



三 生命科学

fastone

AutoDock Vina

生信

云

实

证

Vol.3

速石科技

提速2920倍！

用AutoDock Vina对接2800万个分子

COMPUTER AIDED
DRUG DESIGN





前言



AutoDock Vina是用于分子对接和虚拟筛选的开源程序,由Scripps研究所分子图形实验室的Oleg Trott博士设计和实现,是目前使用最为广泛的分子对接软件之一。

分子对接技术,作为**计算机辅助药物设计 (Computer Aided Drug Design, CADD)**的重要方法,已广泛应用于药物发现阶段的早期虚拟筛选、药物分子设计、先导化合物优化、药物潜在作用靶点发现、药物-靶点相互作用机制、为重要的药物代谢酶寻找特异性配体等。

目前,限于算力,或者高效灵活地调用大规模计算集群的能力,当前的虚拟筛选通常仅采样百万到千万个分子,而事实上目前可用于药物发现的有机分子已经超过10的60次方。



在《15小时虚拟筛选10亿分子, Nature+HMS验证云端新药研发未来》这篇文章里,哈佛大学医学院的研究人员论证了:分子化合物的质量会随着虚拟筛选规模的扩大而提升。



如何在本地资源有限的情况下,提高虚拟筛选规模和质量,把漫长的药物研发周期缩短一点?

我们用实证给你一个答案。

与前两次实证不同,本次生信实证有两大特点:

- ◆ **任务数量多,云上同一地区某种类型机型可能不足,因此会涉及到多区域资源调度;**
- ◆ **可根据用户偏好匹配合适的资源调度策略,满足用户不同需求。**



实证背景信息

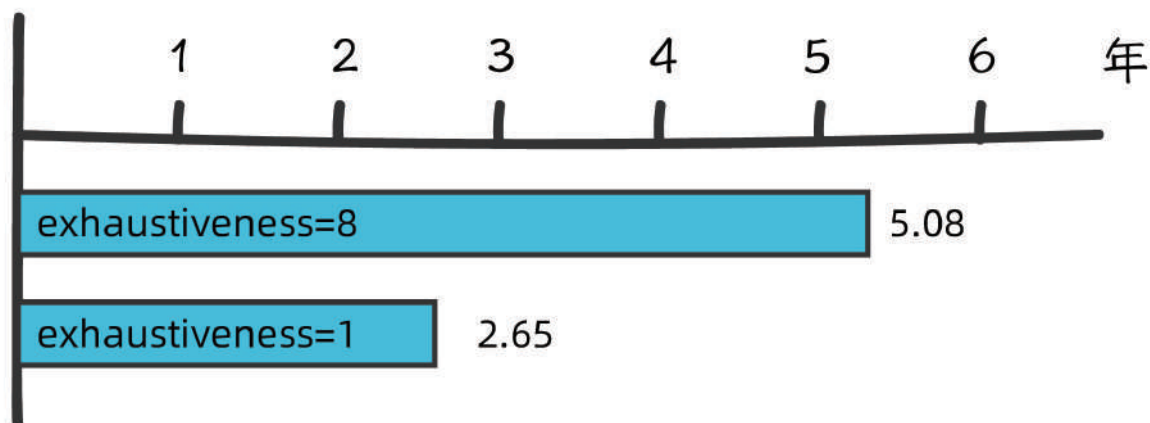


某大型药企在本地建设有机房，计算资源总计为104核。使用AutoDock Vina进行小分子对接：当设定exhaustiveness=8时，筛选56643个原始分子共需90小时；当设定exhaustiveness=1时，耗时需18小时。

参数	原始分子数	运行任务数	本地（104核） 运算时间	平均单个CPU 运算单个分子耗时
exhaustiveness=8	56643	208165	90h	2.6min
exhaustiveness=1	56643	78521	18h	1.4min

(exhaustiveness参数：用来控制对接的细致程度，影响计算时间)

当筛选范围扩大到整个VS数据库(2800万个分子)时，不同参数条件下本地资源所需的运算时间在约2.6-5年不等。



研发负责人认为这么长的时间周期是无法接受的，其本地现有IT架构和资源完全无法满足研发需求。



实证目标



- ◆ AutoDock Vina任务能否在云端有效运行?
- ◆ fastone平台能否大幅度缩短任务运行时间?
- ◆ fastone平台能否有效控制任务运行成本?
- ◆ 针对AutoDockVina任务小, 数量大的特点, fastone平台是否有针对性策略?

