به نام آنکه جان را فکرت آموخت

بخش اول : مقدمه

دکتر عیسی زارع پور

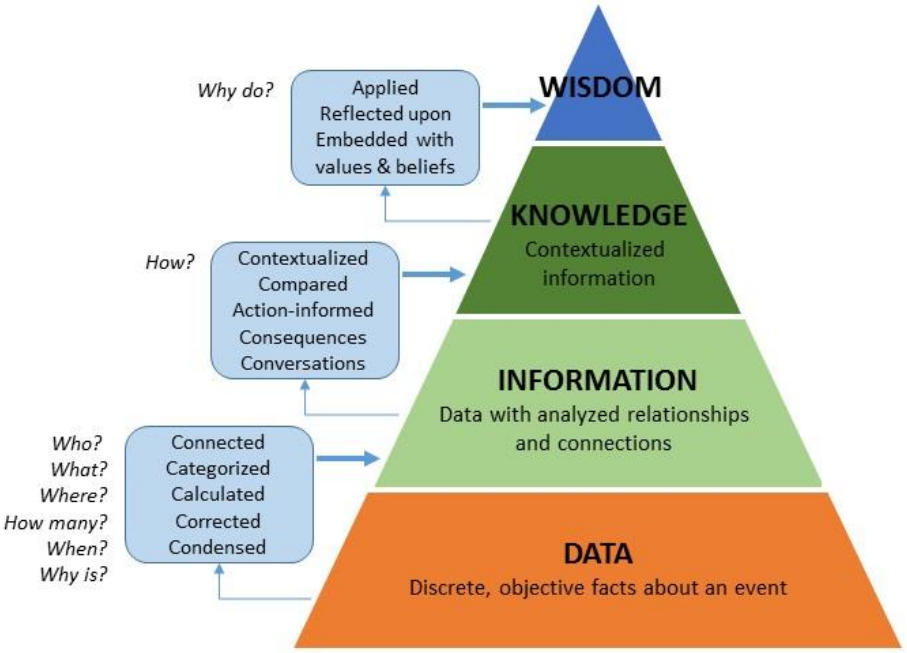
دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت

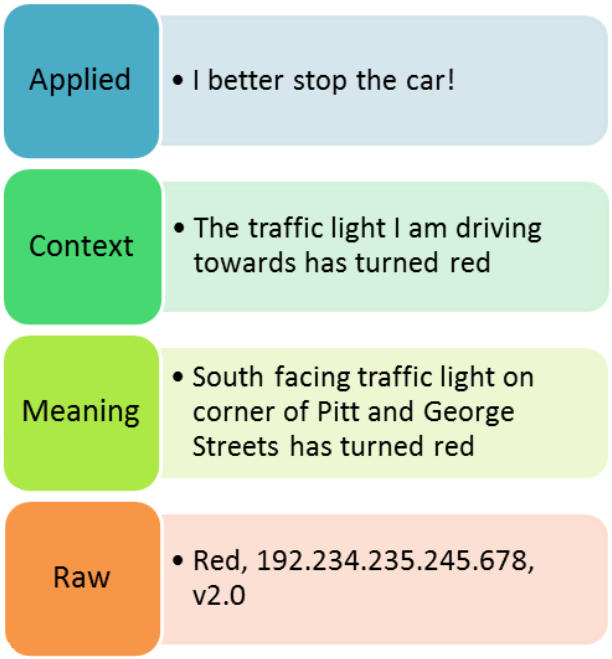
نیمسال اول ۹۹-۹۸

محتویات اسلایدها برگرفته از یادداشت‌های کلاسی استاد محمدتقی روحانی رانکوهی است. اسلایدها توسط آقای دکتر مرتضی امینی(دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف) تهیه شده است.

داده – اطلاع – دانش

داده	اطلاع	دانش	عقل/معرفت
۱۳۹۲/۰۶/۰۸ ؟	• تاریخ تولد یک نفر • سالگرد ازدواج یک نفر • سالگرد انفجار دفتر نخست وزیری • روز مبارزه با تروریزم • ...	او در سن مدرسه است ←	اولیا باید او را در مدرسه ثبت نام کنند ←





داده – اطلاع – دانش

داده (Data)

- تعریف اول ANSI: نمایش بوده‌ها، پدیده‌ها، مفاهیم یا شناخته‌ها به طرزی صوری و مناسب برای برقراری ارتباط، تفسیر یا پردازش توسط انسان یا هر امکان خودکار
- تعریف دوم ANSI: هر نمایشی اعم از کاراکتری (نویسه‌ای) یا کمیت‌های قیاسی که معنایی به آن قابل انتساب باشد (توسط انسان یا یک مکانیسم خودکار)
- By **data**,we mean known facts that can be recorded and that have implicit meaning[Elmasri, 2010].

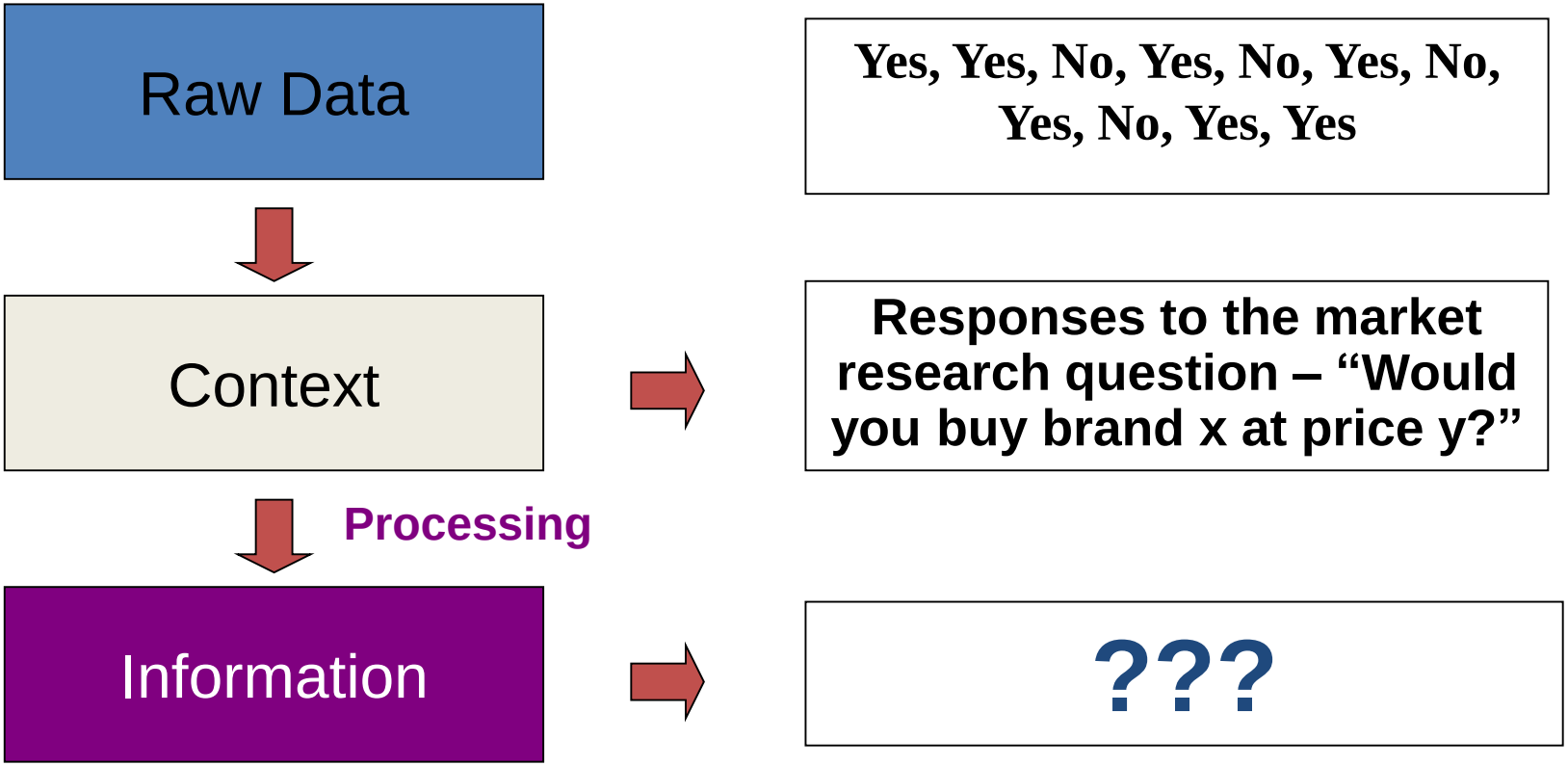
تعریف غیر فرمال

- مقادیر خام شامل حروف، اعداد یا ترکیبی از آنها
- مقادیری که به خودی خود فاقد معنای خاصی هستند
  - ۱۲۳
  - ۲۵۶
  - ادبیات
  - ۱۳۹۸
  - A2020

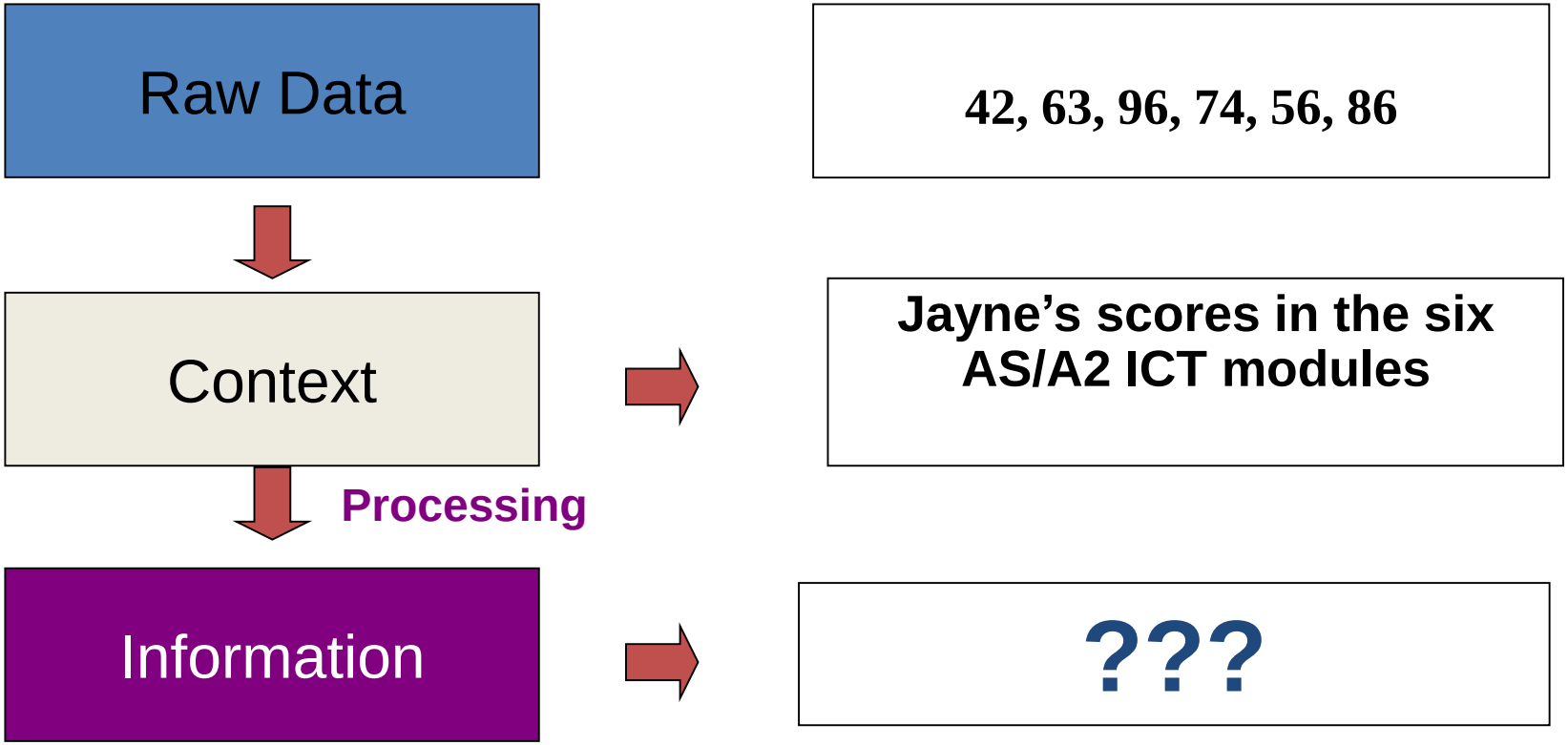
داده – اطلاع – دانش

- اطلاع (Information)
  - تعریف دقیق و جامعی از مفهوم اطلاع وجود ندارد.
  - تعریف اول [LIPS92]: اطلاع، داده پردازش شده است.
  - تعریف دوم [روحا ۷۸-الف]: معنایی که انسان به داده منتسب می‌کند، از طریق قراردادهای شناخته شده‌ای که در نمایش داده به کار می‌روند.
  - برخی داده را همان مقدار واقعا ذخیره شده و اطلاع را معنای آن می‌دانند. بنابراین اطلاع دارای خاصیت **اطلاع‌دهندگی** و ارتباط‌دهندگی است، در حالیکه داده مجرد این خاصیت را ندارد.

