

دوره آموزشی مهندسی داده [Data Engineer]

مدرس: مجتبی بنائی



جلسه: اول

nikamooz;
آموزش برنامه نویسی و اجرای پروژه

مجتبی بنائی



۱. مدرس دوره‌های مهندسی داده موسسه نیک‌آموز
۲. موسس و سردبیر وب سایت مهندسی داده (BigData.ir)
۳. دانشجوی دکترای نرم افزار دانشگاه تهران
(گروه تحقیقاتی پایگاه داده دانشگاه)
۴. عضو تیم مهندسی داده یکتانت
۵. مشاوره و تدریس اصول طراحی سامانه‌های مقیاس پذیر
۶. و...

آنچه خواهیم دید

جلسه اول

۱. دوره مهندسی داده در یک نگاه
۲. نحوه برگزاری دوره
۳. آشنایی با مفاهیم اصلی مهندسی داده
۴. بررسی جایگاه شغلی و مهارت‌های موردنیاز
۵. آشنایی با داکر
۶. کارگاه عملی داکر



زمان بندی



- مهندسی داده و مرور سرفصل
۳۰ دقیقه

- آشنایی با داکر: ۳۰ دقیقه

- شروع کار با داکر: ۱ ساعت

- دستورات پیشرفته داکر: ۱ ساعت

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
- طراحی خطوط پردازش داده با نایفای
- بانک های اطلاعاتی غیررابطه ای
- کار عملی با کافکا / اسپارک استریمینگ
- دیتابیس های تحلیلی کلیک هوس / دروید
- ابزارهای پایش و مانیتورینگ

چند نکته

• مبانی مهندسی داده

- مروری بر ابزار و فناوری‌های مورد نیاز
- معرفی منابع و تمرین‌های مناسب

• دوره های تکمیلی

- دوره پیشرفته اسپارک
- دیتابیس‌های تحلیلی و انبارهای داده نوین
- ابزارهای مدیریت جریان کار و طراحی خطوط پردازش داده
-

نحوه برگزاری دوره

- هر هفته یک جلسه حدودا سه ساعته
 - اختصاص بیشتر وقت هر جلسه به کار عملی
- قرارگیری تمامی کدها و فایلها بر روی گیت
 - آدرس پوشه گیت دوره

https://gitlab.com/nikamooz_bigdata/de

- گروه VIP تلگرام
 - تمرینهای هر جلسه هم در گروه قرار خواهد گرفت
- برگزاری یک جلسه آنلاین برای پرسش و پاسخ
 - در انتهای هر مبحث اصلی

