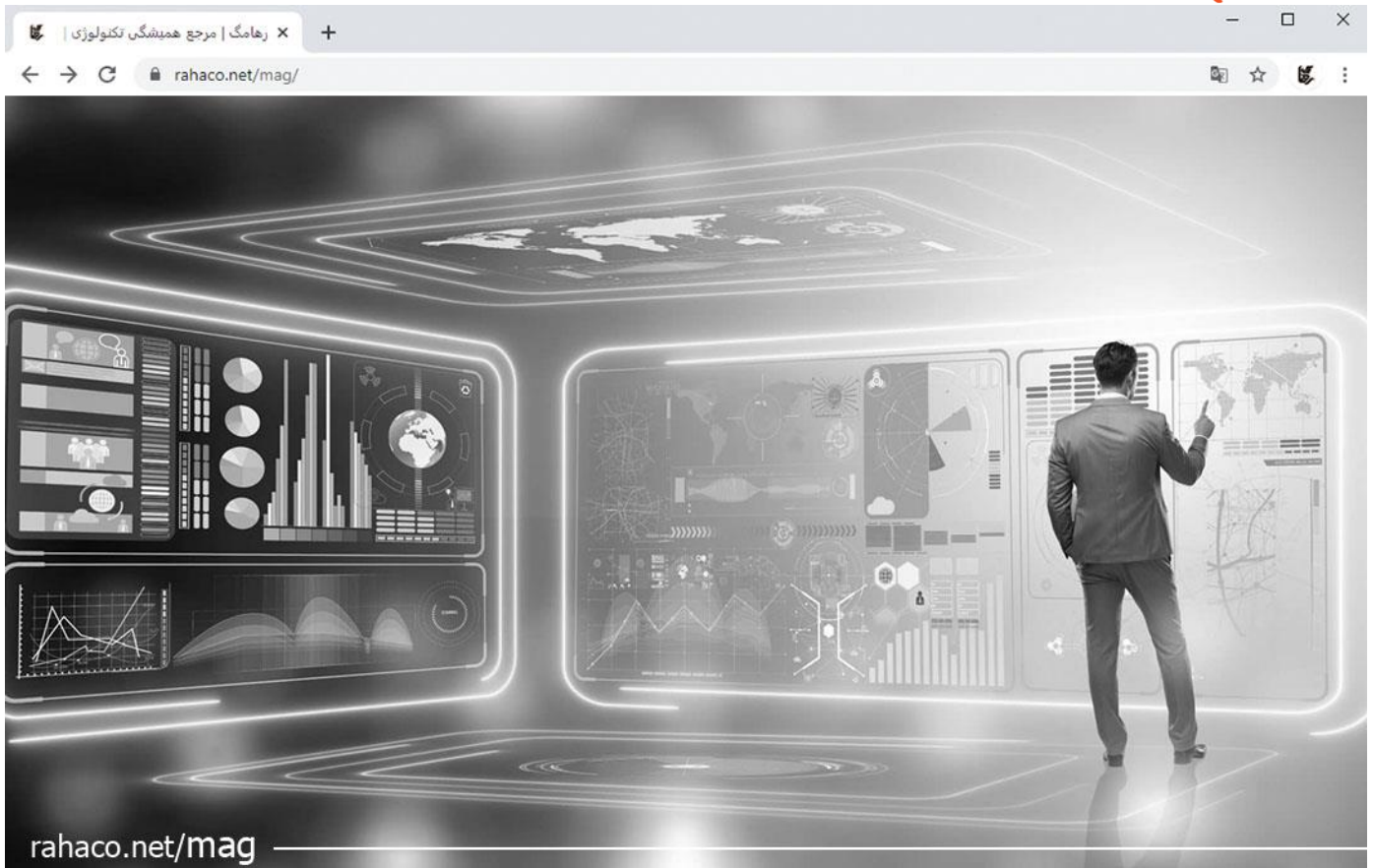




مجموعه شرکت های مهندسی دانش بنیان رها

مهندس داده (data engineer) کیست و چه عملکردی در کسب و کارهای امروزی دارد؟

شرکت رهاکو



فهرست

- 3 تعریف مهندس داده
- 3 یک مهندس داده چه کارهایی را باید انجام دهد؟
- 3 چگونه به یک مهندس داده تبدیل شویم؟
- 3 مسیر شغلی مهندسان داده چگونه است؟
- 4 مهارت هایی که برای تبدیل به یک مهندس داده باید بلد باشید
- 5 مهندسان داده از چه ابزارهایی استفاده می کنند؟
- 5 مسئولیت های مهندس داده
- 6 نتیجه گیری



مهندسان داده اغلب مسئول ساخت الگوریتم‌ها برای دسترسی به داده‌های خام هستند. اما برای انجام این کار، مهندسان داده باید اهداف یا نیازهای یک شرکت یا مشتری را کامل متوجه شوند، زیرا همراستایی استراتژی‌های داده با اهداف تجاری بسیار مهم است، به ویژه زمانی که مهندسان داده درگیر مجموعه داده‌ها و پایگاه‌های داده بزرگ و پیچیده هستند. سازمان‌ها برای جمع‌آوری حجم عظیمی از داده‌ها به Data Engineer نیاز دارند.

تعریف مهندس داده

Data Engineer یک کارشناس فناوری اطلاعات است. وظیفه اصلی کارشناس مهندس داده تهیه داده‌ها برای استفاده‌های تحلیلی یا عملیاتی می‌باشد. مهندسان داده در ارتباط با تیم‌های علم داده کار می‌کنند و شفافیت داده‌ها را بهبود می‌بخشند و کسب و کارها را قادر می‌سازند تا تصمیمات تجاری قابل اعتمادتری بگیرند. آن‌ها داده‌ها را ادغام و پاکسازی می‌کنند.