实证参数



平台:

fastone企业版产品



应用:

AutoDock Vina



适用场景:

分子对接, 研究配体 (药物分子) 与其受体 (已知的靶蛋白或活性位点) 之间的详细相互作用, 预测其结合模式及亲合力, 还可以用来发现并优化药物先导物分子, 进而实现基于结构的药物设计

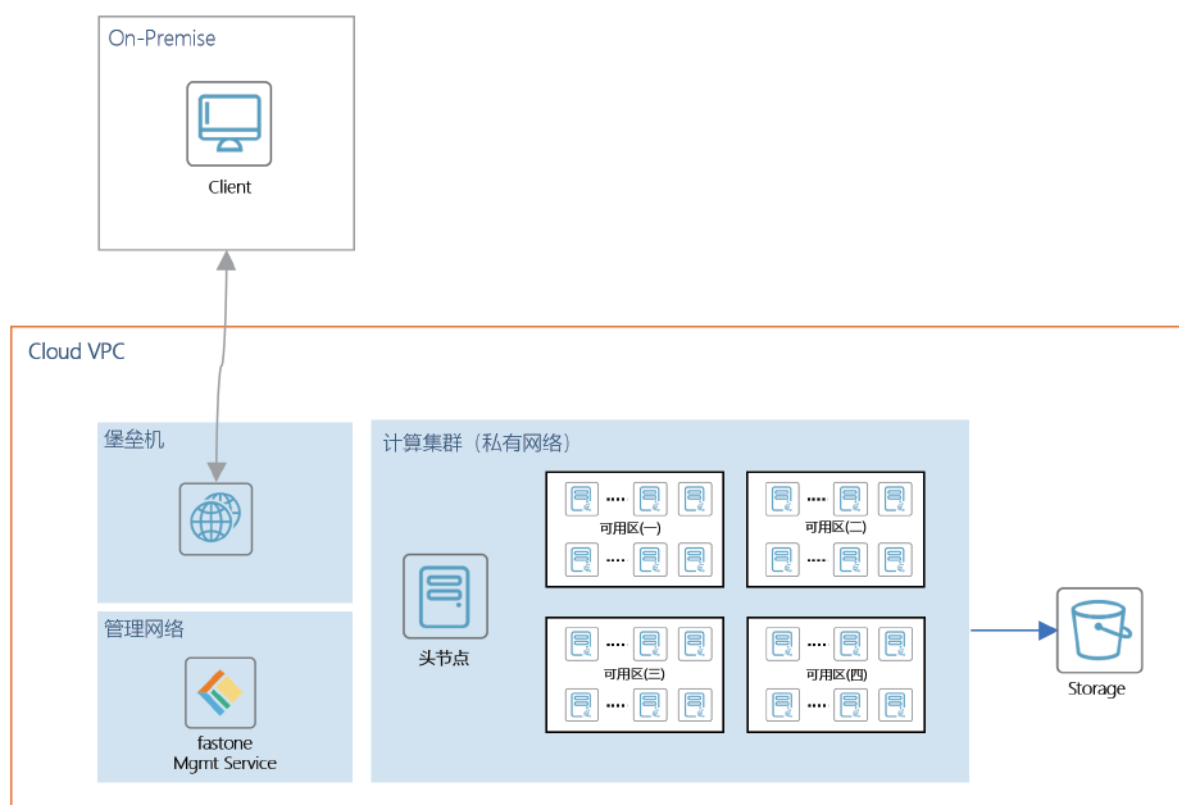


云端硬件配置：

AutoDock Vina在运行时需要对接海量分子,对计算性能要求较高,因此平台为用户推荐选择了匹配其应用特点的计算优化型实例机型。



技术架构图：





以下是两个实证场景。

实证场景一：我们通过10000分子分别进行了AutoDock Vina的云端线性扩展性验证及成本验证；

实证场景二：基于不同用户策略，我们帮用户进行了2800万量级的大规模分子对接。

1、时间优先策略以速度为第一优先级：资源选择以OD按需实例为主，在满足用户时间要求的前提下尽可能通过抢占SPOT实例来优化成本。

2、成本优先策略以成本为第一优先级：资源选择以SPOT实例为主，并在满足用户成本要求的前提下使用OD按需实例来优化时间效率。

SPOT：可被抢占实例，又称竞价实例。价格最低可达到按需实例价格的10%，相当于秒杀，手快有手慢无，价格可高可低波动大，随时可能被抢占中断，需要有一定的技术实力才能使用。

