# Agile, Big Data, and Analytics

Giới thiệu

Phương pháp Agile và Lean là cách tốt nhất hiện nay để phát triển và tạo ra phần mềm phức tạp đáp ứng những rủi ro, yêu cầu thay đổi và sự ko chắc chắn trong việc phát triển phần mềm. Trong bài viết này sẽ phân tích áp dụng phương pháp Agile vào Big Data (và cả dữ liệu nhỏ). Bài tóm tắt này sẽ phân tích các ngành nghiên cứu dữ liệu, việc phát triển dữ liệu, và áp dụng phương pháp Agile.

1. **The Data Scientist, the Software Developer, and the Data Engineer**

Data Scientist (Ngành nghiên cứu dữ liệu) yêu cầu đòi hỏi lập trình, nhưng các nhà nghiên cứu cơ sở dữ liệu lại không được thường xuyên đào tạo về lập trình và thực hành các công nghệ mới, phương pháp làm phần mềm hiện đại. Ngược lại nhiều Software Developer (Người phát triển phần mềm) có kỹ năng về cớ sở dữ liệu, thống kê, nhưng đây không phải là chuyên môn chính của họ. Data Engineer (Kỹ thuật dữ liệu) là mối liên kết giữa Data Scientist và Software Developer.

1. **Analytics vs. App Dev**

* AppDev là sự kết hợp bởi hai kỹ thuật phức tạp và sự thay đổi yêu cầu, trọng tâm là hướng về thực hiện khối chức năng dựa trên yêu cầu kinh doanh phần mềm. App Dev tập trung vào việc thiết kế giải pháp tốt nhất cho mục tiêu kinh doanh.
* Analytics lại rất phức tạp do sự không chắc chắn về thao tác xuất hiện trong cơ sở dữ liệu. Analytics tập trung vào dữ liệu thực nghiệm xác định các dữ liệu tồn tại hỗi trợ các mục tiêu kinh doanh.
* Kết Luận: Cả hai lĩnh vực đều rất phức tạp nhưng theo tính chất khác nhau và đòi hỏi kỹ năng và kinh nghiệm khác nhau.

1. **Vulue-Driven Analytics**

Agile analytics đặc biệt tập trung vào sự phát triển thường xuyên, vòng lặp để duy trì sự hoat động. Mục đích là thực hiện một khám phá nhỏ có được nó trong tay của nhà sản xuất quyết định kinh doanh, và đánh giá tính hữu dụng của các kết quả. Các nhà nghiên cứu cơ sở dữ liệu thường sử dụng dữ liệu cũ để xây dựng mô hình thử nghiệm. Điều này cho phép các nhà nghiên cứu dữ liệu có được một mô hình tốt.

1. **Minimizing Initial Investment**

Agile Analytics xây dựng cơ sở ban đầu khá hiệu quả, phát triển mô hình một cách nhanh chóng, sau đó phát triển nó theo kiểu vòng lặp thay thế nó tốt hơn trước. Agile tạo ra một vòng phản hồi để giám sát tác đông việc sử dụng mô hình, cách tiếp cận này phép đầu tư ban đầu nhỏ nhưng có hiệu quả.

1. **Operationalizing Analytics**

Operation Analytics là quá trình triển khai một mô hình phân tích, dữ liệu sản xuất. Không phải tất cả các pha phân tích đều đảm bảo việc vận hành. Quyết định chiến lược đựa trên những phát hiện, chứ ko phải là việc triển khai các mô hình phân tích hoạt động.

1. **Lab and Factory Model**

Lab (Phòng thí nghiệm) là nơi thử nghiệm và phát hiện xảy ra trong khi đó Factory (Nhà máy) là nơi các nghiên cứu, phân tích được đưa vào sản xuất, là nơi những hiểu biết sâu sắc được mã hóa và thử nghiệm. Trong quá trình đó Agile được áp đụng trong từng giai đoạn khi vừa phát triển và chạy chương trình.

1. **Agility in the Lab**

Trong các phòng nghiên cứu phân tích việc tăng số các thực nghiệm giống với phương pháp Agile, do đó phát triển trong phòng Lab không phải là thường thử nghiệm điều khiển và có thể là một chút hack hơn cũng có thể. Tuy nhiên, một số các mô hình này đủ chất lượng để áp dụng vào thực tiễn.

1. **Agility in the Factory**

Factory(Nhà máy-Công ty) chủ yếu quan tâm đến việc làm thế nào để kết quả phân tích được áp dụng một cách phổ biến và rộng lớn, cải tiến gia tăng với mô hình phân tích và có thế tiếp tục trong phòng thí nghiệm, nhưng nhà máy là nơi mà các mô hình mới nhất được triển khai trực tiếp và có quy mô, dữ liệu lớn. Factory có kiến trúc tốt nhất để hỗi trợ triển khai modul và các mô hình thay thế. Bằng cách này các nhà nghiên cứu dữ liệu có thể tiếp tục phát triển mô hình của họ theo thời gian.

1. **Testing in Agile Analytics**

Kiểm tra và kiểm thử tự động là một vấn đề trong phân tích. Có hai khía cạnh để kiểm thử là

* + - Đảm bảo rằng mã phân tích được viết một cách chính xác.
    - Đảm bảo mô hình này là hợp lệ và chính xác.

Các xét nhiệm có thể được viết đáng tin cậy cho các chức năng và đảm bảo dữ liệu đầy đủ,, nhất quán và đúng đắn. Các mô hình phân tích luôn chứa tính đúng sai và độ chuẩn xác, các nhà Nghiên cứu dữ liệu sử dụng phương pháp để xác nhận mô hình phân tích và xác định độ chính xác của họ, trong khi đó các chuyên gia kinh doanh lại quan trọng những lợi nhuận của họ.

1. **Recommendations**

Nếu giải pháp phần mềm của bạn bao gồm một khả năng phân tích tốt, dữ liệu nghiên cứu là một vai trò đa chức năng chứ không phải là việc phát triển tuần tự. Các nhà nghiên cứu dữ liệu làm việc một hoặc hai lần lặp lại trước khi hoàn thành. Các nhà nghiên cứu dữ liệu làm việc với khách hàng để xác minh rằng mô hình là đúng. Khi đó, mô hình có thể áp dụng vào thực tế để triển khai. Tuy nhiên, hãy cẩn thận để không nhận thức nghiên cứu dữ liệu như là một hoạt động riêng biệt, tách rời. Giống như phương pháp tiến hóa, nghiên cứu dữ liệu có thể chạy trước chỉ là một chút của sự phát triển, nhưng phải liên quan đến sự hợp tác có hiệu quả.