

۲۰۷۴۵۷۴
۹۸۱۰۳۱



مؤسسه فرهنگی هنری
دیباگران تهران

به نام خدا



مؤسسه فرهنگی هنری
دیباگران تهران

ذخیره و بازیابی اطلاعات مشارکت و رویکردها

مؤلف

دکتر علی احسانی

(دکترای مدیریت فناوری اطلاعات)



هرگونه چاپ و تکثیر از محتویات این کتاب بدون اجازه کتبی ناشر ممنوع است. متخلفان به موجب قانون حمایت حقوق مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می گیرند.

«عنوان کتاب: ذخیره و بازیابی اطلاعات مفاهیم و رویکردها»

«مؤلف: دکتر علی بانی»

«ناشر: موسسه رهنمای دیبگران تهران»

«صفحه آرای: فریوش عبدالله»

«طراح جلد: داریوش فرسای»

«نوبت چاپ: اول»

«تاریخ نشر: ۱۳۹۸»

«چاپ و صحافی: درج عقیق»

«تیراژ: ۱۰۰ جلد»

«قیمت: ۷۵۰۰۰۰ ریال»

«شابک: ۹۷۸-۶۲۲-۲۱۸-۲۴۳-۴»

«نشانی واحد فروش: تهران، میدان انقلاب،

خیابان کارگر جنوبی، روبروی پاساژ مهستان،

پلاک ۱۲۵۱

«تلفن: ۰۲۰۸۵۱۱۱-۶۶۴۱۰۰۴۶»

«فروشگاههای اینترنتی دیبگران تهران :

WWW.MFTBOOK.IR

www.dibbook.ir

www.dibagaran-tehran.com

«نشانی تلگرام: @mftbook» «نشانی اینستاگرام دیبیا dibagaran_publishing»

«هر کتاب دیبگران، یک فرصت جدید شغلی.»

«هرگوشی همراه، یک فروشگاه کتاب دیبگران تهران.»

«از طریق سایتهای و اپ دیبگران، در هر جای ایران به کتابهای ما دسترسی دارید.»

سرشناسه: احسانی، علی، ۱۳۶۶-

عنوان و نام پدیدآور: ذخیره و بازیابی اطلاعات، مفاهیم و رویکردها

مشخصات نشر: تهران: دیبگران تهران: ۱۳۹۸

مشخصات ظاهری: ۲۸۴ص: مصور،

شابک: ۹۷۸-۶۲۲-۲۱۸-۲۴۳-۴

وضعیت فهرست نویسی: فیا

موضوع: ذخیره و بازیابی اطلاعات

عنوان: Information storage and retrieval system

موضوع: حافظه داده ها

موضوع: Data collection

موضوع: کامپیوترها- یادگیری ماشین

موضوع: Memory management (computer science)

رده بندی کنگره: ۷۶/۹-Q

رده بندی دیویی: ۰۰۵/۷۴

شماره کتابشناسی ملی: ۶۰۲۸۵۶۹

فهرست مطالب

بخش اول : مفاهیم و اصطلاحات ذخیره و بازیابی ۱۰

فصل اول : رسانه های ذخیره سازی ۱۱

۱-۱- سیستم کامپیوتری ۱۱

۲-۱- حافظه ۱۲

۳-۱- نوار مغناطیسی ۱۳

۴-۱- گپ (GAP) ۱۳

۵-۱- دیسک ساطعی ۱۴

۶-۱- تقسیمات دیسک ۱۵

۷-۱- پارامترهای دیسک ۱۷

۸-۱- طبله (DRUM) ۱۹

فصل دوم: سیستم فایل و بلاک ۲۰

۱-۲- سیستم فایل ۲۰

۲-۲- قالب های رکورد منطقی ۲۱

۳-۲- اشاره گر (نشانه رو) ۲۲

۴-۲- فایل ۲۳

۵-۲- ساختار فایل ۲۳

۶-۲- بلاک بندی ۲۴

۷-۲- فایل در محیط فیزیکی ۳۰

۸-۲- تکنیکهای تهیه پشتیبان ۳۲

۹-۲- سطوح نشانی دهی ۳۴

۱۰-۲- بافر ۳۷

فصل سوم: پارامترهای ارزیابی رسانه ها ۴۱

۱-۳- ظرفیت واقعی نوار ۴۱

۲-۳- محاسبه نرخ انتقال واقعی نوار ۴۱

۴۴ ۳-۳ محاسبه زمان کل پردازش فایل

۴۶ ۳-۴ تکنیک های بهبود کارایی سیستم فایل

۴۷ ۳-۵ توزیع فایل روی چند دیسک

بخش دوم: ساختار فایل ۴۹

..... فصل چهارم: ساختارهای فایل (پایل-ترتیبی) ۵۰

۵۰ ۴-۱ انواع ساختار های فایل

۵۱ ۴-۲ انواع فایل

۵۲ ۴-۳ شرح اصول عملیات ششگانه

۵۵ ۴-۴ فایل با ساختار پایل (بی نظم)