OD：On-Demand，按需实例。针对短期弹性需求，按小时计费，灵活精准，避免浪费，但价格比较高，通常为SPOT实例的3-10倍。



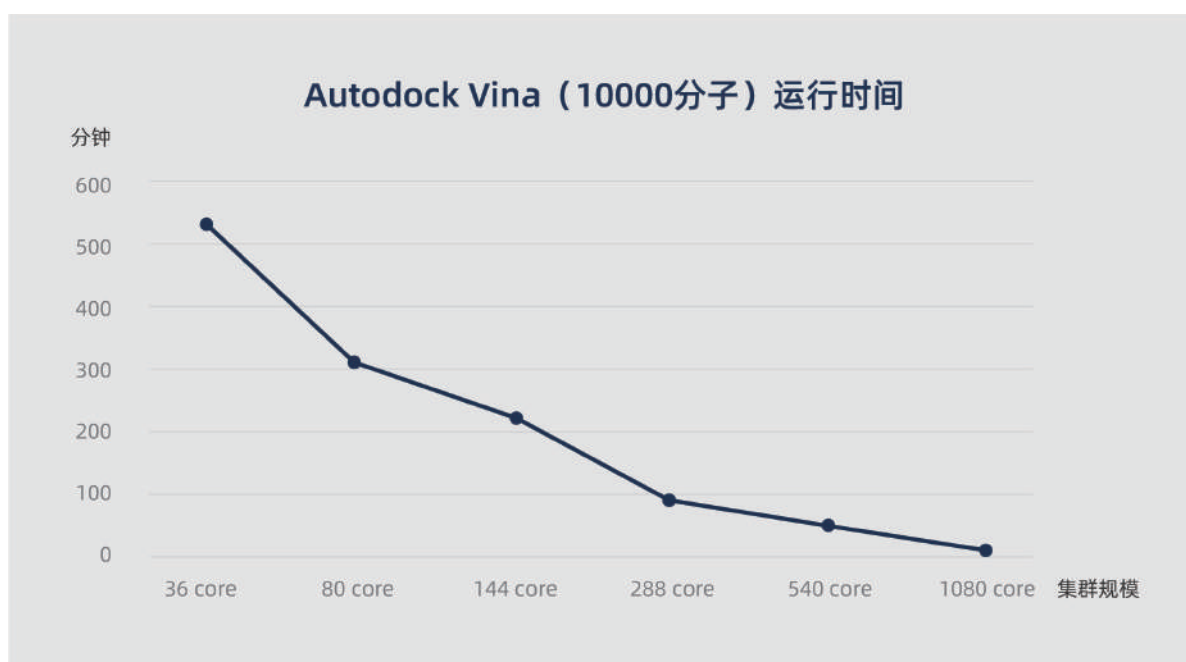
实证场景一：10000分子



AutoDock Vina云端线性扩展性及成本验证

结论一

- 在云端调度不同核数的计算资源对接10000分子, 验证AutoDock Vina在云上具有线性扩展性, 即当处理器数量增加一倍, 运算时间也会缩短一半。



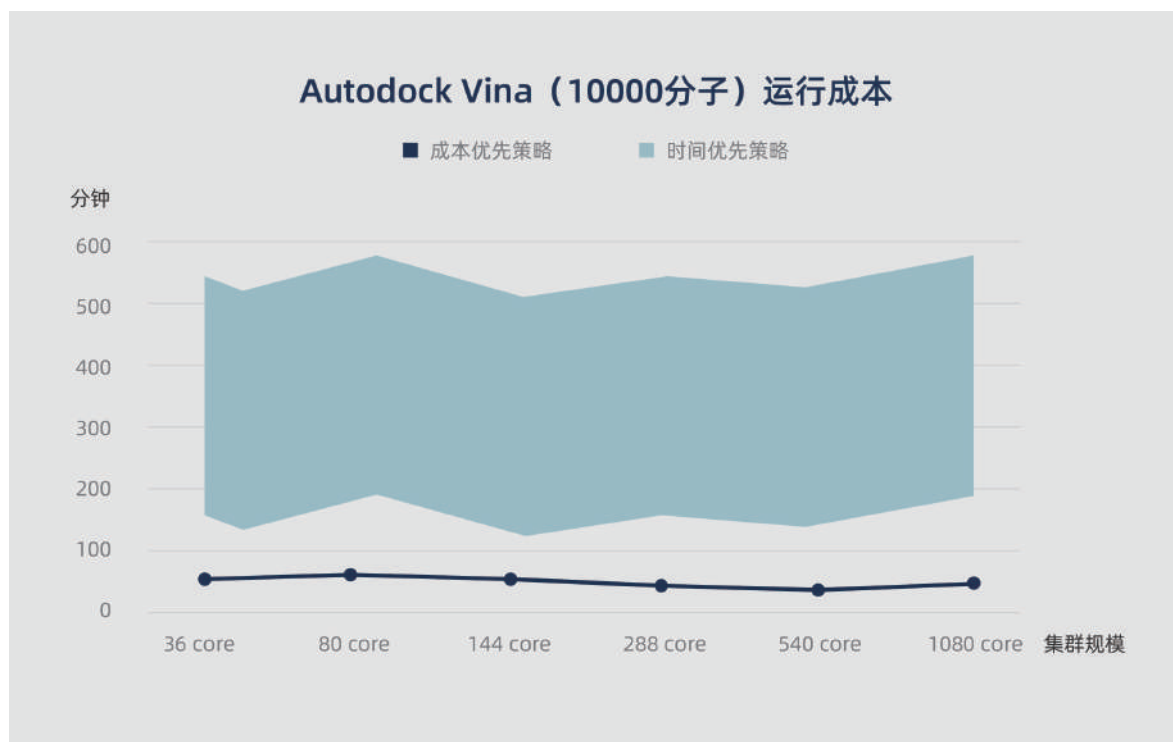
实证过程：

01. 云端调度36核计算资源对接10000分子, 采用时间优先策略需耗时527分钟;
02. 云端调度80核计算资源对接10000分子, 采用时间优先策略需耗时314分钟;
03. 云端调度144核计算资源对接10000分子, 采用时间优先策略需耗时215分钟;
04. 云端调度288核计算资源对接10000分子, 采用时间优先策略需耗时98分钟;
05. 云端调度540核计算资源对接10000分子, 采用时间优先策略需耗时52分钟;
06. 云端调度1080核计算资源对接10000分子, 采用时间优先策略需耗时20分钟。



结论二

- ◆ 在云端调度不同核数的计算资源对接10000分子, 当用户选择成本优先策略时, fastone平台以SPOT实例为主要资源选择, 确保成本为第一优先级。
比时间优先策略, 成本降幅最多可达67%-90%。



实证过程:

以下均采用成本优先策略

01. 云端调度36核计算资源对接10000分子, 抢占SPOT实例, 耗费82元;
02. 云端调度80核计算资源对接10000分子, 抢占SPOT实例, 耗费84元;
03. 云端调度144核计算资源对接10000分子, 抢占SPOT实例, 耗费79元;
04. 云端调度288核计算资源对接10000分子, 抢占SPOT实例, 耗费64元;
05. 云端调度540核计算资源对接10000分子, 抢占SPOT实例, 耗费58元;
06. 云端调度1080核计算资源对接10000分子, 抢占SPOT实例, 耗费68元。



实证场景二：2800万分子



大规模业务验证：基于不同用户策略

fastone基于用户2800万分子对接需求,提供时间优先和成本优先两种策略供用户选择。

一、用户以时间为第一优先级

结论：

- 1、通过fastone平台采用时间优先策略调用10万核计算优化型实例对接2800万个分子,耗时约15.23小时,运算效率提高2920倍;
- 2、fastone平台根据用户计算需求,自动化构建并调度云上10万核大规模算力集群,完成计算任务;
- 3、时间优先策略下,当任务数量达到一定规模时,云上同一地区某种类型机型可能不足,fastone平台可跨区、跨类型自动为用户调度云资源,以最快速度完成计算任务;
- 4、fastone平台自动帮用户确定中断可能性最低的SPOT池,保障任务顺利高效完成,本次实证任务的中断率为0.95%(通常<5%)。

云端部署手动模式 VS 自动模式之间的巨大差异可查看:

