

۱۵۰۸۳۰۲۸  
به نام خدا

# اصول و مبانی علوم داده‌ای با Microsoft SQL Server 2016

مترجمان:

مهندس محمدمعین فاضلی

دکتر یعقوب فرجامی (عضو هیأت علمی دانشگاه قم)

## اصول و مبانی علوم داده‌ای با Microsoft SQL Server 2016

مترجمان: مهندس محمدمعین فاضلی، دکتر یعقوب فرجامی

ویراستار علمی: مهندس رامین مولاناپور

ناشر: انتشارات آتی‌نگر

ناشر همکار: انتشارات وینا

چاپ اول، ۱۳۹۶

شمارگان: ۵۰۰ نسخه

قیمت: ۱۳۰,۰۰۰ ریال

شابک: ۹۶۶-۷۶۳۱-۶۶۰۰-۷۸

ISBN: 978-600-7631-66-9

حق چاپ برای انتشارات آتی‌نگر محفوظ است.

نشانی دفتر فروش: خیابان بهارزاد، جنوبی، روبه‌روی کوچه رشتچی، پلاک ۱۴۴، واحد ۱

تلفن: ۶۶۵۶۵۳۳۶-۸ شابک: ۶۶۵۶۵۳۳۷



www.ati-negar.com \* info@ati-negar.com

سرشناسه: وودی، باک Woody ، Buck

اصول و مبانی علوم داده‌ای با Microsoft SQL Server 2016 / نویسندگان: Buck Woody [آتی‌نگران]:

مترجمان: محمدمعین فاضلی، یعقوب فرجامی

تهران: آتی‌نگر، ۱۳۹۶

۱۲۰ ص: مصور، جدول، نمودار.

ISBN: 978-600-7631-66-9

فیپا.

عنوان اصلی: Data Science with Microsoft SQL Server 2016

یادداشت: مؤلفان باک وودی، دانیل دین، دیراج گوها تاکورتا ...

موضوع: پایگاه‌های اطلاعاتی رابط‌های

شناسه افزوده: فاضلی، محمدمعین، ۱۳۷۲-، مترجم

شناسه افزوده: فرجامی، یعقوب، ۱۳۴۸-، مترجم

رده‌بندی کنگره

رده‌بندی دیویی

شماره کتابشناسی ملی

۱۳۹۶ عالف ۴/س ۷۶/۹ QAY۶

۰۰۵/۷۵۸۵

۴۶۸۴۸۰۲

# فهرست مطالب

## پیشگفتار

۷

## فصل اول: شیوه استفاده از این کتاب

۱۱

- ۱۱..... شیوه استفاده از این کتاب برای دانشمندان علوم داده‌ای و کارشناسان زبان R
- ۱۲..... مثالی کاربردی: ریزش مشتری
- ۱۲..... مثالی کاربردی: پیش‌بینی نیاز به تعمیرات و اینترنت اشیا
- ۱۳..... مثال کاربردی: پیش‌بینی
- ۱۴..... شیوه استفاده از این کتاب برای افراد تازه‌کار در زمینه علوم داده‌ای و زبان R
- ۱۴..... گام اول: بهای حیات
- ۱۶..... گام دوم: SQL Server و SQL تراکنشی (Transact-SQL)
- ۱۷..... گام سوم: محیط برنامه‌نویسی و زبان R

## فصل دوم: سرویس‌های R نرم‌افزار Microsoft SQL Server

۱۹

- ۱۹..... فواید R در SQL Server
- ۲۱..... مروری اجمالی بر معماری سرویس‌های R نرم‌افزار SQL Server
- ۲۱..... سرویس‌های R نرم‌افزار SQL Server
- ۲۳..... شروع به کار با سرویس‌های R نرم‌افزار SQL Server
- ۲۳..... راه‌اندازی و پیکربندی
- ۲۴..... سمت سرور
- ۲۵..... سمت کاربر
- ۳۰..... راه‌حل خود را عملیاتی کنید
- ۳۰..... استفاده از سرویس‌های R نرم‌افزار SQL Server به صورت یک زمینه محاسباتی
- ۳۲..... استفاده از رویه‌های ذخیره‌شده دارای کدهای R

