HPC Workload Characterization Using Feature Selection and Clustering

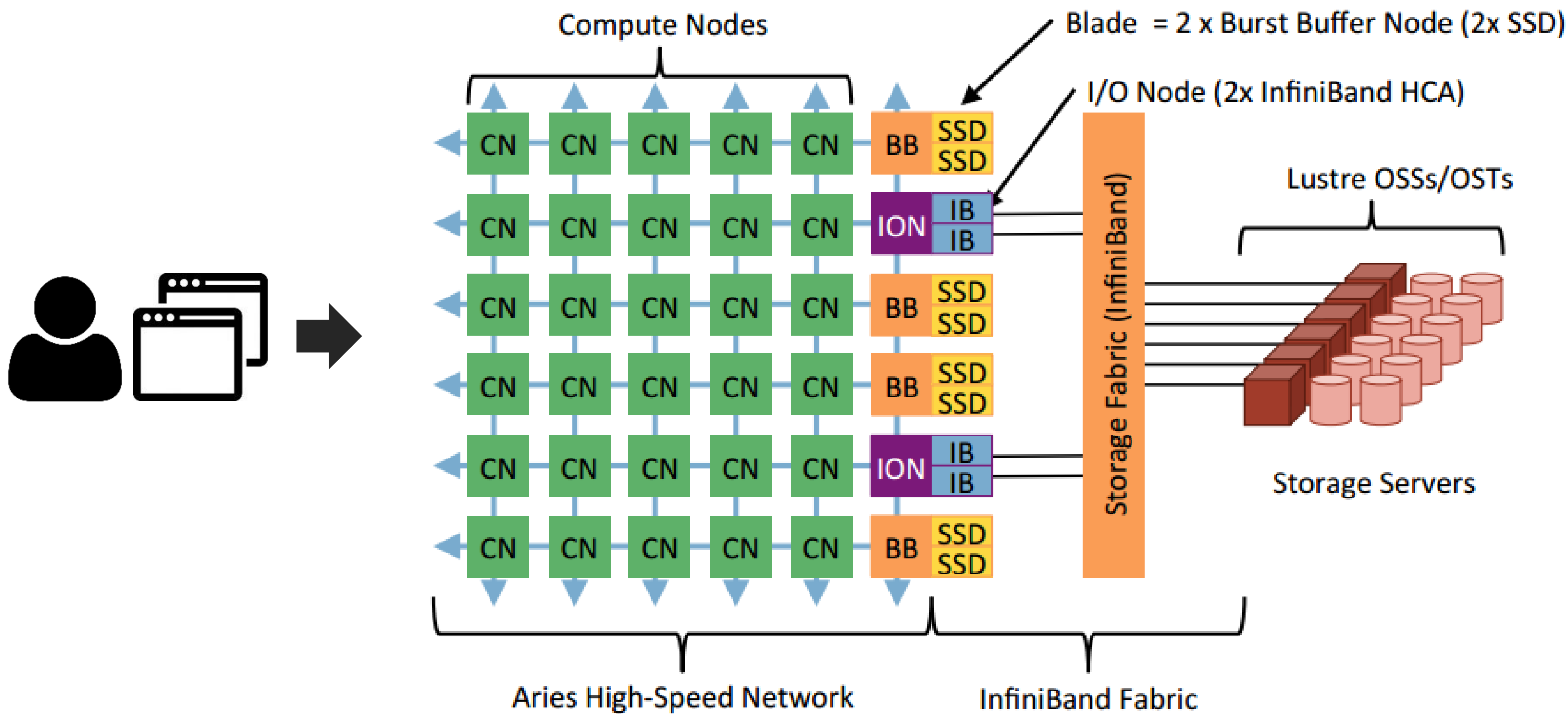
Jiwoo Bang, Chungyong Kim, Kesheng Wu, Alexander Sim, Suren Byna, Hyeonsang Eom

Distributed Computing Systems Laboratory
Department of Computer Science and Engineering
Seoul National University, Korea

Table of Contents

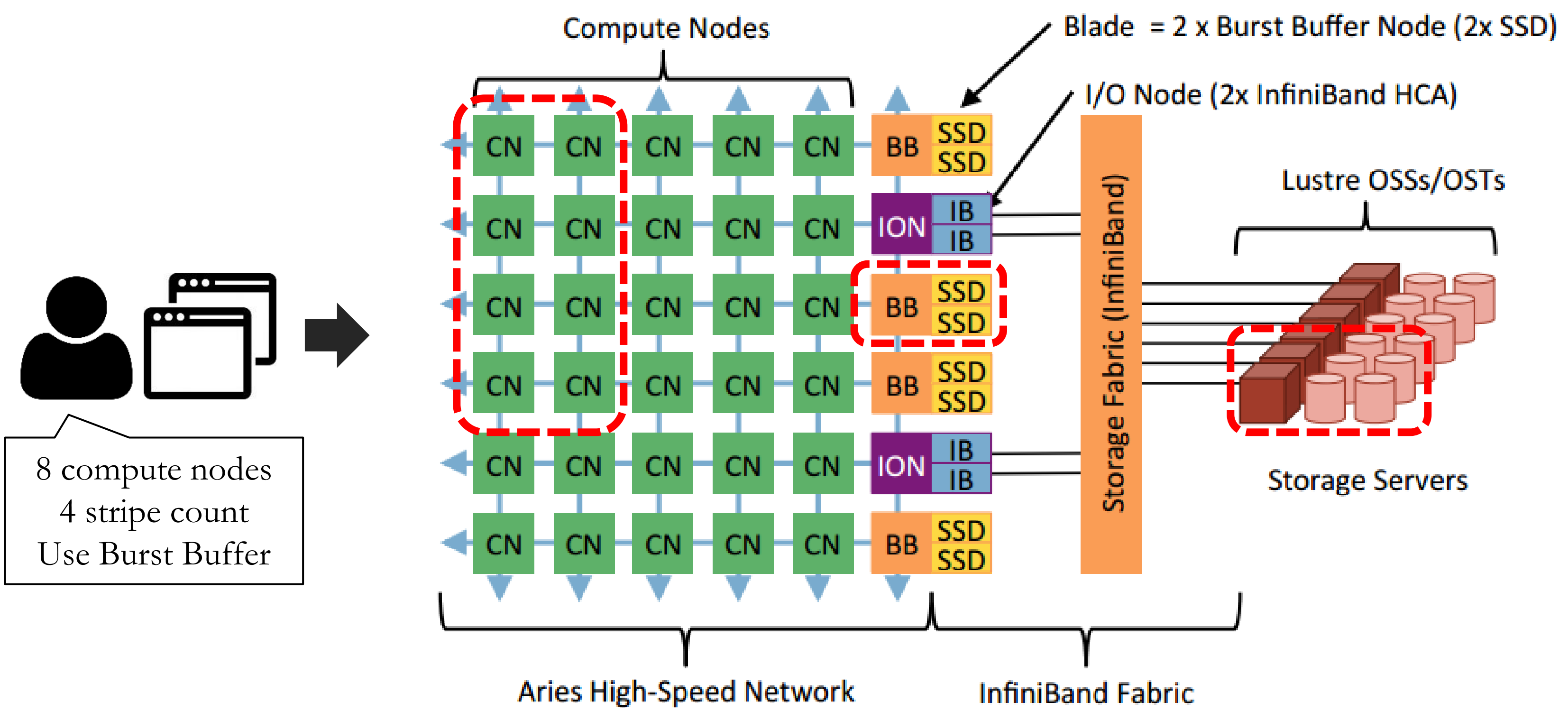
High Performance Computing (HPC) system

- **Applications running on HPC system demand for efficient storage management and high performance computation**



