

STEVE TENDON & WOLFRAM MÜLLER

DOMANDO O FLUXO

Gestão Hiperprodutiva
do Trabalho Intelectual



Versão Brasileira

Adail Retamal

Domando o Fluxo

Gestão Hiperprodutiva do Trabalho Intelectual

Steve Tendon, Wolfram Müller e Adail Retamal

Esse livro está à venda em <http://leanpub.com/tame-the-flow-BR>

Essa versão foi publicada em 2021-04-21



Esse é um livro [Leanpub](#). A Leanpub dá poderes aos autores e editores a partir do processo de Publicação Lean. [Publicação Lean](#) é a ação de publicar um ebook em desenvolvimento com ferramentas leves e muitas iterações para conseguir feedbacks dos leitores, pivotar até que você tenha o livro ideal e então conseguir tração.

© 2013 - 2021 Steve Tendon, Wolfram Müller e Adail Retamal

Tweet Sobre Esse Livro!

Por favor ajude Steve Tendon, Wolfram Müller e Adail Retamal a divulgar esse livro no Twitter!

O tweet sugerido para esse livro é:

[Eu comprei e estou lendo "Domando o Fluxo"! #TameFlow](#)

A hashtag sugerida para esse livro é [#tameflow](#).

Descubra o que as outras pessoas estão falando sobre esse livro clicando nesse link para buscar a hashtag no Twitter:

[#tameflow](#)

Conteúdo

O Blog do Livro	i
Créditos	i
Agradecimentos	i
Sobre os Autores	i
Introdução	v
Quem Deve Ler Este Livro e Por Que	vii
Estrutura do Livro	vii
Como Ler Este Livro	ix
1. Melhorando Durante o Fluxo	1
1.1 Mínima Versão Viável (MVV)	1
1.2 Mínima Versão Viável como uma Unidade de Trabalho Limitadora do WIP	3
1.3 Gerencie o Risco Variando o Tempo, Não o Escopo	5
1.4 O Pulmão MVV	8
1.5 Dimensionamento do Pulmão	11
1.6 Gestão, Uso e Interpretação do Pulmão	15
1.7 Gráficos do Pulmão	21
1.8 Como Construir e Monitorar um Pulmão MVV	29

O Blog do Livro

Você pode ler sobre a evolução dos conceitos apresentados neste livro no blog tameflow.com/blog¹.

Créditos

Imagen da capa: *Fractal Phlegyas on the River Styx*, ©2008, Devin Moore
<http://www.devinmoore.com/>²

Agradecimentos

Agradecemos a Rudi Burkhard por conectar Steve e Wolfram.

Agradecemos a todos os revisores pelos excelentes comentários, percepções, sugestões e até incentivo. Em particular: Michael Burrows, Dimitar Bakardzhiev, Adail Muniz Retamal, Paul Merino, Pascal Van Cauwenberghe, Simon Harris, Andy Carmichael, Rüdiger Wolf e Niranjana Koodavalli.

Sobre os Autores

Vivemos num mundo em constante mudança. Você pode descobrir as últimas sobre os autores nos seus perfis online.

Steve Tendon

- LinkedIn: <http://www.linkedin.com/in/tendon>³
- Twitter: @tendon <https://twitter.com/tendon>⁴
- TameFlow: <https://tameflow.com>⁵

¹<https://tameflow.com/blog>

²<http://www.devinmoore.com/>

³<http://www.linkedin.com/in/tendon>

⁴<https://twitter.com/tendon>

⁵<https://tameflow.com>

- Pessoal: <https://tendon.net>⁶

Steve é um consultor experiente, especializado em *hiperprodutividade organizacional e evolução de processos* tanto para negócios pequenos e em crescimento, como também para organização grandes e estabelecidas. Ele evoluiu sua própria abordagem única para a gestão de negócios, inspirado por uma diversidade de teorias gerenciais contemporâneas e baseado em como a informação flui através da organização moderna (computadorizada), tudo isso enquanto prestando a devida atenção aos fatores humanos, com suas necessidades psicológicas e interações interpessoais.

Ele possui um MSc. em Gestão do Processo de Software (Universidade de Aberdeen).

Alguns clientes atendidos: Wolters-Kluwer (Holanda), CCH Software (Reino Unido), Norsteds-Juridik (Suécia), and IPSOA (Itália).

Wolfram Müller

- LinkedIn (em inglês): <https://www.linkedin.com/in/blue-dolphin>⁷
- XING (em alemão): http://www.xing.com/profile/Wolfram_Mueller⁸
- Site (em alemão): <http://speed4projects.net>⁹
- Site (em inglês): <http://reliable-scrum.info>¹⁰

Wolfram Müller nasceu em 1969. Estudou Mecatrônica e Engenharia Mecânica. Trabalhou por dois anos para a BARD/angiomed e obteve experiência prática no desenvolvimento e manufatura de dispositivos médicos. Nessa época ele aprendeu como funcionam as ferramentas da gestão tradicional de projetos e sobre suas desvantagens. Durante seus estudos ele foi ativo como freelancer em vários projetos de desenvolvimento de software e descobriu a alegria dos projetos rápidos.

De 2000 a 2010 ele trabalhou como desenvolvedor e, posteriormente, como gerente do Escritório de Projetos da 1&1 Internet AG e foi responsável por mais de 500 projetos. Ele usou ideias vindas dos domínios de Lean, TOC e Agilidade. Implementou Corrente Crítica numa das unidades de negócio da 1&1 com grande sucesso. Com base nisso, desenvolveu complementos para os Métodos Ágeis para torná-los compatíveis à Corrente Crítica. Seu foco é sempre na velocidade, ganho, confiabilidade e agilidade.

⁶<https://tendon.net>

⁷<https://www.linkedin.com/in/blue-dolphin>

⁸http://www.xing.com/profile/Wolfram_Mueller

⁹<http://speed4projects.net>

¹⁰<http://reliable-scrum.info>

Desde 2005 Wolfram Müller oferece sua experiência a várias empresas interessadas, na forma de treinamento, coaching e aconselhamento sob o marca “Speed4Projects”.

Adail Retamal

- LinkedIn: <http://www.linkedin.com/in/adail¹¹>
- Twitter: @TheAdail <https://twitter.com/TheAdail¹²>

Adail foi um dos pioneiros em Métodos Ágeis no Brasil, desde os idos de 2001 com a XP, ajudando a trazer uma das primeiras turmas de CSM (Certified Scrum Master) para o país em 2007, e depois trazendo David Anderson pela primeira vez para seus cursos de Gestão Zen (2008) e Kanban (2010). Também é bem conhecido como um dos principais promotores da FDD (Feature-Driven Development) e da Teoria das Restrições (TOC) desde 2003. Trabalhou por dois anos em Cingapura, no PMO de um dos maiores projetos de implantação de SAP na época (2010-2012), para uma das maiores empresas de mineiração do mundo. Desde então ele mora na Austrália, onde ajudou grandes e médias empresas a adotarem a TOC e a Gestão de Projetos pelo Método da Corrente Crítica, com ou sem Métodos Ágeis já em operação. Atualmente é Engenheiro de Sistemas numa grande multinacional na área de transporte.

É graduado em Engenharia Eletrônica/Computação e atuou pela sua empresa Heptagon TI desde 1997 até sua transferência em 2018. Trabalhou na Borland Latin America por 4,5 anos (2002-2006) e quando leu a história contada por Steve no capítulo 1 se identificou muito, o que também o motivou a traduzir este livro. Professor por opção e satisfação, se realiza ensinando e ajudando equipes a se tornarem mais eficazes e eficientes, com muita atenção aos aspectos psicológicos e sociais dos ambientes de projetos.

Aviso

Por favor observe que:

- Nenhum dos autores são falantes nativos em inglês, então tolerem nosso comando da língua. Se você notar erros de qualquer tipo, nós alegremente receberemos sua resposta e faremos as correções. Este livro está sendo inteiramente publicado independentemente. Nós não possuímos os recursos dedicados de uma editora (como revisores e outros profissionais). Confiamos que você apreciará mais as ideias e os conceitos apresentados, em vez da linguagem que possamos ter usado ou abusado.

¹¹<http://www.linkedin.com/in/adail>

¹²<https://twitter.com/TheAdail>

- Quaisquer citações são feitas de boa fé e com base num uso justo, para o propósito de representar fielmente os pensamentos originais e atribuídas aos seus autores. Se você é detentor dos direitos autorais e acha que foi citado de forma inapropriada, por favor entre em contato conosco para esclarecermos o assunto.

Introdução

Este livro é sobre como gerenciar o trabalho intelectual e, em particular, sobre como trazer uma organização intelectual para um estado de *hiperprodutividade*. Equipes e organizações hiperprodutivas sempre existiram, tanto no campo de negócios quanto militar. Evidentemente, elas são pontos raros fora da curva. O foco principal deste livro é como se pode construir e gerenciar uma *organização de trabalho intelectual hiperprodutiva*, tomando como fonte de inspiração a experiência amadurecida no campo da engenharia de software.

Organizações de desenvolvimento de software — como Microsoft, Google e Oracle; mas também organizações de código aberto bastante visíveis, como Apache, Ubuntu, Drupal e outras — podem ser consideradas como *arquétipos de organizações baseadas em conhecimento*, porque *a totalidade de artefatos produzidos por um negócio de software é puramente imaterial*. Organizações de desenvolvimento de software foram as primeiras a confrontar os desafios dos processos plenamente imateriais, inteiramente baseados no conhecimento. Elas também foram as primeiras organizações a experimentar o impacto sócio-tecnológico da tecnologia da informação (especialmente o desenvolvimento das redes de computadores) em seus processos internos de trabalho e estruturas organizacionais.

Devido a esse ritmo frenético e exponencial de inovação, o campo do desenvolvimento de software também tornou-se uma plataforma de testes para uma variedade de abordagens e métodos para gerenciar o trabalho intelectual imaterial, tipicamente sob a forma de processos e metodologias de desenvolvimento de software. Numerosas alternativas evoluíram num período relativamente curto, com muitas competições saudáveis entre elas, com “guerras santas” veementes travadas entre os proponentes de cada abordagem, e com muitos relatos tanto sobre sucessos quanto fracassos.

Percepções valiosas podem ser obtidas ao se estudar as abordagens organizacionais e gerenciais adotadas com sucesso pelo mercado, que produziram a revolução da informação — isto é, aquelas organizações que primeiramente se engajaram no desenvolvimento de software — e então extrapolando esses processos organizacionais e gerenciais para o caso mais genérico das organizações baseadas em conhecimento.

Atualmente a maioria das organizações tornou-se baseada em conhecimento. Mesmo os negócios mais resilientes e tradicionais são forçados a tornarem-se organizações baseadas em conhecimento. Por exemplo, considere o impacto da impressão em 3D em larga escala na indústria da construção (“*Contour Crafting*”), onde tijolos e argamassa literalmente

tornam-se software. Ou as nanotecnologias, onde a manufatura no nível subatômico torna-se software.

O princípio chave deste livro é que a hiperprodutividade organizacional origina-se a partir de dois *Padrões Nobres* de Unidade de Propósito e Comunidade de Confiança. Os líderes e executivos do alto escalão da organização desempenham um papel crítico na criação das condições para que esses padrões se tornem efetivos. A gerência deve contribuirativamente na liderança da organização em direção a essas condições.

No caso de organizações de software (e conhecimento) a gerência terá que obter um entendimento profundo da natureza do software ou do trabalho intelectual, e essa natureza está enraizada no empirismo. Os altos executivos já possuem esse entendimento, embora possam não estar cientes disso e – infelizmente – não agem por conseguinte. A gerência deve criar as condições para o desenvolvimento de uma *organização que aprende*. A responsabilidade financeira deve ser exercitada de forma diferente, e de uma maneira que seja compatível com a natureza empírica do trabalho intelectual. Papéis fundamentais devem ser cultivados, enquanto que cada equipe precisa desenvolver sua própria visão compartilhada do seu propósito.

Ajudas práticas vêm de muitas fontes e nós examinaremos especificamente: o Método Kanban, Scrum e a Teoria das Restrições. A Teoria das Restrições cultiva a Unidade de Propósito e a Comunidade de Confiança de muitas formas. Em particular, a Teoria das Restrições trata diretamente da necessidade de se chegar a métricas comuns, pelas quais tudo e todos são direcionados, e supera os impedimentos estruturais (enraizados no uso da contabilidade de custos e nas métricas de eficiência para propósitos gerenciais) que são as fontes da divergência de propósito e dos objetivos ocultos.

