



دوره آموزشی
آشنایی با مبانی پردازش و ابزارهای نوین

کلان داده

بخش زیرساخت و معماری

مدرس : مجتبی بنائی

جلسه اول



مجتبی بنائی



۱. مدیر بخش دیتای شرکت صباتامین (بازوی سرمایه گذاری شستا)

۲. موسس و سردبیر وب سایت مهندسی داده (BigData.ir)

۳. دانشجوی دکترای نرم افزار دانشگاه تهران

(گروه تحقیقاتی پایگاه داده دانشگاه)

۴. مشاوره و تدریس اصول طراحی سامانه‌های مقیاس پذیر

۵. مدرس دوره‌های مهندسی داده موسسه نیک‌آموز

۶. و ...



پیش نیازها

- مناسب برای همه علاقه‌مندان

- آشنایی با بانک‌های اطلاعاتی رابطه‌ای ضروری است.

- آشنایی با زبان پایتون/مفاهیم کلان داده مزیت محسوب می‌شود.

- توضیح همه ابزار مورد استفاده

- داکر

- دستورات خط فرمان

- زبان پایتون (Airflow)

- SQL (Hive-Hue)

- ابزار ...





آنچه خواهیم آموخت

- آشنایی با داکر / خط فرمان لینوکس
- مفاهیم پایه کلان داده و کار عملی با هدوپ / هایو
- آشنایی با AirFlow
- کار عملی با کافکا / اسپارک استریمینگ
- دیتابیس های تحلیلی کلیک هوس



مهندسی داده

تعریف / جایگاه / مهارتها / ابزارکار



مهندسی داده چیست؟

- مهندسی سامانه‌های مقیاس‌پذیر پردازش داده . (موسسه استاندارد آمریکا)
- مهندسی زیرساخت‌های پردازش داده
- مهندسی : انتخاب بهترین فناوری و ابزار
- زیرساخت‌ها :
- دیتابیس‌ها
- اکوسیستم بیگ‌دیتا
- صف‌های توزیع شده
- ابزار
- پردازش داده



