

پیامبر خدا (ص):
حُسْنُ الْمَسْأَلَةِ نِصْفُ الْعِلْمِ.
خوب پرسیدن، نیمی از دانش است.
تحف العقول : ۵۶ منتخب میزان الحکمة : ۲۶۰

Business Intelligence

درس: هوشمندی کسب و کار (هوش تجاری)
بخش دوم: آشنایی با تکنیک ها و تکنولوژی های هوش تجاری
فصل ششم: انباره سازی (انبارش) داده ها

گروه: مهندسی فناوری اطلاعات دانشگاه نور طوبی - رشته: تجارت الکترونیک
مدرس: حسنعلی نعمتی شمس آباد
دستیار (TA): زهرا سادات طبایی عقدایی

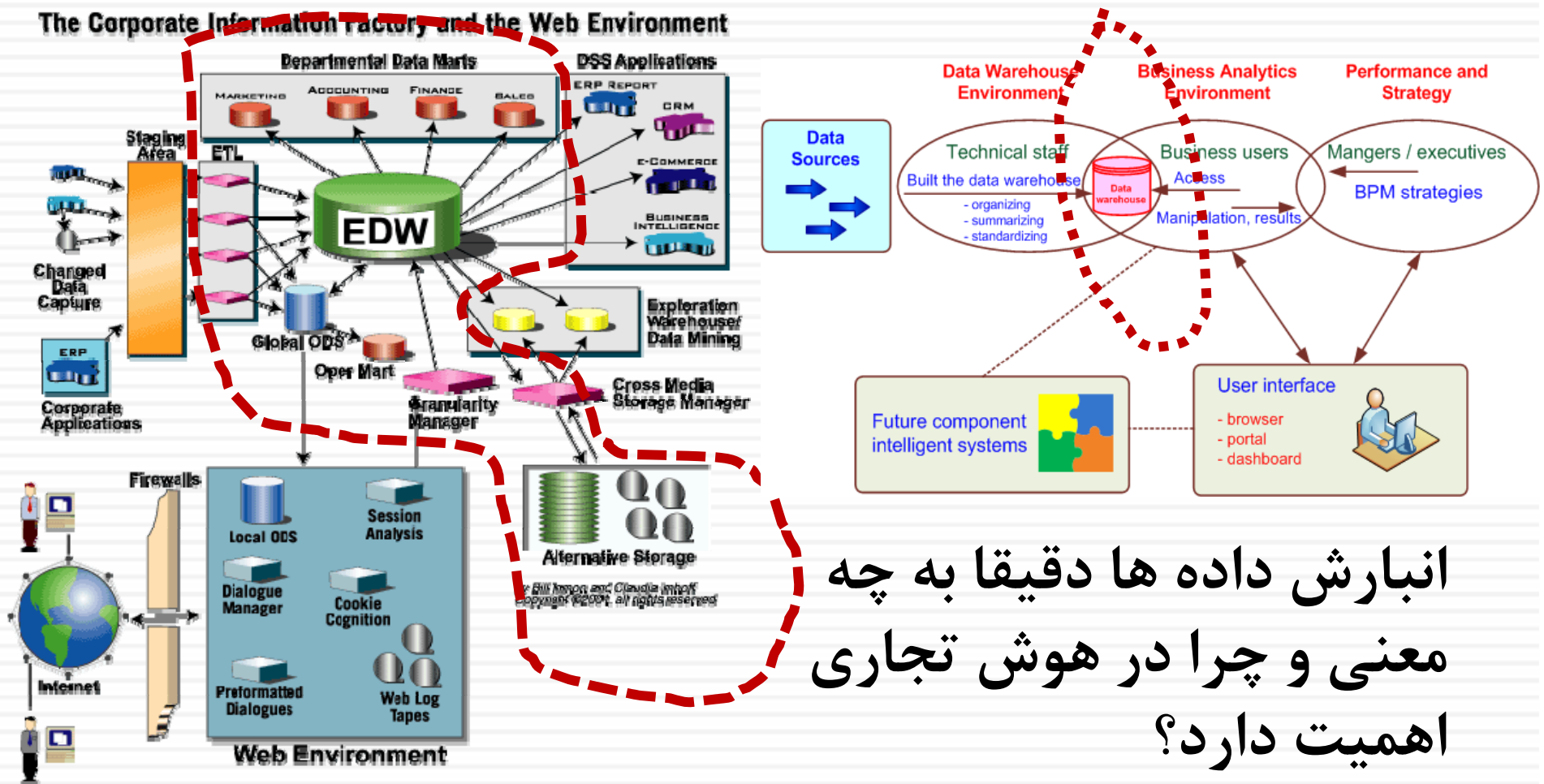
هدف فصل ششم: انباره سازی (انبارش) داده ها

- در این بخش اهداف زیر دنبال می شود
 - آشنایی با مفاهیم اصلی انباره ای داده
 - آشنایی با معماری های انباره سازی داده ها
 - آشنایی با تکنولوژی های انبارش داده ها
 - آشنایی با مسائل اجرایی انبارش داده ها

انبار داده: عنصر اصلی هوش تجاری (۱)

- آنچه که معماری لایه مدیریت داده های سیستم های هوش تجاری را از معماری دیگر سیستم ها متمایز می کند، موضوع مهم و حیاتی انبارسازی و ذخیره دائمی داده های عملیاتی است.
- مفهوم انبارش داده ها، در چارچوب های مفهومی هوش تجاری (مانند کارخانه اطلاعات) از مهمترین مفاهیم است و انبار داده در مرکز معماری منطقی سیستم های هوش تجاری قرار دارد.
- دیگر عناصر یک سیستم هوش تجاری (پردازش های تحلیلی، داده کاوی، ...) از این عنصر مهم بهره برداری می کنند.
- این موضوع آنقدر در حوزه هوش تجاری پررنگ است که گاهی انبار داده ها را معادل هوش تجاری در نظر می گیرند.

انبار داده: عنصر اصلی هوش تجاری (۲)



انبارش داده ها دقیقا به چه معنی و چرا در هوش تجاری اهمیت دارد؟

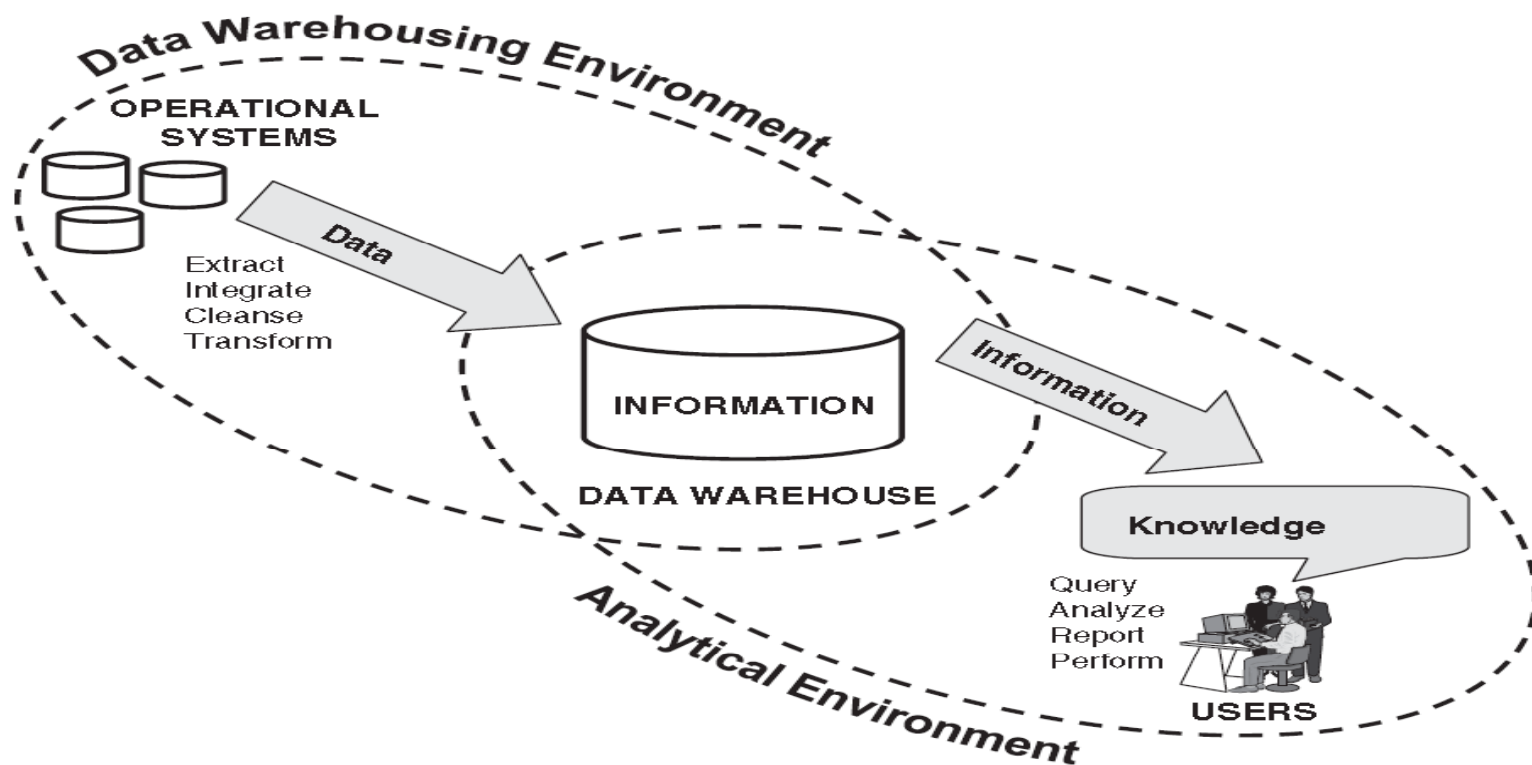
چرا در کنار پایگاه داده به انباره داده نیاز داریم؟

- شکی نیست که برای تصمیم گیری هوشمندانه باید از سوابق کارها و عملیات گذشته آگاهی داشته باشیم.
- آگاهی از سوابق کارها و عملیات گذشته در صورتی ممکن است که اطلاعات کارها و داده های ناشی از عملیات به صورت مناسبی ذخیره شده باشد و دسترسی به آنها از زوایای مختلف وجود داشته باشد.
- انباره داده ها این نقش مهم را بر عهده دارد. یعنی برای هر موضوع تصمیم گیری، خلاصه ای از سوابق داده های مورد نیاز را به صورت یکپارچه و در ابعاد مختلف در اختیار ما قرار می دهد.

هوش تجاری: انبارش داده ها + تحلیل

6

nemati@ut.ac.ir



انباره داده: طرحی نو از پایگاه داده های قدیمی

7

nemati@ut.ac.ir

تصمیم گیری های مبتنی بر BI به داده های گذشته، متراکم و خلاصه شده ای نیاز دارد. این نیاز مهم، نوع خاصی از پایگاه داده را می طلبد.

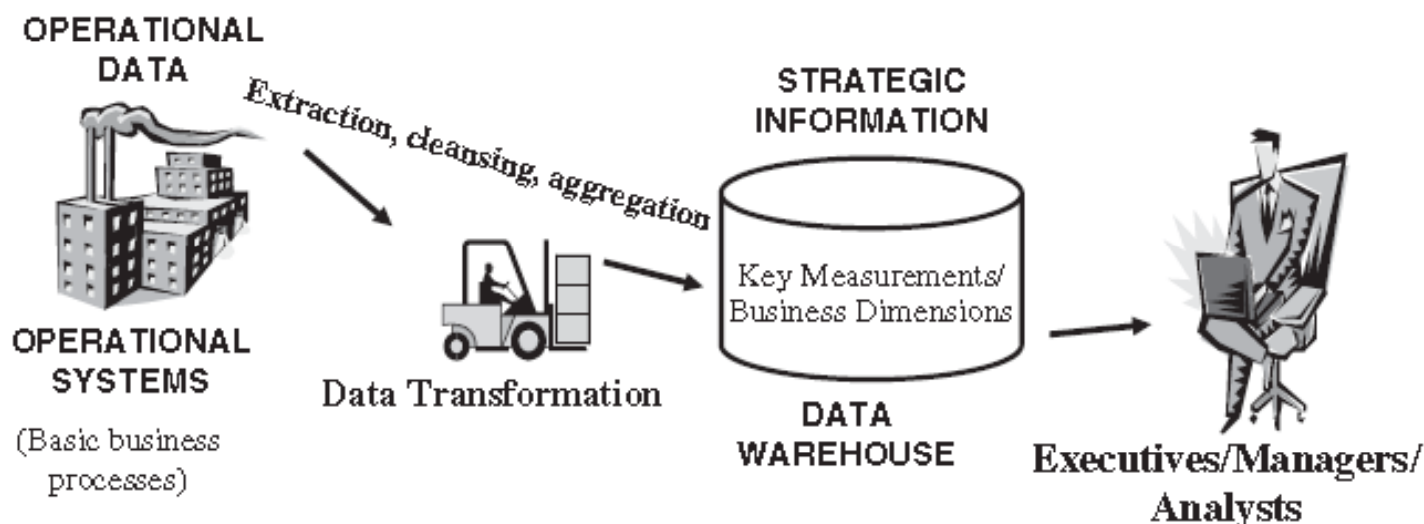
پایگاه داده چندبُعدی جدید همراه با تکنولوژی های BI توانایی برش اطلاعات جهت گزارشگیری و پردازش های تحلیل را خواهد داشت.

دسترسی به داده های ذخیره در این نوع پایگاه داده هم با استفاده از راه های متعارف قبلی (پرس و جوهای SQL) و هم از طریق راه های جدید چندبُعدی (ابزارهای پردازش تحلیل برخط OLAP) امکان پذیری است.

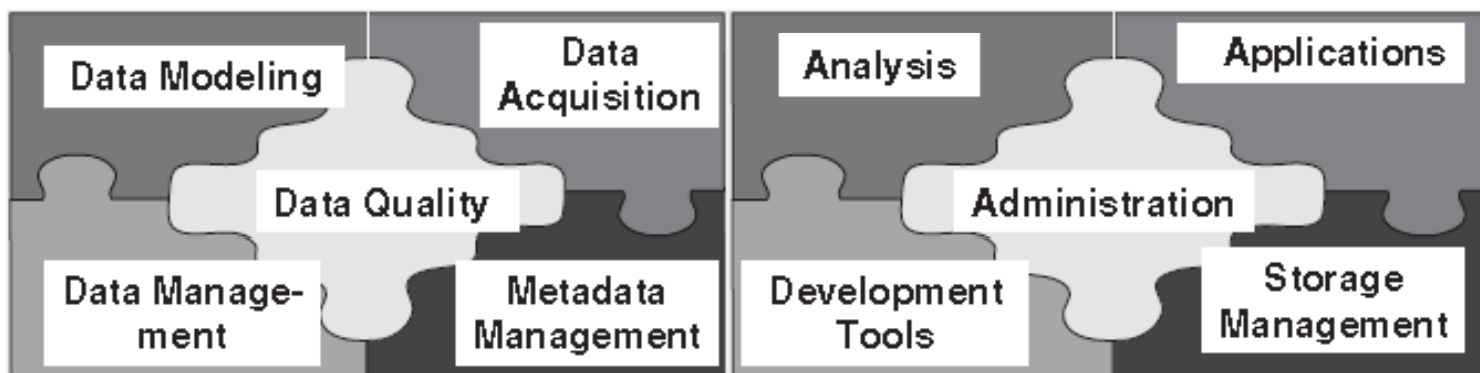
انبار داده ها: ترکیبی از تکنولوژی ها

8

nemati@ut.ac.ir



BLEND OF TECHNOLOGIES



پایگاه داده عملیاتی و مسأله کارآیی

- تمرکز پایگاه داده عملیاتی جلوگیری از ذخیره چند مشخصه همسان در مکان های مختلف و نرمال سازی جداول به منظور پرهیز از افزونگی داده ها است.
- اکثر سیستم های عملیاتی بر اساس فلسفه ثبت داده ها (Data-in) طراحی شده اند به گونه ای که داده ها در فضای بهینه ذخیره شوند و از افزونگی جلوگیری شود.
- تهیه گزارش های تحلیلی استراتژیک از جداول نرمال، نیازمند تجمیع همزمان جداولی بزرگ است که این کار کارایی پایگاه داده را پایین می آورد.
- بنابراین پایگاه داده های BI به صورت چندبُعدی و با جداول غیرنرمال طراحی می شوند.

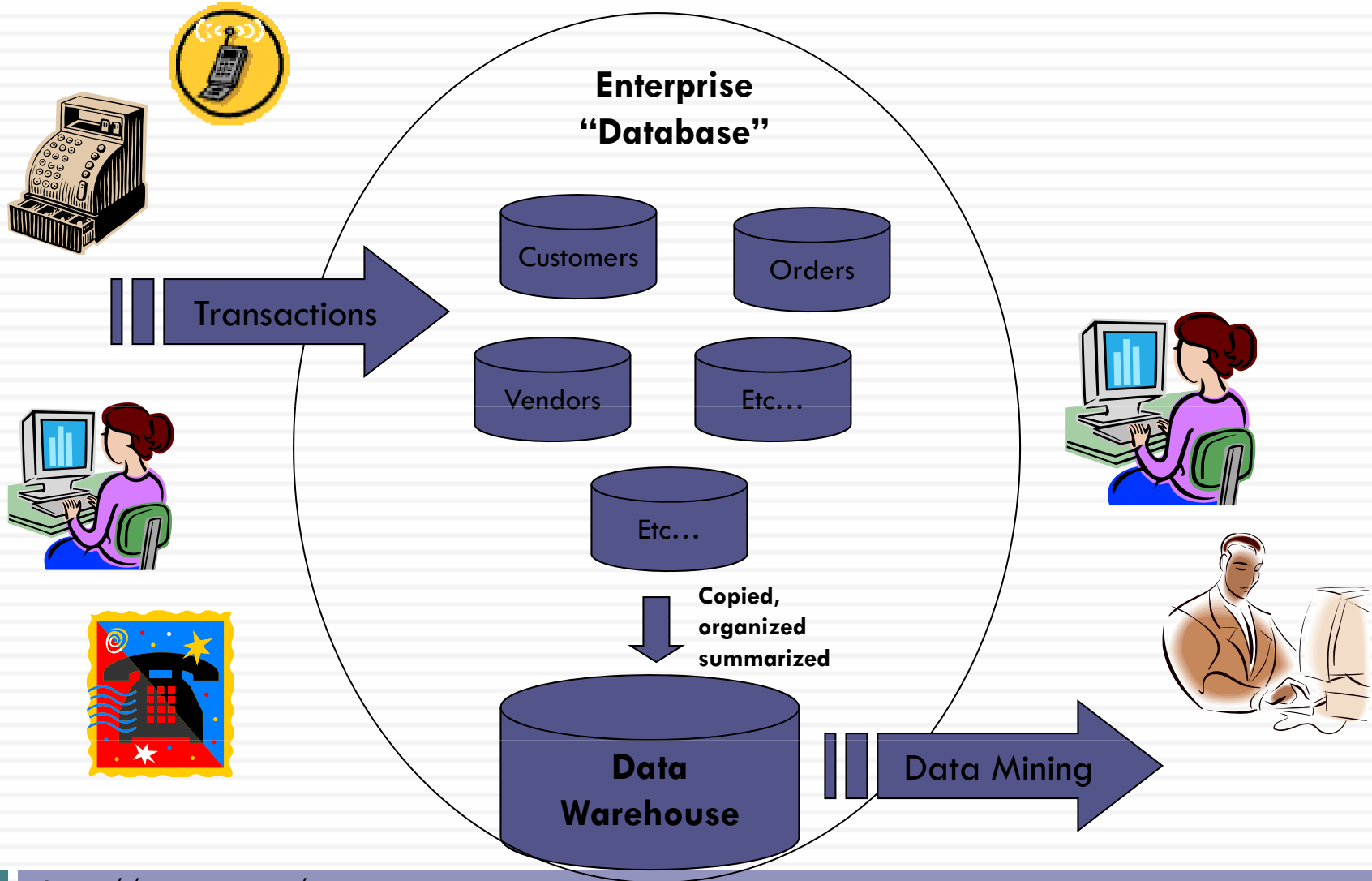
انباره داده یا پایگاه داده هوش تجاری

- بر اساس فلسفه خروج داده (گزارشگیری) شکل گرفته است.
- بر اساس کارایی بالا جهت بازیابی اطلاعات طراحی شده است.
- گرچه حذف افزونگی هدف اصلی طراحی نبوده است اما این افزونگی داده ها کنترل شده است.
- داده ها به گونه ای ذخیره شده که دسترسی به داده های مطلوب برای تحلیل های کلان کسب و کارها به راحتی ممکن است.
- معمولا هیچ داده ای راسا توسط کاربرد به خودی خود تولید نمی شود و کلیه داده ها یا قبلا موجود هستند و یا از طریق منابع داده های عملیاتی داخلی و خارجی به روز می شوند.

انبار داده ها

11

nemati@ut.ac.ir

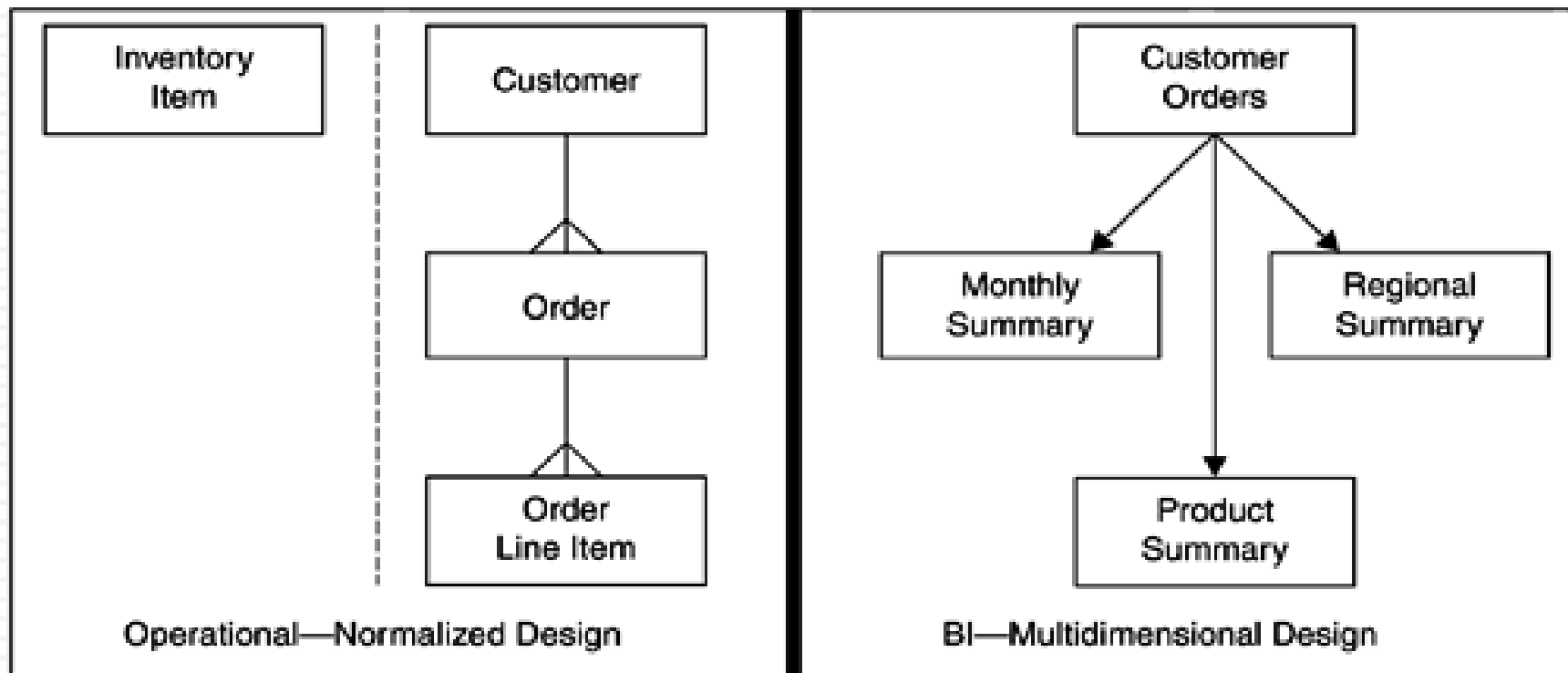


تفاوت طراحی پایگاه داده عملیاتی و پایگاه داده در BI

12

nemati@ut.ac.ir

- در سیستم های هوش تجاری، طراحی پایگاه داده به صورت چند بعدی است. (برای مثال در موضوع سفارشات مشتری، خلاصه اطلاعات هر ماه، محصولات مورد نظر و منطقه سفارش ذخیره شده است).
- در سیستم های عملیاتی پایگاه داده به صورت نرمال طراحی می شود و بدست آوردن این نوع گزارشات چند بعدی به سختی انجام می شود و به پرس و جوهای ترکیبی و سنگینی نیاز می باشد.



