



## **TERMS OF REFERENCE FOR THE SUPPLY, DELIVERY, TRAINING, INSTALLATION, TESTING AND COMMISSIONING OF ADVANCED DATA ANALYTICS SYSTEM**

### **A. OBJECTIVE**

Data analytics is the process of collecting, organizing and analyzing different sets of data to discover patterns, trends and other useful information. These structured and unstructured data are extremely large in volume that is difficult to process using traditional computing systems and techniques. Almost all devices nowadays are IoT-enabled (Internet-of-Things). IoT is simply the network of interconnected things/devices which are embedded with sensors, software, network connectivity and necessary electronics that enables them to collect and exchange data making them responsive. Sensors from automatic weather station can immediately send out real-time data from the readings of barometric pressure, wind speed, precipitation, temperature and humidity which are needed in weather forecasting. Meteorological images, such as satellite and radar data, can be used as input for developing future models by using image/pattern recognition of machine learning algorithms. The data when captured, formatted, manipulated, stored and then analyzed, can provide plenty of information that can be used to improve forecasting operations and make faster, intelligent decisions in providing advanced warning for occurrence of tropical cyclone, storm surge, torrential rain and other extreme weather events.

Weather forecasting has traditionally been done by physical models of the atmosphere, which are unstable to perturbations, and thus are inaccurate for large periods of time. Since machine learning techniques are more robust to perturbations, using the application to weather forecasting may potentially generate more accurate weather forecasts for large periods of time. The project will help forecaster in providing automated weather information using the Artificial Intelligence technology. The AI technology is being applied in weather-related applications in NICT Japan, and with this project, similarly, can be implemented in PAGASA. The project will also provide two AI-driven products: the error correction model for rainfall forecasting; and the automated rainfall prediction based on observed data and satellite images.

### **B. APPROVED BUDGET FOR THE CONTRACT (ABC)**

The Approved Budget for the Contract is **FORTY TWO MILLION NINE HUNDRED NINETY THREE THOUSAND PESOS (Php 42,993,000.00)** inclusive of VAT and all applicable government taxes.

### **C. DELIVERY PERIOD AND PLACE OF DELIVERY**

The winning Bidder shall supply, deliver, install, test and commission the Advanced Data Analytics System facility including the requisite training and technical meeting at PAGASA Data Information Center, WFFC Compound, Agham Road, Diliman, Quezon City within the period of one hundred twenty calendar days (120 c.d.) commencing from the date of the issuance of the Notice to Proceed (NTP). Moreover, AI training workshops, collaboration,

and technical meetings related to the data analytics and advancement of AI technology in PAGASA, including application development, shall be provided by the winning Bidder within one hundred twenty calendar days (120 c.d.).

#### **D. TECHNICAL SPECIFICATIONS AND ADDITIONAL REQUIREMENTS**

To ensure smooth and organized project implementation activities including continuous operation and maintenance of the data analytics system, the winning Bidder shall provide the following:

1. List of their on-site local and warm body technical support personnel to be deployed for the project, accompanied by curriculum vitae and proof of their competency such as, but not limited to diplomas and certificates. The list of technical support personnel must include their certified engineers/technicians highly knowledgeable on the proposed computing hardware, and certified data scientists/analyst highly proficient on the offered software application or data analytic tools.
2. Project management plan of activities and timelines such as Gantt chart showing the proposed schedule of the delivery, training, installation, testing and commissioning of the project. Ensure to conduct required meetings with the end-users and technical support personnel from kick-off meeting to final wrap-up. Meals during the meetings must be shouldered by the winning Bidder.
3. Service Level Agreement (SLA) which clearly defines level of service and indicates the technical support tiers with the name of support personnel, position and role including contact number and email addresses.
4. Engineering Plan and Block Diagram (i.e., electrical and networking) of the data analytics system, its power supply requirement, the grounding system and other accessories with complete sets of brochures. The Engineering Plan and Block Diagram should be duly signed by Professional Electrical Engineer.
5. List of recommended spare parts, both serviceable and disposable, w/ corresponding prices and guarantee of availability in the market within the next five years.

