

# 京都大学新スパコンと クラウド環境

深沢 圭一郎

京都大学学術情報メディアセンター



# 京大スパコンとクラウド

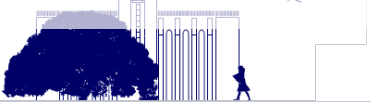
## 京大現行スパコンとクラウド連携

現在の京大スパコンは、2016年にリプレイスされたシステム。

- このシステムには**クラウドに関連する機能はなかった**。

一方で、京大には、汎用コンピュータと呼ばれるシステムがあり、これが2017年にリプレイスされた際に、アカデミッククラウドというオンプレミスとクラウドの構成になった（クラウドはパブリッククラウド京大外のデータセンタにおいたサーバ群）。

- Fujitsu PRIMERGY RX2530 M2×32ノード
- このオンプレ部分が**スパコンと10Gbpsで接続**された。
- オンプレ上のVMにスパコンストレージをマウントするとまあまあ早い（sshfs）。
- 仮想サーバホスティングとして、VMを資源提供をしていた。



# 現行京大スパコン

2016~2022(7/28に稼働停止)

## Camphor 2 (System A)

### CRAY XC40

 Xeon Phi KNL 68cores 1.4GHz x 1 /node  
#nodes = 1,800  
#total cores = 68 cores x 1,800 → 122,400 cores  
Peak performance = 3.05TFlops x 1,800 → 5.48 PFlops  
Memory capacity = (96+16 GB) x 1,800 → 196.9 TB  
Burst buffer = 230 TB, 200 GB/sec 



## Storage

### DataDirect NETWORKS ExaScaler (SFA14K)

Disk capacity = 24 PB  
Bandwidth = 150 GB/sec  
Burst buffer = 230 TB, 250 GB/sec 



高速通信網 InfiniBand EDR/FDR

高速通信網 Omni-Path

## Laurel 2 (System B)

### CRAY CS400 2820XT

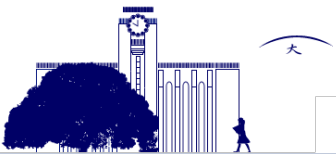
 Xeon Broadwell 18cores 2.1GHz x 2 /node  
#nodes = 850  
#total cores = 36 cores x 850 → 30,600 cores  
Peak performance = 1.21 TFlops x 850 → 1.03 PFlops  
Memory capacity = 128 GB x 850 → 106.3 TB