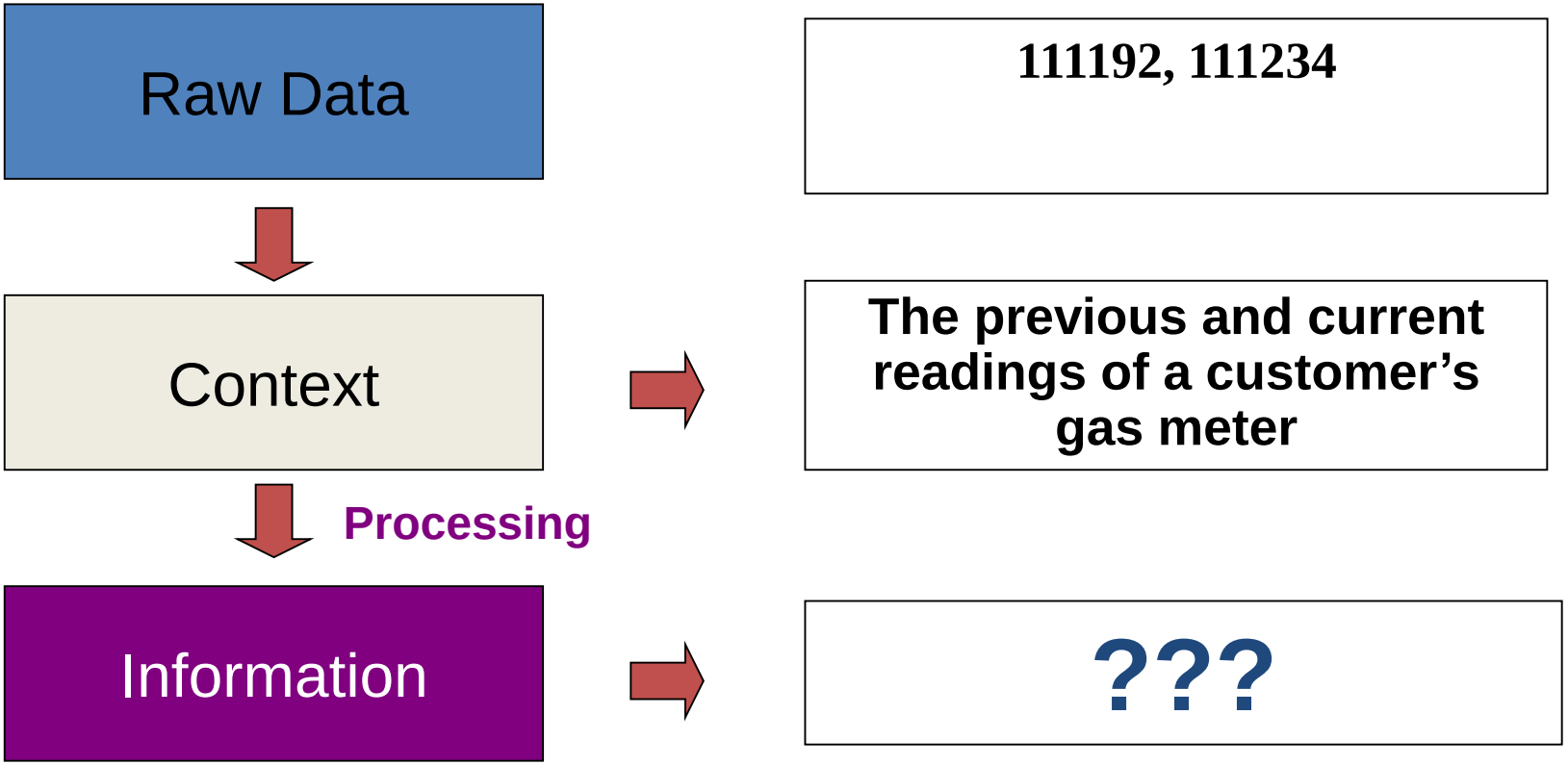
Example 1


Raw Data
Yes, Yes, No, Yes, No, Yes, No, Yes, No, Yes, Yes
Context
Responses to the market research question – “Would you buy brand x at price y?”
Processing
Information
???

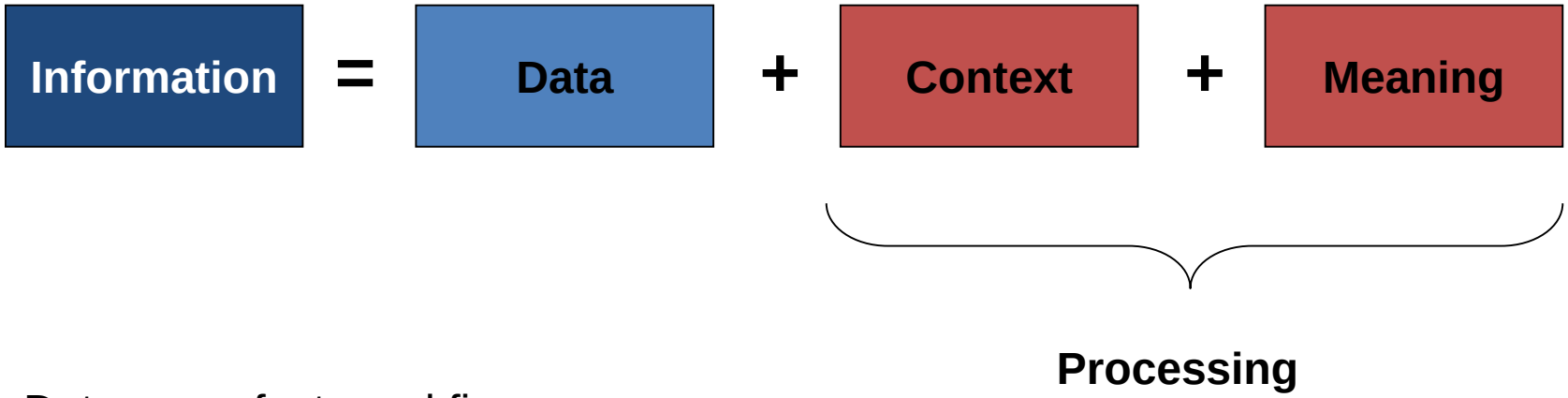
Example 2


Raw Data
42, 63, 96, 74, 56, 86
Context
Jayne's scores in the six AS/A2 ICT modules
Processing
Information
???

Example 3


Raw Data
111192, 111234
Context
The previous and current readings of a customer's gas meter
Processing
Information
???

داده – اطلاع – دانش



Data – raw facts and figures

Information – data that has been processed (in a context) to give it meaning

دانش (Knowledge)

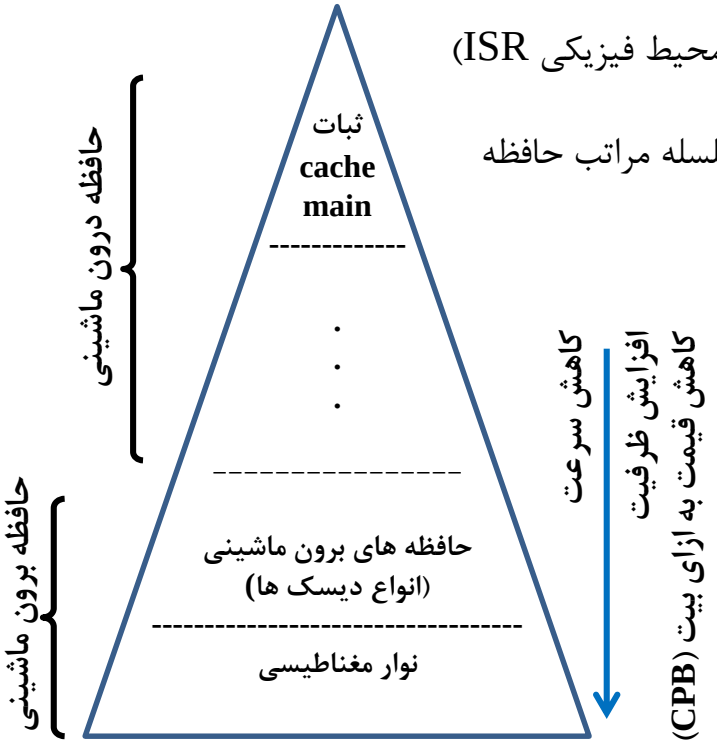
- تعریف [FROS87]: دانش عبارت است از نمایش نمادین جنبه‌هایی از بخشی از جهان واقع (جهان موردنظر یا محیط مطرح)
  - مثال: شنبه هوا بارانی است. حسن فرزند علی است.
- تعریف دوم [روحا ۹۱]: دانش منطقی نوعی شناخت است که از یک مجموعه از اطلاعات بر اساس یک مجموعه از قواعد مشخص، معمولا با روش استقراء حاصل می‌شود. حصول این شناخت می‌تواند توسط انسان یا یک سیستم خودکار انجام شود.

هر **سیستم نرم‌افزاری** از مجموعه ای از **داده‌های ذخیره شده** ممکن است استفاده کند.

در قالب تعدادی فایل (محیط فیزیکی ISR)

در کجا؟ ← در یک سلسله مراتب حافظه

به لحاظ ساختاری:
- ساختمند (structured)
- نیم ساختمند (semi-structured)
- ناساختمند (un-structured)



یادآوری

انواع سیستم نرم‌افزاری:

- بنیادی یا پایه (سیستم‌های عامل)
- نیمه بنیادی (DBMS، DMS، کامپایلرها، اسمبلرها، و ...)
- کاربردی (برنامه‌های کاربردی)
- ابزاری: انواع toolها

Data Management System

Database Management System

```xml
<students>
 <student id="100026">
  <name>Joe Average</name>
  <age>21</age>
  <major>Biology</major>
  <results>
   <result course="Math 101" grade="C-"/>
   <result course="Biology 101" grade="C+"/>
   <result course="Statistics 101" grade="D"/>
  </results>
 </student>

 <student id="100078">
  <name>Jack Doe</name>
  <Birthdate>1990/02/02</Birthdate>
  <major>Physics</major>
  <results>
   <result course="Math 101" grade="A"/>
   <result course="XML 101" grade="A-"/>
   <result course="Physics 101" grade="B+"/>
   <result course="XML 102" grade="A"/>
  </results>
 </student>
</students>
```

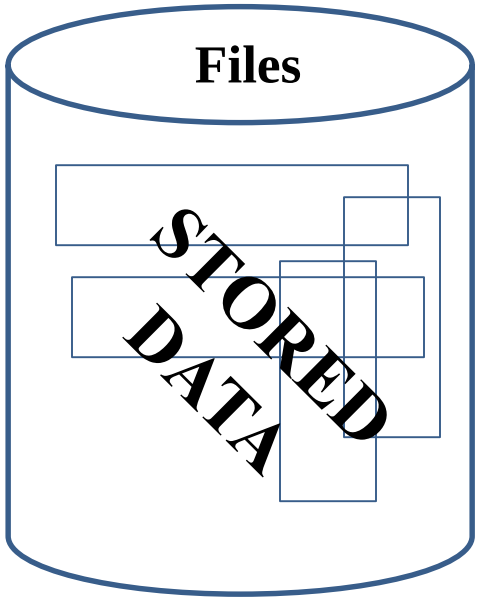
محیط فیزیکی «ذ.ب.ا»

کنجکاوی: دلایل استفاده از این سلسله مراتب حافظه چیست؟

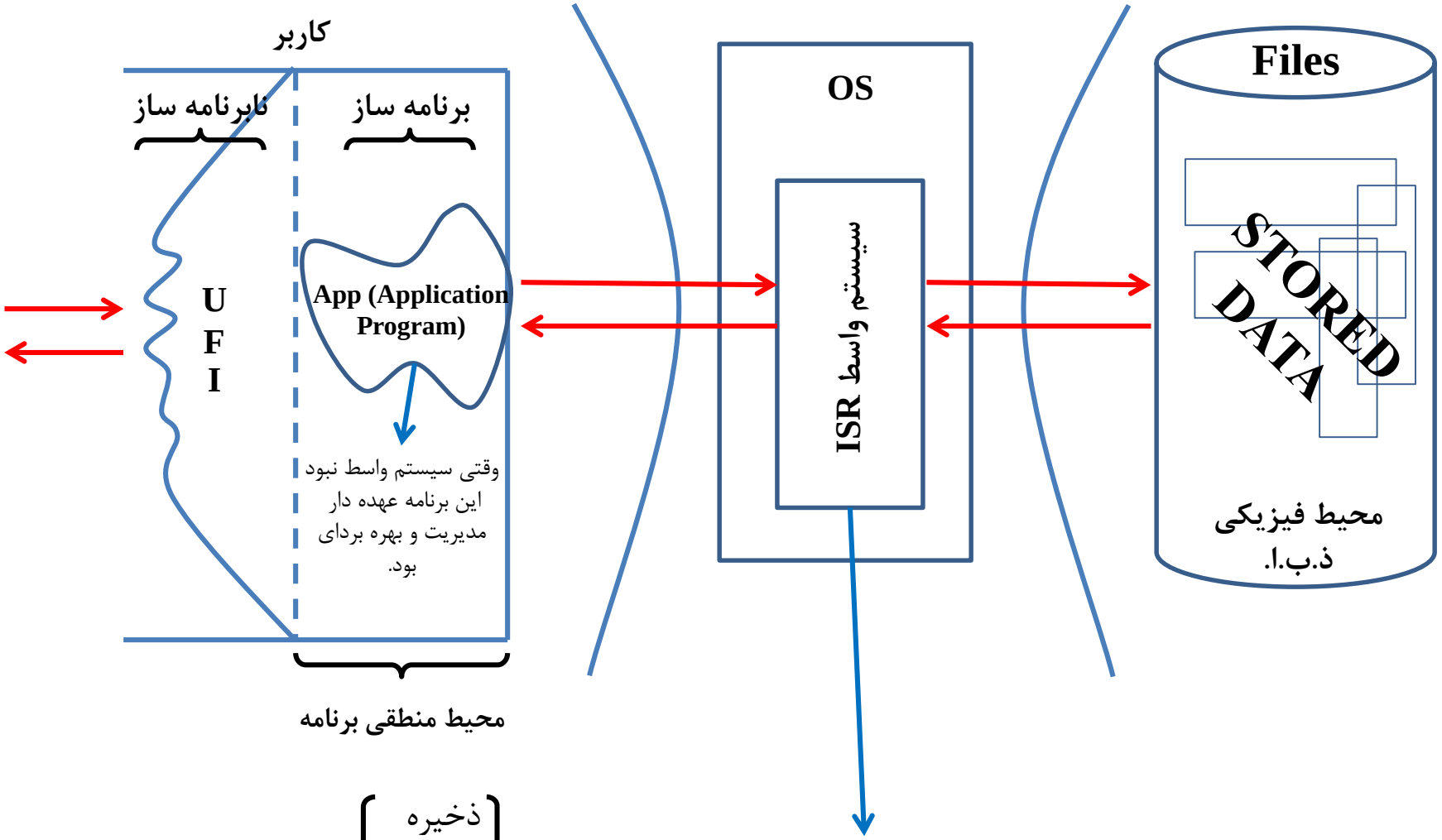
کنجکاوی: چه داده ای، برای چه مدتی، در کدامیک از مراتب سلسله مذکور قرار می‌گیرد؟

- **محیط فیزیکی "ذ.ب.ا." (ذخیره و بازیابی اطلاعات) یا ISR (Information Storage and Retrieval)**

ISR: باید { ایجاد / مدیریت / بهره برداری } شود. ← نیاز به یک سیستم واسط ذ.ب.ا داریم.



سیستم واسط ذ.ب.ا.



