

رایانش ابری

مؤلفین:
دکتر مهدی خلیلی
مهندس مهسان نظری
اعضای هیئت علمی دانشگاه پیام نور

بنام کسی که قلم از قلم
انتشارت امید محمد
به انسان عطا کرد فکر و قلم

سرشناسه: خلیلی، مهدی، ۱۳۵۴ -
عنوان و نام پدیدآور: رایانش ابری / مولفین مهدی خلیلی، مهسا نظری.
مشخصات نشر: تهران: امید مجد، ۱۳۹۴.
مشخصات ظاهری: ۲۶۵ ص.
شابک: ۲-۵۵۰-۷۹۱-۹۶۴-۹۷۸:۶۰۰۰ ریال
وضعیت فهرست نویسی: فیا
موضوع: محاسبات ابری
شناسه افزوده: نظری، مهسا، ۱۳۵۵ -
رده بندی کنگره: ۱۳۹۴/QA۷۶ خ ۳/م ۵۸۵
رده بندی دیویی: ۰۰۶/۷۸
شماره کتابشناسی ملی: ۳۹۶۲۲۳۵

عنوان: رایانش ابری
مؤلفین: دکتر مهدی خلیلی، مهندس مهسا نظری
(اعضای هیئت علمی دانشگاه پیام نور)
ناشر: نشر امید مجد
طراح جلد: علیرضا لکزیان
چاپخانه: موسسه نشاط
صحافی: موسسه نشاط
تیراژ: پانصد نسخه
نوبت چاپ: اول، تابستان ۱۳۹۴
قیمت: شش هزار تومان
شابک: ۲-۵۰-۷۹۱۵-۹۶۴-۹۷۸

آدرس: تهران صندوق پستی ۱۴ ۶۶۵۱۵۵ ۸

تلفن: ۶۹ ۰۶۵۵ ۲۲

Www.OmidMajd.Com

بنام کسی کافرید از عدم
انتشارات امید مجد
به انسان عطا کرد و فکر و قلم

فهرست مطالب

صفحه	عنوان
یک	چکیده
الف	مقدمه
۱	فصل اول: رایانش ابری
۲	۱-۱ رایانش ابری چیست؟
۲	۱-۱-۱ مقدمه
۴	۱-۱-۲ دیدگاه‌های مختلف در خصوص رایانش ابری
۷	۱-۱-۳ شناخت بهتر رایانش ابری
۹	۱-۱-۴ مزایا و نقاط قوت محاسبات ابری
۱۳	۱-۱-۵ نقاط ضعف محاسبات ابری
۱۵	۱-۱-۶ چه کسانی می‌توانند از مزایای محاسبات ابری بهره‌مند شوند؟
۱۶	۱-۱-۷ چه کسانی نباید از رایانش ابری استفاده کنند؟
۱۸	۱-۱-۸ تفاوت رایانش ابری با رایانش مشبک (توری)
۲۰	۲-۱ معماری لایه‌های ابری و انواع آن
۲۰	۱-۲-۱ مقدمه
۲۰	۲-۲-۱ معماری ابرها چیست؟
۲۳	۳-۲-۱ لایه‌های رایانش ابری
۲۴	۱-۳-۲-۱ سرویس‌های برنامه‌کاربردی ابری (SaaS)
۲۵	۲-۳-۲-۱ سرویس پلتفرمی (PaaS)