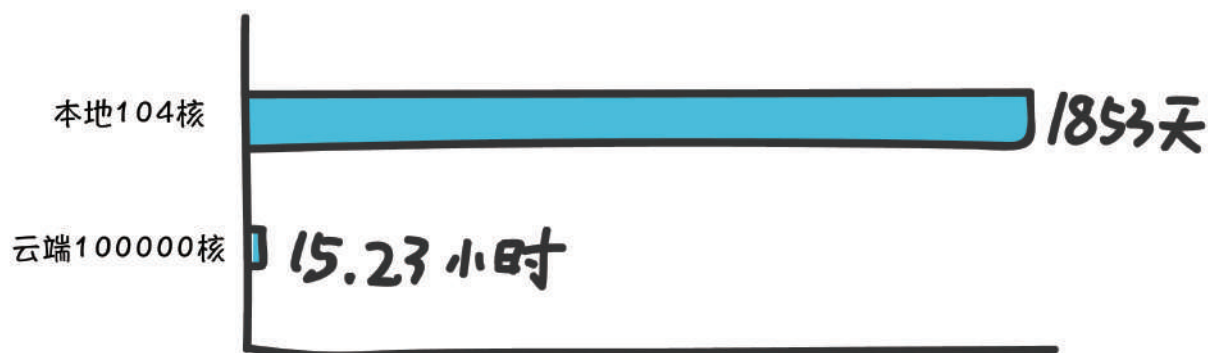
《EDA云实证Vol.1:从30天到17小时,如何让HSPICE仿真效率提升42倍?》



实证过程：

1、设定exhaustiveness=8, 本地104核计算资源对接约2800万个分子, 经估算需耗时约1853天;

2、设定exhaustiveness=8, 云端调度10万核计算资源对接约2800万个分子, 采用时间优先策略需耗时约15.23小时(含配置, 安装, 调度等时间)。



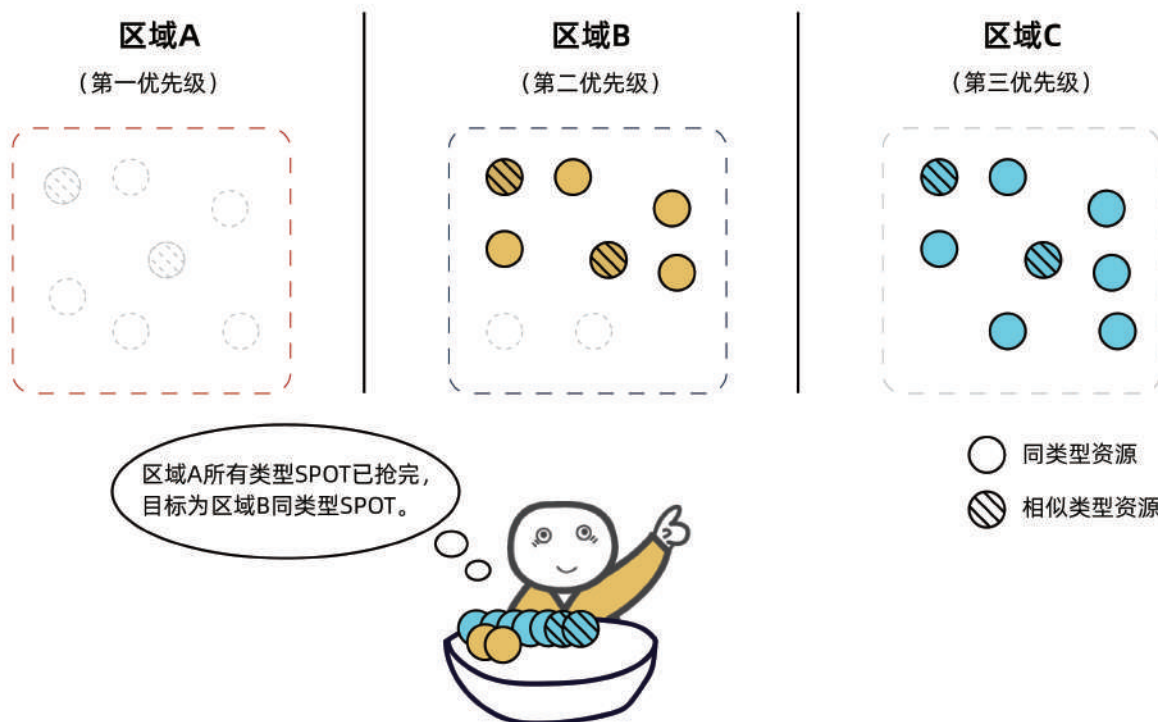
计算资源越多, 运算时间越短。

在满足用户时间要求的前提下, 可通过尽可能抢占SPOT实例来帮助用户优化成本。

当所需的计算资源达到十万核这个数量级以后, 单个区域内我们的目标类型资源可能会瞬间告罄, 造成任务排队, 从而大大拖慢运算时间。



我们需要通过fastone平台的Auto-Scale功能自动调度本区域及其他区域的目标类型或相似类型SPOT实例资源,以最快速地完成任



简单说,就是**优先抢低价的SPOT实例**,抢完同类型的再抢其他类型的,抢完同区域的再抢其他区域的。

这只是Auto-Scale功能的一部分。

fastone的Auto-Scale功能可以自动监控用户提交的任务数量和资源的需求,动态按需地开启所需算力资源,在提升效率的同时有效降低成本。可以让用户根据自身需求,设置调度集群规模上下限,且所有操作都是自动化完成,无需用户干预。



