
Data Mining

Part 1. Introduction

1.1 Information Systems and Data Mining

(سیستم‌های اطلاعاتی و داده کاوی)

Spring 2010

Instructor: Dr. Masoud Yaghini

سیستم‌های اطلاعاتی و داده کاوی

- سیستم اطلاعاتی چیست؟
- اجزا و فعالیتهای یک سیستم اطلاعاتی
- مدیریت منابع داده ای
- ایجاد پایگاه داده
- انبار داده و داده کاوی

سیستم اطلاعاتی چیست؟

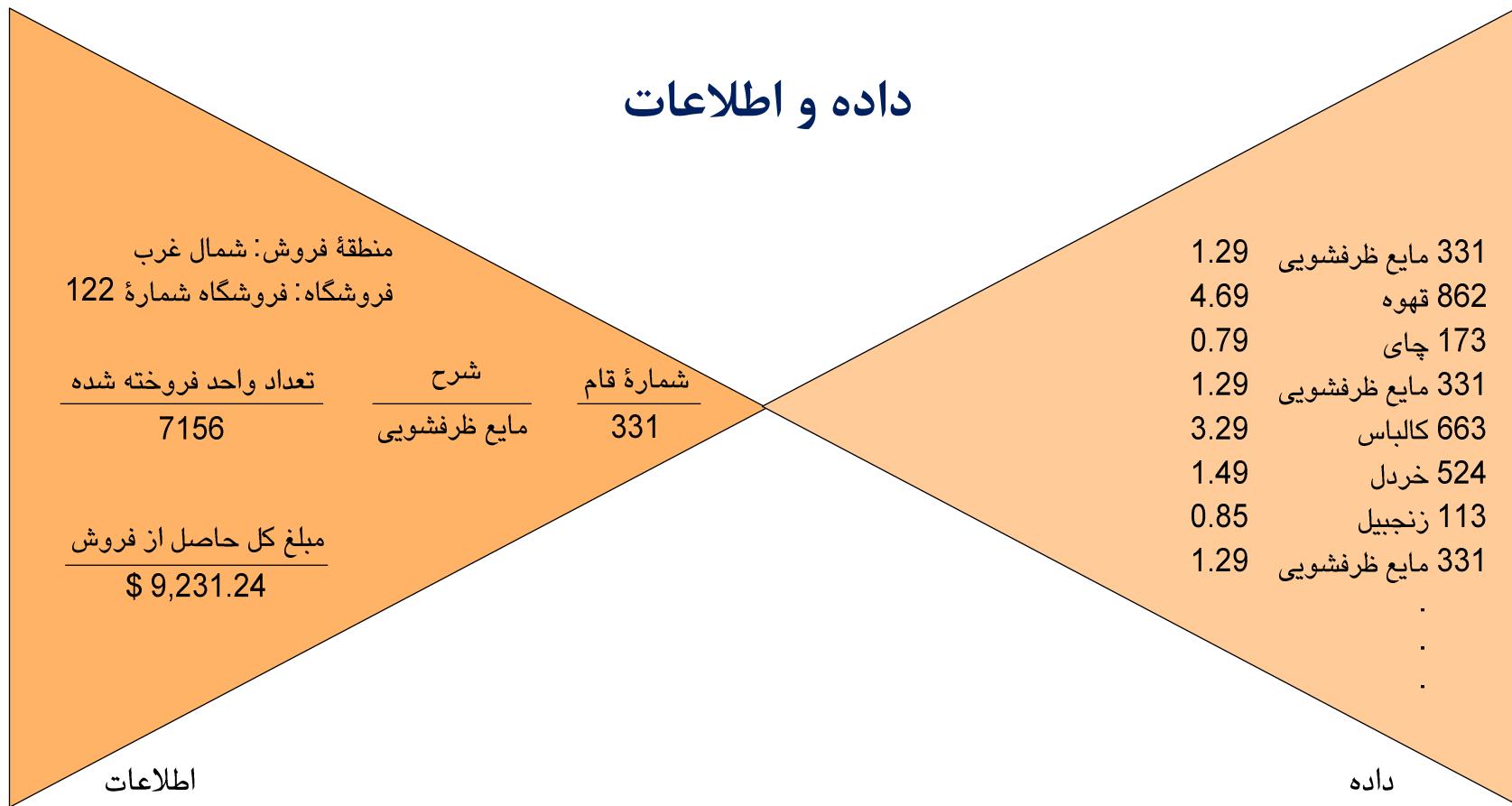
سیستم اطلاعاتی چیست؟

- **سیستم اطلاعاتی:** مجموعه‌ای از اجزای به هم مرتبط برای جمع آوری، ذخیره سازی، پردازش، بازیابی و انتشار اطلاعات جهت پشتیبانی از تصمیم گیری و کنترل در سازمان می‌باشد.

سیستم اطلاعاتی چیست؟

- **اطلاعات:** داده هایی پردازش شده به صورت معنی دار و مفید برای انسانها
- **داده:** وقایع خامی که بیان کننده رویدادهای اتفاق افتاده در سازمان نظیر عملیات کسب و کار

سیستم اطلاعاتی چیست؟



سیستم اطلاعاتی چیست؟

سیستم‌های اطلاعاتی دستی

(Manual information Systems)

- سیستم‌های اطلاعاتی هستند که مبتنی بر فناوری کاغذ و قلم هستند.