جدول تفاوت پایگاه داده عملیاتی و پایگاه داده در BI

پایگاه داده BI	پایگاه داده عملیاتی
با هدف پشتیبانی از انواع گزارشات مربوط به تجارت و کسب و کارهای کلان طراحی شده اند که این گزارش گیری هر روز مورد نیاز نمی باشد. (مثال: تحلیل گزارش های فروش منطقه ای سه ماهه و یا گزارش سفارشات ماهانه)	با هدف حذف افزونگی، هماهنگی در به روز رسانی، تکرار انواع عملیات روزانه مشابه طراحی شده اند. (مثال: عملیات بانکی)
با وجود اهمیت زمان پاسخ، پاسخ های زیر یک ثانیه مورد انتظار نیست و پاسخ ها از چند ثانیه تا چند دقیقه طول می کشد.	نیازمند زمان پاسخ زیر یک ثانیه می باشند.
به نرمال زدایی بالا برای بازیابی سریع حجم زیاد داده ها نیاز است.	به نرمالسازی بالا جهت پشتیبانی از به روز رسانی مداوم و حفظ یکپارچگی نیاز است.
حجم بالایی از داده های مشتق شده به منظور کاهش زمان گزارشگیری ذخیره می شوند.	داده های مشتق شده ی کمی ذخیره می شوند و در صورت نیاز، پویا استنتاج می شوند.
حجم بالایی از داده های تاریخی ذخیره می شوند.	داده های تاریخی ذخیره نمی شوند بلکه رکوردهای تاریخی آرشیو می شوند.
بسیاری از داده ها از قبل محاسبه شده اند و یا خلاصه شده اند.	خلاصه سازی اندک است و اغلب جهت گزارش گیری رخ می دهد.

نمایش چند بعدی داده ها در انبار داده

□ توانایی سازماندهی، ارائه و تحلیل داده به وسیله چند بعد. برای مثال:
فروش صورت گرفته بر اساس منطقه، کالا، فروشنده و زمان فروش (۴ بعد)

زمان	اندازه ها	ابعاد
<ul style="list-style-type: none">• روزانه• هفتگی• ماهانه• فصلی• سالانه	<ul style="list-style-type: none">• پول• میزان فروش• تعداد فروش• سود واقعی یا پیش بینی شده	<ul style="list-style-type: none">• کالاها• فروشنده ها• بخش های بازار• واحدهای تجاری• نواحی جغرافیایی• کانال های توزیع• کشور یا صنعت

دو نوع طراحی پایگاه داده در BI

15

nemati@ut.ac.ir



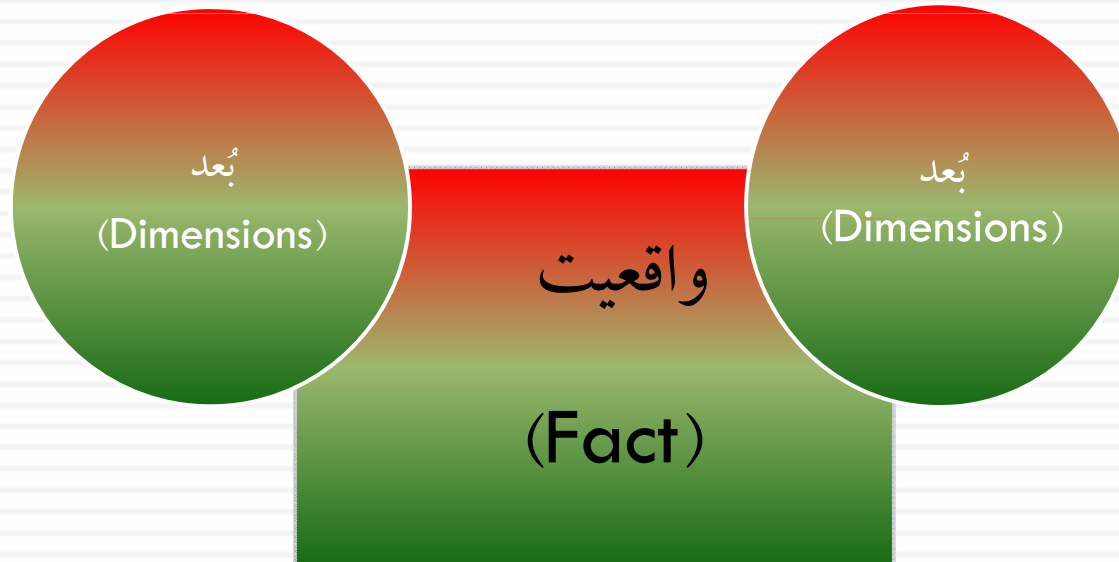
مدل دانه برفی
(Snowflake)



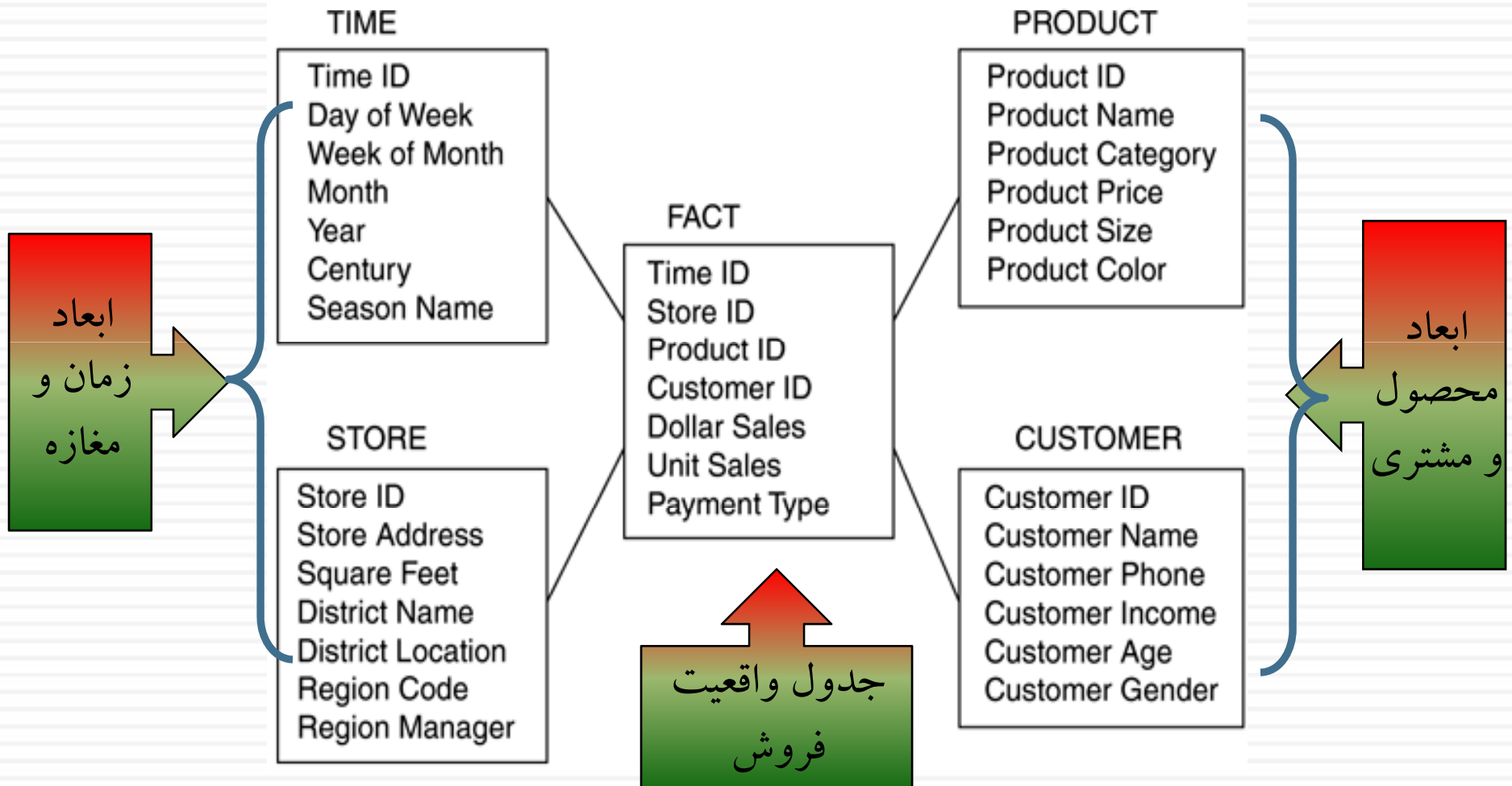
مدل ستاره ای
(Star)

پایگاه داده ستاره ای

- در مدل ستاره ای، یک مدل دوسطحی است که متناسب با هر تحلیل، داده به صورت آرایه ای از مقادیر که واقعیت (Fact) نامیده می شوند بازنمایی می شود. در این مدل هر واقعیت چند بعد (Dimension) دارد.
- مدل ستاره ای نمایی از یک سؤال یا پرس و جو در باره مسأله ای از کسب و کار را منعکس می کند.



مدل ستاره ای (مثال)



ویژگی های جداول واقعیت

- جدول واقعیت، یک واقعه یا رویداد مهم کسب و کار را نشان می دهد. (یک فعالیت یا مبادله در کسب و کار مثل فروش یا تقاضا)
- واقعیت های موجود در جدول، جنبه های کمی رویداد یا فعالیت کسب و کار را نشان می دهند. (ستون های جدول واقعیت)
- در هر جدول واقعیت، کلیدهای خارجی زیادی وجود دارد که به کلید اصلی جداول ابعاد مربوطه مثلا زمان، مشتری، و ... پیوند می خورد.
- ممکن است تعداد زیادی جدول واقعیت برای یک موضوع مشخص همراه با افزونگی داده ها وجود داشته باشد. هر جدول واقعیت می تواند شامل سطوح تجمیع متفاوتی بر روی داده های مشابه باشد. مثلا:
- واقعیت های فروش های مربوط به فروشگاه در منطقه در مدت زمان مشخص
- واقعیت فروش های مربوط به محصول در فروشگاه در مدت زمان معین
- واقعیت های فروش های مربوط به مشتری در منطقه در مدت زمان مشخص
- جداول حقیقت دراز و باریکند یعنی سطرهای زیاد و ستون های کمی دارند.

ویژگی های جدول بعد

19

nemati@ut.ac.ir

جداول بُعد، همیشه دارای کلید اصلی با یک ویژگی (one-attribute primary key) هستند.

جداول بُعد موجودیت ها یا اشیای کسب و کار هستند که هریک از واقعیت های موجود در جدول واقعیت را توصیف می کنند.

جداول بُعد نرمال نشده اند به گونه ای که داده های بخش های گوناگون کسب و کار را در خود جای می دهند و دارای افزونگی می باشند.

جداول بُعد کوتاه و عریض هستند یعنی سطرهای آنها کم و ستون ها زیاد است.

جداول بُعد باید توسط جداول واقعیت به اشتراک گذاشته شده و مطابق باشد.

یک بُعد اغلب زمان است که برچسب زمانی را نشان می دهد و شامل ویژگی های سال، ماه، دوره مالی و ... است.

فواید مدل ستاره ای

کارایی بالای گزارشگیری از داده های تاریخی سالیانه.

انعطاف پذیری بالا برای تحلیل داده های چندبعدی.

پشتیبانی از بسیاری از سیستم های مدیریت پایگاه داده های رابطه ای.

تحلیل ساده تر داده ها نسبت به پایگاه های نرمالیزه شده.

پایگاه داده مدل دانه برفی

21

nemati@ut.ac.ir

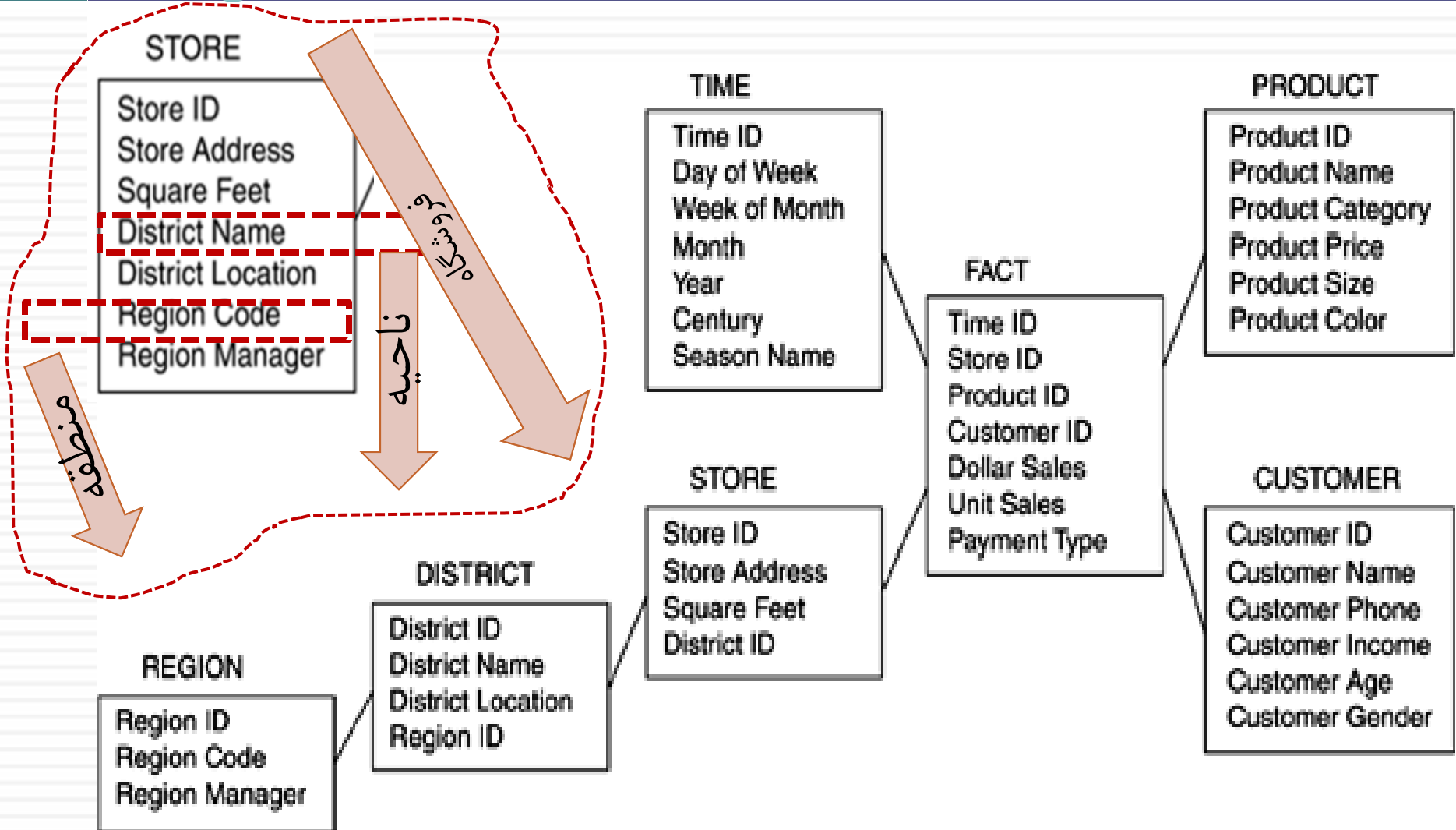
- همان مدل ستاره ای است با این تفاوت که هر بعد می تواند به صورت سطوح سلسله مراتبی به یک جدول دیگر متصل شود.
- جداول سطوح سلسله مراتب، جداول نرمالیزه هستند.
- تعداد جداول در این مدل بسیار زیاد است.



مدل دانه برفی (مثال)

22

nemati@ut.ac.ir



مزایا و معایب مدل دانه برفی

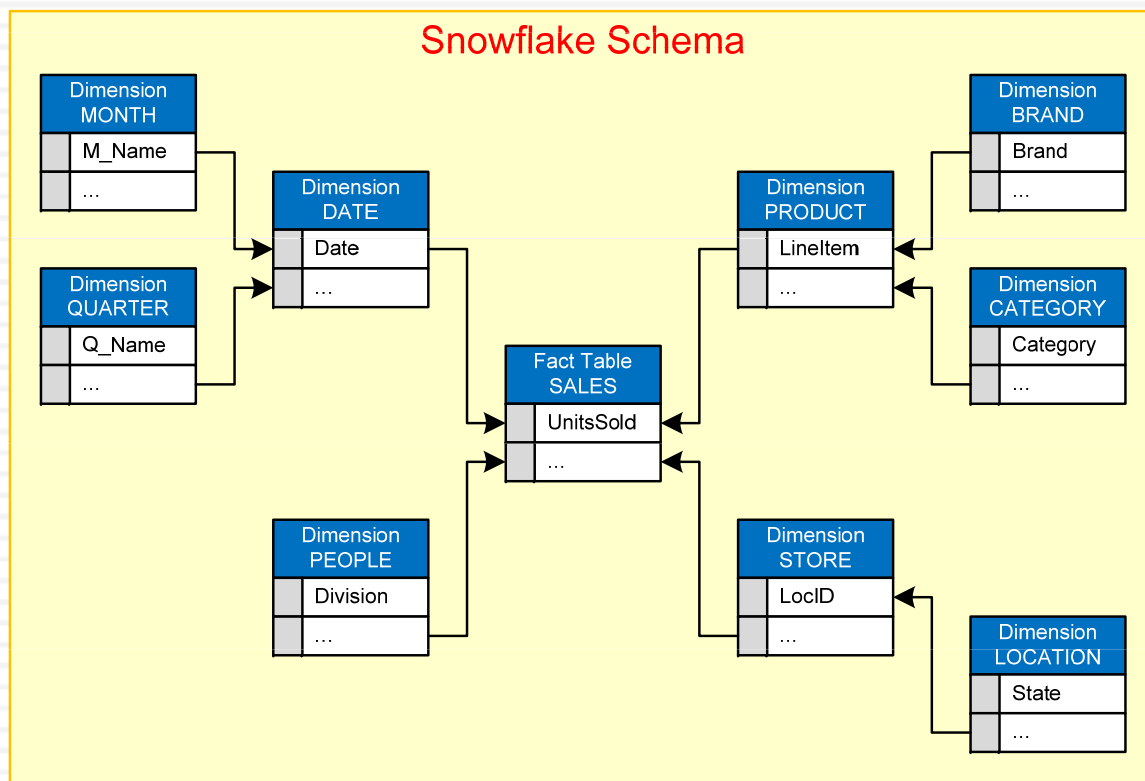
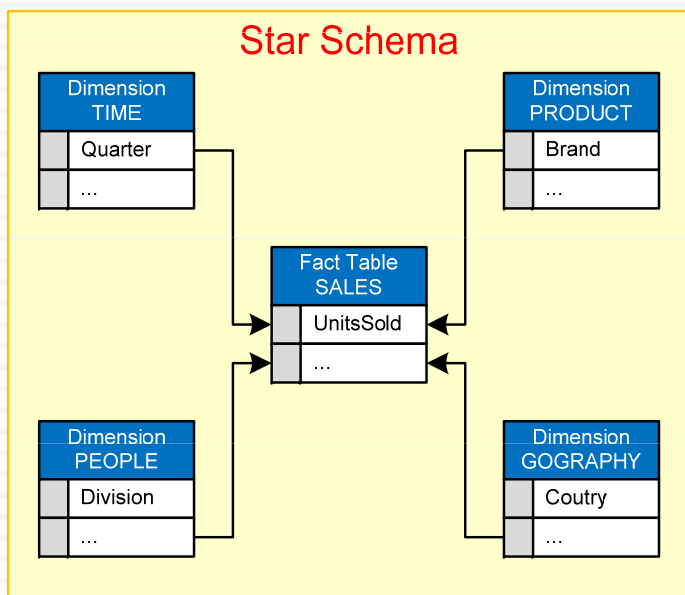
در این مدل اندازه جداول بعد کاهش می یابد و از افزودنی جلوگیری می شود.
برنامه های کاربردی منعطف تر می شوند.



کارایی به خصوص سرعت پایین تر است چرا که با افزایش تعداد جداول به تعداد بیشتری عملگر کند اجتماع نیاز داریم.
به علت افزایش تعداد جداول، هزینه نگهداری بالا می رود.



مقایسه مدل ستاره ای و دانه برفی





سیری تاریخی در تعاریف انبار داده

25

nemati@ut.ac.ir

۱۹۹۲: انبار داده، پایگاه داده ای به منظور تصمیم گیری است که داده های آن غیرقابل تغییر، در یک محدوده وسیع زمانی، یکپارچه و مبتنی بر موضوع می باشد.



۱۹۹۵: مجموعه ای از پایگاه داده های یکپارچه، مبتنی بر موضوع است که به صورت بهینه طراحی شده و برای پشتیبانی از سیستم های تصمیم یار مورد استفاده قرار می گیرد.



۱۹۹۶: منبعی که کاربران نهایی می توانند در آن به داده هایشان دسترسی پیدا کنند.



۱۹۹۷: مجموعه ای از اطلاعات یکی شده حاصل از سیستم های عملیاتی و برخی منابع داده خارجی با هدف پشتیبانی از تصمیم گیریهای تجاری.

انبار داده ها از نگاه اینمون (Inmon) پدر انبار داده!

26

nemati@ut.ac.ir

- A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process.

انبار داده عبارت است از مجموعه ای از داده های
موضوع گرا، یکپارچه، با متغیر زمانی و غیر فرار برای
پشتیبانی از فرآیند تصمیم گیری مدیران.

Bill Inmon

Bill Inmon

انبار داده ها از نگاه کیمبال (Kimball)

27

nemati@ut.ac.ir

- A data warehouse is a copy of transaction data specifically structured for query and analysis.

انبار داده عبارت است از یک رونوشت از داده های تراکنشی که به طور خاص برای پرس و جو و تحلیل ساخت یافته است.

Ralph Kimball

ویژگی های انبار داده

28

nemati@ut.ac.ir

- سازماندهی داده ها بسته به چگونگی ارجاع کاربران

موضوع گرایی

- یکپارچه سازی اسامی
- یکپارچه سازی واحدهای اندازه گیری متغیرها
- یکپارچه سازی ساختارهای داده ای از نظر محدوده مقادیر
- حذف ناسازگاریها و اطلاعات متناقض

یکپارچگی

- محیط های عملیاتی ← گستره زمانی ۶۰-۹۰ روز
- پایگاه داده تحلیلی ← گستره زمانی ۵-۱۰ سال

متغیر زمانی

- داده های فقط خواندنی

غیرفرار

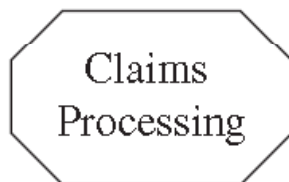
- سطح جزئیات و تفصیلی بودن داده ها

دانه دانه بودن

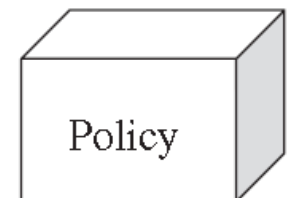
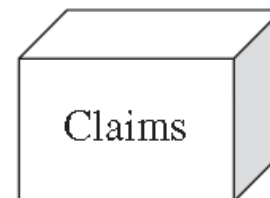
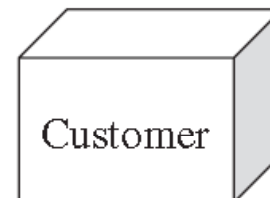
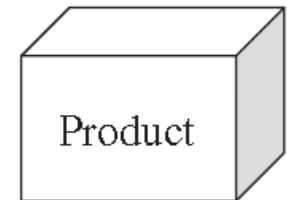
موضوع گرای داده های انبار داده

□ در انبار داده‌ها، داده‌ها بر اساس کاربردهای عملیاتی ذخیره نشده‌اند بلکه بر اساس موضوعات کسب و کار ذخیره شده‌اند.