۲۷	۳-۳-۲-۱ سرویس های زیرساخت ابری (IaaS)
۲۸	۴-۲-۱ انواع ابرها در رایانش ابری
۲۹	۱-۴-۲-۱ ابرهای خصوصی
۲۹	۲-۴-۲-۱ ابرهای عمومی
۲۹	۳-۴-۲-۱ ابرهای هیبریدی (آمیخته)
۳۰	۴-۴-۲-۱ ابرهای گروهی
۳۰	۵-۲-۱ معماری سرویس گرا
۳۴	۶-۲-۱ معماری سرویس گرا و ابرها کجا یکدیگر را ملاقات می کنند؟
۳۵	۱-۶-۲-۱ ارتباط رایانش ابری و SOA
۳۶	۷-۲-۱ سبک معماری سرویس گرای سازمانی
۳۸	۸-۲-۱ ارزیابی نقش استانداردهای باز معماری
۳۹	۳-۱ استفاده از ابرها و مدیریت آنها
۳۹	۱-۳-۱ مقدمه
۴۰	۲-۳-۱ اعمال حاکمیت بر ابرها
۴۲	۳-۳-۱ مدیریت ابرها
۴۶	۱-۳-۳-۱ فناوری اداره کردن / مدیریت
۴۸	۲-۳-۳-۱ بایدها و نبایدهای اداره کردن
۴۹	۳-۳-۳-۱ اهمیت اداره کردن سرویس ها
۵۰	۴-۳-۱ مسئولیتهای مدیریت
۵۳	۵-۳-۱ مدیریت چرخه حیات
۵۳	۶-۳-۱ محصولات مدیریت ابرها
۵۶	۷-۳-۱ استانداردهای مدیریت ابر (DMTF)
۵۷	۸-۳-۱ مدیریت استفاده در رایانش ابری

۶۰	۱-۳-۹ مدیریت مصرف در ابرها
۶۹	۱-۳-۱۰ معماری پیشنهادی
۷۳	۱-۳-۱۱ چالش های مدیریتی
۷۷	فصل دوم: داده کاوی و ضرورت آن در محاسبات ابری
۷۸	۲-۱ مقدمه
۷۹	۲-۲ داده کاوی چیست؟
۸۱	۲-۳ خلاصه سازی و به تصویر درآوردن داده ها
۸۲	۲-۴ انبار داده ها
۸۲	۲-۴-۱ تاریخچه و دلایل استفاده از انبار داده
۸۳	۲-۴-۲ مراحل و نحوه ایجاد انبار داده در سازمان
۸۴	۲-۴-۳ عناصر انبارداری
۸۵	۲-۴-۴ سلسله مراتب انبارها (غرفه های داده)
۸۵	۲-۴-۵ معرفی مخزن داده (DW) و ضرورت بکارگیری آن
۸۷	۲-۵ نظریه فازی و داده کاوی
۸۹	۲-۶ کاربردهای داده کاوی
۹۰	۲-۷ پردازش ابری و داده کاوی
۹۳	۲-۸ الگوریتم های داده کاوی
۹۳	۲-۸-۱ الگوریتم وابستگی (Association algorithm)
۹۴	۲-۸-۲ الگوریتم خوشه بندی (Clustering algorithm)
۹۴	۲-۸-۳ الگوریتم درخت تصمیم (Decision Trees Algorithm)
۹۵	۲-۸-۴ الگوریتم رگرسیون خطی (Linear Regression Algorithm)
۹۵	۲-۸-۵ الگوریتم بیز (Naive Bayes Algorithm)
۹۶	۲-۸-۶ الگوریتم شبکه های عصبی (Neural Network Algorithm)

۹۶	۷-۸-۲ الگوریتم رگرسیون منطقی یا لجستیک (Logistic regression)
۹۷	۸-۸-۲ الگوریتم خوشه‌بندی زنجیره‌ای (Sequence Clustering)
۹۷	۹-۸-۲ الگوریتم سری‌های زمانی (Time Series Algorithm)
۹۸	۹-۲ برترین الگوریتم‌های داده کاوی
۹۹	۱-۹-۲ الگوریتم CART
۱۰۱	۲-۹-۲ الگوریتم Apriori
۱۰۲	۳-۹-۲ الگوریتم EM
۱۰۴	۴-۹-۲ الگوریتم KNN
۱۰۵	۵-۹-۲ الگوریتم Naïve Bayes
۱۰۸	۶-۹-۲ الگوریتم K-means:
۱۰۹	۷-۹-۲ الگوریتم AdaBoost:
۱۱۱	۸-۹-۲ الگوریتم C4.5:
۱۱۴	۹-۹-۲ الگوریتم Page Rank
۱۱۶	۱۰-۹-۲ الگوریتم SVM
۱۱۸	فصل سوم: حملات امنیتی به ابرها و راهکارهای مقابله
۱۱۹	۱-۳ حملات امنیتی به ابرها
۱۱۹	۱-۱-۳ مقدمه
۱۲۱	۲-۱-۳ امنیت ابرها
۱۳۲	۳-۱-۳ معرفی یک ساختار امنیتی
۱۳۳	۲-۳ سیستم‌های تشخیص نفوذ
۱۳۳	۱-۲-۳ نگاهی بر سیستم‌های تشخیص نفوذ (IDS)
۱۳۴	۲-۲-۳ انواع روش‌های تشخیص نفوذ
۱۳۹	۳-۲-۳ معماری سیستم‌های تشخیص نفوذ

