

به نام آنکه جان را فکرت آموخت



بخش اول : مقدمه‌ای بر سیستم‌های پایگاهی

مرتضی امینی

نیمسال اول ۱۴۰۱-۱۴۰۰

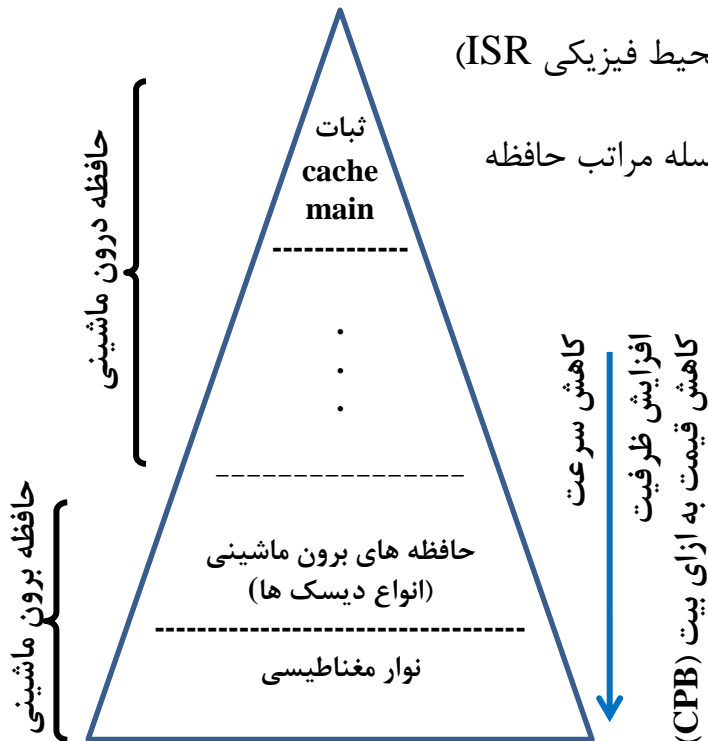
(محتویات اسلایدها برگرفته از یادداشت‌های کلاسی استاد محمدتقی روحانی رانکوهی است.)

هر سیستم نرم‌افزاری از مجموعه‌ای از داده‌های ذخیره شده ممکن است استفاده کند.

فرمت ثابت و از پیش تعیین شده دارد. well-formatted است،
 ساختیافته (structured)
 نیم ساختیافته (semi-structured)
 ناساختیافته (un-structured)

به لحاظ ساختاری

در قالب تعدادی فایل (محیط فیزیکی ISR)
 در کجا؟ ← در یک سلسله مراتب حافظه



آیا نیاز به تحمیل یک ساختار در اینها داریم؟ آیا واقعا داده ناساختیافته داریم؟

انواع سیستم نرم‌افزاری:

بنیادی یا پایه (سیستم‌های عامل)

نیمه بنیادی (DBMS، DMS، کامپایلرها، اسمبلرها، و ...)

Data Management System

کاربردی (برنامه‌های کاربردی)