Operational Applications

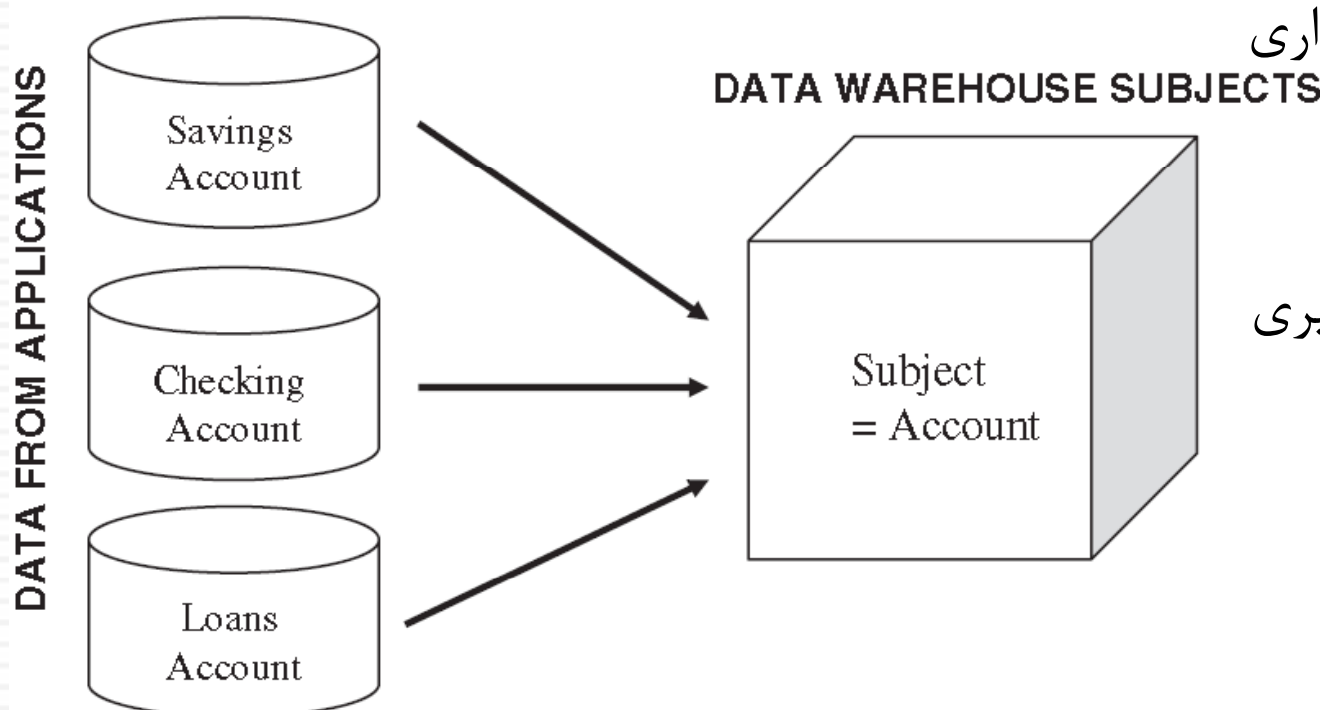


Data Warehouse Subjects



یکپارچه بودن داده ها در انبار داده

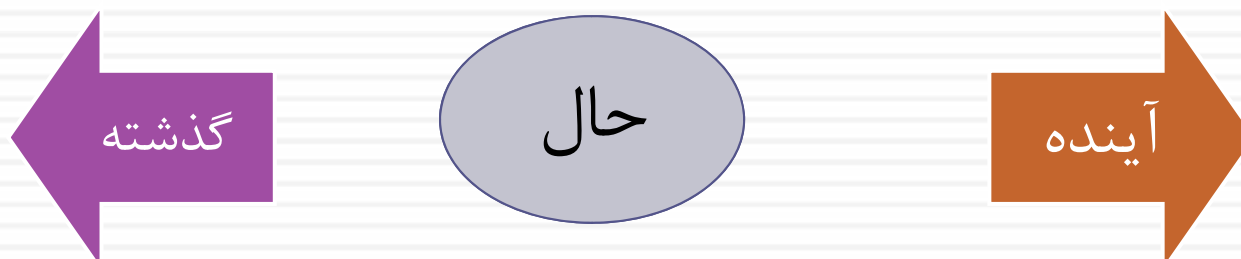
- در انبار داده، داده های ناسازگار حذف شده اند؛ داده ها از برنامه های کاربردی متفاوت یکپارچه شده است.
- از چند لحاظ باید داده ها را استاندارد و سازگار کرد:



- قراردادهای نام گذاری
- کدها
- ویژگی های داده
- واحدهای اندازه گیری

متغیر زمانی بودن داده ها در انباره داده

- ماهیت متغیر زمانی بودن داده ها در انباره داده سبب می شود تا:
 - تحلیل گذشته امکان پذیر باشد.
 - وضعیت جاری تشریح شود.
 - توانایی پیش بینی آینده به وجود آید.



- این ویژگی انباره داده بسیار کلیدی است، به طوری که حتی هوش تجاری را می توان از این منظر تعریف کرد. (اسلاید بعدی)

تعریف هوش تجاری از منظر انبارش داده ها ناشی از ویژگی متغیر زمانی داده ها

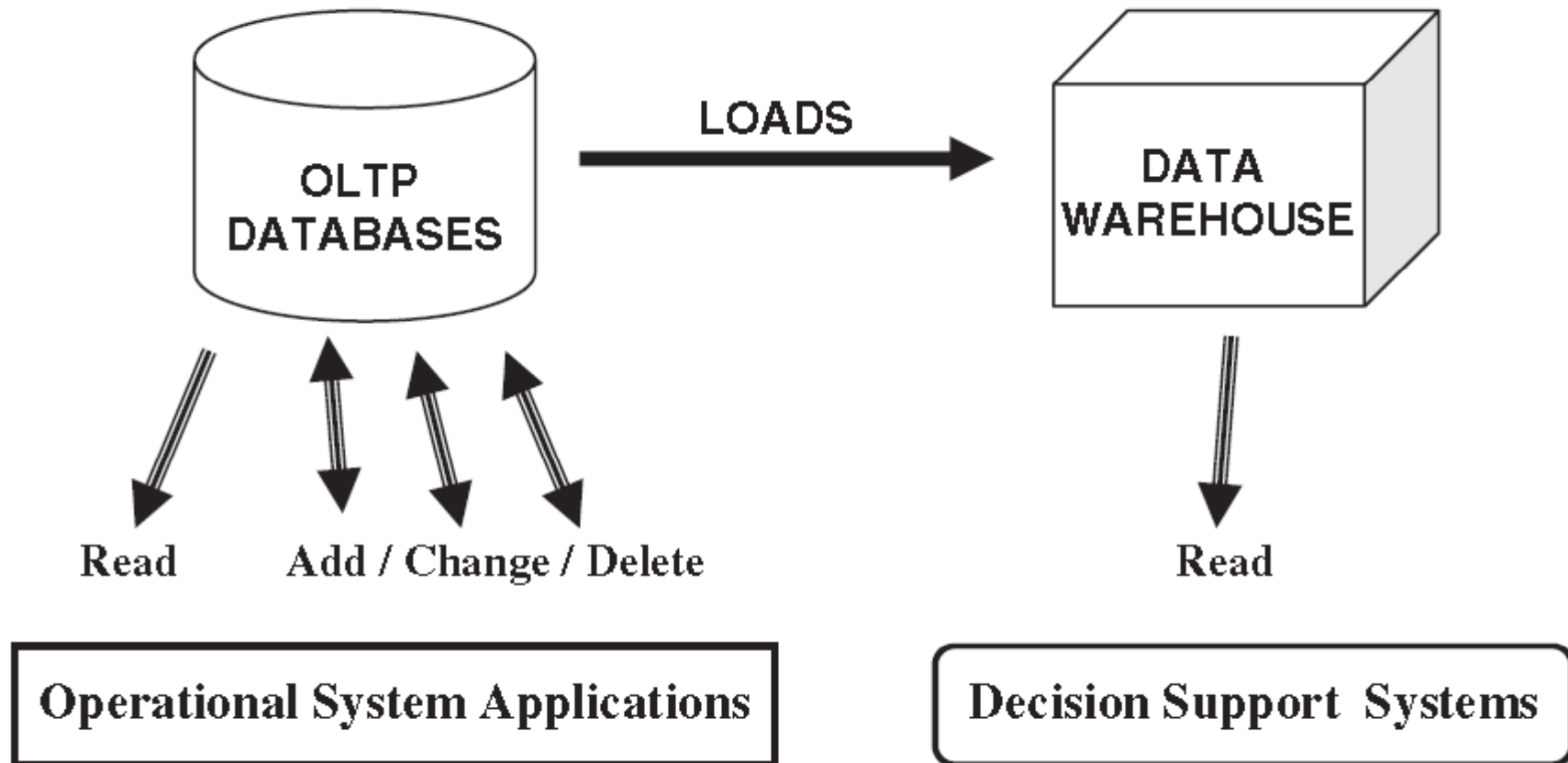
پایگاه داده تحلیلی در هوش تجاری، امکان بررسی و مطالعه رفتارها و کنش های گذشته یک بنگاه به منظور درک موقعیت قبلی سازمان، تعیین وضعیت جاری آن و پیش بینی یا تغییر آنچه که در آینده اتفاق خواهد افتاد، را فراهم می نماید.

فرار نبودن داده ها در انبار داده ها

33

nemati@ut.ac.ir

□ معمولا داده ها در انبار داده ها به روز رسانی و حذف نمی شوند و دائمی هستند.



دانه دانه بودن داده ها در انباره داده

Data Granularity

34

nemati@ut.ac.ir

□ دانه دانه بودن به جزئی و تفصیلی بودن داده ها بر می گردد و بسته به نیازها ممکن است چند سطح تفصیلی وجود داشته باشد. بسیاری از انباره داده ها تا دو سطح دانه بندی می شوند. مثال: سه سطح داده ها در انباره داده بانکداری

جزئیات روزانه

Daily Detail

Account

Activity Date

Amount

Deposit/Withdrawal

خلاصه ماهانه

Monthly Summary

Account

Month

Number of transactions

Withdrawals

Deposits

Beginning Balance

Ending Balance

خلاصه فصلی

Quarterly Summary

Account

Quarter

Number of transactions

Withdrawals

Deposits

Beginning Balance

Ending Balance

دیگر ویژگی های انبار داده

35

nemati@ut.ac.ir

- به علت وسعت قلمرو زمانی حجیم است.

حجیم

- داده های خلاصه شده و مناسب برای تصمیم گیری.

خلاصه شده

- به منظور افزایش کارایی اما افزودنی وجود دارد.

نرمال نشده

- برای کاربردی تحت وب طراحی می شود.

مبتنی بر وب

- ساختار آن یا رابطه ای یا چند بعدی است.

رابطه ای/چند بعدی

- معماری مناسب برای سهولت دسترسی کاربران نهایی

Client/Server

- انبار داده های جدید به صورت بلادرنگ هستند.

بلادرنگ

- حاوی فراداده برای سهولت سازماندهی و استفاده.

فراداده

انواع داده ذخیره شده در انبار داده

36

nemati@ut.ac.ir

داده جاری

- داده های فعلی منابع عملیاتی

داده قدیمی

- داده های قدیمی منابع اطلاعاتی

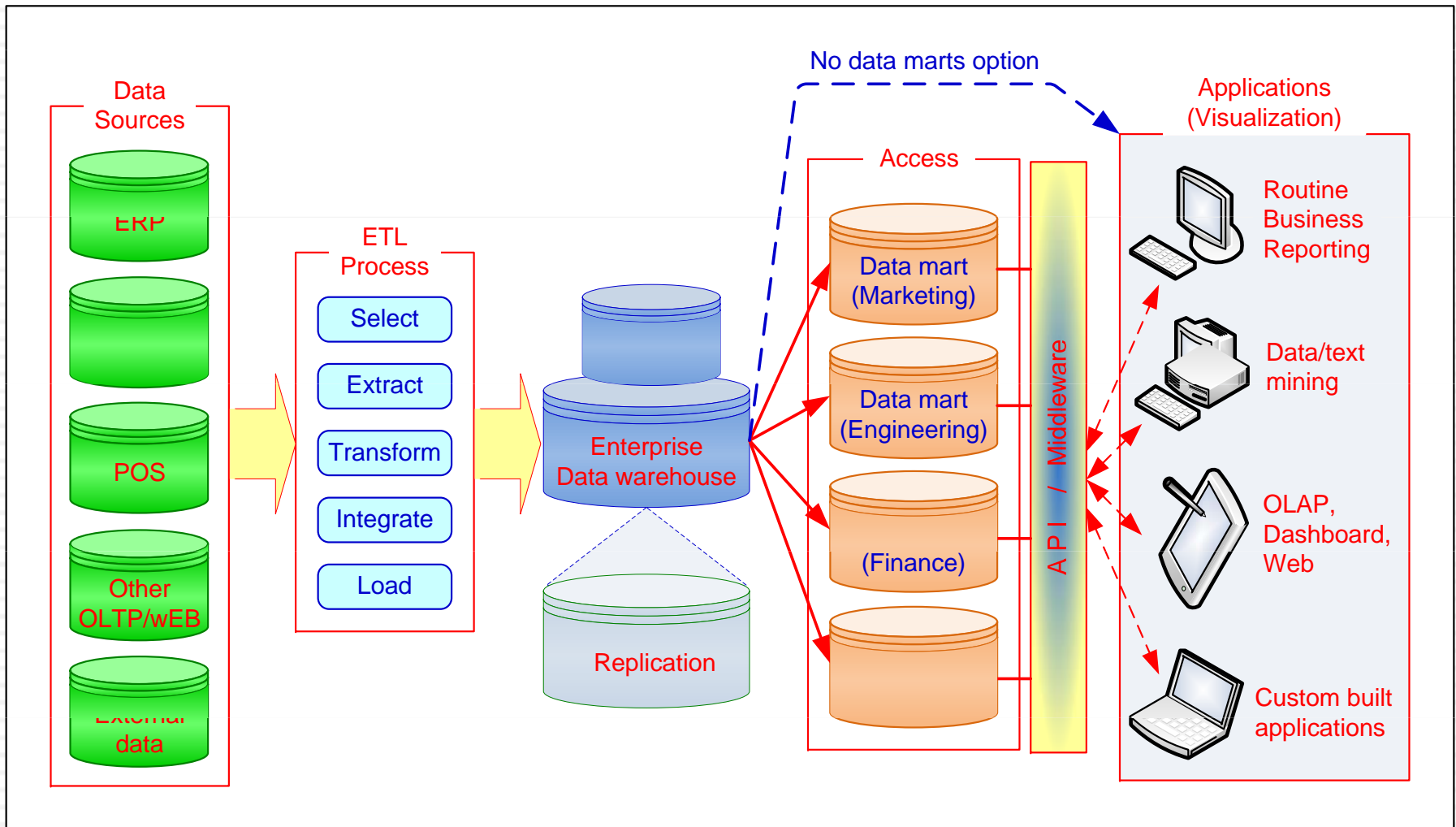
داده خلاصه شده

- داده های نتیجه شده از داده های موجود و فرآیندهای مرتبط

فراداده

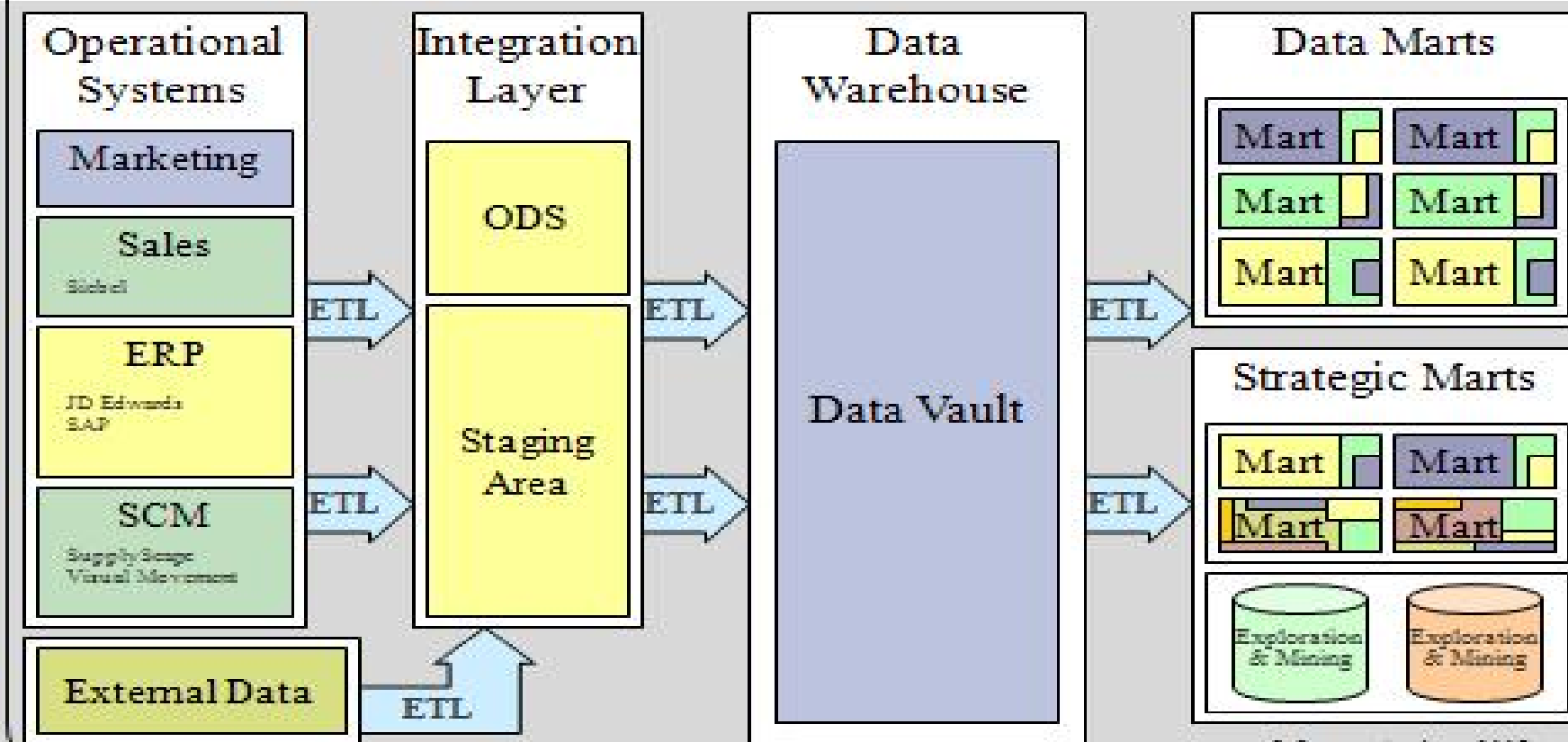
- اطلاعات مربوط به داده ها
- دیکشنری از اطلاعات پایگاه داده
- راهنمایی جهت نگاشت داده از محیط عملیاتی به محیط پایگاه داده تحلیلی
- قواعد استفاده شده برای خلاصه سازی

چارچوب و نماهای انبار داده

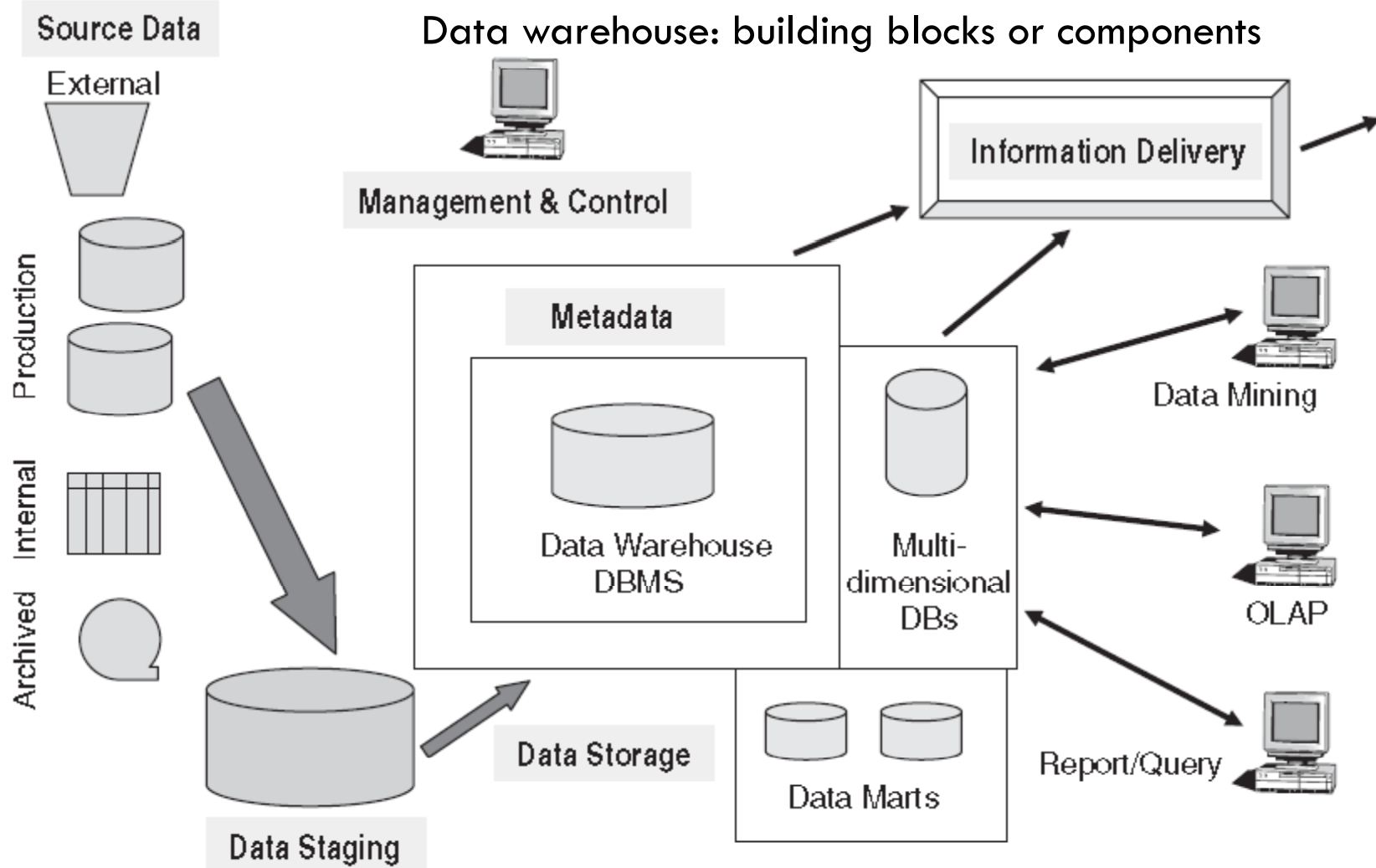


اجزای عمومی انبار داده ها

Data Warehouse



بلوک های سازنده یا مؤلفه های اصلی انبار داده

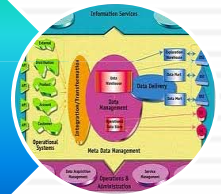


تعریف اجزای مهم و عمومی انبار داده

40

nemati@ut.ac.ir

مخزن داده های عملیاتی (Operational Data Store - ODS): نوعی پایگاه داده که مکانی موقت برای ذخیره داده ها است.



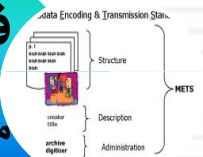
انبارک داده یا مرکز داده عملیاتی (Operational mart): یک مرکز داده یا انباره داده ی عملیاتی.



انبار داده سازمانی (Enterprise DW - EDW): نوعی انبار داده برای سراسر بنگاه.



فراداده (Meta Data): داده درباره داده. در یک انباره داده، فراداده محتوای انباره داده و روش دستیابی و استفاده از آن را شرح می دهد.



مخازن داده های عملیاتی (ODS)

- ODS نوعی پایگاه داده است که به عنوان یک نقطه عملیاتی موقت برای انباره سازی داده های اصلی مورد استفاده قرار می گیرد.
- داده های آن موقت و فرار است و شبیه یک حافظه کوتاه مدت است.
- با تعدادی پایگاه داده ی عملیاتی مرتبط است.
- از جهت عملیات ETL شبیه انباره داده است.
- متناسب و همزمان با عملیات کسب و کار به روز رسانی می شود.
- برای تصمیمات کوتاه مدت و مأموریتی مورد استفاده قرار می گیرد.
- برای تحلیل های چند بعدی بر روی داده ها بسیار مفید است.
- خوراک انبارک داده های عملیاتی (Oper. Mart) است.

