

ساخت زیرساخت مهندسی داده

در فصل قبل، با نقش‌ها و مسئولیت‌های مهندس داده آشنا شدیم و ابزارهای کلیدی مورد استفاده در این حوزه را **Data Pipeline** معرفی کردیم؛ از جمله انواع دیتابیس‌ها، زبان‌های برنامه‌نویسی، و ابزارهای ایجاد و زمان‌بندی

در این فصل، قصد داریم ابزارهای موردنیاز را نصب و پیکربندی کنیم تا در ادامه‌ی مسیر از آن‌ها برای ساخت پایپ‌لاین‌های داده و تحلیل نتایج استفاده کنیم.

ابزارهایی که در این فصل نصب و پیکربندی می‌کنیم

♦ دیتابیس‌ها:

- PostgreSQL
- Elasticsearch

♦ ابزارهای ساخت پایپ‌لاین داده:

- Apache Airflow
- Apache NiFi

♦ ابزارهای مدیریتی و بصری‌سازی داده:

- pgAdmin (برای مدیریت PostgreSQL)
 - Kibana (برای مدیریت و تحلیل داده در Elasticsearch)
-

چرا به این ابزارها نیاز داریم؟

- ✓ اجرای پایپ‌لاین‌های داده: انتقال داده از یک منبع به مقصد
- ✓ ETL مدیریت پردازش داده: زمان‌بندی و هماهنگ‌سازی فرآیندهای
- ✓ بصری‌سازی داده: مشاهده و تحلیل داده‌های پردازش‌شده برای دیباگ و تحلیل
- ✓ برای ذخیره و جستجوی داده‌ها Elasticsearch و PostgreSQL مدیریت و ذخیره‌سازی داده: استفاده از

هرچند با پیشرفت در مسیر یادگیری، ممکن است دیگر نیازی به برخی از این ابزارها نداشته باشید، اما درک پایه‌ای از آن‌ها برای پشتیبانی از سایر کاربران و نقش‌ها در تیم داده ضروری است.

موضوعات این فصل

- ♦ Apache NiFi نصب و پیکربندی
- ♦ Apache Airflow نصب و پیکربندی
- ♦ Elasticsearch نصب و پیکربندی
- ♦ Kibana نصب و پیکربندی
- ♦ PostgreSQL نصب و پیکربندی
- ♦ pgAdmin 4 نصب

در ادامه، گام به گام این ابزارها را نصب و تنظیم خواهیم کرد تا بتوانید از آن‌ها در پروژه‌های مهندسی داده خود استفاده کنید.