O Método Kanban e Scrum são largamente usados nos modernos negócios de trabalho intelectual. A Teoria das Restrições pode estendê-los de maneiras poderosas, trazendo mais previsibilidade do comportamento do sistema como um todo, e também dos indivíduos. A combinação deles torna-se um poderoso terreno fértil para o desenvolvimento da Unidade de Propósito e Comunidade de Confiança. Tanto o Método Kanban quanto Scrum podem ser estendidos com as funcionalidades da Teoria das Restrições, e ajudam a criar uma organização hiperprodutiva. O livro apresentará maneiras práticas pelas quais tal combinação pode ser realizada.

Quem Deve Ler Este Livro e Por Que

O livro é sobre hiperprodutividade organizacional no trabalho intelectual. Durante suas consultorias, Steve geralmente trabalha com empresas numa maneira holística, para fazer com que cheguem a um estado de hiperprodutividade. Para se ter sucesso todos precisam ser envolvidos, desde o Diretor Executivo até o último novato contratado. Tal trabalho tipicamente avança em duas direções paralelas: de baixo para cima e de cima para baixo. Isso explica a estrutura, objetivos e propósito deste livro.

As partes I e II visam proprietários de negócios, diretores e outros executivos de alto nível e a alta gerência. A preocupação é principalmente entender sobre hiperprodutividade e sobre o que precisa-se levar em consideração para criar um novo ou evoluir um negócio existente para uma organização hiperprodutiva. O foco é organizacional e gerencial. O objetivo é chegar ao aumento espetacular no ganho financeiro. Isso corresponde à abordagem de cima para baixo.

As partes III e IV visam os gerentes de nível intermediário, gerentes de projetos, scrum masters e membros de equipes. Em particular: praticantes usando (ou pretendendo usar) o Método Kanban (de David Anderson), Scrum (de Jeff Sutherland) ou CCPM (Corrente Crítica, de Eliyahu Goldratt). A preocupação é sobre como aumentar o “fluxo” do trabalho intelectual. O foco é no operacional e prático. O objetivo é chegar ao espetacular aumento no ganho operacional. Isso corresponde à abordagem de baixo para cima.

Estrutura do Livro

Este livro está dividido nas seguintes partes:

Parte I - O Que e Por Que: Esta parte do livro inicia descrevendo o que é a hiperprodutividade num negócio de software, apresentando o caso da Borland International e da equipe do Quattro Pro for Windows. Essa equipe de desenvolvedores foi a mais produtiva já documentada. A história da Borland International é tomada como um exemplo de como a hiperprodutividade é uma característica que pode ser adquirida, transferida e perdida; e então lança os fundamentos para tornar possível trazer outras organizações para um estado hiperprodutivo. São apresentadas considerações sobre por que pode ser vantajoso importar-se com a hiperprodutividade numa organização intelectual. Dois padrões fundamentais são identificados como Unidade de Propósito e Comunidade de Confiança.

Parte II - Gestão, Liderança e Organização: Esta parte do livro apresenta uma coleção de pensamentos sobre como podemos ligar a percepção empírica da gestão estratégica à

natureza empírica do desenvolvimento de software, e examina que tipo de abordagens gerenciais permitem trazer o empirismo além do nível estratégico e para dentro do exercício diário da gestão operacional. Descreveremos várias abordagens gerenciais que podem contribuir de alguma forma para construir organizações intelectuais hiperprodutivas. A natureza do trabalho intelectual (como software) é investigada, assim como um paralelo pode ser traçado entre a gestão do trabalho intelectual e as atividades de criação de estratégias dos líderes de negócio. O processo iterativo, exploratório, adaptativo e empírico que está na raiz da criação de estratégias é o mesmo usado no trabalho intelectual, portanto a alta gerência está, pelo menos conceitualmente, bem equipada para gerenciar o trabalho intelectual. O processo principal é um processo de aprendizagem social. A responsabilidade e as vulnerabilidades da gerência ao construir uma organização que aprende são enfatizadas. Abordagens de gestão conhecidas, que ajudem neste processo, são examinadas, tais como: planejamento dirigido pela descoberta, além da orçamentação, financiamento incremental e contabilidade de ganhos. A seguir, os papéis críticos das organizações hiperprodutivas são investigados, através das lições aprendidas com o software de código aberto. Além disso, antipadrões para a hiperprodutividade são brevemente considerados, tais como alguns papéis típicos do Scrum. A importância do orgulho do trabalhador, diversão e folga são ressaltados, assim como abordagens práticas para criar visões compartilhadas no nível da equipe, que então se relacionam à Unidade de Propósito no nível organizacional.

Parte III - Na Prática com o Método Kanban: Esta parte do livro entra em detalhes práticos sobre como gerenciar o trabalho intelectual através do Método Kanban. Primeiro mostra-se como existem fortes ligações entre o Método Kanban e a Teoria das Restrições. A Teoria das Restrições é uma das abordagens de maior sucesso para melhorar o desempenho das organizações, embora ela não tenha sido plenamente explorada no contexto do trabalho intelectual. A intenção desta parte do livro é mostrar que a combinação entre o Método Kanban e a Teoria das Restrições é realmente excelente. Por um lado, o Método Kanban fornece as ferramentas para lidar com a gestão do trabalho intelectual. Por outro lado, a Teoria das Restrições fornece o foco — especialmente em termos para determinar a Unidade de Propósito — e a alavancagem que permitirá uma organização melhorar continuamente, em direção a níveis cada vez maiores de desempenho. Algumas das desvantagens fundamentais de implementações comuns do Método Kanban também são examinadas, e resolvidas através da aplicação da Teoria das Restrições.

Parte IV - Na Prática com Scrum: Esta parte do livro, de modo similar à anterior, é de natureza prática e mostra como as melhorias significativas na produtividade podem também ser alcançadas ao se começar com o Scrum, em vez de com o Método Kanban. Aqui também a Teoria das Restrições é usada como um catalisador que possibilita a melhoria contínua de desempenho. A mudança organizacional é considerada a partir do ponto de vista da

Teoria das Restrições, e especialmente como ela se relaciona ao fluxo de trabalho através de uma organização baseada no trabalho intelectual. A natureza das restrições é explicada, e também onde e como começar uma iniciativa de melhoria, indo de grandes versões até às listas de tarefas. A execução do controle torna-se primordial para equilibrar recursos em relação à demanda e conseguir entregar dentro do prazo, finalmente reforçando a Comunidade de Confiança que está crescendo dentro da organização. Ferramentas específicas da Teoria das Restrições — como o agendamento Tambor-Pulmão-Corda e a gestão de pulmões da Corrente Crítica — são descritas de maneira a melhorar ainda mais o Scrum.

Como Ler Este Livro

Se você estiver interessado em entender sobre as principais funcionalidades que caracterizam uma organização hiperprodutiva, você pode ler a Parte I. Ela também contém um capítulo sobre [Padrões e Linguagens de Padrões](#), que você pode pular a menos que esteja interessado nesse tópico específico. Há dois motivos pelos quais os padrões são apresentados. Primeiro, os padrões organizacionais são um dos resultados dos primeiros estudos sobre a hiperprodutividade. Padrões são um fundamento útil para raciocinar sobre a hiperprodutividade. Segundo, a Teoria de Padrões e os Padrões são usados para analisar e decompor vários métodos e abordagens em suas partes constituintes, que então são recombinados em novas configurações. É assim que elementos do Método Kanban, do Kanban do Sistema de Produção Toyota, do Método do Financiamento Incremental e da abordagem de Gestão de Projetos pelo Método da Corrente Crítica são combinados para formar as inovações apresentadas na Parte III.

Se você quer saber o que pode afetar a hiperprodutividade numa organização, a partir de um ponto de vista gerencial, você pode ler a Parte II. Mesmo que os capítulos na Parte II possam referenciar uns aos outros, e sejam apresentados numa sequência importante, você pode escolher ler qualquer capítulo que estiver interessado, em qualquer ordem. Os tópicos cobertos na Parte II não conduzem a um estado de hiperprodutividade por si mesmos, mas todos contêm elementos (Padrões) que podem contribuir para esse fim. Você não deve procurar por receitas na Parte II, mas sim por fontes de inspiração e reflexão. Esses capítulos enfatizarão como você pode chegar a um estado de fluxo *organizacional, psicológico e financeiro*.

As Partes III e IV são de natureza mais prática. Você se interessará pela Parte III se estiver usando (ou pretendendo usar) o Método Kanban, e na Parte IV se estiver usando (ou pretendendo usar) Scrum. As duas Partes mostram como estender o Método Kanban e o Scrum de várias formas, em particular com as ideias vindas da Gestão de Projetos pelo

Método da Corrente Crítica (*Critical Chain Project Management - CCPM*) da Teoria das Restrições. Os capítulos nessas partes são melhor lidos em sequência, na ordem apresentada. Por analogia, se você estiver usando a CCPM, poderá estar interessado em como você pode torná-la mais ágil através do Método Kanban ou do Scrum.

Os métodos apresentados nas Partes III e IV vão ajudá-lo a melhorar o fluxo *operacional* do trabalho através de sua organização. Eles constituem o fundamento prático que *pode* permitir que você chegue num estado de hiperprodutividade, mas você precisará encontrar sua própria solução, combinando o fundamento prático com as ideias, princípios e valores descritos na Parte II. Se precisar de ajuda para conseguir fazer isso, você está convidado a entrar em contato com os autores.

1. Melhorando Durante o Fluxo

NOTA IMPORTANTE: O que examinaremos neste capítulo e no próximo ([Análise de Causa Raiz do Jeito TOC](#)) aplica-se *tanto* ao Kanban convencional (com limites de WIP no estado de trabalho) *quanto* ao Hiper-Kanban (como descrito anteriormente).

Em particular, observe como a abordagem apresentada não altera nada no Método Kanban como é apresentado e ensinado por David Anderson [[ANDERSON-2010](#)⁴]. Esta abordagem é uma melhoria que fornece uma ferramenta poderosa para a melhoria focada e contínua para o Kanban e o Hiper-Kanban, ao focar na variação por causa comum, em vez de na variação por causa especial.

⁴<https://tameflow.com/bibliography/#ANDERSON-2010>

Neste capítulo veremos como podemos melhorar nosso *processo* de trabalho ao mesmo tempo em que o *fluxo* de trabalho está fluindo regularmente, num ritmo sustentável. Esta técnica, que combina ideias do Método Kanban, o Método de Financiamento Incremental (MFI)¹ e a Gestão de Projetos pela Corrente Crítica² da Teoria das Restrições, que também fornecerá a fundação para fazer a análise de causa raiz profunda (como veremos no capítulo [Análise de Causa Raiz do Jeito TOC](#)).

1.1 Mínima Versão Viável (MVV)³

No capítulo [Método de Financiamento Incremental](#) nós aprendemos sobre o conceito de **Mínimas Funcionalidades Viáveis** (MFV)⁴, como inicialmente descritas por [[DENNE-2004a](#)⁵]. Uma MFV é um conjunto *mínimo* de funcionalidades que entregam valor de mercado.

No Kanban o conceito MFV foi reavivado e é conhecido como uma **Mínima Versão Viável**

¹N.T.: IFM — *Incremental Funding Method*

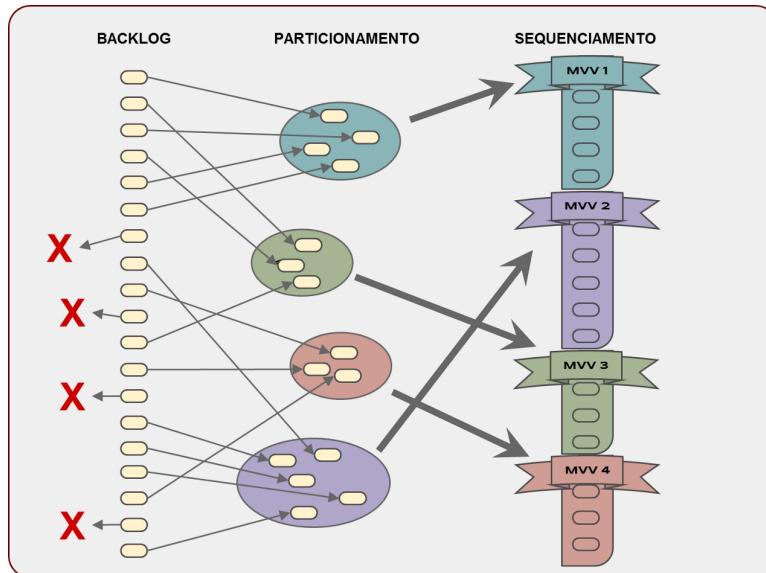
²N.T.: CCPM — *Critical Chain Project Management*

³N.T.: MMR — *Minimum-Marketable Release*

⁴N.T.: MMF — *Minimum-Marketable Features*

⁵<https://tameflow.com/bibliography/#DENNE-2004a>

(MVV)⁶. David Anderson [ANDERSON-2010⁷] define uma MVV como “*alguma unidade atômica de valor para o mercado ou cliente.*” Para todos os propósitos práticos, os dois conceitos são equivalentes e podem ser considerados como sinônimos, porque como veremos, os traços importantes estão na *minimalidade* e *viabilidade*. Kanban não prescreve o uso de MVVs, mas sugere que elas *podem* ser usadas. Às vezes, pelo fato do Kanban tentar enfatizar o fluxo contínuo, o uso de MVVs é desencorajado, porque elas são percebidas como uma interrupção do fluxo. MVVs também podem ser desencorajadas por parecerem implicar num comprometimento *antecipado* com as funcionalidades que elas incluem. Entretanto, com relação à questão do compromisso antecipado, se considerarmos as MVVs como unidades *atômicas* de valor, então o compromisso é com esse valor, com as MVVs como um todo, e não sobre os itens de trabalho individuais que podem fazer parte de uma MVV.