فراداده (Meta Data)

- فراداده، داده ای در مورد داده ها است که کاربردهای فراوانی دارد.
- فراداده در انباره داده شبیه دیکشنری داده ها و یا کاتالوگ داده ها در مدیریت پایگاه داده ها است.
- در دیکشنری داده ها، اطلاعاتی در باره ساختار منطقی داده ه، اطلاعات در باره قایل ها و آدرس ها، اطلاعات در باره ایندکسها نگهداری می شود. فراداده ها در انباره داده تقریبا چنین هستند اما نقشی فراتر دارند. فراداده ها نقش دایرکتوری را برای محتوای انباره داده بازی می کنند.
- فراداده ها به یکپارچه سازی داده های موجود در منابع داده ای متفاوت چه در داخل سازمان و چه در بیرون کمک می کنند و زمینه را برای معماری فراتجاری (MetaBusiness) فراهم می کند.
- اکثر سازمان ها فراداده را به درستی نمی شناسند و از اهمیت آن غافلند.

طبقه بندی انواع فراداده ها

- فراداده ها بر حسب نوع کاربرد دو دسته اند:
 - فراداده تاکتیکی: در یک محدوده خاص و احیانا به طور موقت
 - فراداده استراتژیک: در کل سازمان و به طور دائمی
- فراداده ها بر حسب الگوی توصیف دو دسته اند:
 - فراداده نحوی (Syntactic MD): توصیف ساختار و مدل داده ها
 - فراداده معنایی (Semantic MD): توصیف معانی داده ها
- فراداده ها را می توان به سه دسته تقسیم کرد:
 - فراداده عملیاتی (Operational metadata)
 - فراداده استخراج و تبدیل (Extraction and transformation metadata)
 - فراداده کاربرنهایی (End-user metadata)

فراداده عملیاتی (Operational metadata)

- فراداده عملیاتی با انتقال داده ها از منابع عملیاتی داده ها به انباره داده سر و کار دارد.
- یکپارچه سازی ساختار داده های منابع عملیاتی
- یکپارچه سازی المان های داده ای (طول فیلدها و نوع داده ها)
- شکستن رکوردها، ترکیب بخش هایی از رکوردها و ...

فراداده های استخراج و تبدیل (Extraction and transformation metadata)

45

nemati@ut.ac.ir

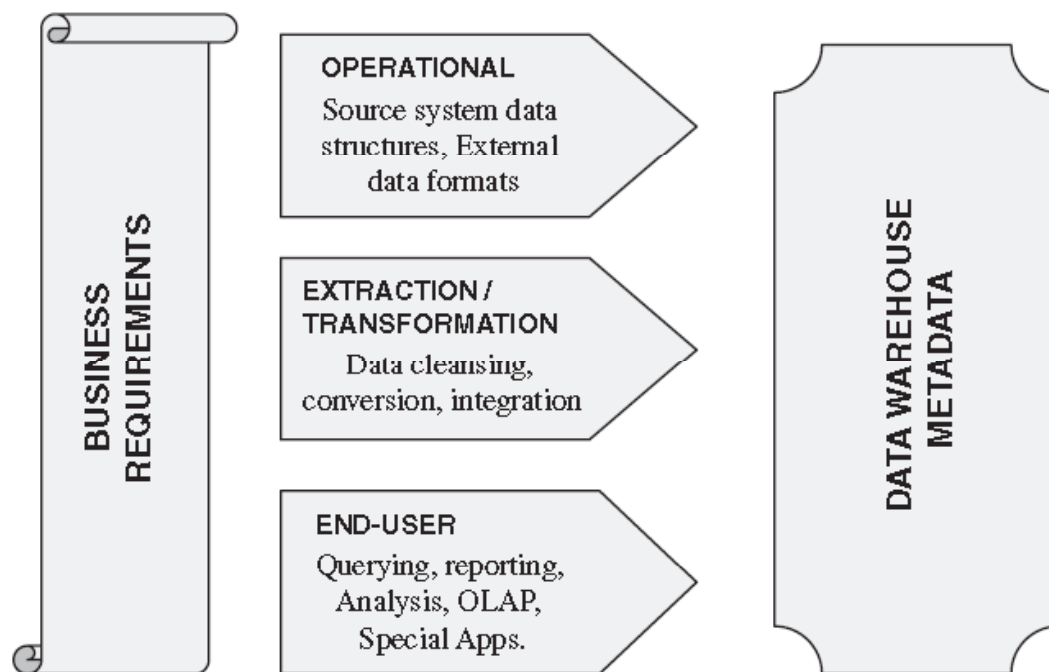
- فراداده های استخراج و تبدیل شامل داده هایی در باره انتقال داده ها از سیستم های منبع است:
- تناوب(فراوانی) استخراج
- روشهای استخراج
- قواعد کسب و کار برای استخراج داده ها
- اطلاعات در باره تمام تبدیل های که در ناحیه عملیات ETL اتفاق می افتد.

فراداده کاربرنهایی (End-user metadata)

46

nemati@ut.ac.ir

□ فراداده کاربرنهایی در واقع نقشه ناوبری (Navigational Map) انباره داده است و باعث می شود تا کاربران نهایی بتوانند اطلاعات موجود در انباره داده را بر اساس اصطلاحات تجاری خاص خودشان پیدا کنند.



مخزن فراداده ها (Metadata Repository)

47

nemati@ut.ac.ir

ناوبری اطلاعات

METADATA REPOSITORY

Information Navigator

Navigation routes through warehouse content, browsing of warehouse tables and attributes, query composition, report formatting, drill-down and roll-up, report generation and distribution, temporary storage of results

فراداده های کسب و کار

Business Metadata

Source systems, source-target mappings, data transformation business rules, summary datasets, warehouse tables and columns in business terminology, query and reporting tools, predefined queries, preformatted reports, data load and refresh schedules, support contact, OLAP data, access authorizations

فراداده های تکنیکی

Technical Metadata

Source systems data models, structures of external data sources, staging area file layouts, target warehouse data models, source-staging area mappings, staging area-warehouse mappings, data extraction rules, data transformation rules, data cleansing rules, data aggregation rules, data loading and refreshing rules, source system platforms, data warehouse platform, purge/archival rules, backup/recovery, security

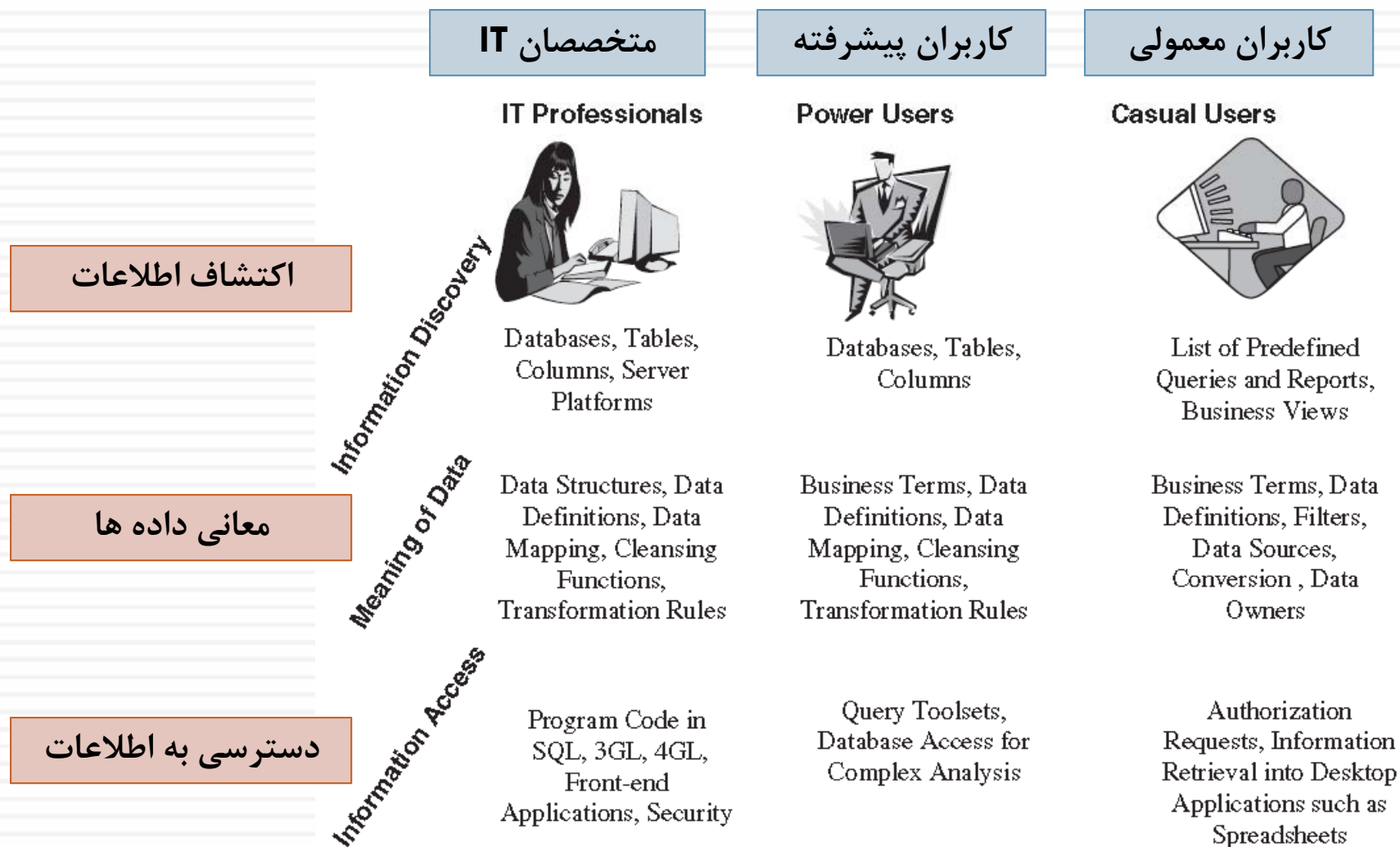
اهمیت فراداده ها برای انباره داده ها

- اهمیت فراداده از این جهت است که:
 - اولاً: فراداده مثل چسبی است که اجزای مختلف انباره داده را به هم متصل می کند. (سریش فراداده!)
 - ثانياً: اطلاعاتی در باره محتوای و ساختار پایگاه داده برای توسعه دهندگان انباره داده فراهم می کند.
 - سرانجام: راهی را برای کاربران نهایی باز می کند تا با استفاده از اصطلاحات خودشان با محتوای انباره داده روبرو شوند.

چه کسی به فراداده ها نیاز دارد؟

49

nemati@ut.ac.ir



انبارک داده (مرکز داده) – Data Mart

50

nemati@ut.ac.ir

- نوع خاصی انباره داده که کوچکتر بوده و برای بخشی از شرکت یا موضوعی خاص طراحی می شود.
- به جدول مقایسه DW و DM

انبارک داده (DM)	انباره داده (DW)
به صورت بخشی	گسترده در پهنه بنگاه یا شرکت
فرآیندی منفرد از کسب و کار	مجموعه ایی متحد از کلیه انبارک های داده ها
اتصال ستاره ای (وقایع و ابعاد)	داده های دریافت شده از ناحیه عملیات (ETL)
بهینه سازی تکنولوژی برای دسترسی و تحلیل	پرس و جوها از طریق منابع بازنمایی (لایه نمایش)
ساختار متناسب با نمای بخشی از داده ها	ساختار برای نمای کامل (کل شرکت) از داده ها
	سازمان دهی شده بر روی مدل موجودیت-رابطه (E-R)

انبارک داده ها(مرکز داده ها-Data Mart)

51

nemati@ut.ac.ir

DM انبار داده ی بخشی است که فقط داده های مرتبط به آن بخش را ذخیره می کند و بر دو نوع است:

Data Mart مستقل

- یک انبار داده کوچک که برای یک واحد تجاری یا دپارتمان طراحی شده است.

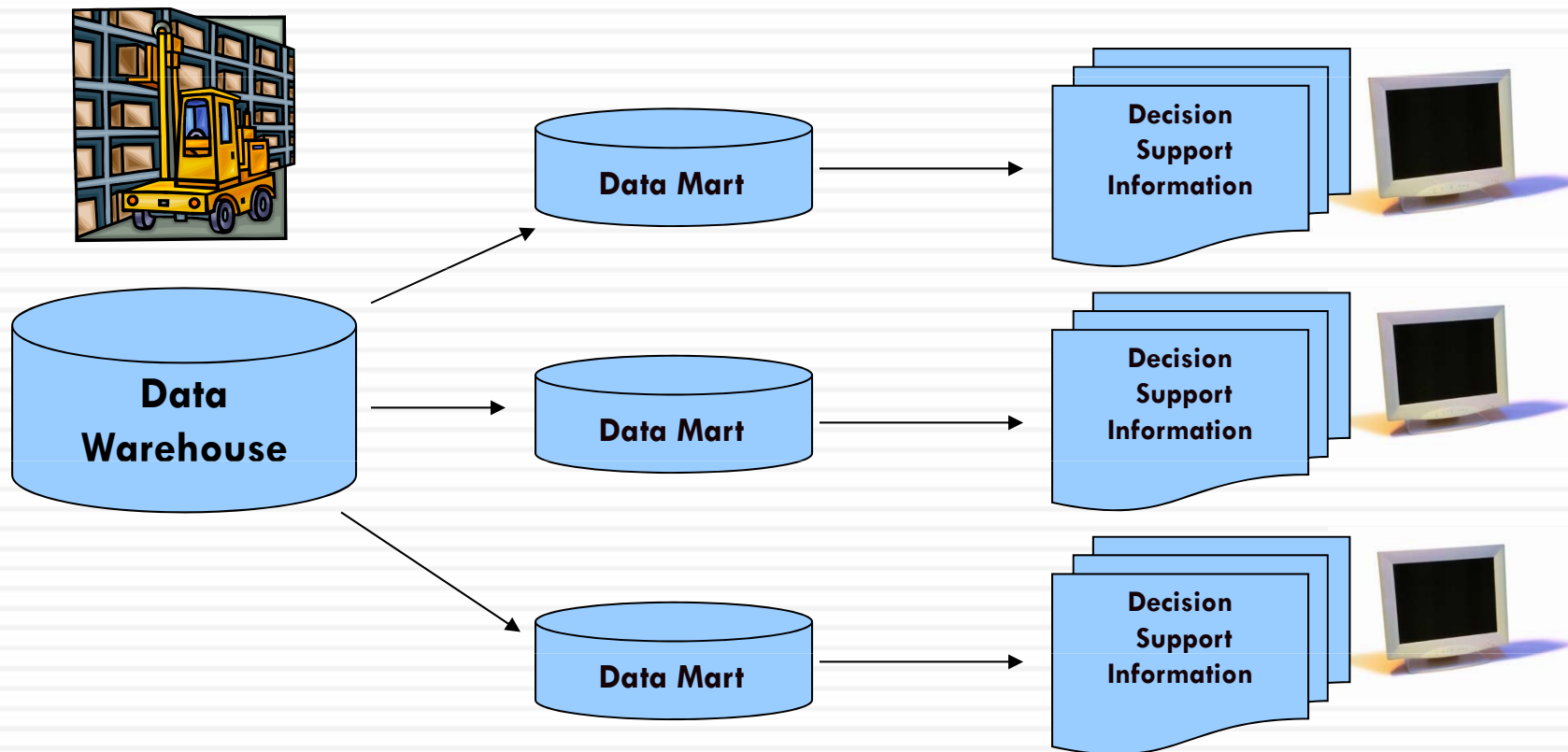
Data mart وابسته

- زیرمجموعه ای که به صورت مستقیم از یک انبار داده به دست می آید.

معماری DM وابسته

52

nemati@ut.ac.ir



چند سؤال اساسی قبل از شروع ساخت

□ قبل از تصمیم گیری در باره ساخت انباره داده ها برای سازمان باید سؤالات اساسی زیر پاسخ دهیم:

□ ساخت با دیدگاه پایین به بالا یا بالا به پایین؟

□ ساخت در گستره کل بنگاه یا در بخشی از آن؟

□ کدام یک اول ساخته شود: انبارک داده یا انباره داده؟

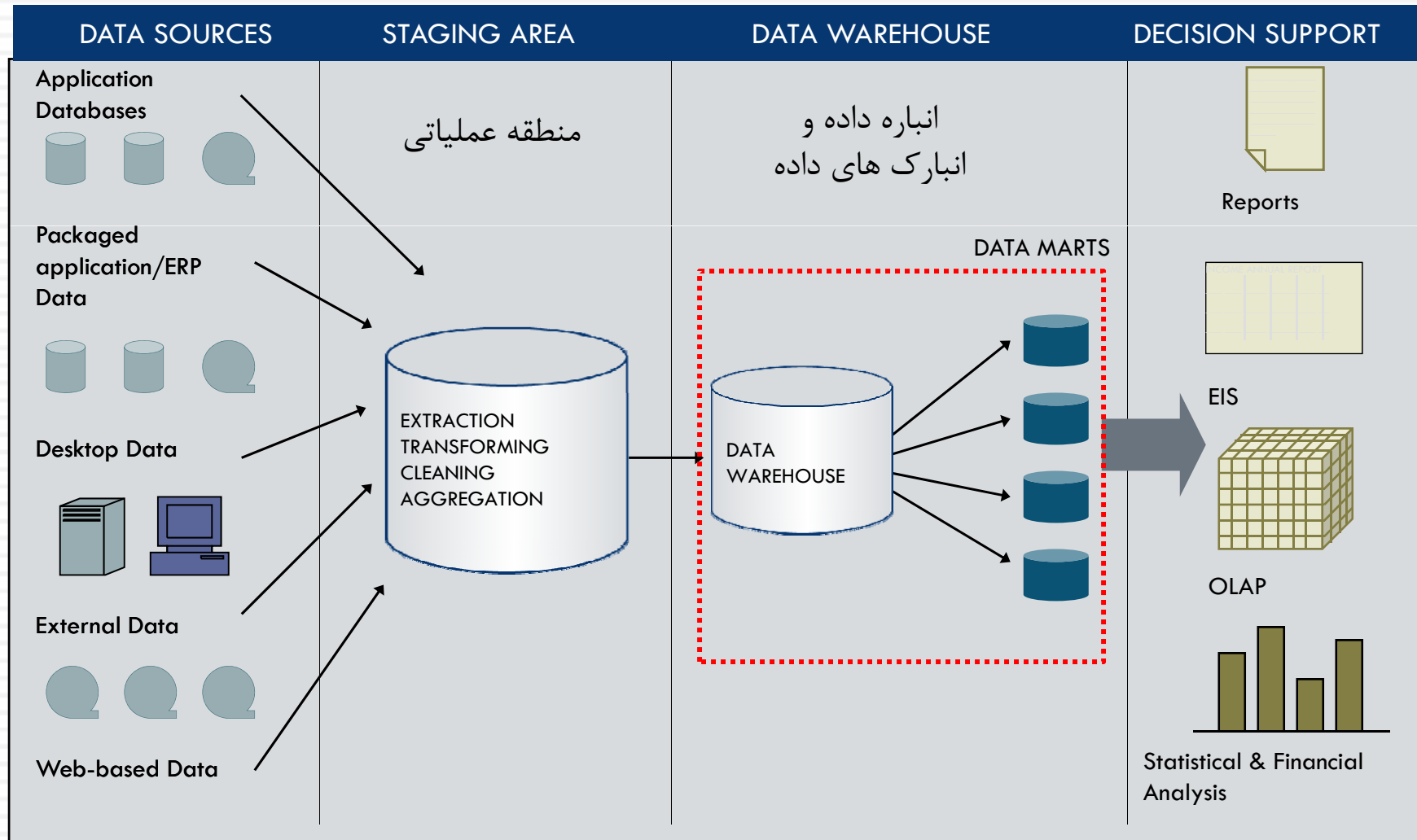
□ ساخت به صورت پایلوت یا به صورت کامل؟

□ ساخت انبارک داده مستقل یا وابسته به انباره داده؟

مروری بر فرآیند انباره سازی داده ها (۱)

54

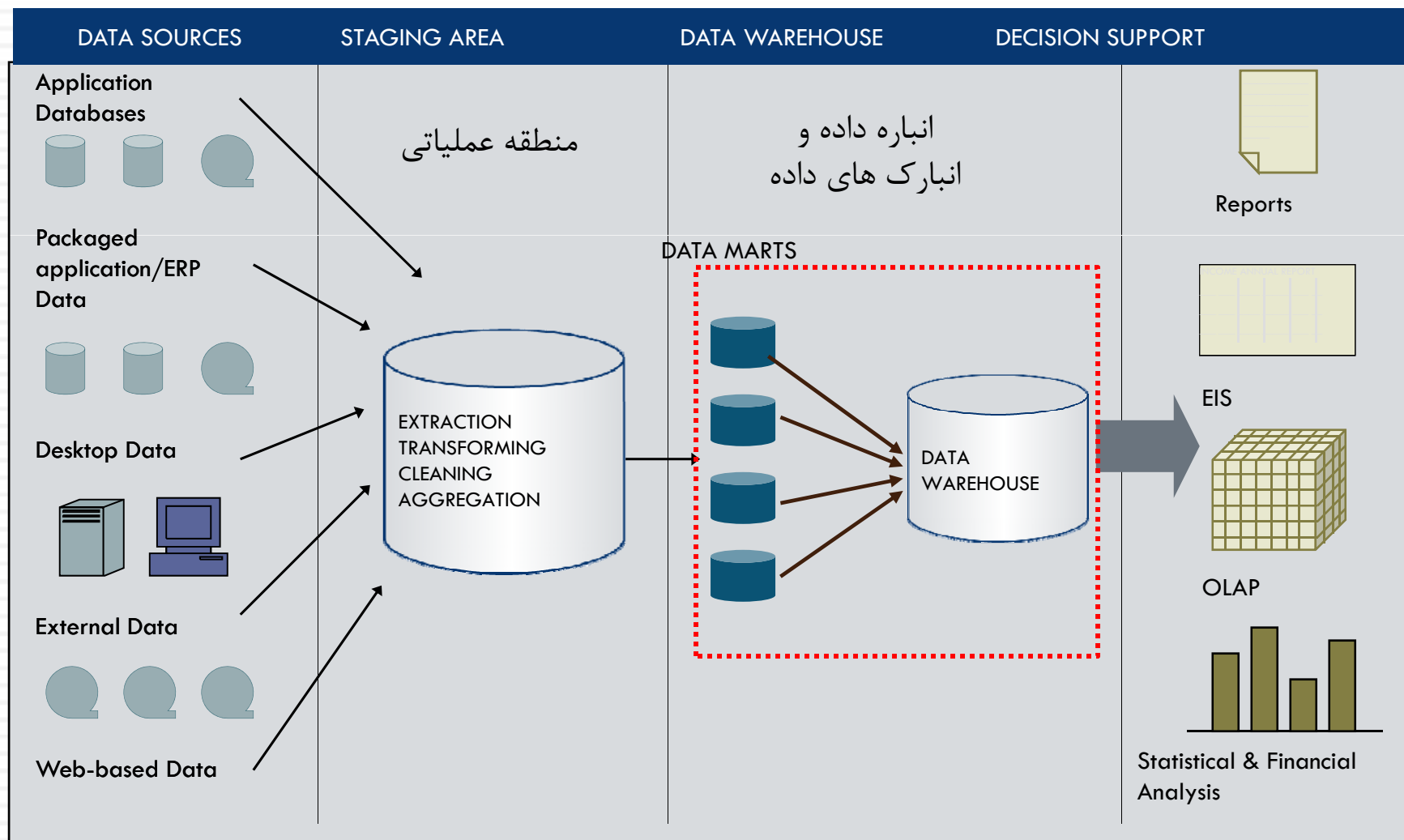
nemati@ut.ac.ir



مروری بر فرآیند انبار داده ها (۲)

55

nemati@ut.ac.ir



معماری انبار داده ها

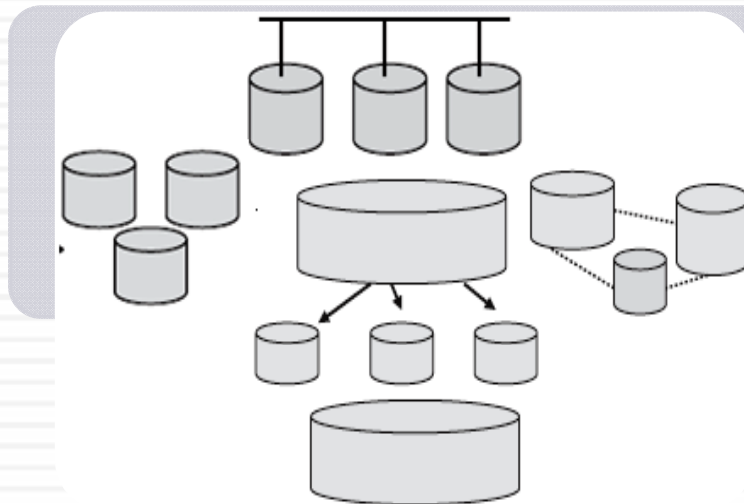
56

nemati@ut.ac.ir

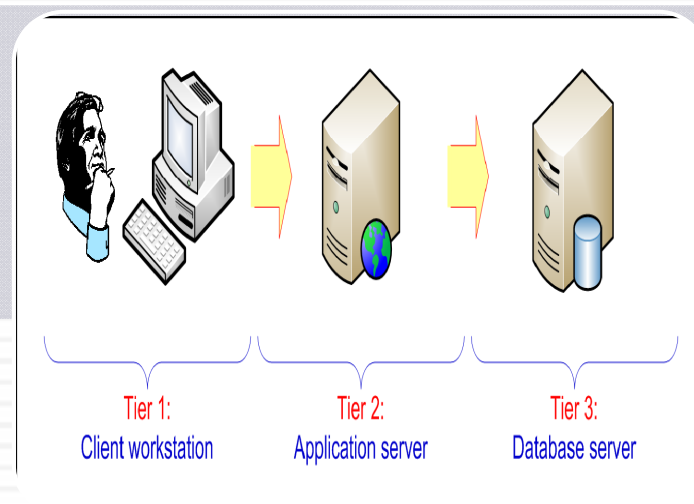
□ دو دسته معماری برای انبار داده می توان در نظر گرفت.