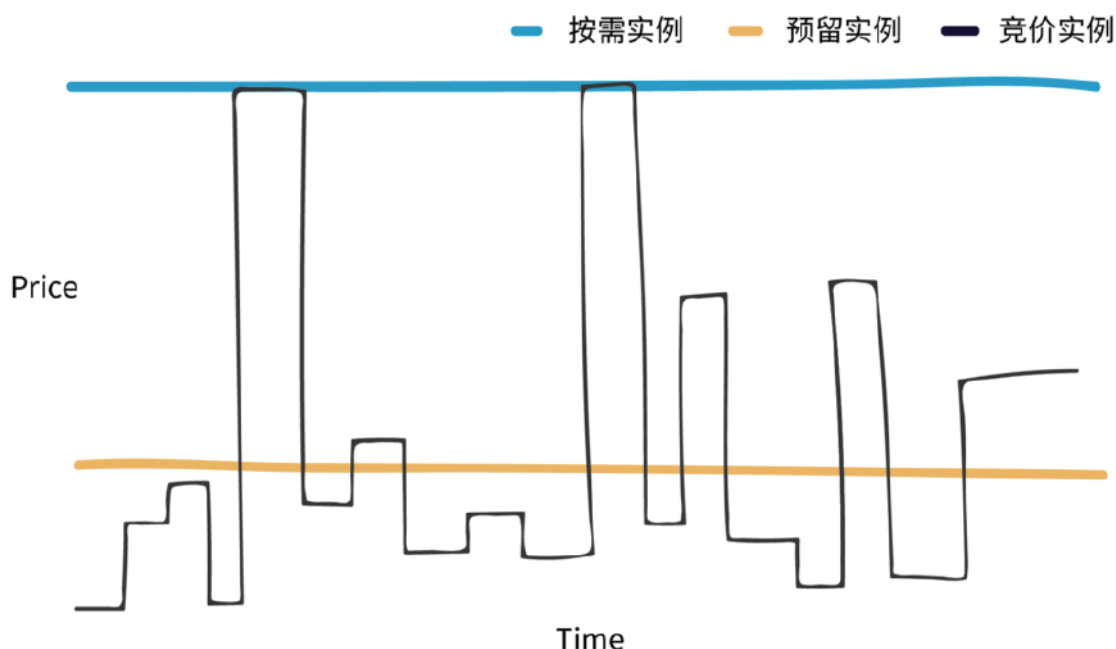
二、用户以成本为第一优先级

使用AutoDock Vina进行分子对接的一大特征是任务数量庞大而单个任务计算时间短, 单个分子对接的时间通常在几分钟以内(与参数设置有关)。

这一特征天然匹配云端的SPOT实例。

云端SPOT实例有四大特点：

- 1、便宜是真便宜。
- 2、不是人人都能用好。
- 3、不是你想要啥就有, 不是你想用的时候就能用。
- 4、或迟或早, 最终一定会被抢走。



OD按需实例价格通常为SPOT实例的3-10倍。

可参考:《云资源中的低成本战斗机——竞价实例, AWS、阿里云等六家云厂商完全用户使用指南》

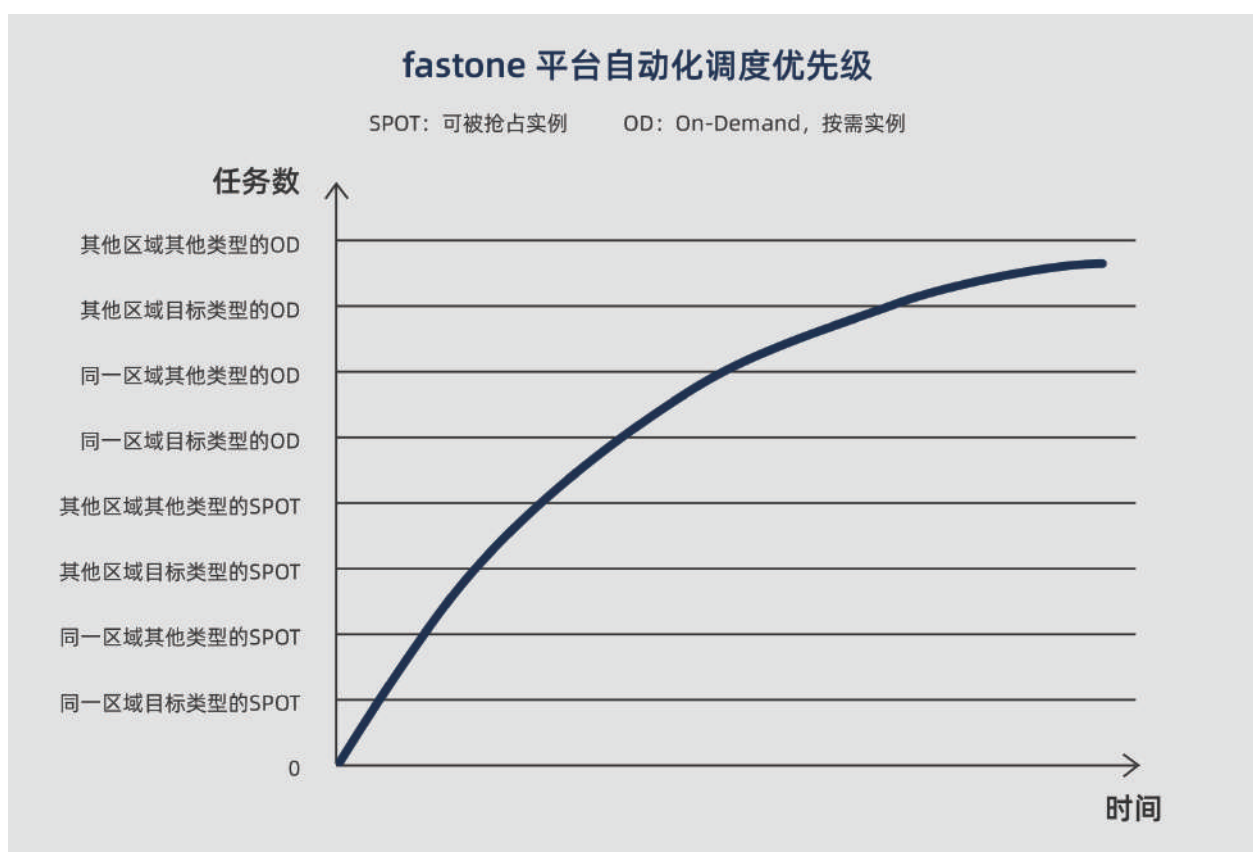


当便宜且随时可能被抢占中断的SPOT实例遇到迷你却海量的分子对接任务，简直就是天造地设的一对。

1、常规分子对接任务几分钟即可算完，特别适合SPOT这种分分钟可能被抢走的状态；