## فصل سوم: مثالی از یک فرآیند علوم داده‌ای سرتاسری

۳۵

- ۳۵..... مرور کلی فرآیندهای علوم داده‌ای
- ۳۷..... فرآیند علوم داده‌ای در سرویس‌های R نرم‌افزار SQL Server
- ۳۷..... کار با داده‌ها و مدل‌سازی

|    |                                                        |
|----|--------------------------------------------------------|
| ۳۹ | فراهم کردن زیرساخت‌ها، محیط توسعه و ابزارها            |
| ۴۶ | داده‌های ورودی و اشیای SQLServerData                   |
| ۴۸ | تحلیل اکتشافی                                          |
| ۴۸ | خلاصه‌سازی داده‌ها                                     |
| ۵۰ | مصورسازی داده‌ها                                       |
| ۵۳ | مهندسی ویژگی‌ها                                        |
| ۵۴ | با استفاده از توابع R                                  |
| ۵۵ | با استفاده از یک تابع SQL                              |
| ۵۸ | ایجاد و ذخیره‌سازی مدل‌ها                              |
| ۵۸ | استفاده از محیط R                                      |
| ۶۰ | با استفاده از T-SQL                                    |
| ۶۱ | استفاده از مدل: آورده‌ی به داده‌ها با یک مدل ذخیره‌شده |
| ۶۵ | ارزیابی دقت مدل                                        |
| ۶۶ | خلاصه فصل                                              |

## فصل چهارم: ایجاد یک راه‌حل برای ه‌شکر ریزش مشتری

|    |                     |
|----|---------------------|
| ۶۷ |                     |
| ۶۷ | بررسی اجمالی        |
| ۶۹ | فهمیدن داده‌ها      |
| ۷۱ | ساخت مدل ریزش مشتری |
| ۷۳ | آموزش گام‌به‌گام    |
| ۸۰ | خلاصه فصل           |

## فصل پنجم: پیش‌بینی نیاز به تعمیرات و اینترنت اشیا

|    |                                                                    |
|----|--------------------------------------------------------------------|
| ۸۱ | اینترنت اشیا چیست؟                                                 |
| ۸۱ | پیش‌بینی نیاز به تعمیرات در حوزه اینترنت اشیا                      |
| ۸۲ | مثال‌های کاربردی پیش‌بینی نیاز به تعمیرات                          |
| ۸۴ | اقدامات موردنیاز قبل از شروع پروژه پیش‌بینی نیاز به تعمیرات        |
| ۸۴ | فرآیند علوم داده‌ای با استفاده از سرویس‌های R نرم‌افزار SQL Server |
| ۸۶ | تعریف هدف                                                          |
| ۸۸ | آشنایی با منابع داده‌ای                                            |
| ۹۰ | کاوش داده‌ها                                                       |
| ۹۲ | ایجاد مجموعه داده تحلیلی                                           |

|          |                         |
|----------|-------------------------|
| ۱۰۰..... | ایجاد مدل یادگیری ماشین |
| ۱۰۲..... | ارزیابی و تنظیم مدل     |
| ۱۰۴..... | استقرار مدل             |
| ۱۰۶..... | خلاصه فصل               |

