1.3 Bắt Đầu - Cơ Bản về Git

**Cơ Bản về Git**

Tóm lại thì, Git là gì? Đây là một phần quan trọng để tiếp thu, bởi vì nếu bạn hiểu được Git là gì và các nguyên tắc cơ bản của việc Git hoạt động như thế nào, thì sử dụng Git một cách hiệu quả sẽ trở nên dễ dàng hơn cho bạn rất nhiều. Khi học Git, hãy cố gắng gạt bỏ những kiến thức mà có thể bạn đã biết về các VCS khác, ví dụ như Subversion và Perforce; việc này sẽ giúp bạn tránh được sự hỗn độn, bối rối khi sử dụng nó. Git "nghĩ" về thông tin và lưu trữ nó khá khác biệt so với các hệ thống khác, mặc dù giao diện người dùng tương đối giống nhau; hiểu được những khác biệt đó sẽ giúp bạn tránh được rất nhiều bối rối.

[**Ảnh Chụp, Không Phải Sự Khác Biệt**](http://git-scm.com/book/vi/v1/B%E1%BA%AFt-%C4%90%E1%BA%A7u-C%C6%A1-B%E1%BA%A3n-v%E1%BB%81-Git#Ảnh-Chụp,-Không-Phải-Sự-Khác-Biệt)

Sự khác nhau cơ bản giữa Git với bất kỳ VCS nào khác (bao gồm Subversion và tương tự là cách Git "nghĩ" về dữ liệu. Về mặt lý thuyết mà nói, phần lớn hệ thống khác lưu trữ thông tin dưới dạng danh sách các tập tin được thay đổi. Các hệ thống này (CVS, Subversion, Perforce, Bazaar,...) coi thông tin được lưu trữ như là một tập hợp các tập tin và các thay đổi được thực hiện trên mỗi tập tin theo thời gian, được minh hoạ trong hình 1-4.



Hình 1-4. Các hệ thống khác hướng tới lưu trữ tập tin dưới dạng các thay đổi so với bản cơ sở của mỗi tập tin.

Git không nghĩ hoặc xử lý dữ liệu theo cách này. Mà thay vào đó Git coi dữ liệu của nó giống như một tập hợp các "ảnh" (snapshot) của một hệ thống tập tin nhỏ. Mỗi lần bạn "commit", hoặc lưu lại trạng thái hiện tại của dự án trong Git, về cơ bản Git "chụp một bức ảnh" ghi lại nội dung của tất cả các tập tin tại thời điểm đó và tạo ra một tham chiếu tới "ảnh" đó. Để hiệu quả hơn, nếu như tập tin không có sự thay đổi nào, Git không lưu trữ tập tin đó lại một lần nữa mà chỉ tạo một liên kết tới tập tin gốc đã tồn tại trước đó. Git thao tác với dữ liệu giống như Hình 1-5.



Hình 1-5. Git lưu trữ dữ liệu dưới dạng ảnh chụp của dự án theo thời gian.

Đây là sự khác biệt lớn nhất giữa Git và hầu hết các VCS khác. Nó khiến Git cân nhắc lại hầu hết các khía cạnh của quản lý phiên bản mà phần lớn các hệ thống khác chỉ áp dụng lại từ các thế hệ trước. Chính lý do này làm cho Git giống như một hệ thống quản lý tập tin thu nhỏ với các tính năng, công cụ vô cùng mạnh mẽ được xây dựng dựa trên nó, không chỉ là một hệ thống quản lý phiên bản đơn giản. Chúng ta sẽ khám phá một số lợi ích đạt được từ việc quản lý dữ liệu theo cách này khi bàn luận về Phân nhánh trong Git ở Chương 3.

[**Phần Lớn Thao Tác Diễn Ra Cục Bộ**](http://git-scm.com/book/vi/v1/B%E1%BA%AFt-%C4%90%E1%BA%A7u-C%C6%A1-B%E1%BA%A3n-v%E1%BB%81-Git#Phần-Lớn-Thao-Tác-Diễn-Ra-Cục-Bộ)

Phần lớn các thao tác/hoạt động trong Git chỉ cần yêu cầu các tập tin hay tài nguyên cục bộ - thông thường nó sẽ không cần bất cứ thông tin từ máy tính nào khác trong mạng lưới của bạn. Nếu như bạn quen với việc sử dụng các hệ thống quản lý phiên bản tập trung nơi mà đa số hoạt động đều chịu sự ảnh hưởng bởi độ trễ của mạng, thì với Git đó lại là một thế mạnh. Bởi vì toàn bộ dự án hoàn toàn nằm trên ổ cứng của bạn, các thao tác được thực hiện gần như ngay lập tức.