۱۴۳	۴-۲-۳ روش‌های برخورد و پاسخ به نفوذ
۱۴۴	۵-۲-۳ اجزای سیستم‌های تشخیص نفوذ
۱۴۶	۳-۳-۳ ارزیابی سیستم‌های تشخیص نفوذ
۱۵۰	فصل چهارم: سیستم‌های تشخیص نفوذ داده کاو محور
۱۵۱	۴-۱ مقدمه
۱۵۱	۴-۲ تعریف مختصری از داده کاوی
۱۵۷	۴-۳-۳ تکنیک‌های داده کاوی
۱۵۷	۴-۳-۱ قوانین انجمنی
۱۶۵	۴-۳-۲ درختان تصمیم‌گیری
۱۷۱	۴-۳-۳ ماشین بردار پشتیبان
۱۷۸	۴-۳-۴ خوشه‌بندی
۱۸۱	۴-۴ مدل‌های تشخیص نفوذ
۱۸۲	۴-۴-۱ تشخیص نفوذ معماری پایگاه داده محور (DAID)
۱۸۳	۴-۴-۲ استخراج و تبدیل و بارگذاری (ETL)
۱۹۶	فصل پنجم: تکنیک‌های خوشه‌بندی داده کاو محور و تشخیص نفوذ
۱۹۷	۵-۱ مقدمه
۱۹۸	۵-۲ داده‌کاوی، KDD و زمینه‌های مرتبط
۱۹۸	۵-۳ تکنیک‌های رایج تشخیص نفوذ
۱۹۹	۵-۴ داده‌کاوی و تشخیص نفوذ
۱۹۹	۵-۴-۱ تشخیص ناهنجاری یا تطبیق نمودار
۲۰۰	۵-۴-۲ تشخیص مخرب یا تطبیق امضا
۲۰۱	۵-۵ موانع سیستم‌های تشخیص نفوذ
۲۰۲	۵-۶ بازبینی تکنیک‌های عملی

۲۰۳	۵-۶-۱ تکنیک‌های رده‌بندی
۲۰۶	۵-۶-۲ تکنیک‌های خوشه‌بندی
۲۰۷	۵-۶-۳ تکنیک‌های آماری
۲۰۷	۵-۷-۷ بررسی اجمالی الگوریتم‌های داده‌کاوی برای سیستم تشخیص نفوذ
۲۰۸	۵-۷-۱ نفوذ و سیستم تشخیص نفوذ
۲۱۰	۵-۷-۲ IDS براساس هدف
۲۱۰	۵-۷-۳ طرز کار سیستم تشخیص نفوذ
۲۱۱	۵-۸ تکنولوژی داده‌کاوی و سیستم تشخیص نفوذ
۲۱۴	۵-۹ الگوریتم‌های داده‌کاوی مختلف برای تشخیص نفوذ
۲۲۳	۵-۱۰ نیاز به داده‌کاوی در تشخیص نفوذ
۲۲۴	۵-۱۱ تکنیک‌های داده‌کاوی برای سیستم‌های تشخیص نفوذ
۲۲۴	۵-۱۱-۱ رده‌بندی
۲۲۴	۵-۱۱-۲ قواعد وابستگی
۲۲۵	۵-۱۱-۳ خوشه‌بندی
۲۲۶	۵-۱۲ تکنیک‌های خوشه‌بندی به کاررفته در سیستم تشخیص نفوذ
۲۲۷	۵-۱۲-۱ خوشه‌بندی K-Means
۲۳۰	۵-۱۲-۲ خوشه‌بندی Y-Means
۲۳۵	۵-۱۲-۳ خوشه‌بندی Fuzzy C-Means
۲۳۹	۵-۱۲-۴ مقایسه خوشه‌بندی‌های K-Means, Y-Means و Fuzzy C-Means
۲۴۲	فصل ششم: بحث و نتیجه‌گیری