مهندسی داده

تعریف / جایگاه / مهارتها / ابزارکار

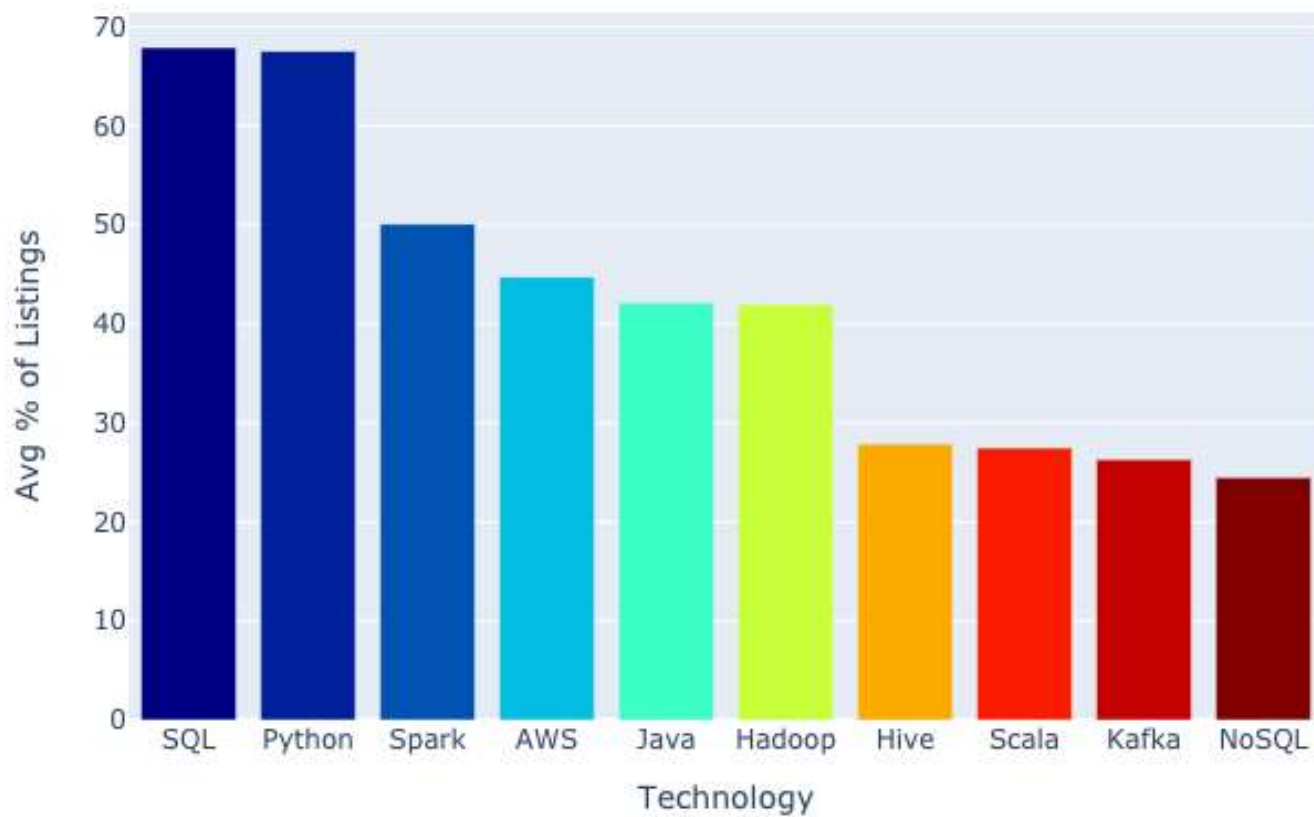
مهندسی داده چیست؟

- 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

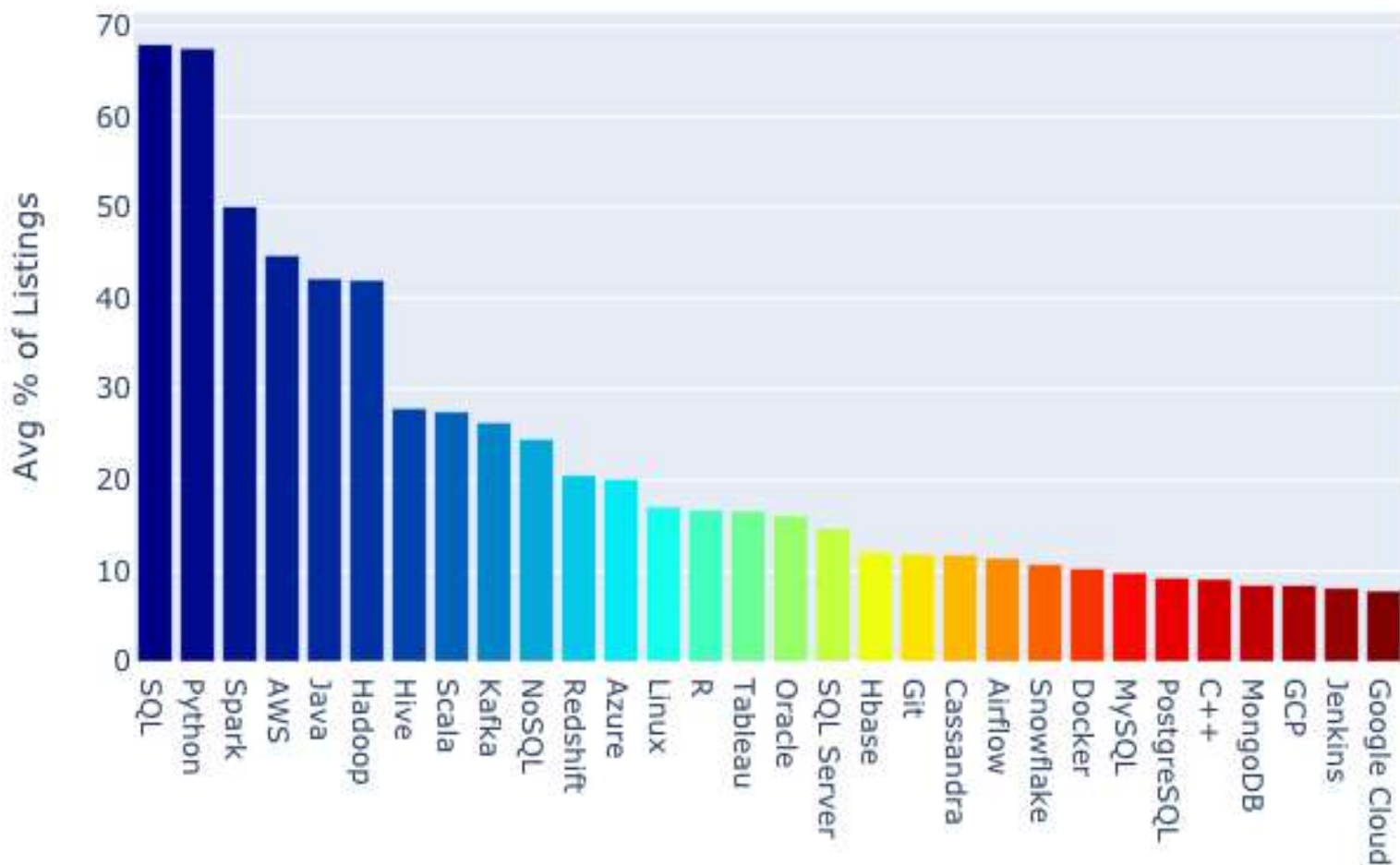
- بانک‌های اطلاعاتی رابطه‌ای و غیررابطه‌ای
- سیستم‌های توزیع شده
- چارچوب‌های پردازش کلان‌داده (آفلاین/جریانی)
 - هدوپ
 - اسپارک
 - بیم / فلینک / اپکس / ایگنایت /
- انبارش داده / دیتابیس‌های تحلیلی
- فرمت‌های ذخیره فایل / سریالیزیشن / Object Storage
- مفاهیم پایه DevOps / مدیریت منابع
- طراحی میکروسرویسی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