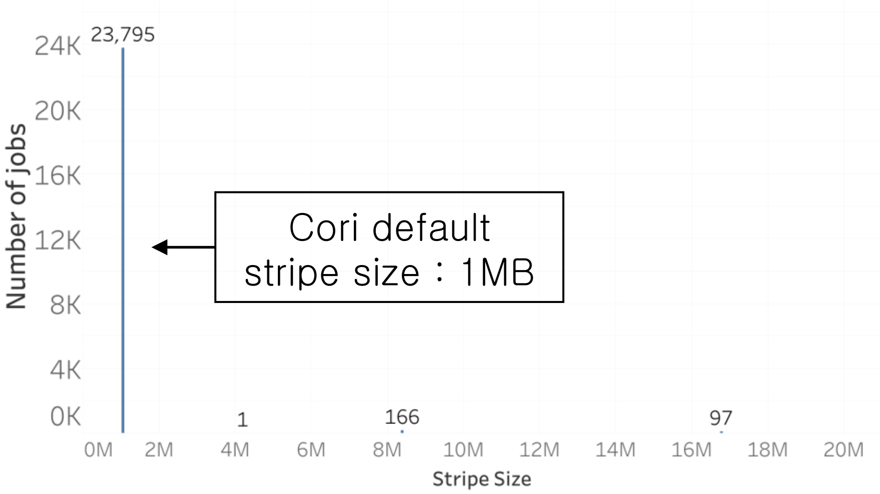
High Performance Computing (HPC) system

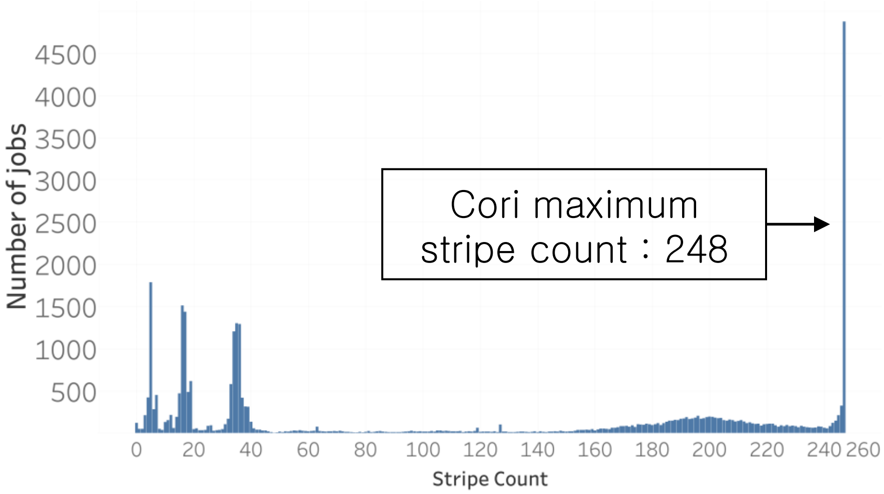
- Applications running on HPC system demand for efficient storage management and high performance computation
- Tunable parameters are provided for higher performance
  - Number of compute nodes, Stripe count, Stripe size, ..



Drawbacks on deploying HPC environment

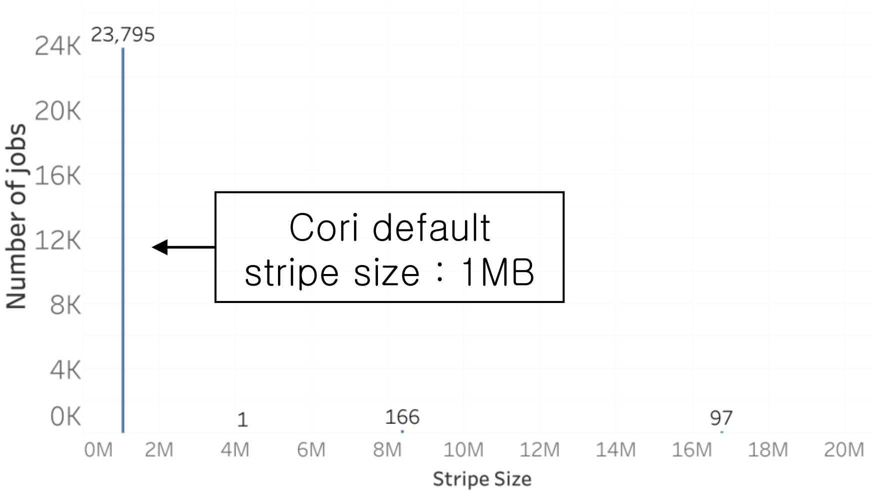
- **Users are not familiar with using tunable parameters**
  - They use default configurations the system provides or maximum available resource

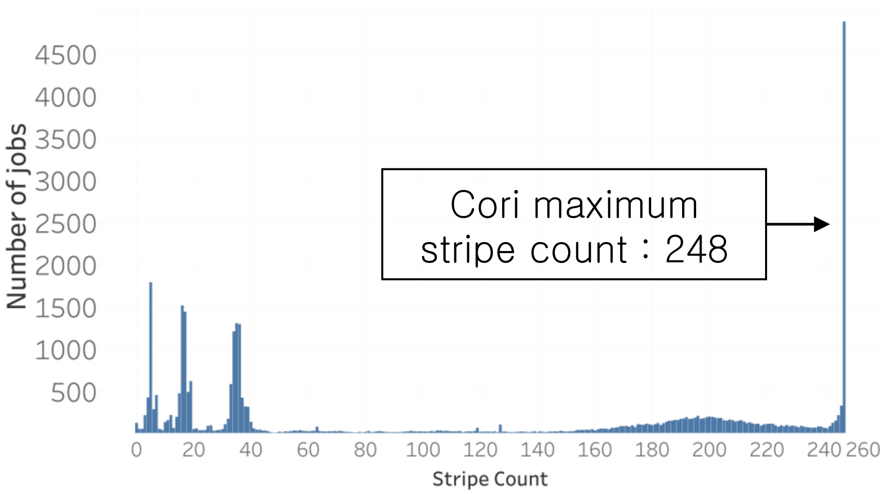




Drawbacks on deploying HPC environment

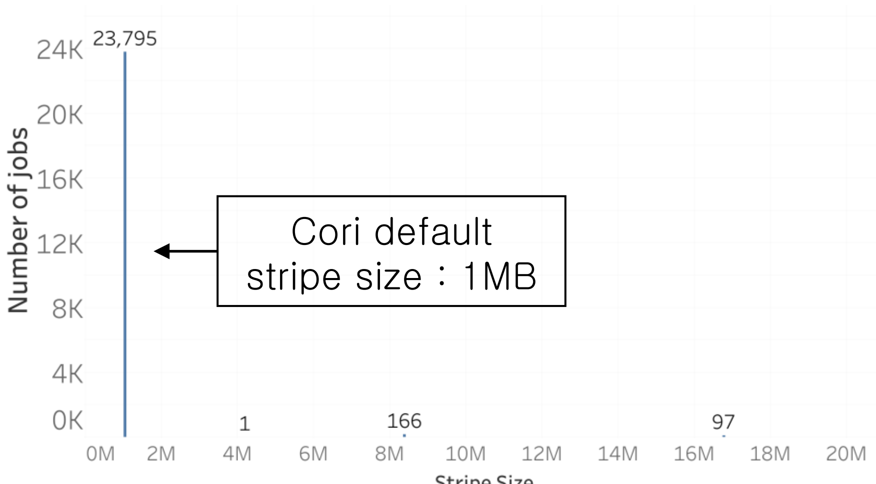
- **Users are not familiar with using tunable parameters**
  - They use default configurations the system provides or maximum available resource
- **Some of the HPC applications do not meet I/O demands**
  - I/O characteristics for each applications are different
  - I/O performance differs depending on the HPC system