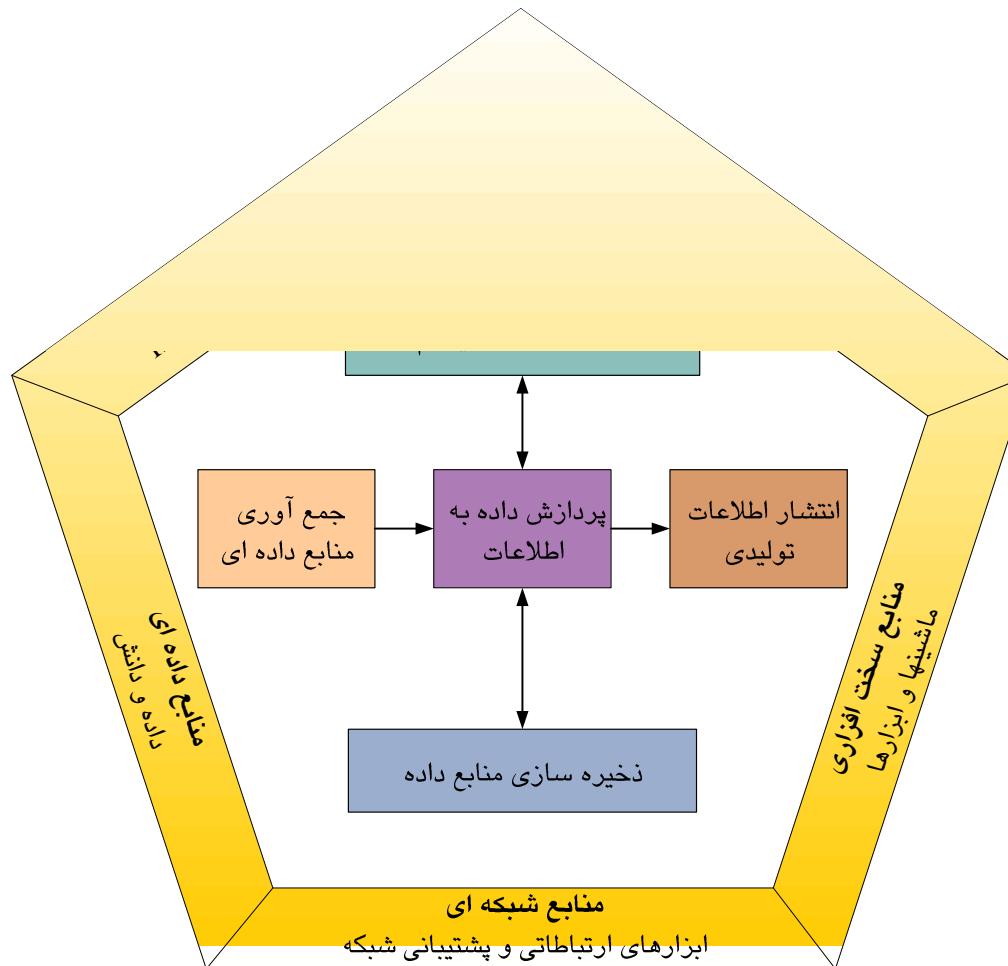
سیستم‌های اطلاعاتی مبتنی بر کامپیووتر (Computer-Based Information Systems)

- سیستم‌هایی هستند که برای پردازش و انتشار اطلاعات متکی فناوری اطلاعات می باشند.

اجزا و فعالیتهای یک سیستم اطلاعاتی

اجزا و فعالیتهای یک سیستم اطلاعاتی

اجزا و فعالیتهای یک سیستم اطلاعاتی



اجزا و فعالیتهای یک سیستم اطلاعاتی

منابع سیستمهای اطلاعاتی:

- منابع انسانی
- منابع سخت افزاری
- منابع نرم افزاری
- منابع داده ای
- منابع شبکه ای

اجزا و فعالیتهای یک سیستم اطلاعاتی

انواع منابع انسانی سیستم‌های اطلاعاتی

- کاربران نهایی: از سیستم اطلاعاتی یا اطلاعات تولید شده توسط آن استفاده می‌کنند.
- متخصصین سیستم‌های اطلاعاتی: سیستم‌های اطلاعاتی را ایجاد و مدیریت می‌کنند.

اجزا و فعالیتهای یک سیستم اطلاعاتی

انواع منابع سخت افزاری سیستم‌های اطلاعاتی

- کامپیوترها: حاوی واحد پردازش مرکزی، و تجهیزات جانبی متنوعی
- تجهیزات جانبی کامپیوترها: صفحه کلید، ماوسهای کامپیوتری، صفحه نمایش، چاپگر، و دیسکهای مغناطیسی و نوری

اجزا و فعالیتهای یک سیستم اطلاعاتی

منابع نرم افزاری سیستمهای اطلاعاتی

- **نرم افزار سیستمی:** یک برنامه سیستم عامل که عملیات یک کامپیوتر را کنترل و پشتیبانی می کند.
- **نرم افزار کاربردی:** برنامه ای که پردازش را توسط کاربران هدایت می کند.
- **رویه ها:** دستور العملهایی اجرایی برای افرادی که از یک سیستم اطلاعاتی استفاده می کنند.

اجزا و فعالیتهای یک سیستم اطلاعاتی

منابع داده ای سیستمهای اطلاعاتی

- **پایگاه های داده ای:** داده های پردازش شده و سازماندهی شده را نگهداری می کند.
- **پایگاه های دانش:** دانش را به شکل های متنوعی نظیر وقایع، قوانین، و نمونه های موردنی کسب و کار موفق را نگهداری می کنند.