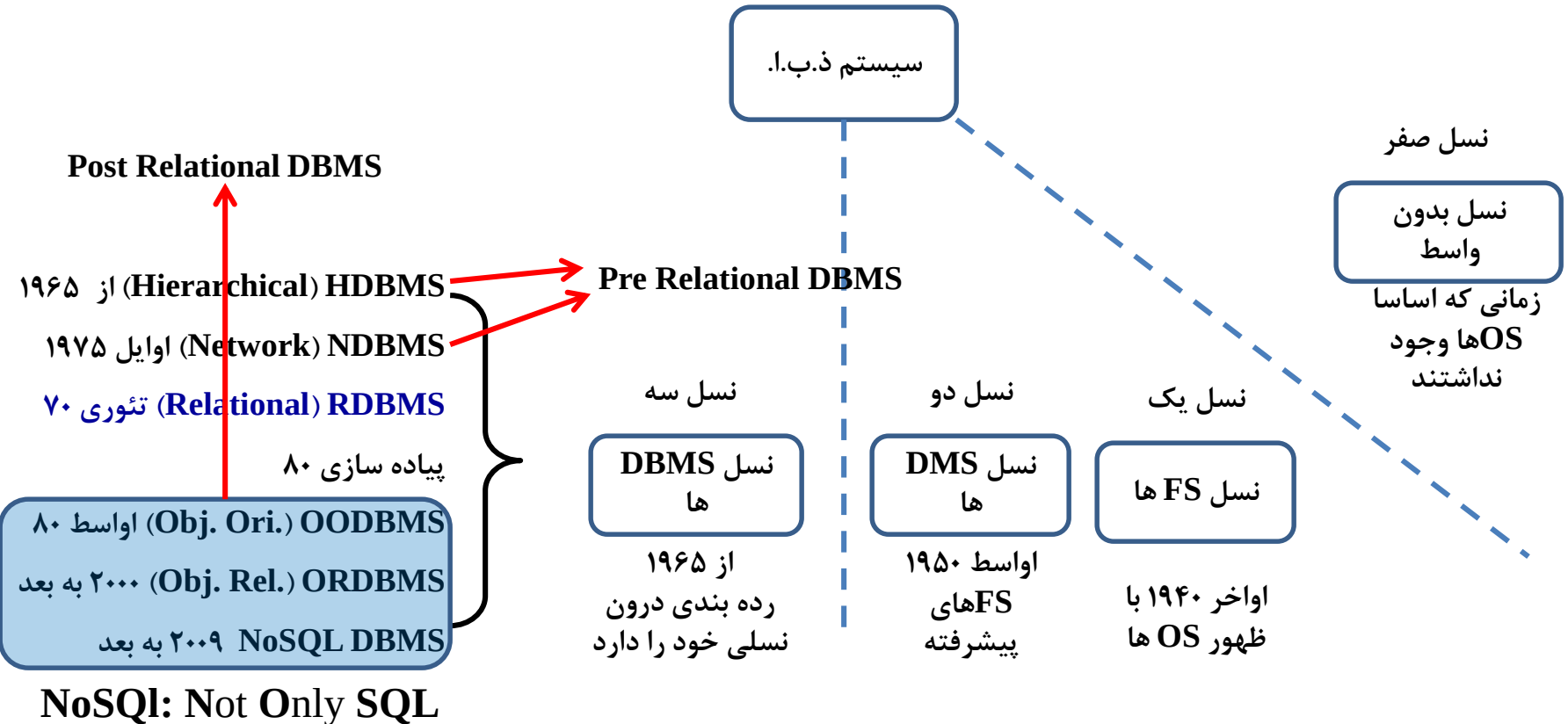
این سیستم این امکان را می‌دهد تا کاربر داده‌های خود را { ذخیره / بازیابی / پردازش } کند.

...

سیر تحول سیستم ذ.ب.ا. (2)

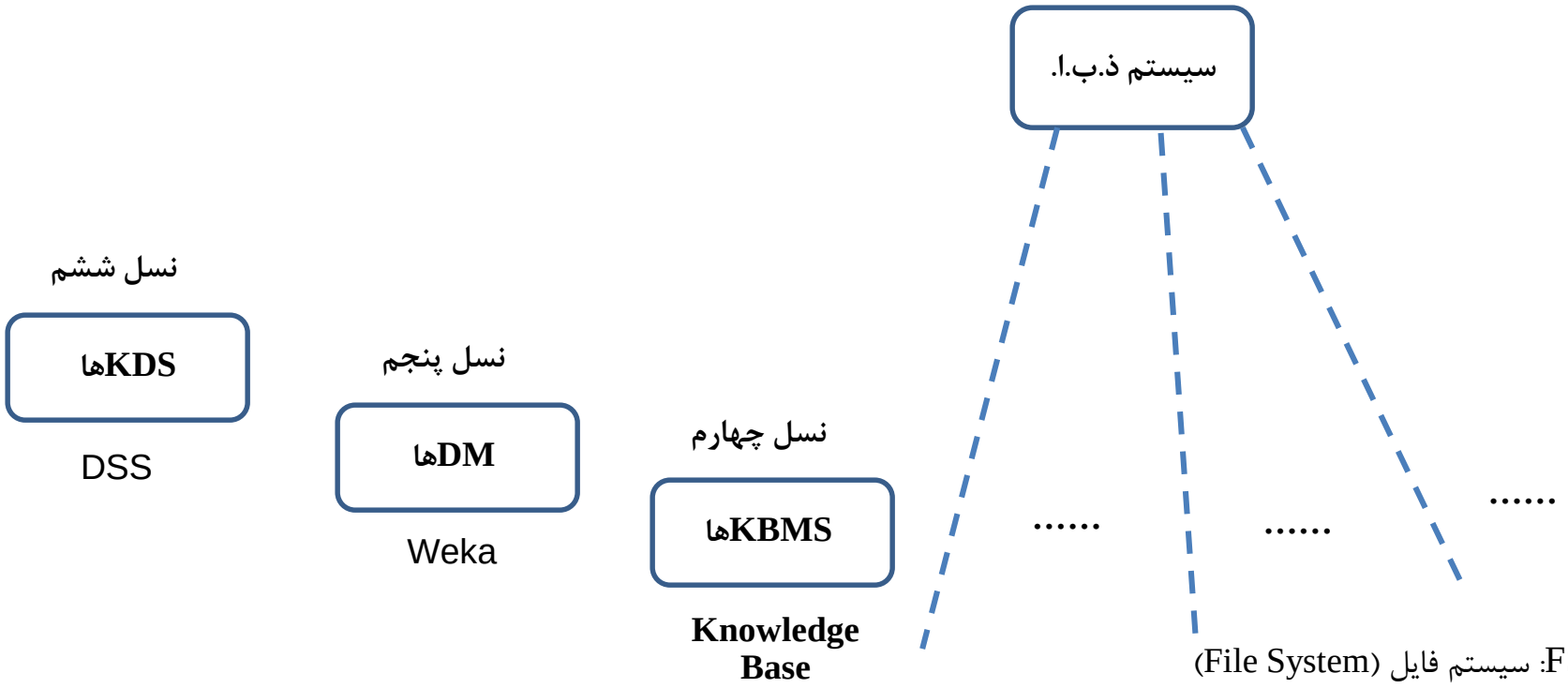
کنجکاوی: رده بندی از مفهوم کاربر ارایه کنید؟ به بیان دیگر گونه های دیگر کاربر کدامند؟

- **سیستم واسط "ISR" سیر تحول خاص خود را دارد :**
- ۶ نسل تکنولوژیک قابل بازیابی است (به طور کلی) [دیدگاه نرم‌افزاری]



NoSQl: Not Only SQL

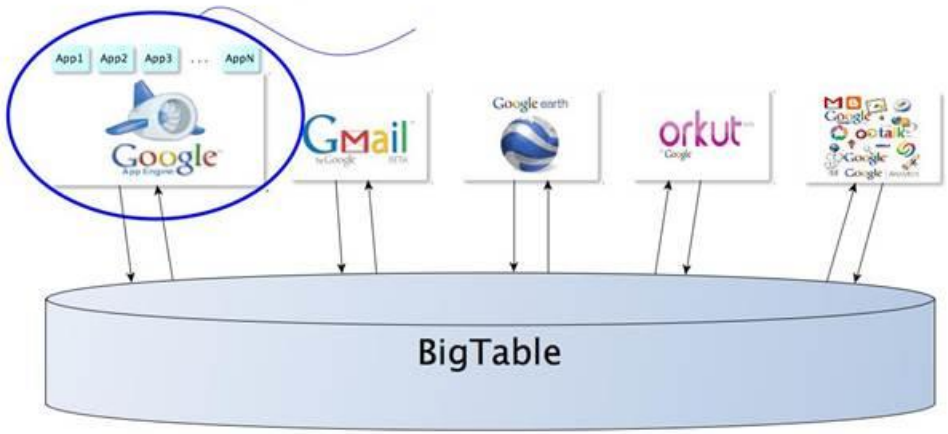
سیر تحول سیستم ذ.ب.ا. (3)



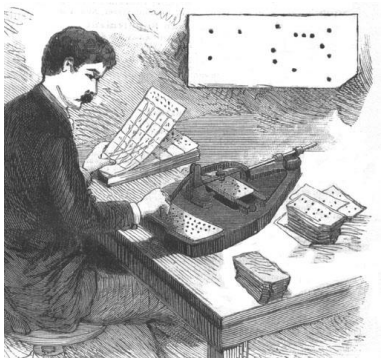
- FS: سیستم فایل (File System)
- DMS: سیستم مدیریت داده‌ها (Data Management System)
- DBMS: سیستم مدیریت پایگاه داده‌ها (DataBase Management System)
- KBMS: سیستم مدیریت پایگاه دانش (Knowledge Base Management System)
- DM: سیستم داده‌کاوی (Data Mining System)
- KDS: سیستم کشف دانش (Knowledge Discovery System)

سیر تحول سیستم ذ.ب.ا. (4)

- در این نسل‌بندی، نسل بعدی نسل قبلی را منسوخ نمی‌کند. نسل بعدی نسل قبلی را تکمیل می‌کند و از آن استفاده می‌کند. هنوز از سیستمهای شبیه پانچ کارت استفاده میشود. کجا؟
- انواع نیازهای پردازشی، کنترلی، و عملیاتی سبب ایجاد نسل‌های سیستم «ذ.ب.ا.» شد.
- مثلا گوگل از سیستم «ذ.ب.ا.» جدیدی با نام BigTable برای ذخیره داده های مربوط به سیستمهای خود استفاده می کند.
  - داده های بسیار حجیم
  - برچسب زمانی
  - برچسب مکانی
  - استفاده همزمان توسط برنامه های مختلف



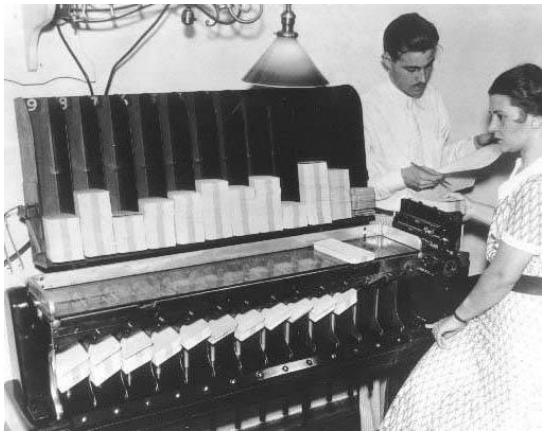
سیر تحول سیستم ذ.ب.ا. (5)



Data entry

Storage and retrieval

Query processing



Sorting

IBM 305 RAMAC

The first commercial computer that used a moving-head hard disk drive (magnetic disk storage) for secondary storage.[1] The system was publicly announced on September 14, 1956,

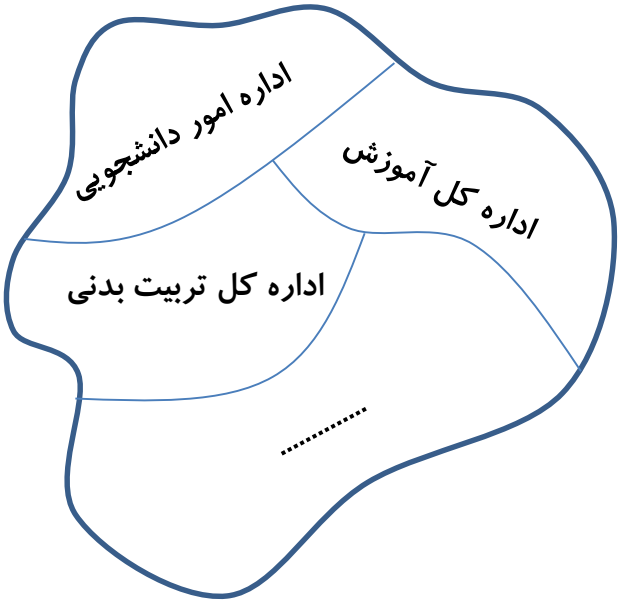
تعریف پایگاه داده:

مجموعه‌ای است از داده‌های ذخیره شده، پایا ، مجتمع ، به هم مرتبط ، و حتی الامکان فاقد افزونگی، (دارای معماری خاصّ خود، مبتنی بر یک مدل داده ای مشخص) ، تحت کنترل یک سیستم متمرکز، مورد استفاده یک یا چند کاربر در یک سازمان (در یک محیط) به طور اشتراکی و همروند.

Persistent
Integrated
Interconnected /Related
Redundancy
Data Model
Shared
Concurrent
Centralized System
Multi User

پایگاه داده – مثال مقدماتی

مثال کاربردی



محیط عملیاتی: دانشگاه

بخشی از جهان واقعی که قصد ایجاد سیستم برای آن را داریم.

Micro Real World (خرد جهان واقع)
Mini World
Universe of Discourse (جهان مطرح)

- نکته: هر محیط از تعدادی زیر محیط تشکیل شده است.
- در هر محیط مجموعه‌ای از نوع موجودیت‌ها(Entities) وجود دارند که نیازهای کاربران ناظر به { داده‌ای / پردازشی } آنهاست (یعنی به داده‌هایی در مورد آنها نیاز دارند).

مثال مقدماتی

- **نکته:** زیرمحیط های یک محیط معمولا با هم اشتراک دارند در نوع موجودیت‌ها (Entity Type یا Object Type)
  - مثال: در محیط دانشگاه دانشجو، استاد، درس، کلاس، و ...
  - مثال: نوع موجودیت دانشجو در هر سه زیر محیط مطرح است.

- **مسئله (خواسته): ایجاد سیستم(های) کاربردی برای این زیر محیط‌ها**
  - برای این منظور در اساس دو مَشی-روش (approach) وجود دارد.
    - مشی فایلینگ [سنتی یا کلاسیک] یا ناپایگاهی
    - مشی پایگاهی Database Approach

    یعنی ممکن است مشی‌های بینابینی نیز وجود داشته باشد.

ادامه مثال مقدماتی (مشی فایلینگ)

کارهای لازم در مشی فایلینگ به طور خلاصه:

- توجه: این کارها معمولا برای هر زیرمحیط به طور جداگانه انجام می شود. ← تعدادی سیستم کاربردی جدا (نامجتمع) و بی ارتباط در یک محیط ...

۱- مطالعه و شناخت محیط

۲- انجام مهندسی نیاز ها Requirement Engineering ← ۱- تشخیص نیاز های داده‌ای
۲- تشخیص نیاز های پردازشی
۳- مستندسازی نیازها
۴- دریافت تایید سازمان

۳- تعیین مشخصات سیستم کاربردی System Specification

۴- [انتخاب پیکربندی سخت افزار و نرم افزار H/S]

۵- [انتخاب یک FS و/یا DMS]سیستم واسط ISR

۶- طراحی تعدادی فایل (طبق مشخصات سیستم)

طراحی فایل

طراحی

۶-۱- تعیین فرمت رکورد

۶-۲- تعیین ساختار فایل

ساختار فایل: ساختاری که براساس آن فقره داده ها (رکوردها) در سطح منطقی [و/یا فیزیکی] با یکدیگر مرتبطند.

ساختار فایل یک امکان برای نمایش ارتباط بین فقره داده‌هاست (Data Items) خواه در سطح نمایش منطقی باشد یا فیزیکی.