□ معماری لایه ای

□ معماری ذخیره سازی



معماری ذخیره سازی



معماری لایه ای

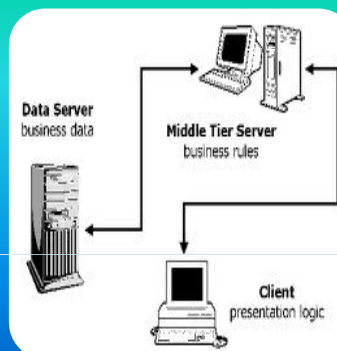
معماری های لایه ای انبار داده

57

nemati@ut.ac.ir

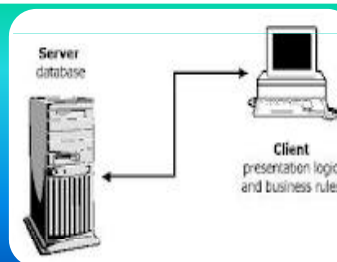
معماری سه لایه

- نرم افزار اکتساب داده (Bach-end)
- یک انبار داده که شامل داده و نرم افزار است.
- نرم افزار سرویس گیرنده (front- End) که به کاربران اجازه می دهد به داده های موجود در انبار داده ها دسترسی داشته و به تحلیل پردازند.



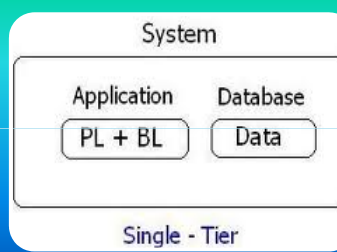
معماری دو لایه

- دو لایه اول در معماری ۳ لایه ای را در یک لایه قرار می دهد.



معماری یک لایه

- کل ۳ لایه را در یک لایه قرار می دهد.



معماری سه لایه

58

nemati@ut.ac.ir

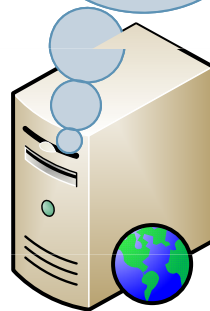
سرویس گیرنده هایی که با عملیات داده کاوی، پرس و جو و تحلیل، داده دریافت می کنند.

سرویس دهنده های پردازش تحلیلی بر خط که داده چند بعدی را برای ارائه به کاربران نهایی در اختیار قرار می دهند.

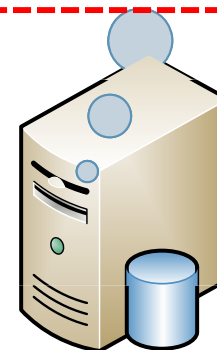
پایگاه های داده رابطه ای، فایل های مسطح و منابع دیگر



Tier 1:
Client workstation



Tier 2:
Application server

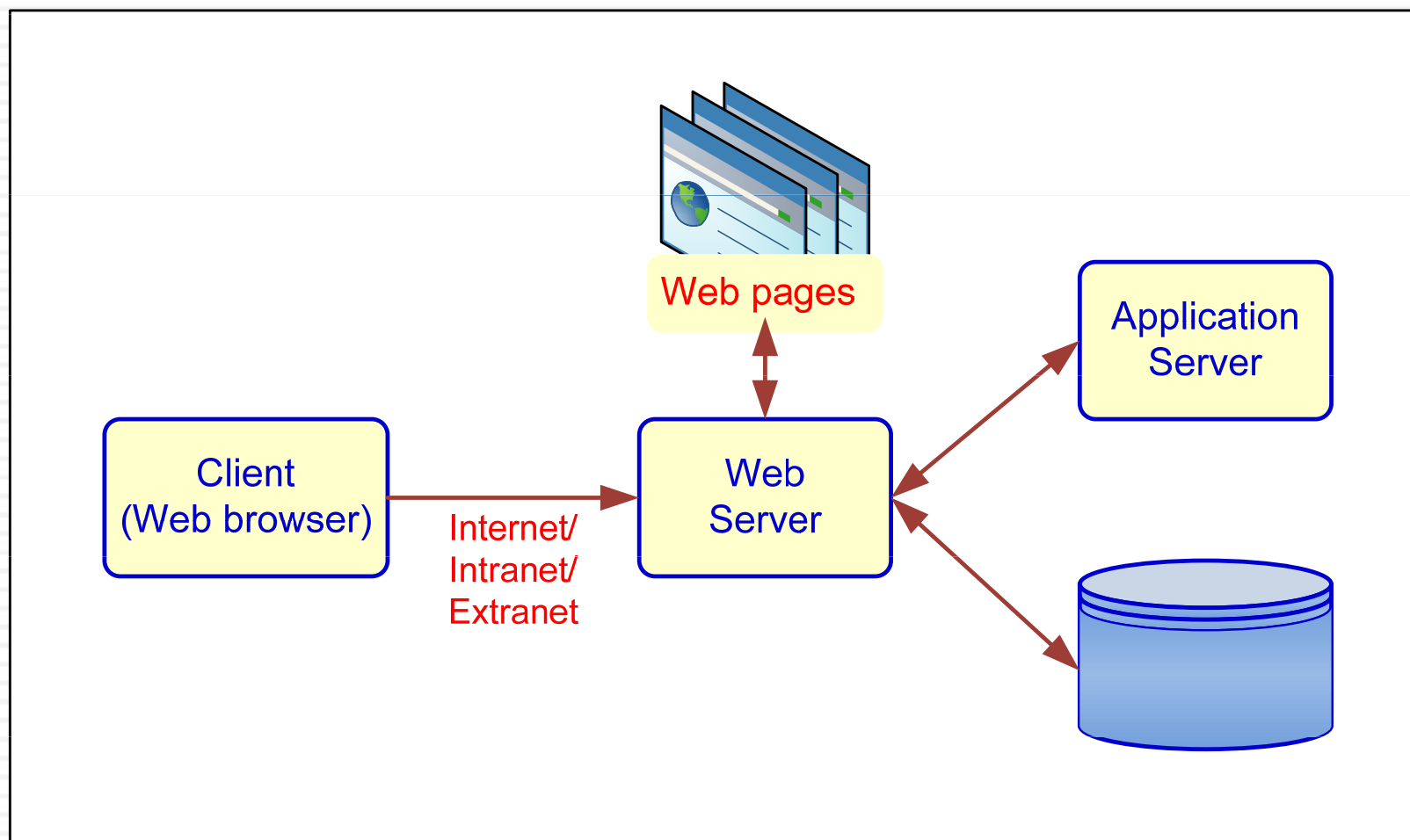


Tier 3:
Database server

معماری انبار داده مبتنی بر وب

59

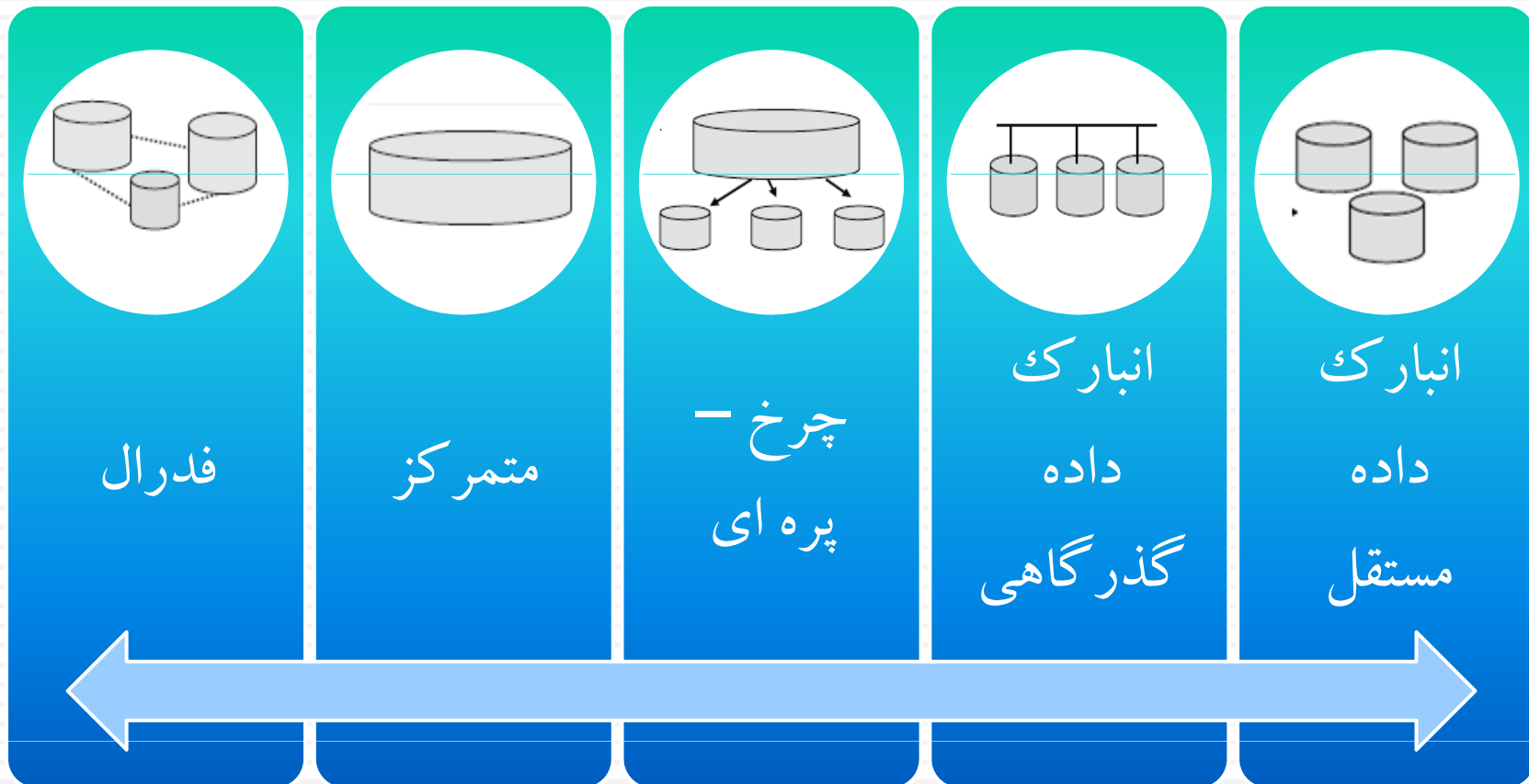
nemati@ut.ac.ir



معماری انبار داده بر اساس نحوه ذخیره داده و ارتباطات بین آنها

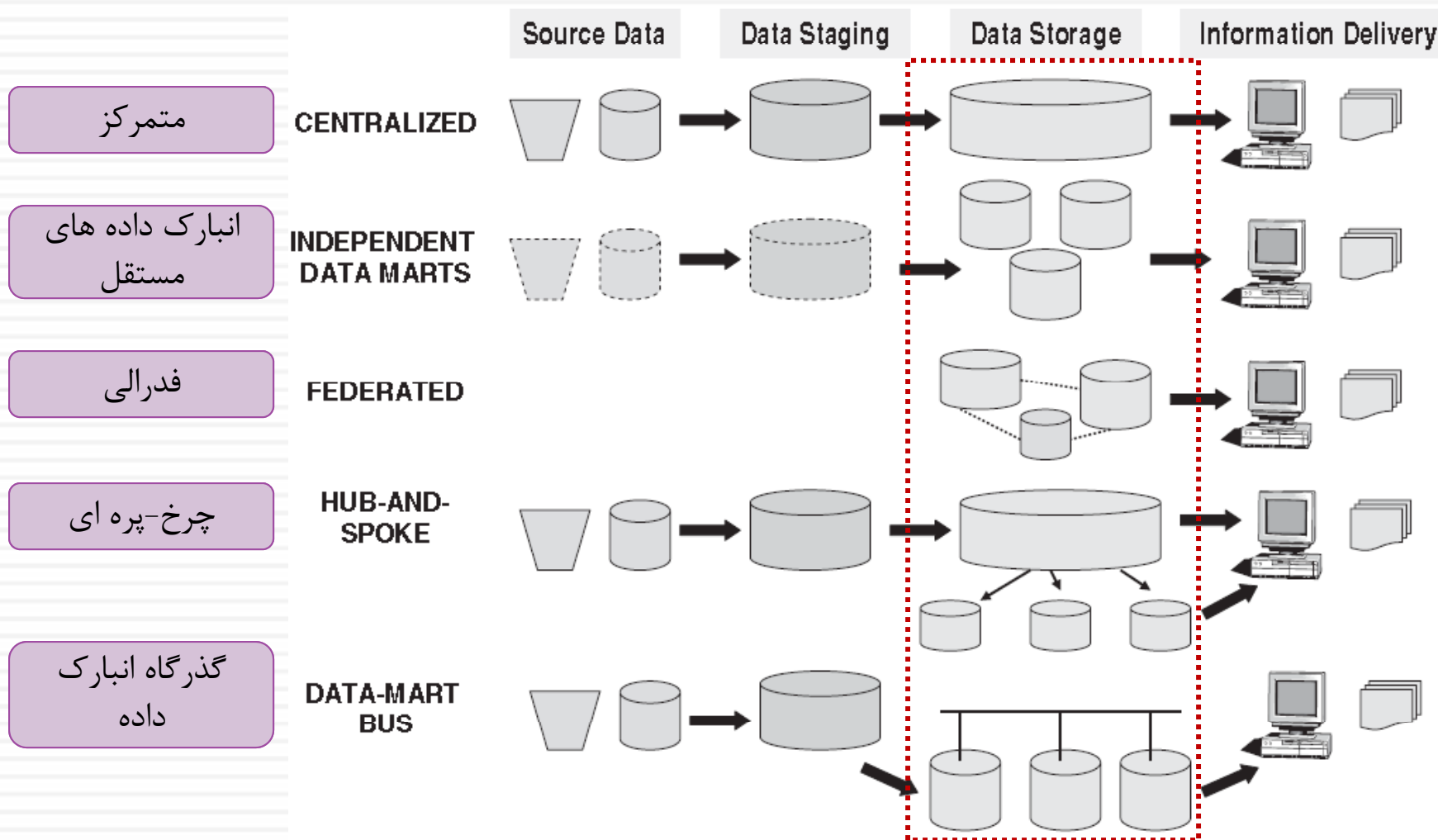
60

nemati@ut.ac.ir

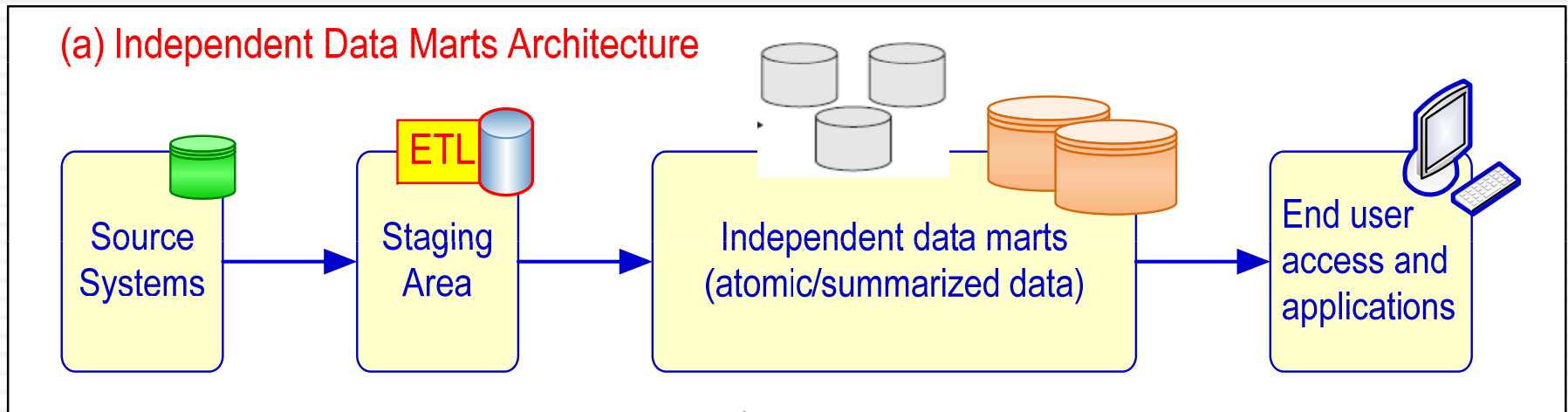


منابع مورد استفاده: [۱] [۳]

انواع معماری انبار داده (ذخیره سازی)



معماری Data Mart های مستقل

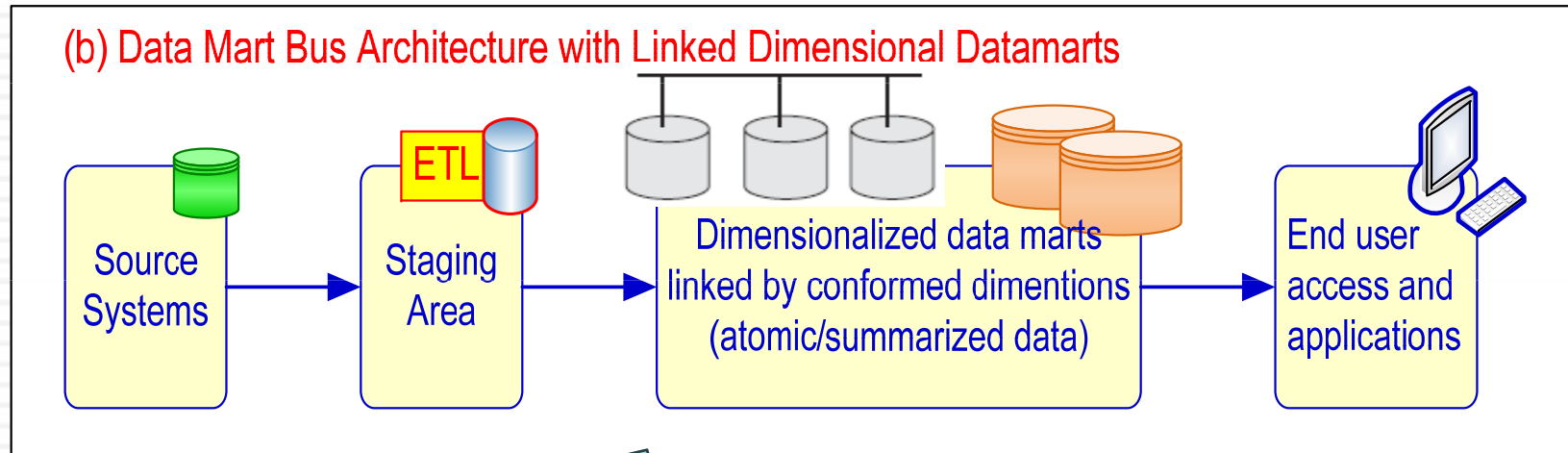


- زمانی به کار می رود که واحدهای سازمانی، Data Mart هایی را برای مقاصد خاص خود توسعه داده اند.
- Data Mart ها کاملاً به صورت مستقل بوده و دارای تعاریف و استانداردهای ناسازگار می باشند.
- تحلیل سراسری غیرممکن است.

معماری گذرگاهی (Data Mart Bus)

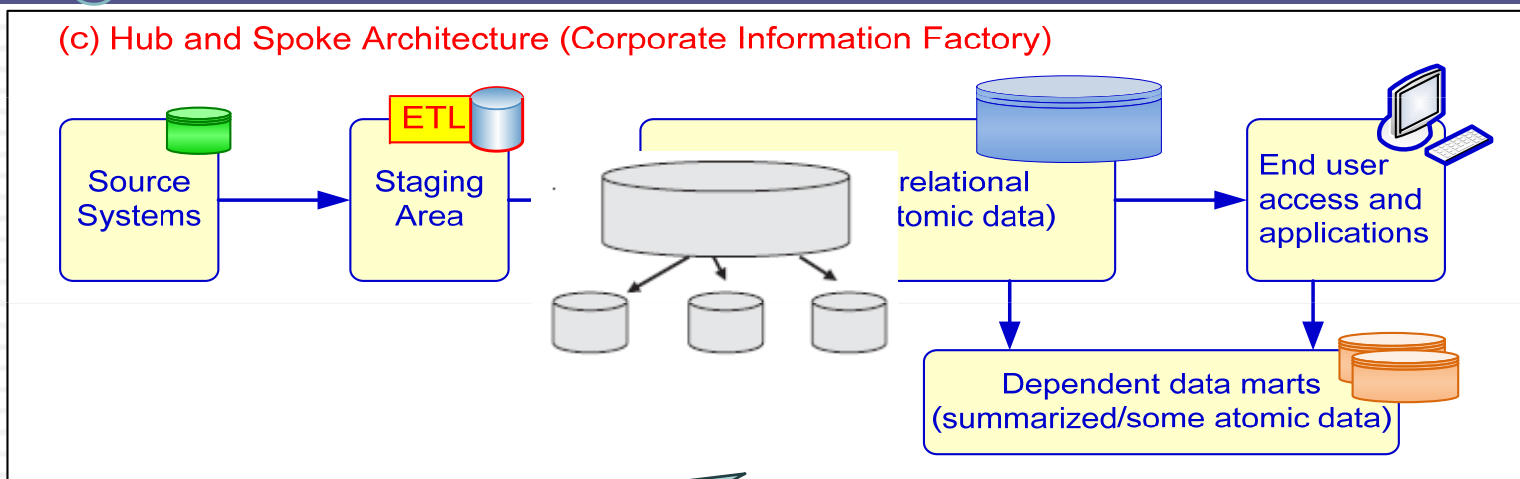
63

nemati@ut.ac.ir



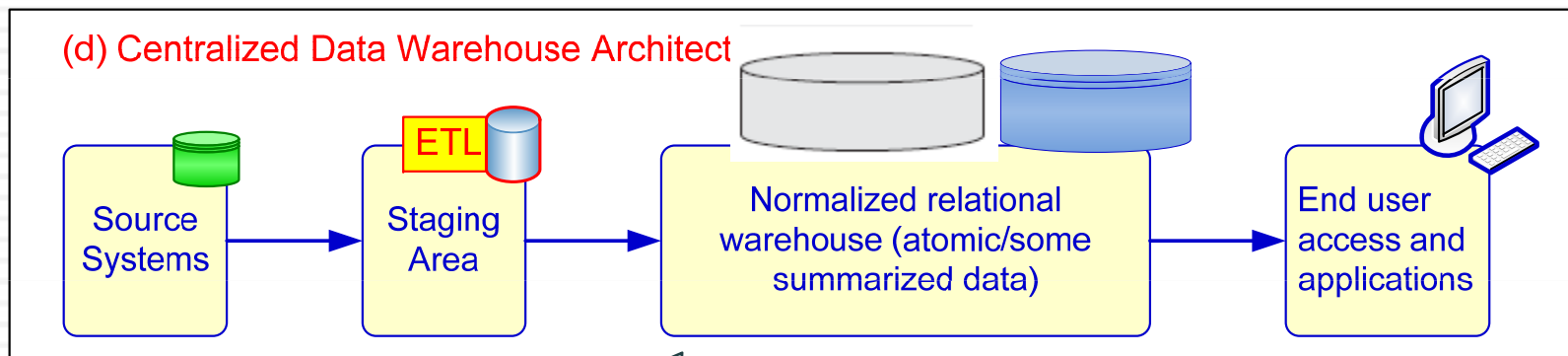
- ✓ در این معماری ابتدا موضوعات موجود در هر کسب کاری تجزیه و تحلیل می شود.
- ✓ اولین Data mart تحت عنوان Supermart بر اساس ابعاد تجاری و معیارهایی ساخته می شود. ابعاد تجاری در آینده بین Data Mart ها مشترک می شوند.
- ✓ در نتیجه ابعاد بین Data Mart ها، Supermart های یکپارچه ای است و یک دید سازمانی از داده فراهم می آورد.
- ✓ این معماری بر اساس یک دیدگاه پایین به بالا شکل گرفته است.

