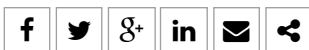




## Engenheiros de dados: a ascensão de uma nova profissão

Nas áreas emergentes nas tecnologias de informação, é natural que os papéis dos intervenientes inicialmente se confundam, em especial quando as próprias áreas não estão claramente definidas



Na área de **Data Science** não só não é claro em muitos contextos qual o papel de um *data scientist*, como este é, frequentemente considerado um super-homem (ou super-mulher) que tem que conhecer desde os aspetos mais complexos de plataformas como Hadoop, MapReduce, Hive, Pig, MySQL, MongoDB, Cassandra, até um portfólio completo de algoritmos de *machine learning* e ser fluente em linguagens como R, Python ou Scala.

Não é viável ou até desejável que as mesmas pessoas tenham responsabilidades tão diversas - e que implicam, por vezes, caminhos de atualização tão diferentes? Na verdade, não.

O que tenderá a acontecer é um separar das águas, que se delinea já.

Por um lado, aqueles que irão estar focados na componente de modelação e análise. Por outro, os que trabalharão sobretudo nas infraestruturas de dados. Esta função caberá aos **engenheiros de dados** (ou *data engineers*).

Os engenheiros de dados são os profissionais que preparam a infraestrutura para suportar os dados que serão posteriormente analisados por cientistas de dados. Terão habitualmente uma formação em engenharia de software e integram dados de múltiplos recursos e fazem também a gestão de **big data**. Criam *queries* complexas e o seu objetivo é otimizar o desempenho do ecossistema de big data da companhia para a qual trabalham. Poderão executar ETL (Extract, Transform and Load) no topo dos grandes datasets e criar data warehouses que poderão ser analisados ou reportados por data scientists.

Para além disso, e porque os engenheiros de dados se focam mais no projeto e na arquitetura, tipicamente não é expectável que sejam também especialistas em machine learning ou em análise para big data. Ou seja, estaremos num campo já conhecido noutros contextos: o da distinção entre ciência e engenharia. São funções complementares, ambas indispensáveis numa estratégia empresarial de boa governança dos dados. Não há subalternização de uma função em relação a outra. O que é fundamental é que as empresas entendam que se tratam de funções diferentes, para as quais são necessárias competências diferentes e trajetórias profissionais diferentes.

Mas, mesmo para os atuais profissionais na área de sistemas de informação, nem sempre é clara a distinção quando tentam recentrar carreiras para estas novas áreas dos sistemas de informação. Na verdade, não poderá existir data science de excelência sem uma excelente data engineering. Quanto mais cedo as empresas focarem os seus esforços de contratação, formação e organização tendo em conta estas distinções, tanto melhor poderão maximizar os investimentos que vierem a fazer em big data e *analytics*.

Henrique Carreiro,

*Docente de cloud computing e mobilidade empresarial na Nova Information Management School*



## Recomendado pelos leitores

### ☰ OPINIÃO

O custo real dos consumíveis de impressão falsificados

LER MAIS

(<http://www.itchannel.pt/news/opiniao/o-custo-real-dos-consumiveis-de-impressao-falsificados->)

### ☰ OPINIÃO

Inteligência Artificial: a arma (pouco) secreta da Google para a próxima geração da cloud

LER MAIS

(<http://www.itchannel.pt/news/opiniao/inteligencia-artificial-a-arma-pouco-secreta-da-google-para-a-proxima-geracao-da-cloud>)