کنجکاوی: چند نوع ساختار فایل وجود دارد؟

۶-۳- نحوه دسترسی به رکوردها – استراتژی دسترسی

۶-۴- اندازه فایل ها

۶-۵- میزان گسترش چه میزان باشد

۶-۶- ارتباط با فایل های دیگر

۶-۷- عملیات مجاز در فایل‌ها + کاربران

مثال مقدماتی (مشی فایلینگ)

- **کارهای لازم در مشی فایلینگ به طور خلاصه : (ادامه)**

۷- طراحی واسط‌های کاربری (UFI)

۸- طراحی تعدادی برنامه کاربردی (Application Program) [ضمن تعیین تراکنش(ها)]

۹- تولید برنامه‌های { ایجاد / کنترل / پردازش } فایل‌ها

۱۰- ایجاد محیط فیزیکی «ذ.ب.ا.» به طور آزمایشی (برای داده های تست)

۱۱- ایجاد محیط فیزیکی «ذ.ب.ا.» با داده‌های واقعی اما حجم محدود و انجام تست مرحله دوم

[۱۲- ایجاد محیط فیزیکی «ذ.ب.ا.» با داده‌های واقعی و حجم واقعی و انجام تست مرحله سوم]

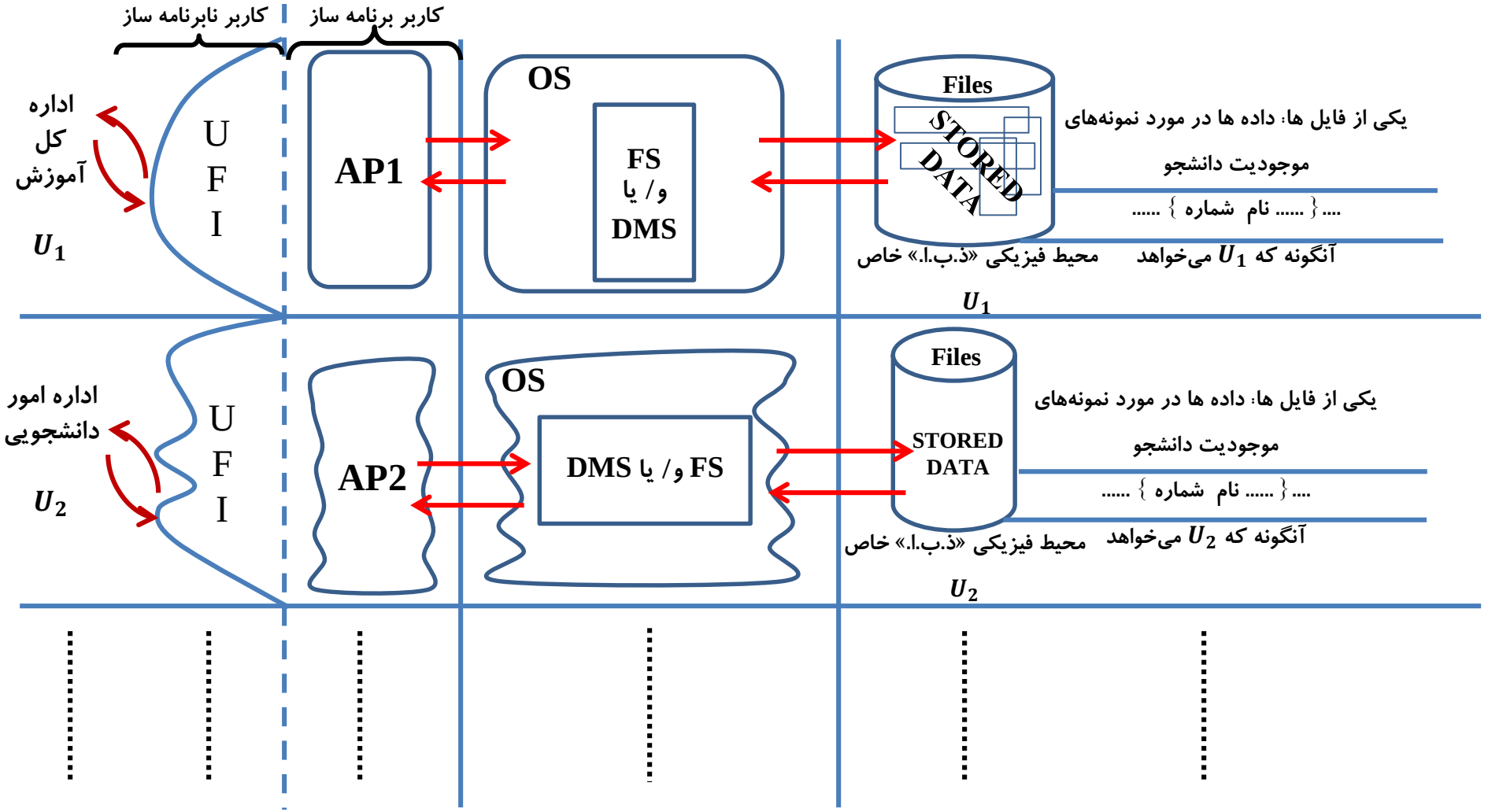
۱۳- رفع اشکال‌ها در هر مرحله

۱۴- ایجاد محیط فیزیکی واقعی با نصب، پیکربندی و ورود داده‌های اولیه (Data Entry)

۱۵- آغاز بهره برداری و نگهداری سیستم

۱۶- رفع معایب و بهینه‌سازی سیستم

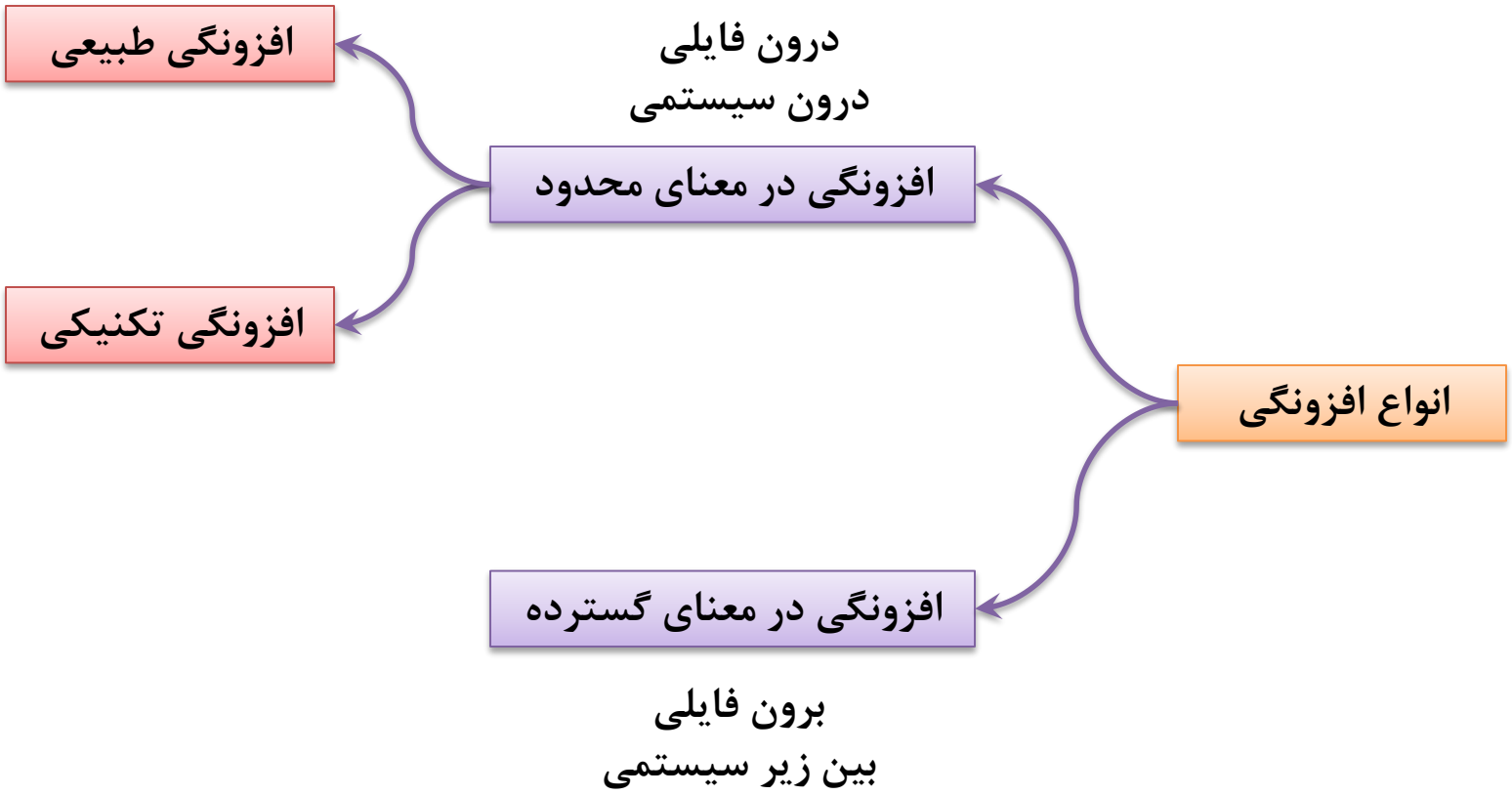
نمایش شماتیک مشی فایلینگ



معایب مشی فایلینگ

- **برخی از معایب مشی فایلینگ:**
  - وجود سیستم های نامجتمع در یک سازمان [محیط] و نامرتبط به هم
  - عدم وجود یک سیستم کنترل متمرکز روی کل داده‌های سازمان
  - وجود افزونگی زیاد (تکرار در ذخیره سازی داده ها)
  - خطر بروز ناسازگاری داده ها (Data Inconsistency) ← کنجکاوی: جنبه های بروز ناسازگاری کدامند؟
  - عدم امکان اعمال ضوابط حفظ امنیت داده‌ها (Data Security)
  - عدم امکان اشتراکی شدن داده ها (Data Sharing) [یا در حداقل و یا با دشواری]
  - مصرف نابهینه سخت افزار (به ویژه سخت افزار ذخیره‌ساز)
  - وابسته بودن برنامه ها به جنبه های فایلینگ محیط ذخیره‌سازی، به گونه‌ای که اگر قرار باشد در فایلینگ تغییراتی ایجاد شود، برنامه ها هم متناسبا باید تغییر یابد. (به طور مثال فرمت ساختار یا نحوه دسترسی (Access Strategy) را تغییر دهیم)

افزونگی Data Redundancy



افزونگی (ادامه)

- **افزونگی در معنای محدود** (یعنی درون فایلی -intrafile redundancy – در مباحث فایلینگ)
- عبارت است از تکرار ذخیره سازی مقادیر (value) یک صفت یا بیش از یک صفت در فایل داده‌ای یا فایل کمکی آن.
- این نوع افزونگی گونه‌هایی دارد:

۱- **طبیعی**: ناشی از ماهیت داده های محیط (مثل صفت رشته دانشجو که برای دانشجویان مختلف می‌تواند یکسان و در نتیجه تکراری باشد یا نمره دانشجویان)

- کنجکاوی: برای کاهش مصرف حافظه در حالت افزونگی طبیعی چه باید کرد؟

۲- **تکنیکی**: ناشی از استفاده از یک تکنیک *معمولا* برای افزایش سرعت (مثل نمایه سازی [شاخص بندی Indexing] و یا Replication در پایگاه داده ها)

افزونگی (ادامه)

- **افزونگی در معنای گسترده** (یعنی برون‌فایلی- در مباحث پایگاه داده)
- عبارت است از تکرار ذخیره‌سازی داده‌ها در مورد نمونه‌های یک یا بیش از یک نوع موجودیت از یک محیط.
  - این نوع افزونگی نه از نوع طبیعی و نه از نوع تکنیکی است بلکه ناشی از رهیافت انتخاب شده برای طراحی و تولید سیستم‌های کاربردی است.
  - به طور مثال تکرار اطلاعات دانشجویان در دو زیرسیستم اداره کل آموزش و زیرسیستم اداره امور دانشجویی.

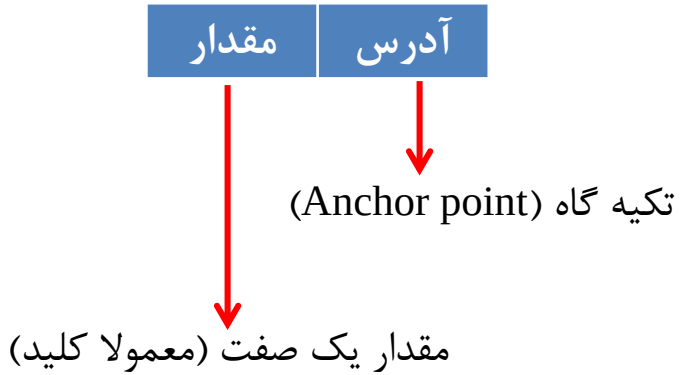
نکته: افزونگی از نوع طبیعی و تکنیکی در پایگاه داده هم می‌تواند وجود داشته باشد.

یادآوری: نمایه‌سازی (نمونه‌ای از افزونگی تکنیکی)

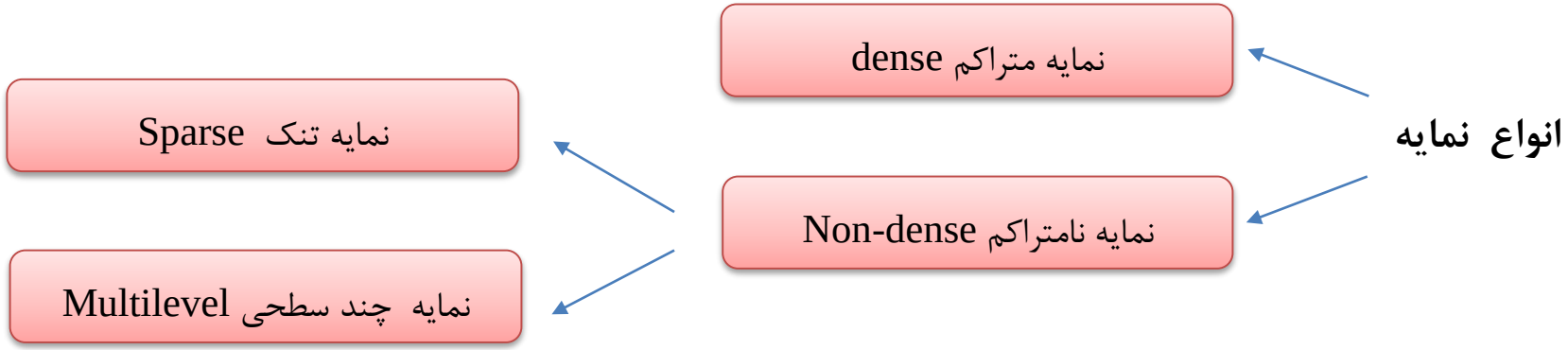
یادآوری

نمایه سازی مکانیزمی است که امکان **دسترسی مستقیم** (direct access)به داده های بانک اطلاعاتی می دهد.