2、fastone平台具备自动重试功能，一个任务被中断可以自动重新提交，任务之间互相不影响，重新提交单个任务影响很小。

fastone平台会按以下顺序依次进行自动化调度：





实证小结



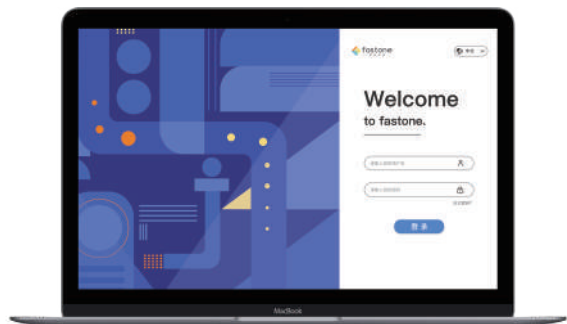
- ◆ AutoDock Vina任务能在云端有效运行；
- ◆ fastone平台能够大幅度缩短任务运行时间；
- ◆ fastone平台能够有效控制任务运行成本；
- ◆ fastone平台的Auto-Scale功能可完美匹配AutoDock Vina任务小、数量大的特点；
- ◆ fastone平台能根据用户不同需求, 为用户提供不同的自动化调度策略。

本次生信行业Cloud HPC实证系列Vol.3就到这里。

在下一期的实证中, 我们将为大家带来Amber上云实证, 这次涉及到了云端GPU资源的使用。

未来我们还会带给大家更多领域的用云“真香”实证, 请保持关注哦!