معماری چرخ-پره ای (Hub-and-Spoke) (Hub-and-Spoke)



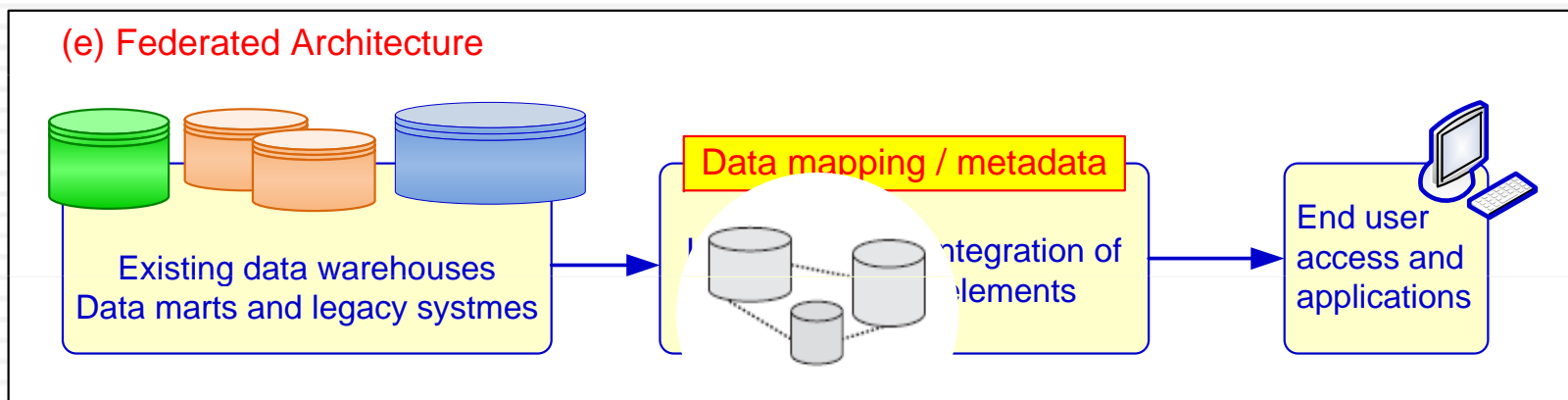
- یک پایگاه داده مرکزی سازمانی وجود دارد.
- داده ها تا سطح سوم نرمال شده اند.
- دارای Data Mart های مجزا می باشد که مزیت مهمی محسوب می شود.
- پایگاه، مرکزی مانند یک گذرگاه (هاب) Mart ها را تغذیه می کند.
- Data Mart ها برای مقاصد گوناگونی توسعه یافته اند: نیازهای تحلیلی دپارتمان های مختلف، پرس و جوهای اختصاصی، داده کاوی و ...
- هر Data Mart بر اساس الزامات منحصر به فرد ممکن است شامل ساختار داده های نرمال، غیر نرمال، خلاصه شده و یا چندبعدی باشد.
- این معماری بر اساس یک دیدگاه بالا به پایین شکل گرفته است.

معماری متمرکز (Centralized)



- اطلاعات در سطح سازمانی به صورت یکپارچه مورد نیاز است.
- داده ها تا سطح سوم نرمال شده اند.
- داده های اتمیک و خلاصه شده وجود دارند.
- پرس و جوها و برنامه های کاربردی به داده های ذخیره شده در انبار داده ها به صورت مستقیم دسترسی دارند.
- هیچ Data Mart مجزایی وجود ندارد.

معماری فدرال



- برخی شرکت ها به دنبال نوعی از پایگاه داده های تحلیلی هستند که همه سیستمی قبلی از سازه های تصمیم- پشتیبان که در قالب سیستم های عملیاتی، پایگاه داده های استخراجی و Data Mart های اولیه می باشد، تبعیت کنند.
- معماری فدرال باعث می شود تا داده ها از نظر فیزیکی یا منطقی از طریق کلیدها، فراداده های سراسری، پرس و جوهای توزیعی و دیگر روش ها با یکدیگر یکپارچه شوند.
- در این معماری انبار داده سراسری وجود ندارد.

عوامل مؤثر بر نوع معماری داده در انبار داده

✓ استقلال اطلاعات بین واحدهای سازمانی.

✓ نیازمندی اطلاعاتی مدیریت ارشد.

✓ فوریت نیاز به پایگاه داده عملیاتی.

✓ طبیعت وظایف کاربر نهایی.

✓ محدودیت در منابع.

✓ دید استراتژیک به انبار داده قبل از پیادسازی.

✓ سازگاری با سیستم های موجود.

✓ توانایی ادراکی کارکنان فناوری اطلاعات.

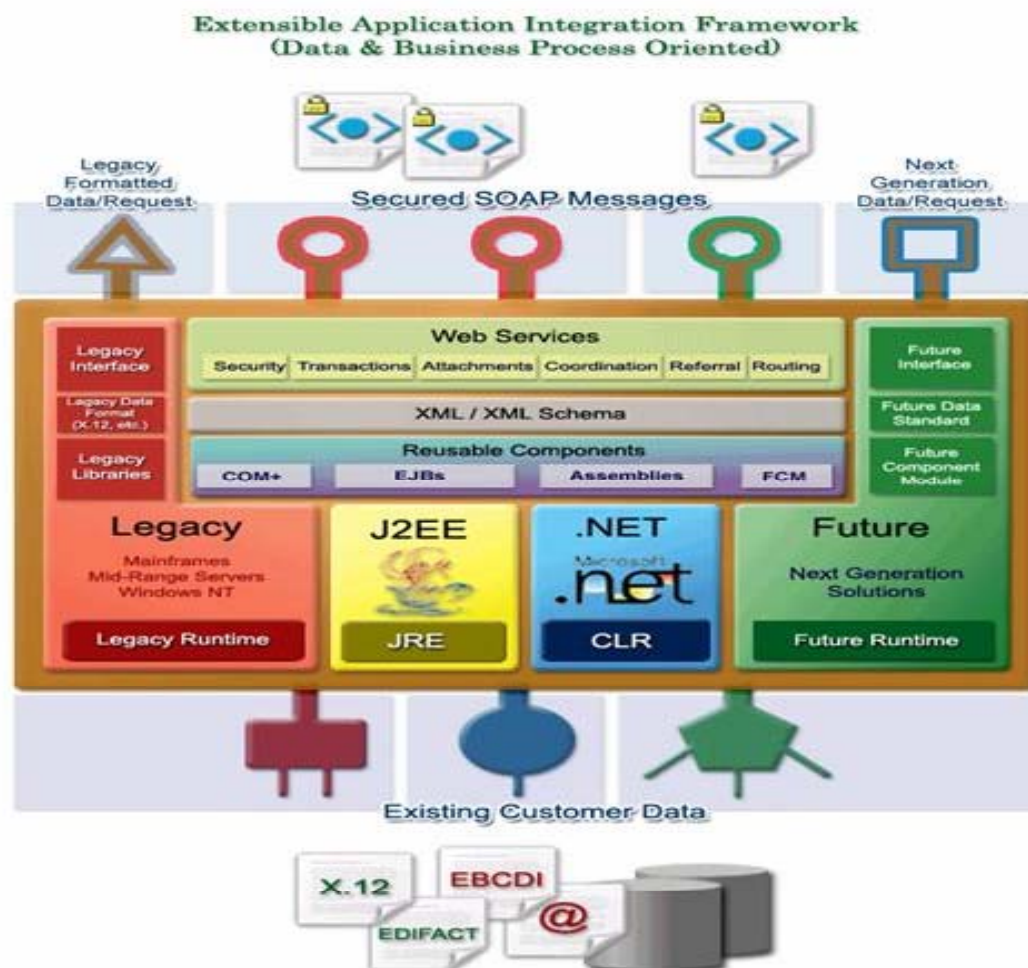
✓ موضوعات تکنیکی.

✓ عوامل سیاسی و اجتماعی.

تکنولوژی های یکپارچه سازی داده ها

- مقصود اصلی هر انباره داده این است که داده ها را در چند سیستم یکپارچه کند.
- تکنولوژی های یکپارچه سازی گوناگونی برای یکپارچه سازی داده ها و فراداده ها وجود دارد:
 - یکپارچه سازی کاربردی مؤسسه (EAI) - Enterprise Application Integration
 - معماری سرویس گرا (SOA) - Service Oriented Architecture
 - یکپارچه سازی اطلاعات مؤسسه (EII) - Enterprise Information Integration
 - استخراج، تبدیل و بارگذاری (ETL) - Extraction, Transformation and Load
- در هر حال یکپارچه سازی داده ها شامل سه فرآیند اصلی است:
 - دسترسی به داده ها
 - جمع آوری داده ها
 - تغییر داده ها

یکپارچه سازی کاربردی مؤسسه (EAI)



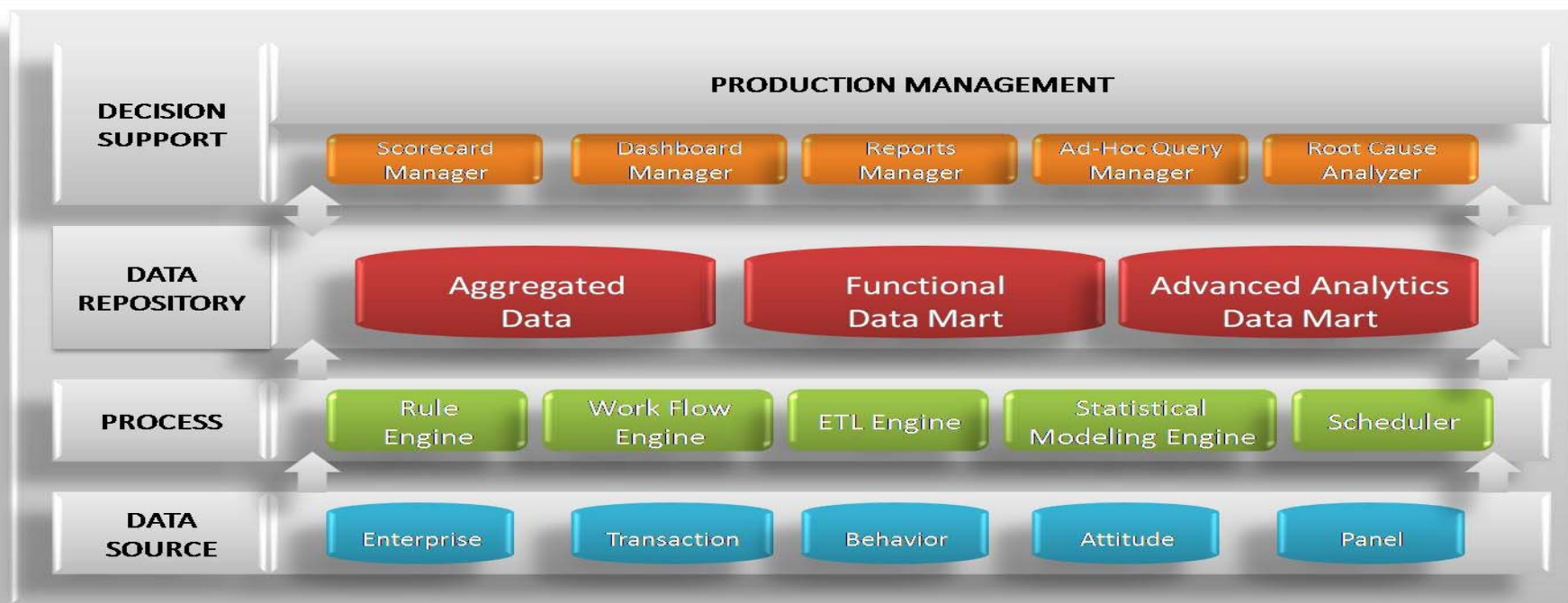
□ EAI مجموعه تکنیک ها و تکنولوژیهای است که محمل لازم برای انتقال داده ها از سیستم های منبع را به انبار داده ها فراهم می کند.

یکپارچه سازی اطلاعات سازمانی (EII)

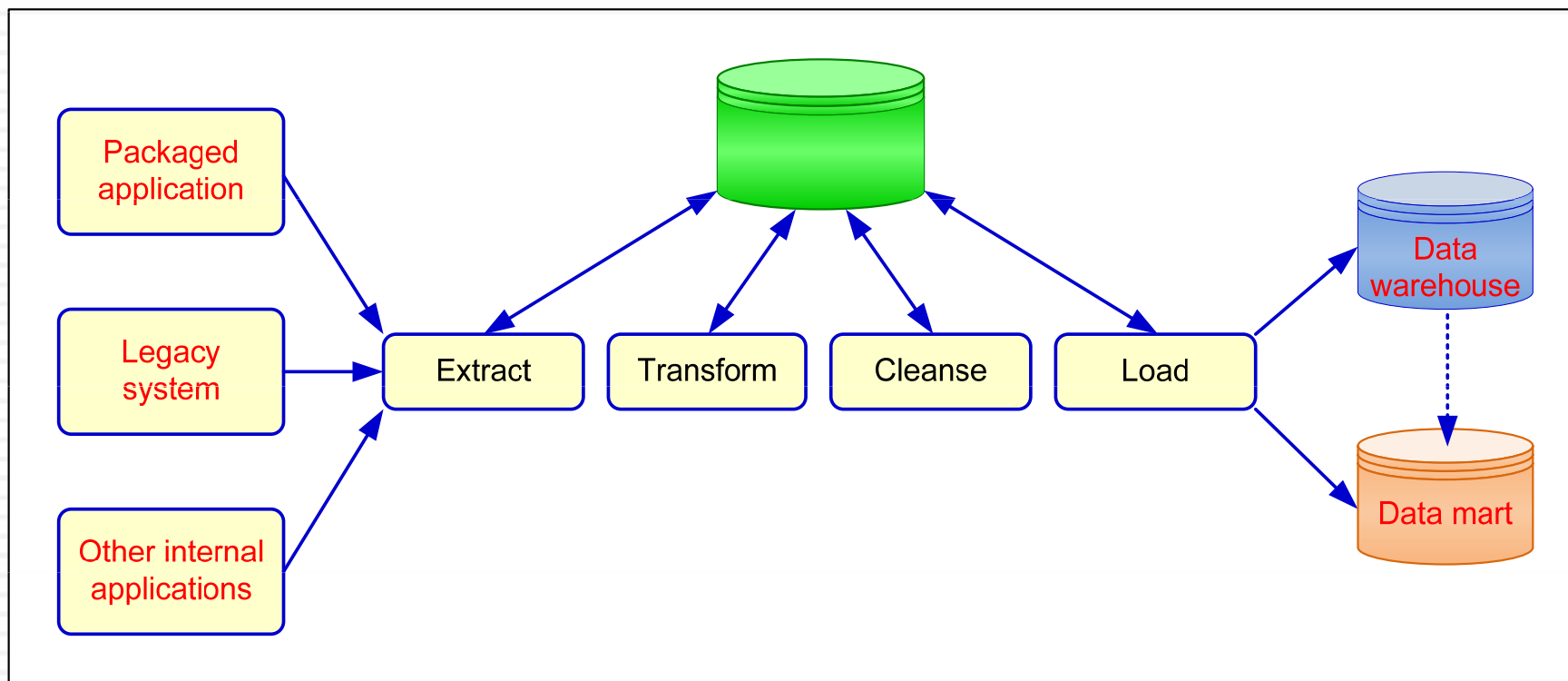
70

nemati@ut.ac.ir

□ ابزارهایی در حال تکامل است که یکپارچه سازی بلادرنگ داده ها از منابع گوناگون مانند پایگاه داده های رابطه ای، سرویس های وب و پایگاه داده های چندبعدی را فراهم می کند.



استخراج، انتقال و بارگذاری داده (ETL)



ابزارهای ETL

- ابزارهای تبدیل داده گران هستند.
- ابزارهای تبدیل داده ممکن است به یادگیری طولانی مدت نیاز داشته باشند.

عوامل مؤثر بر خرید ابزار ETL

- توانایی خواندن و نوشتن در تعداد نامحدودی منابع داده با معماری های گوناگون.
- دریافت و تحویل خودکار داده ها.
- توانایی انطباق با استانداردها.
- واسط کاربری آسان برای توسعه دهنده و کاربر عملیاتی.

معیارهای مهم در انتخاب یک ابزار ETL

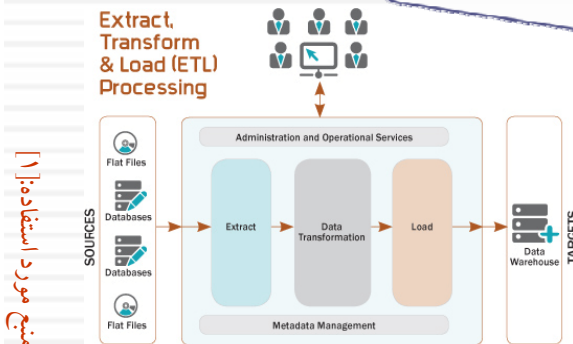


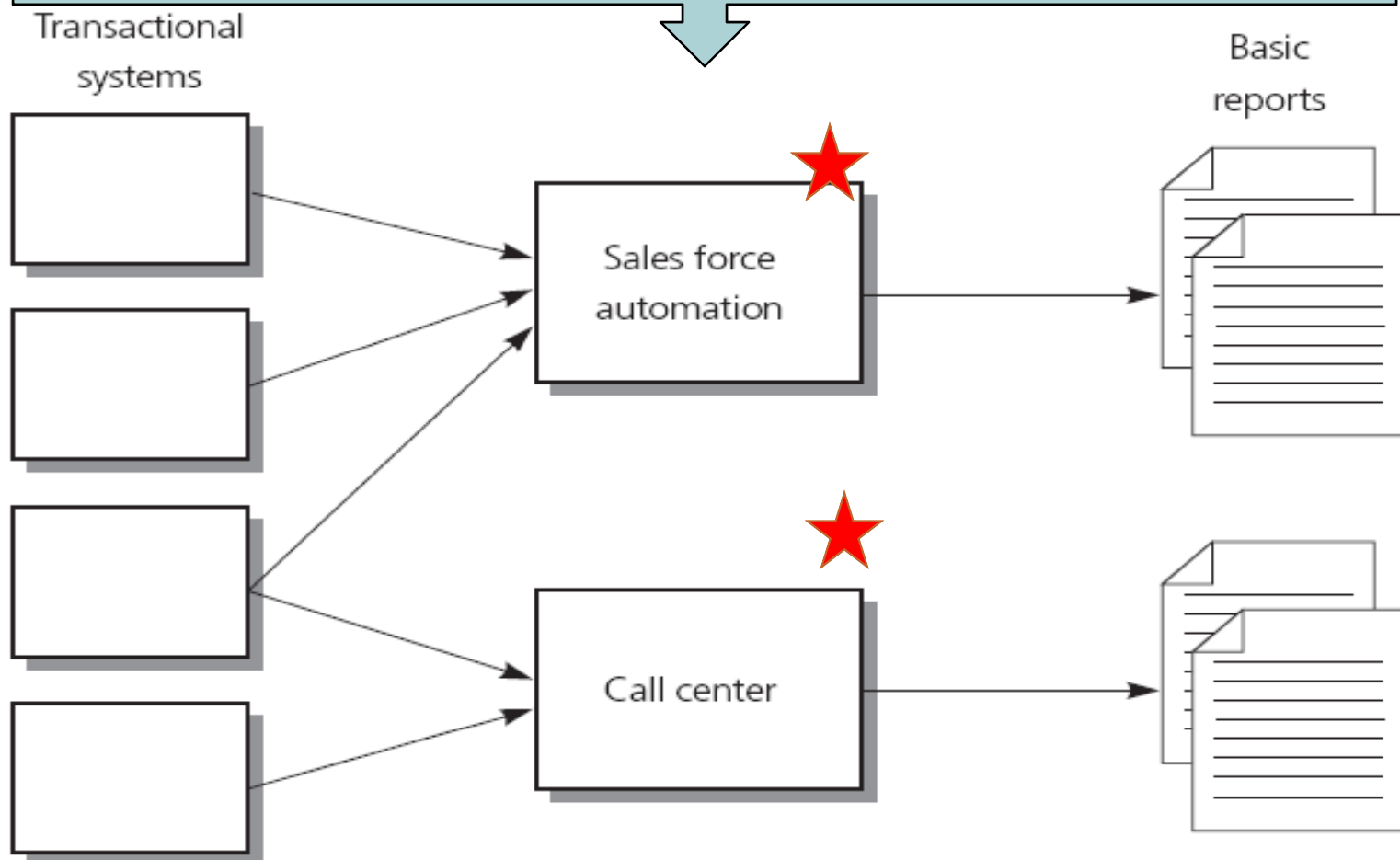
FIGURE 1: BASIC ETL FUNCTIONALITY

مثالی برای درک روند توسعه انبار داده در حوزه CRM: فاز ۱: استقرار اولیه سیستم اصلی

73

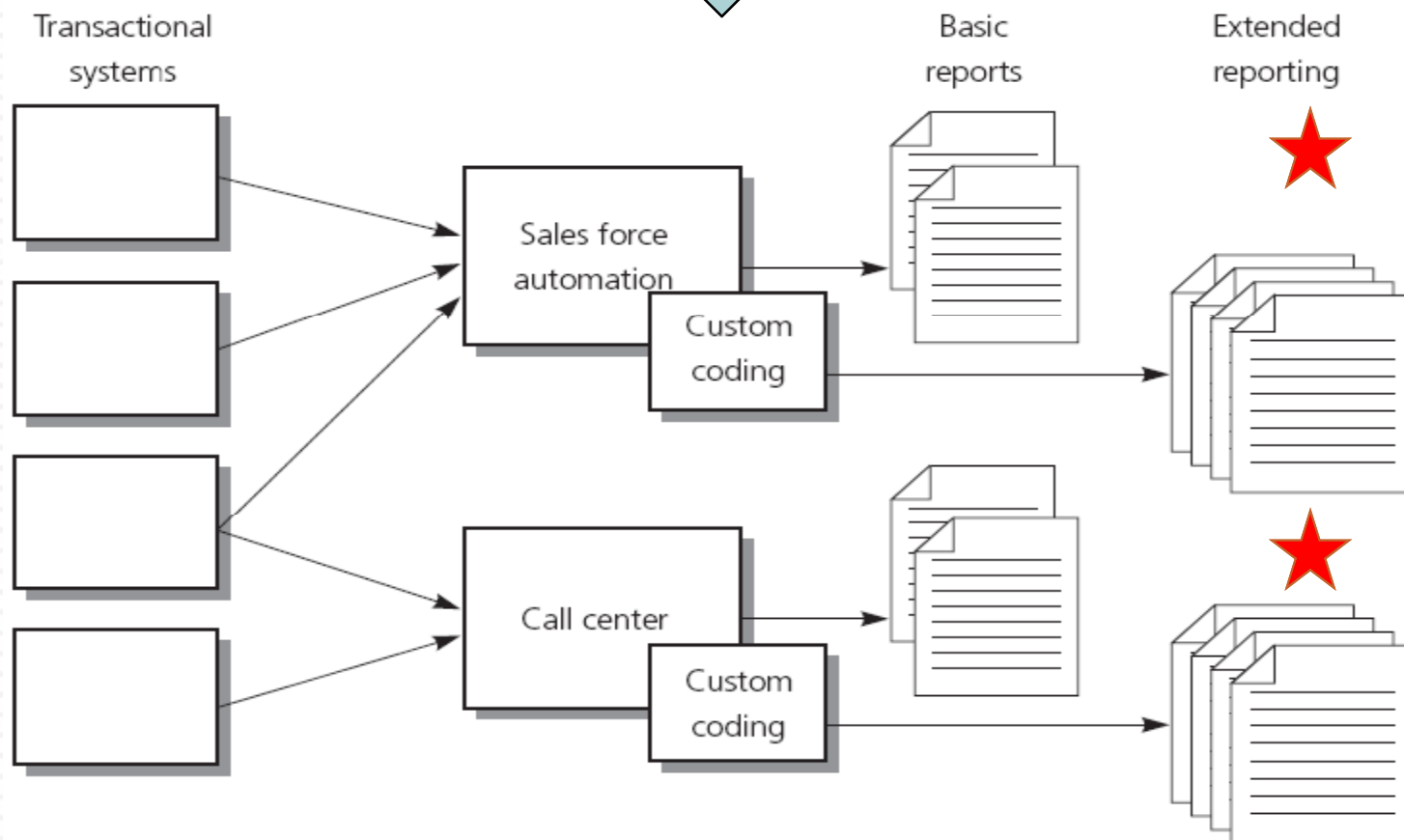
nemati@ut.ac.ir

۱. توسعه اولیه سیستم محوری (مرکزی) CRM



مثالی برای درک روند توسعه انبار داده در حوزه CRM: فاز ۲: گسترش گزارشات سیستم مرکزی