با گسترش روز افزون تبادل اطلاعات و استفاده از سیستم‌های بر خط، میزان حملات و نفوذ در سیستم‌های اطلاعاتی افزایش یافته است. رایانش ابری گام بعدی تکامل سرویس‌های فناوری اطلاعات بر حسب تقاضا می‌باشد. به منظور جلوگیری از سوءاستفاده از اطلاعات و رخنه در سیستم‌های حساس و مهم باید سیستم‌های مذکور امن شوند.

با گسترش سیستم‌های پایگاهی و حجم بالای داده‌های ذخیره شده در این سیستم‌ها، به ابزاری نیاز است تا بتوان این داده‌ها را پردازش کرده و اطلاعات حاصل از آن را در اختیار کاربران قرار داد. یکی از روش‌های بسیار مهمی که با آن می‌توان الگوهای مفیدی را در میان داده‌ها تشخیص داد، داده‌کاوی است، این روش که با حداقل دخالت کاربران همراه است اطلاعاتی را در اختیار آنها و تحلیل گران قرار می‌دهد تا براساس آنها تصمیمات مهم و حیاتی را در سازمانشان اتخاذ نمایند.

در امنیت اطلاعات، کار عملیات تشخیص، تشخیص نفوذ است که برای امکان شناسایی یکپارچگی، قابلیت اطمینان یا قابلیت دسترسی به منابع تلاش می‌کند. بنابراین، این فرآیند بازبینی و تحلیل وقایع به وجودآمده در یک سیستم رایانه به منظور تشخیص مشکلات امنیتی می‌باشد.

در این کتاب، پس از تشریح تکنیک‌های مختلف خوشه‌بندی داده‌کاوی محور و تشخیص نفوذ، بطور اجمالی الگوریتم‌های داده‌کاوی برای سیستم تشخیص نفوذ را بیان نموده و در پایان، الگوریتم‌های داده‌کاوی خوشه‌بندی K-Means، خوشه‌بندی Y-Means و خوشه‌بندی Fuzzy C-Means را به منظور ارتقاء سیستم‌های تشخیص نفوذ و افزایش سطح امنیت اطلاعات در ابرها پیشنهاد می‌نماییم.

❖ **کلمات کلیدی:** رایانش ابری- حملات امنیتی به ابر- روش‌های تشخیص نفوذ به ابر-

سیستم‌های تشخیص نفوذ- داده‌کاوی- سیستم‌های تشخیص نفوذ داده‌کاوی محور- الگوریتم K-

Means- الگوریتم Y-Means- الگوریتم Fuzzy C-Means.

رایانش ابری، به گونه‌ای از سیستم‌های توزیع شده و موازی اطلاق می‌گردد که مجموعه‌ای از کامپیوترهای مجازی را که به یکدیگر متصل هستند شامل می‌شود. این کامپیوترها به طور پویا عرضه شده و به عنوان یک یا چند منبع محاسباتی یکپارچه بر اساس توافقات سطح سرویس دیده می‌شوند و این توافقات در طول مذاکرات سرویس دهندگان و مصرف کنندگان برقرار می‌گردند. رایانش ابری سعی دارد ایجاد پویای نسل جدیدی از مراکز داده ای را، با ارائه کردن سرویس‌ها و خدمات در ماشین های مجازی شبکه شده به صورت پویا، به گونه ای ممکن سازد که کاربران بتوانند از هر جایی از دنیا به برنامه‌های کاربردی دسترسی داشته باشند. واضح است که رایانش ابری گام بعدی تکامل سرویس های فناوری اطلاعات برحسب تقاضا می‌باشد. با این حال مسئله امنیت در رایانش ابری یکی از مسائل پیچیده به شمار رفته که تمامی سه لایه ابر، مسئولیت هایی که بین کاربران و ارائه دهندگان تقسیم می‌شود و حتی شخص ثالث را درگیر می‌کند. مشکلات امنیتی به عنوان یک مانع بزرگ در مقابل استفاده کاربران از سیستم های رایانش ابری قلمداد می‌شود.

