター，「研究データ管理基盤 (GakuNin RDM)」，https://rcos.nii.ac.jp/service/rdm/（2019 年 6 月 7 日閲覧）