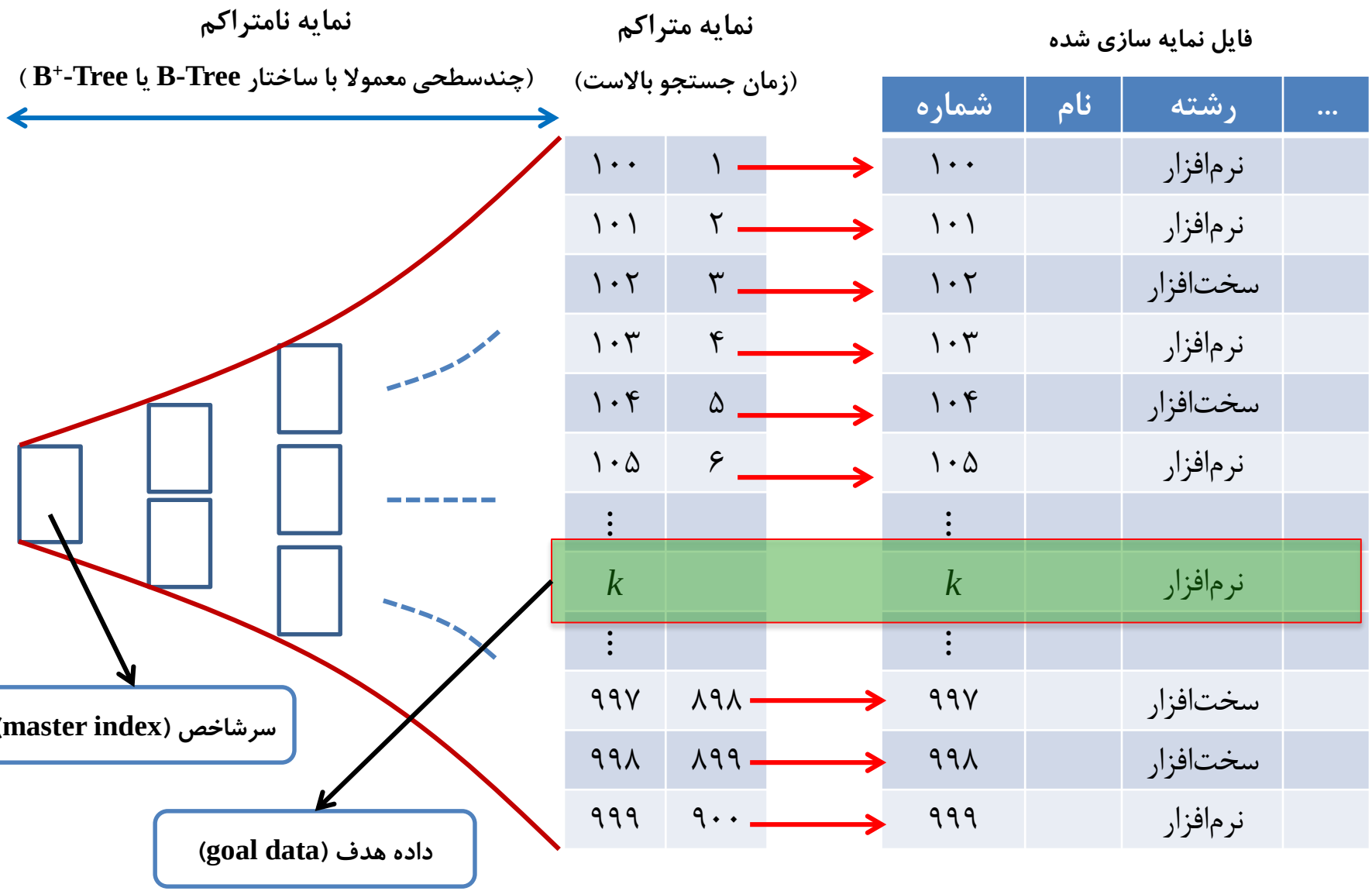
- تشکیل شده از تعدادی درایه (مدخل-entry)



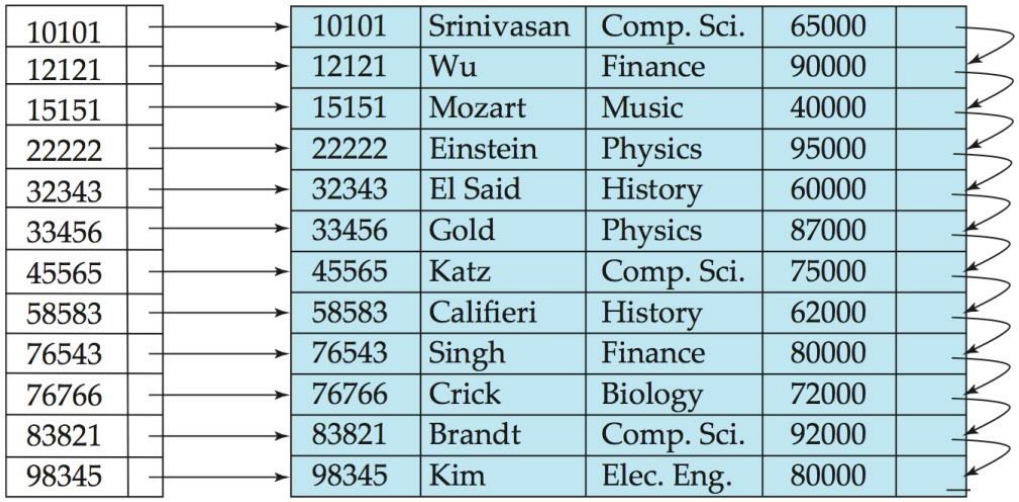
هر مدخل اشاره دارد به یک یا گروهی از رکورد ها به صورت یک یا چند سطحی



یادآوری: نمایه‌سازی،نمونه‌ای از افزونگی تکنیکی(ادامه)



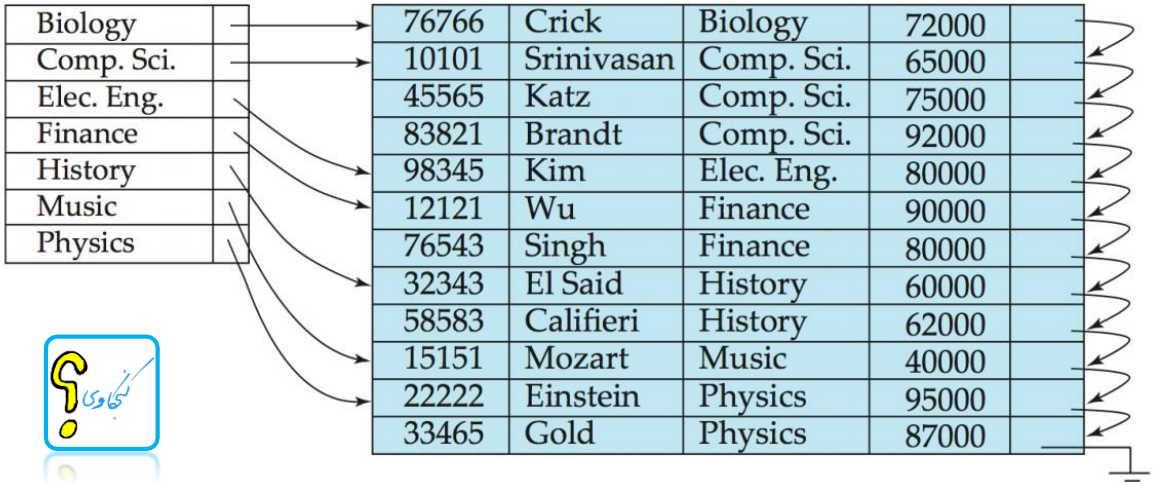
Dense indexing



10101	10101	Srinivasan	Comp. Sci.	65000
12121	12121	Wu	Finance	90000
15151	15151	Mozart	Music	40000
22222	22222	Einstein	Physics	95000
32343	32343	El Said	History	60000
33456	33456	Gold	Physics	87000
45565	45565	Katz	Comp. Sci.	75000
58583	58583	Califieri	History	62000
76543	76543	Singh	Finance	80000
76766	76766	Crick	Biology	72000
83821	83821	Brandt	Comp. Sci.	92000
98345	98345	Kim	Elec. Eng.	80000

نمایه سازی متراکم

وقتی بر روی ستونی می خواهیم عمل اندیس گذاری را انجام دهیم می بایست به ازای هر کلید جستجو (search-key) غیر تکراری در ستون مورد نظر، یک رکورد در فایل نمایه مربوط به آن ستون اضافه کنیم.

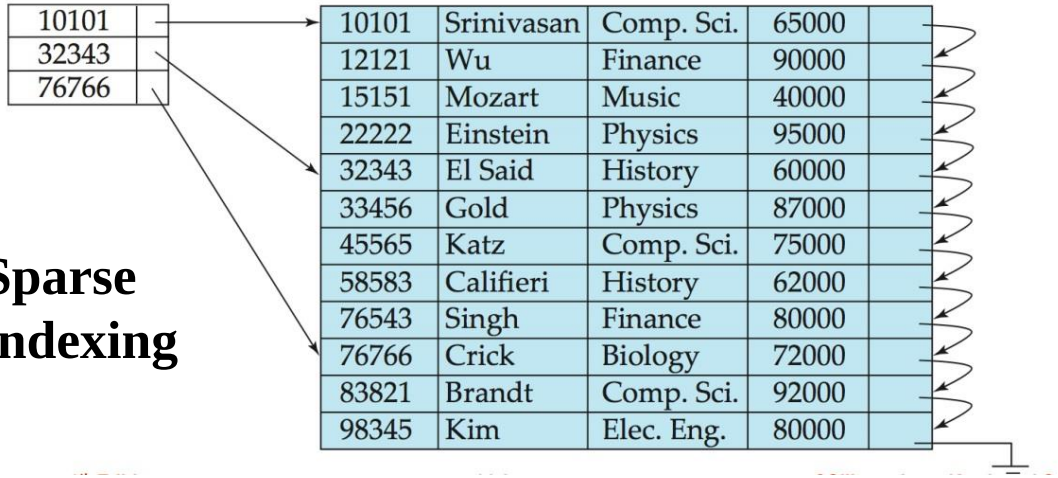


Biology	76766	Crick	Biology	72000
Comp. Sci.	10101	Srinivasan	Comp. Sci.	65000
Elec. Eng.	45565	Katz	Comp. Sci.	75000
Finance	83821	Brandt	Comp. Sci.	92000
History	98345	Kim	Elec. Eng.	80000
Music	12121	Wu	Finance	90000
Physics	76543	Singh	Finance	80000
	32343	El Said	History	60000
	58583	Califieri	History	62000
	15151	Mozart	Music	40000
	22222	Einstein	Physics	95000
	33465	Gold	Physics	87000

در نمایه سازی متراکم زمان جستجو بالاست

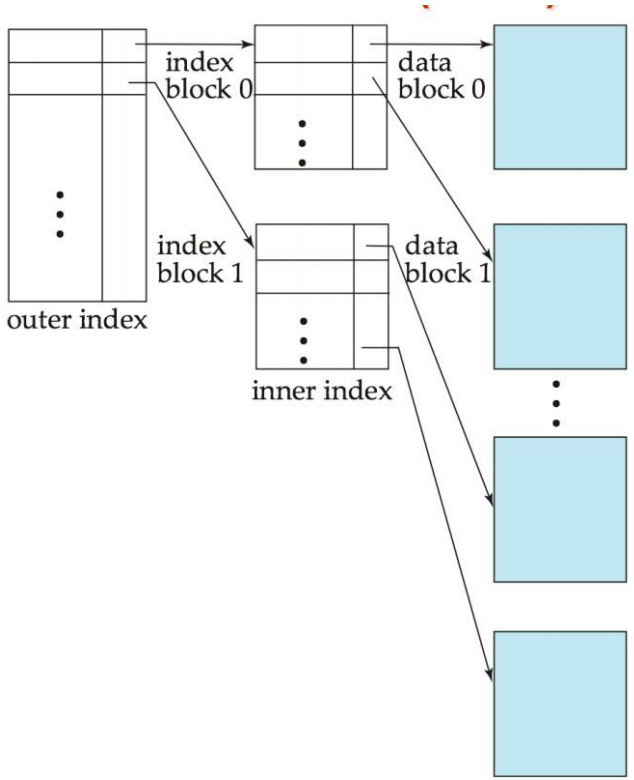
نکته: مرتب بودن فایل نمایه چه کمکی می کند؟

Non-dense indexing



Sparse indexing

Multilevel Index



مثال مقدماتی : اما در مشی پایگاهی

- **کارهای لازم در انجام یک «پروژه پایگاهی»**: (فعلا نه در جزئیات)

به منظور ایجاد یک سیستم یکپارچه

۱- مطالعه و شناخت محیط

۲- انجام مهندسی نیاز ها Requirement Engineering ← ۱- تشخیص نیاز های داده ای؛ ۲- تشخیص نیاز های پردازشی؛ ۳- مستندسازی نیاز ها؛ ۴- دریافت تایید سازمان

۳- تعیین مشخصات سیستم یکپارچه

[۴- انتخاب (حداقل) یک پیکربندی H/S]

[۵- انتخاب (حداقل) یک DBMS (یک تصمیم گیری حیاتی است)]

۶- مدلسازی معنایی داده ها (data semantic modeling) ← در مشی فایلینگ انجام نمی‌شود.

۷- طر احی منطقی پایگاه داده Logical Data Base Design

۸- طراحی فیزیکی پایگاه داده Physical Data Base Design

۹- طراحی UFI ها

} مجتمع

۱۰- طراحی AP ها [ضمن تعریف تراکنش ها (Transactions)]

ادامه مثال مقدماتی (مشی پایگاهی)

ادامه:...