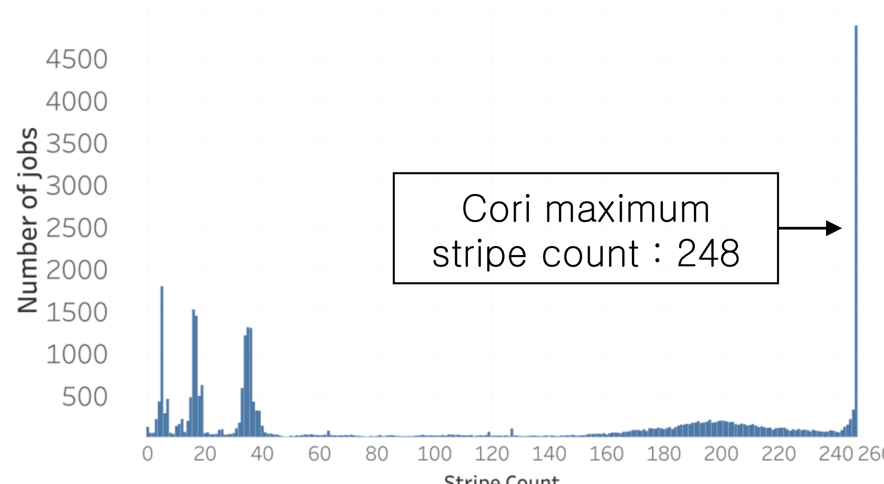




Drawbacks on deploying HPC environment

- **Users are not familiar with using tunable parameters**
  - They use default configurations the system provides or maximum available resource
- **Some of the HPC applications do not meet I/O demands**
  - I/O characteristics for each applications are different
  - I/O performance differs depending on the HPC system





Understanding the different I/O demands of HPC applications is important

Used Dataset

- **Real-world user log data from Oct. 2017 to Jan. 2018**
  - Total 4-month Darshan log data is used
    - Darshan I/O profiling tool captures I/O behaviors of applications run on Cori system
    - Darshan interacts with Slurm workload manager
  - Parser is used to extract meaningful information from Darshan log and Lustre monitoring tool
    - Total 78 features are obtained from the parser

Used Dataset

- **Real-world user log data from Oct. 2017 to Jan. 2018**
  - Total 4-month Darshan log data is used
    - Darshan I/O profiling tool captures I/O behaviors of applications run on Cori system
    - Darshan interacts with Slurm workload manager and Lustre monitoring tool
  - Parser is used to extract meaningful information from Darshan log
    - Total 78 features are obtained from the parser
- **I/O throughput (*writeRateTotal*) is the target variable**
  - HPC applications are categorized based on their I/O behaviors

Data Preprocessing

- **User logs with less than 1GB I/O are dropped**
  - They cannot capture the relationship between features and the target variable

Data Preprocessing

- **User logs with less than 1GB I/O are dropped**
  - They cannot capture the relationship between features and the target variable
- **Data having negative values are all set to zero**

Data Preprocessing

- **User logs with less than 1GB I/O are dropped**
  - They cannot capture the relationship between features and the target variable
- **Data having negative values are all set to zero**
- **The features with zero variance are eliminated**
  - The features with the constant value are not meaningful at all

DCSLAB

Data Preprocessing

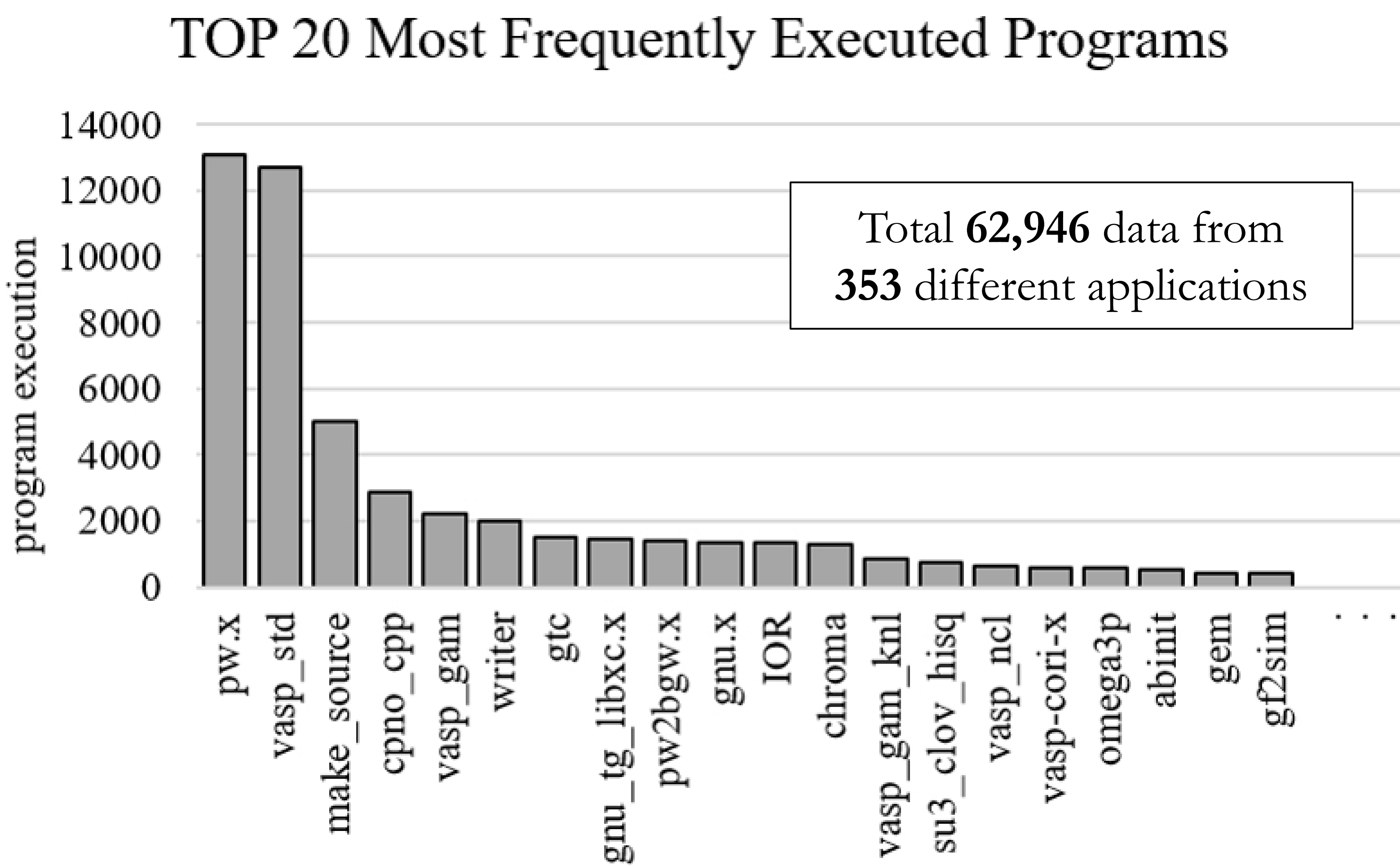
- **User logs with less than 1GB I/O are dropped**
  - They cannot capture the relationship between features and the target variable
- **Data having negative values are all set to zero**
- **The features with zero variance are eliminated**
  - The features with the constant value are not meaningful at all
- **The features having highly correlated value with other features are eliminated**
  - The correlation value threshold is set to 0.8
  - It is to reduce redundancy among the feature selection results

