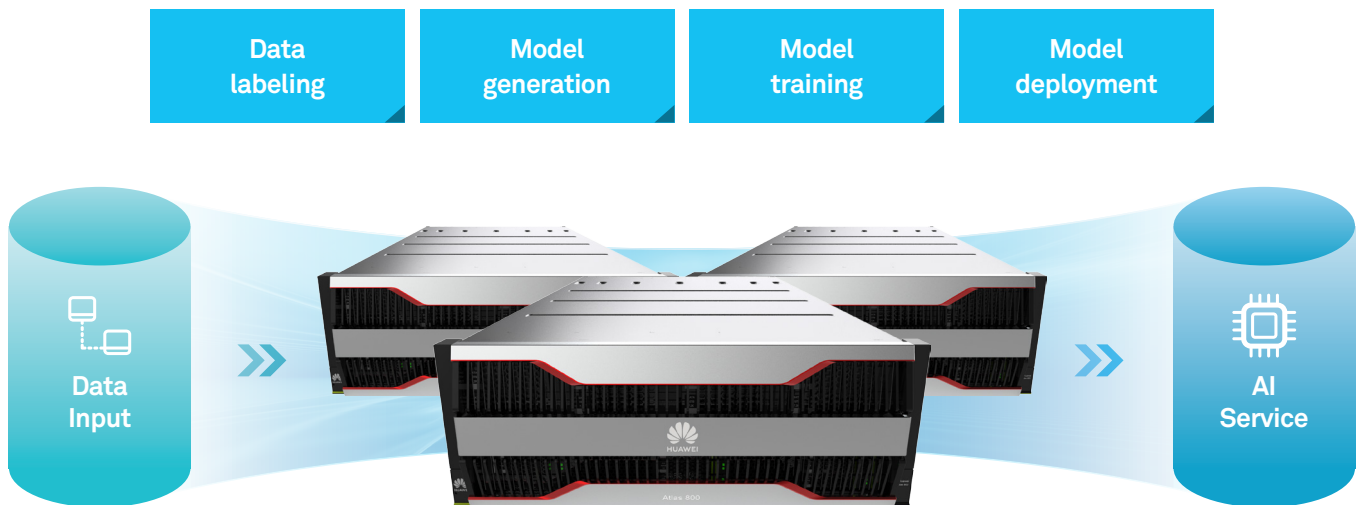


## Enterprise-Level Deep Learning Platform, One-Stop AI Enabling Solution

Tailored for AI developers and data researchers, the Huawei Atlas 800 AI appliance provides end-to-end capabilities in deploying the data labeling, model generation, model training, and model inference services. Available in integrated delivery of software and hardware, this appliance reduces the entry technical requirements of AI application and enables quick development and rollout of AI services for customers.



Deep Learning | Model Training | Recommendation



### Out-of-the-Box Installation

Ready to work in 2 hours with a preinstalled AI development environment, underlying software library, and development framework.

Automatic model generation with AutoDL, automatic hyperparameter tuning, and one-click model deployment.



### Ultimate Performance

Optimized AI environment based on a standard framework and programming environment.

High-performance scheduling algorithms, improving resource utilization by over 15%.

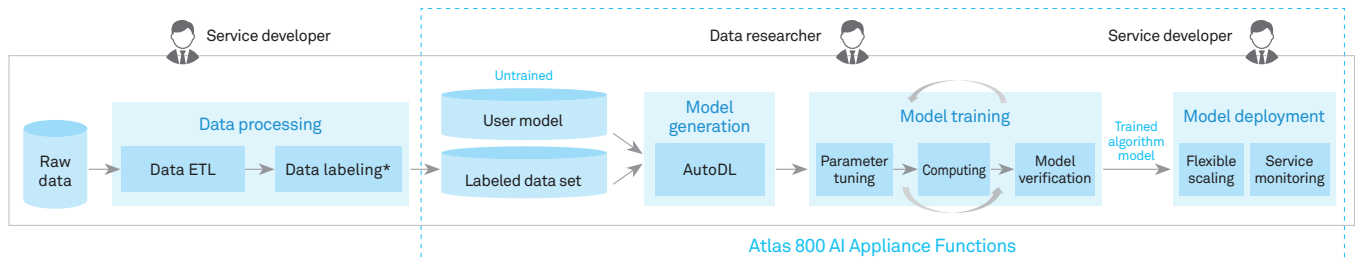


### Integrated Management

Comprehensive management of resource utilization, health monitoring, and job scheduling.

Easy-to-use WebUI, more intuitive and efficient than the command line interface (CLI).

## Atlas 800 AI Training Process



|   |   |   |  |   |
|---|---|---|--|---|
| <b>Code Development</b><br>Integrated the Jupyter Notebook development environment for code compilation and document creation and sharing, simplifying workflows and facilitating collaboration | <b>Preset Models</b><br>Trained models are preconfigured for typical service scenarios. Users can use their own data to perform secondary training on the preconfigured models to obtain optimization models. | <b>Training Management</b><br>Supports mainstream deep learning frameworks such as TensorFlow, PyTorch, Caffe, and mxnet. Integrates multiple visualization tools to facilitate real-time tracking of training processes. | <b>Model Management</b><br>Data models obtained after training jobs are complete can be deployed as online model inference services and can be published as RESTful API services by one click. | <b>O&amp;M</b><br>Real-time status monitoring supports fault diagnosis, alarming, and isolation. Cluster management and one-click firmware upgrade are supported, improving O&M efficiency. |
|---|---|---|--|---|

## Atlas 800 Product Specifications

|  |                          |                        |                     |  |
|--|--------------------------|------------------------|---------------------|--|
| Cluster                                | Max. nodes               | 128                    | AI Accelerator Card | Each full-width AI accelerator card: 8 full-height full-length AI accelerator cards (PCIe or NVLink) |
|  | Users                    | 1000                   |                     | Server   |
| Recommended jobs (per user)            | Training                 | 10                     | Hard Drives         |  |
|  | Inference                | 5                      |                     | RAID   |
|  | Visualization            | 5                      | I/O                 | 4*PCIe x16 LP+2*10GE LOM   |
| WebUI & Services Processing Capability | Development environment  | 4                      | Power Supply Units  | 4 hot-swappable 2200 W AC or 240 V HVDC PSUs, with support for N+N redundancy                        |
|  | Maximum online users     | 30                     |                     | Fan Modules  |
|  | Max. concurrent requests | 30 per second          | Temperature         |  |
| Request processing latency             | 3 seconds                | Dimensions (H x W x D) |                     | 175 mm x 447 mm x 790 mm (6.89 in. x 17.60 in. x 31.10 in.)  |
| API Processing Capability              | Max. online users        |                        | 30                  |  |
|  | Max. concurrent requests | 30 per second          |                     |  |
| Request processing latency             | 3 seconds                |                        |                     |  |

\*Note: Data labeling is planned to release in 2019Q3.

For more information, please visit

<https://e.huawei.com/en/solutions/business-needs/data-center/atlas>