هدف مهندسان داده دسترسی آسان به داده‌ها و بهینه‌سازی اکوسیستم کلان داده سازمان‌ها است. مهندسان داده علاوه بر دسترسی به داده‌ها، تجزیه و تحلیل داده‌های خام را برای ارائه پیش‌بینی و نشان دادن روندهای کوتاه مدت و بلند مدت ایجاد می‌کنند. بدون مهندسی داده، درک حجم عظیمی از داده‌هایی که در دسترس کسب و کارها هستند غیر ممکن خواهد بود.

یک مهندس داده چه کارهایی را باید انجام دهد؟

مهندسان داده برای ساختن سیستم‌هایی کار می‌کنند، که داده‌های خام را جمع‌آوری و مدیریت کنند تا بتوانند اطلاعات قابل استفاده را برای تفسیر دانشمندان داده و تحلیلگران تجاری ارسال کنند. هدف نهایی یک Data Engineer این است که داده‌ها را در دسترس قرار دهد، تا سازمان‌ها بتوانند از آن برای ارزیابی و بهینه‌سازی عملکرد سازمان استفاده کنند. مهندسان داده سیستم‌هایی را برای جمع‌آوری، اعتبارسنجی، آماده‌سازی داده‌های با کیفیت بالا می‌سازند.

چگونه به یک مهندس داده تبدیل شویم؟

با کسب مجموعه‌ای از مهارت‌ها و دانش‌ها، می‌توانید در حرفه مهندسی داده پیشرفت چشمگیری داشته باشید. بسیاری از مهندسان داده دارای مدرک لیسانس در علوم کامپیوتر یا یک رشته مرتبط هستند. با کسب مدرک مورد نظر، می‌توانید پایه و اساس دانشی را که در این زمینه به آن نیاز دارید، بسازید.