Data Preprocessing

- **User logs with less than 1GB I/O are dropped**
  - They cannot capture the relationship between features and the target variable
- **Data having negative values are all set to zero**
- **The features with zero variance are eliminated**
  - The features with the constant value are not meaningful at all
- **The features having highly correlated value with other features are eliminated**
  - The correlation value threshold is set to 0.8
  - It is to reduce redundancy among the feature selection results
- **The feature data is normalized to range from 0 to 1**
  - The features can have same scale and weight when calculated by feature selection methods

Data Preprocessing

- **Top 20 mostly executed programs after preprocessing step**



DCSLAB

Feature Selection for Dimension Reduction

- **Feature selection methods**

Feature Selection for Dimension Reduction

- **Feature selection methods**
  - Mutual Information regression
  - F Regression
  - Decision Tree
  - Extra Tree

Feature Selection for Dimension Reduction

- **Feature selection methods**
  - Mutual Information regression
  - F Regression
  - Decision Tree
  - Extra Tree
  - Min-max Mutual Information (the new feature selection method)

Feature Selection for Dimension Reduction

- **Feature selection methods**
  - Mutual Information regression
  - F Regression
  - Decision Tree
  - Extra Tree
  - Min-max Mutual Information (the new feature selection method)
    - The data preprocessing step of removing features that have highly correlated value with other features is not applied
    - Min-max mutual information selects features that are less correlated to each other
    - The first feature that has highest correlation value with *wrtieRateTotal* is selected, and then this process is repeated

DCSLAB

Feature Selection for Dimension Reduction

- Analysis of Feature Selection results

Feature	Score
seqWritePct	1.233037
totalFile	1.138258
totalOpenReq	1.096744
totalIOReq	1.082124
numProc	1.005036
runTime	0.973189
totalReadReq	0.937343

(a) Mutual Information

Feature	Score
totalFileSTDIO	25849.56
totalFile	3074.00
numProc	1281.61
numOST	957.34
readLess1m	464.63
ossWriteMean	411.40
ossWriteHigher1g	294.13

(b) F Regression

Feature	Score
totalFileSTDIO	0.353953
runTime	0.079576
runProc	0.075793
totalFile	0.054656
totalReadReq	0.053308
seqWritePct	0.050254
writeTimePOSIXonly	0.049154

(c) Decision Tree

Feature	Score
totalFileSTDIO	0.357795
runProc	0.073166
runTime	0.069633
totalFile	0.054624
totalReadReq	0.051944
seqWritePct	0.049576
writeTimePOSIXonly	0.046418

(d) Extra Tree

Feature	Score
readMore1m	-
metaTimePOSIXonly	-
readMore1k	-
ossWriteHigher4g	-
writeLess1k	-
stripeSize	-
totalReadReq	-

(e) Min-max Mutual Information

Application of Clustering Model

- **Clustering models**
  - KMeans Clustering
  - Gaussian Mixture Model
  - Ward Linkage Clustering

Application of Clustering Model

- **Clustering models**
  - KMeans Clustering
  - Gaussian Mixture Model
  - Ward Linkage Clustering

- **Cluster Validity Metrics**
  - Davies-Bouldin index (DBI) metric
  - Silhouette score metric
  - Combined score metric
    - $$CombinedScore(x) = \begin{cases} \frac{Silhouette(x)}{DBI(x)}, & \text{if } DBI(x) \neq 0 \\ \text{undefined}, & \text{otherwise} \end{cases}$$

Application of Clustering Model

- **Clustering models**
  - KMeans Clustering
  - Gaussian Mixture Model
  - Ward Linkage Clustering

- **Cluster Validity Metrics**
  - Davies-Bouldin index (DBI) metric
  - Silhouette score metric
  - Combined score metric
    - $$CombinedScore(x) = \begin{cases} \frac{Silhouette(x)}{DBI(x)}, & \text{if } DBI(x) \neq 0 \\ \text{undefined}, & \text{otherwise} \end{cases}$$

For **DBI**, the lower the better the cluster quality
For **Silhouette** and **Combined score**, the higher the better the cluster quality

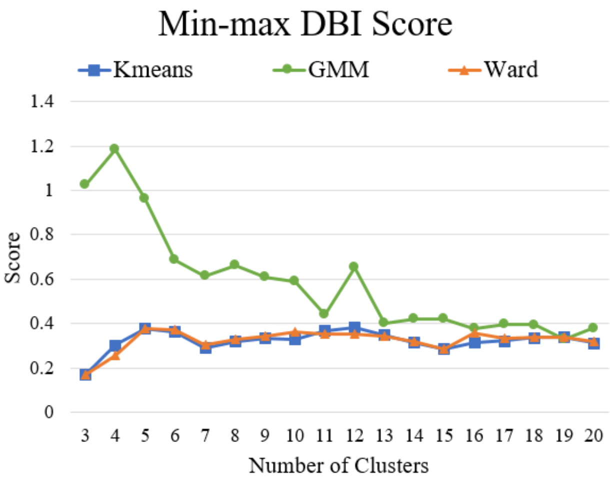
Performance Evaluation

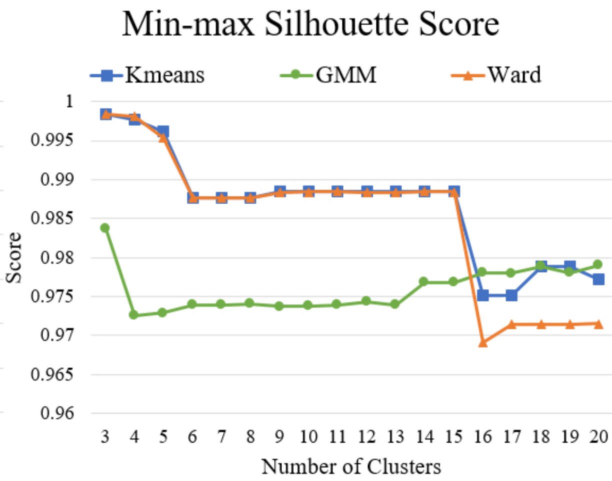
- **Selecting Best Clustering method**
  - The features selected from Min-max mutual information are used
    - The most suitable feature selection method for our dataset's characteristic: every feature is considerably correlated to each other
  - The number of clusters varies from 3 to 20