سیر تکاملی پردازش ابری به گونه‌ای است که می‌توان آن را پس از آب، برق، گاز تلفن به عنوان عنصر اساسی پنجم فرض نمود. در چنین حالتی، کاربران سعی می‌کنند بر اساس نیازهایشان و بدون توجه به اینکه یک سرویس در کجا قرار دارد و یا چگونه تحویل داده میشود، به آن دسترسی یابند. نمونه های متنوعی از سیستم های محاسباتی ارائه شده است که سعی دارند چنین خدماتی را به کاربران ارائه دهند. برخی از آنها عبارتند از: محاسبات کلاستری، محاسبات توری و اخیراً رایانش ابری.

در پایین ترین سطح که لایه زیرساخت به عنوان سرویس (IaaS) نامیده می‌شود، پردازنده، حافظه و اجزای سخت افزاری توسط فراهم کننده سرویس ابری ارائه می‌گردد. لایه میانی یا پلتفرم به عنوان سرویس (PaaS) میزبان محیط های مختلف برای ارائه خدمات می‌باشد. در آخر بالاترین لایه نرم افزار کاربردی به عنوان سرویس (SaaS) است که دسترسی به این خدمات از طریق وب سرویس ها و مرورگرهای وب صورت می‌گیرد. Google App, Amazon Ec2 و Salesforce به ترتیب نمونه های لایه های ارائه شده رایانش ابری می‌باشند.

رایانش (پردازش/محاسبات) ابری ساختاری شبیه یک توده ابر دارد که به واسطه آن کاربران می توانند به برنامه های کاربردی از هر جایی از دنیا دسترسی داشته باشند. بنابراین، رایانش ابری می تواند با کمک ماشین های مجازی شبکه شده، به عنوان یک روش جدید برای ایجاد پویای نسل جدید مراکز داده مورد توجه قرار گیرد. بدین ترتیب، دنیای محاسبات به سرعت به سمت توسعه نرم افزارهایی پیش می رود که به جای اجرا بر روی کامپیوترهای منفرد، به عنوان یک سرویس در دسترس میلیون ها مصرف کننده قرار می گیرند. بررسی ها نشان می دهد که، محاسبات کلاستری در حال حاضر نسبت به دو مورد دیگر مقبولیت کمتری دارد، محاسبات توری در رتبه دوم قرار گرفته است، و پردازش انبوه با فاصله زیادی از آنها در حال افزایش جلب توجه افراد بیشتری است.

با گسترش روز افزون تبادل اطلاعات و استفاده از سیستم های بر خط، میزان حملات و نفوذ در سیستم های اطلاعاتی افزایش یافته است. به منظور جلوگیری از سوءاستفاده از اطلاعات و رخنه در سیستم های حساس و مهم باید سیستم های مذکور امن شوند. مفهوم امنیت به معنای محافظت از محرمانگی، یکپارچگی و دسترس پذیری است. اما محرمانگی، یکپارچگی و دسترس پذیری چیست؟

محرمانه بودن داده ها به معنای این است که داده های در حال انتقال در شبکه تنها باید توسط افرادی که به شیوه مناسب احراز هویت شده اند مورد دستیابی قرار بگیرد. حفظ جامعیت داده به این معنی است که داده ها از لحظه ای که منتقل می شوند تا لحظه ای که دریافت می شوند نباید هیچ اختلال و فقدانیه چه به دلیل رویدادهای تصادفی و چه به دلیل فعالیت های خرابکارانه داشته باشند. دسترس پذیری در حالت مشهود به معنای آن است که سیستم در مقابل حملات انکار سرویس باید قدرتمند و ضدضربه باشد.