۲. گسترش گزارشات سیستم مرکزی CRM، با برنامه نویسی سفارشی

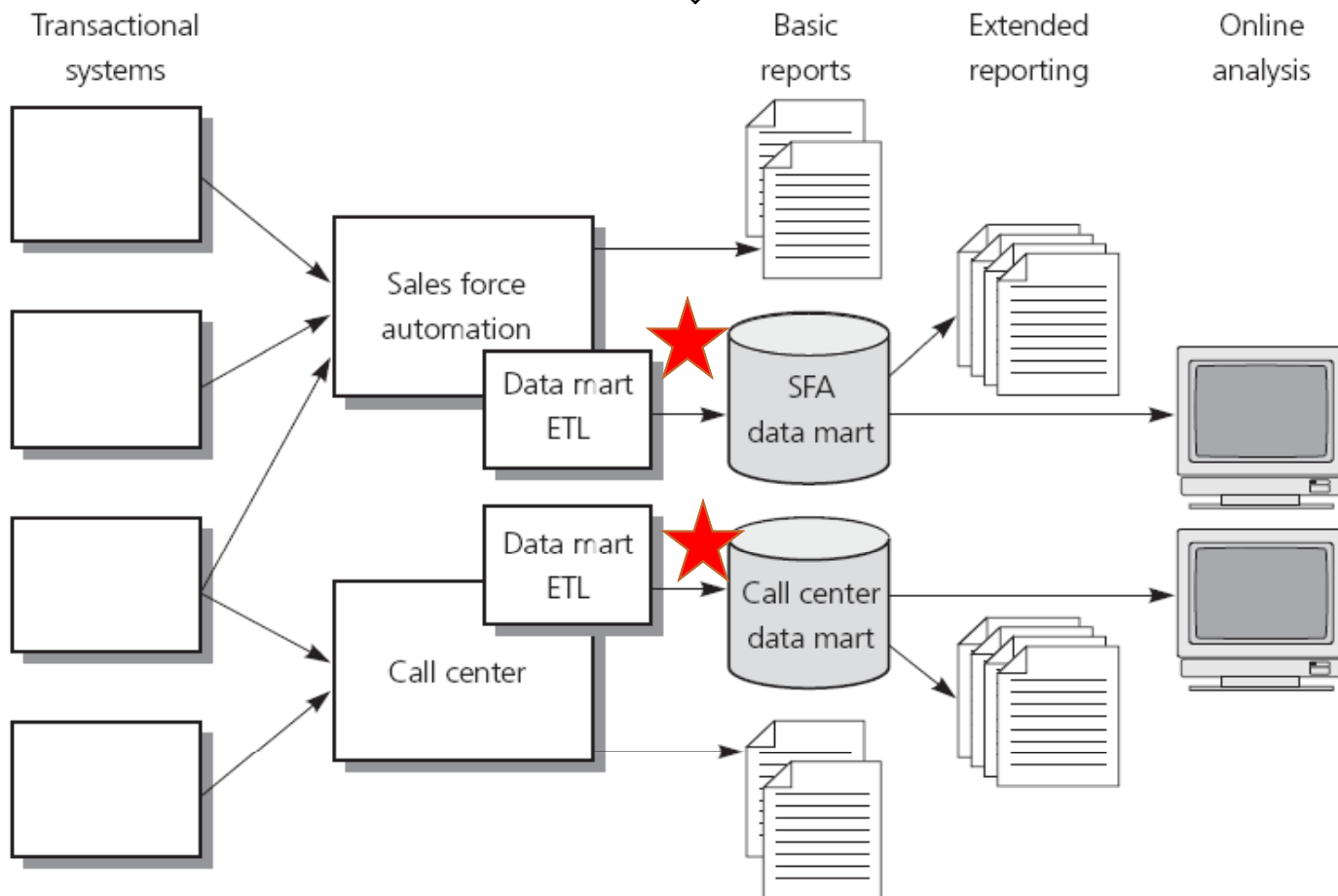


مثالی برای درک روند توسعه انباره داده در حوزه CRM: فاز ۳: ساخت Data Mart های عملیاتی جهت پشتیبانی از CRM

75

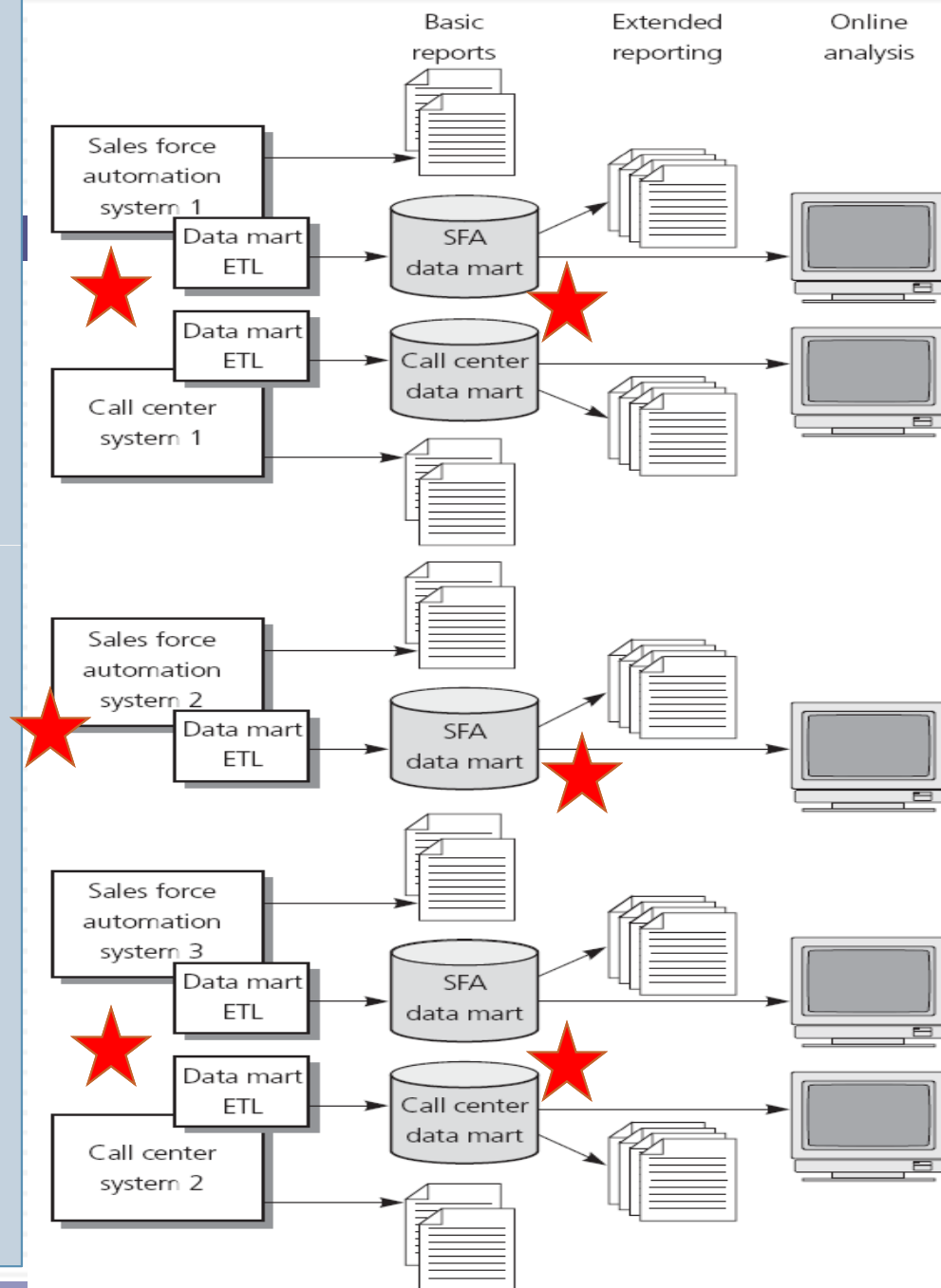
nemati@ut.ac.ir

۳. ساخت Data Mart های عملیاتی جهت پشتیبانی از CRM

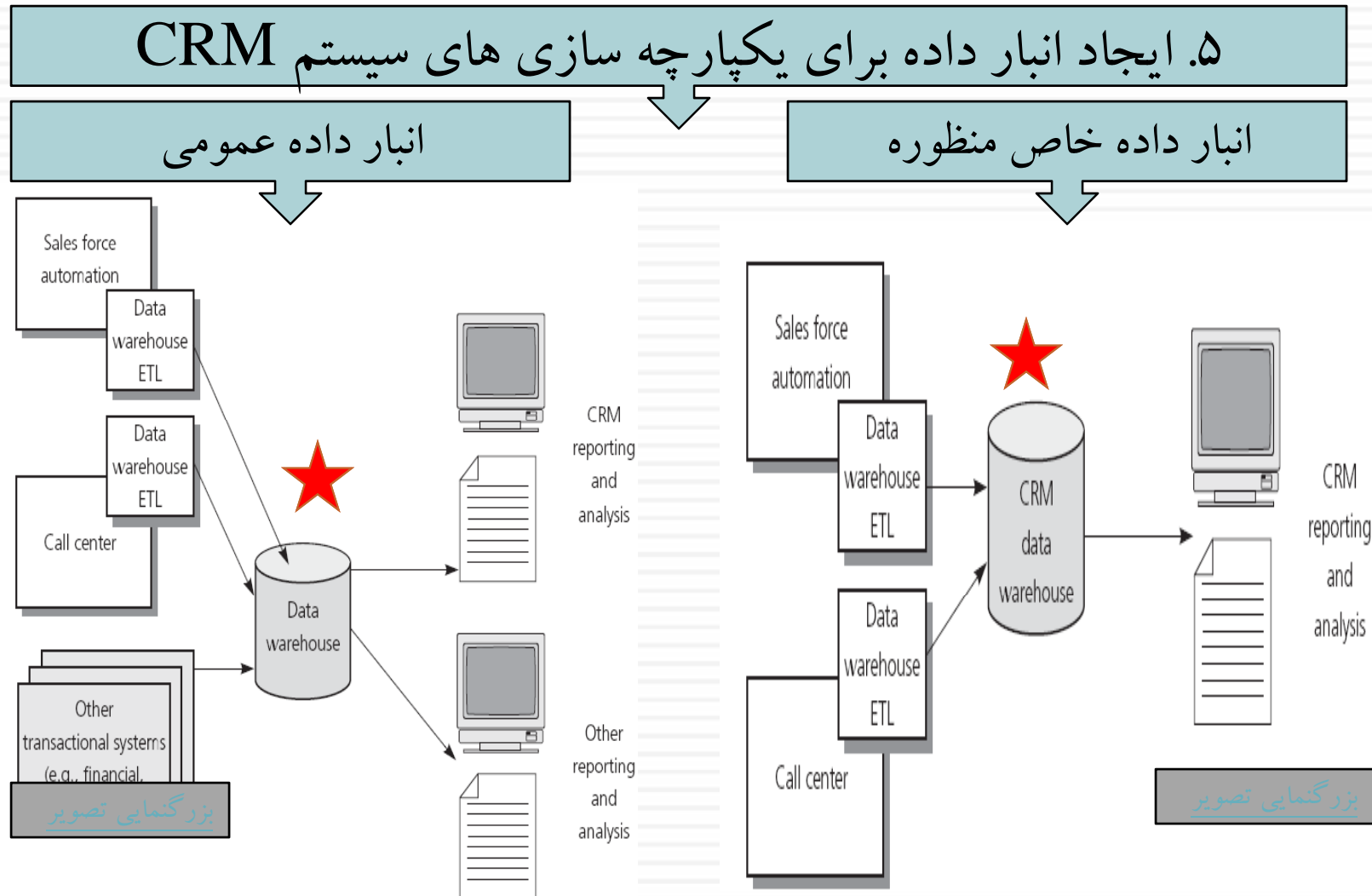


سه سیستم
اتوماسیون نیروی
فروش و دو سیستم
مرکز تماس که
انبارک های داده
(DM) آنها در با
یکدیگر یکپارچه
نیست.

مثالی برای درک روند توسعه انباره داده در حوزه CRM:
فاز ۴: چند سیستم مرکزی CRM و Data Mart های غیر یکپارچه



مثالی برای درک روند توسعه انبار داده در حوزه CRM: فاز ۵: دو رویکرد برای یکپارچه سازی CRM با تمرکز بر انبار داده DW

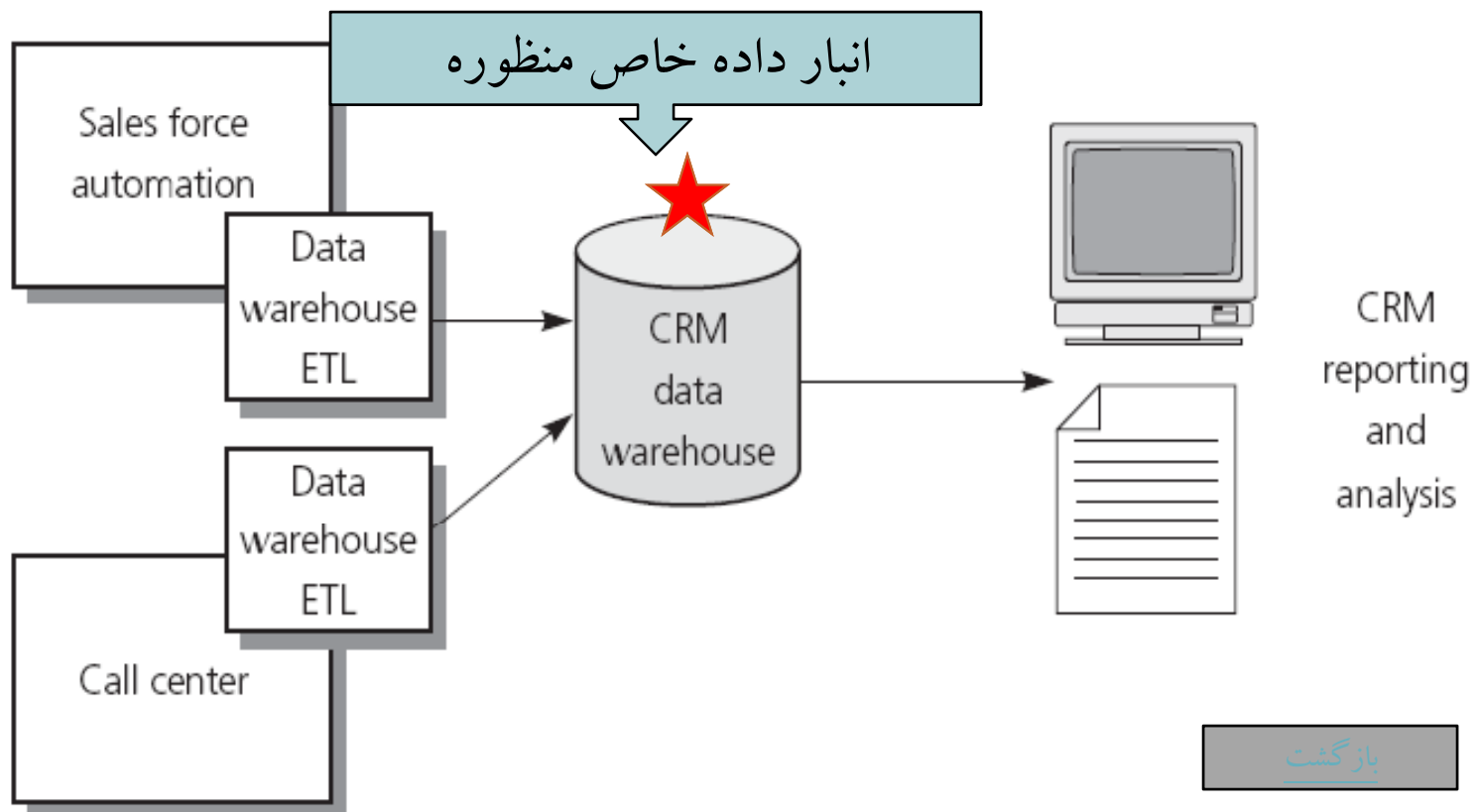


مثالی برای درک روند توسعه انبار داده در حوزه CRM: فاز ۵: رویکرد اول یکپارچه سازی: انبار داده سازی خاص

78

nemati@ut.ac.ir

۵. ایجاد انبار داده برای سیستم CRM



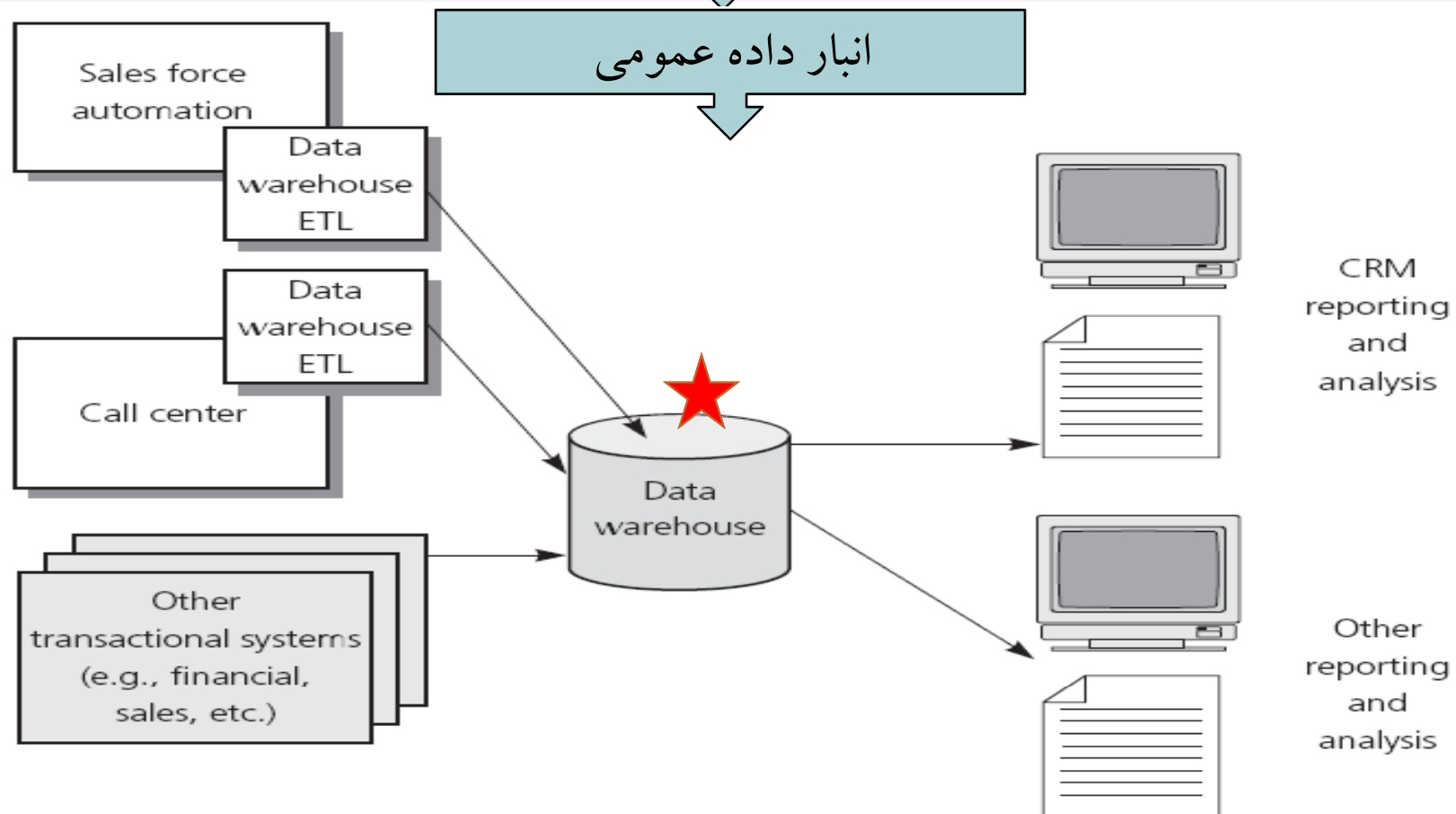
بازگشت

مثالی برای درک روند توسعه انبار داده در حوزه CRM: فاز ۵: رویکرد دوم یکپارچه سازی: انبار داده عمومی

