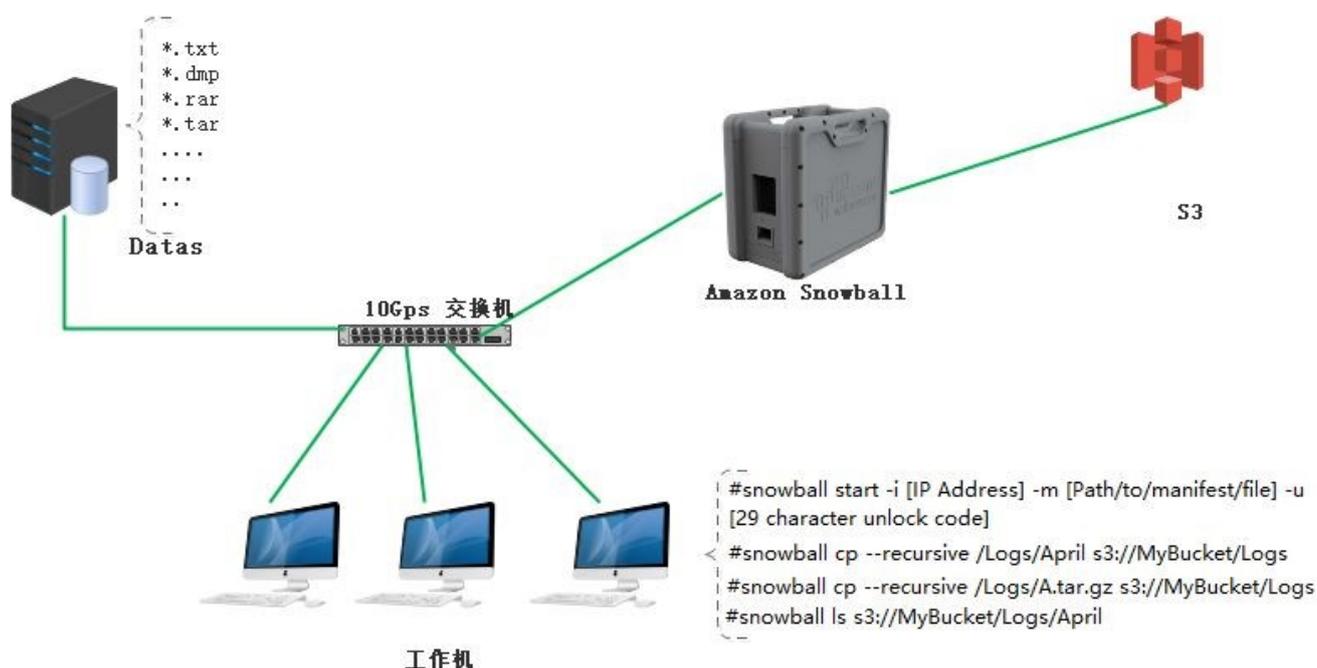


# 云享家 | 宝典在手，轻松玩转AWS Snowball.

原创：胡守跃 BespinGlobal 4天前



AWS Snowball 是一种数据传输解决方案，它能够使用以物理方式安全传输的存储设备，加快 TB 级到 PB 级数据传入或传出 AWS 服务的速度，帮助客户在中国更快、更经济高效地将数据迁移到云中。使用 Snowball 有助于解决进行大规模数据传输时会遇到的难题，包括网络成本高、传输时间长和安全问题。除大规模数据传输外，AWS Snowball 还支持从 Hadoop 分布式文件系统 (HDFS) 导入数据。



## 场景适用

### 云迁移

适合大量数据导入 Amazon S3，及从Amazon S3导出。

### 灾难恢复

借助 AWS Snowball，我们将执行灾难恢复工作所需的时间从几周缩短到了几天，而且实现了 100% 完全恢复。

### 数据中心停用

适合数据中心停用，将数TB或PB的数据更快速的传输至Snowball，确保有价值的不会丢失，完成数据迁移。

## Snowball的优势

### 传输速度快

借助Snowball设备，可以在不到一周的时间内完成50TB数据传输。

### 高扩展性

单个Snowball 设备可以传输数TB 的数据，可以并行使用多台设备将数据传入或传出 Amazon S3 存储桶。

### 简单且可兼容

可以通过行业标准网络接口（RJ45 、 铜缆SFP+ 、 光纤SFP+ 、 10Gb以太网）和存储协议（HDFS、S3 API和S3 CLI）连接到AWS Snowball设备。

