TAMIR, Mike; MILLER, Steven; GAGLIARDI, Alessandro. The Data Engineer. **Ssrn Electronic Journal**, [S.L.], 2015. Elsevier BV. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2762013. Acesso em: 29 maio 2021.

* Data science is better thought of as a broad field with numerous sub-fields(...)(p.1)
* Data engineering is at least a distinct sub-field of data science if not it’s own field altogether.
* A few have pigeon holed the data engineer as data wranglers or plumbers who clean up data and make it ready to analyze. That is a key part of the work, but is only a component. The data engineers work includes a broad range of knowledge and skill:
  + Extract, clean, and integrate data (wrangling)
  + Bridge between the data science models and production systems
  + Implement machine learning & computational algorithms at scale
  + Put the right data system to work for the job at hand. Meaning they need a deep understanding of transactional ACID databases along with a growing variety of NoSQL databases including JSON document, graph, column stores, and partitioned row.
  + Demonstrate a deep understanding of distributed computing and database considerations for consistency, scalability, and security.
  + Protect customer privacy and anonymity.
* A data engineer makes it possible to take huge amounts of data and translate it into insights.(...) A savvy data engineer can provide an interface to those data that make them useful.(p.5)
* Data engineering is not simply about maintaining a repository for huge volumes of data, it is about creating possibilities for everyone from developers to data scientists to executives. (p.5)