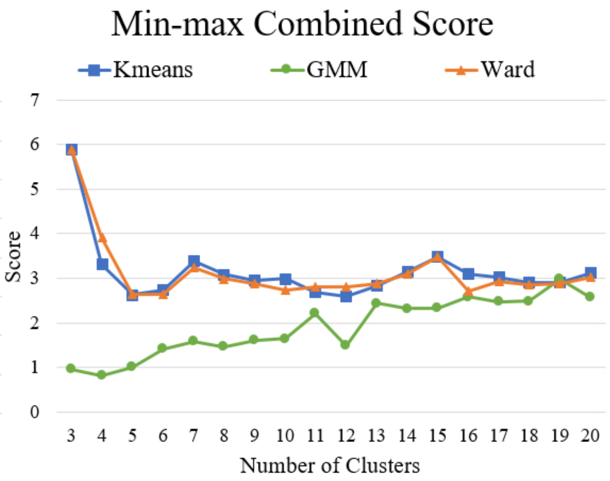
DCSLAB

Performance Evaluation

- **Selecting Best Clustering method**
  - The features selected from Min-max mutual information are used
    - The most suitable feature selection method for our dataset's characteristic: every feature is considerably correlated to each other
  - The number of clusters varies from 3 to 20

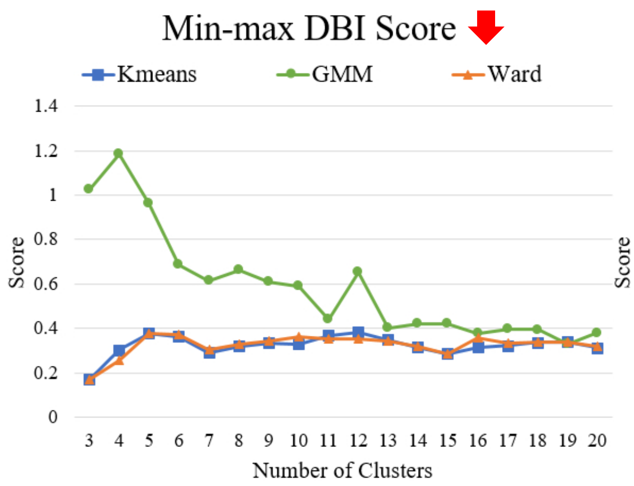


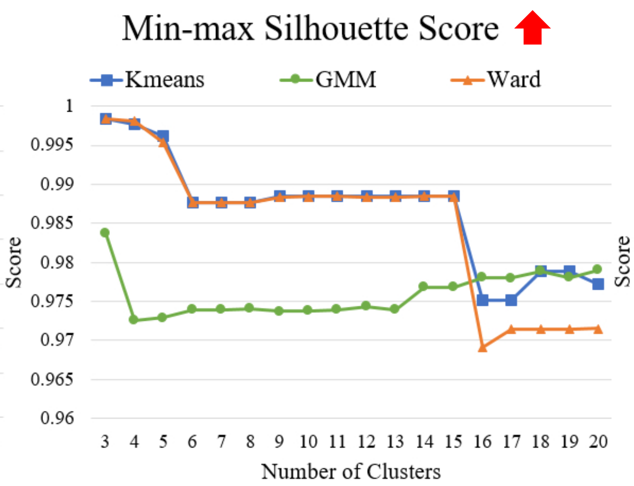


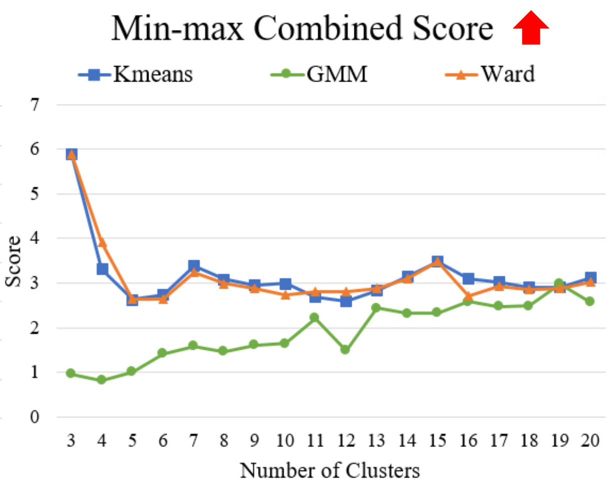


Performance Evaluation

- **Selecting Best Clustering method**
  - The features selected from Min-max mutual information are used
    - The most suitable feature selection method for our dataset's characteristic: every feature is considerably correlated to each other
  - The number of clusters varies from 3 to 20

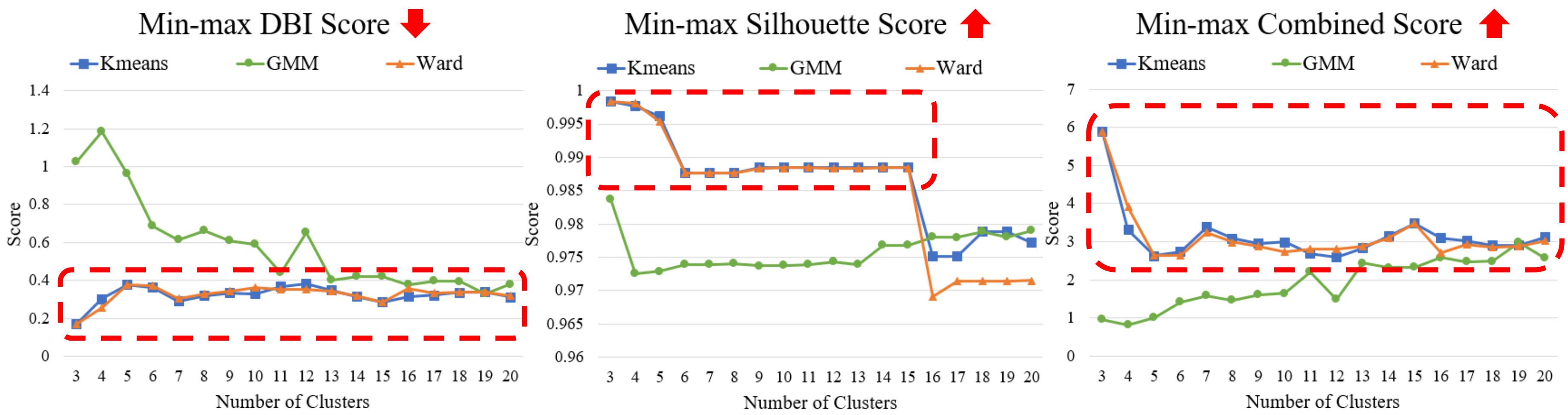






Performance Evaluation

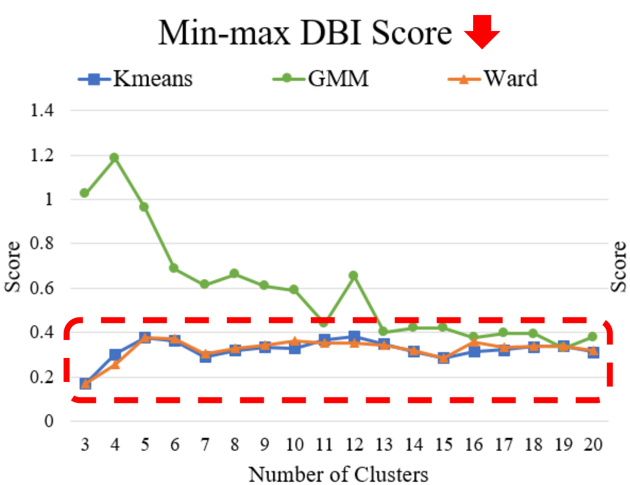
- **Selecting Best Clustering method**
  - The features selected from Min-max mutual information are used
    - The most suitable feature selection method for our dataset's characteristic: every feature is considerably correlated to each other
  - The number of clusters varies from 3 to 20

