



مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

# تجزیه و تحلیل داده‌ها در ابر

مدل‌ها، تکنیک‌ها و کاربردها

نویسندگان: دومینیکو تالیا، پائولو ترانفیو، فابریزیو ماروزو

مترجم: دکتر بهاره پهلوان‌زاده

عضو هیأت علمی مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

سرشناسه	: تالیا، دومنیکو Talia, Domenico
عنوان و نام پدیدآور	: تجزیه و تحلیل داده‌ها در ابر: مدل‌ها، تکنیک‌ها و کاربردها/ نویسندگان دومنیکو تالیا، پائولو ترانفیو، فابریزیو ماروزو؛ ترجمه بهاره پهلوان‌زاده.
مشخصات نشر	: شیراز: مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری، انتشارات، ۱۴۰۰.
مشخصات ظاهری	: ۱۴۴ ص: جدول، مصور (بخشی رنگی).
شابک	: ۹۷۸-۶۲۲-۹۷۷۵۹-۶-۷
وضعیت فهرست نویسی	: فیپا
یادداشت	: عنوان اصلی: Data analysis in the Coloud
یادداشت	: کتابنامه.
موضوع	: رایانش ابری
موضوع	: Cloud Computing
موضوع	: داده‌کاوی
موضوع	: Data mining
شناسه افزوده	: ترانفیو، پائولو
شناسه افزوده	: Trunfio, Paolo
شناسه افزوده	: ماروتسو، فابریزیو
شناسه افزوده	: Marozzo, Fabrizio
شناسه افزوده	: پهلوان‌زاده، بهاره، ۱۳۶۱-، مترجم
شناسه افزوده	: ایران. وزارت علوم، تحقیقات و فناوری. مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری
شناسه افزوده	: Iran. Ministry of Science, Research and Technology Regional information center for science and technology
رده بندی کنگره	: ۵۸۵/۷۶QA
رده بندی دیویی	: ۶۷۸۲/۰۰۴
شماره کتابشناسی ملی	: ۷۶۴۰۱۵۴
وضعیت رکورد	: فیپا



انتشارات مرکز منطقه‌ای  
اطلاع‌رسانی علوم و فناوری

عنوان: تجزیه و تحلیل داده‌ها در ابر (مدل‌ها تکنیک‌ها و کاربردها)

مترجم: دکتر بهاره پهلوان‌زاده

ناشر: انتشارات مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

شابک: ۹۷۸-۶۲۲-۹۷۷۵۹-۶-۷

تیراژ: ۲۰۰ نسخه

اداره انتشارات مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

شیراز بلوار جمهوری اسلامی خیابان جام جم Publication@ricest.ac.ir

تلفن: ۰۷۱۳۶۴۶۸۴۵۲

لقد علم به

خانواده و پشتیبان همیشگی عزیزم  
کلیه‌ی مدرسان، دانشجویان و پژوهشگران حوزه‌ی رایانش ابری،  
تجزیه و تحلیل داده‌های بزرگ و داده‌کاوی

بهاره پهلوان‌زاده

RICES

RICEST

## فهرست

- پیش‌گفتار مترجم ..... ۱  
پیشگفتار نویسندگان ..... ۳

### فصل ۱

- مقدمه ای بر داده‌کاوی ..... ۷  
۱.۱ مفاهیم داده‌کاوی ..... ۷  
۱.۱.۱ طبقه‌بندی ..... ۱۱  
۱.۱.۲ خوشه بندی ..... ۱۶  
۱.۱.۳ قوانین انجمنی ..... ۲۲  
۱.۲ داده‌کاوی موازی و توزیع شده ..... ۲۵  
۱.۲.۱ طبقه‌بندی موازی ..... ۲۶  
۱.۲.۲ خوشه‌بندی موازی ..... ۲۷  
۱.۲.۳ موازی‌سازی در قوانین انجمنی ..... ۲۹  
۱.۲.۴ داده‌کاوی توزیع شده ..... ۳۰  
۱.۳ خلاصه ی فصل ..... ۳۴  
مراجع ..... ۳۶

### فصل ۲

- مقدمه ای بر رایانش ابری ..... ۳۹  
۲.۱ رایانش ابری: تعریف، مدل‌ها و معماری‌ها ..... ۳۹  
۲.۱.۱ مدل‌های سرویس ..... ۴۱  
۲.۱.۲ مدل‌های استقرار ..... ۴۲  
۲.۱.۳ محیط‌های رایانش ابری ..... ۴۶  
۲.۲ سیستم‌های رایانش ابری برای کاربردهای داده حجیم ..... ۵۲  
۲.۲.۱ نیازمندی‌های عملکردی ..... ۵۲  
۲.۲.۲ نیازمندی‌های غیرعملکردی ..... ۵۴  
۲.۲.۳ مدل‌های رایانش ابری برای تجزیه و تحلیل داده‌های توزیع شده ..... ۵۶  
۲.۳ خلاصه ی فصل ..... ۵۹