The winning Bidder shall supply, deliver, train, test and commission the Advanced Data Analytics System at PAGASA WFFC Building, Quezon City based on the following minimum hardware specifications:

##### **1. HARDWARE DELIVERABLES**

<b>ITEM NO.</b>	<b>QTY &amp; UNIT</b>	<b>DESCRIPTION</b>	<b>SPECIFICATIONS</b>
1	1 Unit	Login Master Node	Each node have at least: <ul style="list-style-type: none"><li>• 2 x Intel Xeon Gold 5218 16C 2.3GHz Processor</li></ul>

			<ul style="list-style-type: none"> <li>• 12 x 16GB (192GB) TruDDR4 2933MHz RDIMM</li> <li>• 1 x 100GbE PCIe Adapter</li> <li>• 1 x 2.5" SATA/SAS 8-Bay Backplane</li> <li>• 2 x 2.5" 600GB 15K SAS 12Gb HS 512n HDD</li> <li>• 4 x 1GbE Ports RJ45</li> <li>• 1 x RAID 530-8i PCIe 12Gb Adapter</li> <li>• 2 x Platinum Hot-Swap Power Supply</li> <li>• With complete accessories &amp; peripherals</li> </ul>
2	3 Units	High-Speed GPU Computing Node	<p>Each node have at least:</p> <ul style="list-style-type: none"> <li>• 4 x Intel Xeon Gold 6248 20C 2.5GHz Processor</li> <li>• 12 x 16GB (192GB) TruDDR4 2933MHz RDIMM</li> <li>• 1 x 100GbE PCIe Adapter</li> <li>• 4 x NVIDIA Tesla V100 32GB PCIe Passive GPU</li> <li>• 2 x 2.5" 300GB 10K SAS 12Gb HS 512n HDD</li> <li>• 2 x 2.5" 480GB 5200 Entry SATA 6Gb HS SSD</li> <li>• 1 x 2.5" SATA/SAS 8-Bay Backplane</li> <li>• 2 x 1GbE Port RJ45</li> <li>• 1 x RAID 530-8i PCIe 12Gb Adapter</li> <li>• 2 x Platinum Hot-Swap Power Supply</li> <li>• With complete accessories &amp; peripherals</li> </ul>
3	2 Units	Filesystem Storage Node	<p>Each node have at least:</p> <ul style="list-style-type: none"> <li>• Intel Xeon Gold 5218 16C 2.3GHz Processor</li> <li>• 6 x 16GB (96GB) TruDDR4 2933MHz RDIMM</li> <li>• 1 x 100GbE PCIe Adapter</li> <li>• 2.5" 600GB 15K SAS 12Gb HS 512n HDD</li> <li>• 1 x 2.5" SATA/SAS 8-Bay Backplane</li> <li>• 4 x 1GbE Ports RJ45</li> <li>• 2 x Platinum Hot-Swap Power Supply</li> <li>• With complete accessories &amp; peripherals</li> </ul>
4	2 Units	Hybrid Flash Storage Array Subsystem	<p>Each storage subsystem unit:</p> <ul style="list-style-type: none"> <li>• Must have at least 500TB usable capacity</li> <li>• Must be a scalable parallel filesystem solution.</li> <li>• Should support and implemented with high performance interconnect of Omni Path Architecture (OPA) or Infiniband.</li> <li>• Should support low latency RDMA communication for data transfer between compute node and fileservers.</li> <li>• Should have no single point of failure in the overall architecture and the solution should be highly available so that there is no loss of access to data even in the event of a complete node or controller failure.</li> <li>• Must support detecting access pattern as sequential, random, and fuzzy sequential, or strided, &amp; prefetches striped data in parallel.</li> <li>• Must support using multiple physical disks spread across multiple nodes to store a file in separate blocks.</li> <li>• Must support distributed lock management with fine-grained block-level locking and byte-range locking.</li> <li>• Must support data replication (mirroring) at a file, directory tree, or file system level.</li> </ul>