اجزا و فعالیتهای یک سیستم اطلاعاتی

منابع شبکه ای سیستمهای اطلاعاتی

- **تجهیزات ارتباطاتی:** کابل زوجی، کابل فیبر نوری، و فناوری موج کوتاه، ماهواره ای، و بی سیم.
- **زیرساختار شبکه:** پردازشگرهای ارتباطاتی نظریر مودم و پردازشگرهای بین شبکه ای، و نرم افزارهای کنترل ارتباطات نظریر سیستم عاملهای شبکه

اجزا و فعالیتهای یک سیستم اطلاعاتی

فعالیتهای سیستمهای اطلاعاتی:

- جمع آوری
- پردازش
- انتشار
- ذخیره سازی
- کنترل عملکرد سیستم

مدیریت منابع داده ای

مدیریت منابع داده ای

سلسله مراتب داده

- **بیت (Bit):** کوچکترین واحد داده است؛ یک رقم باینری که می تواند 0 یا 1 باشد.
- **بایت (Byte):** گروهی از بیتها که نمایانگر یک کاراکتر است.
- **فیلد (Field):** گروهی از لغات یا اعداد کامپیوتری
- **رکورد (Record):** گروهی از فیلدهای مرتبط
- **فایل (File):** گروهی از رکوردهای هم نوع
- **پایگاه داده (Database):** گروهی از فایلهای مرتبط

مدیریت منابع داده ای

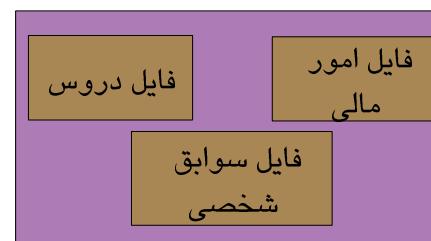
سلسله مراتب داده

سلسله مراتب

پایگاه داده

مثال

پایگاه داده دانشجو



فایل دروس

نام	درس	تاریخ	نمره
جان استوارت	IS 101	پائیز 2003	B+
کارن تیلور	IS 101	پائیز 2003	A
امیلی وینسنت	IS 101	پائیز 2003	C

فایل

رکورد

نام	درس	تاریخ	نمره
جان استوارت	IS 101	پائیز 2003	B+

فیلد

جان استوارت (فیلد نام)

بايت

ASCII حرف J بر اساس کد 01001010

بیت

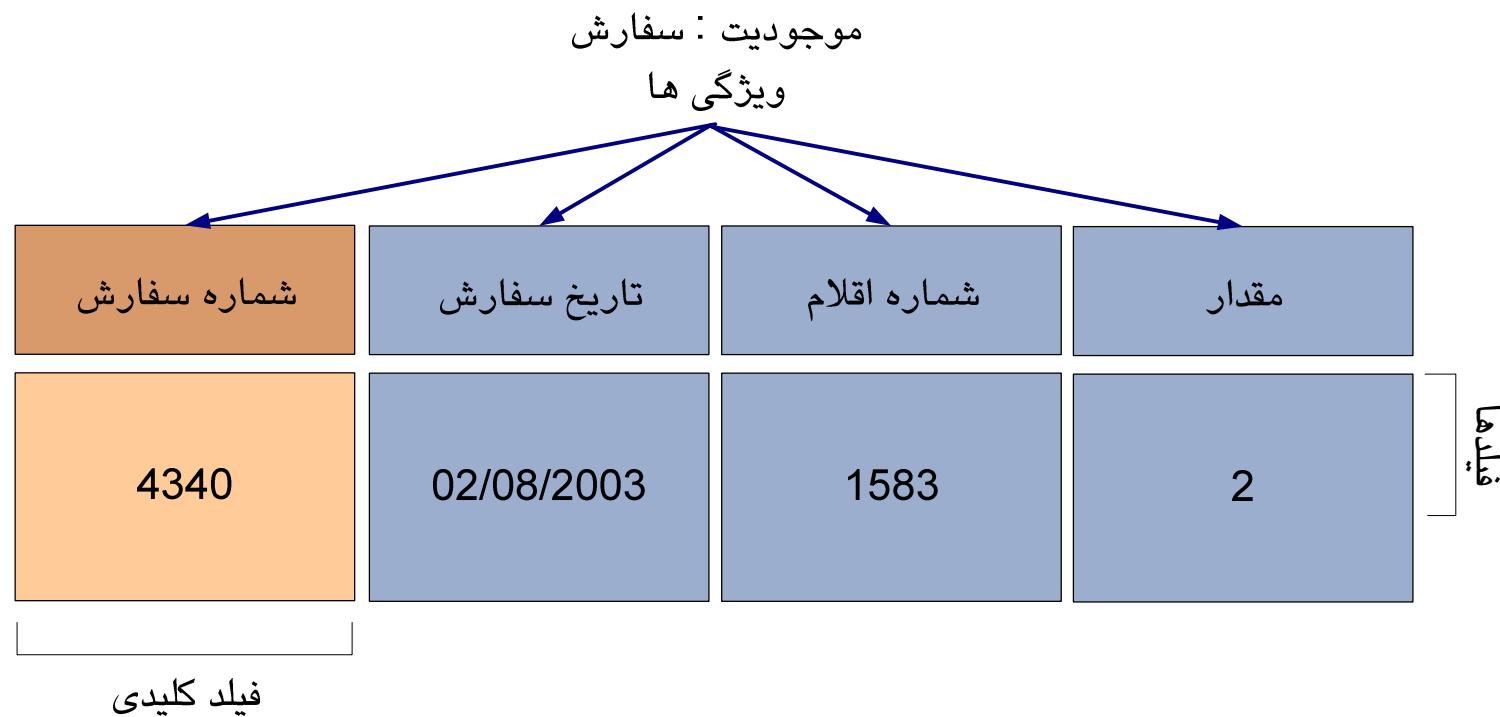
0

مدیریت منابع داده ای

- **موجودیت (Entity)**: فرد، مکان، شیء، یا رویدادی که درباره آن اطلاعات جمع آوری می شود.
- **ویژگی (Attribute)**: قطعه ای از اطلاعات که یک موجودیت بخصوص را توصیف می کند.
- **فیلد کلیدی (Key Field)**: فیلدی منحصر به فردی که هر رکورد از یک فایل را شناسایی می کند.