MVVs são uma sequência ordenada de subconjuntos de um backlog. Cada MVV é mínima e viável. A sequência é tal que ela maximiza o valor de negócios do projeto. Observe que alguns itens do backlog podem ser descartados no processo.

Se o conjunto de trabalho a ser feito é considerado como um backlog (como no Scrum e outras

⁶O movimento *Lean Startup* (Empreendedorismo Enxuto) popularizou recentemente o conceito de “*Produto Mínimo Viável*” (PMV) (*Minimum Viable Product* – MVP) e geralmente ele é igualado às MFVs e MVVs. Observe, porém, que os dois conceitos, PMV por um lado, e MFV/MVV por outro, são inteiramente diferentes. Um PMV é usado principalmente para testar uma *hipótese* sobre a viabilidade comercial de um produto mínimo. Uma MFV/MVV considera a viabilidade como condição certa que é assumida. Use PMVs se você precisa explorar sobre as preferências dos novos mercados ou a aceitação de novas ideias de produtos. Use MFVs/MVVs se você tem um produto estabelecido com um backlog de funcionalidades que foram requisitadas (e validadas) pelos seus mercados existentes.

⁷<https://tameflow.com/bibliography/#ANDERSON-2010>

abordagens ágeis), então as MVVs são simplesmente um *particionamento* desse conjunto em subconjuntos, e então *ordenação* desses subconjuntos numa sequência que deve maximizar o valor, como ilustrado na figura. É útil pensar sobre as MVVs desta forma.

O MFI ensina como descobrir a *sequência ótima de implementação* de tais subconjuntos, para maximizar o *valor presente líquido* de um projeto. Embora possamos aplicar as técnicas do MFI, nós consideramos tais técnicas de sequenciamento como opcionais. De fato, no Kanban, a seleção para puxamento é mais apropriadamente tratada considerando-se o custo da demora e outros fatores de risco. O foco da presente discussão, todavia, não é sobre *estratégias de seleção para puxamento*. É sobre as características das MVVs que as tornam valiosas de uma perspectiva de gestão de riscos e melhoria de processos, quando combinadas com outros elementos da TOC.

1.2 Mínima Versão Viável como uma Unidade de Trabalho Limitadora do WIP

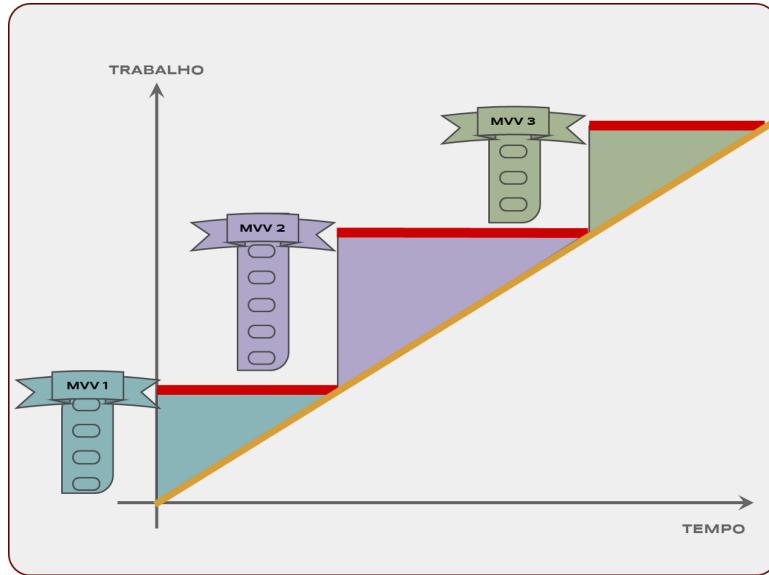
Uma MVV é uma *unidade de trabalho* que é uma coleção — um lote, de fato — de itens de trabalho. Os itens individuais de trabalho podem pertencer a classes de serviço diferentes, se necessário. O princípio fundamental é que todos os itens de trabalho pertencentes à uma MVV devem ser entregues juntos, como uma unidade só. Nós vamos proveitosamente considerar uma MVV primariamente como uma *unidade de trabalho que é liberada para o fluxo de trabalho*, e secundariamente como uma *unidade de valor que é liberada para o mercado* quando finalizada.

Uma MVV é como um Projeto de Escopo Fixo

Uma MVV representa a *quantidade mínima de funcionalidade* que pode ser oferecida com sucesso ao mercado. *Minimalidade* e *viabilidade* juntas possuem uma consequência muito importante: uma MVV torna-se como um pequeno *projeto de escopo fixo*. Pela definição de MVV nós não podemos retirar itens de trabalho dela, ou ela perderia sua viabilidade. Analogamente, não podemos adicionar itens a uma MVV, ou ela não seria mais mínima. Portanto, já que não podemos retirar ou adicionar itens, a MVV é similar a um miniprojeto de escopo fixo.

Uma MVV pode ser considerada como uma *unidade de trabalho com escopo fixo e valor de mercado*, que deve ser entregue na sua totalidade. Nem mais, nem menos.

Nós colocamos uma forte ênfase no **Escopo Fixo**: não podemos mudar o escopo de uma MVV, para que o próprio propósito da definição e identificação da MVV, em primeiro lugar, se torne em vão. (Naturalmente, se houver um erro na definição inicial da MVV, então ela precisa ser alterada, mas essa é outra questão. Vamos assumir que a definição do escopo está apropriada e correta.)



Cada MVV é como um mini-projeto de escopo fixo (como o Herbie caminhando pela trilha em estágios). Uma MVV naturalmente limita o WIP, de forma que a demanda (sendo fixa) é menor que a capacidade. O Tempo de Fluxo é proporcional ao WIP, e pelo fato do WIP ser limitado existe um “fim” para o mini-projeto.

Uma MVV Limita o Trabalho em Processo

Num diagrama de fluxo de trabalho acumulado, a MVV aparece como uma linha de demanda horizontal. Em outras palavras, o *trabalho em processo é limitado desde o início*. A linha de capacidade cresce progressivamente até interceptar a linha da demanda. Nesse ponto todo o trabalho foi realizado e pode ser entregue.

Em virtude da Lei de Little, podemos assumir que existe um *tempo de fluxo finito* para a entrega da MVV. Em outras palavras, porque a MVV interrompe um fluxo (ideal) contínuo de trabalho, nós temos um ponto futuro no tempo quando a MVV será entregue. (Veremos em breve como chegar nesse ponto no tempo.)

Não obstante uma MVV interromper o fluxo contínuo de trabalho, ela traz a importante

vantagem de *limitar* o trabalho em processo, um mecanismo de controle fundamental que favorece o fluxo *melhor*.

1.3 Gerencie o Risco Variando o Tempo, Não o Escopo

Kanban não exige uma gestão explícita de riscos, pois existem muitas competências para tratamento de riscos “embutidas” e “gratuitas”, mesmo com a adoção mais superficial do Kanban. A gestão de riscos dirigida por eventos, que é possibilitada pelos limites de WIP nos estados de trabalho do Kanban (não obstante suas muitas desvantagens, mencionadas várias vezes neste livro) ou por sinais do Hiper-Kanban, vão adequadamente tomar conta da maioria dos riscos de *variação por causa especial*.

Embora não prescreva a gestão de riscos, Kanban sugere fortemente que a gestão explícita de riscos deva ser realizada mesmo assim. Por outro lado, o Método de Financiamento Incremental, onde as MVVs foram inicialmente conceitualizadas, *exige* a gestão explícita de riscos precisamente por causa da natureza (escopo fixo!) das MVVs.

Consequentemente, se nós usamos o Kanban com MVVs — que também têm escopo fixo — então a gestão explícita de riscos torna-se obrigatória.

Além disso, o MFI obriga a gestão explícita de riscos, mas não prescreve ou define como realizá-la. Portanto torna-se interessante ver como o risco pode ser gerenciado.

“Cortar o Backlog” Não o Corta

Deve-se observar, neste ponto, que a abordagem ágil mais comum e popular para gestão de riscos, a de *cortar o backlog*, não pode ser aplicada. Cortar o backlog não funciona com uma MVV de escopo fixo, por causa das propriedades de minimalidade e viabilidade, como já explicado. O risco deve ser controlado de alguma outra forma.

Lições da Gestão de Projetos pela Corrente Crítica

Se não podemos variar o escopo por causa da natureza da MVV, precisamos melhorar na gestão do *tempo*. De fato, é possível *lidar com riscos considerando o tempo!* É aqui que podemos ser inspirados pela abordagem da Gestão de Projetos pela Corrente Crítica, da Teoria das Restrições. Vamos nos familiarizar com a GPCC.

A **Gestão de Projetos pela Corrente Crítica** (GPCC) é caracterizada por uma rede de atividades do projeto, tal como no tradicional *Método do Caminho Crítico* (CPM - *Critical Path Method*). O CPM é basicamente o método que é suportado pela maioria dos pacotes de software para gestão de projetos, como o Microsoft Project.

Diferentemente do Método do Caminho Crítico, a GPCC dá atenção explícita às *dependências de recursos* além das *dependências de tarefas*. (Sem a contenção de recursos, a GPCC e o CPM seriam essencialmente iguais, no que se refere à rede do projeto.)

Na GPCC a *Corrente Crítica* é definida por [SULLIVAN-2012⁸] como

A sequência mais longa de eventos dependentes através de uma rede de projeto, considerando tanto as dependências de tarefas quanto de recursos para completar o projeto. A corrente crítica é a restrição do projeto.

Uma percepção chave aqui é a seguinte: não importa que tipo de processo ou metodologia seja empregada, ao final do projeto todo o trabalho deve ter atravessado uma “*rede de tarefas executada pelos recursos*.” Um fato inegável permanece: para chegar à entrega final, a rede do projeto precisa ser atravessada. Dentro dessa rede do projeto a Corrente Crítica ainda é a restrição que limita a capacidade do projeto de entregar, não importa qual metodologia ou abordagem seja usada. Mesmo quando se usa o Kanban.

Na GPCC a rede do projeto é usada para *planejar* o projeto, tal como no Método do Caminho Crítico, mas a lógica usada para agendar as atividades é diferente. Para nossos propósitos, a *rede de projeto da GPCC não nos interessa*. Nós simplesmente assumimos que as atividades envolvidas na entrega de uma MVV são equivalentes às atividades representadas dentro de uma tal rede de projeto.

A outra grande diferença entre a GPCC e o CPM é que na GPCC um único **Pulmão de Projeto**⁹ é colocado no final da rede do projeto. No CPM o “acolchoamento”¹⁰ tipicamente ocorre no nível das tarefas.

Esse pulmão de projeto é um *pulmão de tempo*. No Kanban estamos muito preocupados com os aspectos temporais: normalmente usamos as métricas do prazo de entrega (*lead time*) e do tempo de fluxo para muitos propósitos, sendo um não menos importante deles o de estabelecer os acordos de nível de serviço associados às classes de serviço. Em breve veremos como vamos fazer uso dessa observação.

⁸<https://tameflow.com/bibliography/#SULLIVAN-2012>

⁹N.T.: *Project Buffer*

¹⁰N.T.: *Padding*

A TOC ensina-nos a identificar, proteger e elevar a restrição. Na rede do projeto, a restrição (isto é, a corrente crítica) é protegida e gerenciada pelo monitoramento constante do pulmão do projeto. Uma explicação superficial de porque o pulmão de projeto é necessário é que ele está lá para absorver qualquer variabilidade que possa afetar negativamente a entrega do projeto. Isso seria verdade se o pulmão fosse utilizado para o propósito de “enchimento”. Como veremos, esse não é o caso, pois usaremos o pulmão para outro propósito. Nós vamos monitorar constantemente o pulmão para obter sinais *adiantados* e *operacionais* sobre problemas vindouros, que podem afetar a saúde do projeto.

