




# INTERNET DAS COISAS NAS NUVENS

por Tito Ocampos

.....  
A COMPUTAÇÃO EM NUVEM É UM DOS PRINCIPAIS MEIOS DE  
SERVIÇO, INFRAESTRUTURA, PLATAFORMA DE SOFTWARE E  
ANÁLISE DE DADOS DA INTERNET DAS COISAS.  
.....





COMPUTAÇÃO EM NUVEM é uma das tecnologias mais importantes do mundo dos negócios atualmente. Com um mercado global que deve alcançar 120 bilhões de dólares em 2018 (Forbes, 2015), Computação em Nuvem e Internet das Coisas caminham juntas para estabelecer um novo cenário de tecnologia mundial. A Internet das Coisas demanda Computação em Nuvem em diversos níveis de serviço, incluindo infraestrutura, plataforma, software e análise de dados. Em um mundo com bilhões de dispositivos gerando dados, serão necessários serviços escaláveis, robustos e de alta disponibilidade para armazenar, processar, personalizar e entregar informações de alto valor agregado para os clientes, a qualquer momento, em qualquer lugar.

Infraestrutura como Serviço (*Infrastructure as a Service - IaaS*) deverá ir muito além da provisão de computação e de armazenamento em servidores. Toda a infraestrutura de Internet das Coisas pode ser virtualizada, criando acesso onipresente a todos os dispositivos através de uma única interface. Você quer testar um software que recupera dados de mil sensores de temperatura e exibe a informação resultante em um *smartwatch* de um determinado fabricante? Virtualize! Infraestruturas de cidades inteligentes inteiras poderão ser virtualizadas, criando um mercado inimaginável para plataformas, software e análise de dados. Em 2015, o mercado global de IaaS deve alcançar 16,5 bilhões de dólares (Gartner, 2015).

No nível de Plataforma como Serviço (*Platform as a Service - PaaS*), interoperabilidade é a chave. PaaS de propósitos específicos tornarão o acesso a informações de áreas diversas, tais como saúde, governo, indústria de manufatura, independentes dos tipos de dispositivos, protocolos utilizados e seus fabricantes, criando uma camada de abstração para a implantação de Internet das Coisas em ambientes heterogêneos. Imagine comprar sensores diversos, implantar em sua fábrica e direcionar os dados para um endereço em nuvem. Pronto! A partir daí, seu software

de apoio à tomada de decisão já pode se beneficiar das informações e serviços da sua PaaS. O mercado global de PaaS deve alcançar 14 bilhões de dólares em 2017 (IDC, 2013).

Em uma época de dispositivos clientes cada vez mais leves e versáteis, Software como Serviço (*Software as a Service – SaaS*) se torna um diferencial. Novas tecnologias que permitam o desenvolvimento de SaaS de forma integrada a dispositivos portáteis e vestíveis serão criadas, de forma cada vez mais transparente e eficiente para os desenvolvedores. Segundo estimativas, *Cloud Apps* serão responsáveis por 90% do total de tráfego de dados móveis em 2019 (Forbes, 2015).



Análise de dados como serviço (*Analytics as a Service – AaaS*) completa o pacote de serviços indispensáveis à Internet das Coisas que serão providos pela computação em nuvem. Com a enorme quantidade de dados armazenados, analisá-los é uma atividade que demanda muitos recursos, o que tem tudo a ver com computação em nuvem.

No exemplo da fábrica, até mesmo a geração dos gráficos e relatórios estatísticos sobre os dados de produção coletados pelos sensores seria provida pela nuvem. O mercado estimado de AaaS para 2020 é de mais de 23 milhões de dólares (marketsandmarkets.com, 2015), mesmo considerando que é um modelo de negócio recente e ainda em maturação.

Mas a computação em nuvem para a Internet das Coisas também traz muitos desafios. Como sempre, segurança e privacidade estão entre os principais. Com os mais de 1,6 zettabytes de dados trafegados globalmente em 2015, cerca de 133 exabytes por mês, quanto de informação pode ser gerada? Como garantir o acesso autorizado aos dados, serviços e informações? Qual o poder dos provedores com tanta informação passível de ser processada e cruzada? Pode-se argumentar que atualmente muita informação sensível já é armazenada em servidores ao redor do mundo. Porém, a quantidade de dados e a possibilidade de descoberta de

conhecimento a partir destes dados atingirá uma escala nunca antes vista na História.