مدیریت منابع داده ای

موجودیتها و ویژگیها



مدیریت منابع داده ای

پایگاه داده

Database

- مجموعه ای است از داده های متمرکز.
- افزونگی داده را کنترل می کند.
- داده ها ذخیره شده اند تا برای کاربر در یک مکان واحد جلوه گر شوند.
- به چندین سیستم کاربردی خدمت رسانی می کند.

مدیریت منابع داده ای

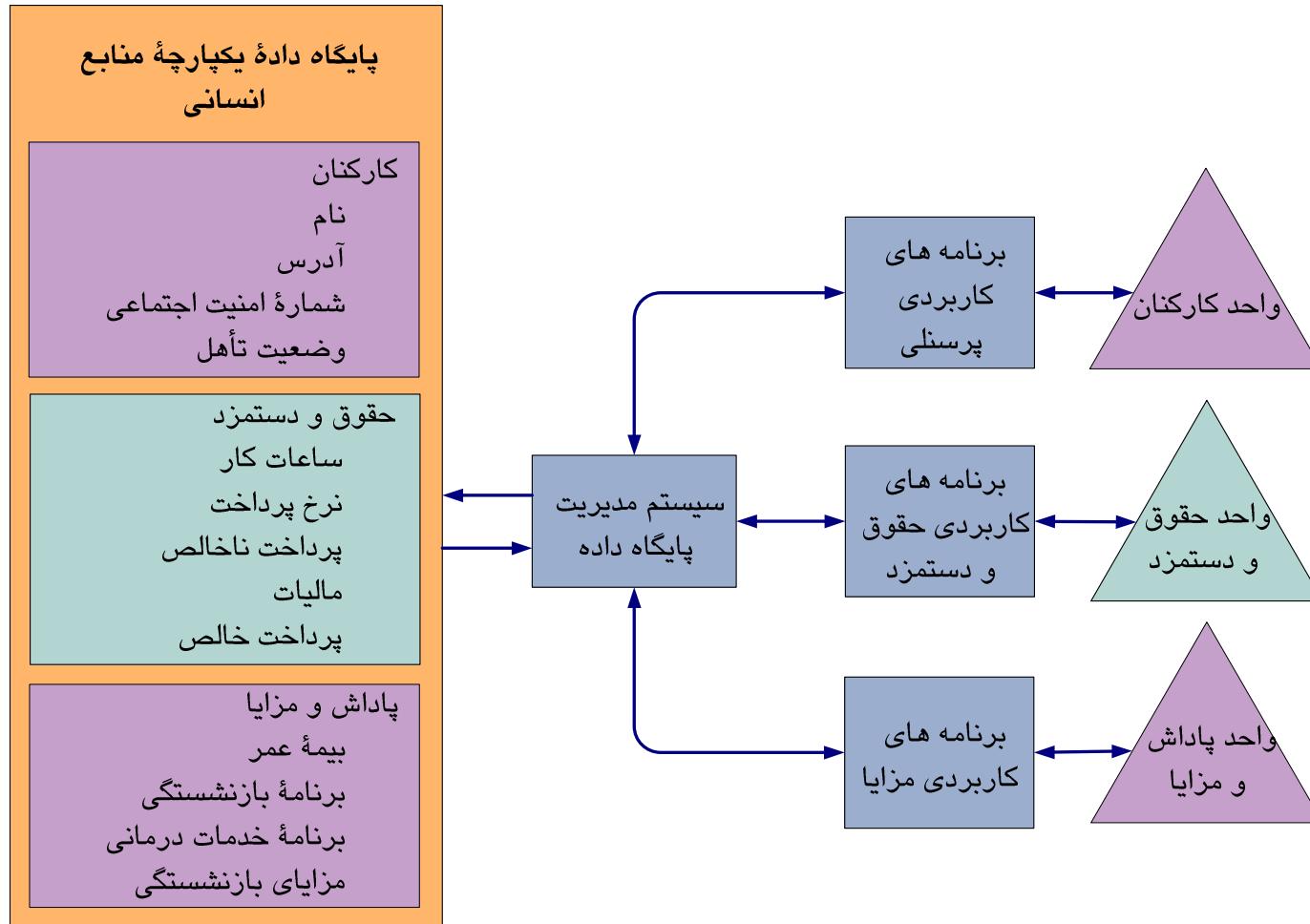
سیستم مدیریت پایگاه داده

Database Management System (DBMS)

- پایگاه های داده را ایجاد و از آنها نگهداری می کند.
- نیاز به تعریف داده را از بین می برد.
- حکم واسطی بین برنامه های کاربردی و فایل های داده های فیزیکی را دارد.
- دیدهای فیزیکی و منطقی داده را از هم جدا می کند.

مدیریت منابع داده ای

محیط یک پایگاه داده



مدیریت منابع داده ای

مزایای استفاده از DBMS:

- کاهش افزونگی داده
- حذف عدم سازگاری داده
- حذف وابستگی برنامه به داده
- افزایش دسترسی و دسترس پذیری داده
- فراهم آوردن امکان مدیریت متمرکز داده، استفاده از داده، و امنیت