速石科技 为应用定义的云平台



免费试用

用这个能干嘛？
跑任务。



即刻获得TOP500超级算力

2 分钟
自动开通

在线即时开通
不用人工介入, 无需等待

300 元
体验金

主动奉上体验金
兵马未动, 粮草先行

样本数据

提前安排好了
样本数据

Demo

部分应用提供了
相应的操作视频



立即免费试用

- 01 平台集成多行业应用, 大量任务多节点并行
- 02 应对短时间爆发性需求, 连网即用
- 03 跑任务快, 原来几个月甚至几年, 现在只需几小时
- 04 5分钟快速上手, 拖拉点选可视化界面, 无需代码
- 05 支持高级用户直接在云端创建集群

速石资料库

《2020新版-六大云厂商资源价格对比报告》

《半导体行业云解决方案白皮书》

《超算 VS 云计算全方位终极PK》

《EDA云实证Vol.1-从30天到17小时，HSPICE仿真效率如何提升42倍?》

《CAE云实证Vol.2-从4天到1.75小时，如何让Bladed仿真效率提升55倍?》

《【2021版】全球44家顶尖药企AI辅助药物研发行为白皮书》

.....



添加小F微信
并备注:资料获取



.....

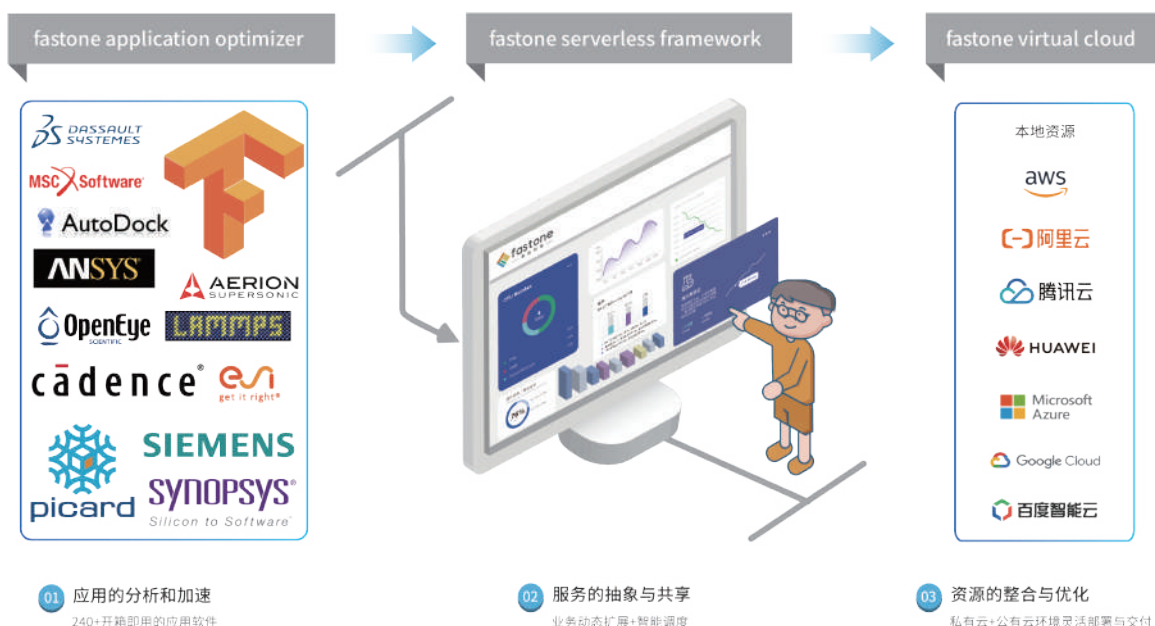
联系我们

速石科技 (fastone) 致力于构建为应用定义的云, 让任何应用程序, 始终以自动化、更优化和可扩展的方式, 在任何基础架构上运行。

我们为有高算力需求的用户提供一站式多云算力运营解决方案。基于本地+公有混合云环境的灵活部署及交付, 帮助用户提升10-20倍业务运算效率, 降低成本达到75%以上, 加快市场响应速度。

我们提供为应用优化的一站式交付平台, 对药物研发/基因分析/CAE/EDA/AI等行业应用进行分析与加速, 通过Serverless框架屏蔽底层IT技术细节, 实现用户对本地和公有云资源无差别访问。

我们的产品包括: fastone COMPUTE PLATFORM、fastone COMPUTE CLOUD。



电话: +86-21-31263638

邮箱: marketing@fastonetech.com

网站: www.fastonetech.com

地址: 上海市杨浦区国通路118号天盛广场A栋
1801室