			<ul style="list-style-type: none"> <li>• Must support filesystem journaling where all filesystem operations are logged.</li> <li>• Improved performance and data protection with Dynamic Disk Pools (DDP) technology, as well as support for traditional RAID.</li> <li>• Have storage management functions which include Dynamic Disk Pools, SSD read cache, snapshots, volume copy, and thin provisioning.</li> <li>• Intuitive, web-based GUI for easy system setup and management.</li> <li>• Dual-active controller modules with automatic load balancing and failover</li> <li>• Designed for 99.999% availability with redundant hot-swap components, including controllers and I/O modules, power supplies, proactive maintenance, and non-disruptive firmware upgrades.</li> </ul>
5	1 Unit	GPU Visualization Node	<p>Each node have at least:</p> <ul style="list-style-type: none"> <li>• 2 x Intel Xeon Gold 5218 16C 2.3GHz Processor</li> <li>• 24 x 16GB (384GB) TruDDR4 2933MHz RDIMM Memory</li> <li>• 2 x NVIDIA Quadro P6000 24GB GPU</li> <li>• 1 x 100GbE PCIe Adapter</li> <li>• 2 x 2.5" 600GB 15K SAS 12Gb HS 512n HDD</li> <li>• 4 x 1GbE Ports RJ45</li> <li>• 1 x RAID 530-8i PCIe 12Gb Adapter</li> <li>• 2 x Platinum Hot-Swap Power Supply</li> <li>• With complete accessories &amp; peripherals</li> </ul>
6	1 Unit	High Speed Interconnect Switch	<p>100Gbps Low Latency Interconnect</p> <ul style="list-style-type: none"> <li>• Must include a scalable system interconnect that provides quick data access with low system overhead and efficient scheduling</li> <li>• Interconnect should support parallel filesystem and capable of unidirectional throughput of at least 100Gb/s, and bi-directional throughput of at least 198Gb/s</li> <li>• Interconnect of messages rates of at least 150 million/second</li> <li>• With complete cable accessories &amp; peripherals</li> </ul>
7	1 Unit	Management Switch	<ul style="list-style-type: none"> <li>• Must have at least 48 x 10/100/1000BASE-T RJ-45 ports and 4 x 10 Gigabit Ethernet SFP+ uplink ports</li> </ul>
8	1 Unit	42U Server Rack Cabinet	<ul style="list-style-type: none"> <li>• Must be industry standard, adjustable 19" rack</li> <li>• Must be perforated, lockable front &amp; rear doors</li> <li>• Rear cable management channels must have space for mounting 0U strips PDUs (power distribution units) and vertically mounted</li> <li>• Must have KVM console for server management</li> <li>• With complete power cable accessories &amp; peripherals</li> </ul>
9	1 Unit	VPN Firewall	<ul style="list-style-type: none"> <li>• 14 x GE RJ-45 Ports + 2 x GE RJ-45 WAN Ports</li> <li>• 1/2/1 x GE RJ-45 Management/HA/DMZ Ports</li> <li>• Firewall throughput at least 6.6 Mpps</li> <li>• IPsec VPN Throughput (512 byte) of 4Gbps</li> </ul>