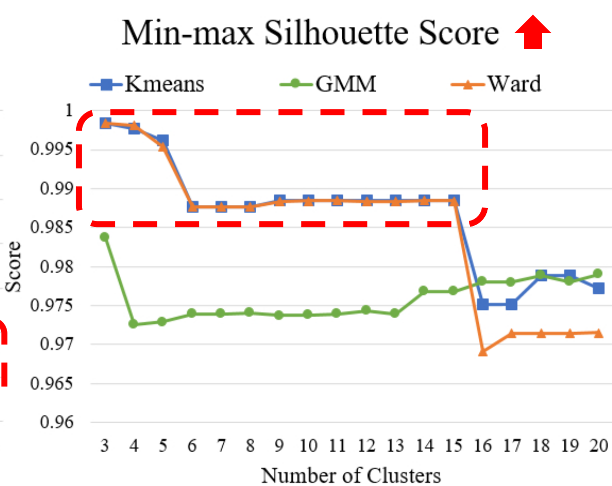


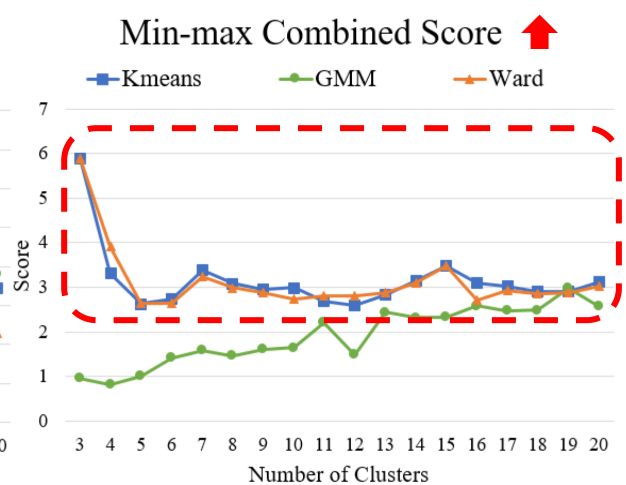
DCSLAB

Performance Evaluation

- **Selecting Best Clustering method**
  - The features selected from Min-max mutual information are used
    - The most suitable feature selection method for our dataset's characteristic: every feature is considerably correlated to each other
  - The number of clusters varies from 3 to 20







Kmeans and **Ward linkage** show high cluster performance
The performance is highest when the number of clusters is **3**

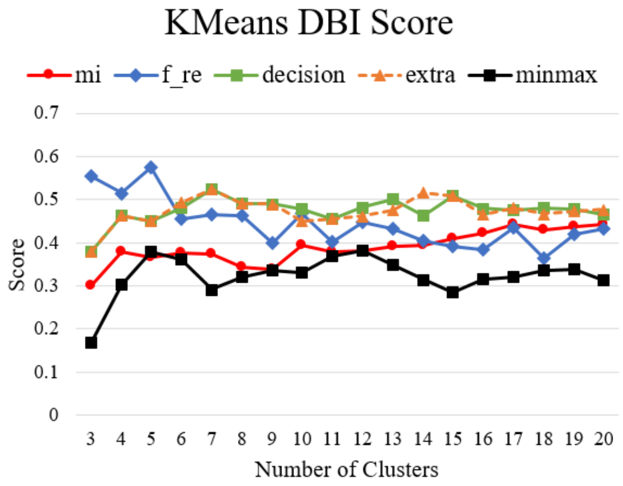
DCSLAB

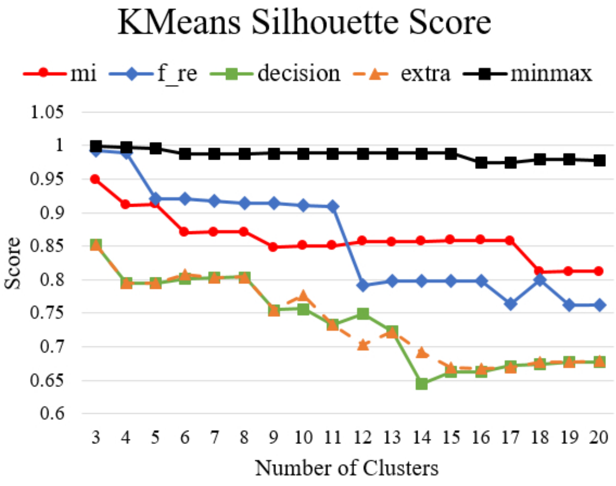
Performance Evaluation

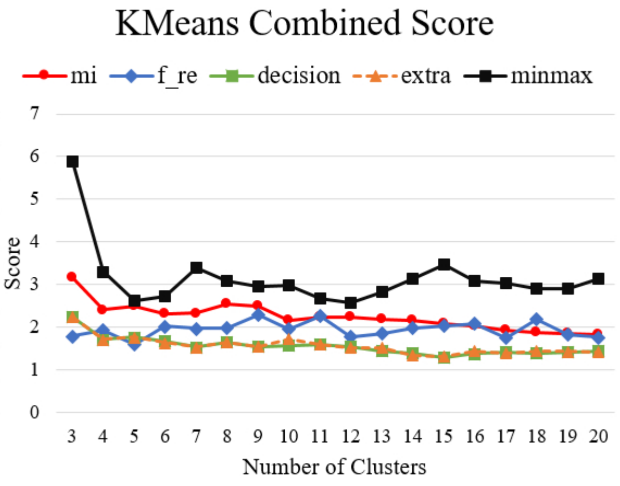
- **Feature Selection methods comparison**
  - The impact the five feature selection methods have on Kmeans clustering method is evaluated
    - Mutual information, F-regression, Decision tree, Extra tree, and Min-max mutual information

Performance Evaluation

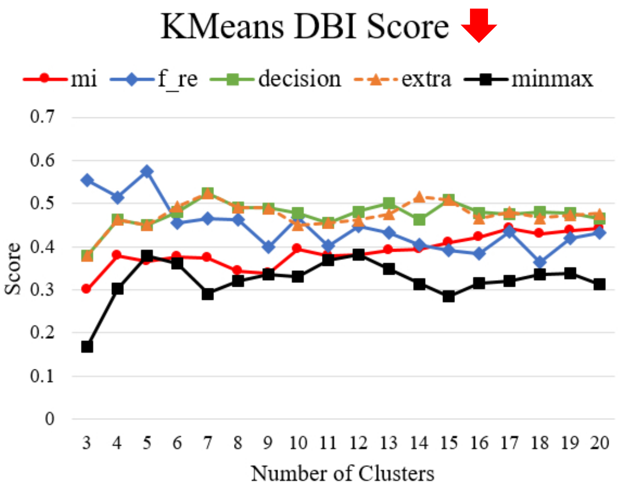
- **Feature Selection methods comparison**
  - The impact the five feature selection methods have on Kmeans clustering method is evaluated
    - Mutual information, F-regression, Decision tree, Extra tree, and Min-max mutual information