79

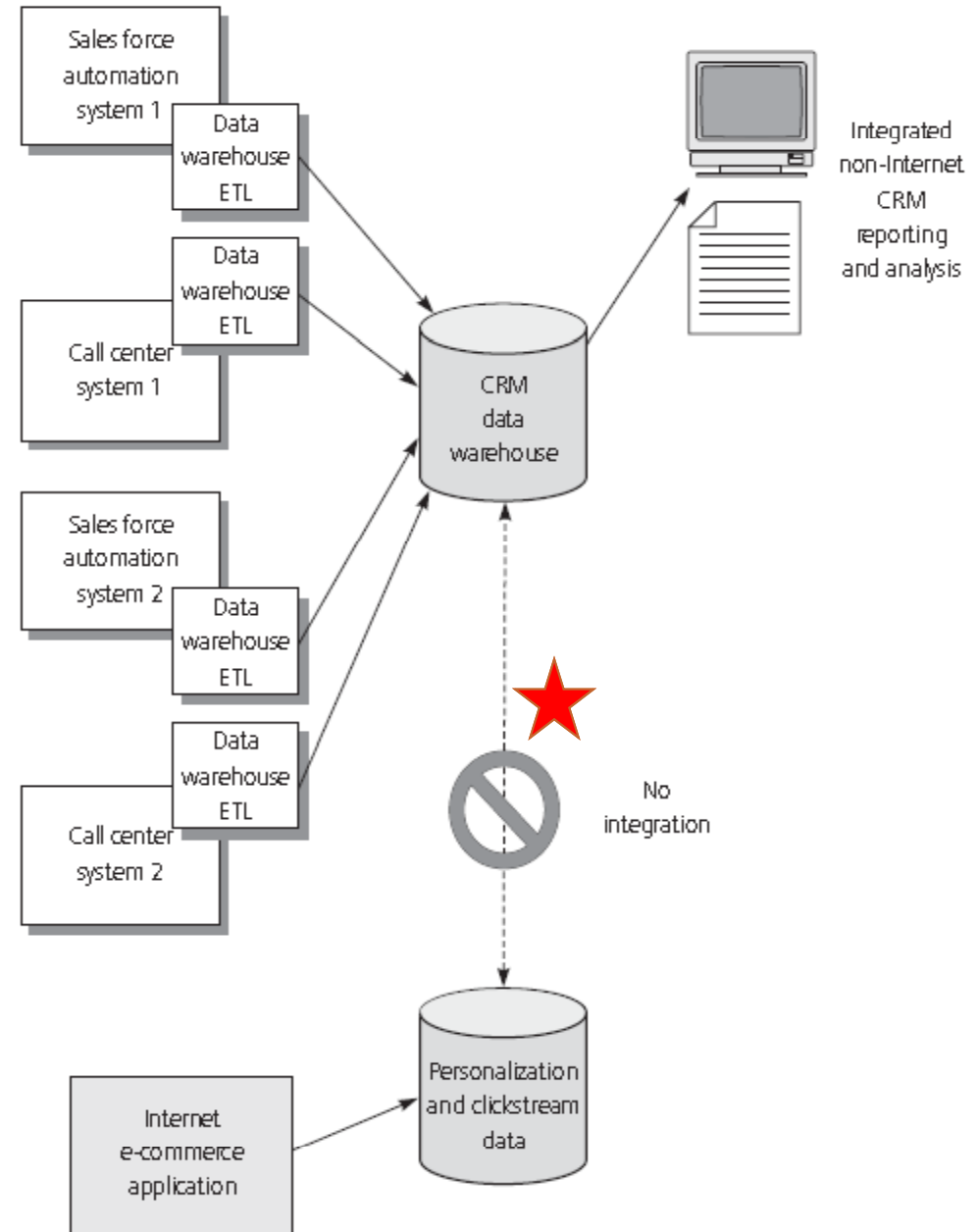
nemati@ut.ac.ir

۵. ایجاد انبار داده برای سیستم CRM

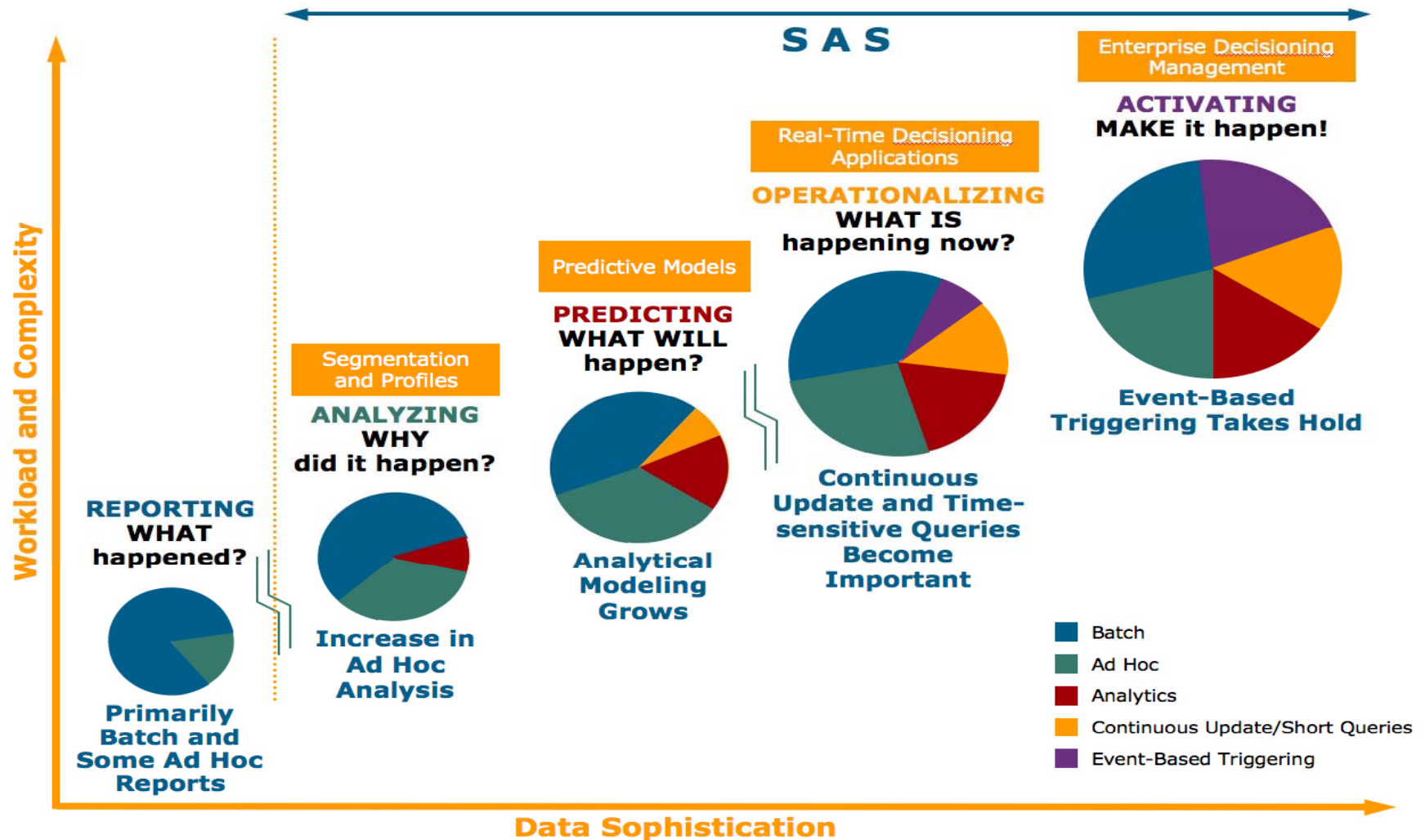


انباره داده ویژه CRM با داده های جریان کلیک و شخصی سازی که از برنامه های کاربردی تجارت الکترونیک تغذیه می شود، یکپارچه نیست و لذا سیستم تحلیل و گزارش دهی CRM با اینترنت یکپارچه نشده است.

مثالی برای درک روند توسعه انباره داده در حوزه CRM:
فاز ۶: انباره داده خاص CRM بدون یکپارچگی با eCommerce



سیر تکاملی تصمیمات سازمانی و انبار داده ها



انبار داده ها به صورت بلادرنگ یا فعال



به روز رسانی بلادرنگ داده، جهت تحلیل ها و تصمیم گیری های فوری تجاری امری ضروری است که به پیاده سازی بلادرنگ و فعال انبار داده ها نیاز دارد که گاهی بسیار گران و حتی غیرممکن است.

انبار داده های بلادرنگ یا فعال

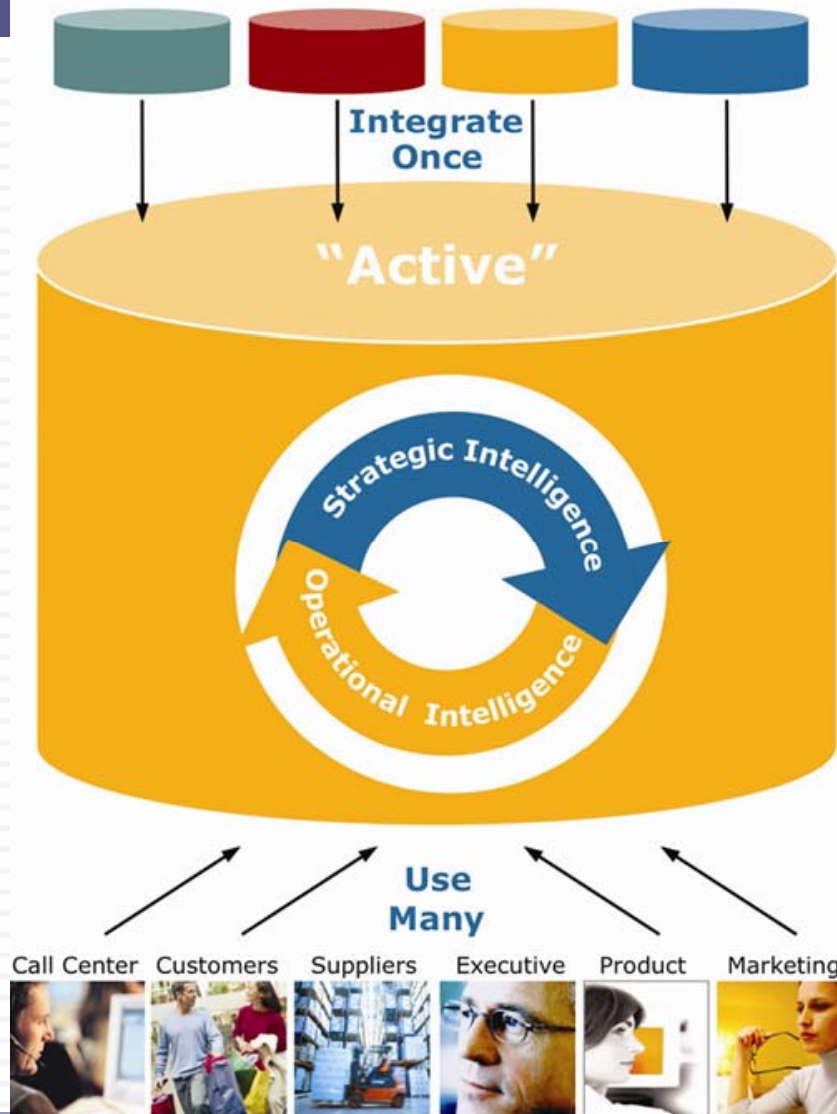
83

nemati@ut.ac.ir

مدیریت کاری فعال:
مدیریت فعال منابع سیستمی جهت کارایی بهینه و بهره برداری مناسب از منابع.

یکپارچگی سازمانی فعال:
یکپارچگی معماری سازمانی جهت ارائه خدمات تصمیم گیری هوشمند.

دسترسی پذیری فعال:
تداوم تجارت (هر ۷ روز هفته و ۲۴ ساعته).



دسترسی فعال:
تصمیمات اساسی عملیاتی یا خدمات حمایتی با دسترسی تقریباً بلادرنگ زیر ۵ ثانیه.

بارگذاری فعال:
بارگذاری داده در سیستم های دسته ای در یک دقیقه یا حتی یک ثانیه.

رویداد فعال:
نظارت پیش فعال فعالیت های تجاری بر اساس قوانین و موضوعات، و پشتیبانی سیستماتیک از فرآیندهای عملیاتی.

تفاوت انبار داده سنتی و فعال

انبار داده فعال

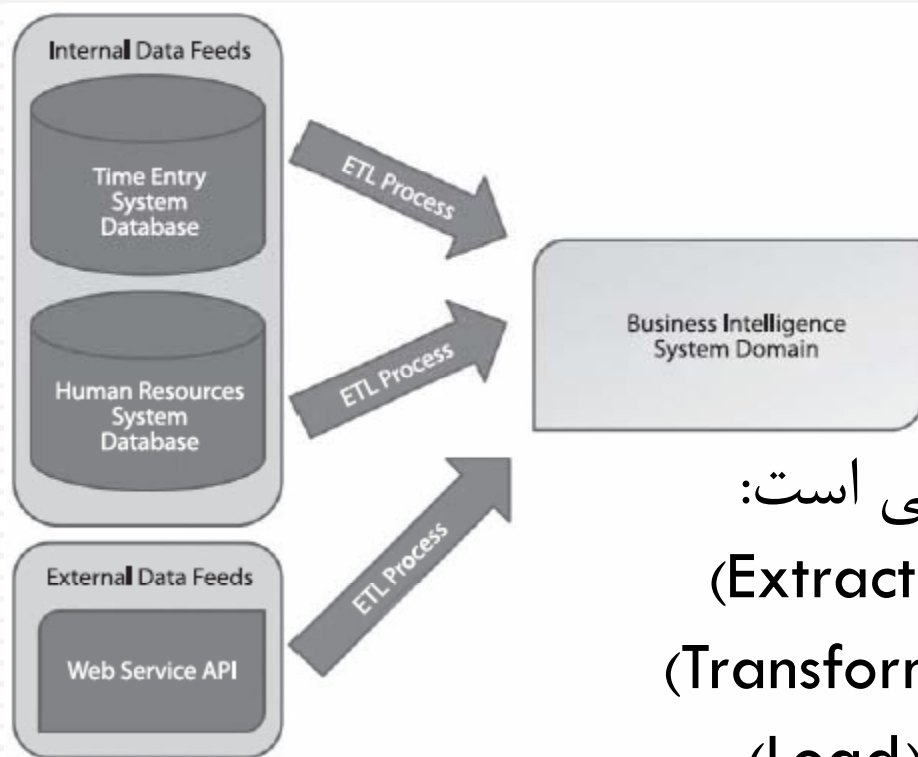
- تصمیمات استراتژیک و تاکتیکی گرفته می شوند.
- نتایج توسط عملگرهایی قابل اندازه گیری هستند.
- تعداد زیادی از کاربران (۱۰۰۰ یا بیشتر) به صورت همزمان به سیستم دسترسی دارند.
- گزارشگیری با انعطاف بالا جهت کشف فرضیات و ارتباطات صورت می گیرد.
- کارمندان عملیاتی، مراکز تماس و کاربران خارجی به سیستم دسترسی دارند.

انبار داده سنتی

- تنها تصمیمات استراتژیک گرفته می شود.
- گاهی نتایج به سختی قابل اندازه گیری هستند.
- تعداد متوسطی از کاربران به صورت همزمان امکان دسترسی دارند.
- گزارشات بسیار محدودی جهت تأیید و بررسی فرآیندها و الگوها استفاده می شود.
- کاربران قوی، کارگران دانشی و کاربران داخلی حق دسترسی دارند.

خوراک داده های سیستم هوش تجاری

□ سیستم های BI از طریق دو منبع با فرآیند ETL تغذیه می شوند:



□ تغذیه داده های داخلی

■ پایگاه داده های سیستم های داخلی

□ تغذیه داده های خارجی

■ رابط برنامه کاربردی وب سرویس

□ فرآیند ETL شامل سه مرحله اصلی است:

□ استخراج داده ها از منابع داده ها (Extract)

□ تبدیل داده ها به فرمت مناسب (Transform)

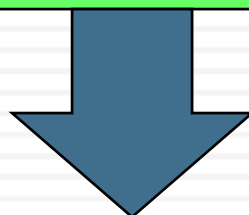
□ بارگذاری داده ها در انبار داده ها (Load)

به نظر شما فواصل زمانی تغذیه اطلاعاتی سیستم هوش تجاری چگونه است؟

هوش تجاری تقاضا محور و بلادرنگ

تقاضا برای دسترسی به موقع و آنی به اطلاعات پراکنده (در قالب نیاز به کاهش شکاف بین داده های عملیاتی و اهداف استراتژیک) رو به رشد است و فشار برای رسیدن به آن شدیدتر می شود.

نیاز تنها تهیه گزارش نیست، بلکه کاربران می خواهند عملکرد کسب و کار را در هر لحظه دنبال، آنرا تحلیل نمایند و سریعاً دلایل بروز وقایع را درک نمایند.

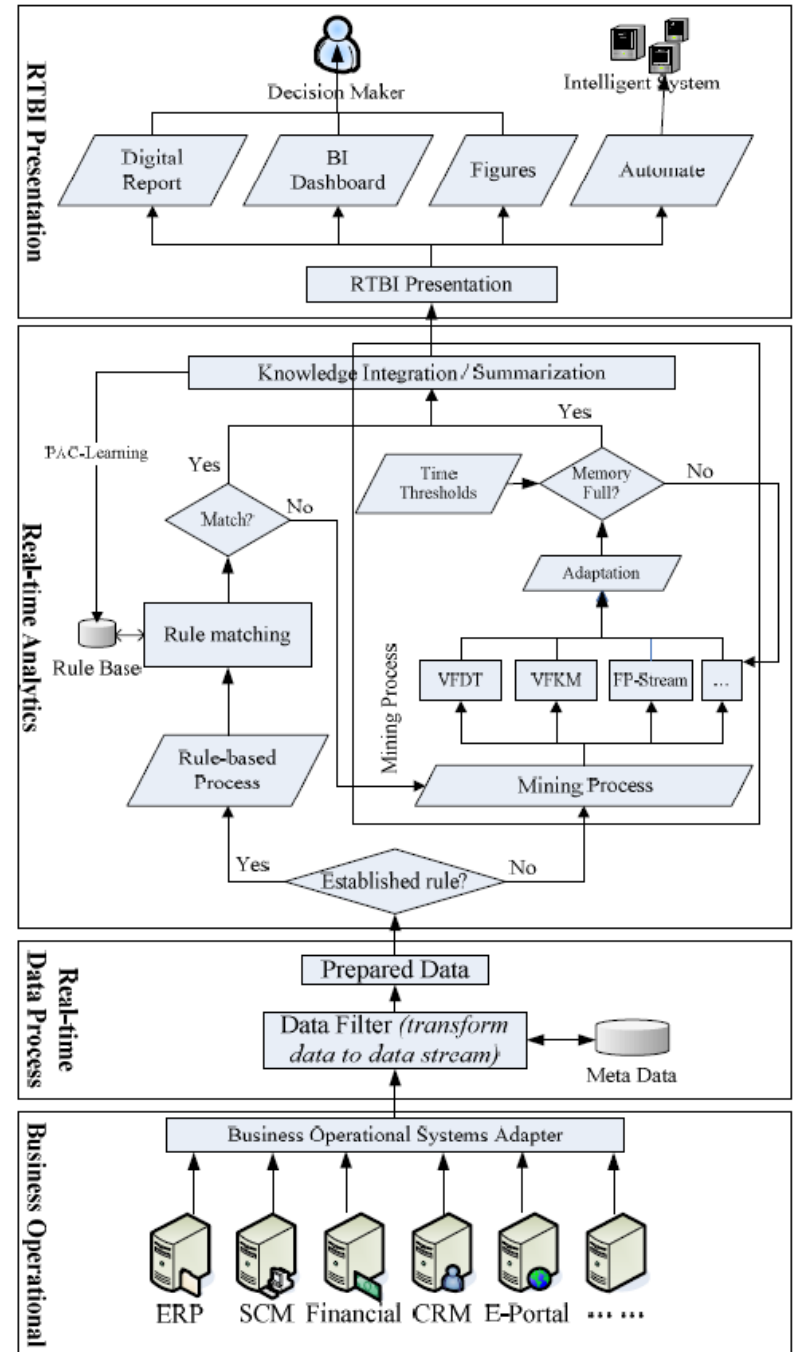
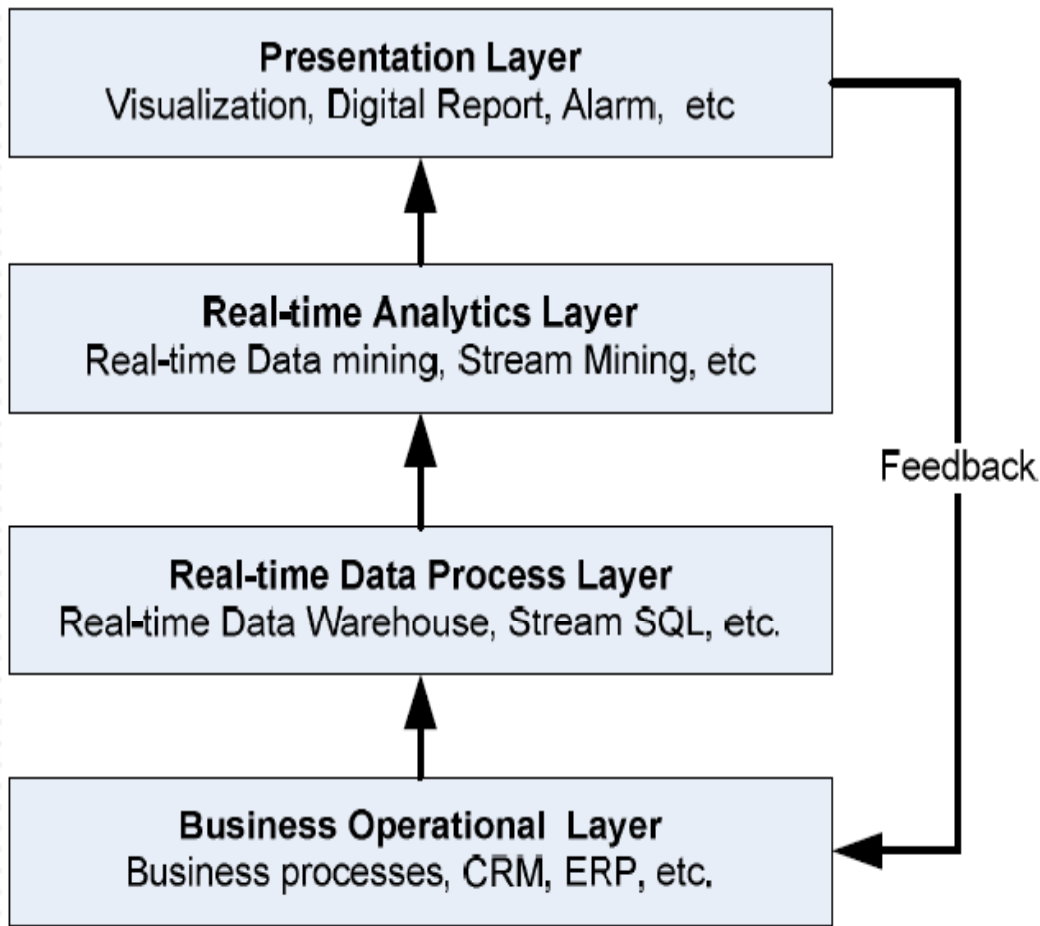


کاربران باید به موقع، تغییرات در داده ها را درک کنند و یا به گزارش ها و علائم هشدار دهنده با توجه به روندهای در حال ظهور در وب، پست الکترونیکی و یا پیام فوری دسترسی داشته باشند و از طرفی وقایع را به سرعت درک کنند. که این مهم از طریق **هوش تجاری بلادرنگ** قابل دستیابی است.

Real_Time BI System Architecture

87

nemati@ut.ac.ir



یازده وظیفه اصلی برای موفقیت پیاده سازی پروژه انبار داده

88

nemati@ut.ac.ir

ایجاد توافقات در سطح سرویس ها و الزامات تازه سازی داده.

شناسایی منابع داده و سیاست های دولتی آنها.

برنامه ریزی های کیفی داده.

انتخاب ابزار مناسب برای ETL.

طراحی مدل داده ای.

انتخاب نرم افزار و پلتفرم پایگاه داده رابطه ای.

انتقال داده ها.

تبدیل داده ها.

فرآیندهای اصلاح.

برنامه ریزی برای آرشیو نمودن و پاکسازی داده ها.

پشتیبانی از کاربر نهایی

دلایل شکست پروژه های توسعه انبار داده ها

89

nemati@ut.ac.ir

عدم حمایت مدیران اجرایی.

مشخص نبودن اهداف تجاری.

نادیده گرفتن موضوعات فرهنگی همچون مدیریت تغییرات.

انتظارات غیر واقعی.

معماری نامناسب.

کیفیت پایین داده و اطلاعات مفقود.

بارگذاری هر نوع داده ی در دسترس بدون توجه به ضرورت وجود آن و ایجاد هرج و مرج.

رهنمودهایی در باره پیاده سازی پروژه انباره داده ها

پروژه باید با استراتژی شرکت و اهداف کسب و کار متناسب باشد.

باید انتظارات کاربر درباره پروژه اجرایی، مدیریت شود.

انبار داده ها باید به صورت مرحله ای و افزایشی ساخته شود.

پایگاه داده باید به صورت قابل انعطاف، سازگار و مقیاس پذیر ساخته شود.

پروژه باید توسط متخصصان IT و تجاری، توأمأ مدیریت شود.

تنها داده های پاک سازی شده که از نظر سازمانی دارای کیفیت مناسبی هستند، بارگذاری شود.

الزامات آموزشی حتماً مورد توجه قرار گیرند.

سیاست ها مدنظر قرار گیرند.

تمرین ۶-۱

- گزارش مختصری در باره یکپارچه سازی برنامه های کاربردی مؤسسه (EAI) و یکپارچه سازی اطلاعات سازمانی (EII) تهیه کنید؟
- کدام شرکت ها در این زمینه خدمات مناسبی ارائه می دهند؟ چه نوع سرویس هایی؟
- جایگاه این تکنیک ها و تکنولوژی ها در کدام بخش از معماری هوش تجاری قرار می گیرد؟
- یکپارچه سازی داده ها و ETL را با یکدیگر مقایسه کنید. ارتباط این دو چگونه است؟

تمرین ۲-۶

- دیدگاه های اینمون (Innmon) و کیمبال (Kinmball) را در در باره انباره سازی داده ها بررسی نموده و آنها را با هم مقایسه کنید.
- تذکر:
- برای پاسخ به این سؤال از کتاب ها و منابع دست اولی که توسط این دو صاحب نظر در زمینه انباره سازی داده ها منتشر شده استفاده کنید.

تمرین ۳-۶

□ بررسی کنید آینده انباره سازی داده ها چه سمت و سویی دارد؟

□ تذکر:

□ مقالاتی که در این زمینه نوشته شده است را پیدا کنید و با مطالعه آنها دیدگاه خود را در باره آینده انباره سازی داده ها بیان کنید.

تمرین ۴-۶

- با بررسی و مطالعه دو نمونه پروژه انباره سازی داده ها (یکی موفق و یکی ناموفق) عوامل مؤثر بر موفقیت و شکست پروژه های انباره سازی داده ها را یافته و آنها را در قالب یک مدل مفهومی ارائه دهید.
- ابزاری تهیه کنید که بتوان با استفاده از آن موفقیت و شکست پروژه های انباره را سنجید.

- [1] E. Turban, R. Sharda, J. E. Aronson, and D. King (2007) “Business Intelligence A managing approach”, Chapter 2.
- [2] Alan R. Simon, Steven L. Shaffer (2001) “Data warehousing and business intelligence for e-commerce “, Published by Scott Norton, 2001.
- [3] PAULRAJ PONNIAH (2010) “DATA WAREHOUSING FUNDAMENTALS FOR IT PROFESSIONALS”, Second Edition Published by John Wiley & Sons, Inc.