			<ul style="list-style-type: none"> <li>• 2K Gateway-to-Gateway IPsec VPN Tunnels</li> <li>• 10K Client-to-Gateway IPsec VPN Tunnels</li> <li>• 250 Mbps SSL-VPN Throughput</li> </ul>
10	10 Units	Unix/Linux Notebook	<p>Each unit have at least:</p> <ul style="list-style-type: none"> <li>• 13.3-inch LED-backlit Display</li> <li>• Resolution display up to 2560 x 1600 pixels</li> <li>• Intel Core i7 Processor with 16 GB Memory</li> <li>• 512 GB SSD Storage</li> <li>• Unix or Linux Operating System</li> <li>• With installed Office productivity applications (Microsoft Office alternatives)</li> <li>• Jupyter with Python and R</li> <li>• With complete accessories &amp; peripherals</li> </ul>
11	3 units	Unix/Linux Desktop	<p>Each unit have at least:</p> <ul style="list-style-type: none"> <li>• Processor Intel Core i7 10<sup>th</sup> Generation</li> <li>• Memory 16GB DDR4 or any compatible</li> <li>• Intel motherboard or any compatible</li> <li>• Integrated graphic processor or external</li> <li>• Storage 512GB SSD (OS) + 1TB HDD</li> <li>• Dual GbE LAN port RJ45</li> <li>• Wireless LAN Interface supports IEEE 802.11g/n/ac</li> <li>• Unix or Linux Operating System</li> <li>• With installed office productivity applications (Microsoft Office alternatives)</li> <li>• Jupyter with Python and R</li> <li>• 3 x LED 21" monitor with single desk stand</li> <li>• UPS 480Watts / 800VA</li> <li>• Complete power cables and accessories</li> </ul>
12	2 units	Ink Tank Printer (A3)	<p>Each unit have at least:</p> <ul style="list-style-type: none"> <li>• High capacity integrated ink tanks</li> <li>• Technology: On-demand inkjet (Piezoelectric)</li> <li>• Maximum Print Resolution: 5760 x 1440 dpi</li> <li>• Bi-directional or Uni-directional printing</li> <li>• Print Speed: Approx. 30 ppm (Draft, A4, Black) and 17 ppm ( Draft, A4, Colour)</li> <li>• Standard Paper Input Capacity: Up to 100 sheets, A4 Plain paper</li> <li>• Standard Paper Output Capacity: Up to 50 sheets, A4 Plain paper</li> <li>• Maximum Paper Size: 12.95 x 44"</li> <li>• Paper Sizes: A3+, A3, B4, A4, A5, A6, B5, 10x15cm (4x6), 13x18cm (5x7"), 16:9 wide size, Letter (8.5x11"), Legal (8.5x14"), Half Letter (5.5x8.5"), 9x13cm (3.5x5"), 13x20cm(5x8"), 20x25cm(8x10")</li> <li>• Paper Feed Method: Friction feed</li> <li>• Additional 1 set of ultra-high capacity ink bottle/refill (black, cyan, magenta &amp; yellow)</li> </ul>

## 2. SOFTWARE APPLICATION TOOLS

<b>I.     Orchestration and management software</b>
<ul style="list-style-type: none"> <li>a) The proposed hardware solution must include the integrated software tool that simplifies the deployment of data analytics and AI tasks in model development including the management of all system hardware components.</li> <li>b) It must include an open source cluster management software stack, consolidating the management, monitoring and scheduling functions into a single platform.</li> <li>c) It should include a user-friendly web-based portal to execute, monitor and manage analytics and AI jobs on a distributed cluster.</li> <li>d) It should include workflow templates to provide an intuitive starting point for less experienced users.</li> <li>e) It should include management of private space on shared storage through the GUI</li> <li>f) It should be capable of monitoring of job progress and log access.</li> <li>g) It should include console access for users with advanced command-line skills.</li> <li>h) It must have a single management portal consolidating the monitoring, alarms, and reporting with generation of notifications and alarms based on system status.</li> <li>i) It should include user management and multi-user support with user and groups.</li> <li>j) It should be compatible with popular shared file systems such as Spectrum Scale, NFS, Lustre, and Ceph.</li> <li>k) Enhanced end-user functionality to support AI model training and management</li> <li>l) Management of private space on shared storage through the GUI and monitoring of job progress and log access.</li> <li>m) AI Runtime module definition and pre-loading with job execution.</li> <li>n) Accelerated AI templates to provide training and inference capabilities for many common AI use cases.</li> <li>o) TensorBoard visualization tools integrated into the interface (TensorFlow-based).</li> <li>p) Container-based user management of supported AI frameworks and applications (through Singularity).</li> </ul>
<b>II.    AI-driven application development</b>
<b>1. Error correction model for rainfall forecasting</b>
<ul style="list-style-type: none"> <li>a) Must be able to suggest correction adjustments to existing weather forecasting models for rainfall.</li> <li>b) The software must be able to consume grib2, satellite image, and bin files as input to the model.</li> <li>c) Must be able to produce a prediction in less than 2 hours only for specified location domain.</li> <li>d) Must be able to accept new observed data for training daily.</li> <li>e) Must be able to produce prediction that can be easily visualized using GIS or GrADS.</li> <li>f) Must be able to access and organize necessary data for training.</li> <li>g) Each model should be versioned.</li> <li>h) Models deployed can be changed to a different existing version by admins.</li> <li>i) Models should be interpretable through statistical means.</li> <li>j) The architecture of the models should be designed as needed by the analyst.</li> <li>k) The model should be able to predict the succeeding error corrections for the next 3, 6, 9 and 12 and 24 hrs.</li> </ul>