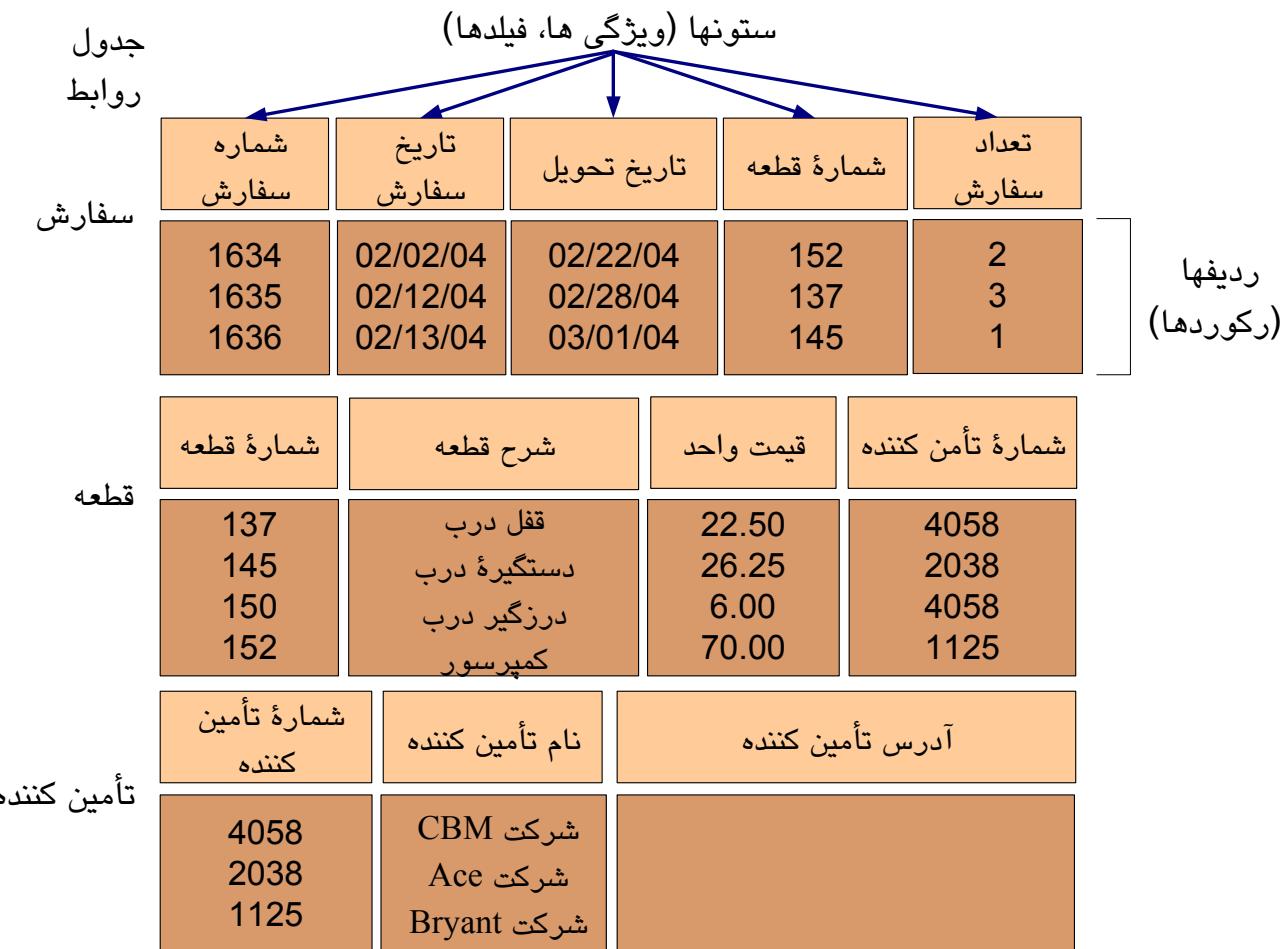
مدیریت منابع داده ای

DBMS رابطه ای Relational DBMS

- داده را در قالب جداول دو بعدی نمایش می دهد.
- داده را از طرق جداولی بر مبنای عناصر داده ای مشترک به هم مرتبط می سازد.
- مثال:
 - Oracle برای سیستمهای بزرگ
 - MS SQL Server برای سیستمهای حد متوسط
 - Oracle Lite و Microsoft Access برای سیستمهای کوچک

مدیریت منابع داده ای

مدل داده رابطه ای



ایجاد پایگاه داده

ایجاد پایگاه داده

دو فعالیت در طراحی یک پایگاه داده

- **طراحی مفهومی یا منطقی (Conceptual or Logical)**: مدل خلاصه شده ای است از پایگاه داده از زاویه دید کسب و کار.
- **طراحی فیزیکی (Physical Design)**: مدلی که نشان می دهد پایگاه داده چگونگی بر روی تجهیزات ذخیره سازی، سازماندهی شده است.