Ví dụ, khi bạn muốn xem lịch sử của dự án, Git không cần phải lấy thông tin đó từ một máy chủ khác để hiển thị, mà đơn giản nó được đọc trực tiếp từ chính cơ sở dữ liệu cục bộ của bạn. Điều này có nghĩa là bạn có thể xem được lịch sử thay đổi của dự án gần như ngay lập tức. Nếu như bạn muốn so sánh sự thay đổi giữa phiên bản hiện tại của một tập tin với phiên bản của một tháng trước, Git có thể tìm kiếm tập tin cũ đó trên máy cục bộ rồi sau đó so sánh sự khác biệt cho bạn. Thay vì việc phải truy vấn từ xa hoặc "kéo về" (pull) phiên bản cũ của tập tin đó từ máy chủ trung tâm rồi mới thực hiện so sánh cục bộ.

Điều này còn đồng nghĩa với có rất ít việc mà bạn không thể làm được khi không có kết nối Internet hoặc VPN bị ngắt. Nếu bạn muốn làm việc ngay cả khi ở trên máy bay hoặc trên tầu, bạn vẫn có thể commit bình thường cho tới khi có kết nối Internet để đồng bộ hoá. Nếu bạn đang ở nhà mà VPN lại không thể kết nối được, bạn cũng vẫn có thể làm việc bình thường. Trong rất nhiều hệ thống khác, việc này gần như là không thể hoặc rất khó khăn. Ví dụ trong Perforce, bạn gần như không thể làm gì nếu như không kết nối được tới máy chủ; trong Subversion và CVS, bạn có thể sửa tập tin nhưng bạn không thể commit các thay đổi đó vào cơ sở dữ liệu (vì cơ sở dữ liệu của bạn không được kết nối). Đây có thể không phải là điều gì đó lớn lao, nhưng bạn sẽ ngạc nhiên về sự thay đổi lớn mà nó có thể làm được.

[**Git Mang Tính Toàn Vẹn**](http://git-scm.com/book/vi/v1/B%E1%BA%AFt-%C4%90%E1%BA%A7u-C%C6%A1-B%E1%BA%A3n-v%E1%BB%81-Git#Git-Mang-Tính-Toàn-Vẹn)

Mọi thứ trong Git được "băm" (checksum or hash) trước khi lưu trữ và được tham chiếu tới bằng mã băm đó. Có nghĩa là việc thay đổi nội dung của một tập tin hay một thư mục mà Git không biết tới là điều không thể. Chức năng này được xây dựng trong Git ở tầng thấp nhất và về mặt triết học được coi là toàn vẹn. Bạn không thể mất thông tin/dữ liệu trong khi truyền tải hoặc nhận về một tập tin bị hỏng mà Git không phát hiện ra.

Cơ chế mà Git sử dụng cho việc băm này được gọi là mã băm SHA-1. Đây là một chuỗi được tạo thành bởi 40 ký tự của hệ cơ số 16 (0-9 và a-f) và được tính toán dựa trên nội dung của tập tin hoặc cấu trúc thư mục trong Git. Một mã băm SHA-1 có định dạng như sau:

24b9da6552252987aa493b52f8696cd6d3b00373

Bạn sẽ thấy các mã băm được sử dụng ở mọi nơi trong Git. Thực tế, Git không sử dụng tên của các tập để lưu trữ mà bằng các mã băm từ nội dung của tập tin vào một cơ sở dữ liệu có thể truy vấn được.

[**Git Chỉ Thêm Mới Dữ Liệu**](http://git-scm.com/book/vi/v1/B%E1%BA%AFt-%C4%90%E1%BA%A7u-C%C6%A1-B%E1%BA%A3n-v%E1%BB%81-Git#Git-Chỉ-Thêm-Mới-Dữ-Liệu)

Khi bạn thực hiện các hành động trong Git, phần lớn tất cả hành động đó đều được thêm vào cơ sở dữ liệu của Git. Rất khó để yêu cầu hệ thống thực hiện một hành động nào đó mà không thể khôi phục lại được hoặc xoá dữ liệu đi dưới mọi hình thức. Giống như trong các VCS khác, bạn có thể mất hoặc làm rối tung dữ liệu mà bạn chưa commit; nhưng khi bạn đã commit thì rất khó để mất các dữ liệu đó, đặc biệt là nếu bạn thường xuyên đẩy (push) cơ sở dữ liệu sang một kho chứa khác.