در دستگاه ها و سایر سیستم های اطلاعاتی جهت حفظ امنیت از ۵ مرحله حسابرسی استفاده می کنند که به ترتیب زیر است:

۱. هویت: به معنای داشتن یک شناساگر برای افراد به جهت دسترسی کاربران معتبر به اطلاعات است که نمونه بارز آن در حال حاضر نام کاربری است.

۲. احراز هویت: به معنای بررسی صحت نام کاربری فردی است که مدعی مالکیت آن است، به کمک پسورد یا رمز عبور می‌توان صحت هویت فرد مدعی را بررسی کرد.

۳. اجازه: فرایندی است که طی آن کاربران و یا تجهیزات متقاضی دسترسی به منابع، امکان استفاده از منبع یا منابع مستقر بر روی شبکه داده می‌شود. به بیان دیگر این عمل برای مدیران شبکه امکان تعیین نوع دسترسی به هر یک از منابع شبکه، برای تک تک متقاضیان دسترسی و یا گروهی از آنها را فراهم می‌کند. لیست‌های کنترل دسترسی مانند دیوارهای آتش و سیستم‌های تشخیص نفوذ در این مرحله قرار دارند. در واقع لیست‌های کنترل دسترسی می‌توانند ترافیک داخلی و خارجی شبکه را کنترل کنند از این رو سیستم‌های تشخیص نفوذ در این لیست قرار می‌گیرند.

۴. حسابرسی یا ممیزی کردن: به معنای جمع آوری اطلاعات شبکه و رکوردهای ترافیک شبکه است تا به کمک ابزارهای مناسب بتوان این داده ها را بررسی کرد و در جهت تشخیص عملیات نفوذ و حتی پی بردن به عملکرد سیستم از آنها استفاده کرد. این داده ها می‌توانند درباره سیستم عامل، کاربردها و یا فعالیت کاربران باشد.

۵. پاسخگویی: در حالت کلی به معنای بیان نتایج حاصل از درخواست کاربر احراز هویت شده است. با توجه به اهمیت سیستم‌های تشخیص نفوذ در ارتقا امنیت سیستم‌های اطلاعاتی و موقعیت آن در مراحل ایمن سازی می‌توان به کمک این سیستم ها، عملیات نفوذ در یک سیستم را نافرجام باقی گذاشت. اهمیت سیستم‌های تشخیص نفوذ در آنجا به اوج خود می‌رسد که قبل از اینکه حمله به اتمام برسد وقوع آن را گزارش کند و مانع از آسیب‌های احتمالی شود. بنابراین تشخیص نفوذ عبارت است از فرایند شناسایی و پاسخ به فعالیت های مخرب که به صورت هدفمند، منابع شبکه را مورد تهاجم قرار می‌دهد.

براساس تعریف مؤسسه استاندارد و تکنولوژی «فرآیند نظارت بر رویدادهایی که در سیستم کامپیوتر و یا شبکه رخ می‌دهد و تجزیه و تحلیل آنها در جهت یافتن

نشانه‌هایی از نفوذ، که به معنای دورزدن محرمانگی و جامعیت و دسترسی پذیری و یا عبور از مکانیزم‌های امنیتی یک کامپیوتر یا شبکه است، می‌باشد.

با توجه به تکامل در عرصه رایانش، روش‌های بسیاری جهت توزیع منابع و پیشرفت استفاده

داده‌ها از قبیل خوشه‌بندی داده‌ها، رایانش توری و سیستم مدیریت پایگاه داده‌های توزیع شده معرفی شده‌اند. امروزه رایانش ابری مکانیزم در حال ظهور برای محاسبات سطح بالا به عنوان یک سیستم

ذخیره سازی تلقی می‌شود که در آن ابرها به کاربران خود بر مبنای میزان استفاده از منابع هزینه دریافت کرده و سرویس‌های خود را در اختیار آن‌ها قرار می‌دهند. از این رو می‌توان سرویس‌های ابری را در ایجاد انگیزه برای شروع یک کسب و کار با هزینه‌های مالی پایین‌تر سهم دانست. رایانش ابری بسته به نوع توزیع منابع از سه لایه زیرساخت به عنوان سرویس، پلتفرم به عنوان سرویس و نرم افزار کاربردی به عنوان سرویس تشکیل شده است.