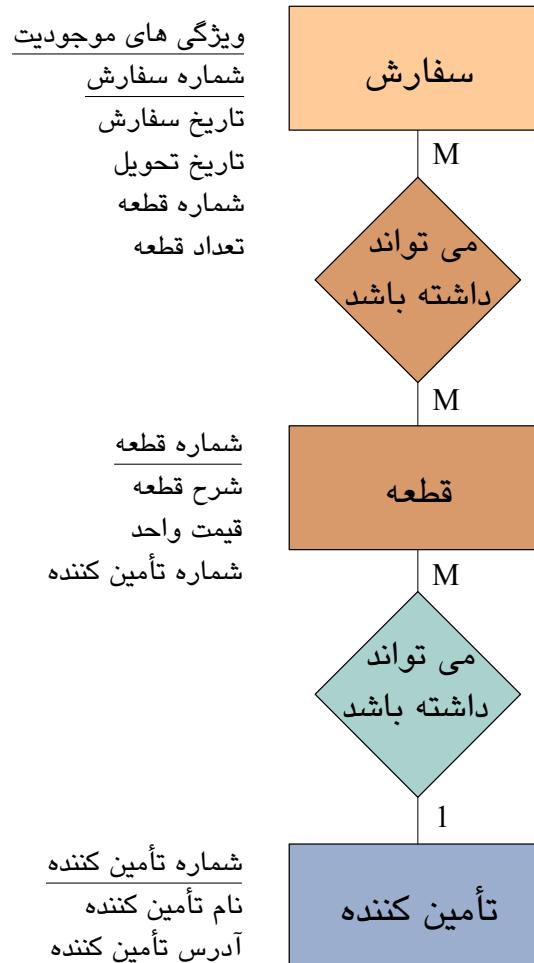
ایجاد پایگاه داده

طراحی مفهومی پایگاه داده Conceptual Database Design

- رابطه بین عناصر داده ای را شناسایی می کند.
- کاراترین راه گروه بندی عناصر داده ای را شناسایی می کند.
- عناصر داده ای تکراری را شناسایی می کند.
- گروه های داده ای مورد نیاز سیستمهاي کاربردي بخصوصی را شناسایی می کند.
- نمودار رابطه موجودیت (**Entity-Relationship Diagram**) متدولوژی برای مستند سازی پایگاه های داده که رابطه بین موجودیتهای مختلف پایگاه داده را نشان می دهد.

ایجاد پایگاه داده

یک نمودار رابطه موجودیت‌ها



ایجاد پایگاه داده

نرمال سازی (Normalization)

- فرآیند ایجاد ساختارهای داده کوچک، پایدار، و در عین حال منعطف از گروههای پیچیده داده، زمانی که یک پایگاه داده ای رابطه ای در حال طراحی است.

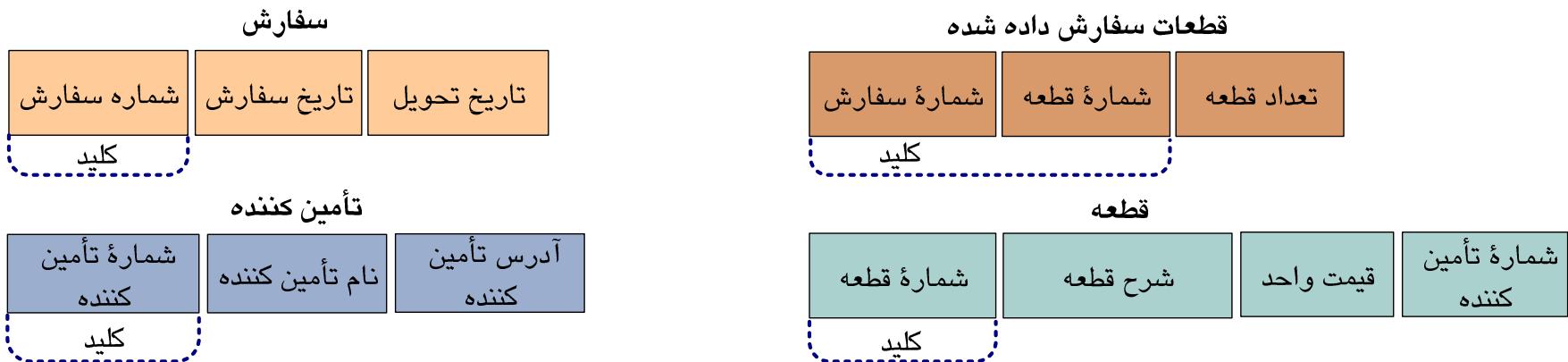
ایجاد پایگاه داده

یک رابطه غیر نرمال برای موجودیت سفارش

سفارش									
شماره سفارش	تعداد قطعه	شماره قطعه	شرح قطعه	قیمت واحد	شماره تأمین کننده	نام تأمین کننده	آدرس تأمین کننده	تاریخ سفارش	تاریخ تحویل