A parte essencial que preservaremos da GPCC é o pulmão de projeto ao final da rede do projeto; não nos interessam os aspectos de agendamento da GPCC e da real construção da rede do projeto. Diferentemente da rede, que é usada durante o *planejamento*, o pulmão do projeto é usado durante a *execução* do projeto. Nós executaremos todas as atividades de planejamento como sugerido pelo Kanban. O que vamos transferir da GPCC para o Kanban é a interpretação e uso do pulmão do projeto. Usaremos esse pulmão exatamente da mesma forma como é feito na GPCC e, assim, obteremos as vantagens bem documentadas, típicas da execução de projetos pela GPCC.

Tais vantagens incluem:

- Indicadores adiantados da materialização iminente do risco
- Compilação do registro de riscos *just-in-time*
- Análise de Frequência para identificar as fontes da variação por causa comum
- Análise de causa raiz, identificando a fonte daqueles problemas que são mais caros ou frequentes
- Processo de melhoria contínua, dirigido pelo foco e alavancagem de soluções para aqueles problemas caros ou frequentes

O resultado mais importante desta abordagem é que ela permitirá a melhoria do processo empregado, resultando em maior ganho. Nós seremos capazes de aumentar a capacidade da equipe e/ou processo, com melhorias que serão benéficas não apenas para o projeto que estivermos trabalhando no momento, mas também para todos os futuros projetos que assumirmos. As *melhorias são sistêmicas*.

O Melhor de Dois Mundos

É importante observar que as melhorias induzidas por esta abordagem não substituem aquelas alcançadas apenas com o Kanban ou Hiper-Kanban, pois as duas são complementares.

Uma completa a outra de forma significativa. Mais especificamente, nós alcançamos os benefícios de ambas assim:

- A gestão de riscos dirigida por eventos, típica do Kanban ou Hiper-Kanban, que nos permite reagir muito rapidamente à variação por causa especial (seja através dos limites de WIP por estado de trabalho ou através dos sinais do Hiper-Kanban); e
- A gestão de riscos dirigida por monitoramento do pulmão, típica da GPCC, que nos permite não apenas identificar as fontes da variação por causa comum, mas também discernir e selecionar aquelas que afetam negativamente as competências da equipe e/ou processo, e começar iniciativas de melhoria em torno delas.

Do mesmo modo, esta proposta pode ser vista a partir da perspectiva da Teoria das Restrições. O desenvolvimento de software viu uma proliferação de métodos Ágeis nos últimos 15-20 anos, que agora estão sendo adotados mais geralmente por todos os tipos de profissionais intelectuais. A GPCC está, de muitas formas, ancorada a abordagens que não podem se beneficiar dos avanços observados com os métodos Ágeis ou outros métodos modernos de gestão. Portanto, esta abordagem pode ser vista como uma melhoria que introduz o Kanban ou Hiper-Kanban dentro das práticas operacionais da Teoria das Restrições para execução de projetos (Gestão de Pulmão)¹¹.

Os praticantes da TOC ganharão as vantagens complementares oferecidas pelo Kanban ou Hiper-Kanban, que são muito importantes nas suas competências para gestão de riscos dirigida por eventos. Mais ainda, eles reconhecerão e apreciarão que os princípios, práticas e valores do Kanban estão surpreendentemente alinhados com o que Eliyahu Goldratt definiu como a *Visão Viável* e a *Empresa Sempre Próspera*.¹²

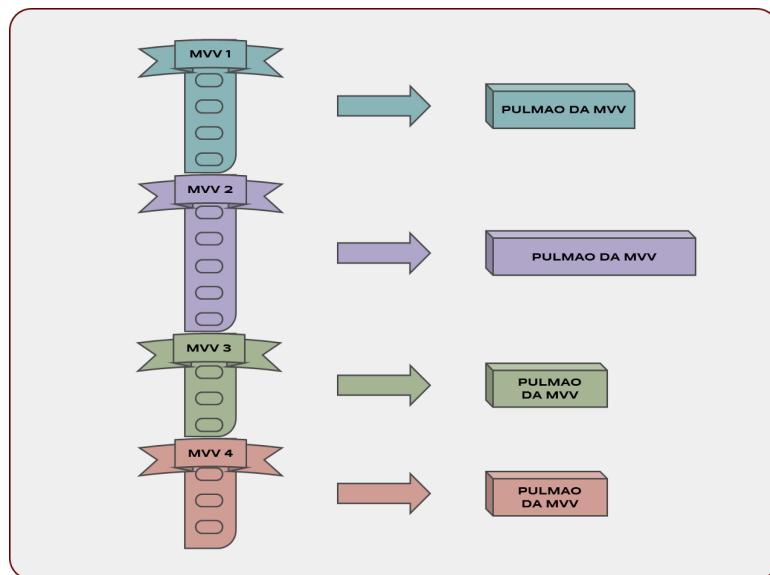
1.4 O Pulmão MVV

Da mesma forma que a GPCC protege as atividades na corrente crítica com um pulmão de (tempo de) projeto, também protegeremos a implementação de uma MVV ao (conceitualmente) adicionar um pulmão ao seu final e monitorá-lo constantemente.

Ao particionar um projeto em MVVs podemos considerar cada MVV como um projeto de escopo fixo. Para cada uma dessas MVVs sabemos desde o início quanto trabalho precisamos realizar. Ao somar os tempos médios de fluxo dos itens de trabalho que pertencem a uma MVV nós obtemos um tempo médio de fluxo esperado para ela.

¹¹N.T.: *Buffer Management*

¹²N.T.: *Viable Vision e Ever Flourishing Company*

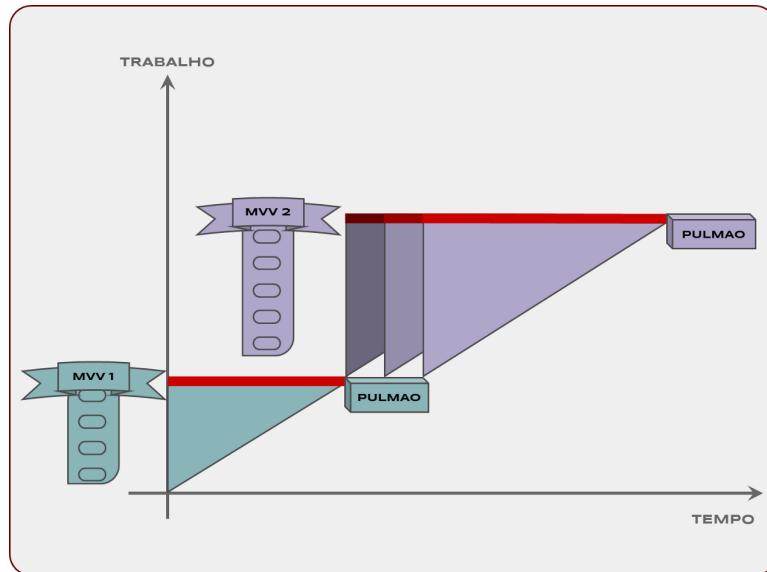


Um pulmão de tempo é associado a cada MVV. Cada pulmão é dimensionado como 50% do tempo médio de fluxo esperado para a MVV correspondente.

Naturalmente, porque cada MVV varia em tamanho, cada uma terá um pulmão de tamanho diferente. Isto também significa que nós estamos realmente esperando que o tempo de início do trabalho nas MVVs subsequentes varie, dependendo se terminarmos uma MVV mais cedo ou mais tarde.

Observe que este pulmão extra *não* é um tipo de *enchimento* que adicionamos “só pra garantir” se algo der errado. Reinertsen [REINERTSEN-2009¹³] refere-se a isso como o princípio “V11: O Princípio do Pulmão: pulmões trocam dinheiro por redução da variabilidade”. É típico do CPM, onde o “enchimento” (explícito ou não) é usado para “comprar segurança*.

¹³<https://tameflow.com/bibliography/#REINERTSEN-2009>

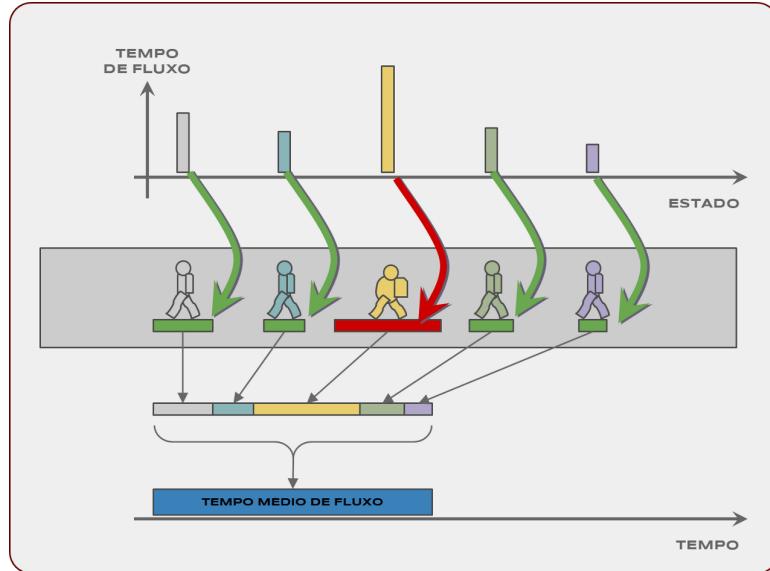


Incluir um pulmão ao final do tempo (médio) de fluxo esperado de uma MVV protege datas de entrega confiáveis, absorve incerteza (risco) e fornece uma ferramenta de relatório para os interessados.

Na GPCC (e em nossa abordagem) o pulmão não está lá para esse propósito, embora seja geralmente apresentado assim e possa até ser (ab)usado dessa maneira. O principal propósito do pulmão MVV é fornecer uma ferramenta de monitoramento operacional que dê sinais adiantados significativos da materialização de riscos vindouros. A Corrente Crítica é a restrição e deve ser explorada ao máximo; o objetivo num projeto GPCC ou em nossa abordagem é entregar qualquer item de trabalho *o mais rápido possível*. O pulmão não é usado para *reduzir a variabilidade*, como no caso do enchimento, mas para *detectar variabilidade desfavorável* cedo o bastante para sermos capazes de tomar quaisquer medidas corretivas a tempo, antes que ela afete negativamente todo o projeto. O pulmão é dimensionado de uma forma intencional; e, o mais importante, ele é *usado ativamente* para guiar a execução do projeto e obter *sinais* de problemas futuros. Para dimensionar o pulmão nós usamos a heurística simples de dividir o tempo de fluxo esperado ao meio. O motivo pelo qual isso faz sentido é explicado pela GPCC e nós examinaremos a lógica disso na próxima seção.

Esse pulmão também é importante sob a perspectiva da hiperprodutividade: não se trata apenas de absorver a incerteza ou gerenciar riscos, mas também de construir uma Comunidade de Confiança ao tornar a entrega de projetos mais confiável.

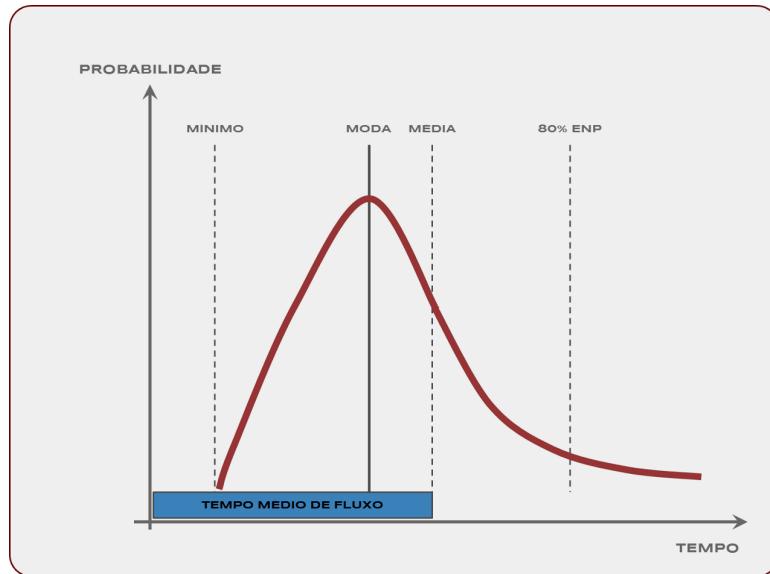
1.5 Dimensionamento do Pulmão



Considere o tempo médio de fluxo para um item de trabalho como a soma dos tempos (médios) de fluxo de cada estado de trabalho. (Faça isso para cada classe de serviço.)