Além de privacidade e segurança, infraestruturas de computação em nuvem deverão atingir cada vez mais um nível de interface amigável ao usuário final. Mecanismos de *self-service* serão indispensáveis, assim como gerenciamento e balanceamento de carga, bilhetagem e análise de tráfego e utilização. Usar sistemas de computação em nuvem será tão simples quanto ler um e-mail ou acessar redes sociais.

Uma solução já adotada por muitas empresas é a implantação de nuvens privadas. Estima-se que o mercado de nuvens privadas deverá atingir 69 bilhões de dólares em 2018 (tbri.com, 2015). Soluções baseadas em OpenStack são atualmente as mais utilizadas, como a Fusion Cloud, da Huawei.

Computação em nuvem atualmente já possui seu mercado estabelecido, mas terá na Internet das Coisas um grande aliado para massificação e popularização nos próximos anos. A Internet das Coisas, por outro lado, depende de computação em nuvem para se consolidar, ganhar escala e se tornar uma realidade nos diversos setores da economia, incluindo a indústria e a eletrônica de consumo. O casamento destas duas gigantes pode se tornar a maior (r)evolução tecnológica da história!●



**TITO OCAMPOS** | É bacharel em Informática pelo IM/UFRJ e trabalha há 15 anos em projetos na área de telecomunicações. Desde 2014 atua como gerente do Centro de Pesquisa e Desenvolvimento da Huawei do Brasil. Interconectado com os demais centros de P&D da Huawei no mundo, os principais focos de pesquisa do centro de P&D no Brasil são projetos em Telecomunicações, Internet das Coisas, Computação em Nuvem, Big Data e Redes Ágeis de Computadores.



# BIG DATA: **UM NOVO DESAFIO À NOSSA PORTA**

por Fabio Porto, Artur Ziviani e Eduardo Ogasawara

O PRINCIPAL DESAFIO NA ÁREA DE INTERNET DAS COISAS É A COLETA, TRANSMISSÃO, PROCESSAMENTO E ANÁLISE DE DADOS, JUNTAMENTE COM A EXTRAÇÃO DE INFORMAÇÕES RELEVANTES.



**H**á grandes desafios na Internet das Coisas ligados às expectativas relacionadas ao imenso volume de dispositivos envolvidos e de dados gerados por estes. A coleta, a transmissão, o processamento e a análise de dados em larga escala, bem como a extração de informação relevante nesse contexto, também se apresentam como um grande desafio na área de Internet das Coisas para os próximos anos. Para dimensionar a magnitude destas expectativas, estudos recentes do IDC estimam para 2020 um mercado global ligado à Internet das Coisas de 4 trilhões de dólares, envolvendo a existência de mais de 25 bilhões de dispositivos e sistemas inteligentes conectados gerando mais de 50 trilhões de GBs de dados, um volume que deve ser processado de alguma forma aproximadamente 10 vezes mais do que atualmente presenciamos em 2015. Para refletirmos sobre os desafios no processamento deste grande volume de dados produzidos, precisamos caracterizar sua natureza. Primeiramente, fontes de dados em Internet das Coisas produzem dados em alta frequência e com baixo conteúdo informacional. Por exemplo, podemos citar o monitoramento de pacientes, em um contexto de apoio à saúde, no qual um sistema baseado em sensores captura, continuamente, informações sobre o estado do paciente, tais como: frequência cardíaca, temperatura e pressão, além da identificação do paciente e a sua localização. Um agente de monitoramento recebe essas informações e reage com mensagens a outros objetos, quando o comportamento observado apresenta cenários fora do estado normal. Neste exemplo, podem-se observar vários elementos importantes do contexto de Internet das Coisas. Inicialmente, a identificação da fonte de informação, juntamente com sua localização espaço-temporal, aparece de forma ubíqua com as variáveis monitoradas. Em seguida, ações a serem tomadas a partir das informações recebidas dependem de sua relevância para o receptor. Requer-se então que se possa expressar, de modo declarativo, os cenários de



