Effectively Task Containerization in High Performance Clustering using Artificial Intelligence

आर्टिफिशियल इंटेलिजेंस का उपयोग करके

उन्नत सक्षम क्लस्टरिंग में प्रभावी ढंग से कार्य कंटेनरकरण

Thesis submitted in partial fulfillment
for the Award of Degree

Doctor of Philosophy

by

Chitranjan Singh

चितरंजन सिंह

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY
(BANARAS HINDU UNIVERSITY)
VARANASI - 221005

Roll No. 17071512

Year 2023

CERTIFICATE

It is certified that the work contained in the thesis titled ***"Effectively Task Containerization in High Performance Clustering using Artificial Intelligence"*** by ***Chitranjan Singh*** has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all requirements of Comprehensive Examination, Candidacy, and SOTA for the award of Ph.D. Degree.

Supervisor

06/11/23

Dr. Hari Prabhat Gupta
Associate Professor,
Department of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi,
Uttar Pradesh, INDIA 221005.

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

DECLARATION BY THE CANDIDATE

I, ***Chitranjan Singh***, certify that the work embodied in this Ph.D. thesis is my own bonafide work carried out by me under the supervision of ***Dr. Hari Prabhat Gupta*** from ***December 2017*** to ***October 2023*** at ***Department of Computer Science and Engineering***, Indian Institute of Technology (BHU) Varanasi. The matter embodied in this thesis has not been submitted for the award of any other degree/diploma. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not willfully copied any other's work, paragraphs, text, data, results, *etc.* reported in journals, books, magazines, reports, dissertations, theses, *etc.*, or available at websites and have not included them in this thesis and have not cited as my own work.

Date: 20/10/2023

Place: Varanasi

Chitranjan S.

(Chitranjan Singh)

CERTIFICATE BY THE SUPERVISOR

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

06/11/2023

(Dr. Hari Prabhat Gupta)
Associate Professor,
Dept. of Computer Science and Engineering,
Indian Institute of Technology (BHU) Varanasi

पर्यवेक्षक/Supervisor
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(काशी हिन्दू विश्वविद्यालय)
(Banaras Hindu University)
वाराणसी/Varanasi-221005

Signature of Head of Department

अध्यक्ष एवं विभागाध्यक्ष
Professor & Head
संगणक विज्ञान एवं अभियांत्रिकी विभाग
Department of Computer Sc. & Engg
भारतीय प्रौद्योगिकी संस्थान
Indian Institute of Technology
(बनारस हिन्दू यूनिवर्सिटी)
(Banaras Hindu University)
वाराणसी-221005 / Varanasi-221005

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Effectively Task Containerization in High Performance Clustering using Artificial Intelligence

Name of the Student: Chitranjan Singh

Copyright Transfer

The undersigned hereby assigns to the Institute of Technology (Banaras Hindu University) Varanasi all rights under copyright that may exist in and for the above thesis submitted for the award of the *Doctor of Philosophy*.

Date: 20/10/2023

Place: Varanasi **(Chitranjan Singh)**

Dedicated to my family,

Mrs. Sharda Devi, Mr. Shio Shankar Singh (parents)

Mrs. Neha Singh (wife)

and

Ms. Adwita Singh, Ms. Akshita Singh (Daughters)

ACKNOWLEDGEMENT

First and foremost, I would like to thank my supervisor, Dr. Hari Prabhat Gupta, for his valuable asset and assistance. I regard it as tremendous happiness to express my profound sense of gratitude and sincere regard for his constant feedback and expertise during all these years. I am eternally grateful to have had the opportunity to work on my thesis under his supervision. He supported my research directions and allowed me to explore new ideas in the domain of Machine Learning, DevOps, and HPC Cluster. My cordial thanks to all the members of the Department of Computer Science and Engineering for creating an excellent working atmosphere.

I would also like to thank my co-supervisor Dr. Sanjay S Kadam, consultant, C-DAC and other members of my Doctoral committee, Dr. Bhaskar Biswas, Department of Computer Science and Engineering, and Dr. N.S Rajput, Department of Electronics Engineering for their help and support throughout the tenure of my studies. Special thanks to Centre for Development of Advanced Computing (C-DAC), Pune for supporting my doctoral-ship. I would also like to convey my sincere gratitude to Dr. S.K Singh, Head of Department of Computer Science and Engineering and all the RPEC and DPGC members for their suggestions and endorsement to this work. I am deeply grateful to acknowledge the National Supercomputing Mission (NSM) for providing computing resources of PARAM Smriti, which is implemented by C-DAC and funded by the Ministry of Electronics and Information Technology (MeitY) and Department of Science and Technology (DST), Government of India.