واضح است که رایانش ابری گام بعدی تکامل سرویسهای فناوری اطلاعات برحسب تقاضا می‌باشد. با این حال مسئله امنیت در رایانش ابری یکی از مسائل پیچیده به شمار رفته که تمامی سه لایه ابر، مسئولیت‌هایی که بین کاربران و ارائه دهندگان تقسیم می‌شود و حتی شخص ثالث را درگیر می‌کند. مشکلات امنیتی به عنوان یک مانع بزرگ در مقابل استفاده کاربران از سیستم‌های رایانش ابری قلمداد می‌شود.

با گسترش سیستم‌های پایگاهی و حجم بالای داده‌های ذخیره شده در این سیستم‌ها، به ابزاری نیاز است تا بتوان این داده‌ها را پردازش کرد و اطلاعات حاصل از آن را در اختیار کاربران قرار داد. معمولاً کاربران پس از طرح فرضیه‌ای بر اساس گزارشات مشاهده شده به اثبات یا رد آن می‌پردازند، در حالی که امروزه به روشهایی نیاز داریم که به اصطلاح به کشف دانش پردازند یعنی روش‌هایی که با کمترین دخالت کاربر و به صورت خودکار الگوها و رابطه‌های منطقی را بیان نمایند.

یکی از روش‌های بسیار مهمی که با آن می‌توان الگوهای مفیدی را در میان داده‌ها تشخیص داد، داده کاوی است، این روش که با حداقل دخالت کاربران همراه است اطلاعاتی را در اختیار آنها و تحلیل گران قرار می‌دهد تا براساس آنها تصمیمات

مهم و حیاتی را در سازمانشان اتخاذ نمایند. باید توجه داشت که اصطلاح داده کاوی زمانی به کار برده می شود که با حجم بزرگی از داده ها، در حد مگابایت یا ترابایت، مواجه باشیم. در تمامی منابع داده کاوی بر این مطلب تاکید شده است. هرچه حجم داده ها بیشتر و روابط میان آنها پیچیده تر باشد دسترسی به اطلاعات نهفته در میان داده ها مشکلتر می شود و نقش داده کاوی به عنوان یکی از روش های کشف دانش، آشکارتر می گردد.

داده کاوی از چندین رشته علمی (نظیر: تکنولوژی پایگاه داده، هوش مصنوعی، شبکه های عصبی، آمار، سیستم های مبتنی بر دانش، بازیابی اطلاعات و غیره) به طور همزمان بهره می برد. که برای پرهیز از اطاله کلام می توان به لحاظ تاریخی به اختصار به مراحل زیر تقسیم کرد:

- **مرحله اول:** گردآوری و ایجاد پایگاه اطلاعاتی (تا دهه ۱۹۶۰)
- **مرحله دوم:** نظام های مدیریتی مبنی بر پایگاه اطلاعاتی (دهه ۱۹۷۰ و اوایل دهه ۱۹۸۰)
- **مرحله سوم:** نظام های پایگاه اطلاعاتی پیشرفته (اواسط دهه ۱۹۸۰ تا زمان حاضر)
- **مرحله چهارم:** انبارش اطلاعات و داده کاوی (اواخر دهه ۱۹۸۰ تا به امروز)
- **مرحله پنجم:** نظام پایگاه اطلاعاتی مبنی بر شبکه (دهه ۱۹۹۰ تا کنون)
- **مرحله ششم:** نسل نوین نظام های اطلاعاتی یکپارچه شده (از ۲۰۰۰ به بعد)

بدین ترتیب فعالیتی که از دهه ۱۹۶۰ شروع شده بود در دهه ۱۹۹۰ گام های بلندی برداشت و انتظار می رود در این قرن به رشد و بالندگی خود ادامه دهد. در امنیت اطلاعات، کار عملیات تشخیص، تشخیص نفوذ است که برای امکان شناسایی یکپارچگی، قابلیت اطمینان، یا قابلیت دسترسی به منابع تلاش می کند. درکل، تشخیص نفوذ منجر به جلوگیری از نفوذهای نمی شود. تقاضاهای تشخیص نفوذ پیشرفته با مشکلات پیچیده ای روبرو هستند. این تقاضاها برای مدیریت به توسعه پذیری، معتبر بودن و آسان بودن نیاز داشته و هزینه نگهداری پایینی دارند.