interesse para tomada de decisão. Ações derivadas da interpretação de dados produzidos por sensores ocorrerão no encontro dos dados monitorados com os critérios de interesse expressos pelos receptores. Deriva-se desta última a necessidade da seleção e armazenamento de dados na Internet das Coisas para avaliação de seu interesse pelos múltiplos agentes de recepção e ação. Dado o grande número de sensores produzindo continuamente informações, bem como o volume de dados por eles gerados, a detecção amostral dos dados aparece como fundamental para que seja viável seu armazenamento e avaliação de registros de interesse. O desafio apresenta-se em conceber uma arquitetura escalável tanto no tratamento de transações de produção de dados quanto na interpretação e ação sobre esses dados, de forma inteiramente distribuída. Um exemplo nesta direção é o sistema S-Store [Meehan et al. 2015], desenvolvido sobre o HBase, que oferece alta escalabilidade e processamento de grande volume de transações em memória.

Esse desafio culmina na demanda pela descoberta de conhecimento a partir desses dados (KDD) [Tsai et al. 2014]. O processo geral de KDD na Internet das Coisas engloba as atividades de seleção, pré-processamento, mineração de dados e análise de dados. Pela perspectiva dos aplicativos, a maior parte das tecnologias em KDD foi projetada para execução em um sistema centralizado. Em função disso, essas não são diretamente capazes de manipular os dados produzidos pela Internet das Coisas. De fato, essas limitações são típicas de cenários de Big Data. Uma das principais ações para superar tais limitações consiste em reduzir a complexidade dos dados. Tal redução ocorre tanto por meio de redução de dimensionalidade quanto pelo seu volume. Considere o exemplo de monitoramento da saúde previamente apresentado. Imagine que um médico deseja identificar se um paciente apresenta uma determinada doença. O conjunto de dados monitorados pode ser muito amplo e há uma necessidade de preparação



dos dados coletados. Para tanto, faz-se uso de análise de componentes principais (PCA), métodos de seleção de atributos e transformação de dados (discretização, alisamento), amostragem e agregação para redução e síntese dos dados. Posteriormente, pode-se fazer uso de diferentes métodos de mineração (classificação, agrupamento, padrões frequentes) para corroborar as hipóteses levantadas pelo médico. Em função da complexidade dos problemas tratados, os cenários típicos de Internet das Coisas comumente demandam uma composição ordenada de diferentes métodos de mineração (classificação, agrupamento, padrões frequentes).

Neste sentido, fica claro que estamos em face de mais uma grande (r)evolução na computação impulsionada pela Internet, e potencializada pela integração de dispositivos à rede. A autonomia na produção de dados retira a restrição da capacidade humana como agente da produção de dados e geração de conhecimento para tomada de decisão. Nesse novo contexto, novos componentes computacionais serão necessários, bem como revisitados serão aqueles hoje tidos como estado da arte. Parece que o termo Big Data ganha uma nova dimensão como uma nova Ciência de Dados.●

---

### Referências Bibliográficas

- Atzori, L., Iera, A., Morabito, G., The Internet of Things: A Survey, *Computer Networks*, vol. 54, pp. 2787-2805, 2010.
- Chen, M., Mao, S., Liu, Y., Big Data: A Survey, *Mobile Network Applications*, vol. 19, pp. 171-209, 2014.
- Meehan, J., Tatbul, N., Zdonick, S., et al., S-Store: Streaming Meets Transaction Processing, *Proc. of Very Large Databases*, vol. 8, no. 13, Hawaii, USA, 2015
- Porto, F., Ziviani, A., Ciência de Dados, III Seminário de Grandes Desafios da Computação no Brasil, Rio de Janeiro, RJ, Setembro de 2014.
- Tsai, C.-W., Lai, C.-F., Chiang, M.-C., Yang, L. T., Data Mining for Internet of Things: A Survey, *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, 2014.