Nós devemos escolher o tamanho do pulmão com cuidado. Para começar, considere o tempo médio de fluxo esperado, a partir de métricas históricas de fluxo. Conceitualmente, dimensionar o pulmão MVV é como ir na direção contrária ao que acontece na Gestão de Projetos pela Corrente Crítica. Na GPCC nós cortamos ao meio a duração da Corrente Crítica para obter uma probabilidade de 50% de entrega no prazo, e então adicionamos o pulmão do projeto ao seu final.

No caso do Pulmão MVV, nós começamos com os dados históricos dos tempos médios de fluxo. Vamos considerar um único item de trabalho que pode estar incluído na MVV. O item de trabalho pertencerá a uma pré-determinada classe de serviço e, consequentemente, terá um tempo médio de fluxo esperado (o tempo médio de fluxo associado com aquela classe de serviço).

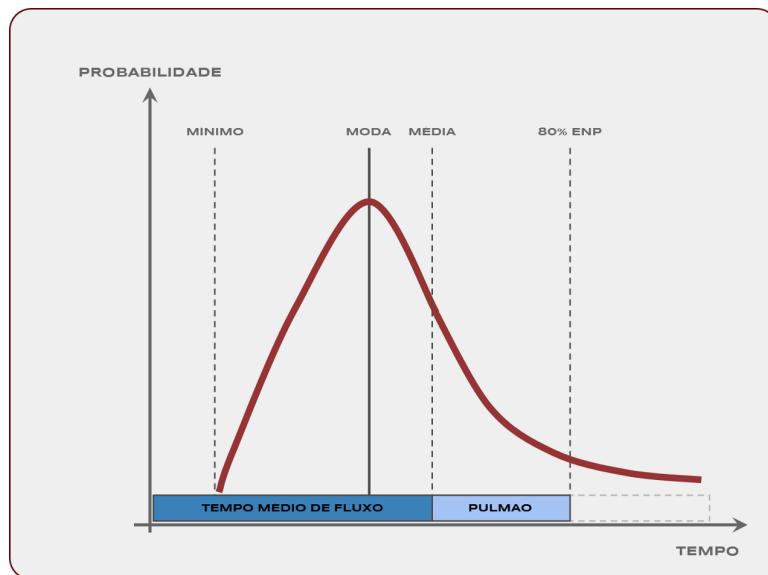


Distribuição da probabilidade do tempo de fluxo. Observe que ela se inclina para a esquerda. A barra azul é o tempo médio de fluxo.

A distribuição da probabilidade dos tempos de fluxo é geralmente distorcida (como ilustrado na figura). O tempo “médio” não é o tempo mais provável. Em muitos casos vamos precisar de menos do que o tempo médio de fluxo para entregar qualquer item de trabalho. Nós também teremos alguns poucos itens de trabalho que levarão “muito mais tempo” que a média. É por isso que temos uma distribuição enviesada.

Outra consequência do fato da distribuição ser distorcida é que teremos uma maior probabilidade de entregar no prazo quanto mais progredirmos ao longo do eixo do tempo. A ideia é adicionar um pulmão ao tempo médio de fluxo para atingir um ponto no tempo onde temos um excelente desempenho em relação à data de entrega.¹⁴

¹⁴N.T.: Due Date Performance



O pulmão é colocado após o tempo médio de fluxo e é dimensionado para 50% dele. Observe que o final do pulmão coincide muito provavelmente com a zona de mais alto desempenho em relação à data de entrega (80% ou mais).

Como uma primeira aproximação nós podemos dimensionar o pulmão como a metade desse tempo médio de fluxo esperado. Ao adicionarmos metade do tempo médio de fluxo é muito provável que já estejamos alcançando as áreas de probabilidade muito alta para a entrega no prazo, na ordem de 80% ou mais.

Nós veremos como esse pulmão (como o pulmão no Hiper-Kanban) será usado para dar sinais significativos de direcionamento. Pode ser vantajoso experimentar com diferentes técnicas de dimensionamento para o pulmão MVV, porque o tamanho do pulmão é crítico para a eficácia dos sinais.

Se o pulmão for muito grande, os sinais virão muito tarde. Se for muito pequeno, os sinais ocorrerão muito frequentemente, com muitos “falsos positivos”. Eventualmente pode-se usar as técnicas avançadas de dimensionamento dinâmico do pulmão, da Teoria das Restrições, para conseguir o tamanho ótimo para o pulmão, com relação às características únicas do nosso projeto. Determinar o tamanho do pulmão é de importância crítica, e ele pode ser dimensionado de várias maneiras.

Por exemplo: [GEEKIE-2006¹⁵] revê sete métodos comuns. [FALLAH-2010¹⁶] usa a incerteza

¹⁵<https://tameflow.com/bibliography/#GEEKIE-2006>

¹⁶<https://tameflow.com/bibliography/#FALLAH-2010>

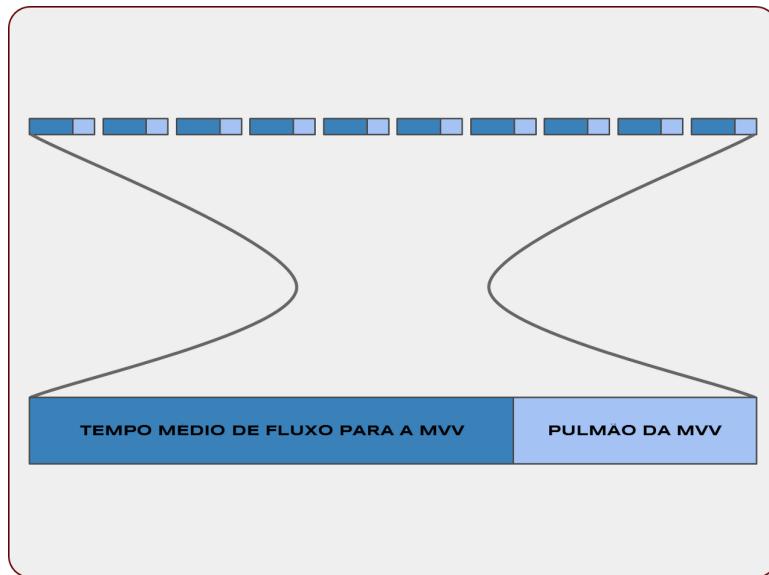
para dimensionar o pulmão. [COX-2010¹⁷] descreve o *dimensionamento dinâmico do pulmão*, onde o tamanho é alterado dinamicamente durante o devido curso do projeto.

Embora conhecer sobre as técnicas avançadas para dimensionamento do pulmão seja útil, é também fora do escopo da (e pouco importante para a) apresentação do conceito aqui. O que realmente importa é a abordagem que o pulmão *possibilita* em termos de *gestão de riscos*.

Em termos mais práticos, o pulmão deve ser “*dimensionado apropriadamente*” — O “*corte ao meio*” mencionado acima é apenas uma simplificação grosseira, usada principalmente como uma heurística simplificadora, que funciona surpreendentemente bem na maioria dos casos.

Dimensionamento Agressivo do Pulmão com o Hiper-Kanban. Ao se usar o Hiper-Kanban nós já estaremos monitorando os tempos médios de fluxo de todos os estados de trabalho, porque os usamos para encontrar “Herbie” (a restrição) no fluxo de trabalho. Neste caso, em vez de considerar o tempo médio de fluxo de cada fluxo de trabalho, nós consideraremos o tempo médio de fluxo *na restrição*. Isso resultará num pulmão muito menor; e pode ser muito pequeno para ser funcional. Contudo, isso nos dá uma boa heurística. Podemos tomar as metades de ambos os tempos de fluxo e considerá-las como os pontos extremos para determinar nosso tamanho de pulmão. Em outras palavras, podemos encontrar um bom tamanho de pulmão em algum lugar entre esses extremos.

¹⁷<https://tameflow.com/bibliography/#COX-2010>



O tempo médio de fluxo do MVV e seu pulmão são a agregação de todos os tempos médios de fluxo dos itens de trabalho e seus pulmões.

Uma vez que os tempos médios de fluxo e pulmões de todos os itens de trabalho na MVV forem determinados eles são agregados para encontrar o tempo de fluxo e o pulmão correspondentes para a MVV.

Lembre-se que os itens de trabalho podem ser de tamanhos diferentes e pertencer a classes de serviço diferentes; todos eles podem estar associados com seus próprios tempos médios de fluxo. Nós precisamos levar isso em consideração ao dimensionarmos o pulmão.

(Observe, porém, que, para efeitos de ilustração, a figura mostra itens de trabalho pertencentes a apenas uma e a mesma classe de serviço e, portanto, todos mostram os mesmos tempos de fluxo!)

Novamente, é importante enfatizar que esse pulmão não é criado para o propósito de “inchar” o prazo de entrega. Ele é uma ferramenta operacional que será usada durante a execução. A seguir veremos como usar e interpretar o pulmão para esse propósito.

1.6 Gestão, Uso e Interpretação do Pulmão

As técnicas de **gestão do pulmão** que iremos explorar agora são uma parte integrante da TOC em geral e da GPCC em particular (no contexto da gestão de projetos). O pulmão do

projeto é um instrumento de segurança. Ele protege o projeto (isto é, a implementação de uma MVV em nosso caso) de interrupções que possam acontecer quando as atividades de trabalho (na Corrente Crítica/na MVV) são realizadas.

O pulmão é *constantemente monitorado* para detectar variabilidade desfavorável.

Um pulmão dimensionado apropriadamente dará bons *indicadores adiantados* de problemas futuros. Assim como o pulmão TPC no quadro Kanban dá sinais sobre problemas que estão acontecendo em estados anteriores a Herbie *antes* que eles afetem o próprio trabalho de Herbie, os sinais do pulmão de tempo da MVV são pistas confiáveis de que problemas estão acontecendo *antes* que eles realmente impactem negativamente na confiabilidade da entrega.

O pulmão dá um sinal *enquanto ainda há tempo* para recuperação. É claro, deve-se também agir e aproveitar a oportunidade para fazê-lo.

Taxa de Consumo do Pulmão

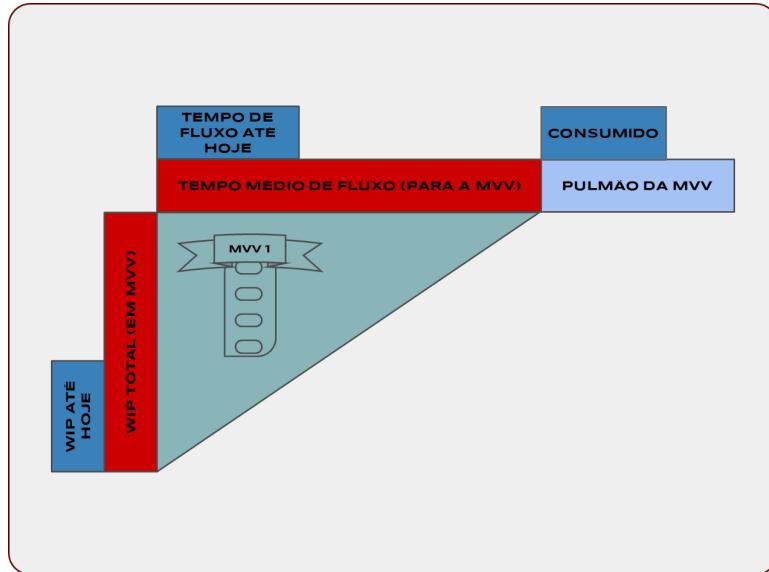
A gestão do pulmão desempenha um papel crítico e a chave é o conceito da **Taxa de Consumo do Pulmão**¹⁸ (também conhecida como *Consumo do Pulmão*). Ela é definida por [SULLIVAN-2012¹⁹] assim:

A taxa na qual o pulmão do projeto está sendo consumido [...] A taxa é calculada como a razão entre a porcentagem de penetração no pulmão do projeto e a porcentagem de conclusão da corrente crítica.

Para nosso propósito, a conclusão da corrente crítica corresponde à quantidade de itens de trabalho na MVV que foi completada. A porcentagem do consumo do pulmão pode ser calculada em virtude da Lei de Little. Podemos determinar quando a MVV será entregue usando o ganho atual e então comparar esse ponto no tempo aos pontos de início e fim do pulmão. (Um exemplo disso será dado mais tarde.) Um alto consumo do pulmão em relação à conclusão é um sinal certo de que algo está errado.

¹⁸N.T.: *Buffer Burn Rate*

¹⁹<https://tameflow.com/bibliography/#SULLIVAN-2012>



A taxa de consumo do pulmão é a razão entre a porcentagem de penetração no pulmão e a porcentagem de conclusão da MVV.