داده کاوی توجه زیادی به علت افزایش، تولید، انتقال و ذخیره مقدار داده دارد و برای خلاصه کردن به جذب اطلاعات و دانش سودمند از آن ها نیاز دارد. در سال های گذشته، تحقیق درخصوص امکان استفاده از تکنیک های داده کاوی در زمینه برآوردن امنیت اطلاعات مخصوصاً در مشکل چالش انگیز تشخیص نفوذ شروع شده است. تشخیص نفوذ فرآیند یافتن وقایع مهم به وجودآمده در یک سیستم رایانه و تحلیل آن ها برای وجود نفوذ است. بنابراین، این فرآیند بازبینی و تحلیل وقایع به وجودآمده در یک سیستم رایانه به منظور تشخیص مشکلات امنیتی می باشد. یک IDS براساس کیفیت و کارایی به یک آرایه با مؤلفه ها و خصوصیات مختلف نیاز دارد شامل:

- متمرکز کردن دیدگاه داده
 - توانایی انتقال داده
 - روش های داده کاوی و تحلیلی
 - گسترش کشف کننده انعطاف پذیر شامل زمان بندی که فعال می شود.
 - تشخیص زمان واقعی و سازنده علامت هشدار
 - قابلیت های گزارش گیری
 - فرآیند توزیع
 - قابلیت دسترسی بالا به سیستم
 - مقیاس پذیری با بارگذاری سیستم
- در کل دو نوع حمله وجود دارد:
- حمله داخلی، حمله ای است که یک مزاحم، همه امتیازی برای دسترسی به درخواست دارد.
 - حمله خارجی، حمله ای است که مزاحم، حقوق مناسبی برای دسترسی به سیستم ندارد. تشخیص حمله داخلی معمولاً در مقایسه با حمله خارجی مشکل تر است.

در فصل نخست این کتاب، به تشریح ابعاد مختلف پردازش ابری شامل تعریف ابرها و انواع آنها، نقاط قوت و ضعف محاسبات ابری و همچنین معماری لایه های ابری و انواع آن، نحوه مدیریت ابرها می پردازیم.

در فصل دوم، به تعریف داده کاوی، الگوریتم‌های مورد استفاده در آن و ضرورت استفاده از داده کاوی در محاسبات ابری خواهیم پرداخت.

در فصل سوم نیز، نکات امنیتی ابرها شامل حملات امنیتی به ابرها، انواع روش‌های تشخیص نفوذ به ابرها و راهکارهای مقابله، انواع معماری سیستم‌های تشخیص نفوذ، اجزای سیستم‌های تشخیص نفوذ، ارزیابی سیستم‌های تشخیص نفوذ را بیان می‌نماییم.

در فصل چهارم، سیستم‌های تشخیص نفوذ داده کاو محور، تکنیک‌های داده کاوی و درختان تصمیم‌گیری و مدل‌های تشخیص نفوذ را مورد بحث و بررسی قرار می‌دهیم.

و در انتها، به بیان تکنیک‌های خوشه‌بندی داده کاو محور و تشخیص نفوذ پرداخته، به طور اجمالی الگوریتم‌های داده کاوی برای سیستم تشخیص نفوذ را بیان نموده و در پایان، الگوریتم‌های داده کاوی خوشه‌بندی K-Means، خوشه‌بندی Y-Means و خوشه‌بندی Fuzzy C-Means را به منظور ارتقاء سیستم‌های تشخیص نفوذ و افزایش سطح امنیت اطلاعات در ابرها پیشنهاد کرده و نقاط قوت ضعف هریک را تشریح می‌نماییم.