## Cinnamon 2 (System C)

### CRAY CS400 4840X

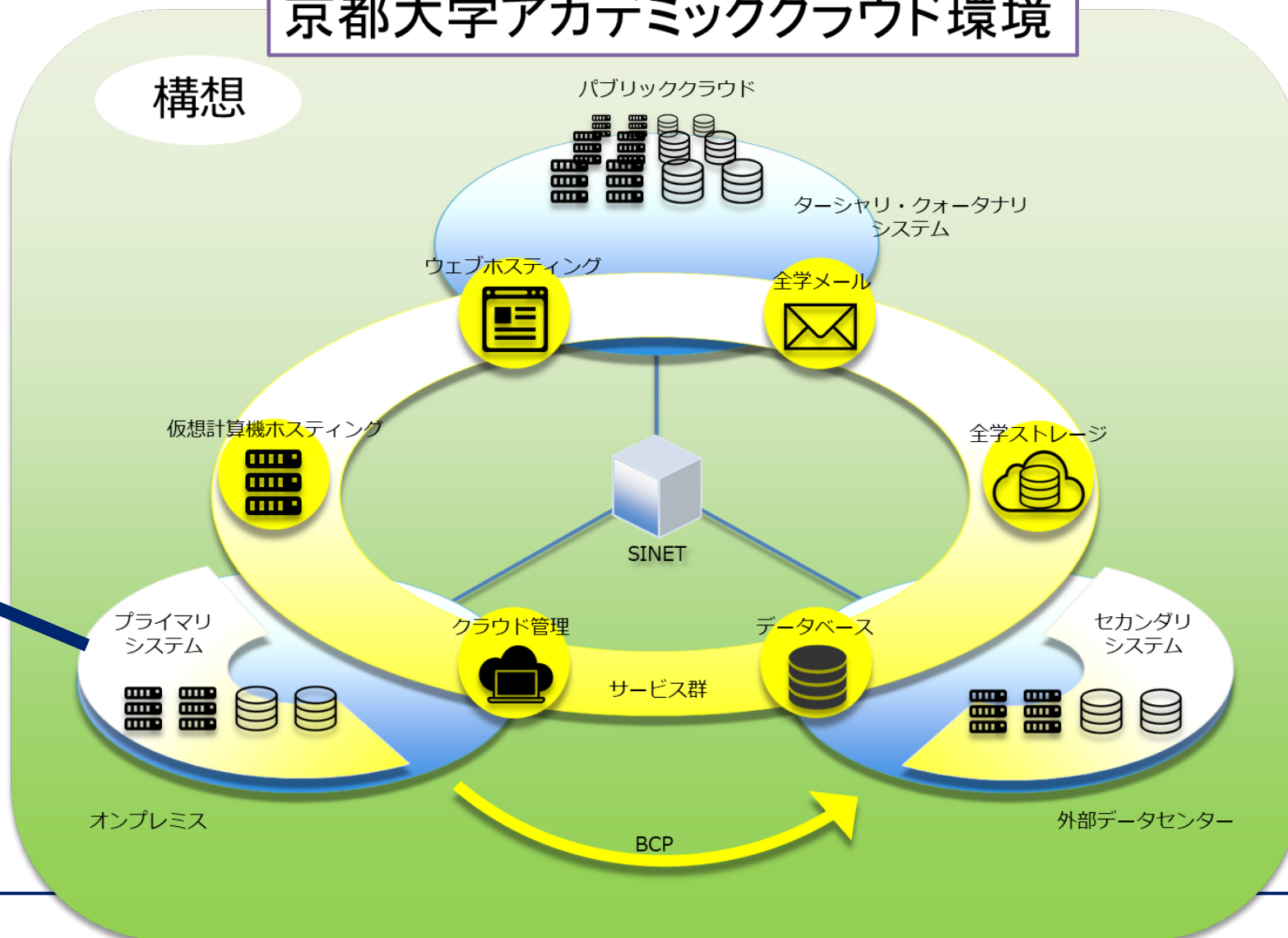
 Xeon Haswell 18cores 2.3GHz x 4 /node  
#nodes = 16  
#total cores = 72 cores x 16 → 1,152 cores  
Peak performance = 2.65 TFlops x 16 → 42.4 TFlops  
Memory capacity = 3 TB x 16 → 48.0 TB



# 汎用コンピュータ(2017~2021)

## 京都大学アカデミッククラウド環境

構想



スパコン



# 汎用コンピュータからアカデミッククラウドへ

## 汎用コンがアカデミッククラウド(ARCS)と正式名称変更

2021年からアカデミッククラウドとしてリプレイスされる。

- オンプレとクラウドの構成は変わらないが、クラウドはAWSへ。
- オンプレミス:Dell PowerEdge R640 ×16ノード
- クラウド:AWS

Amazon EC2 (計 540コア)、Amazon EBS(50TiB/月)、  
Amazon S3(120TiB/月)、Amazon Glacier Deep Archive(200TiB/月)、  
Amazon Aurora(2TiB/月)、Amazon EFS(10TiB/月)

- オンプレミスがスパコンと**10GbE×2で接続**となり、2倍の帯域へ。
- 汎用コンと同様に、仮想サーバホスティングとしてVMの資源提供をしている。

<https://www.iimc.kyoto-u.ac.jp/services/comp/pdf/futankin.pdf>



# VMサーバホスティング

## 各種OS環境を提供

| 仮想化環境      | VMWare                                                                                   |
|------------|------------------------------------------------------------------------------------------|
| OS         | Alma Linux 8<br>RHEL 8<br>CentOS Stream 9<br>Ubuntu 20<br>Windows Server 2019 Datacenter |
| 標準スペック     | CPU:2コア,メモリ:4GB,ディスク:100GB                                                               |
| SSHによるログイン | 可能(Root権限も付与)                                                                            |

|          | 利用負担金                                    |
|----------|------------------------------------------|
| VMサーバ    | 12,000円/年(京大内基本料金)                       |
| システム資源増量 | CPU:2コア単位で3,000円/年<br>メモリ:4GB単位で3,000円/年 |
| ディスク増量   | 100GB単位で6,000円/年                         |