A figura acima mostra os elementos relacionados a uma MVV e seu respectivo pulmão, que são necessários para calcular a taxa de consumo do pulmão como se segue:

$$\text{GANHO ATUAL} = \text{WIP ATÉ HOJE} / \text{TEMPO DE FLUXO ATÉ HOJE}$$

$$\text{TEMPO DE FLUXO ESPERADO} = \text{WIP TOTAL} / \text{GANHO ATUAL}$$

$$\text{DIFERENÇA NO TEMPO DE FLUXO} = \text{TEMPO DE FLUXO ESPERADO} - \text{TEMPO MÉDIO DE FLUXO}$$

$$\% \text{CONSUMO DO PULMÃO} = \text{DIFERENÇA NO TEMPO DE FLUXO} * 100 / \text{TAMANHO DO PULMÃO MVV}$$

$$\% \text{COMPLETADA DA MVV} = \text{WIP ATÉ HOJE} * 100 / \text{WIP TOTAL}$$

$$\text{TAXA DE CONSUMO DO PULMÃO} = \% \text{CONSUMO DO PULMÃO} / \% \text{COMPLETADA DA MVV}$$

Nota: Os cálculos acima serão mostrados com números reais na seção [Como Construir e Monitorar um Pulmão MVV](#)

Na GPCC o pulmão é usado para coordenar os recursos na Corrente Crítica (e tê-los prontos quando necessário) e priorizar o trabalho. O monitoramento do pulmão é o motivo pelo qual

o pulmão é criado. É a atividade principal para saber se o projeto está saudável. [WOEPPEL-2005²⁰] descreve o papel do pulmão muito claramente:

A execução do projeto é A parte mais importante para se alcançar o sucesso. [...] Monitorar e responder à condição dos pulmões é a chave para isso. Em vez de responder às tarefas individuais, a equipe do projeto responde à condição dos pulmões. [...] A taxa [de consumo do pulmão] nos diz quando um projeto está em perigo de não ser completado no prazo. [...] Ao identificar quais tarefas estão criando a maior taxa de consumo do pulmão, o gerente do projeto sabe em quais tarefas focar nesse instante.

A qualquer tarefa que consome o pulmão é dada a mais alta prioridade. Em nosso caso, já que não estamos lidando com tarefas, mas com itens de trabalho numa MVV, quando o pulmão der sinais nós examinaremos a configuração do quadro Kanban para saber onde intervir e identificar os itens de trabalho que *podem* estar tendo problemas. Se o quadro Kanban não estiver dando quaisquer sinais especiais (isto é, tudo ainda está “no fluxo”), então usaremos as técnicas explicadas posteriormente para encontrar maneiras de neutralizar a *variação por causa comum*.

O monitoramento do pulmão é uma *atividade operacional* realizada durante a execução do projeto. A Corrente Crítica é usada para *planejar* o projeto, mas o pulmão é usado para *gerenciar a execução* do projeto. É por isso que nós dispensamos os métodos de planejamento da Corrente Crítica (nós usamos Kanban ou Hiper-Kanban, no lugar) e tiramos vantagem do pulmão pelo o que ele dá em termos de gestão da execução.

A habilidade de ter um indicador *adiantado* — a proporção de consumo do pulmão — é a mais forte contribuição da GPCC. Este indicador *não relata a quantidade de trabalho completado*. Isso é muito diferente da maioria dos outros métodos para gestão de projetos, que tendem a reportar o status do projeto em termos de trabalho feito (“Nós temos 90% feito!” “Oba! É isso aí!”). Este indicador representa *trabalho feito em relação a quanto tempo foi reservado (o pulmão) para absorver problemas imprevistos*. O grau em que esta margem é consumida é uma indicação da saúde do projeto. Consequentemente, há muitas vantagens operacionais, que possibilitam a melhor gestão de riscos. Por exemplo, uma tal vantagem relaciona-se à frequência dos reportes. [LEACH-2004²¹] comenta:

Para que a gestão do pulmão seja plenamente útil, o tempo de monitoramento do pulmão deve ser pelo menos tão frequente quanto a menor duração de tarefa.

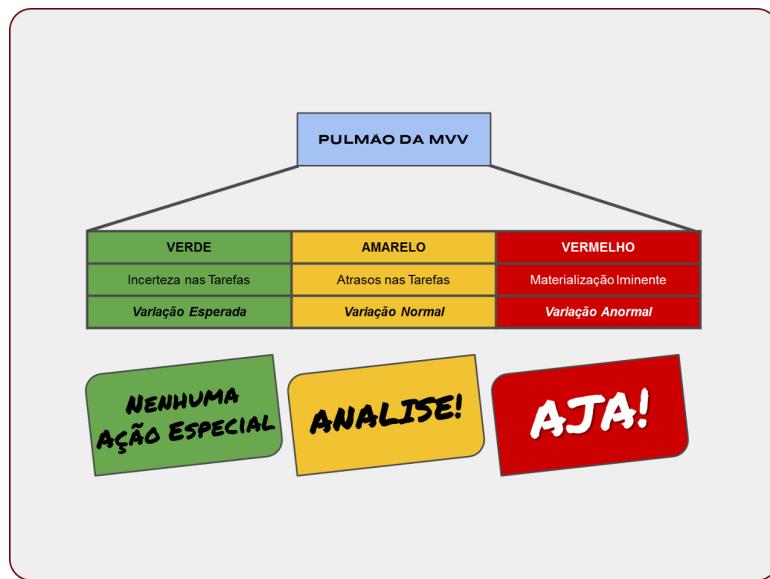
²⁰<https://tameflow.com/bibliography/#WOEPPEL-2005>

²¹<https://tameflow.com/bibliography/#LEACH-2004>

Podemos equacionar a “menor duração de tarefa” ao menor tempo médio de fluxo de qualquer classe de serviço presente na MVV. Se monitorarmos o pulmão nessa frequência, então a frequência de reporte pode ser bem maior. Se a menor tarefa planejada for medida em horas, então o reporte pode ser feito a cada hora — em vez de semanalmente, como é típico na maioria dos ambientes de projeto — dando sinais mais precoces sobre problemas.

Zonas do Pulmão

O pulmão do projeto é dividido em três zonas. Elas são geralmente representadas em Verde, Amarelo e Vermelho. Tipicamente essas zonas são dimensionadas em um terço do pulmão, embora tamanhos relativos possam ser alterados dinamicamente nas aplicações mais avançadas (isso se relaciona a como *dimensionar apropriadamente* um pulmão).



As três zonas do pulmão.

As três zonas dão um controle mais granular para saber quando agir. C. Spoede Budd and J. Cerveny (in [COX-2010²²]) oferecem uma percepção crucial: as três zonas são representantes, respectivamente, de *Variação Esperada*, *Variação Normal* e *Variação Anormal*.

Monitorar o consumo do pulmão com relação às três zonas fornece visibilidade e sinais providenciais:

²²<https://tameflow.com/bibliography/#COX-2010>

- **Variação Esperada** (Zona Verde): Tudo está funcionando “*de acordo com o plano*”. A zona verde absorve a incerteza inerente à tarefa. **Nenhuma ação especial** é necessária. De fato, qualquer interferência nesta zona provavelmente será prejudicial, pois poderia produzir o que [DEMING-1993²³] chama de “violação” (*tampering*): um desperdício de tempo produtivo que causa perda de foco.
- **Variação Normal** (Zona Amarela): Tudo está sob controle, mas precisamos **investigar e preparar para a ação**. A zona amarela absorve a incerteza inerente à previsão da duração da tarefa. O tempo é consumido para cobrir excedentes da tarefa: prepare planos para recuperar o tempo perdido, *mas não tome nenhuma ação ainda* (evite a violação). Foque no entendimento do que está causando o consumo do tempo e o que pode ser feito.
- **Variação Anormal** (Zona Vermelha): Quando a zona vermelha for atingida **nós devemos agir**. Implemente os planos preparados enquanto o consumo do pulmão estava na zona amarela. Muito provavelmente, eventos extraordinários fora do curso normal das operações do projeto causaram o problema.

Causas Comuns e Especiais: A terminologia da TOC pode ser mapeada para a tradicional variação por causa comum e por causa especial. A *Variação Normal* e a *Variação Anormal* podem ser pensadas, respectivamente, como *Variação por Causa Comum* e *Variação por Causa Especial*. A distinção será esclarecida mais tarde. Já que as zonas do pulmão nem sempre representam isso fielmente, nós usaremos outras ferramentas para distinguir entre causas comuns e especiais. O que mais importa é o interesse operacional das três zonas do pulmão.

A divisão do pulmão do projeto em três zonas resulta numa ferramenta poderosa para antecipar e agir sobre os riscos. Outras práticas de gestão de projetos não detectam os problemas até que seja tarde. [LEACH-2004²⁴] conclui:

Através deste mecanismo, a gestão do pulmão fornece uma ferramenta antecipatória excepcional para gestão de projetos, com critérios claros de decisão.

²³<https://tameflow.com/bibliography/#DEMING-1993>

²⁴<https://tameflow.com/bibliography/#LEACH-2004>

1.7 Gráficos do Pulmão

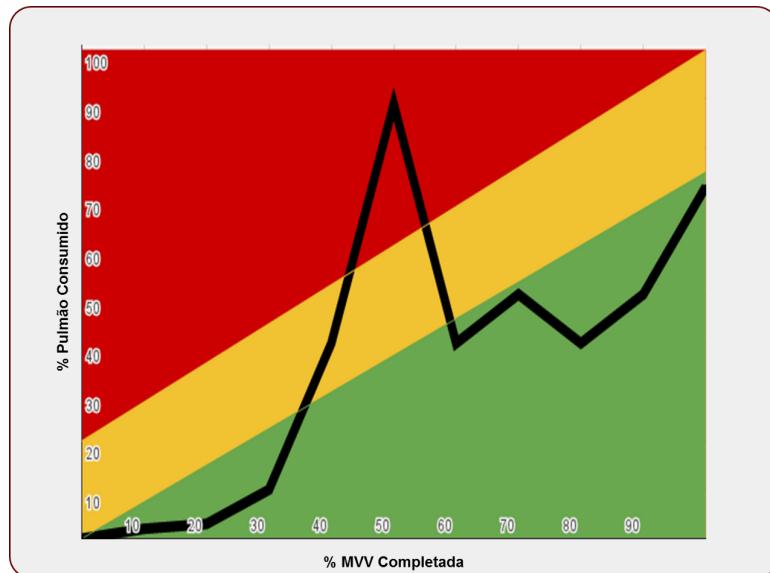
Sempre que tivermos dados quantitativos é natural usarmos diagramas para obter percepções sobre o significado real dos números. A taxa de consumo e a porcentagem de consumo do pulmão podem ser usadas como base para plotar dois diagramas importantes, respectivamente, o gráfico do termômetro do pulmão e o gráfico de controle do pulmão.

Gráficos do Termômetro do Pulmão

Para visualizar o status de um pulmão nós tipicamente usamos um *Gráfico do Termômetro*, onde plotamos o consumo do pulmão (como uma porcentagem) em relação à conclusão do projeto (novamente, como uma porcentagem). Por exemplo, podemos ter um diagrama que parece com o ilustrado na figura.

Nela a linha representa o progresso da execução do projeto. Podemos ver que ela começou na zona verde, progrediu para a zona amarela e depois enfrentou problemas quando penetrou a zona vermelha. Os problemas foram tratados, a execução voltou para a zona verde, e permaneceu lá até o final do projeto, significando que o prazo de entrega foi aceitável (“verde”).

(Nota: esta figura é apenas ilustrativa. A localização real e a inclinação das duas linhas limites dependem de como dimensionamos o pulmão e, especialmente, de como dimensionamos as três zonas.)



O Gráfico do Termômetro mostra o consumo do pulmão (como uma porcentagem) em relação à conclusão da MVV (também como uma porcentagem).

O gráfico do termômetro do pulmão é excelente para reportar a saúde do projeto aos interessados. Ele também pode ser usado em ambientes multi-projetos, onde o status de todos os projetos podem ser representados num único gráfico.

Gráficos de Controle do Pulmão

Nós podemos conseguir melhores refinamentos via **Gráficos de Controle**²⁵. Com efeito, [LEACH-2004]²⁶ sugere isso:

Plote tendências da utilização do pulmão. A medição do pulmão torna-se, em essência, um gráfico de controle e pode usar regras semelhantes. [...] dados de tendência do pulmão preservam o histórico do tempo dos dados e mostram a tendência do consumo do pulmão em relação ao tempo do projeto.