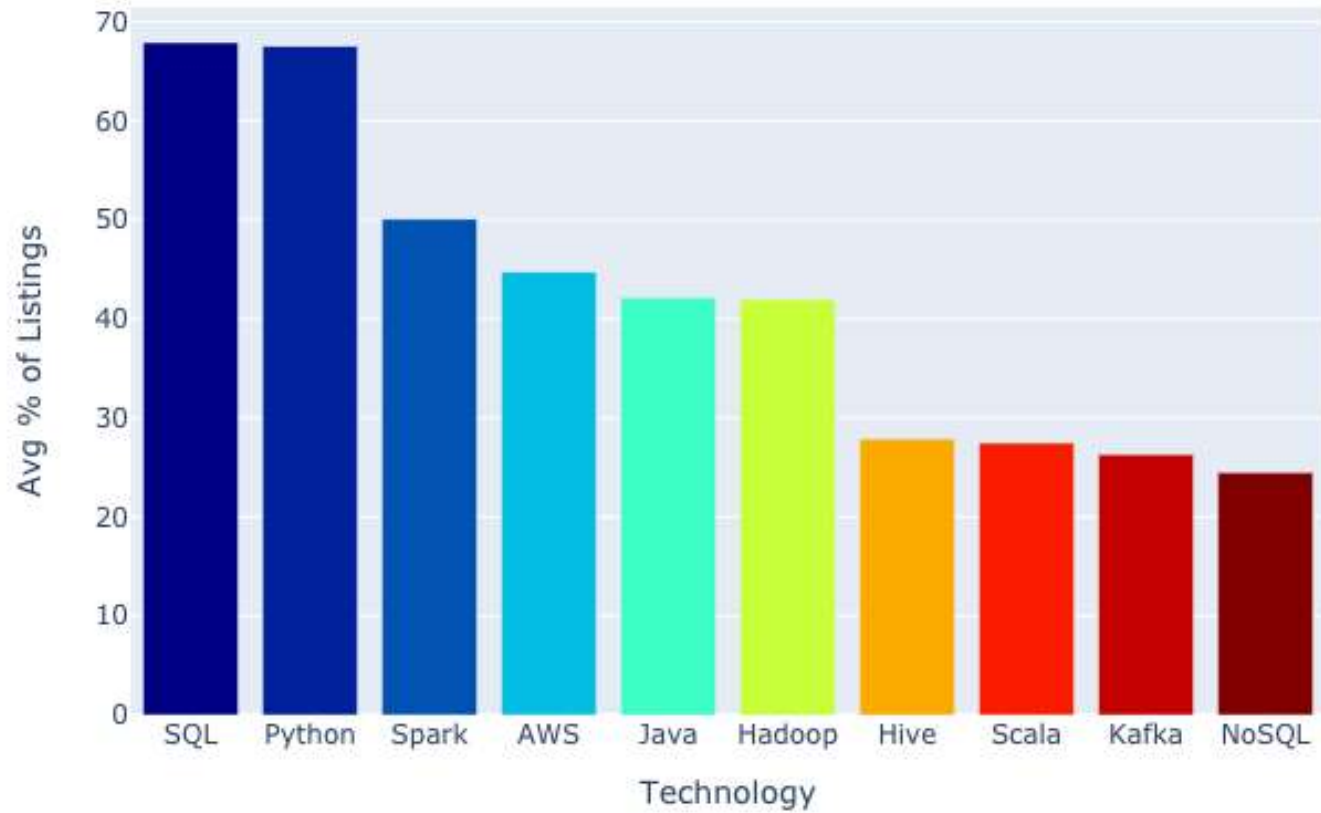
مهندس داده

- بانک‌های اطلاعاتی رابطه‌ای و غیررابطه‌ای
- سیستم‌های توزیع شده
- چارچوب‌های پردازش کلان‌داده (آفلاین/جریانی)
 - هدوپ
 - اسپارک
 - بیم / فلینک / اپکس / ایگنایت /
- انبارش داده / دیتابیس‌های تحلیلی
- فرمت‌های ذخیره فایل / سریالیزیشن / Object Storage
- مفاهیم پایه DevOps / مدیریت منابع
- طراحی میکروسرویسی / معماری چندلایه / Gateways



مهارت‌پرور

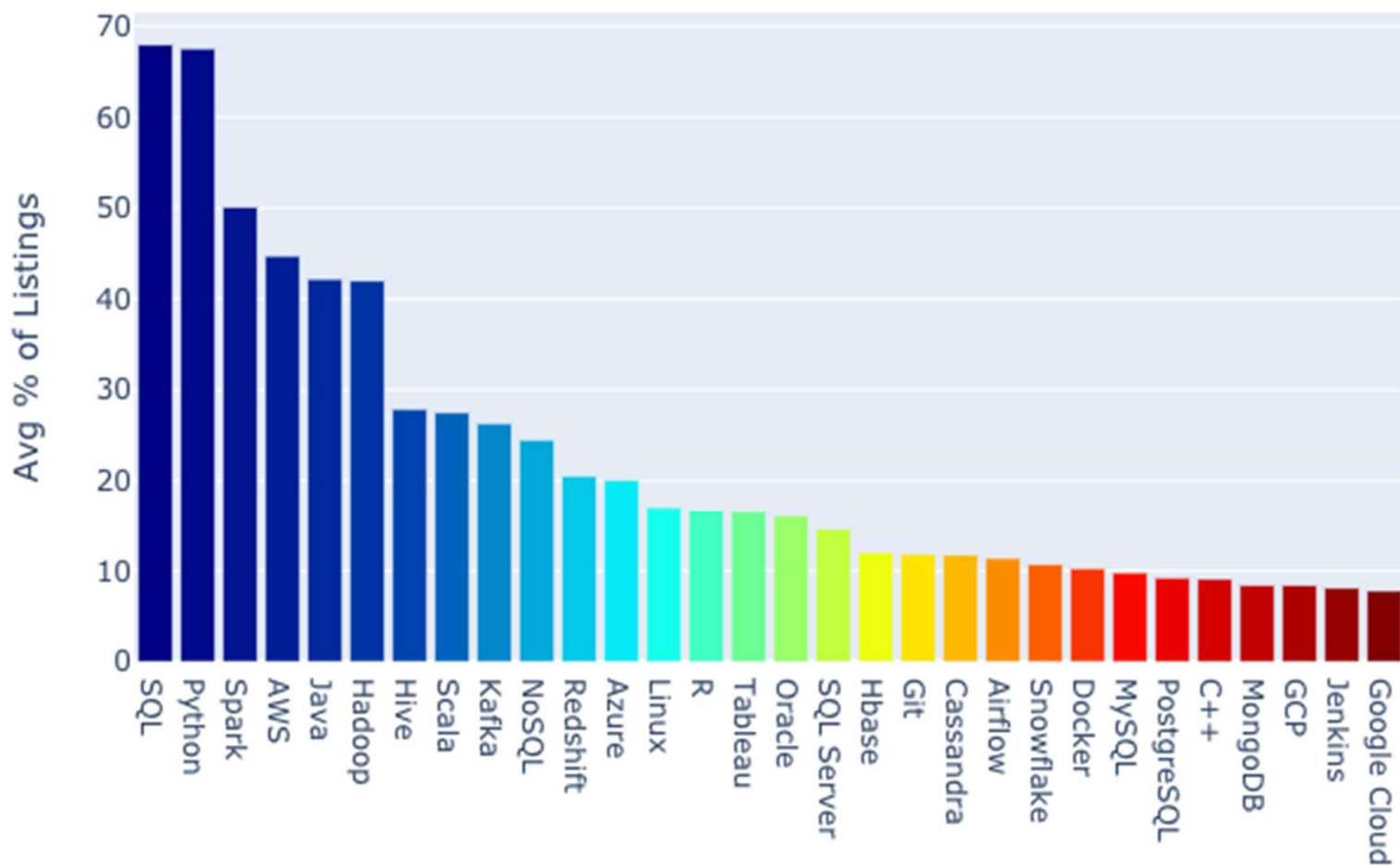
Technologies in Data Engineer Job Listings 2020