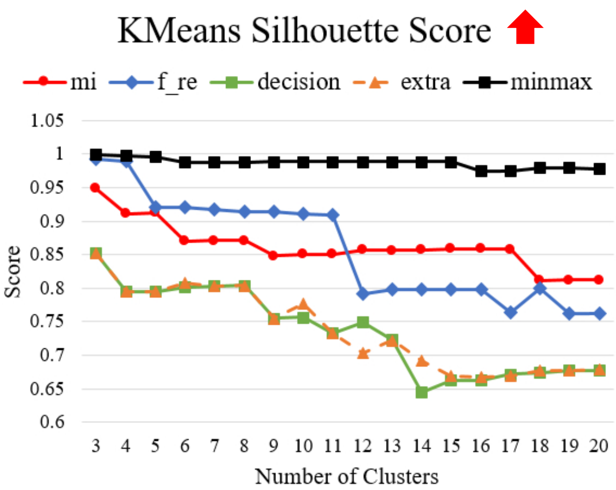


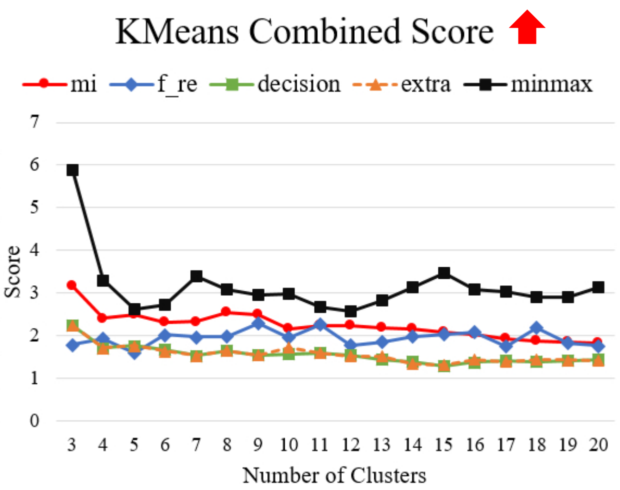




- **Feature Selection methods comparison**
  - The impact the five feature selection methods have on Kmeans clustering method is evaluated
    - Mutual information, F-regression, Decision tree, Extra tree, and Min-max mutual information



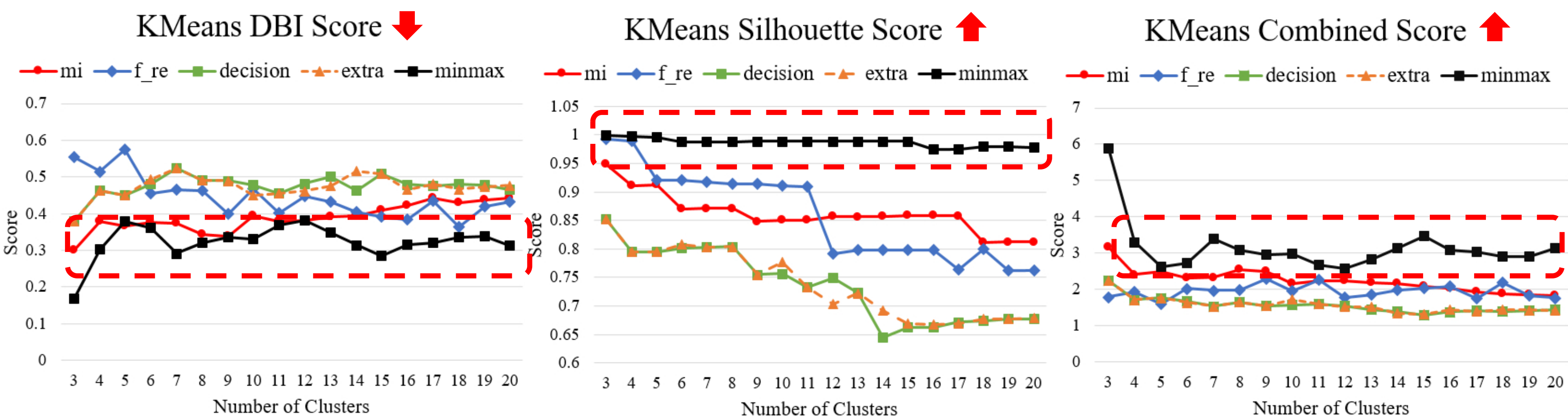




DCSLAB

Performance Evaluation

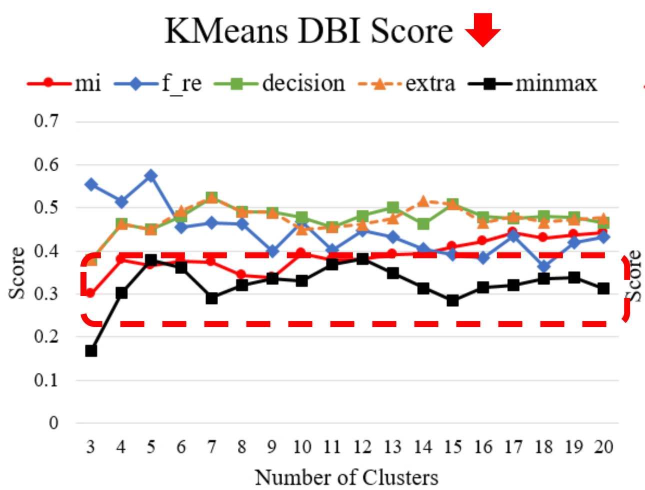
- **Feature Selection methods comparison**
  - The impact the five feature selection methods have on Kmeans clustering method is evaluated
    - Mutual information, F-regression, Decision tree, Extra tree, and Min-max mutual information

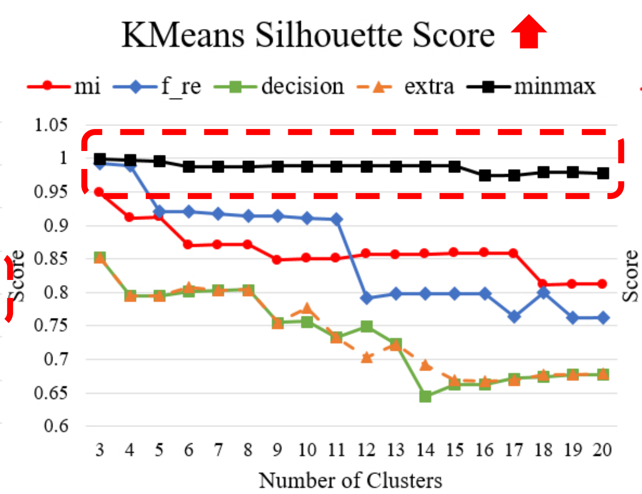


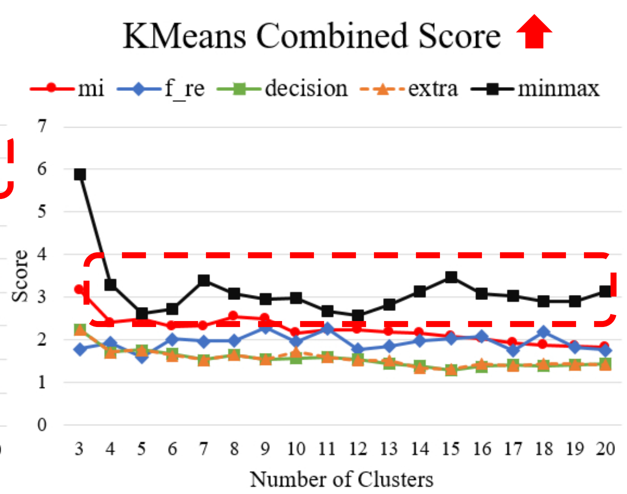
DCSLAB

Performance Evaluation

- **Feature Selection methods comparison**
  - The impact the five feature selection methods have on Kmeans clustering method is evaluated
    - Mutual information, F-regression, Decision tree, Extra tree, and Min-max mutual information