Technologies in Data Engineer Job Listings 2020

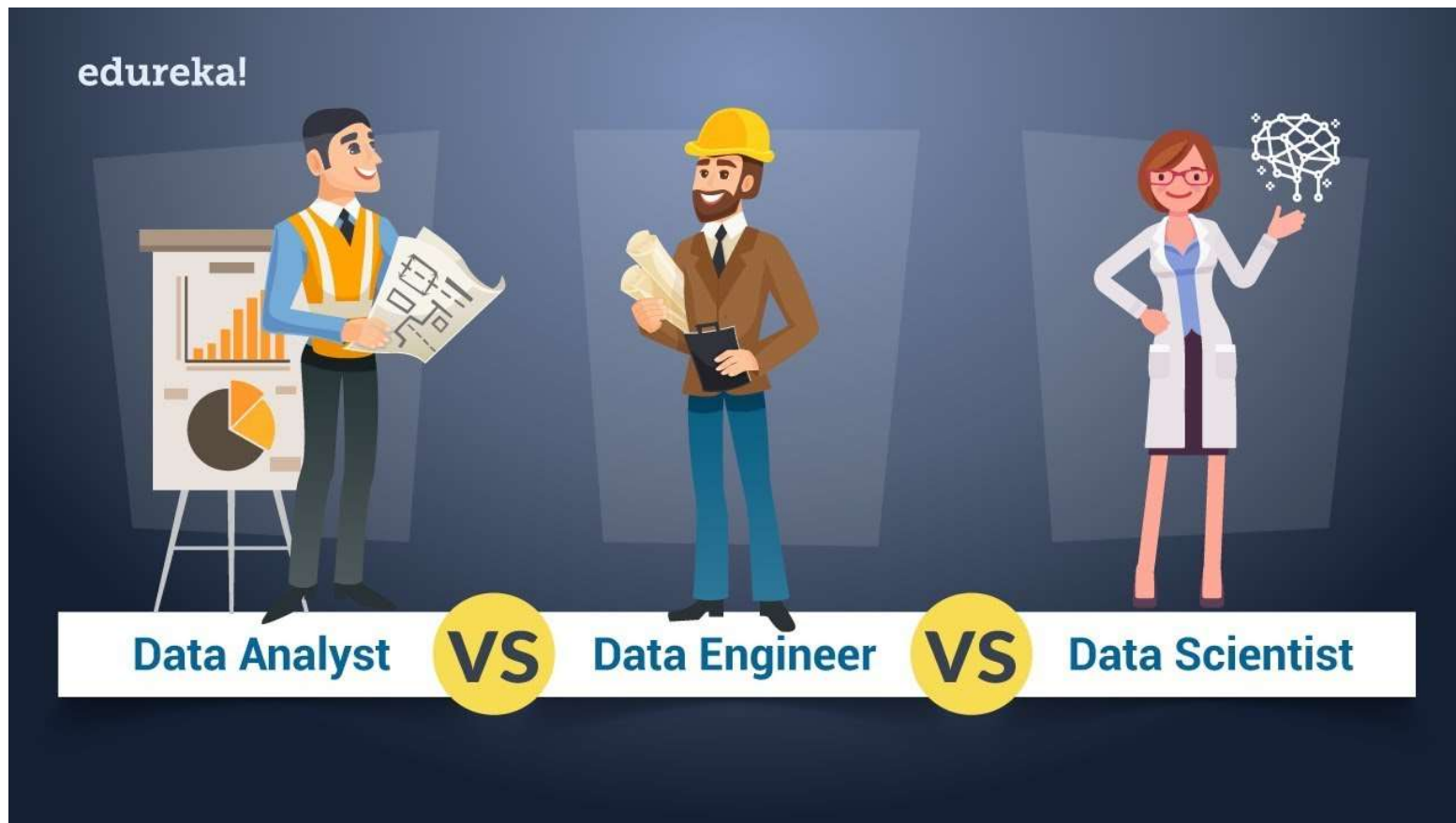


مهارت‌های ۲۰

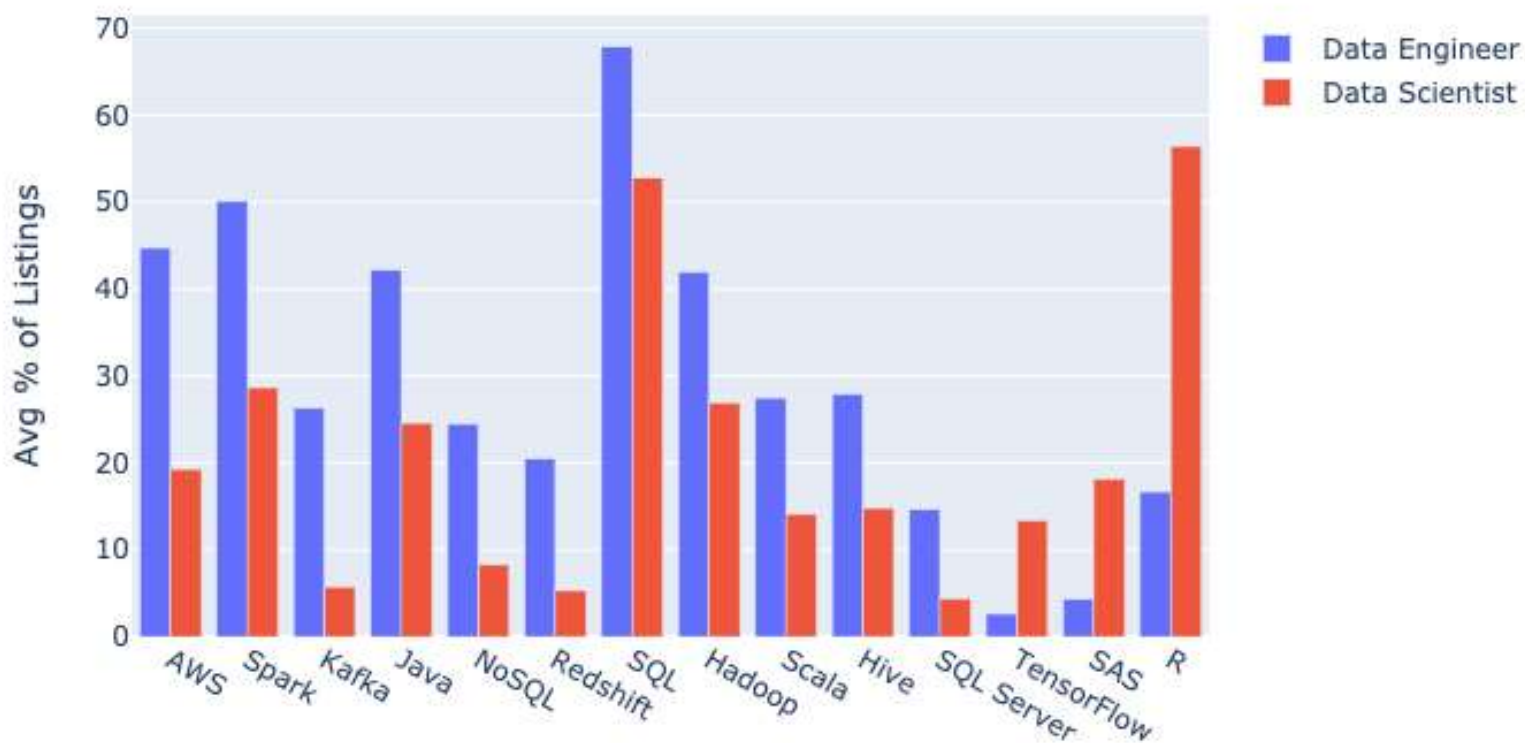
Technologies in Data Engineer Job Listings 2020



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 (گرافا/کیبانا)
- Data Flow/Work Flow
- مستندسازی، دیتاکاتالوگ