۱۱- تولید برنامه‌های تعریف (ایجاد) و کنترل DB

۱۲- تولید برنامه‌های عملیات در داده‌ها (پردازش داده‌ها)

۱۳- ایجاد محیط فیزیکی «ذ.ب.ا.» با داده های تستی و رفع اشکال ها (تست مرحله اول)

۱۴- ایجاد محیط فیزیکی «ذ.ب.ا.» با داده های واقعی اما حجم محدود و انجام تست مرحله دوم

۱۵- ایجاد محیط فیزیکی «ذ.ب.ا.» با داد های واقعی و حجم واقعی و انجام تست مرحله سوم

۱۶- تنظیم سیستم پایگاهی (Data Base System Tuning)

۱۷- آغاز بهره برداری و نگهداری از سیستم

۱۸- گسترش سیستم

۱۹- رفع معایب و بهینه‌سازی سیستم

کنکاوی؟

مزایا و معایب جداسازی این دو دسته برنامه

تعریف و کنترل و عملیات در داده‌ها چیست؟

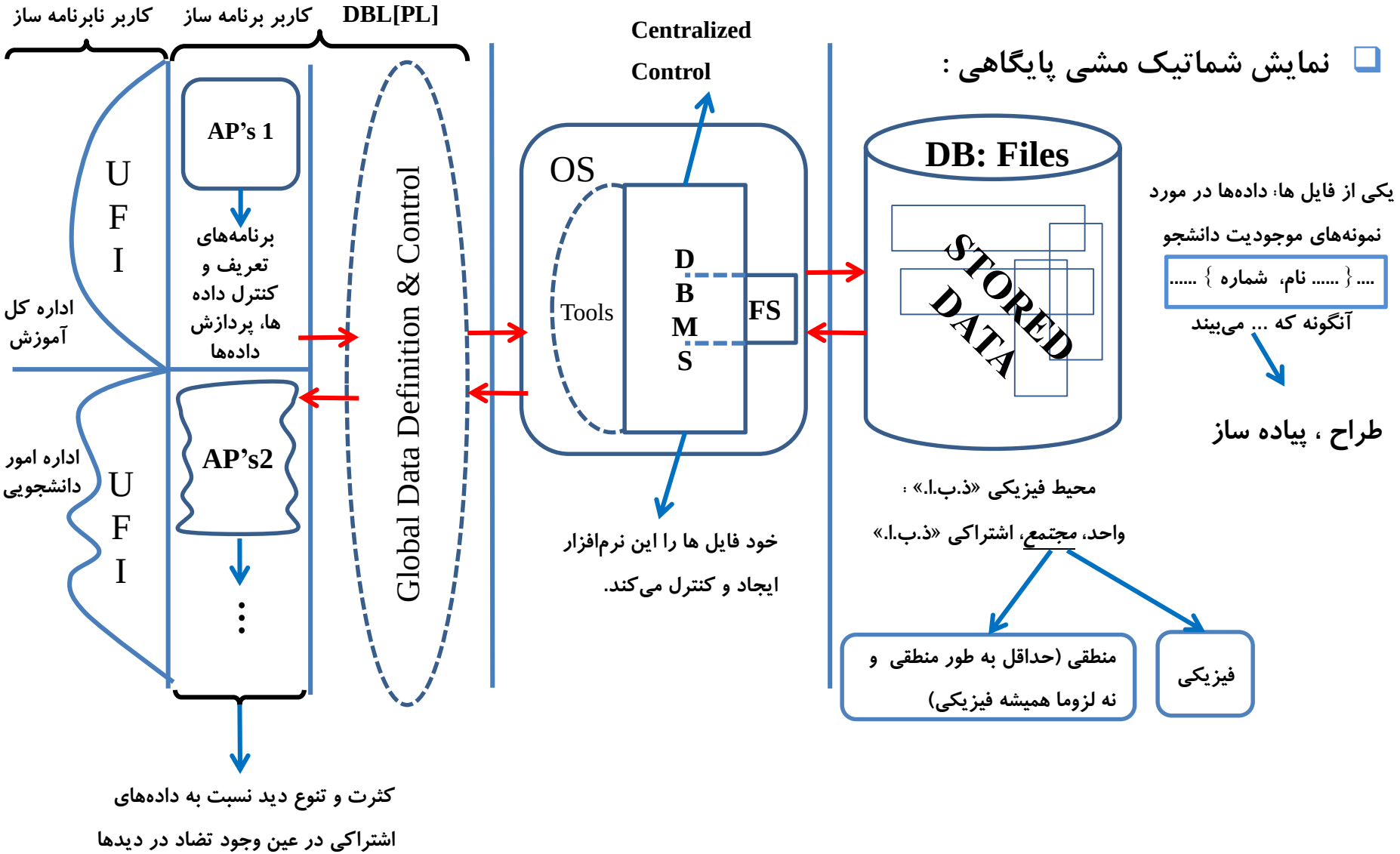
۱- از دیدگاه عملیات در داده‌ها

۲- از دیدگاه زبان‌های برنامه‌سازی

به طور مثال به منظور افزایش کارایی

یکی از ویژگی های DBMS گسترش پذیری سیستم است.

ادامه مثال مقدماتی (مشی پایگاهی)



- نمایش شماتیک مشی پایگاهی :

یکی از فایل ها: داده‌ها در مورد نمونه‌های موجودیت دانشجو

{ ..... نام، شماره } .....

آنگونه که ... می‌بیند

طراح ، پیاده ساز

ادامه مثال مقدماتی (مشی پایگاهی)

چگونه از این کثرت دید می توان به آن «وحدت» رسید؟ ← ۱- خود نرم افزار DBMS

۲- معماری پایگاه داده

- تمرین: مزایای مشی پایگاهی چیست؟ ← (طبق معلومات فعلی: عکس معایب مشی فایلینگ)
- تمرین: چند سطح تعریف داده داریم؟

عناصر محیط پایگاهی

عناصر اصلی محیط پایگاهی:

۱- سخت افزار ←
- ذخیره سازی
- پردازشگر
- ارتباطی (همرسانی) Data Communication

۲- نرم افزار

۳- کاربر

۴- داده

عناصر محیط پایگاهی – (۱) سخت‌افزار

- **سخت افزار ذخیره سازی:**
  - رسانه اصلی: دیسک ترجیحا با تکنولوژی RAID (Redundant Array of Independent /Inexpensive Disk)
  - رسانه فرعی: نوار مغناطیسی [از جمله برای تولید نسخه های پشتیبان]

- اغلب DBMSهای امروزی تکنیک های تولید Back up را دارا هستند.

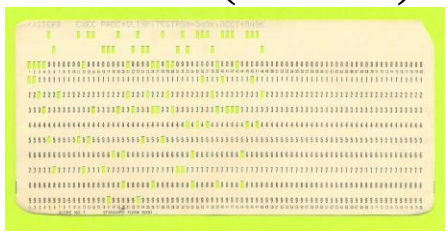
- تکنیک های تولید نسخه پشتیبان؟
- سطوح مختلف Back up

- **سخت افزار پردازشگر:**
  - کامپیوتر های معمولی از هر رده [PC, main,…]
  - اما ماشین‌های خاص DB هم داریم : DB Machines

- **سخت افزار ارتباطی (همرسانی):**
  - امکانات محلی: برای ارتباط دستگاه‌های جانبی با پردازنده
    SAN (storage area network)
    NAS (Network attached storage)
  - امکانات شبکه‌ای: برای ایجاد شبکه در سیستم پایگاهی نامتمرکز

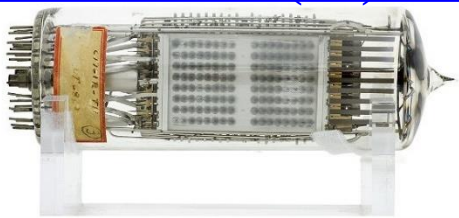
سیر تحول رسانه های ذخیره سازی

1- Punch cards (1920-40):70B



2- Punch tapes (1950-1960):300B

3- Selectron tube (2K):$500



4- Magnetic tape (1960): 500KB

5-Cassette tape

6-MAGNETIC DRUM

7-FLOPPY DISK

8-HARD DRIVE

9-COMPACT DISC (1982): 700 MB

10- Digital Versatile Disc, DVD(1995, 4GB)

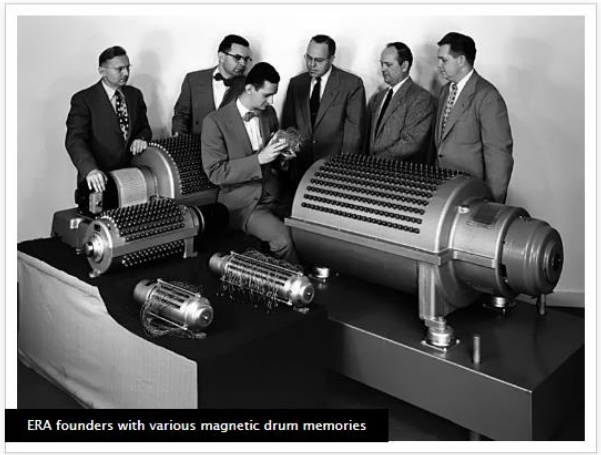
11-Flash memories (1980-1998)

12-Holographic Versatile Disc, HVD (4 TB)

سیر تحول رسانه های ذخیره سازی

3000 bits

Magnetic drum memory



1932

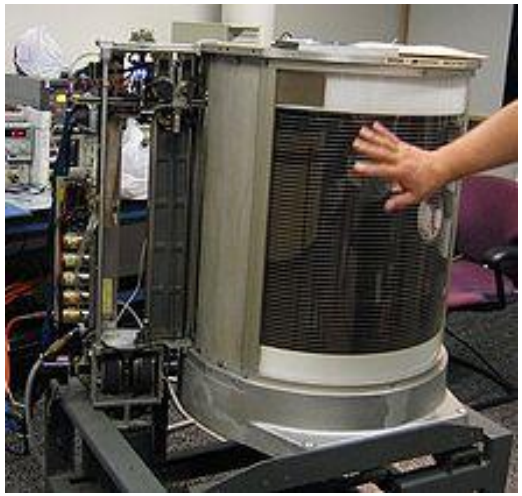
2.3 MB

IBM 726 Magnetic tape

1952

5 MB

IBM RAMAC 305 disk



1956

سیر تحول رسانه های ذخیره سازی ۲

DVD

4.7 GB

=

90 000 000 50 Byte

6 000 720 kB

4 500 1MB

0.2% 2 TB

2014: IBM 3592/1150 tape cartridge **Up to 7 TB**

http://www.zetta.net/about/blog/history-data-storage-technology
http://www.computerhistory.org/timeline/memory-storage/

October 2014:

Sony announced in 2014, that they had developed a tape storage technology with the highest reported magnetic tape data density, **148 Gbit**/in², potentially allowing tape capacity of **185 TB**.

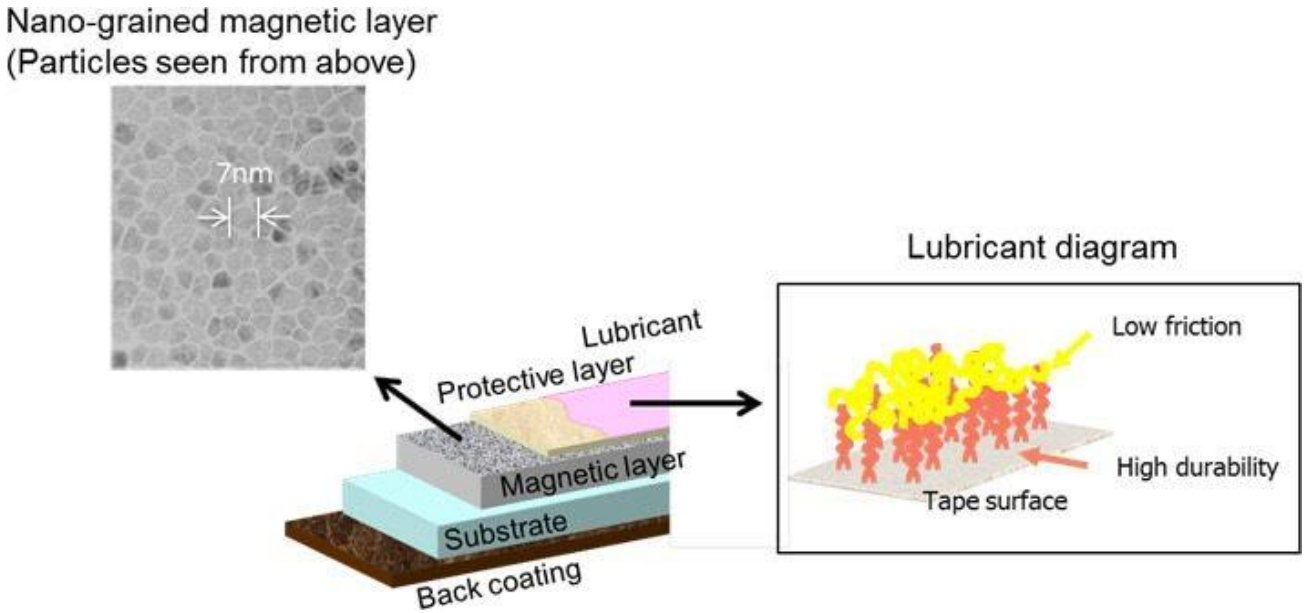
https://newatlas.com/sony-ibm-magnetic-tape-density-record/50743/

August 2, 2017:

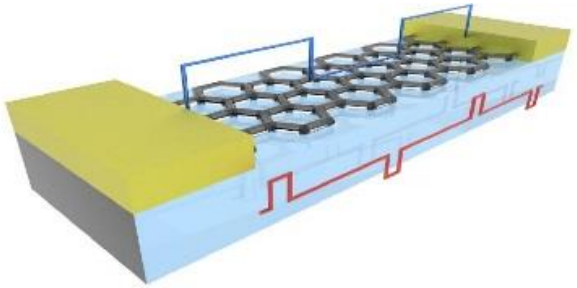
Sony Develops Magnetic Tape Storage Technology with the Industry's Highest*1 Recording Areal Density of 201 Gbit/in²

20x Greater Density Than Conventional Technology for Supporting High-Capacity **330 TB** Recording

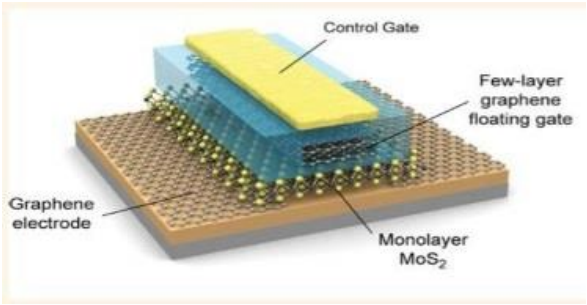
https://www.sony.net/SonyInfo/News/Press/201708/17-070E/index.html



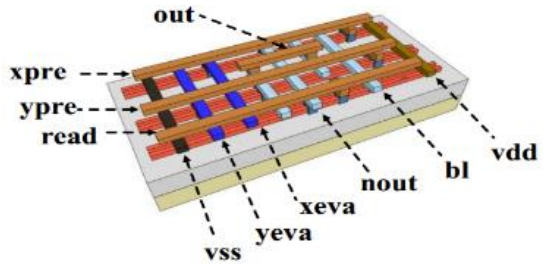
نسل آینده حافظه ها: Nano-scale Memory



- Graphene nanoribbon memory cell
- Reaching the 10 nm scale



Nano-scale flash memory using graphene



8T-Nanowire RAM Array

Database Machine

ORACLE EXADATA DATABASE MACHINE X6-2, (April 2016)

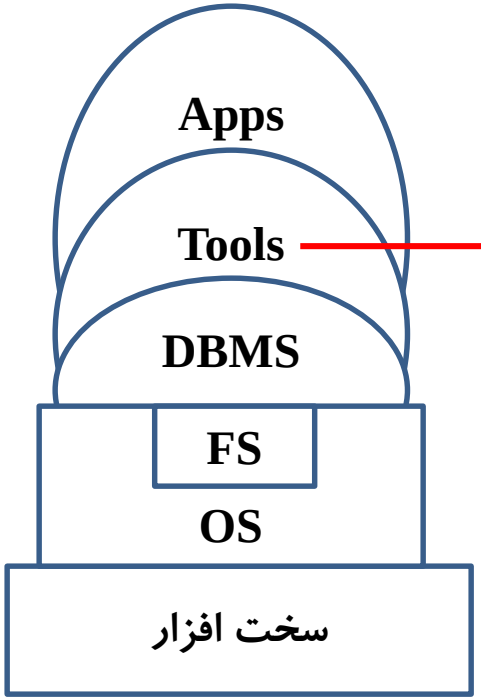
KEY FEATURE S

- Up to 836 CPU cores and 28.5TB memory per rack for database processing
- Up to 360 CPU cores per rack dedicated to SQL processing in storage
- From 2 to 19 Database Servers per rack
- From 3 to 18 Storage Servers per rack
- Up to 460 TB of flash capacity (raw) per rack
- Up to 1.7 PB of disk capacity (raw) per rack
- Hybrid Columnar Compression often delivers
- 10X-15X compression ratios
- 40 Gb/second (QDR) InfiniBand Network
- Complete redundancy for high availability

https://www.oracle.com/engineered-systems/exadata/index.html

عناصر محیط پایگاهی – (۲) نرم افزار

- **انواع نرم افزارهای مطرح در محیط پایگاهی:**
  - سیستم عامل و سیستم فایل (OS و FS)
  - سیستم مدیریت پایگاه داده‌ها (DBMS)
  - ابزارها (Tools)
  - برنامه‌های کاربردی (Apps)



تسهیلات نرم افزار ← یا با خود DBMS می فروشند، یا جداگانه خریداری می‌شود و به امکانات آن اضافه می‌شود.