## فصل ششم: پیش‌بینی

|          |                                                                    |
|----------|--------------------------------------------------------------------|
| ۱۰۷..... | مقدمه‌ای بر پیش‌بینی                                               |
| ۱۰۸..... | پیش‌بینی مالی                                                      |
| ۱۰۸..... | پیش‌بینی تقاضا                                                     |
| ۱۰۸..... | پیش‌بینی موجودی                                                    |
| ۱۰۹..... | دقت پیش‌بینی                                                       |
| ۱۰۹..... | ابزارهای پیش‌بینی                                                  |
| ۱۱۰..... | مدل‌های آماری برای پیش‌بینی                                        |
| ۱۱۰..... | تحلیل سری‌های زمانی                                                |
| ۱۱۲..... | پیش‌بینی سری‌های زمانی                                             |
| ۱۱۳..... | پیش‌بینی با استفاده از سرویس‌های ابری: ابزار SQL Server            |
| ۱۱۴..... | بارگذاری داده‌ها در SQL Server                                     |
| ۱۱۵..... | تقسیم داده‌ها به دو بخش آموزش و آزمون                              |
| ۱۱۶..... | آموزش مدل‌های پیش‌بینی سری‌های زمانی و ارزیابی با استفاده از آن‌ها |
| ۱۱۸..... | تولید معیارهای سنجش دقت مدل                                        |
| ۱۱۹..... | خلاصه فصل                                                          |

## پیشگفتار

در دنیای پیرامون ما تقریباً تمامی صنایع و بنگاه‌ها به‌وسیله فناوری متحول شده‌اند. دلیل این تغییر و تحول به سه موضوع زیر برمی‌گردد: انفجار حجم عظیم اطلاعات، هوشمندی که ناشی از فرآیندهای یادگیری ماشین و رویکرد تحلیلی پیشرفته اطلاعات است و در نهایت چابکی و ظرفیت‌های اقتصادی رایانش‌های ابری ...

اگرچه پایگاه‌های داده تقریباً تمامی جنبه‌های کسب‌وکار امروزی را قدرتمند ساخته‌اند ولی اساساً برای موضوعاتی که در بالا بیان شده طراحی نشده‌اند. پایگاه داده‌های سنتی برای ثبت و بازگردانی تراکشن‌هایی نظیر سفارش‌ها و پرداخت‌ها مناسب هستند. آن‌ها برای ایجاد قابلیت اطمینان، امنیت و ... در برنامه‌های کاربردی مرتبط با تراکشن‌ها در مقیاس کوچک تا متوسط، آن هم فقط برای یک مراکز داده مستقر در مکان مشخص، شخص طراحی شده‌اند.

پایگاه داده‌هایی که برای پاسخگویی به چالش‌های امروزی ساخته شده‌اند، علاوه بر تراکشن‌های ثبت شده قادر به تحلیل داده‌های آنلاین در حافله نیز هستند. آن‌ها تأخیر بسیار اندکی در تحلیل‌های پیشرفته یادگیری ماشین مانند مدل‌های پیش‌بینی و پیش‌بینانه روی داده‌های یکسان دارند بنابراین برنامه‌های کاربردی می‌توانند با سادگی هوشمندی مناسبی را در زمینه داده‌ها به‌صورت درون نهاده داشته باشند. در این روش، پایگاه‌های داده می‌توانند به‌صورت سرویسی کاملاً مبتنی بر ابر ارائه شوند که ساخت و انتشار برنامه‌های کاربردی هوشمند SaaS (نرم‌افزار به‌عنوان یک سرویس) را به سادگی ممکن می‌سازد.

این پایگاه‌های داده ویژگی‌های امنیتی خلاقانه‌ای نیز برای مقابله با تهدیدات امروز که بیشتر داده‌ها از طریق اینترنت در دسترس قرار می‌گیرد فراهم کرده‌اند. آن‌ها پشت بانی ۲۴ ساعته در ۷ روز هفته با دسترس‌پذیری بالا، مدیریت مناسب و مدیریت پایگاه داده از طریق رابط‌های مختلف را ارائه می‌کنند. آن‌ها همچنین قابلیت ایجاد و مدیریت برنامه‌های کاربردی هوشمند را هم به‌صورت مبتنی بر ابر و هم در یک محل فیزیکی فراهم می‌کنند. این پایگاه داده‌ها خبر از دنیایی جدید با محیطی هوشمند می‌دهند.