۵۷ ۴-۵ معرفی ساختار ترتیبی

..... فصل پنجم: ساختارهای شاخص دار ۶۴

۶۴ ۵-۱ معرفی ساختار های شاخص دار

۶۴ ۵-۲ انواع شاخص

۶۶ ۵-۳ ساختار ترتیبی شاخص دار (INDEXED SEQUENTIAL)

۷۳ ۵-۴ ساختار چند شاخصی (فایل شاخص دار)

..... فصل ششم: ساختار فایل مستقیم ۷۷

۷۷ ۶-۱ معرفی ساختار مستقیم

۷۷ ۶-۲ روش های حل مشکل تصادف و درج سرریزی ها

۷۹ ۶-۳ موارد استفاده فایل مستقیم

۸۱ ۶-۴ پاکت بندی

بخش سوم: مفاهیم تکمیلی در ذخیره سازی اطلاعات ۸۳

..... فصل هفتم: آشنایی با ذخیره اطلاعات ۸۴

۸۵ ۷-۱ ذخیره اطلاعات

۹۲ ۷-۲ تکامل معماری ذخیره سازی

۹۳ ۷-۳ زیرساخت مرکز داده

۹۷ ۷-۴ مجازی سازی و رایانش ابری

فصل هشتم: مرکز داده ۹۹

- ۱-۸- برنامه کاربردی ۹۹
- ۲-۸- سیستم مدیریت پایگاه داده ۱۰۰
- ۳-۸- میزبان (محاسب) ۱۰۰
- ۴-۸- اتصال ۱۰۸
- ۵-۸- ذخیره سازی ۱۱۱
- ۶-۸- دسترسی میزبان به داده ها ۱۱۲
- ۷-۸- ذخیره سازی اتصال مستقیم (DAS) ۱۱۴
- ۸-۸- صف فرمان ۱۱۶
- ۹-۸- شنایابی با درایوهای فلش ۱۱۷

فصل نهم: حفاظت از داده ها: RAID ۱۲۰

- ۱-۹- روش های اجرای RAID ۱۲۱
- ۲-۹- مولفه های ارائه RAID ۱۲۲
- ۳-۹- روشهای RAID ۱۲۲
- ۴-۹- سطوح RAID ۱۲۶
- ۵-۹- تأثیر RAID بر عملکرد دیسک ۱۳۳
- ۶-۹- مقایسه RAID ۱۳۵

فصل دهم: سیستم های ذخیره سازی هوشمند ۱۳۷

- ۱-۱۰- مؤلفه های یک سیستم ذخیره سازی هوشمند ۱۳۷
- ۲-۱۰- تأمین ذخیره ساز ۱۴۴
- ۳-۱۰- انواع سیستم های ذخیره سازی هوشمند ۱۵۱

فصل یازدهم: فن آوری های ذخیره سازی تحت شبکه ۱۵۵

- ۱-۱۱- کانال فیبر: بررسی اجمالی ۱۵۶
- ۲-۱۱- SAN و تکامل آن ۱۵۷
- ۳-۱۱- معماری کانال فیبر ۱۵۸
- ۴-۱۱- توپولوژی های FC SAN ۱۶۱
- ۵-۱۱- مجازی سازی در SAN ۱۶۴
- ۶-۱۱- IP SAN ۱۶۸

۱۶۹.....	۷-۱۱- ذخیره سازی متصل به شبکه (NAS)
۱۷۰.....	۸-۱۱- سرورها با اهداف عمومی در مقابل دستگاه های NAS
۱۷۱.....	۹-۱۱- مزایای NAS
۱۷۱.....	۱۰-۱۱- مؤلفه های NAS
۱۷۲.....	۱۱-۱۱- پروتکل به اشتراک گذاری فایل NAS
۱۷۴.....	۱۲-۱۱- عوامل مؤثر بر عملکرد NAS
۱۷۵.....	۱۳-۱۱- مجازی سازی سطح فایل

فصل دوازدهم: رایانش ابری ۱۷۸

۱۷۸.....	۱-۱۲- فن آوری های ایجاد CLOUD
۱۷۹.....	۲-۱۲- ویژگی های رایانش ابری
۱۸۰.....	۳-۱۲- مزایای محاسبات ابری
۱۸۱.....	۴-۱۲- مدل های سرویس ابری
۱۸۳.....	۵-۱۲- مدل استقرار ابر
۱۸۶.....	۶-۱۲- زیرساخت رایانش ابری
۱۹۰.....	۷-۱۲- چالش های ابری
۱۹۱.....	۸-۱۲- ملاحظات پذیرش ابر

فصل سیزدهم: ایمن سازی زیرساخت های ذخیره سازی ۱۹۴

۱۹۴.....	۱-۱۳- چارچوب امنیت اطلاعات
۱۹۵.....	۲-۱۳- ریسک سه گانه
۱۹۹.....	۳-۱۳- دامنه های امنیتی ذخیره سازی
۲۰۸.....	۴-۱۳- ایمن سازی زیرساخت های ذخیره سازی در محیط های مجازی و ابری