## 2. Rainfall Prediction based on observed data and satellite images

- a) The software should be able to accept numerical data from PAGASA stations and international models together with corresponding satellite images on a certain period.
- b) The software can only be used to predict rainfall on areas where ground data is available.
- c) The software should be able to convert raw data to images and use such images as input to the training model.
- d) The output forecasts of the models should be comparable to that of the international models.
- e) The software must be able to utilize grib2, satellite image, and bin files as input to the model.
- f) Must be able to produce a prediction in less than 2 hours only for specified location domain.
- g) Must be able to accept new observed data for training daily.
- h) Must be able to produce prediction that can be easily visualized using GIS or GrADS.
- i) Must be able to access and organize necessary data for training.
- j) Each model should be versioned.
- k) Models deployed can be changed to a different existing version by sysadmins.
- l) Models should be interpretable through statistical means.
- m) The architecture of the models should be designed as needed by the input data.
- n) Only authorized analysts should have access to modify deployed models.
- o) The model should be able to predict the succeeding amount of rainfall for the next 3, 6, 9 and 12 and 24 hrs.
- p) The model should be able to predict the amount for rainfall specifically for selected region/cities.

## **E. NETWORKING AND DATA CABLING**

The winning Bidder shall supply, install and configure all necessary networking devices, peripherals, cables and accessories for the connectivity of the system to the Weather Division local area network (LAN).

## **F. ELECTRICAL SYSTEM**

The prospective bidders shall coordinate with the electrical group of PAGASA Engineering and Technical Services Division (ETSD) regarding the electrical system requirements before submitting the required Engineering Plan and Block Diagram.

## **G. FACTORY ACCEPTANCE TEST (FAT) AND FACTORY TRAINING**

The Factory Acceptance Test (FAT) shall be conducted at the factory site of the advanced data analytics system, to be witnessed and accepted by three (3) PAGASA personnel from any of the middle to top management staff. The FAT shall be conducted within five (5)



calendar days. Simultaneously, a 5-day Factory Training on advanced systems management and maintenance shall also be conducted and to be attended by three (3) qualified technical personnel of PAGASA who are directly involved in the operation, management and maintenance of the AI analytics system facility.

All related expenses during the FAT and Factory training, such as, but not limited to, round trip airfare, local transportation, lodging/accommodation, training materials and allowable travel expenses based on the prevailing UNDP-DSA rates for each participant shall be borne by the winning Bidder.

## **H. SITE ACCEPTANCE TEST (SAT) AND ON-SITE TRAINING**

Site Acceptance Test (SAT) shall be conducted on-site at WFFC Compound, Agham Road, Diliman, Quezon City. The purpose of the test is to verify the performance of the system in accordance with the specifications and functional requirements. The series of tests must be based on detailed checklist test provided by the winning Bidder and to be performed by the end-user for acceptance. Any defect or deviation discovered during the site acceptance test shall be rectified by the winning Bidder immediately or within a maximum period of one (1) month from the completion of the test. After such rectification, another testing shall be made to verify the rectification. The SAT shall be witnessed and accepted by three (3) PAGASA personnel from middle to top management staff, end-user, and technical staff. The activity shall be conducted within a total of seven (7) calendar days.