عناصر محیط پایگاهی – (۳) کاربر

- به طور عام کاربر یا استفاده‌کننده از سیستم پایگاهی است (**کاربرنهایی/ کاربر نابرنامه ساز)** و یا در طراحی و پیاده سازی سیستم پایگاهی نقش دارد (**کاربر برنامه ساز**).
- هر کدام را می توان بر اساس معیارهای مختلف دسته بندی کرد. مثلا کاربر نهایی را می توان بر اساس شیوه تعامل با سیستم پایگاهی به کاربران برخط (Online)، گروهی (Batch) یا تعاملی (Interactive) تقسیم بندی کرد.

یک دسته بندی برای کاربران برنامه ساز

- تحلیلگر داده ها
- مدل ساز داده ها
- طراح واسط کاربری
- مدیر پایگاه داده (Database administrator)
  - مدیر داده Data Administrator
  - مدیر کنترل کارایی
  - مدیر سیستمهای کاربردی
  - مدیر ارتباط با کاربران
  - ....
- مدیر شبکه
- ....

در ادامه درس منظور ازکاربر، استفاده کننده یا End User است

عناصر محیط پایگاهی – (۴) کاربر (انواع)

- **انواع کاربر از نظر اسلوب عملیاتی:**
  - **Batch- یکجا** (تعدادی برنامه یا پرس‌وجو جمع آوری می‌شود و به صورت یکجا به سیستم داده می‌شود و جواب آن بر می‌گردد.)
  - **Online – برخط – پیوسته** (یک برنامه یا پرس‌وجو به سیستم داده می‌شود، اجرا می‌شود، و جوابش برمی‌گردد.)
  - **Interactive-تعاملی –** بسته به اینکه چه جوابی داده شود عمل دیگری از کامپیوتر درخواست می‌شود.
    - Online لزوما Interactive نیست اما Interactive لزوما Online است.

- سیستم پایگاهی به صورت پیش فرض چند کاربره (multi-user) است.

عناصر محیط پایگاهی – (۴) داده

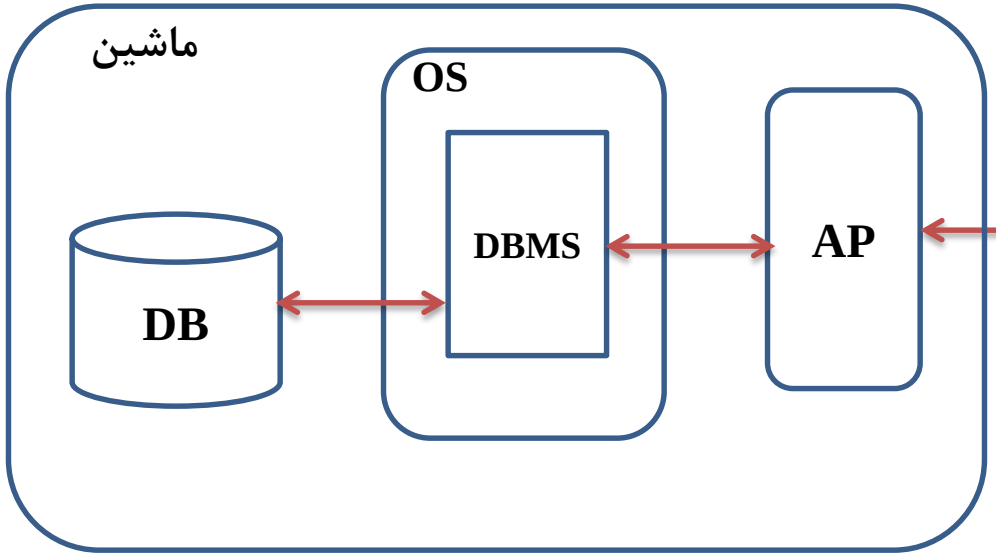
- **داده‌های ذخیره شده** در یک سیستم پایگاهی عبارتند از:
  - داده‌های کاربران
  - داده‌های سیستمی
- مباحث مرتبط با داده در محیط پایگاهی در ادامه درس مطرح می‌گردد.

انواع معماری سیستم پایگاهی

- **سوال:** می‌خواهیم یک سیستم کاربردی پایگاهی ایجاد کنیم.بر اساس کدام معماری ایجادکنیم؟
- **در توصیف معماری یک سیستم باید مشخص کنیم که**
  - از چه مولفه‌هایی، از هر مولفه چند عدد و با چه کیفیتی تشکیل شده است،
  - مولفه‌ها چگونه با هم ترکیب شده‌اند (جنبه ساختاری سیستم)،
  - مولفه‌ها چگونه با یکدیگر در تعامل هستند (جنبه رفتاری سیستم).
- **انواع معماری سیستم پایگاهی:**
  - معماری متمرکز
  - معماری نامتمرکز
    - معماری مشتری-خدمتگزار
    - معماری توزیع‌شده
    - معماری چندپایگاهی
    - معماری با پردازش موازی

معماری متمرکز

- در این معماری یک پایگاه داده (متمرکز و مجتمع) روی یک سیستم کامپیوتری و بدون ارتباط با سیستم کامپیوتری دیگر ایجاد می‌شود.
- معمولا به صورت تک کاربری و برای کاربردهای کوچک و با امکانات محدود از این معماری استفاده می‌شود.



معماری مشتری– خدمتگزار

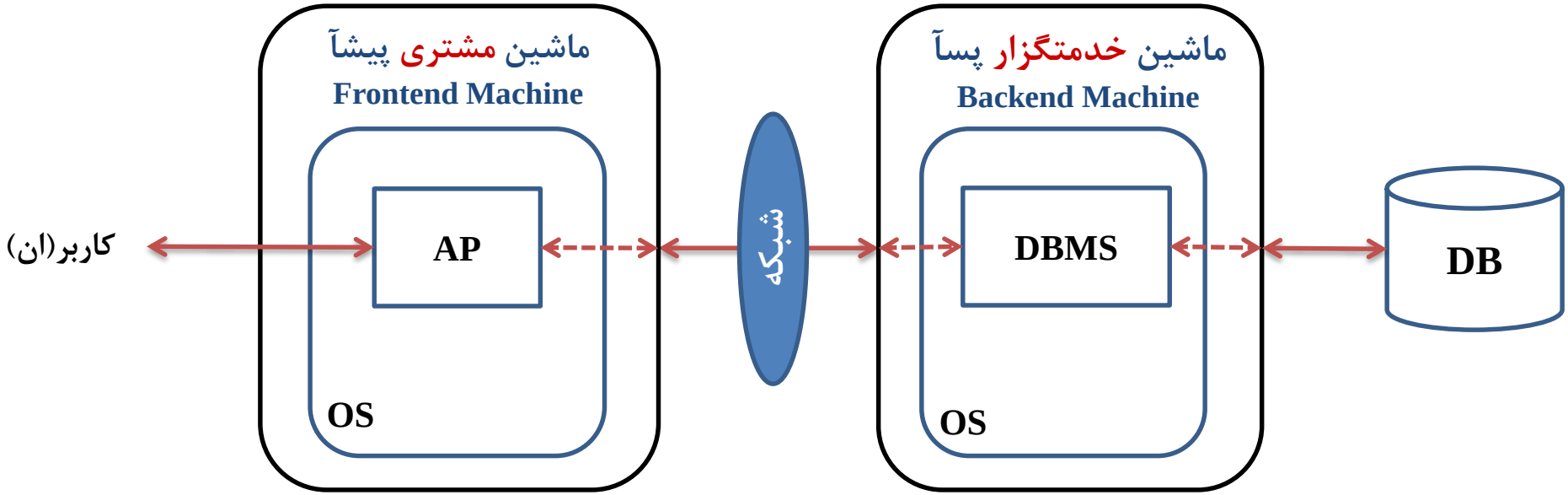
- **دلیل:** دلیل اصلی استفاده از معماری مشتری-خدمتگزار (Client-Server): تقسیم وظایف سیستم
- **تعریف:** هر ماشینی (فیزیکی یا منطقی) که خدمتی را به ماشین دیگر بدهد، **خدمتگزار** نامیده می‌شود.

نمونه‌هایی از انواع خدمتگزارها: File Server، Print Server، Message Server، DB Server

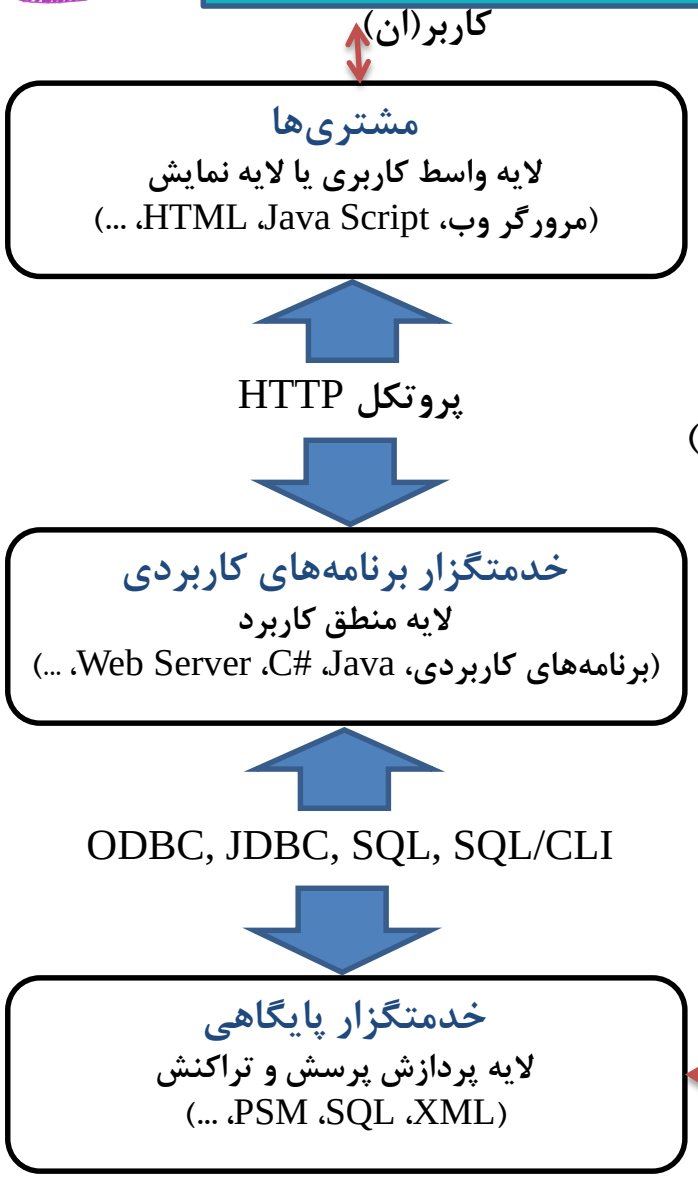
- **انواع معماری مشتری- خدمتگزار**
  - معماری تک مشتری- تک خدمتگزار
  - معماری چند مشتری- تک خدمتگزار
  - معماری تک مشتری- چند خدمتگزار
  - معماری چند مشتری- چند خدمتگزار

معماری مشتری –خدمتگزار دو لایه

- معمولا شامل دو سایت:
  - **سایت مشتری:** تمام برنامه‌های کاربردی در آن اجرا می‌شوند.
  - **سایت خدمتگزار:** تمام داده‌ها در آن ذخیره می‌شوند
- به این معماری، **معماری دولایه** (2-tier) نیز گویند.



معماری مشتری– خدمتگزار سه لایه

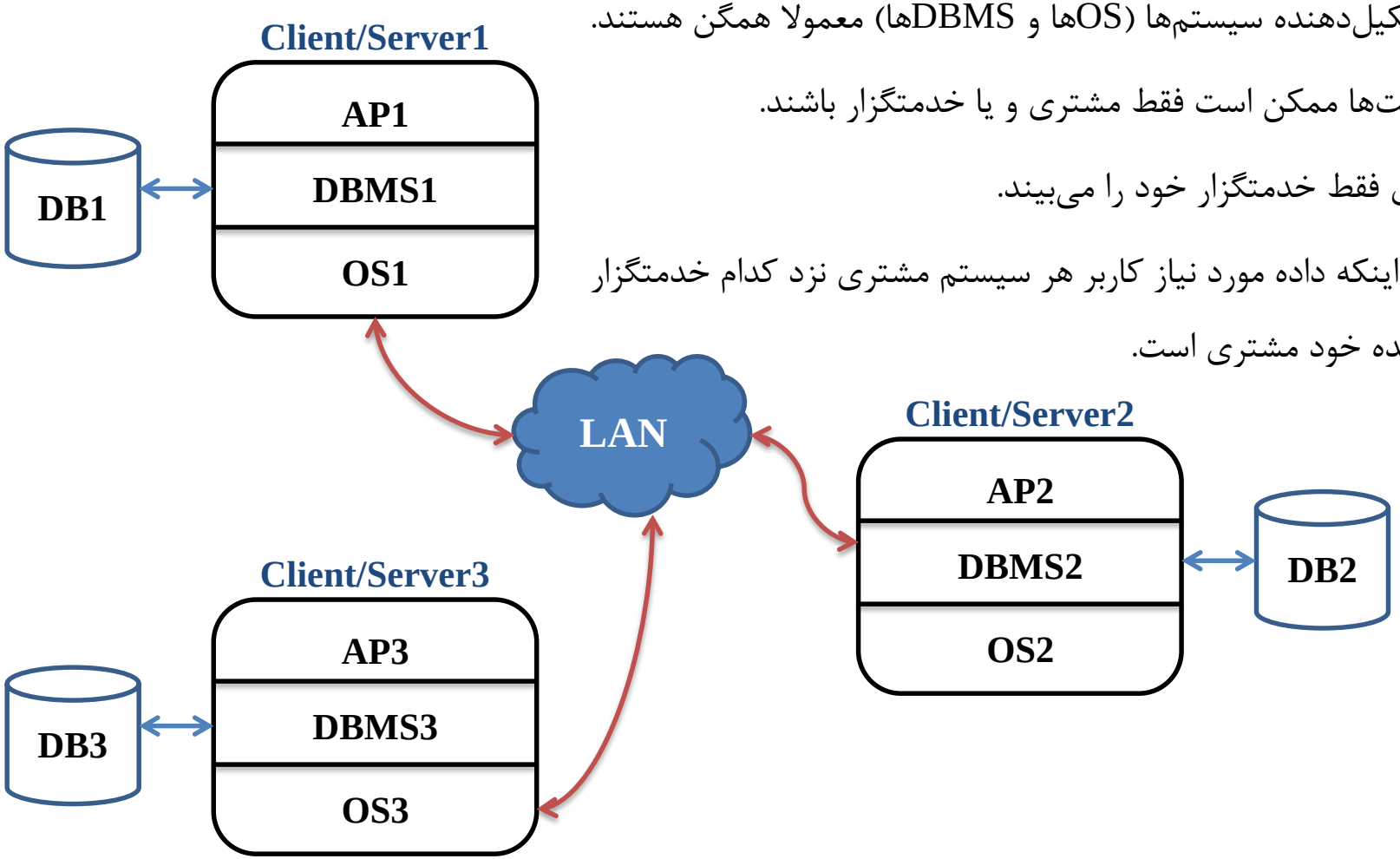


ماشین‌های ساده، ارزان و حتی بدون دیسک (thin client)