## 工作站规格

Item	建议规格
处理能力	16 核 CPU
内存	16 GB RAM
	重要
	对于内存密集型任务（如执行 snowball cp 命令），客户端和/或适配器的每个正在运行的实例最高需要 7GB 的专用 RAM。
Microsoft Windows 支持（仅 64 位）	Windows 7 Windows 8 Windows 10
Mac 支持	Mac OS X 版本 10.10 或更高版本
Linux 支持（仅 64 位）	Ubuntu 版本 12 或更高版本 Red Hat Enterprise Linux (RHEL) 版本 6 或更高版本
用户界面支持	键盘 鼠标

	显示器
<b>网络 I/O 支持</b>	RJ45
	SFP+ 铜缆
	SFP+ 光纤

## AWS Snowball的性能

因为每个本地环境的运营方式各不相同 – 不同的网络技术、不同的硬件、不同的操作系统、不同的过程等。

下表概述了您的网络传输速率如何影响向 Snowball 填充数据所需的时间。如果小文件未事先批处理成大文件即进行传输，则会因增加开销而减慢您的传输速度。

速率 (MB/s)	42-TB 传输时间	72-TB 传输时间
800	14 小时	1 天
450	1.09 天	1.8 天
400	1.16 天	2.03 天
300	1.54 天	2.7 天
277	1.67 天	2.92 天
200	2.31 天	4 天
100	4.63 天	8.10 天
60	8 天	13 天
30	15 天	27 天
10	46 天	81 天

## 加快数据传输速度

一般来说，您可通过以下方法加快数据从数据源到 Snowball 的传输速度，这些方法是根据对性能产生的积极影响由大到小进行排序的：

- 1 **使用最新的 Mac 或 Linux Snowball 客户端** - 最新的 Snowball 客户端 for Mac 和 Snowball 客户端 for Linux 都支持 x86 指令集架构的高级加密标准新指令 (AES-NI) 扩展。此扩展将加快在 Snowball 与您的 Mac 或 Linux 工作站之间传输时加密或解密数据的速度。

2

- 批量传输小型文件** - 因为要进行加密，每个复制操作都会产生一些开销。因此，相比传输大型文件中的相同数据，对独立文件执行多次传输可能会降低整体性能。您可以大大加快小文件的传输速度，方式是通过一个 `snowball cp` 命令批处理小文件。默认情况下会启用小文件的批处理。在导入 Amazon S3 的过程中，这些批处理后的文件将自动提取到其原始状态。
- 3 一次执行多个复制操作** - 如果工作站的功能足够强大，则可一次执行多个 `snowball cp` 命令。您还可通过在单独的 Snowball 客户端实例中从一个单独的终端窗口运行每条命令完成此操作，所有这些实例都连接到同一个 Snowball。
- 4 从多个工作站进行复制** - 您可将单个 Snowball 连接到多个工作站。各工作站都可托管一个单独的 Snowball 客户端实例。
- 5 传输目录而不是文件** - 由于每个 `snowball cp` 命令都存在开销，因此建议不要排列大量单个复制命令。排列许多命令会对传输性能产生严重的负面影响。例如，假设您拥有称为 `C:\\MyFiles` 的目录，该目录只包含三个文件：`file1.txt`、`file2.txt` 和 `file3.txt`。假设您发起以下三个命令。

```
snowball cp C:\\MyFiles\\file1.txt s3://mybucket  
snowball cp C:\\MyFiles\\file2.txt s3://mybucket  
snowball cp C:\\MyFiles\\file3.txt s3://mybucket
```

在这种情况下，您的开销是使用以下复制命令传输整个目录的三倍。

```
Snowball cp -r C:\\MyFiles\\* s3://mybucket
```
- 6 请勿在传输过程中对文件执行其他操作** - 在传输过程中重命名文件，更改其元数据，或在执行复制操作期间在文件中写入数据，都会对传输性能产生严重的负面影响。我们建议您在传输文件时将文件保持静态
- 7 减少本地网络使用** - 您的 Snowball 将跨本地网络进行通信。因此，降低 Snowball、它连接到的交换机和承载您的数据源的工作站之间的其他本地网络流量可以提高数据传输速度
- 8 消除不必要的跃点** - 建议您设置 Snowball、数据源和工作站，使其成为通过单个交换机通信的仅有计算机。这样做可大大提高数据传输速度。

更多关于：

- 通过实验获得更好的性能
- 使用AWS Snowball将数据导入到AWS S3
- 将Snowball设备连接到本地网络
- 使用Snowball客户端传输数据
- Snowball cp命令的语法

请点击“阅读原文”下载《使用手册》

①

点击下方  
阅读原文



阅读原文