

How does Data Engineering differ from traditional Web / Desktop / Mobile Development?

Dr. James Pizagno
(Software/Data Engineer)

27 February 2018

People

- The people are an important part of Engineering
 - Both Data Scientists and engineers
- Data Scientist are more technical than web surfers
 - typically scientists doing analysis
 - Data Scientists are your coworkers
- Data Engineering Team is more diverse:

	Data Engineering	Web/Desktop/Mobile
Experience	0 to 1+ year. 100+ lines of Scala/Spark Code -> „Team Tech Lead” :>	Years to decades. (grey hair, was „Webmaster” in 1990s)
Training	Online, on-the job, and new college programs	IT + few tech. degrees
Roles	Harder to define. Software Engineering + Systems Engineering + feel for data	Well-defined: Frontend, Backend, DBA, Sys.-Admin.

Data

- The data in „Data Engineering” is more diverse, and undefined

	Data Engineering	Web/Desktop/Mobile
Velocity	10TB/Month	1KB/sec
Variety	Self-defined from source (JSON, CSV, Parquet, Logs)	Defined by engineers internally (Web form+CSS+JPA)
Volume	PB Otto: 2PB	GB Gelbeseiten.de: 20GB
Veracity	Sparse data (1% filled), a- lot-of noise (80% trash)	Well defined (i.e. order basket, or entry form)
Goal	Data Transport, Cleaning, Security, Availability	Data Transport, Cleaning, Security, Availability

Technology

- Data Engineering technology is
 - more diverse
 - younger technology

	Data Engineering	Web/Desk/Mobile
Languages	Python (2.7->2010), Scala(2.11->2014)	Java (1.6->2006), JS(1.7->2008)
Frameworks	Spark(1.6->2016), Hadoop (HDFS, Hive/Impala. 2.6->2016)	Spring (2002), ReactJS (2011), Bootstrap (0-2011)
Platforms	Cloudera (4.7->2014) AWS (S3+EC2 ->2006)	Android (2.0->2009), iOS(2009/2010)
years	2010-2014	2000s+

- What are your thoughts?
- How are your teams organized?
 - Everybody does everything?
 - DevOps vs. Data Science?
 - Each team does both Data Science and DevOps?