Clustering result using features selected from
Min-max mutual information shows highest cluster performance

Cluster Characterization

- **Cluster Characterization**
  - Min-max mutual information feature selection
  - KMeans (or Ward linkage) clustering algorithm
  - Clustering with 3 clusters scores highest

Cluster Characterization

- **Cluster Characterization**
  - Min-max mutual information feature selection
  - KMeans (or Ward linkage) clustering algorithm
  - Clustering with 3 clusters scores highest

Features \ Cluster index	1	2	3
numProc	412	1140	71693
totalFileMPIIO	6.69	1.31	14.33
totalSeekReq	8037382	34852	103872479
totalStatReq	17116	1804	34260138
totalOpenReq	12778	6183	34558780
readLess1m	67.20	0	33.33
writeLess1m	49.13	0	33.33
readMore1m	0	100	0
writeMore1m	2.20	20.75	0
writeRateTotal	7844	47794	65226

DCSLAB

Cluster Characterization

- **Cluster Characterization**
  - Min-max mutual information feature selection
  - KMeans (or Ward linkage) clustering algorithm
  - Clustering with 3 clusters scores highest

Features \ Cluster index	1	2	3
numProc	412	1140	71693
totalFileMPIIO	6.69	1.31	14.33
totalSeekReq	8037382	34852	103872479
totalStatReq	17116	1804	34260138
totalOpenReq	12778	6183	34558780
readLess1m	67.20	0	33.33
writeLess1m	49.13	0	33.33
readMore1m	0	100	0
writeMore1m	2.20	20.75	0
writeRateTotal	7844	47794	65226

DCSLAB

Cluster Characterization

- ## Cluster Characterization

Features \ Cluster index	0	1	2
consecWritePct	■		
ossWriteHigher4g	■		
readLess1k	■		
readLess1m	■		
totalFileSTDIO	■		
mdsOPSMin		■	
ossReadHigher4g		■	
ossWriteHigher1g		■	
readMore1k		■	
readMore1m		■	
stripeSize		■	
totalFile		■	
totalFilePOSIX		■	
writeMore1k		■	
writeMore1m		■	
metaTimePOSIXonly			■
numProc			■
readTimePOSIXonly			■
totalFileMPIIO			■
totalIOReq			■
totalMetaReq			■
totalOpenReq			■
totalReadReq			■

Features \ Cluster index	0	1	2
totalSeekReq			■
totalStatReq			■
totalWriteReq			■
writeTimePOSIXonly			■
consecReadPct			
mdsCPU95			
mdsCPUMean			
mdsOPS95			
mdsOPSMean			
numOST			
ossRead95			
ossReadHigher1g			
ossReadLargest			
ossReadMean			
ossWrite95			
ossWriteLargest			
ossWriteMean			
runTime			
seqReadPct			
seqWritePct			
writeLess1k			
writeLess1m			
writeRateTotal			

Cluster Characterization

- Cluster Characterization

Features \ Cluster index	0	1	2
consecWritePct			
ossWriteHigher4g			
readLess1k			
readLess1m			
totalFileSTDIO			
mdsOPSMin			
ossReadHigher4g			
ossWriteHigher1g			
readMore1k			
readMore1m			
stripeSize			
totalFile			
totalFilePOSIX			
writeMore1k			
writeMore1m			
metaTimePOSIXonly			
numProc			
readTimePOSIXonly			
totalReadReq			

Features \ Cluster index	0	1	2
totalSeekReq			
totalStatReq			
totalWriteReq			
writeTimePOSIXonly			
consecReadPct			
mdsCPU95			
mdsCPUMean			
mdsOPS95			
mdsOPSMean			
numOST			
ossRead95			
ossReadHigher1g			
ossReadLargest			
ossReadMean			
ossWrite95			
ossWriteLargest			
ossWriteMean			
runTime			
writeRateTotal			

Cluster 1

- workloads with less than 1MB size read/write operations, mostly on stdio units

- Average I/O throughput is only a few MB/s

Cluster Characterization

- ## Cluster Characterization

Features \ Cluster index	0	1	2
consecWritePct	■		
ossWriteHigher4g	■		
readLess1k	■		
readLess1m	■		
totalFileSTDIO	■		
mdsOPSMin		■	
ossReadHigher4g		■	
ossWriteHigher1g		■	
readMore1k		■	
readMore1m		■	
stripeSize		■	
totalFile		■	
totalFilePOSIX		■	
writeMore1k		■	
writeMore1m		■	
metaTimePOSIXonly			■

Features \ Cluster index	0	1	2
totalSeekReq			■
totalStatReq			■
totalWriteReq			■
writeTimePOSIXonly			■
consecReadPct			
mdsCPU95			
mdsCPUMean			
mdsOPS95			
mdsOPSMean			
numOST			
ossRead95			
ossReadHigher1g			
ossReadLargest			
ossReadMean			
ossWrite95			
ossWriteLargest			

Cluster 2

- workloads with more than 1MB size read/write operations
- lots of I/O operations during the processing time
- likely to use 8MB stripe size in average, which is 8 times the default size
- use the relatively small number of cores

Cluster Characterization

- Cluster Characterization

Features \ Cluster index	0	1	2
consecWritePct	■		
ossWriteHigher4g	■		
readLess1k	■		
readLess1m	■		
totalFileSTDIO	■		
mdsOPSMin		■	
ossReadHigher4g		■	
writeMore1k		■	
writeMore1m		■	
metaTimePOSIXonly			■
numProc			■
readTimePOSIXonly			■
totalFileMPIIO			■
totalIOReq			■
totalMetaReq			■
totalOpenReq			■
totalReadReq			■

Features \ Cluster index	0	1	2
totalSeekReq			■
totalStatReq			■
totalWriteReq			■
writeTimePOSIXonly			■
consecReadPct			
mdsCPU95			
mdsCPUMean			
ossReadMean			
ossWrite95			
ossWriteLargest			
ossWriteMean			
runTime			
seqReadPct			
seqWritePct			
writeLess1k			
writeLess1m			
writeRateTotal			

Cluster 3

- workloads use more than 70,000 MPI ranks on average
- Use 62 times more processors on average
- Issue a large number of I/O requests

Conclusion

- **Summary**
  - We extracted the features highly related to the I/O performance
    - We implemented new feature selection method, Min-max mutual information, in order to get meaningful information from real HPC workload data
  - We clustered the HPC applications and evaluated with cluster quality score
  - We could identify meaningful clusters from the large set of application logs
- **Future work**
  - We aim to give applications in each cluster detailed guidance to improve their performance