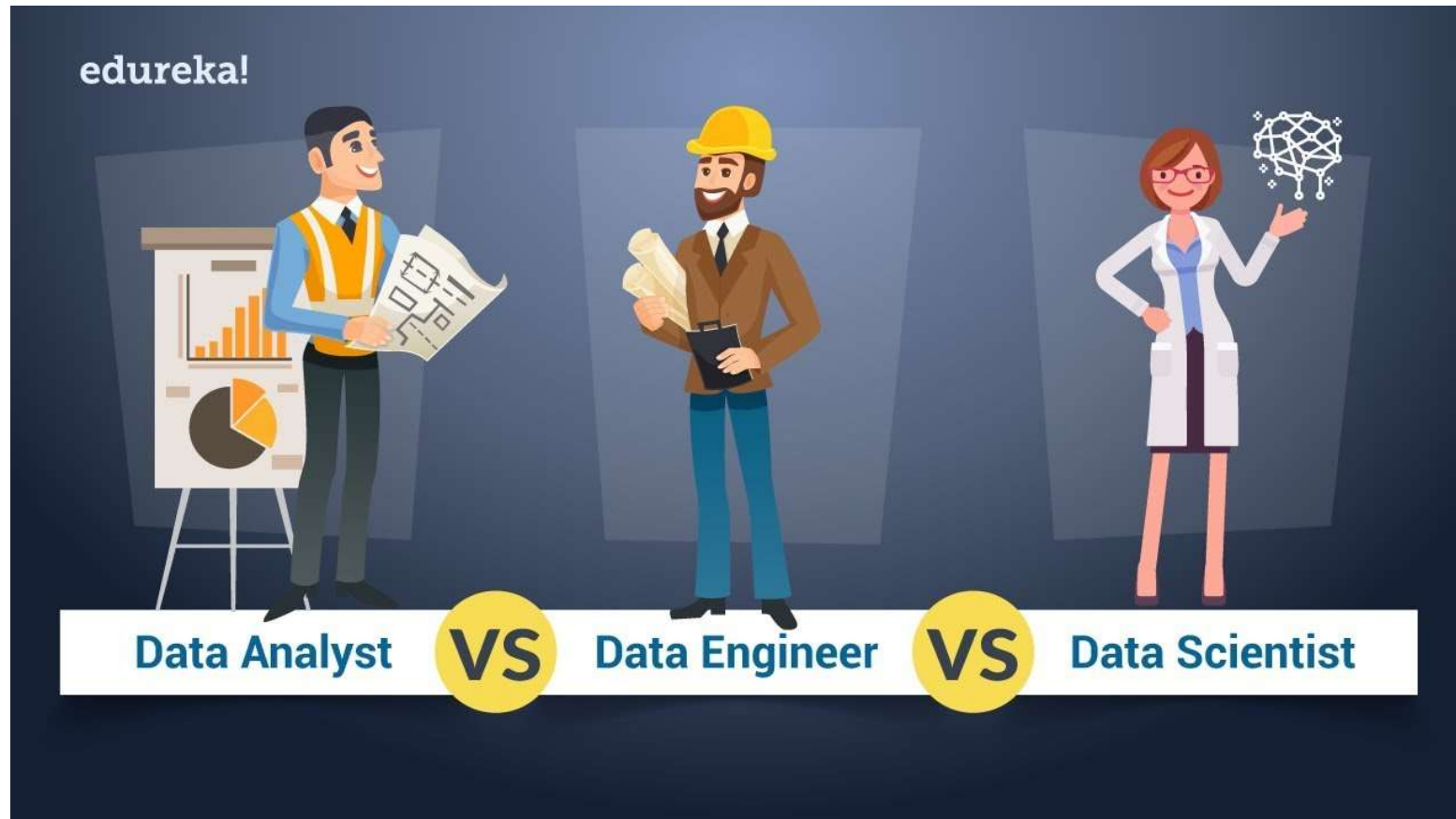
معماری ۲۰

Technologies in Data Engineer Job Listings 2020



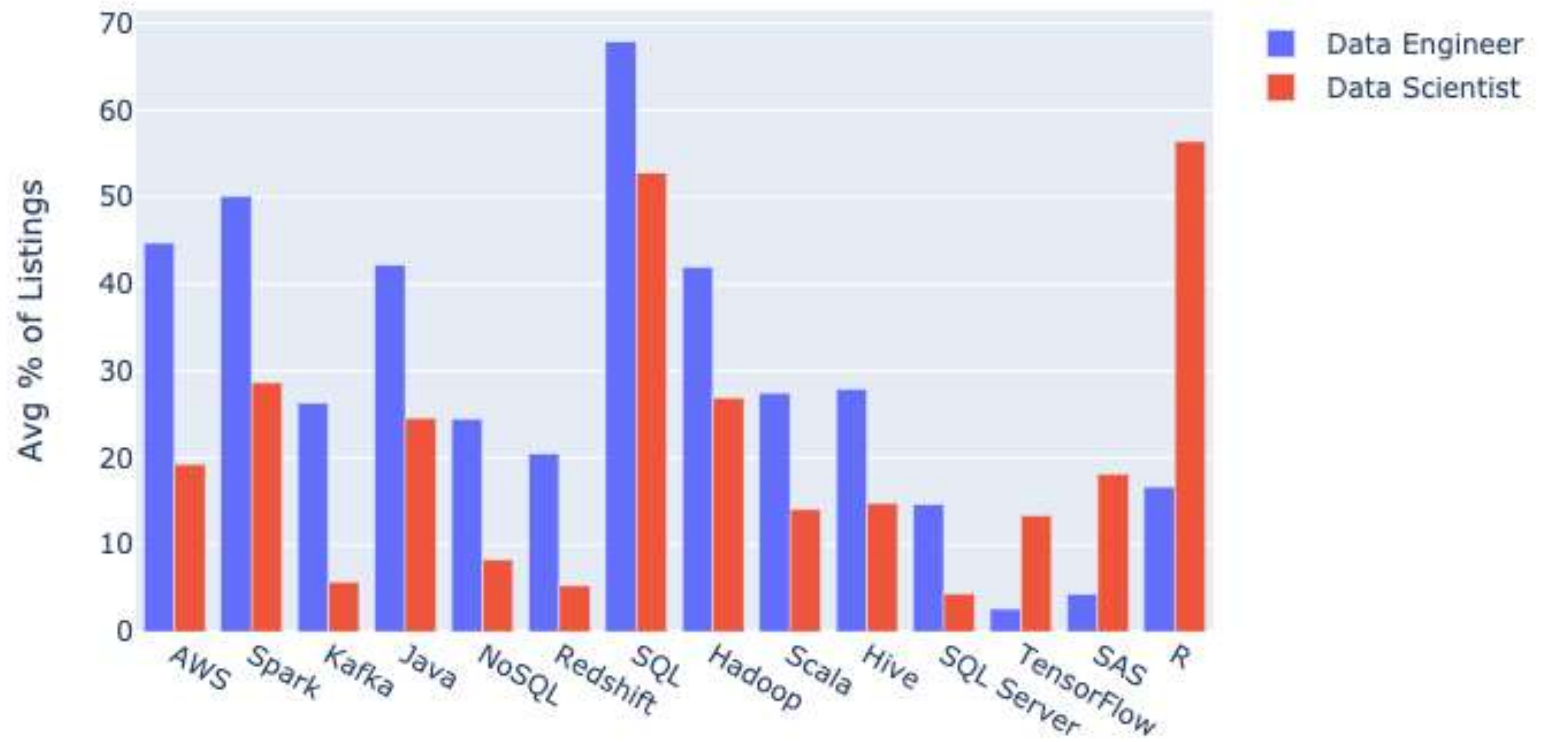


مهندس داده / تحلیلگر داده / دانشمند داده





Largest Differences for Data Engineer vs. Data Scientist Job Listing Keywords





ابزارهای یک مهندس داده

- خط فرمان لینوکس/ویندوز
- SQL
- ابزارهای کار با دیتابیس ها ← Dbeaver, Hue
- Git (کدها / خطوط پردازش/SQL)
- ابزارهای مانیتور و پایش (گرافانا/کیبانا)
- Data Flow/Work Flow
- مستندسازی، دیتاکاتالوگ
- CDC Tools
- Data Versioning
- Visualization
-