**FABIO PORTO** | É tecnologista do Laboratório Nacional de Computação Científica (LNCC) do MCTI, onde coordena o Laboratório DEXL, no desenvolvimento de pesquisa e inovação na área de gerência e análise de grandes volumes de dados. Obteve o doutorado em Informática em 2001, pela PUC-Rio, com período sanduíche no INRIA, França. Realizou estágio pós-doutorado na EPFL, Suíça, entre 2004 e 2006. É associado da SBC e da ACM, e sócio-fundador da start-up DBS2.



**ARTUR ZIVIANI** | É tecnologista sênior do Laboratório Nacional de Computação Científica (LNCC), onde é um dos coordenadores do grupo de P&D MARTIN (Mecanismos e ARquiteturas de TeleINformática). Possui mestrado em Eng. Elétrica pela COPPE/UFRJ e doutorado em Computação pela Université Paris VI, Sorbonne Universités, França. É associado da SBC, membro afiliado da Academia Brasileira de Ciências (ABC) e membro Sênior da ACM e do IEEE.



**EDUARDO OGASAWARA** | É professor do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca (CEFET/RJ), onde lidera o Grupo de Pesquisa em Ciência de Dados. Tem doutorado em Engenharia de Sistemas e Computação pela COPPE/UFRJ. É membro da SBC, ACM, INNS, IEEE e realiza pesquisas na área de Banco de Dados.



# ANALYTICS: DADOS E ATENÇÃO

por Nazareno Andrade

.....  
O CONCEITO DE ANALYTICS NA INTERNET DAS COISAS  
BASEIA-SE NA CONSTANTE GERAÇÃO DE DADOS.  
.....



## COISAS CRIAM DADOS

**S**ETE DA NOITE, subo na balança, que mede meu peso, massa muscular e se conecta na rede sem fio para salvar as medições na nuvem. Enquanto como algo antes de sair, verifico o histórico de minhas medições no smartphone ao mesmo tempo que ele toca música no som da sala. Configuro com o smartphone quais luzes da casa devem ficar acesas na minha ausência, checo mais uma vez

**A presença desses dados cria a necessidade de novos mecanismos para aproveitá-los de maneira eficaz.**

a lotação e nível de poluição medidos por sensores na praça em que vou correr, e saio me certificando de que meu smartwatch tem bateria para monitorar meu exercício. Este exemplo de Internet das Coisas no cotidiano de alguém interessado em saúde ressalta uma perspectiva fundamental na Internet das Coisas: as coisas nos geram dados; novos dados, dados de mais fontes e muitas

vezes dados em fluxo contínuo. Smartphones atualmente possuem entre cinco e dez sensores que geram dados em formato numérico, de coordenadas geográficas, em áudio, vídeo e fotografias. Dispositivos de smart homes e smart cities cada vez mais estão disponíveis e fornecendo dados em fluxo contínuo.

Comumente, a presença desses dados cria a necessidade de novos mecanismos para aproveitá-los de maneira eficaz. Por exemplo, além de saber a leitura de meu peso atual, há valor em entender qual a tendência desse número no tempo, ou em relação com outras variáveis: estou emagrecendo? Há uma relação entre os dias que realizo mais exercício segundo medido pelo meu smartphone e as medições de peso da minha balança

inteligente? Para responder a estas perguntas, tipicamente precisamos de um modelo e da comunicação dos resultados deste modelo de maneira visual. Precisamos de *analytics*.

## DADOS PEDEM MODELOS

Analytics, análise de dados ou Ciência de Dados são os termos atualmente utilizados para definir a interseção entre a Ciência da Computação, Estatística e Visualização de Dados aplicados. Tipicamente, modelos de Estatística e Mineração de Dados são utilizados em sistemas computacionais para detectar padrões e realizar previsões com base em dados vindos de fontes diversas, como sensores, redes sociais e logs de sistemas online. Tipicamente, comunicar um grande número de valores e variáveis, assim como suas relações e o resultado de modelos de maneira tabular ou textual é muito ineficiente. Nossos cérebros são muito mais eficientes na leitura de formas em imagens do que no processamento de texto e, portanto, a confecção de visualizações é uma etapa essencial da análise de dados.