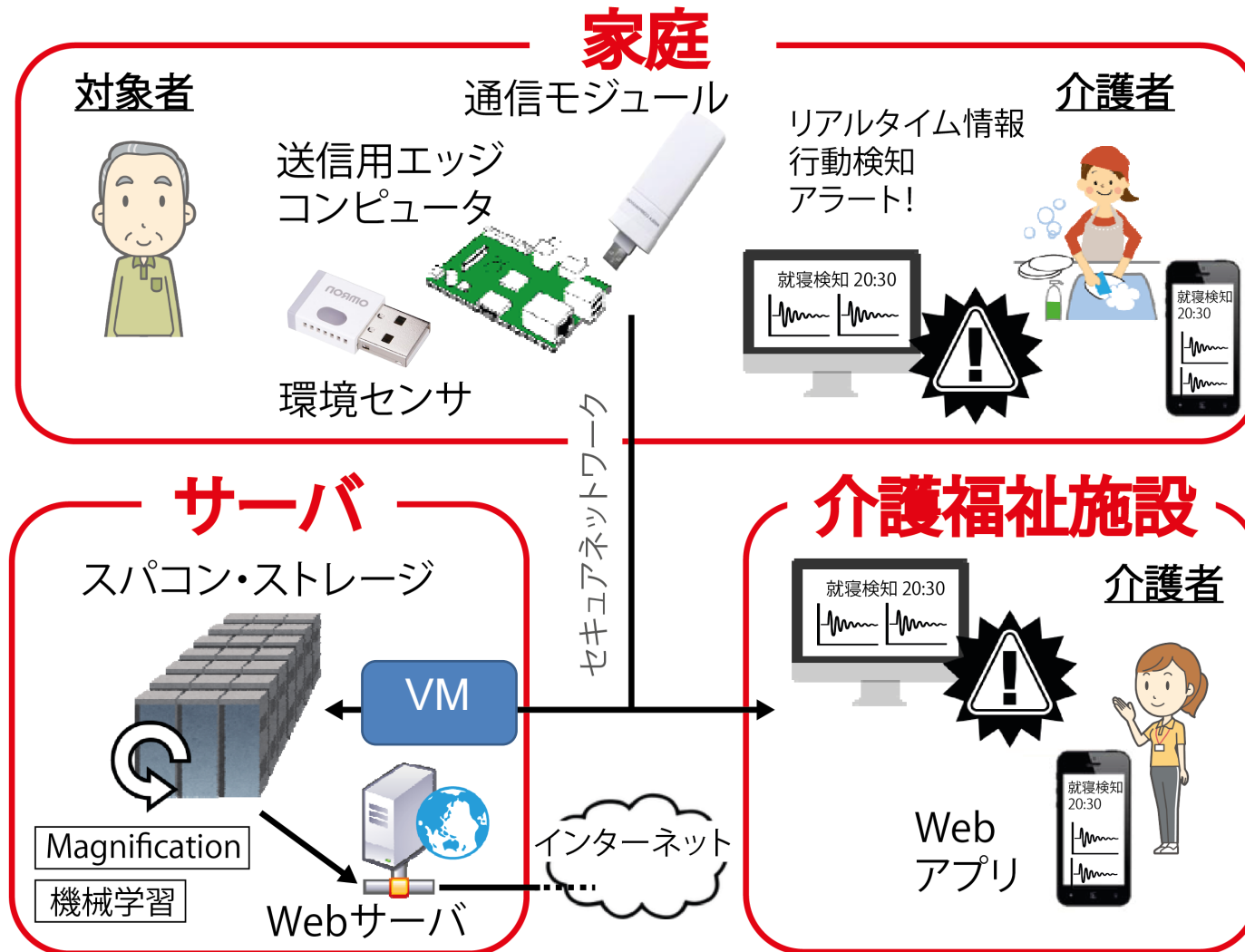
※CPUコア数・メモリ容量・ディスク容量は増量可能。RHELは別途費用がかかる

価格はさくらインターネットのVPSと同程度かちょっと安い位？



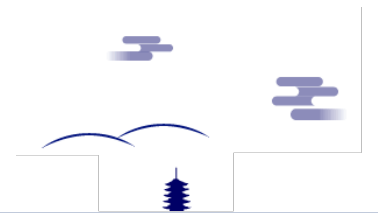
# VMとスパコンを利用した研究例

## IoT機器を使った見守りシステム



数値シミュレーションだけでなく、新しいスパコンの利用方法へのチャレンジ。

センサに繋がったIoT機器から、データがVMからスパコンに送られて、スパコンで解析し、検知、予測などを行うシステム。

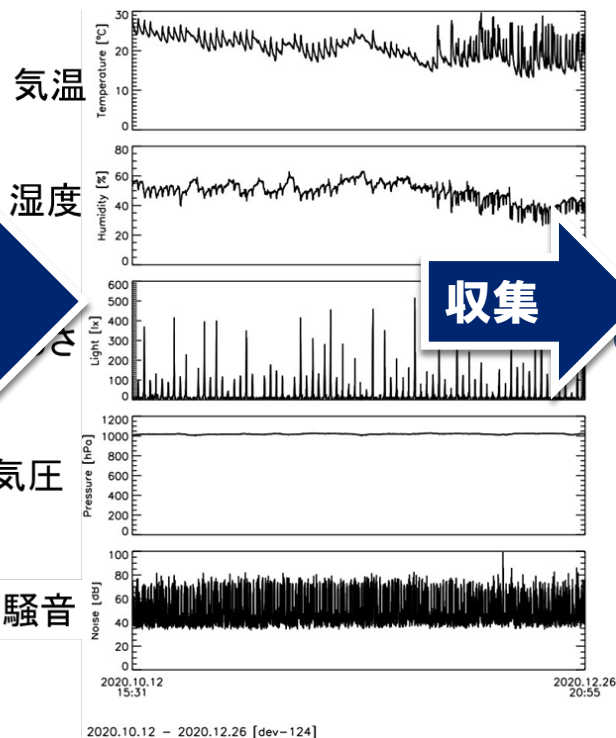


# VMとスパコンを利用した研究例

## IoT機器からVM経由で収集されたデータを使いスパコンで処理



測定



収集

VM

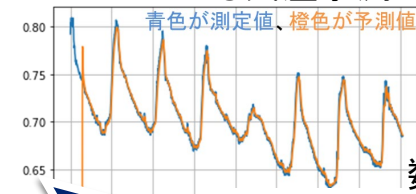


予測

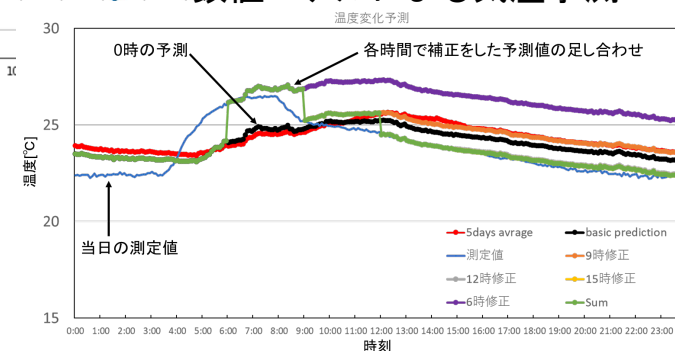
VM

検知