مسیر شغلی مهندسان داده چگونه است؟

Data Engineer برای همیشه نقش ثابتی ندارد بلکه بسیاری از مهندسان داده به عنوان مهندس نرم افزار یا تحلیلگر هوش تجاری می‌توانند تغییری در مسیر شغلی خود داشته باشند. بنابراین هنگامی که در حال پیشرفت در رشته مهندس داده هستید، این احتمال وجود دارد که بتوانید به سمت نقش‌های مدیریتی نیز حرکت کنید یا یک معمار داده شوید.



مهارت‌هایی که برای تبدیل به یک مهندس داده باید بلد باشید

کدنویسی: مهارت در زبان‌های برنامه نویسی برای نقش مهندس داده بسیار ضروری است، بنابراین برای یادگیری و تمرین مهارت‌های خود دوره‌هایی را در نظر بگیرید. زبان‌های برنامه نویسی رایج برای مهندسی داده عبارتند از: SQL، NoSQL، Python، Java، R و Scala.

سیستم‌های ETL (استخراج، تبدیل و بارگذاری): فرآیندی است که توسط آن داده‌ها را از پایگاه‌های داده و سایر منابع به یک مخزن واحد مانند: یک انبار داده منتقل می‌کند. ابزارهای رایج ETL عبارتند از: Talend و Alooka، Stitch، Xplenty.

ذخیره سازی داده‌ها: تمام داده‌ها نباید به یک شکل ذخیره شوند، به خصوص وقتی صحبت از کلان داده می‌شود. همانطور که راه حل‌های داده را برای یک شرکت طراحی می‌کنید، باید بدانید که برای مثال: چه زمانی از یک دریاچه داده در مقابل یک انبار داده استفاده کنید.

اتوماسیون و اسکریپت نویسی: اتوماسیون بخشی ضروری از کار با کلان داده است، زیرا سازمان‌ها قادر به جمع آوری اطلاعات زیادی هستند. شما باید بتوانید اسکریپت‌هایی بنویسید که کارهای تکراری را به صورت خودکار انجام دهد.

ابزارهای کلان داده: مهندسان داده فقط با داده‌های معمولی کار نمی‌کنند. آن‌ها اغلب وظیفه مدیریت کلان داده را بر عهده دارند. امروزه ابزارها و فناوری‌ها در حال تکامل هستند و بسته به شرکت‌ها متفاوت می‌باشند، اما برخی از محبوب‌ترین آن‌ها عبارتند از: Hadoop، MongoDB و Kafka.

Hadoop یک چارچوب مهندسی داده محسوب می‌شود. که برای ذخیره و محاسبه مقادیر زیادی داده با استفاده از شبکه‌ای از رایانه‌ها استفاده می‌شود.

آپاچی کافکا یک ابزار محبوب برای ساخت خطوط لوله استریم است و بیش از 80 درصد از شرکت‌ها از آن استفاده می‌کنند. به عنوان مثال: نتفلیکس از کافکا برای پردازش بیش از 500 میلیارد رویداد در روز استفاده می‌کند، از فعالیت‌های مشاهده کاربر گرفته تا گزارش‌های خطا.

Striim یک پلتفرم ادغام داده است. که توسط مهندسان داده برای ادغام داده‌ها استفاده می‌شود.

رایانش ابری: شما باید فضای ذخیره سازی ابری و رایانش ابری را درک کنید زیرا شرکت‌ها به طور فزاینده‌ای سرورهای فیزیکی را با خدمات ابری مبادله می‌کنند. مبتدیان بهتر است یک دوره آموزشی در زمینه خدمات وب آمازون (AWS) یا Google Cloud در نظر بگیرند.

امنیت داده‌ها: در حالی که برخی از شرکت‌ها ممکن است تیم‌های اختصاصی امنیت داده داشته باشند، اما بسیاری از مهندسان داده وظیفه دارند تا داده‌ها را به طور ایمن مدیریت و ذخیره کنند تا بتوانند از گم شدن یا سرقت داده‌ها محافظت کنند.



