**OpenStation**

**技术白皮书**

**元物质（深圳）智能科技有限公司**

**2022年3月**

**致谢与声明**

尊敬的用户：

衷心感谢您选用了元物质openstation系统！

本手册介绍了元物质openstation系统所能实现的功能，可使您更好地了解本软件的使用范围和使用方法。 元物质（深圳）智能科技有限公司拥有本手册的版权。 未经 元物质（深圳）智能科技有限公司许可，任何单位和个人不得以任何形式复制本用户手册。 元物质（深圳）智能科技有限公司保留随时修改本手册的权利。

本手册中的内容如有变动恕不另行通知。

如果您对本手册有疑问或建议，请向元物质（深圳）智能科技有限公司垂询。

**声明**

**最终用户软件许可协议**

请务必仔细阅读并理解以下最终用户许可协议（以下简称《协议》）。一旦安装或使用随此《协议》提供的软件（以下简称"软件"），即表明您同意本《协议》的各项条款。如果您不同意，则请不要使用"软件"。

**通则**

本《协议》是您（个人或单位实体）与 元物质（深圳）智能科技有限公司（"元物质"）之间达成的法律协议。本《协议》适用于元物质可能向您提供或使您可以得到的"软件"的更新、补充，附加组件或基于 Internet 的服务组件，除非元物质对这些更新、补充、附加组件或基于 Internet 的服务组件规定了其他条款。元物质有权停止通过使用"软件"而向您提供或使您可以得到的任何基于 Internet 的服务。与"软件"有关的任何产品支持服务也受此《协议》的约束，除非在您和元物质之间达成的其他协议中另有规定。"软件"中可能附带了对本《协议》的修正或补充。

**许可证的授予**

1.在元物质授权使用的机器上使用，出于备份或档案管理的目的，可以制作本软件的拷贝。

2.本软件只限购买者本人（本单位）使用，如转让或转卖，本公司将追究其法律责任。

3.不得通过非正常途径，强制性进行注册授权，增加受控节点管理数。

**被授权人所应该履行的义务**

1.禁止复制和扩散光盘。

2.禁止对本软件进行逆向工程、反汇编或解体拆卸。

3.禁止以任何方式将本软件中的部分或全部数据用于商业目的。

4.您将本软件或拷贝的全部或局部转手给另一使用方时，您的许可权即自行终止。

**本软件的版权和所有权：**

本软件及其所有拷贝的名称，与光盘上或本软件中注明的公司同在。本软件及文档享有版权，并受国家版权法及国际条约条款的保护。您不得从本软件中删除其版权声明。并保证为本软件的拷贝（全部或部分）复制版权声明。您同意制止以任何形式非法拷贝本软件。

**售后担保：**

元物质（深圳）智能科技有限公司担保，在正常使用的情况下，自售出之日起九十天内，其软件载体无材料或工艺缺陷。经验证确有缺陷时， 元物质（深圳）智能科技有限公司的全部责任就是退换其软件载体。也是给您的唯一补偿。您若是通过购买 元物质（深圳）智能科技有限公司其它产品而获得此软件的免费赠送，则按相关产品的售后服务条款处理担保事项。因事故、滥用或错误应用导致的载体缺陷，或者导致其它损坏，售后担保无效。退换的载体享受原担保期剩余时间，或三十天的担保。取其长者优先。

**商标声明：**

是 元物质（深圳）智能科技有限公司的注册商标。其他商标分别属于其相应的注册公司。 责任限制：

上述担保，无论是明示的或是暗示的，均为担保的全部内容，包括对特殊应用目的的商品性和适应性担保。无论遵循本协议与否，就使用本软件而产生的：利润损失、可用性消失、商业中断，或任何形式的间接、特别、意外或必然的破坏，或任何其他方的索赔， 元物质（深圳）智能科技有限公司及其代理、销售人员均不承担 IV 任何赔偿责任。即使 元物质（深圳）智能科技有限公司事先被告知此类事有可能发生， 元物质（深圳）智能科技有限公司及其代理、销售人员亦不承担赔偿责任。

**许可终止：**

您若违反本协议的任一条款与条件， 元物质（深圳）智能科技有限公司可能随时会终止许可。终止许可之时，您必须立即销毁本软件及文档的所有拷贝，或归还给 元物质（深圳）智能科技有限公司。

**适用法律：**

本协议适用中华人民共和国法律，包括但不限于《知识产权海关保护条例》、《著作权法》、《专利法》、《商标法》等。 至此，您确认已经详细阅读并已理解本协议全部条款，特别是对于限制或免除 元物质（深圳）智能科技有限公司及其代理、销售人员有关责任的免责条款已经详细阅读并充分理解其内容，并同意严格遵守各条款和条件。