مراجع ..... ۵۹

### فصل ۳

مدل‌ها و تکنیک‌های تجزیه و تحلیل داده‌ها مبتنی بر رایانش ابری.....	۶۱
۳.۱ MapReduce برای تجزیه و تحلیل داده‌ها.....	۶۱
۳.۱.۱ الگوی.....	۶۳
۳.۱.۲ چارچوب‌های.....	۶۵
۳.۱.۳ الگوریتم‌ها و برنامه‌های MapReduce.....	۶۸
۳.۲ جریان‌کاری‌های تجزیه و تحلیل داده.....	۷۰
۳.۲.۱ برنامه‌نویسی جریان‌کاری.....	۷۲
۳.۲.۲ سیستم‌های مدیریت جریان‌کار.....	۷۷
۳.۲.۳ سیستم‌های مدیریت جریان‌کار برای ابرها.....	۷۸
۳.۳ مدل‌های NOSQL برای تجزیه و تحلیل داده‌ها.....	۸۲
۳.۳.۱ ویژگی‌های اصلی NoSQL.....	۸۳
۳.۳.۲ طبقه‌بندی پایگاه داده‌های NoSQL.....	۸۴
۳.۳.۳ سیستم‌های NoSQL.....	۸۶
۳.۳.۴ موارد استفاده.....	۹۰
۳.۴ خلاصه‌ی فصل.....	۹۲
مراجع.....	۹۳

### فصل ۴

طراحی و پشتیبانی تجزیه و تحلیل داده‌های مقیاس‌پذیر.....	۱۰۱
۴.۱ سیستم تجزیه و تحلیل داده‌ها برای رایانش ابری.....	۱۰۱
۴.۱.۱ سیستم Pegasus.....	۱۰۲
۴.۱.۲ زبان برنامه‌نویسی Swift.....	۱۰۳
۴.۱.۳ بستر Hunk.....	۱۰۴
۴.۱.۴ Sector / Sphere.....	۱۰۵
۴.۱.۵ BigML.....	۱۰۷

۱۰۸	..... بستر تجزیه و تحلیلی	۴.۱.۶
۱۰۹	..... Mahout چارچوب	۴.۱.۷
۱۱۰	..... Spark چارچوب	۴.۱.۸
۱۱۱	..... Microsoft Azure Machine Learning سرویس	۴.۱.۹
۱۱۲	..... ClowdFlows بستر	۴.۱.۱۰
۱۱۳	..... شیوهی طراحی یک چارچوب مقیاس پذیر تجزیه و تحلیل داده در رایانش ابری	۴.۲
۱۱۴	..... معماری و مکانیزم های اجرایی	۴.۲.۱
۱۱۶	..... پیاده سازی در مایکروسافت اژر	۴.۲.۲
۱۱۸	..... برنامه نویسی تجزیه و تحلیل داده ها مبتنی بر جریانکار	۴.۳
۱۲۱	..... VL4Cloud	۴.۳.۱
۱۲۶	..... JS4Cloud	۴.۳.۲
۱۲۹	..... الگوهای جریانکار در	۴.۳.۳
۱۳۸	..... مطالعات موردی تجزیه و تحلیل داده ها	۴.۴
۱۳۸	..... جریانکار مسیرکاوی با استفاده از VL4Cloud	۴.۴.۱
۱۴۱	..... جریانکار یادگیری گروهی با استفاده از JS4Cloud	۴.۴.۲
۱۴۶	..... طبقه بندی موازی با استفاده از MapReduce	۴.۴.۳
۱۴۸	..... طبقه بندی موازی با استفاده از	۴.۴.۴
۱۵۳	..... مراجع	

## فصل ۵

۱۵۷	..... روندهای پژوهش در تجزیه و تحلیل داده های بزرگ	
۱۵۷	..... محاسبات داده فشرده	۵.۱
۱۵۹	..... مقیاس پذیری exascale در تجزیه و تحلیل داده ها	۵.۱.۱
۱۶۰	..... مسائل برنامه نویسی برای تجزیه و تحلیل exascale داده ها	۵.۱.۲
۱۶۵	..... تجزیه و تحلیل عظیم شبکه ی اجتماعی	۵.۲
۱۶۹	..... حوزه های پژوهشی کلیدی	۵.۳
۱۷۱	..... خلاصه ی فصل	۵.۴
۱۷۳	..... مراجع	

RICEST