مهندسان داده از چه ابزارهایی استفاده می کنند؟

هیچ ابزار یکسانی وجود ندارد که مهندسان داده از آن استفاده کنند. در عوض، هر سازمانی بر اساس نیازهای تجاری که دارد از ابزارهایی متفاوت استفاده می کند. با این حال، در متن زیر برخی از ابزارهای محبوبی که مهندسان داده استفاده می کنند آورده شده است. احتیاج نیست در اینجا بر همه ابزارها تسلط داشته باشید، اما توصیه می کنیم اصول هر ابزار اصلی را بیاموزید.

پایگاه داده: در دنیای امروز ابزارها و فناوری ها دائماً در حال تکامل هستند، SQL ابزاری اساسی برای مهندسان داده است. SQL زبان برنامه نویسی استاندارد برای ایجاد و مدیریت سیستم های پایگاه داده رابط های (مجموعه ای از جداول که از سطر و ستون تشکیل شده) است.

پایگاه های داده NoSQL غیر جدولی هستند و بسته به مدل داده هایشان می توانند به شکل یک نمودار یا یک سند باشند. پایگاه داده های معروف SQL شامل: MySQL، PostgreSQL، Oracle و MongoDB، Cassandra و Redis. نمونه هایی از پایگاه های داده محبوب NoSQL هستند.

پردازش داده ها: یکی از وظایف مهندسان داده اهمیت پردازش داده ها، برای کسب و کارهای امروزی در زمان مناسب برای ارتقای تصمیمات تجاری است. در اصل مهندسان داده مسئول ساخت خطوط لوله انتقال داده و پردازش داده ها در زمان مناسب هستند. آپاچی اسپارک یک موتور تحلیلی است که برای پردازش هایی با سرعت بالا استفاده می شود.

زبان های برنامه نویسی: مهندسان داده حداقل باید به یک زبان برنامه نویسی، برای ایجاد راه حل های نرم افزاری برای چالش های داده، مسلط باشند. پایتون به عنوان محبوب ترین و پرکاربردترین زبان برنامه نویسی در جامعه مهندسی داده در نظر گرفته می شود.

مسئولیت های مهندس داده

- ساخت و نگهداری زیر ساخت داده برای استخراج، تبدیل و بارگذاری بهینه داده ها از منابع متنوعی مانند: خدمات وب آمازون (AWS) و پلتفرم های کلان داده ابری گوگل.
- اطمینان از دسترسی به داده ها در هر زمان و اجرای سیاست ها، حفظ داده های شرکت با توجه به حریم خصوصی و محرمانه بودن داده ها.
- بهبود قابلیت اطمینان، سرعت و عملکرد سیستم های داده.
- ایجاد ابزارهای داده، خطوط لوله (اصلی ترین وظیفه یک مهندس داده ساخت کانال های انتقال اطلاعات است که به آن خطوط لوله می گویند) و سیستم های گزارش دهی بهینه برای حل مشکلات تجاری.



نتیجه گیری

مهندسان داده مسئول طراحی، نگهداری و بهینه سازی زیر ساخت داده برای جمع آوری و مدیریت داده ها هستند. آن ها مسئول ایجاد خطوط لوله هستند که داده های خام را به فرمت های قابل استفاده برای دانشمندان داده و سایر مصرف کنندگان داده تبدیل می کنند. نقش مهندس داده برای رسیدگی به جنبه های داده اصلی، با مهندسی نرم افزار و علم داده تکامل یافته است.

مهندسان داده از اصول مهندسی نرم افزار برای توسعه الگوریتم هایی استفاده می کنند که فرآیند جریان داده ها را به صورت خودکار انجام می دهد. مهندسان داده همچنین با دانشمندان داده همکاری می کنند تا زیر ساخت یادگیری ماشین و تجزیه و تحلیل داده را از آزمایش کنند.