بخش چهارم: بازیابی اطلاعات ۲۱۳

فصل چهاردهم: الگوی معماری بانک های اطلاعاتی و NOSQL ۲۱۴

۲۱۴.....	۱-۱۴- وضعیت بانک های اطلاعاتی از گذشته تا کنون
۲۱۶.....	۲-۱۴- الگوهای معماری داده NOSQL

فصل پانزدهم: آشنایی با سیستم های بازیابی اطلاعات ۲۴۳

۲۴۴.....	۱-۱۵- تعریف سیستم بازیابی اطلاعات
۲۴۶.....	۲-۱۵- اهداف سیستم های بازیابی اطلاعات

۲۴۶.....	۱۵-۳- روش ها و معیارهای ارزیابی الگوریتم های بازیابی اطلاعات
۲۴۹.....	۱۵-۴- جستجوی اطلاعات
۲۵۳.....	۱۵-۵- پردازش زبان طبیعی (NLP)
۲۵۷.....	۱۵-۶- متن کاوی (TEXT MINING)
۲۶۳.....	۱۵-۷- انواع ایندکس یا اندیس در پایگاه داده (INDEX)

فصل شانزدهم: کلان داده ها ۲۷۰

۲۷۰.....	۱۶-۱- خصوصیات کلان داده
۲۷۲.....	۱۶-۲- چالش های حوزه کلان داده
۲۷۳.....	۱۶-۳- فناوری های مرتبط با کلان داده
۲۸۰.....	۱۶-۴- یادگیری عمیق

www.ketab.ir

مقدمه مولف:

بار الها سپاس بیکران تو را، که به ما توفیق خدمتگزاری به نسل جوان و آینده ساز ایران اسلامی را عطا فرمودی و همواره چتر لطف و عنایت بی نهایت تو بر ما، سایه مهر خود را گسترده است.

پیشرفت فناوری، گسترش اینترنت و استفاده از خدمات الکترونیکی و فناوری اطلاعات منجر به تولید حجم انبوهی از داده ها و اطلاعات در دامنه ها و زمینه های گوناگون مانند پزشکی، علمی، تجارت، نجوم و غیره شده است. با توجه به این حجم گسترده و رو به رشد، نمی توان آن ها را به وسیله پایگاه داده ها و ابزارهای سنتی پردازش و تحلیل کرد و نیازمند معماری جدید برای گردآوری، انتقال، ذخیره سازی و پردازش داده ها خواهیم بود. در کتاب حاضر معطوف به بحث ذخیره و بازیابی اطلاعات خواهد بود. فعالیت هایی که برای تحلیل و سازماندهی مدارک و منابع صورت می گیرد، ذخیره اطلاعات، و تلاش هایی که برای یافتن یک یا چند مدرک از میان بوه مدارک ذخیره شده انجام می شود، بازیابی اطلاعات نام دارد. نظام هایی که این جریان ها در آنها روی می دهد نظام های ذخیره و بازیابی اطلاعات خوانده می شوند. کتاب حاضر به گونه ای تالیف گردیده است که بتواند پاسخگوی نیاز تمام کاربران چه در سطح دانشگاهی و چه در سطح کاربردی باشد. فصل بندی کتاب به شکلی صورت پذیرفته است که بیان کننده سیر رشد فناوری های ذخیره و بازیابی اطلاعات باشد. فصول ابتدایی با بیان سیستم های ذخیره اطلاعات و فایده سیستم ها آغاز می گردد و این رشد تا فصول انتهایی و بخش کلان داده ها ادامه خواهد یافت. کتاب حاضر برای دانشجویان رشته فناوری اطلاعات، مهندسی کامپیوتر و علوم کامپیوتر و تمامی علاقه مندان این حوزه چه در سطح دانشگاهی و چه سطح عمومی مفید می باشد اما مطالعه این کتاب برای خوانندگانی که با مفاهیم و رویکردهای جدید بازیابی اطلاعات آشنایی داشته و قصد مطالعه این مباحث را در سطح پیشرفته دارند توصیه نمی گردد.

کتاب فوق الذکر پس از چندین سال بررسی، پژوهش و تدوین در اینجانب در زمینه ذخیره و بازیابی اطلاعات و مطالعه و تدریس درس ذخیره و بازیابی اطلاعات در دانشگاه تبریز به نظر می رسد. محضر اساتید بزرگی که شاید در این مختصر جای بیان تک تک اسامی این عزیزان مقدور نباشد، صورت پذیرفته است. در انتها جای دارد از مدیریت محترم انتشارات مجتمع فنی تهران و تمامی همکارانی که در گردآوری و تدوین کتاب بنده را یاری نمودند قدردانی نمایم.

امید است این اثر بتواند نقشی هر چند کوچک در خدمت به جامعه تعلیم و تربیت کشور داشته باشد.