SQL 2016 برای این دنیای جدید و یاری‌رسانی به بنگاه‌ها برای مواجهه با چالش‌های کنونی ایجاد شده است. این نرم‌افزار از پردازش‌های ترکیبی تراکشنی/تحلیلی، رویکرد تحلیلی پیشرفته و تراکشن‌های در حافله با ماندگاری بالا پشتیبانی می‌کند. این نرم‌افزار فرآیندهای تحلیلی پیشرفته را در پایگاه داده‌ها ادغام کرده است و قابلیت‌های انقلابی برای ساخت برنامه‌های کاربردی هوشمند و با کارایی بالا فراهم می‌کند.

تصور کنید برنامه کاربردی اصلی یک شرکت با استفاده از یک پایگاه داده نظیر SQL Server ایجاد شده است. چقدر خوب است اگر بتوانید اقدامات هوشمندی مانند الگوریتم‌های تحلیل پیشرفته، تبدیل داده‌ها درون خود پایگاه داده و هوشمندسازی هر تراکنش به صورت بلادرنگ را به آن بیفزایید. این کار اکنون برای اولین بار با کمک زبان R و فرآیندهای یادگیری ماشین که در SQL Server 2016 ایجاد شده است امکان‌پذیر می‌باشد. با ترکیب فناوری پردازش تحلیلی آنلاین (OLTP) نرم‌افزار SQL Server با زبان R و یادگیری ماشین، برنامه‌های کاربردی به کارایی خارق العاده‌ای در رویکرد تحلیلی و تحلیل دست می‌یابند در حالی که از فوایدی نظیر توان عملیاتی بالا، موزایی سازی عملیات، امنیت، قابلیت اطمینان، گواهی‌های انطباق و قابلیت مدیریت قدرتمند پایگاه داده‌های صنعتی نیز بهره‌مند می‌شوند.

این کتاب برای اولین بار نحوه ایجاد برنامه کاربردی هوشمند با استفاده از SQL Server و زبان R را به شما آموزش خواهد داد. این کتاب محتوایی ارزشمند است که به توسعه‌دهندگان قدرت آزادسازی توان و نیروی انسانی از روتین‌های مبتنی بر داده‌ها را در سازمان‌شان می‌دهد.

## مقدمه

R یکی از قدرتمندترین و محبوب‌ترین زبان‌ها و محیطی‌های برنامه‌نویسی برای تحلیل داده‌ها است که توسط دانشمندان علوم داده‌ای مورد استفاده قرار می‌گیرد. داده‌های عملیاتی بنگاه‌های مختلف عموماً در سیستم‌های مدیریت پایگاه داده رابطه‌ای (RDBMS) ذخیره می‌شوند و یکی از پرکاربردترین RDBMS‌ها Microsoft SQL Server است. فراتر از یک سرور پایگاه داده بوده و اکوسیستمی غنی با قابلیت‌های تحلیلی پیشرفته دارد. سرویس زبان R نرم‌افزار SQL Server با ادغام محیط این زبان در خود، اجازه تعامل بین داده‌های روی RDBMS و زبان R را می‌دهد، در عین حال، امنیت و سلامت RDBMS نیز حفظ خواهد شد. در این کتاب چگونگی ادغام این دو محیط توسط مایکروسافت در یکدیگر و نحوه بهره‌گیری از این قابلیت جدید را خواهید آموخت. همچنین با مثال‌هایی عملی و کاربردی از نحوه استفاده از سرویس زبان R نرم‌افزار SQL Server برای ایجاد راه‌حل‌هایی برای مسائل دنیای واقعی روبه‌رو خواهید شد.

## ساختار این کتاب

کتابی که در دست شماست به سه بخش اصلی تقسیم شده است: مقدمه‌ای بر سرویس زبان R نرم‌افزار SQL Server و آموزشی در مورد کلیت SQL Server، شرح و توضیح نحوه کار دانشمندان علوم داده‌ای در این محیط جدید و مثال‌هایی عملی از مسائل دنیای واقعی. خوانندگان این کتاب می‌توانند این مثال‌ها را مرور و بررسی کنند و یا همراه با کتاب روی آن‌ها به صورت عملی کار کنند.