I am grateful to Dr. Rahul Mishra, Dr. Preti Kumari, Mr. Himanshu Sahu, all my dear lab members for the long discussions and their brilliant insights that have helped me to overcome the challenges I have faced in the development of this work. I also want to thank Mr. Sanjay Wandhekar, Mr. Ankur Agnihotri, Mr. Shubham Mehta, Ms. Divya Chaudhary, Mr. Abhishek Patel, C-DAC-HRD team, and my dear friends in C-DAC and IIT(BHU) Varanasi for supporting me in my hard times.

Finally, I express my heartfelt gratitude to my parents Mrs. Sharda Devi and Mr.

Shio Shankar Singh, my wife Mrs. Neha Singh, my daughters Ms. Adwita Singh and Ms. Akshita Singh, my uncle Mr. Rajballav Singh and Mr. Rajesh Singh, my brother Mr. Vishwaranjan Singh and Mr. Abhiranjan Singh, and my entire family for their constant support, love, encouragement, and sacrifices. Their affectionate love and care cannot be expressed in words.

(Chitranjan Singh)

List of Figures

List of Tables

List of Symbols

Symbol	Description
$\mathcal{N}$	Set of WMs
N	Number of WMs
τ_n	Fraction of task
Z_n	Number of containers
α	Security threshold
n	Index of WM
w_n	WM n
α_n	Probability of attack
$\gamma_{n,z}$	Price of WM n
$Q_{n,z}$	Quality of data
β	Deadline threshold
$\mathcal{P}$	Set of participants
$\mathcal{D}_i$	Dataset at participant p_i
$\mathcal{C}$	Set of clusters
M_k	Model for cluster C_k
Q_i	Total instances in $\mathcal{D}_i$
$\mathcal{F}i$	Freshness of data at p_i
W_i	Weight matrix at p_i
d	Dropout
B_i	Available Bandwidth
ΔT_i	Time elapsed of data collection
C	The set of available containers
M	The set of available shared memory locations
P	The set of performance metrics
x	Binary variable for container selection and placement
$w(p)$	Weight assigned to each performance metric

Symbol	**Description**
(p,x)	Function for system performance calculation
y	Variable matrix for container placement
y_{ij}	Variable indicating container c_i placement in m_j
c_i	Size of container c_i
m_j	Capacity of shared memory m_j
p	Threshold value for performance metric p
N'	Number of available containers
η	Learning rate
R	Number of communication rounds
E	Number of epochs
K	Count of leaf devices
S	Shared memory
t	Iteration index for communication rounds
Cj	Selected container in iteration
D_i	Local dataset of party P_i
M	Global machine learning model
Θ_i	Local model parameters of party P_i
t	Communication round index
$\Delta\theta_{it}$	Update to local model parameters of party P_i
$\Delta\theta_t$	Global model update aggregated from all parties
$\Delta\Theta_t$	Global update communicated to all parties
T	Fixed number of communication rounds
F(w)	Objective function representing the global loss
n_k	Number of data points on client i
$x \in D_i$	Data point x belonging to local dataset D_i
$f_i(w)$	Local loss function with respect to Θ_i
$F(w)$	Equation for global loss in federated learning
$\{x_{ji}, y_{ji}\}$	Radiology image data and class label
F_i	Data freshness at participant p_i
T_i	Time elapsed since radiology image collection
Q_i	Number of images in participant's dataset
C	Set of clusters in the system

Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
ANN	Artificial Neural Network
ML	Machine Learning
MM	Master Machine
WM	Worker Machine
MAR	Maximum Allowable Response time
FL	Federated Learning
KD	Knowledge Distillation
HPC	High Performance Computing
CE	Community Edition
EE	Enterprise Edition
CLI	Command-line Interface
CNCF	Cloud Native Computing Foundation
CRD	Custom Resource Definition
CI/CD	Continuous Integration/Continuous Delivery
CNN	Convolution Neural Network
RNN	Recurent Neural Network
LSTM	Long Short-Term Memory
IoT	Internet of Things
LN	LoRa Nodes
LG	LoRa Gateway
CR	Coding Rates
SF	Spreading Factors
LoRaWAN	Long Range Wide Area Network
CPU	Central Processing Unit
IoMT	Internet of Medical Things
DDoS	Distributed Denial Of Service

Abbreviation	**Description**
DNN	Deep Neural Network
GRU	Gated Recurrent Unit
FLOP	Floating Point Operations Per Second
WPM	Weight Parameter Matrix
SGD	Stochastic Gradient Descent
CASAS	Center for Advanced Studies in Adaptive Systems
OS	Operating System
VM	Vrtual Machines
TFLOPS	TeraFLOPS
C-DAC	Centre for Development of Advanced Computing
SLURM	Simple Linux Utility for Resource Management
IID	Independently and Identically Distributed
SMPC	Secure Multi-party Computation
PFS	Parallel File System