ایجاد پایگاه داده

یک رابطه نرمال شده برای موجودیت سفارش



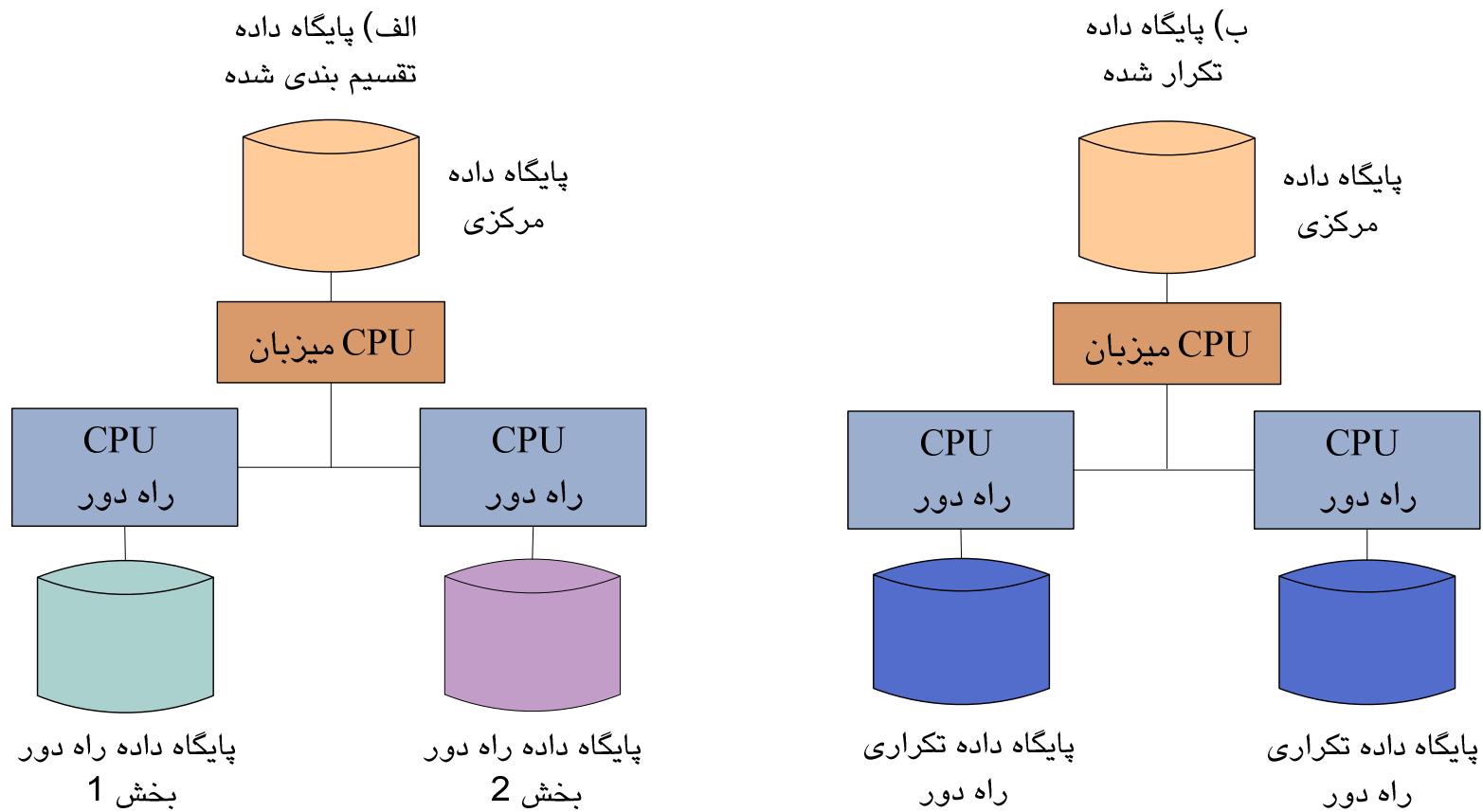
ایجاد پایگاه داده

پایگاه داده توزیع شده

- **پایگاه داده توزیع شده (Distributed Database):** پایگاه داده ای است که در بیش از یک مکان فیزیکی واقع شده باشد؛ به بیش از یک مکان تقسیم بندی یا تکرار شده است.

ایجاد پایگاه داده

پایگاه های داده توزیع شده



ایجاد پایگاه داده

پایگاه داده توزیع شده

- مزایا:

- افزایش خدمت رسانی و پاسخگویی به کاربران محلی
- کاهش آسیب پذیری به یک سایت واحد مرکزی بزرگ

- معایب:

- وابستگی به خطوط ارتباط از راه دور
- ریسک امنیتی از طریق توزیع داده های حساس
- به روز رسانی داده های مرکزی با داده های محلی

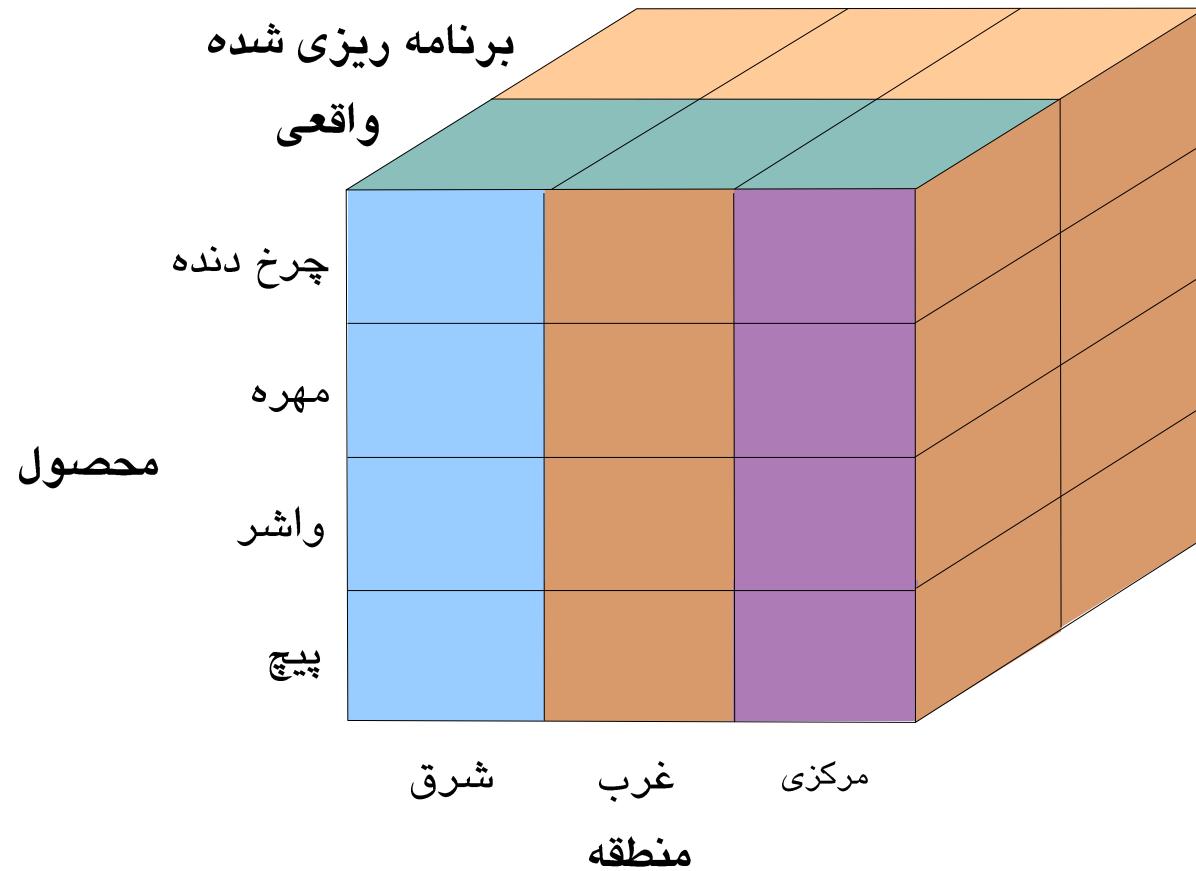
انبار داده و داده کاوی

انبار داده و داده کاوی

- انبار داده (**Data Warehouse**): داده های جاری و قدیمی را برای گزارش دهی و تحلیل ذخیره می سازد. داده ها بصو
- بازار داده (**Data Mart**): زیر مجموعه ای است از انبارهای داده با خلاصه ای از داده برای کاربران خاص.