- **برخی مزایای معماری سه لایه نسبت به دولایه:**
  - گسترش‌پذیری بهتر
  - کارایی بالاتر
  - امنیت داده‌ای بیشتر (عدم ارتباط مستقیم مشتری‌ها با کارگزار داده)
  - قابلیت کاهش هزینه سخت افزاری (با استفاده از thin client)

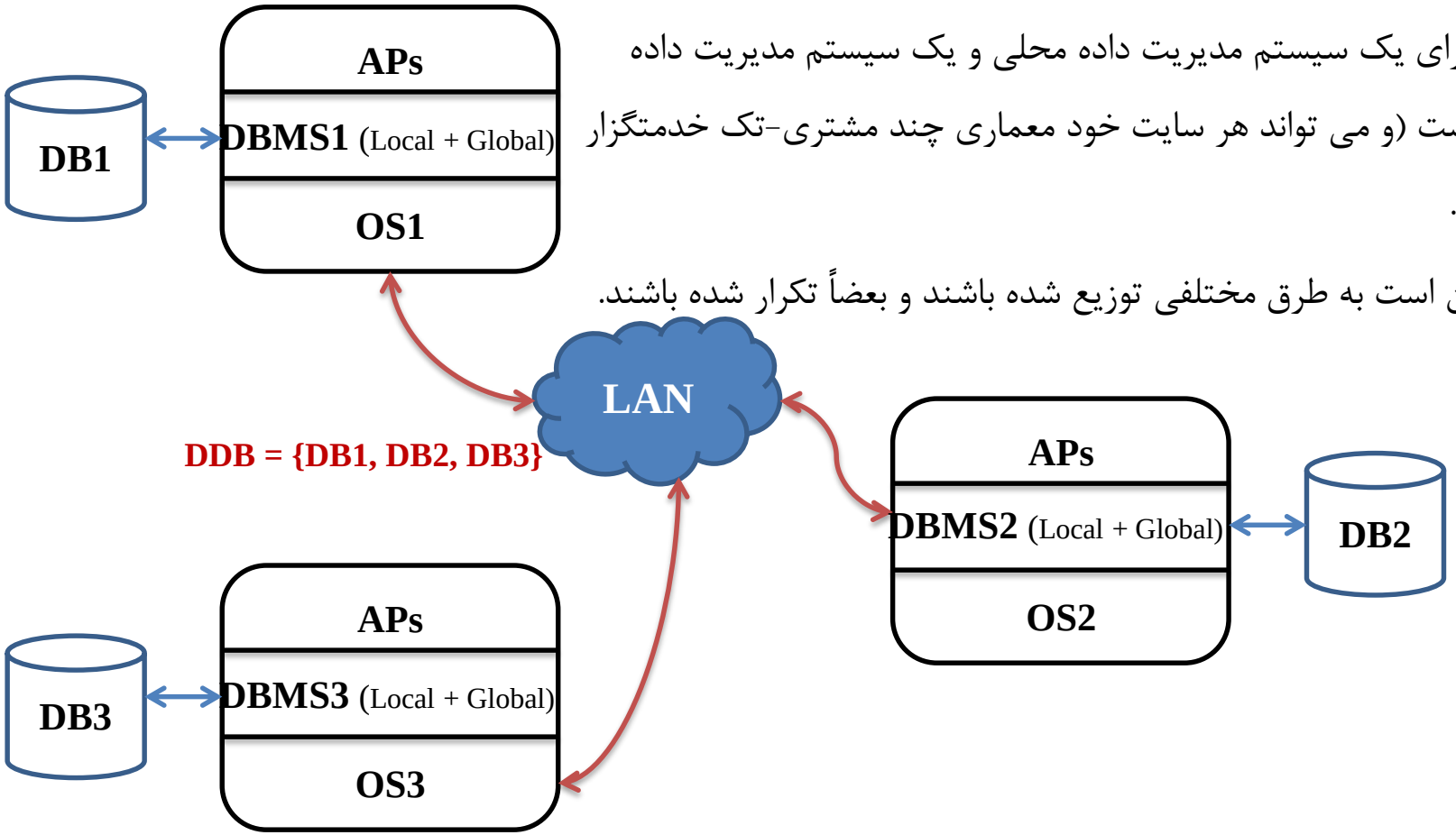
معماری چند مشتری–چند خدمتگزار

- سیستم‌های پایگاهی همزمان یا ناهمزمان ایجاد می‌شوند.
- اجزای تشکیل‌دهنده سیستم‌ها (OSها و DBMSها) معمولا همگن هستند.
- برخی سایت‌ها ممکن است فقط مشتری و یا خدمتگزار باشند.
- هر مشتری فقط خدمتگزار خود را می‌بیند.
- مسئولیت اینکه داده مورد نیاز کاربر هر سیستم مشتری نزد کدام خدمتگزار است برعهده خود مشتری است.



معماری توزیع‌شده

- مجموعه‌ای است از چند پایگاه داده منطقاً یکپارچه (مجتمع)، ولی به طور فیزیکی توزیع شده روی یک شبکه کامپیوتری.
- توزیع شدگی از دید برنامه‌ها و کاربران پایگاه داده پنهان است.
- هر سایت دارای یک سیستم مدیریت داده محلی و یک سیستم مدیریت داده توزیع‌شده است (و می تواند هر سایت خود معماری چند مشتری-تک خدمتگزار داشته باشد).
- داده‌ها ممکن است به طرق مختلفی توزیع شده باشند و بعضاً تکرار شده باشند.



گوگل از بیش از یک میلیون سرور در قالب بیش از ۱۵ مرکز داده در سه قاره بهره می گیرد.

Google Data Centers

G+1 280

Search this site

Data centers › Inside look › Locations

Data center locations

We own and operate data centers around the world to keep our products running 24 hours a day, 7 days a week. Find out more about our data center locations, community involvement, and job opportunities in our locations around the world.

Americas

Berkeley County, South Carolina
Council Bluffs, Iowa
Douglas County, Georgia
Jackson County, Alabama
Lenoir, North Carolina
Mayes County, Oklahoma
Montgomery County, Tennessee
Quilicura, Chile
The Dalles, Oregon

Asia

Changhua County, Taiwan
Singapore

Europe

Dublin, Ireland
Eemshaven, Netherlands
Hamina, Finland
St Ghislain, Belgium

تراکنش Transaction:

- دنباله ای از عملیات («قطعه برنامه») که معمولاً حد اقل یک عمل تغییردهنده (درج، حذف، به روزرسانی) در محیط ذخیره‌سازی داده‌ها انجام می‌دهد و باید یا به تمامی اجرا شود و یا اجرا نشده تلقی شود.
- **تراکنش ها می توانند به صورت همروند در سیستم اجرا شوند.**

یک مثال بانکی: ۵۰ هزار تومان از حساب A برداشت و به حساب B واریز کن.

```
BEGIN TRANS
        READ (A)
        A = A – 50
        UPDATE (A)
        READ (B)
        B = B + 50
        UPDATE (B)
END TRANS
```

- **دو مشکل اساسی که ممکن است اجرای تراکنشها را با چالش روبرو کند:**
  - **بروز انواع خطا در سیستم (سخت افزاری، نرم افزاری و ...) که منجر به عدم اجرای کامل دستورات شود**
  - **اجرای همزمان چندین تراکنش مرتبط**

تراکنش

- یک تراکنش می بایست دارای خواص ACID باشد که توسط DBMS تضمین می شود:

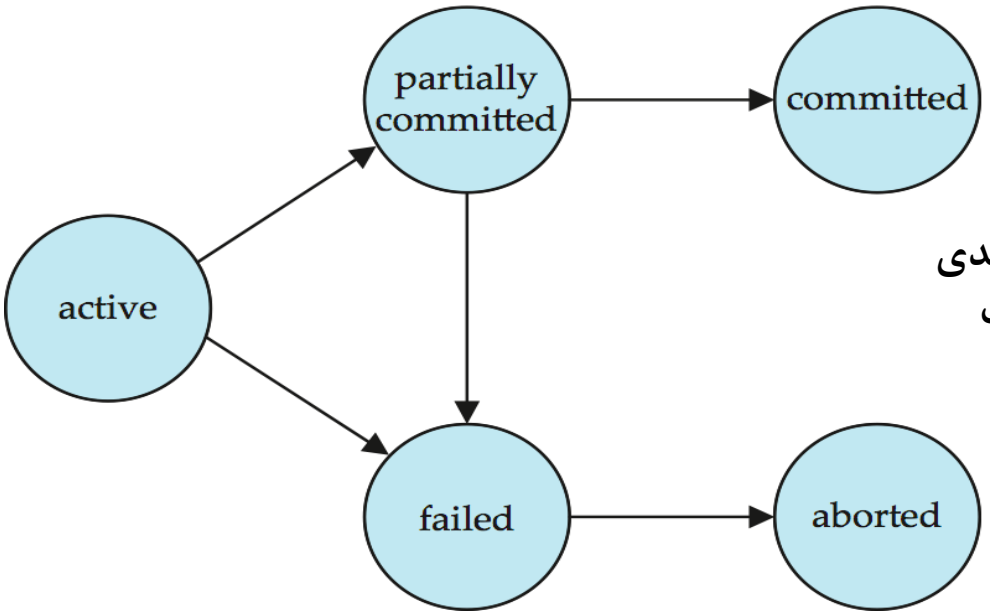
ACID : **A**tomicity, **C**onsistency, **I**solation, **D**urability

- **A**tomicity: یا همه یا هیچ
- **C**onsistency: سازگاری
- **I**solation: انفراد و جدایی
- **D**urability: دوام (پایداری)

مثال

شرط سازگاری پایگاه داده در این مثال : A+B ثابت باشد

```
BEGIN TRANS
        READ (A)
        A = A – 50
        UPDATE (A)
        READ (B)
        B = B + 50
        UPDATE (B)
END TRANS
```



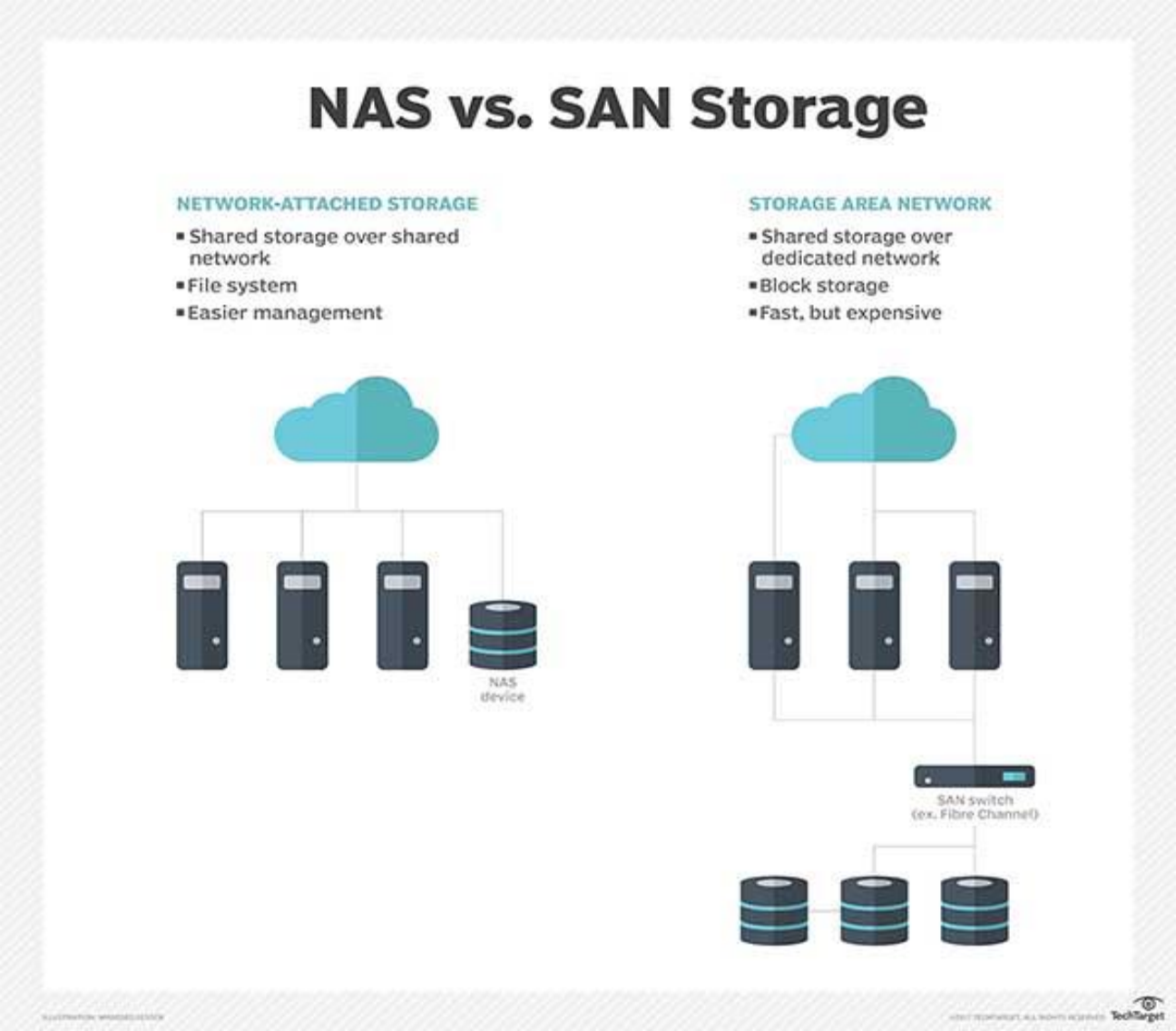
یک دسته بندی از حالات یک تراکنش

پرسش و پاسخ . . .

ایمیل : zarepour@iust.ac.ir

ارتباط حضوری: ساعت مشخص شده در برنامه هفتگی به عنوان رفع اشکال دانشجویی (روزهای شنبه و دوشنبه ساعت ۹:۳۰ تا ۱۱ صبح)

www.ezarepour.ir


NAS vs. SAN Storage
NETWORK-ATTACHED STORAGE
Shared storage over shared network
File system
Easier management
NAS device
STORAGE AREA NETWORK
Shared storage over dedicated network
Block storage
Fast, but expensive
SAN switch
(ex. Fibre Channel)
TechTarget