²⁵http://en.wikipedia.org/wiki/Control_chart

²⁶<https://tameflow.com/bibliography/#LEACH-2004>

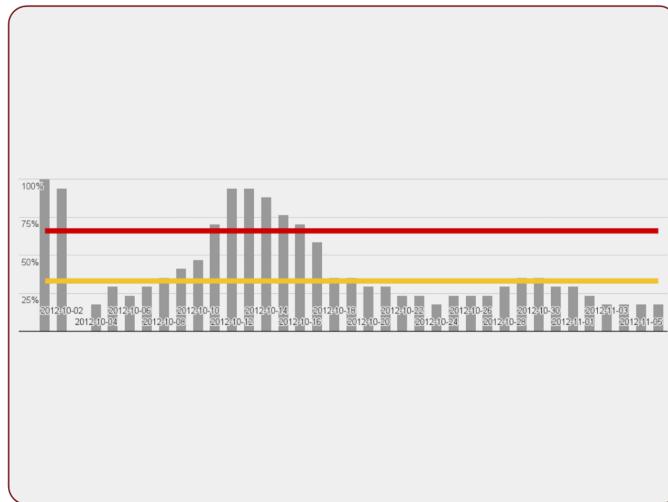


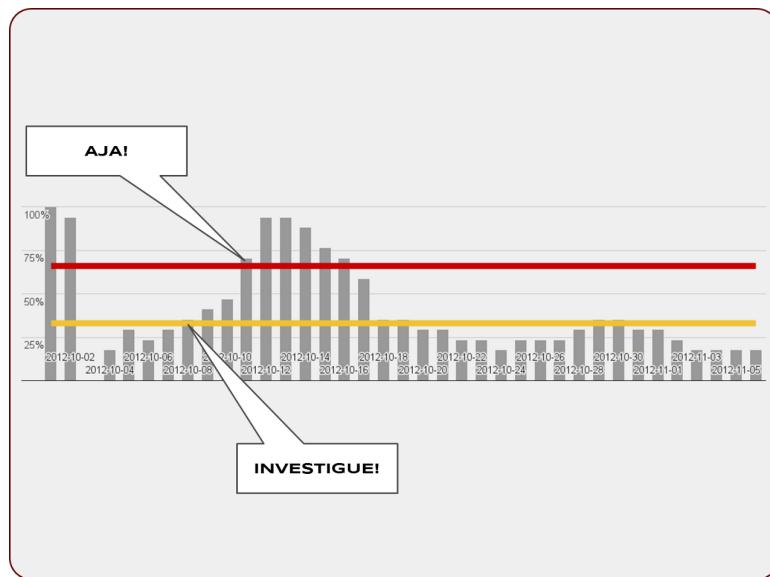
Gráfico de controle do pulmão.

Esse tipo de gráfico certamente ajuda a melhorar o controle (do trabalho em relação ao tempo). Um gráfico de controle dá uma indicação muito clara sobre a saúde do projeto com relação ao tempo (real) no calendário, em vez de em relação ao consumo do pulmão. Ele pode ser usado para saber se o projeto está saudável dia-a-dia, e pode ser usado como base para percepções e reflexões durante as reuniões diárias (como as reuniões em pé) e também nas retrospectivas e *katas*.

Como veremos em breve, pelo fato do gráfico de controle compartilhar a mesma linha do tempo que pode ser usada com os diagramas de fluxo cumulativo, a combinação dos dois é uma visualização poderosa.

Límiates e Sinais

O gráfico é usado como qualquer gráfico de controle, com dois limites de controle nos limites entre as zonas verde-amarela e amarela-vermelha, que são desenhados, respectivamente, com uma linha amarela e uma linha vermelha.



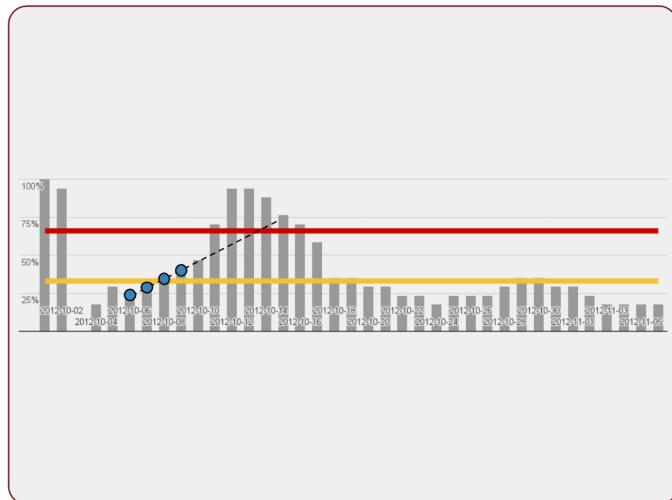
Linhas de limiares de sinal e os sinais no gráfico de controle do pulmão.

A transição através do primeiro limiar (entrando na zona amarela) sinaliza que precisamos começar a investigar sobre potenciais problemas. Nós não devemos agir ainda neste ponto, mas simplesmente começar a pensar sobre o que pode ser um problema. As técnicas de análise de causa raiz que serão apresentadas no próximo capítulo serão muito úteis, pois elas descobrirão até fontes de *variação por causa comum*, e não apenas fontes de *variação por causa especial*.

A transição através do segundo limiar (entrando na zona vermelha) sinaliza que há um problema real e que deve-se agir imediatamente. Ainda há tempo para se recuperar quando este limiar for ultrapassado, mas não há tempo a perder! A ação deve ser rápida e firme. A natureza da ação já deve ter sido decidida enquanto o projeto estava passando pela zona amarela.

Tendências

Podemos refinar o uso do gráfico de controle por meio da Análise de Tendência.



Quatro pontos em sequência revelam uma tendência.

Por exemplo, quatro pontos em sequência tendendo em direção a um limiar podem ser suficientes para se agir. Nós detectamos o problema emergente *ainda mais cedo*. Use seu discernimento para decidir se vai ou não agir sobre tais sinais, mas pelo menos agora temos a possibilidade de ponderar sobre esses sinais antecipados.

Como observado por [SHEWHART-1986²⁷], se nosso processo *não* estiver sob controle estatístico, então o uso de dados de tendência é ainda mais importante.

Processos empíricos raramente estão sob controle estatístico e, como vimos num capítulo anterior, o desenvolvimento de software e o trabalho intelectual podem ser considerados como sistemas adaptativos complexos dinâmicos, que podem ser controlados apenas por processos empíricos. Reconhecer tendências nos diagramas pode ser difícil e requerer experiência, mas é uma ferramenta poderosa que pode fornecer percepções adicionais e especialmente precoces e adiantadas.

Diagramas de Fluxo Cumulativo

Não é coincidência que o Scrum utiliza os gráficos de consumo²⁸ e que o Kanban empregue os diagramas de Fluxo Cumulativo²⁹ como o ilustrado na figura. Ambos os tipos de gráficos

²⁷<https://tameflow.com/bibliography/#SHEWHART-1986>

²⁸N.T.: Burn Down Charts

²⁹N.T.: Cumulative Flow Diagrams

revelam tendências num processo empírico. Todos esses tais gráficos podem ser usados para revelar tendências emergentes e ajudar na tomada de decisão baseada em tais tendências.

Com as técnicas de gestão do pulmão da Teoria das Restrições (e as demais técnicas que exploraremos no próximo capítulo), tais decisões podem ser tomadas com uma percepção melhor sobre a natureza dos problemas que podemos estar enfrentando, especialmente com relação à *variação por causa comum* ou à *variação por causa especial*.

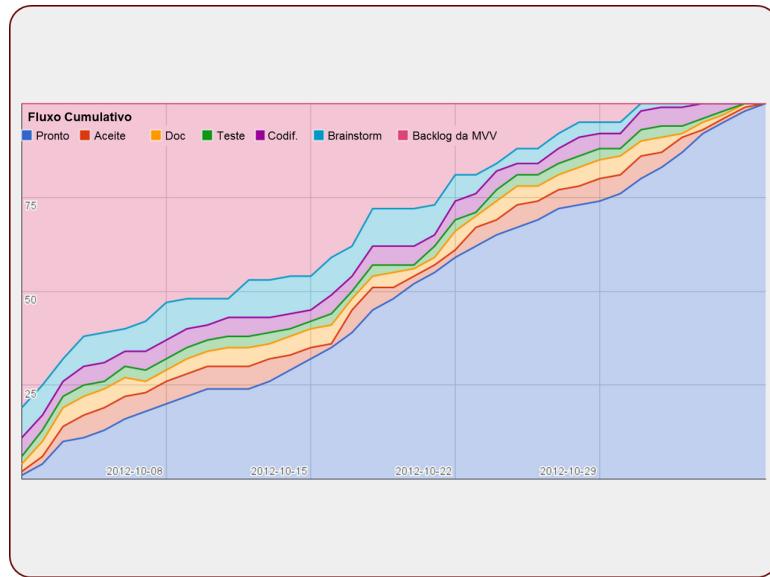


Diagrama de fluxo cumulativo.

Os Diagramas de Fluxo Cumulativo contêm muitas informações³⁰ e fornecem pistas visuais sobre a carga de WIP através de cada estado de trabalho. A formação de filas e inanição são fáceis de observar. O problema, entretanto, é que tais sinais geralmente chegam *tarde*; eles são uma indicação de que o risco já se materializou.

Os sinais do pulmão são indicadores *adiantados* e sua sinalização pode ser reforçada se houver uma tendência se desenvolvendo no diagrama de fluxo cumulativo. Portanto, os melhores resultados podem ser alcançados quando o Diagrama de Fluxo Cumulativo é combinado com um Gráfico de Controle do Pulmão, como descrito a seguir.

³⁰Veja o excelente artigo no blog do Paweł Brodzinski [Cumulative Flow Diagram](#).

Combinando Diagramas e Gráficos

Uma representação particularmente poderosa do progresso do projeto é a combinação de um Diagrama de Fluxo Cumulativo com um Gráfico de Controle do Pulmão. Ambos compartilham o mesmo eixo horizontal, a linha do tempo.

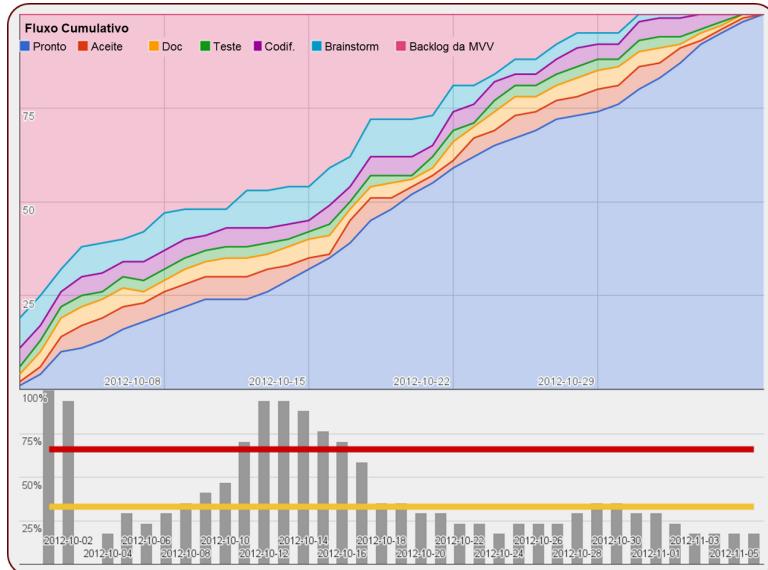


Diagrama de fluxo cumulativo combinado com um gráfico de controle do pulmão. Observe que eles compartilham a mesma linha do tempo (o eixo X).

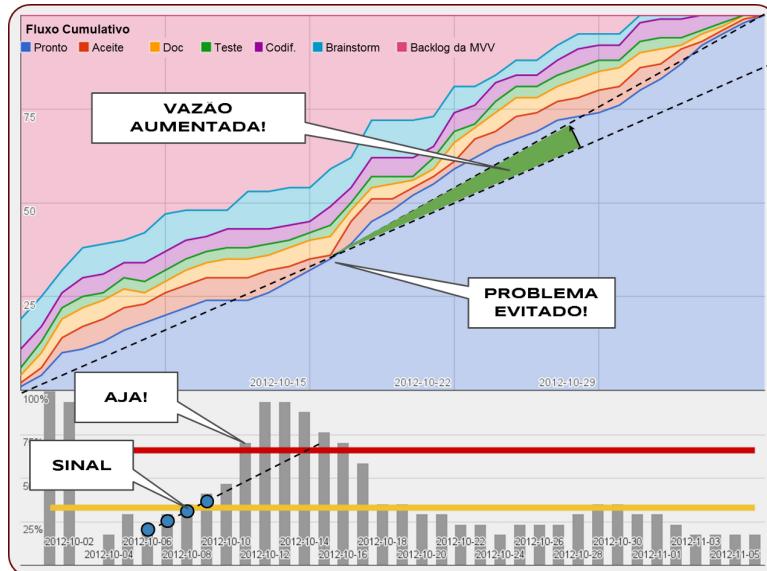
Com esta combinação os *sinais adiantados* fornecidos pelos pulmões tornam-se bem aparentes. Na figura podemos ver esta combinação. Os dois diagramas representam a mesma situação, na mesma linha do tempo. As tendências se desenvolvendo no diagrama de fluxo cumulativo podem ser usadas para apontar aonde os problemas estão e para confirmar os sinais antecipados do gráfico de controle do pulmão.