Điều này khiến việc sử dụng Git trở nên thích thú bởi vì chúng ta biết rằng chúng ta có thể thử nghiệm mà không lo sợ sẽ phá hỏng mọi thứ. Để có thể hiểu sâu hơn việc Git lưu trữ dữ liệu như thế nào hay làm sao để khôi phục lại dữ liệu có thể đã mất, xem Chương 9.

[**Ba Trạng Thái**](http://git-scm.com/book/vi/v1/B%E1%BA%AFt-%C4%90%E1%BA%A7u-C%C6%A1-B%E1%BA%A3n-v%E1%BB%81-Git#Ba-Trạng-Thái)

Bây giờ, hãy chú ý. Đây là điều quan trọng cần ghi nhớ về Git nếu như bạn muốn hiểu được những phần tiếp theo một cách trôi chảy. Mỗi tập tin trong Git được quản lý dựa trên ba trạng thái: committed, modified, và staged. Committed có nghĩa là dữ liệu đã được lưu trữ một cách an toàn trong cơ sở dữ liệu. Modified có nghĩa là bạn đã thay đổi tập tin nhưng chưa commit vào cơ sở dữ liệu. Và staged là bạn đã đánh dấu sẽ commit phiên bản hiện tại của một tập tin đã chỉnh sửa trong lần commit sắp tới.

Điều này tạo ra ba phần riêng biệt của một dự án sử dụng Git: thư mục Git, thư mục làm việc, và khu vực tổ chức (staging area).



Hình 1-6. Thư mục làm việc, khu vực khán đài, và thư mục Git.

Thư mục Git là nơi Git lưu trữ các "siêu dữ kiện" (metadata) và cơ sở dữ liệu cho dự án của bạn. Đây là phần quan trọng nhất của Git, nó là phần được sao lưu về khi bạn tạo một bản sao (clone) của một kho chứa từ một máy tính khác.

Thư mục làm việc là bản sao một phiên bản của dự án. Những tập tin này được kéo về (pulled) từ cơ sở dữ liệu được nén lại trong thư mục Git và lưu trên ổ cứng cho bạn sử dụng hoặc chỉnh sửa.

Khu vực khán đài là một tập tin đơn giản được chứa trong thư mục Git, nó chứa thông tin về những gì sẽ được commit trong lần commit sắp tới. Nó còn được biết đến với cái tên "chỉ mục" (index), nhưng khu vực tổ chức (staging area) đang dần được coi là tên tiêu chuẩn.

Tiến trình công việc (workflow) cơ bản của Git:

1. Bạn thay đổi các tập tin trong thư mục làm việc.
2. Bạn tổ chức các tập tin, tạo mới ảnh của các tập tin đó vào khu vực tổ chức.
3. Bạn commit, ảnh của các tập tin trong khu vực tổ chức sẽ được lưu trữ vĩnh viễn vào thư mục Git.

Nếu một phiên bản nào đó của một tập tin ở trong thư mục Git, nó được coi là đã commit. Nếu như nó đã được sửa và thêm vào khu vực tổ chức, nghĩa là nó đã được staged. Và nếu nó được thay đổi từ khi checkout nhưng chưa được staged, nó được coi là đã thay đổi. Trong Chương 2, bạn sẽ được tìm hiểu kỹ hơn về những trạng thái này cũng như làm thế nào để tận dụng lợi thế của chúng hoặc bỏ qua hoàn toàn giai đoạn tổ chức (staged).

[prev](http://git-scm.com/book/vi/B%E1%BA%AFt-%C4%90%E1%BA%A7u-S%C6%A1-L%C6%B0%E1%BB%A3c-L%E1%BB%8Bch-S%E1%BB%AD-c%E1%BB%A7a-Git) | [next](http://git-scm.com/book/vi/v1/B%E1%BA%AFt-%C4%90%E1%BA%A7u-C%C3%A0i-%C4%90%E1%BA%B7t-Git)