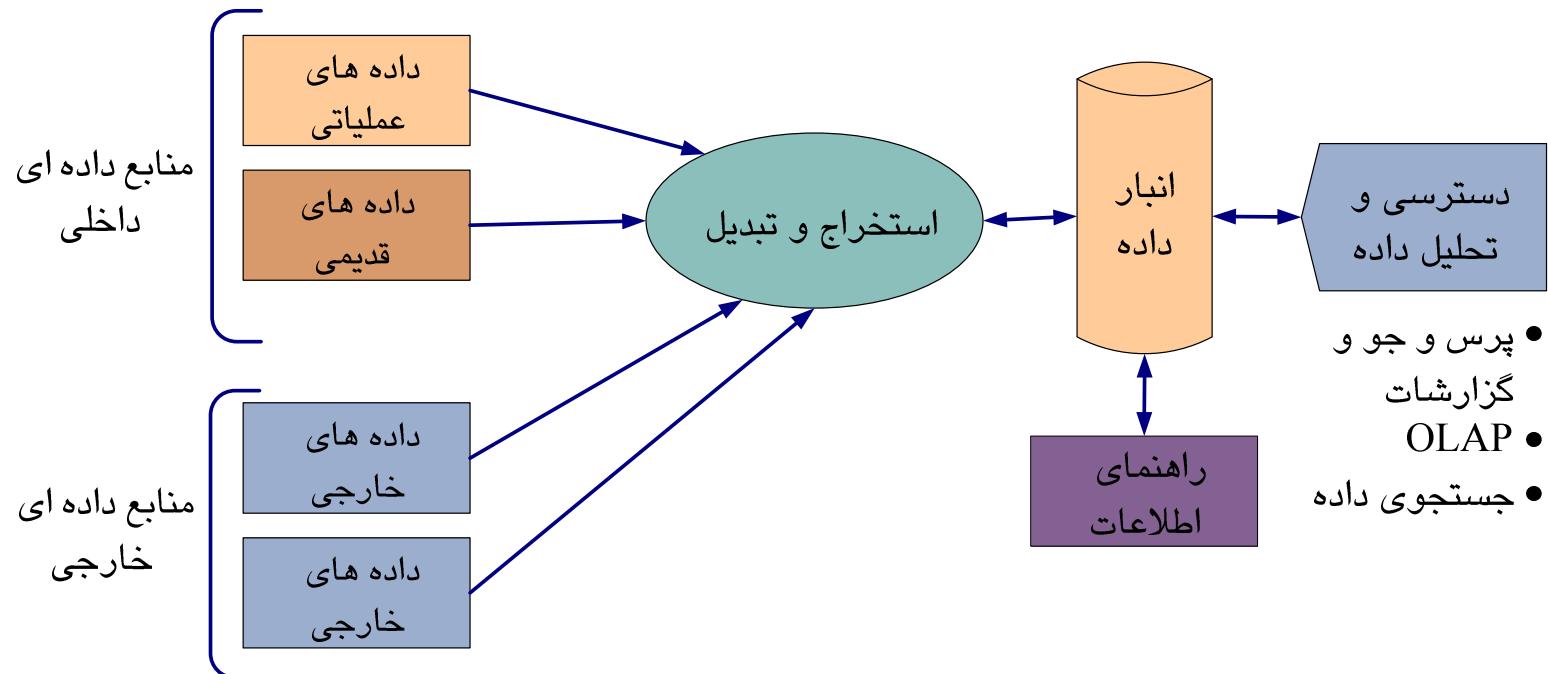
انبار داده و داده کاوی

مدل چند بعدی داده



انبار داده و داده کاوی

اجزای یک انبار داده



انبار داده و داده کاوی

مزایای انبارهای داده

- بهبود اطلاعات و قابلیت دسترسی به آنها
- قابلیت مدل کردن و مدلسازی مجدد داده
- قادر ساختن دسترسی به داده بدون تأثیر گذاری بر عملکرد سیستم های عملیاتی موروثی زیربنایی سازمان

انبار داده و داده کاوی

داده کاوی

- **داده کاوی (Data Mining):** تکنیکهایی برای یافتن الگوهای روابط پنهان در حجم عظیمی از داده برای استنتاج قوانینی برای پیش بینی روندهای آینده.

انبار داده و داده کاوی

نمونه هایی از کاربرد داده کاوی

- **Disco S.A.**: فروشگاه های زنجیره ای آرژانتین از استخراج داده برای تحلیل الگوهای خرید بیش از ۵/۱ میلیون مشتری که دارای برنامه خرید متنابی در بیش از ۲۰۰ فروشگاه هستند، استفاده می کند.
- **Red Robin**: این رستورانها از استخراج داده برای تحلیل منوهای ۸۷ رستوران خود استفاده می کند. با تحلیل قیمت، هزینه، و کیفیت هر یک از موادی که در غذاهای سرو شده به کار می رود، Red Robin می تواند اقلام منو را که هم دارای تقاضای زیاد و هم سود حاشیه ای بالاست، شناسایی کند.

References

References

- K.C. Laudon and J.P. Laudon, **Essentials of Management Information Systems, Managing the Digital Firm**, Prentice Hall, Sixth Edition, 2005.

The end