Concretamente, os modelos mais comuns no contexto de analytics podem ser divididos em cinco tipos. Modelos de *regressão* buscam entender que fatores têm impacto em uma variável de interesse, e possivelmente prever valores futuros para essa variável. Um modelo deste tipo pode, por exemplo, afirmar se o novo tipo de treinamento que passei a fazer no último mês me proporciona maior perda de peso que o anterior, considerados os demais fatores que também variaram entre os dois meses. Pode também estimar que peso terei em duas semanas caso mantenha o treinamento e demais fatores relevantes constantes.

Modelos de *classificação* funcionam de maneira semelhante, porém lidam com entender ou prever uma variável que descreve a que categoria um item pertence. Aplicações recentes de moni-





toramento de atividade física usam modelos de classificação para identificar via acelerômetro se o usuário está andando, correndo ou dentro de um veículo. Aqui, o modo de transporte é a variável de interesse.

Um modelo de *recomendação* busca prever quais itens devem ser oferecidos a um usuário em um dado momento. Por exemplo, ao pedir ao smartphone indicações de locais para visitar, ele

**Ao pedir ao smartphone indicações de locais para visitar, ele poderá usar um modelo de recomendação para combinar informações de minha localização atual.**

poderá usar um modelo de recomendação para combinar informações de minha localização atual, meu histórico de movimentação, preferências e dados climatológicos para recomendar que há um parque próximo que eu provavelmente não conheço.

Um agrupamento é um modelo que identifica padrões de semelhança e diferença entre itens de interesse, criando grupos coesos e distintos entre si. Por exemplo, considerando o histórico de consumo de energia obtido através de smart meters, é possível identificar quais os principais cinco grupos de consumidores segundo seu padrão de consumo.

Por fim, um modelo de *visualização* transforma informações numéricas, de ligações e espaciais em formas e símbolos que permitem que seres humanos compreendam padrões na informação original de maneira rápida e eficiente. Uma visualização de uma linha de tendência de meu peso, da distribuição geográfica dos locais recomendados para que eu visite ou uma representação visual da série temporal de meu consumo de energia são fundamentais ao entendimento dos exemplos dados até agora.

## QUESTÕES PRÁTICAS

No Laboratório Analytics da UFCG, temos trabalhado em várias aplicações na interseção entre Internet das Coisas e a análise de dados. Entre as aplicações, há o uso de smartphones e câmeras para contribuir no monitoramento do transporte público, e plataformas de sensores que nos permitam acessar dados sobre espaços públicos tanto para planejamento quanto para o desenvolvimento de aplicações utilizando, por exemplo, a API da Praça da Bandeira.

Julgamos a partir dessa experiência que há duas questões práticas relacionadas à formação de nossos profissionais na Computação que merecem discussão. A primeira é no tocante à formação em Analytics que permita ao profissional compreender desde aspectos computacionais e de rede, passando pela criação de modelos matemáticos e estatísticos com os dados, até a criação de visualizações de dados combinando habilidades computacionais e de design. Essa formação é eminentemente multidisciplinar e sua relevância merece atenção de nossos cursos de Computação. Em segundo lugar, a necessidade de lidar com a privacidade dos dados coletados e processados aponta a necessidade de discussões sólidas sobre esse tópico em nossas graduações. Atentemos para esses dois pontos, e as oportunidades na união de Internet das Coisas e Analytics serão muitas e frutíferas. ●



**NAZARENO ANDRADE** | É professor do Departamento de Sistemas e Computação da UFCG, onde coordena projetos no Laboratório Analytics e no Laboratório de Sistemas Distribuídos. Atua principalmente nas áreas de Computação Social, Visualização e Mineração de Dados, e Computação e Música. É Tecnólogo em Telemática pelo IFPB e Doutor em Engenharia Elétrica pela UFCG. Foi *visiting scholar* da University of British Columbia e da Delft University of Technology.