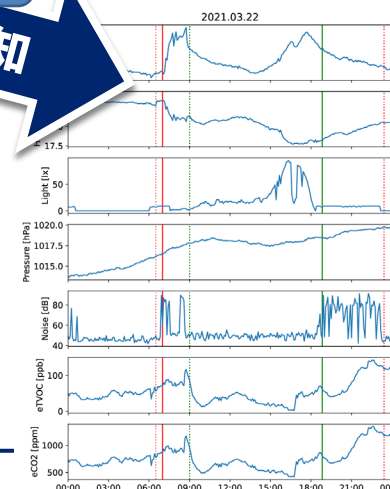
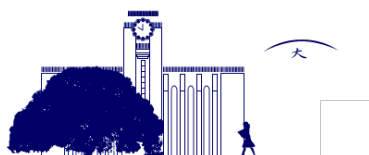
RNNによる気温予測



数値モデルによる気温予測



見守りの負担軽減



起床：○，外出：○，就寝：○

# 次期京大スパコンとクラウド

## 次期京大スパコンとクラウド連携

半導体不足、円安という厳しい状況の中、2022年にスパコンをリプレイス（中）。

仕様書に阪大スパコンのような**クラウドバーステイング**を想定して、「**クラウドサブシステム**」の要求要件を記載し、クラウド連携機能が導入される。

- プライベートネットワークでスパコンシステムのファイルシステムをNFSマウント可能
- クラウド資源がSINETに20Gbps以上の専用線で接続している
- 他にも細かな縛りを仕様に記載

その結果、クラウド部分はOracle Cloud Infrastructureになっている。  
運用をどうするか検討中だが、クラウド利用となるジョブクラスを作る  
可能性が高い？資源量と支払方法も検討中。



# 次期京大スパコン構成

2022/11~(予定?)

