



THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學

UNIVERSITY
RESEARCH
FACILITY



**BIG
DATA
ANALYTICS**

UBDA



大數據分析中心實驗室

010011
101001
000100



INTRODUCTION

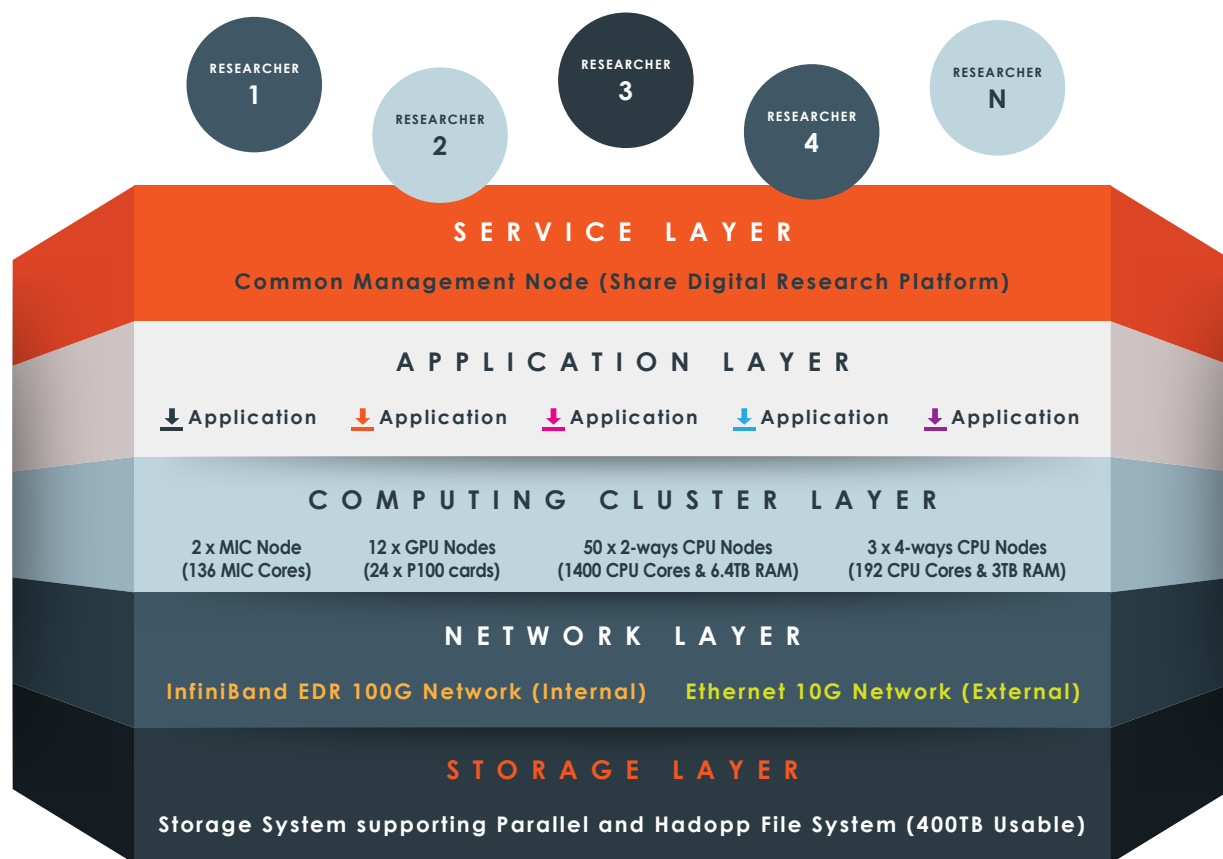
The advances in technologies of communication, IoT, social networks and other related areas have led to the recent exponential growth of data. They are generated not only from devices, sensors, Web pages but also human activities, not only about the product/service itself but also how the product/service is used. The increase in large volume, great variety and real-time data allows us to improve measurement and monitoring and, consequently, to see things with a larger scope, more details and greater clarity, in a way we could not before. Having multiple sources of data available and with several orders of magnitude more fine-grained, big data also enables us to formulate new questions and discover hidden knowledge.

While big data provides us unprecedented opportunities to generate insights for optimising decision-making and predictions, it also imposes great challenges, not only in processing and analytics techniques but also in business and application models. It calls for close collaboration between users, domain experts, data scientists and computer scientists. The main objective of establishment of UBDA is to meet the increasing demand for computing resources and expertise in big data analytics. UBDA offers an advanced infrastructure including the computing platform, data repository and data analytics tools and libraries and provides a platform for cross-disciplinary collaboration among PolyU researchers and external partners to develop, support, service and sustain research into big data analytics. It has significant value in promoting open innovation in all aspects of human, social, and technology development.

THE MISSIONS OF THE UBDA ARE:

- Build infrastructure and develop the framework, methodology and software tools to support big data management, processing and analytics.
- Foster collaboration among researchers in different disciplines for solving important and challenging problems in different application sectors and technical areas.
- Establish strong partnership with industries and organisations that may leverage big data analytics to their advantages.
- Educate and train staff, students, and professionals in the cutting-edge technology of big data analytics.

