

۱۳۹۶۸۹۷

داده کاوی، OLAP

دکتر جمال شهرابی

عضو هیئت علمی دانشگاه صنعتی امیرکبیر

و دبیر کنفرانس ملی داده کاوی ایران

مهندس ونوس شهنورنژاد

انتشارات سروش گیتا

سرشناسه	: شهرابی، جمال، ۱۳۴۵ -
عنوان و نام پدیدآور	: داده کاوی و OLAP/تالیف جمال شهرابی، ونوس شکورنیا؛ [برای] کنفرانس داده کاوی ایران، موسسه پژوهشی داده پردازان گیتا.
مشخصات نشر	: تهران: سروش گیتا، ۱۳۹۴.
مشخصات ظاهری	: ۱۴۴ص: جدول، نمودار.
شابک	: ۱۲۰۰۰ریال ۷-۶-۹۵۶۱۴-۶۰۰-۹۷۸:
وضعیت فهرست نویسی	: فیا
یادداشت	: چاپ قبلی: جهاد دانشگاهی، واحد صنعتی امیرکبیر، ۱۳۹۰.
یادداشت	: کتابنامه.
موضوع	: سرور اس. کیو. ال
موضوع	: داده کاوی - نرم افزار
موضوع	: تکنولوژی اولپ
موضوع	: سامانه های پشتیبانی تصمیم گیری - نرم افزار
شناسه افزوده	: سر نیا ونوس، ۱۳۶۵ -
شناسه افزوده	: کنفرانس داده کاوی ایران
شناسه افزوده	: موسسه پژوهشی داده پردازان گیتا
رده بندی کنگره	: ۱۶۴ ۹۷ ۵۲/۵۸ QA
رده بندی دیویی	: ۰۵/۷۴
شماره کتابشناسی ملی	: ۴۱۱۷۲۴۸



داده کاوی، OLAP

دکتر جمال شهرابی

مهندس ونوس شکورنیا

چاپ: اول، ۱۳۹۴، شمارگان: ۱۰۰۰ نسخه، چاپ: قلم، قیمت: ۹۵۰۰۰ ریال
تمام حقوق این اثر برای انتشارات سروش گیتا محفوظ است.

خیابان اقدسیه، بزرگراه ارتش، مجتمع میلاد، بلوک ۳، واحد ۱۶، کد پستی: ۳۵۴۴۱ - ۱۶۹۵۸

تلفن: ۲۲۴۹۷۲۸۳ و ۲۲۴۹۷۲۹۳ - تلفکس: ۲۲۴۹۶۶۹۳

Publication@ irandatamining.com

www.irandatamining.com

پیشگفتار انتشارات سروش گیتا:

دانش نوین داده‌کاوی، هوشمندی کسب و کار و هوش مصنوعی زمینه‌های تخصصی انتشارات سروش گیتا است. این انتشارات با انتشار اولین و بیشترین کتب داده‌کاوی در کشور مقام اول را در گسترش دانش داده‌کاوی دارا است.

انتشارات سروش گیتا تلاش دارد با همکاری تمامی متخصصان و علاقه‌مندان دانش داده‌کاوی و هوشمندی کسب و کار و هوش مصنوعی، در انتشار مستمر آخرین یافته‌های علمی در این زمینه، نقشی ارزشمند در گسترش این زمینه‌های تخصصی در کشور داشته باشد.

انتشارات سروش گیتا از تمامی فرهیختگان این عرصه‌ها جهت چاپ و انتشار اثر ارزشمندشان دعوت به عمل می‌آورد.

فهرست مطالب

فصل اول - معرفی داده کاوی و OLAP	۱۱
داده کاوی چیست؟	۱۱
معرفی OLAP	۱۴
مقدمه	۱۴
معرفی OLAP و OLTP	۱۵
اسکیمای ستاره ای و انواع آن	۱۸
مفهوم Dimension و سلسله مراتب	۲۰
مفهوم Measure Groups و Measure	۲۲
پردازش Cube و ذخیره سازی	۲۴
استفاده از ذخیره گاه بلادرنگ	۲۶
پرس و جوی Cube	۲۸
انجام محاسبات	۳۰
مرور و مشاهده Cube	۳۲
مفهوم مدل های ابعادی یکپارچه شده	۳۴
فصل دوم - راهکار SQL Server برای OLAP	۴۱