**Camphor 3 (System A)** (To be provide in FY2023)

## DELL PowerEdge C6620

Intel Xeon w/ HBM x 2 /node

#nodes = 1,120  
#total cores = 112 cores x 1,120 → 125,440 cores  
Peak performance = 5.19 TFlops x 1,120 → 5.82 PFlops  
Memory capacity = 128 GiB x 1,120 → 140 TiB  
Memory bandwidth = 3.2 TB/sec x 1,120 → 3.6 PB/sec

## Storage

### DDN EXAScaler (ES400NVX2)

HDD capacity = 40 PB  
HDD Bandwidth = 280 GB/sec  
SSD capacity = 4 PB (To be provide in FY2023)  
SSD Bandwidth = 768 GB/sec

高速通信網 InfiniBand HDR/NDR

クラウド

**Laurel 3 (System B)**

## DELL PowerEdge C6620

Intel Xeon x 2 /node

#nodes = 370  
#total cores = 112 cores x 370 → 41,440 cores  
Peak performance = 5.91 TFlops x 370 → 2.19 PFlops  
Memory capacity = 512 GiB x 370 → 185 TiB  
Memory bandwidth = 614 GB/sec x 370 → 227 TB/sec

**Gardenia (System G)**

## DELL PowerEdge XE8545

AMD EPYC Milan 32cores 2.6GHz x 2 /node

#nodes = 16  
#total cores = 64 cores x 16 → 1,024 cores  
Peak performance = 2.66 TFlops x 16 → 42.6 TFlops  
Memory capacity = 512 GB x 16 → 8.2 TB  
Memory bandwidth = 409 GB/sec x 16 → 6.5 TB/sec

**Cinnamon 3 (System C)**

## DELL PowerEdge C6620

Intel Xeon x 2 /node

#nodes = 16  
#total cores = 112 cores x 16 → 1,792 cores  
Peak performance = 5.91 TFlops x 16 → 94.6 TFlops  
Memory capacity = 2 TiB x 16 → 32 TiB  
Memory bandwidth = 563 GB/sec x 16 → 9 TB/sec

## Accelerator

NVIDIA A100 80GB SXM x 4 /node

#total GPUs = 4 GPU x 16 → 64 GPU  
Peak performance = 282TFlops x 64  
(FP16) → 18.0 PFlops  
Memory capacity = 80 GiB x 64 → 5.1 TiB  
Memory bandwidth = 2.04 GB/sec x 64 → 130 TB/sec



# 次期京大スパコン

## これまでの構成を踏襲+拡張

### システムA

- メニーコア、自作コードを持った研究者がプロダクトランに使うイメージ。

### システムB

- Xeon+DDRの汎用計算機、ISVアプリやシングルノードユーザ向け。

### システムC

- 大容量メモリ、共有メモリアプリ向け。

### システムG

- GPUマシン、単ノード利用を想定（大規模並列GPUは他サイトで）。

### クラウドシステム

- オンプレリソース不足、オンプレとは違う環境を使いたい人向け。



# 現状のスパコンとクラウドの連携での課題

## 扱うデータサイズに関する課題が大きい気がする

スパコンを使うようなユーザだと、ローカルからクラウドへ（その逆も）のデータの移動負荷が高い（データが大きいため）

- リソースが無い時に使ってもらはずが、データ転送待ちで結局待ちが発生。  
数日待つようなリソース枯渇時は問題無いと思う。

データが大きいので通信にチャージされるとコストがきつい

- 少データやクラウドで閉じる計算には問題無い。

クラウド上にデータを保存する際のRead/Writeコストがあるとつらい

- 速度次第で色々コストは変わるが、高い。

スパコンとクラウドシステム連携より、スパコンをmoreクラウド化が近道？



# まとめ

## 京大スパコンはもうすぐリプレイス

- ✓ スパコンと汎用コンピュータ・アカデミッククラウドと連携を行っており、VMとスパコンがある程度物理的に近い環境で運用されている。
- ✓ 次期スパコンでは、クラウドバーステイングを導入。
- ✓ 運用方法は未決定でこれから議論と、運用しながら色々と変更していく予定。
- ✓ 次期スパコン自体は、これまでの京大スパコンの構成を引き継いでいる。
- ✓ 東大のストレージ別調達やシステム毎の時期をずらした調達など、今後とも色々と検討して、利用者にとって良いシステムを目指す。