FACILITY AND SERVICE



The UBDA infrastructure has five layers.

The bottom layer is the storage layer, which consists of a 400TB storage system with parallel file systems to allow researchers to store and process their research data in a reliable environment.

The Network Layer contains two types of networks (InfiniBand and Ethernet). The InfiniBand is a high-speed 100G (EDR) network, which is mainly for internal interconnection of computing nodes with low latency and non-blocking data transfer to support big data analysis. The Ethernet network is mainly for external connection to the campus network and public internet.

The Computing Cluster Layer consists of a pool of various types of hardware server nodes, including general CPU nodes, MIC nodes and GPU nodes, which can be configured to form the required clusters for different kinds of big data analytics projects, supporting both data intensive and computation intensive processing tasks. The total initial capacity is planned to be able to support around 123 Tflops.

The Application Layer provides modelling and programming support for developing applications of different areas and is composed of domain-specific models, languages and algorithms, some of which are represented as software tools and libraries. Researchers can also install their own software to support their research.

The Service Layer is a common management layer providing the interface for accessing and using the underlying big data facility. It allows the users to log in to the UBDA system to manage their profile, access to the allocated resources, install and configure their applications, and manage their jobs through the job scheduler.

UBDA provides service for researchers to access and use the facility. The operation framework of UBDA consists of the charging model, mechanism, procedure and policy for accessing and using the central facilities.

RESEARCH AND COLLABORATION

Big data analytics provides an alternative approach, the data-driven approach, to understand the behaviour of complex systems and phenomena that cannot easily be modelled. It involves collecting and processing large datasets to correlate the data with the targeted behaviours via machine learning and distributed computing techniques. UBDA will provide technical support for a wide range of research and applications using the data-driven approach to solve important and challenging problems. UBDA will also serve PolyU academics and external partners on their needs of collaborations, providing consultancy and advisory services as well as linkage among the staff from different disciplines and application sectors.

Three main important areas are identified to be the initially focused areas of development by the UBDA, namely Smart City, Finance, and Life Science and Healthcare.

SMART CITY

The goal of smart city development is to improve the quality of life based on modern technology. It involves integrating urban data from various information systems to manage a city's assets and improve the efficiency of services within the community. Big data in smart city involves a wide range of urban and human dynamic problems such as transportation and traffic management, human mobility, tourism and hospitality, online retail and entertainment, ageing population, and online/offline social network, among others. For different sources and types of urban and human data, various kinds of data analytic tools in statistics, optimisation and data mining can be applied to extract useful information and patterns for better planning and decision-making that could improve the efficiency of the city and the life quality of residents.

FINANCE

Facing increasing competition, regulatory constraints and customer needs, financial institutions are seeking new ways to leverage technology innovation to enhance efficiency and improve services. Big data analytics can be used to gain

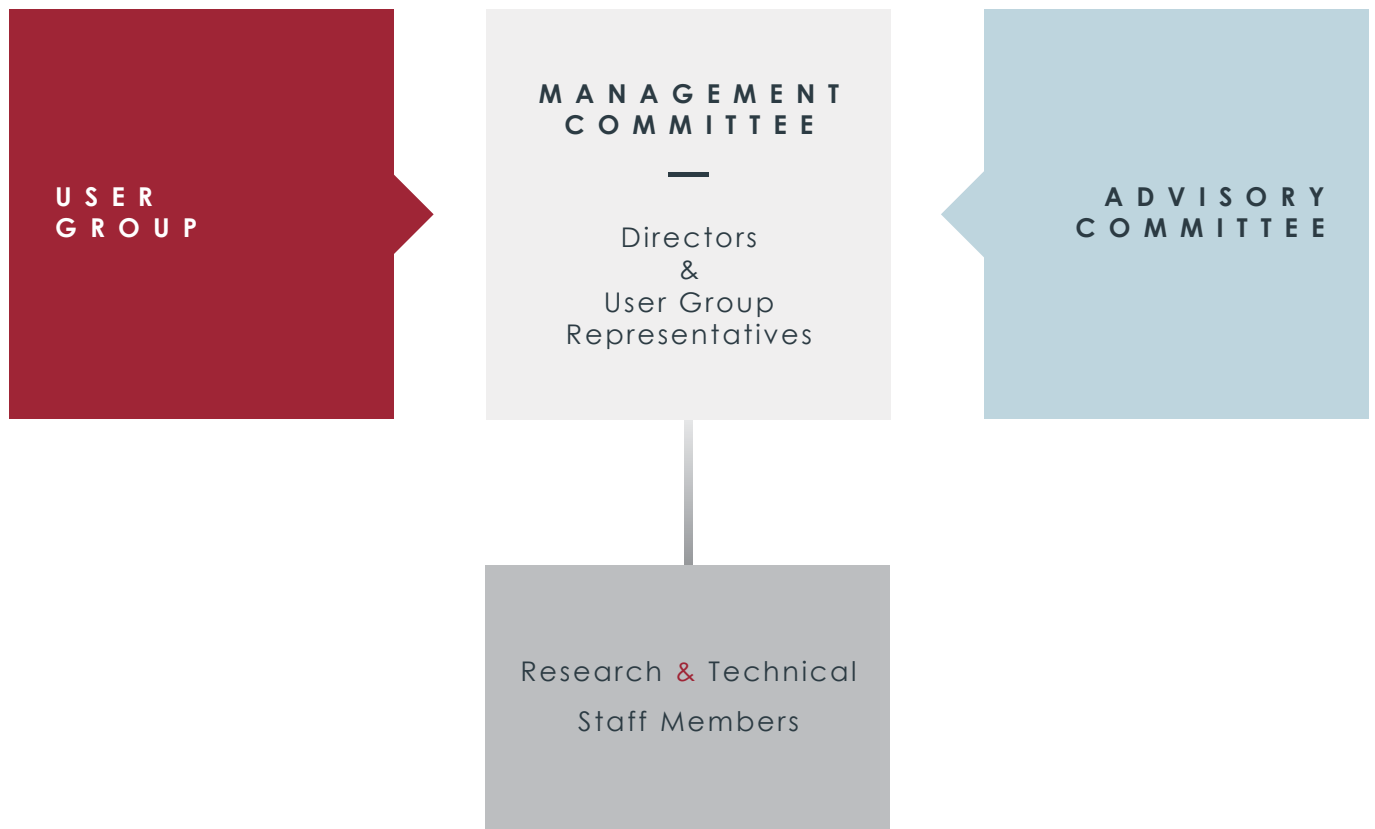
a competitive advantage. Financial services, in particular, have widely adopted big data analytics to inform better investment decisions with consistent returns. Apart from trading, analysis of potential financial crises is another important issue. As they arise from the complex interplay of decisions made by many individuals, analysing detailed market data and the complex networks that underlie them have attracted a lot of scientific attention.