## این کتاب برای چه افرادی مناسب است

مخاطبین این کتاب افراد فنی (به‌ویژه دانشمندان علوم داده‌ای) هستند و فرض بر این است که با زبان و محیط برنامه‌نویسی R آشنایی دارند. گرچه ما به‌صورت مختصر علوم داده‌ای و زبان R را معرفی می‌کنیم، ولی منابع فراوانی برای مطالعه‌کنندگان برای فراگیری این موضوعات وجود دارد. این کتاب همچنین برای مدیران پایگاه داده‌ها، توسعه‌دهندگان نرم‌افزاری و سایر کارشناسان داده‌ها مناسب است.

اگرچه ما تمامی مباحث مرتبط با SQL Server را در این کتاب پوشش نمی‌دهیم ولی منابعی در این زمینه معرفی خواهند شد و برخی مفاهیم در بخش‌هایی از SQL Server که شاید شما با آن آشنایی نداشته باشید (که عموماً در موضوعات علوم داده‌ای است) ارائه شده است.

## تقدیر و تشکر

از Xibin Gao و Hang Zhang, Gopi Kumar, Fang Zhao, Brad Severtson که در توسعه و نگارش محتوای فصول ۳ و ۴ ما را با این رساننده‌ها متشکریم.

## کتاب‌های الکترونیکی رایگان مایکروسافت

کتاب‌های رایگان مایکروسافت از بررسی‌های تکنیکال گرفته تا ارائه اطلاعاتی عمیق و غنی در موضوعات ویژه، گستره‌ای عظیمی از موضوعات را پوشش می‌دهند. این کتاب‌های الکترونیکی به‌صورت EPUB, PDF و فرمت Mobi برای Kindle در آدرس <http://aka.ms/spreefree> در دسترس شما برای دانلود است. هرچند مدت یک بار برای بررسی مطالب جدید این سایت بررسی کنید.

## غلطنامه، به‌روزرسانی و پشتیبانی کتاب

ما تمام تلاش خود را برای اطمینان از دقت مطالب این کتاب به کار گرفتیم. شما می‌توانید با مراجعه به آدرس <https://aka.ms/IntroSQLServerR/errata> به به‌روزرسانی‌های این کتاب (به‌صورت لیستی از اشتباهات ارسال شده و اصلاحات صورت گرفته) دسترسی یابید. اگر با خطایی مواجه شدید که در حال حاضر در لیست موجود نیست لطفاً آن را به همان آدرس برای ما ارسال کنید.

اگر به پشتیبانی بیشتری نیاز دارید، از طریق آدرس ایمیل [mspinput@microsoft.com](mailto:mspinput@microsoft.com) مربوط به پشتیبانی کتاب‌های مایکروسافت با ما در ارتباط باشید.



لطفاً توجه کنید که پشتیبانی از محصولات سخت‌افزاری و نرم‌افزاری میکروسافت در آدرس بیان شده صورت نمی‌گیرد. برای بهره‌گیری از پشتیبانی سخت‌افزاری و نرم‌افزاری میکروسافت به آدرس <http://support.microsoft.com> مراجعه کنید.

## ما می‌خواهیم نظرات شما را بشنویم

در بخش انتشارات میکروسافت، رضایت شما اولویت اصلی ما و بازخوردهای شما ارزشمندترین دارایی ما است. لطفاً نظر خود را در مورد این کتاب از طریق آدرس <http://aka.ms/tellpress> به ما بیان کنید. نظرسنجی مورد در آدرس فوق بسیار کوتاه است و ما تک تک نظرات و ایده‌های شما را می‌خوانیم. پیشاپیش از نظرات شما سپاس‌گذاریم...

## ارتباط مستمر با ما

بیایید به گفتگوی خود هم ادامه دهیم! ما در Twitter به آدرس <http://twitter.com/MicrosoftPress> منتظر شما هستیم...