Por exemplo, na figura podemos ver que quando o gráfico de controle sinaliza que o consumo do pulmão entrou na zona vermelha, o WIP no estado “Aprovação” está gradualmente ficando cada vez menor: está tendendo à inanição.

Sem o sinal adiantado do gráfico de controle do pulmão esta tendência no diagrama de fluxo cumulativo passaria despercebida até muito mais tarde. É evidente como o diagrama de fluxo cumulativo revela um problema que é resolvido mais ou menos na metade da linha do tempo. Todavia, se observarmos o gráfico de controle correspondente, vemos que um sinal amarelo foi levantado sete dias antes, e o sinal vermelho, quatro dias antes. Seis dias antes o gráfico

de controle exibiu quatro pontos em sequência, revelando uma tendência em direção à zona vermelha.

Em outras palavras, comparado à situação sem o gráfico do pulmão, neste caso temos entre quatro e seis dias de aviso antecipado de problemas a caminho.



Evidência de sinais antecipados. Os sinais para investigar, com a linha de tendência, e o sinal para agir antecipam problemas que se tornarão visíveis no diagrama de fluxo cumulativo bem mais tarde.

O efeito da intervenção que foi feita em virtude desses sinais antecipados também pode ser visto no diagrama de fluxo cumulativo, onde um notável crescimento do ganho é evidente após aquele contratempo no ponto médio.

Uma regra prática geral é que nós não precisamos examinar o diagrama de fluxo cumulativo a menos que estejamos recebendo sinais do gráfico de controle do pulmão. Tão logo tais sinais apareçam, procure por qualquer tendência que você vir se desenvolvendo no diagrama de fluxo cumulativo para entender qual(is) estado(s) de trabalho podem esconder algum problema.

Reação ao Sinal Tratada por Políticas Normais do Kanban

Sempre que o pulmão MVV levantar um sinal nós devemos identificar a Causa Raiz (veremos como fazer isso no próximo capítulo) que é a fonte do problema. Se a resolução exigir ação da equipe, então crie um cartão de Problema/Impedimento correspondente. Já que estamos

mapeando o fluxo de trabalho com o Kanban, devemos atribuir uma classe de serviço ao cartão de impedimento. Consequentemente, o cartão de impedimento será tratado através das políticas normais de classe de serviço do Kanban.

1.8 Como Construir e Monitorar um Pulmão MVV

Nós já vimos que dimensionar um pulmão para uma MVV é simples. Nós adicionamos os tempos de fluxo de todos os itens de trabalho e tomamos a metade disso. O que pode não estar muito claro, neste ponto, é como usar o pulmão *na prática*.

Para propósitos de ilustração, vamos assumir que temos 100 itens de trabalho numa MVV e que os tempos médios de fluxo somam 34 dias. Em outras palavras, nós esperamos entregar os 100 itens de trabalho em 34 dias, com um nível de confiança de 50%.

Nós dividimos 34 por dois e obtemos um pulmão de 17 dias. Assim, por exemplo, se o projeto começa no dia 1º de Outubro, a data final esperada seria 34 dias depois, no dia 4 de Novembro. (Nesta ilustração não consideramos finais de semana ou feriados, embora, obviamente, num caso real todas as datas precisariam ser ajustadas apropriadamente.) O pulmão se estenderia por outros 17 dias, de 5 a 21 de Novembro.

Vamos examinar a situação no dia 7 de Outubro. Suponha que 18 itens de trabalho tenham sido entregues. A porcentagem de trabalho completado é 18% (nós tínhamos 100 itens no início).

Para entender se estamos num estado saudável ou não precisamos descobrir o status do pulmão. É aqui onde invocamos a Lei de Little. Nós sabemos que o tempo decorrido é de 7 dias. Então o ganho real até esta data é a entrega de 18 itens dividido pelo tempo de fluxo de 7 dias, resultando em 2,57 itens de trabalho por dia.

Nós agora assumimos que este é o desempenho *sustentável* do ganho da equipe, e que ele será mantido até a entrega total. Vamos ver quanto tempo seria necessário, com esse ganho atual, para entregar todos os 100 itens de trabalho na MVV. Novamente invocamos a Lei de Little. O tempo necessário é o WIP original de 100 dividido pelo ganho atual de 2,57, que nos dá 39 dias (arredondando para dias inteiros).

Os 39 dias esperados são 5 dias a mais do que os 34 estabelecidos através do nosso tempo de fluxo histórico. Se o ganho atual permanecer inalterado nós consumiríamos 5 dias do nosso pulmão de 17 dias. A penetração no pulmão é 5 / 17 ou 29%. Isso é menor que 33%, então estamos na *zona verde* do pulmão. Não há nada com o que se preocupar.

Suponhamos agora que no dia seguinte, 8 de Outubro, dois itens de trabalho foram entregues, totalizando 20. O tempo decorrido é de 8 dias. O ganho atual é de 20 itens dividido por 8 dias, ou 2,5 itens por dia. O tempo estimado para completar os 100 itens de trabalho na MVV é o WIP original de 100 itens dividido pelo último ganho atual de 2,5, ou 40 dias. Agora, 40 dias são 6 dias a mais do que os 34 dias do nosso histórico tempo médio de fluxo. A penetração no pulmão é de $6 / 17$ ou 35%. Isso fica acima do limiar de 33%. Nós acabamos de passar para a *zona amarela* do pulmão. Precisamos investigar o que pode estar afetando negativamente o desempenho do ganho.

Nós ilustramos duas vezes o cálculo básico. Aproveite a Lei de Little e calcule a data esperada de entrega baseada no WIP original e no ganho real medido. Olhe aonde a data cai com relação ao pulmão, e se ela for posterior à data de início do pulmão, calcule a porcentagem de penetração. Uma vez que se tenha uma planilha de cálculo pronta com as fórmulas não é mais difícil do que fornecer a quantidade total de WIP entregue até a data de hoje.

Os números usados nesse exemplo são os mesmos que podem ser vistos nas figuras com o diagrama de fluxo cumulativo e o gráfico de controle do pulmão, apresentadas acima.

A Lei de Little e a Premissa do Fluxo em Estado Estável/Ideal

Naturalmente, o que estamos fazendo é confiar pesadamente na lei de Little para fazer tanto o dimensionamento do pulmão quanto os cálculos subsequentes significativos. A lei de Little é relevante apenas quando o sistema estiver num estado estável, e isso significa que a entrada (média) está equilibrada com a saída (média).

Nós vimos como podemos alcançar o equilíbrio dinâmico através do Tambor-Pulmão-Corda, fichas Kanban, e fichas de Reposição usadas no Hiper-Kanban. A abordagem do Hiper-Kanban é mais estável do que a original usada no Kanban, com limites de WIP nos estados de trabalho. Isto pode ser visto literalmente ao compararmos os diagramas de fluxo cumulativo dos projetos Kanban com aqueles dos projetos Hiper-Kanban. Os projetos Kanban apresentam um diagrama de fluxo cumulativo que é mais dentado, com muitas paradas e continuações, planícies e picos; essas são todas consequências da natureza dos limites de WIP nos estados de trabalho. Um diagrama de fluxo cumulativo de um projeto Hiper-Kanban (como o mostrado acima) é mais suave, com acelerações/desacelerações mais lentas de fluxos, que são detectadas e atuadas mais cedo, tanto por causa do mecanismo TPC e ainda mais por causa das técnicas de gestão do pulmão MVV apresentadas aqui.

Como já mencionamos várias vezes, os limites de WIP nos estados de trabalho introduzem interrupções artificiais no fluxo de trabalho. A interrupção artificial do fluxo torna a aplicação da Lei de Little menos confiável. A implicação é que, mesmo que as técnicas do

pulmão MVV funcionem muito bem com o Kanban (de fato, ela foi primeiramente concebida com o Kanban, antes que o Hiper-Kanban surgisse), elas funcionam muito melhor com o Hiper-Kanban, porque ele se empenha em interferir o mínimo possível na Lei de Little.

A Lei de Little e as Condições para o Ritmo Máximo Sustentável

O método funciona mesmo quando a equipe estiver trabalhando no seu ritmo *máximo sustentável*. Observe que um estado estável (isto é, quando a Lei de Little é aplicável) pode ser alcançado mesmo num nível de ganho que é *menor* do que o máximo sustentável. Isso acontece, tipicamente, quando o fluxo é equilibrado com o ganho médio exibido por Herbie, em vez de com a capacidade real de Herbie, como mostrado na seção [A Miragem do Equilíbrio do Fluxo](#) no capítulo [O Método Kanban](#). É muito fácil cair nesta armadilha quando se busca “equilibrar o fluxo”. E é uma situação que o mecanismo tambor-pulmão-corda evitará.

Mesmo se nós, muito provavelmente, não estivermos chegando nem perto do ritmo máximo sustentável da equipe com o Kanban (embora nós estaríamos com o Hiper-Kanban), a peculiaridade a ser observada é que quando a equipe estiver no seu ritmo *máximo* sustentável então não seria possível aplicar a estratégia normal do Kanban para aumentar os limites de WIP para obter mais ganho. Isso iria efetivamente introduzir uma interrupção de variação por causa especial no fluxo de trabalho. Herbie iria tropeçar e cair — por assim dizer — e pararia o progresso da tropa toda.

Como não queremos perder *sustentabilidade*, a única maneira de melhorar o ganho é diminuir o tempo de fluxo — e, como aprendemos nos capítulos anteriores, tal redução no tempo de fluxo deve envolver Herbie, a restrição.

Este é um ponto importante. Não podemos aumentar o ganho indefinidamente aumentando o WIP. Chegará num ponto quando o aumento nos limites de WIP tornará insustentável a habilidade da equipe de lidar com a carga adicional de trabalho. É neste ponto que qualquer tentativa de aumentar o ganho pelo aumento dos limites de WIP falhará. Tais tentativas serão uma “violação” com os limites de WIP e uma “derrota” para o sistema.

A falha certamente aparecerá no Quadro Kanban, mas a aplicação dos Cinco Passos para Focalização revelará apenas o motivo de causa especial pelo qual o ritmo maior não poderia ser sustentado (WIP adicional). Não será revelado o motivo inerente pelo qual a tropa não pode andar mais rápido, apesar de Herbie ainda continuar como a restrição geral (porque a capacidade está limitada). Descobrir que estamos indo mais devagar *porque* aumentamos a carga não nos dá qualquer informação a mais. Nós mesmos adicionamos a carga extra. Em

vez disso, descobrir porque a capacidade está limitada pode produzir frutos bem melhores. Se pudermos aumentar a capacidade, *então* seremos capazes de aumentar a carga também.

É como um atleta praticando esportes de resistência. Um aumento no limite de WIP irá apenas resultar numa *corrida de curta distância*³¹ (que os leitores praticantes de Scrum perdoem o trocadilho aqui!), que fará todos ficarem sem fôlego após um breve período. Mas nós queremos correr maratonas, não corridas de curta distância. Queremos manter um ritmo sustentável pelo máximo de tempo possível, e aumentar o ritmo sem cair ao chão sem fôlego.

Quando estivermos no estado *máximo sustentável*, os limites de WIP devem ser mantidos intocados, para *manter* esse nível de desempenho sustentável. É nesta situação que usaremos a gestão do pulmão para revelar o que está impedindo a equipe de ser ainda mais produtiva. A melhoria que resultará irá afetar o processo inteiro. A melhoria não apenas removerá um impedimento temporário (e talvez artificial) do projeto ou fluxo de trabalho atual. A melhoria será benéfica para todos os *futuros* projetos também, porque ela será uma melhoria permanente para o *processo*.

O foco de qualquer tentativa de melhoria precisa mudar dos recursos individuais para a totalidade da equipe, vista como um sistema. Nós queremos descobrir o que impede a equipe de aumentar seu ritmo máximo sustentável, enquanto mantemos o processo em seu estado ideal/sustentável! É isso o que significa melhorar *durante o fluxo!* Mas para conseguir isso precisamos caçar as fontes da variação por causa comum. Reagir apenas à variação por causa especial não nos permitirá fazer tais melhorias fantásticas. A natureza desses aperfeiçoamentos são verdadeiramente a quebra de restrições sistêmicas, e elas são alavancas Arquimedianas que aumentarão o desempenho para níveis muito mais altos.

³¹N.T.: *Sprint*