LIFE SCIENCE AND HEALTHCARE

Life science is a sector that has been early adopter of big data because of the rapid generation of large and complex biomedical data created by devices across the globe every day. Life sciences data comes in many forms, which includes information about molecular pathways, genomic sequences and populations of people. Scientists need to overcome the challenges of handling such complexity of information. Different analytical tools and techniques for big data have been developed which will allow researchers to gain a better understanding of the basic biological mechanisms and how the results can be applied in healthcare. The complexity of analysing big data stems from combining different types of information, such as data from genomics, proteomic, cellular signalling, clinical research and environmental studies. The results could lead to new insights into the treatment of diseases and the prediction of the most likely response to treatment. The healthcare industry also is on the brink of a major transformation through the use of advanced and big data analytics technologies, providing value-based, patient-centric care, real-time monitoring of patients and reducing fraud, waste, and abuse.

At PolyU, our colleagues have carried out research and development projects with external collaborators in the three areas and accumulated remarkable experiences and results. Our collaborators include Alibaba, Huawei, Airport Authority Hong Kong, The Hong Kong R&D Centre for Logistics and Supply Chain Management Enabling Technologies (LSCM), and GogoVan.

ORGANISATION OF UBDA



The UBDA organisation structure consists of the Directors, Advisory Committee, Management Committee and User Group.

DIRECTOR

Jiannong CAO

Chair Professor
Department of Computing
The Hong Kong Polytechnic University

ASSOCIATE DIRECTOR

Xiaojun CHEN

Chair Professor
Department of Applied Mathematics
The Hong Kong Polytechnic University

ADVISORY COMMITTEE MEMBERS

Jianqing FAN

Frederick L. Moore '18 Professor of Finance
Professor of Statistics
Professor of Operations Research and
Financial Engineering
School of Engineering and Applied Science
Princeton University

Dean
School of Data Science
Fudan University

Daniel LAI

Director
Digital Technology and Consultancy Co. Ltd.
Distinguished Fellow and Past President
Hong Kong Computer Society

Hui LEI

Director & Chief Technology Officer
IBM Watson Health Cloud

Zhiquan LUO

Director
Shenzhen Research Institute of Big Data
Vice President (Academic)
The Chinese University of Hong Kong, Shenzhen

Helen MENG

Professor & Chairman
Department of Systems Engineering and
Engineering Management
The Chinese University of Hong Kong
Director
Stanley Ho Big Data Decision Analytics
Research Centre
The Chinese University of Hong Kong

Qiang YANG

New Bright Professor of Engineering &
Chair Professor
Department of Computer Science and
Engineering
Hong Kong University of Science and
Technology

Peter YEUNG

Head, ICT & GreenTech Cluster
Hong Kong Science and Technology Parks
Corporation

Philip YU

UIC Distinguished Professor &
Wexler Chair in Information Technology
Department of Computer Science
University of Illinois at Chicago

Jingren ZHOU

Vice President
Alibaba Group

Albert ZOMAYA

Chair Professor of High Performance
Computing and Networking
School of Information Technologies
The University of Sydney

Director
Centre for Distributed and High Performance
Computing

University Research Facility in Big Data Analytics (UBDA)

P505, The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

ubda.enquiry@polyu.edu.hk

www.ubda.polyu.edu.hk