۴۱	مقدمه
۴۳	انبار داده و OLAP - مفاهیم و تعاریف
۴۳	مولفه های راه حل
۴۳	منابع داده
۴۴	استخراج، تبدیل و بارگذاری (ETL)
۴۴	انبار داده
۴۵	OLAP
۴۵	گزینه های طراحی معماری
۴۶	راه کار SQL Server برای OLAP
۴۷	سرویس های تحلیلی SQL Server برای OLAP
۴۸	معماری SSAS
۵۱	اشیای پایگاه داده SSAS
۵۲	ابزار مورد بحث
۵۳	SQL Server Management studio
۵۳	پردازش اشیای سرویس های تحلیلی SSAS
۵۳	مرور اشیای سرویس های تحلیلی
۵۴	ساختن پرس و جو

۵۴		اسکرپت کردن اشیای سرویس های تحلیلی SSAS
۵۴		مدیریت پایگاه داده های سرویس های تحلیلی SSAS
		استودیوی توسعه هوشمندی کسب و کار Microsoft Business Intelligence
۵۴		Development Studio (BIDS)
۵۵		محیط جستجوگر Analysis Services Solution Explorer
۵۵		محیط طراحی Analysis Services Designers
۵۵		منوهای سرویس های تحلیلی Analysis Services Menus
۵۵		Analysis Services Tools/Options
۵۶		گزینه های ذخیره سازی در OLAP
۵۶		ROLAP و HOLAP، MOALP
۵۷		تنظیمات از پیش تعیین شده ذخیره سازی
۵۷		MOLAP
۵۸		MOLAP زمان بندی شده
۵۸		MOLAP اتوماتیک
۵۸		MOLAP با تاخیر متوسط
۵۸		MOLAP با تاخیر کم
۵۹		HOLAP بلادرنگ

۵۹	ROLAP بلادرنگ
۵۹	مقایسه میان MOALP ، HOLAP و ROLAP
۶۱	مثال ROLAP در SSAS
۶۴	آزمون مقیاس پذیری
۶۵	محیط آزمون
۶۵	آزمون داده و گزینه های ذخیره سازی
۶۵	رویه آزمون
۶۶	نتایج آزمون مقیاس پذیری
۶۷	چکیده
۶۸	ساخت Cube
۶۸	نیازمندیها
۶۹	ساخت پروژه SSAS
۷۰	ایجاد منبع داده
۷۱	ایجاد نمای منبع داده
۷۴	ایجاد Cube با استفاده از Wizard
۷۶	افزودن مشخصه ها به Dimension ها
۷۸	ساخت Dimension های سلسله مراتبی

۷۹	تعریف کلید برای مشخصه‌های Dimension
۸۰	تعریف نام ستون برای صفاتی که دارای ستونهای کلید چندگانه هستند
۸۱	تعریف سلسله‌مراتب
۸۳	ساخت روابط مشخصه‌ها
۸۴	استقرار و Cube
۸۶	مرور و مشاهده داده‌ها در Cube
۸۹	فصل سوم - رابطه بین OLAP و داده‌کاوی
۹۵	کاوش داده تجمیع شده
۹۶	نیازمندیهای استخراج الگوهای OLAP
۹۸	داده‌کاوی با OLAP در مقایسه با داده‌بنی رابطه‌ای
۱۰۰	ایجاد مدل‌های کاوشی OLAP با استفاده از ابزارها و ویرایشگرها
۱۰۱	استفاده از ویزارد داده‌کاوی
۱۰۱	ایجاد مدل بخش‌بندی مشتریان
۱۰۱	الگوریتم خوشه بندی
۱۰۳	پردازش و روش عملکرد خوشه‌بندی
۱۰۵	سناریوی ساخت مدل بخش‌بندی
۱۱۱	ساخت مدل تحلیل سبد بازار

۱۱۱	تحلیل سبد خرید
۱۱۲	سناریوی ساخت مدل
۱۱۹	ایجاد یک مدل پیش بینی فروش و برنامه ریزی تقاضا
۱۲۰	سناریوی ساخت مدل
۱۲۶	استفاده از محیط طراحی داده کاوی
۱۲۷	مفهوم Dimension های داده کاوی
۱۳۳	استفاده توام بر سر دجه های MDX و DMX
۱۳۵	تعیین نام ستون ها از MDX
۱۳۵	Using AMO for the OLAP Mining Model
۱۳۵	استفاده از اشیاء مدیریت تحلیل در مدل های کاوشی OLAP