、

**1 前言**

**1.1** 文档用途

本文档用于元物质人工智能平台OpenStation的介绍。

**1.2** 适用范围

本文档为元物质OpenStation产品线内部参考文档。

**1.3** 适用对象

本文档提供给以下相关人员使用：

 软件研发工程师

 系统运维工程师

其他获得授权的人员

**2 OpenStation功能概述**

**2.1 OpenStation产品介绍**

元物质OpenStation科研计算平台，是一个面向高校、科研机构的计算管理平台，他统一集成各个科研领域主流的计算框架和常用环境（TensorFlow、PyTorch、Caffe、MxNet、PaddlePaddle等），并将计算资源（CPU、GPU、内存、存储）统一管理起来的平台，简称OpenStation。

**2.2 OpenStation产品功能**

**多框架模型统一管理**

　　支持Tensorflow、Pytorch、Caffe、MXnet等主流框架模型，支持ONNX格式模型

　　支持自定义镜像，通过页面配置快速部署，以及云原生方式进行自定义编排

　　支持模型导入、查询、删除等操作

　　支持个人模型、组模型的权限管控

**应用服务全周期管理**

　　支持Tensor Serving、Torch Serving、Triton 推理引擎

　　支持离线测试，方便开发人员在模型发布前验证模型准确性

　　支持A/B测试与滚动发布进行模型迭代升级

　　支持服务弹性伸缩，根据流量自动调整服务实例数量

　　提供周期部署与定时部署的服务调度策略，解决不同时段的业务算力需求

　　支持流量调节、批量处理、业务下线等操作

支持边缘场景下的推理服务部署，状态查看及管理操作

**应用请求快速响应**

　　优化网络模型，减少网络代理延迟

　　深度调优网络参数，减少网络延迟

　　服务延迟1ms，业内领先水平

**资源性能监控**

　　简介直观的集群监控界面，核心参数一目了然

　　强大的扩展能力，支持用户自定义页面

　　丰富的日志检索能力，支持检索模版

　　日志分析能力出众，支持分析结果可视化

**3 OpenStation 技术原理**

　 支持平台服务的健康监控及容错，自动感知服务资源异常情况，并为线上服务选取可用计算资源进行主动迁移。

**轻松部署推理服务，提速AI生产交付**

　　提供自动快捷的模型部署方式及服务调用方式，一键实现滚动发布、A/B测试、服务回滚等高级服务部署功能，模型部署上线从2-3天缩短到几分钟；

**灵活开放，ISV应用统一适配**

　　支持TensorFlow、Pytorch等主流深度学习框架，同时兼容用户自定义框架，无缝对接企业自研或外购AI模型，满足企业多样化的应用场景；

**计算资源统一管理，敏捷高效调配**

　　支持英伟达、寒武纪、昆仑芯、高通等6家国际国内厂商的12款AI和GPU芯片的多元算力支持，针对不同AI应用的计算特点按需动态调配计算资源，支持GPU单卡算力细粒度切分，集群资源利用率从40% 提高到 80%；

**低延迟高吞吐，秒级响应线上业务**

　　高并发高吞吐，服务响应平均延迟低于1ms；服务资源弹性扩缩，应对突发场景可在几分钟内自动完成服务扩容，提升整体性能；

**智能容错，保障业务平稳运行**

在训练数据管理方面，统一管理个人数据及开发数据集，支持数据共享协同。对接企业数据存储系统，支持NFS、HDFS、BeeGFS等文件系统。支持文件上传、下载、压缩、解压、新建和删除等。支持用户访问及下载权限设置，保证数据安全可靠。

在AI计算资源集中管理调度方面，GPU服务器及GPU卡池化管理，使用状态一览展示。GPU资源使用情况实时统计分析。开发用户按需申请资源、动态分配。GPU多维细粒度调度。智能化资源调度策略。

在GPU性能监控方面，AI服务器性能及状态监控，CPU利用率、MEM利用率、网络IO、IB流量、磁盘分区、节点负载。GPU性能监控，GPU利用率、显存利用率、时钟频率、功耗、温度。存储系统及磁盘分区使用监控，各用户存储使用量监控。

在AI计算资源运维管理方面，实时统计用户组/用户资源使用。实时查看计算节点资源占用情况。管理员控制各用户资源使用配额。管理员快速清理回收资源，再分配。

在交互式模型开发训练过程方面，集成主流AI开发框架。打通数据集、计算资源、框架工具。提供丰富完整的AI开发工具链。对接pycharm、vscode等IDE工具。

在分布式训练任务编排方面，不同深度学习框架模板，通过模板快速提交训练任务。多机分布式任务跨节点自动分配计算资源，自动编排计算进程。按需申请计算资源的CPU核数和GPU卡数内存和训练集等。提供CLI方式任务提交及管。

**4 总结**

元物质秉承以客户为关注焦点，技术创新是原动力理念，关注客户需求，通过技术创新和产品创新，为客户提供完善的产品和方案，OpenStation即是元物质众多产品的核心。通过本文档我们了解到通用虚拟化技术的原理，OpenStation的功能和架构，OpenStation在资源虚拟化、人工智能计算和开放性三个方面的技术特性。OpenStation是元物质为客户倾力打造的战略级产品，为客户提供应用向虚拟化平台的部署或迁移，构建人工智能计算数据中心的关键能力。

关于更多 OpenStation的资料和内容，请联系相关行业和区域元物质技术人员。