Moreover, a 5-day on-site, end-users training on Advanced Data Analytics system shall also be conducted. All expenses related to the on-site training, including meals shall be provided by the winning Bidder and to be attended by ten (10) participants or end-users. The Site Acceptance Test (SAT) may be conducted simultaneously with the On-Site Training.

## **I. DATA ANALYTICS AND AI TRAINING WORKSHOP**

A 5-day training workshop on Analytics and AI shall be attended by five (5) PAGASA personnel and shall be done in the AI Technology Center of the Manufacturer's country of origin. The workshop includes AI Technology Update activities and AI Technology hub site visit. All related expenses, such as, but not limited to the training materials, round trip airfare, local transportation, lodging/accommodation and allowable travel expense based on the prevailing UNDP-DSA rates for each participant shall be borne by the winning Bidder.

An in-depth training on the developed model shall be conducted on-site. The said training shall be activity-based, to promote effective and active learning. Its core premise shall include hands-on activities not only geared toward familiarizing more with the AI application used in the model, but likewise running the model in the analytics computing system facility. The training shall be attended by at least ten (10) PAGASA end-users. Training materials and meals shall be provided to the participants by the Winning Bidder. Moreover, it shall also provide self-paced, online intensive training with facilitated lab from online courses provider on Python, Jupyter and R, the programming language commonly used in data analytics.



## **J. SYSTEM COMMISSIONING**

After the satisfactory conclusion of the Site Acceptance Test, the Winning Bidder shall show and demonstrate the capability and performance of the operational data analytics/AI system. Acceptance test shall also be conducted on all components and peripherals. The successful demonstration based on acceptance checklist thereof, shall mean that the project Advanced Data Analytics System has been commissioned.

## **K. WARRANTIES**

All workmanship, system parts, accessories, other materials and equipment and services shall be warranted by the winning Bidder and shall have 3-year maintenance support services warranty on server nodes, switches, and rack cabinet, 2-year standard warranty on VPN firewall, and 1-year warranty on notebook, desktop, and printer. The warranty shall also include the following:

- 1) Software and firmware updates, replacement of parts or the units itself, repair and troubleshooting within the respective subscription period.
- 2) Availability of Technical Support services 24/7 via telephone, text, and email which include Remote Access Assistance thru Internet web or VPN access.
- 3) Prospective bidders shall provide the end-user a copy of the Service Level Agreement (SLA) clearly indicating the technical support tiers.
- 4) For severe or critical issues, the Service Level Agreement (SLA) should indicate that response time must be available within 4 hours.
- 5) The winning Bidder should not assign, transfer, pledge nor subcontract any part or interest therein.

The winning Bidder shall assure that any manufacturing defects found on any hardware shall be corrected within the warranty period.

## **L. AFTER SALES SUPPORT**

The winning Bidder shall include in its bid a commitment for five (5) years support to PAGASA for the repair and maintenance of the equipment to be supplied specifically the server nodes and switches. It shall include in its commitment a provision of a reliable, swift and efficient on-site 24/7 support, availability of ticketing and response system, and ensure a quick and readily available supply of spare and replacement parts especially for the server nodes where the computing activities are extensive and critical.

## **M. SYSTEM DOCUMENTATION**

The winning Bidder shall provide installation, operations and maintenance manuals to the end-user. The manuals shall also include the system configuration of the nodes, switches and routers. It shall also contain among others the complete and detailed schematic diagrams, theory of operations, calibration and maintenance procedures. All other hardware and software requirements shall also be turned-over to PAGASA prior to the issuance of the Final Inspection and Acceptance report. In addition, the winning Bidder shall provide a complete list of deliverables and installation materials.