پیشگفتار

داده کاوی و OLAP هر دو از اعضای کلیدی خانواده فن آوری BI بشمار می آیند. این دو مکمل یکدیگر بوده و از مشخصات و امکانات همدیگر جهت انجام تحلیل ها استفاده می کنند. OLAP می تواند در مرحله انتقال داده از طریق موتور تجمیع داده برای داده کاوی سودمند باشد. در بسیاری از موارد، الگوها تنها در داده های تجمیع شده یافت می شوند و کشف الگوها به طور مفید از جداول Fact بسیار دشوار است. از سویی دیگر فن آوری OLAP در زمینه تجمیع مقادیر عظیم داده های تراکنشی بر پایه تعاریف ابعادی مناسب بوده و بر پایه محاسبات تجمیعی عمل می کنند. در حالی که بسیاری از تکنیک های OLAP، از خانواده پایگاه داده به شمار می آیند؛ تکنیک های داده کاوی از سه زمینه آکادمیک منتج شده اند: آمار، یادگیری ماشین، فن آوری های پایگاه داده.

با توجه به اینکه امروزه، بیشتر حوزه ها با حجم بالای ذخیره سازی داده های تراکنشی مواجه هستند، استفاده مستقیم از داده کاوی بر روی این داده ها ممکن است به استخراج الگوهای مفید و قابل توجهی منجر نگردد؛ از این رو شایسته است این داده ها در ساختارهای تحلیلی ابعادی، خلاصه و تجمیع شده و سپس با استفاده از تکنیک های داده کاوی به استخراج الگوها پرداخته شود. در این زمینه فقدان کتابی با رویکردی کاملاً کاربردی و تخصصی جهت استفاده کارشناسان IT در حوزه های بانک های اطلاعاتی تحلیلی و از ریز کارشناسان آمار و کلیه افرادی که در هر زمینه کاربردی در داده کاوی با حجم عظیمی از داده ها سروکار دارند، احساس می شود. دبیرخانه کنفرانس داده کاوی ایران با در نظر گرفتن نیازهای کارشناسان امور IT، آمار و همچنین با توجه به آخرین دستاوردهای پژوهشی در زمینه داده کاوی و OLAP، اقدام به تهیه و انتشار کتابی با رویکردی کاربردی و تخصصی نموده است. در فصل اول این کتاب به معرفی فن آوری OLAP، ساختار و اجزای ذخیره سازی آن پرداخته شده است. در فصل دوم راهکار SQL Server برای OLAP تشریح شده و نحوه

ساخت Cube ها در قالب یک پروژه نمونه بصورت گام به گام توضیح داده شده است. در فصل سوم ارتباط آن با دانش نوین داده کاوی تشریح گردیده است و ساخت سه مدل کاربردی بر روی ساختار OLAP در بانک اطلاعاتی نمونه ای حاوی اطلاعات فروش فروشگاه های زنجیره ای، توصیف شده است. این سه مدل عبارتند از مدل بخش بندی بازار با استفاده از تکنیک خوشه بندی، مدل پیش بینی فروش و برنامه ریزی تقاضا با استفاده از سری های زمانی و مدل تحلیل سبد بازار با استفاده از تکنیک قوانین وابستگی. این مدل ها با استفاده از محیط کسب و کار هوشمند در SQL Server پیاده سازی شده اند. بانک اطلاعاتی Focul Mart که شامل اطلاعات فروش فروشگاه های زنجیره ای است، به عنوان بانک اطلاعاتی نمونه در ساخت مدل ها استفاده شده است. این بانک اطلاعاتی از بانک های نمونه - آموزشی SQL Server می باشد.

نمونه پروژه ها و بانک اطلاعاتی در استفاده در CD ضمیمه کتاب قرار دارند.

شایان ذکر است به منظور درک سریعتر و استفاده کاربردی مطالب، از معادل فارسی کلیه واژه های کلیدی و مورد استفاده در نرم افزار، اجتناب شده است.

جا دارد از مدیریت محترم انتشارات جهاد دانشگاهی راجد صنعتی امیرکبیر بواسطه تلاش های انجام یافته شان در انتشار این کتاب سپاسگزاری نمایم.

دکتر جمال شهبابی	ونوس شکورنیاژ
عضو هیات علمی دانشکده مهندسی صنایع دانشگاه صنعتی امیرکبیر و دبیر کنفرانس داده کاوی ایران jamalshahrabi@aut.ac.ir	کارشناسی ارشد مهندسی صنایع دانشگاه صنعتی امیرکبیر venus.shakoorniaz@gmail.com