Database Management System

ابزاری: انواع toolهای خاص منظوره

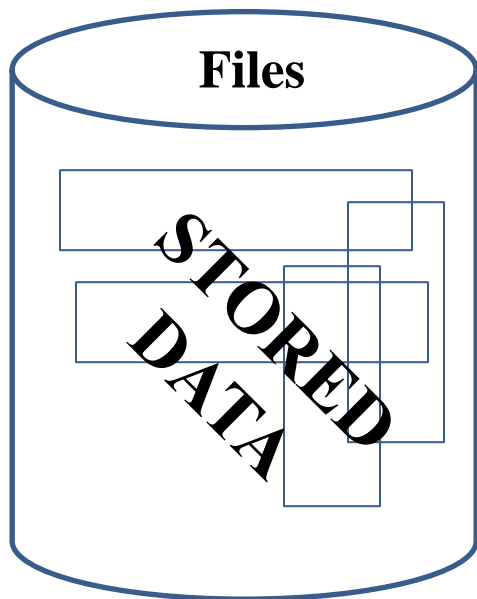


کنجکاوی: دلایل استفاده از این سلسله مراتب حافظه چیست؟

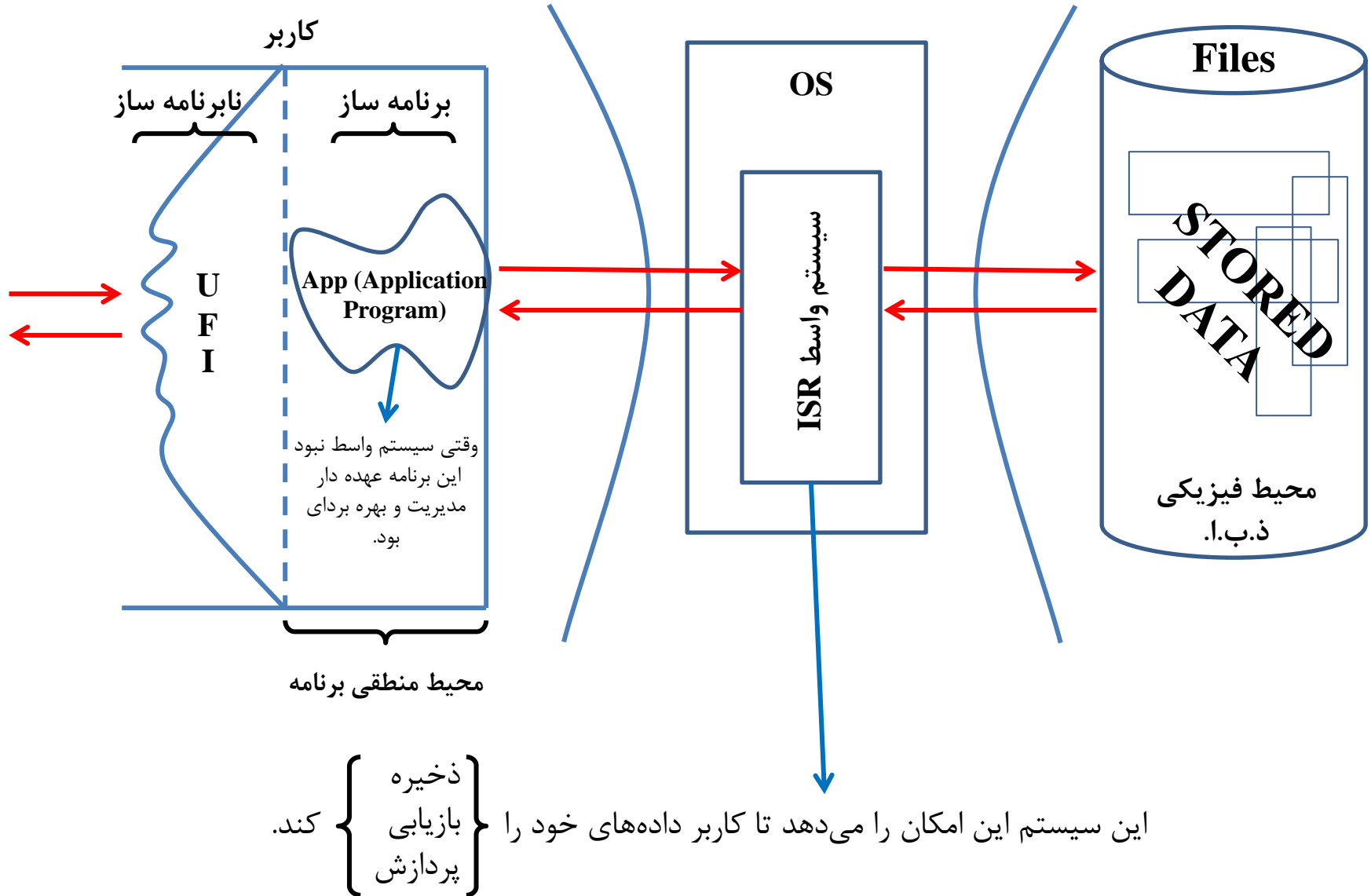
کنجکاوی: چه داده ای، برای چه مدتی، در کدامیک از مراتب سلسله مذکور قرار می گیرد؟

کنجکاوی: خصوصیات عمومی فایل ها چیست؟

□ 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.

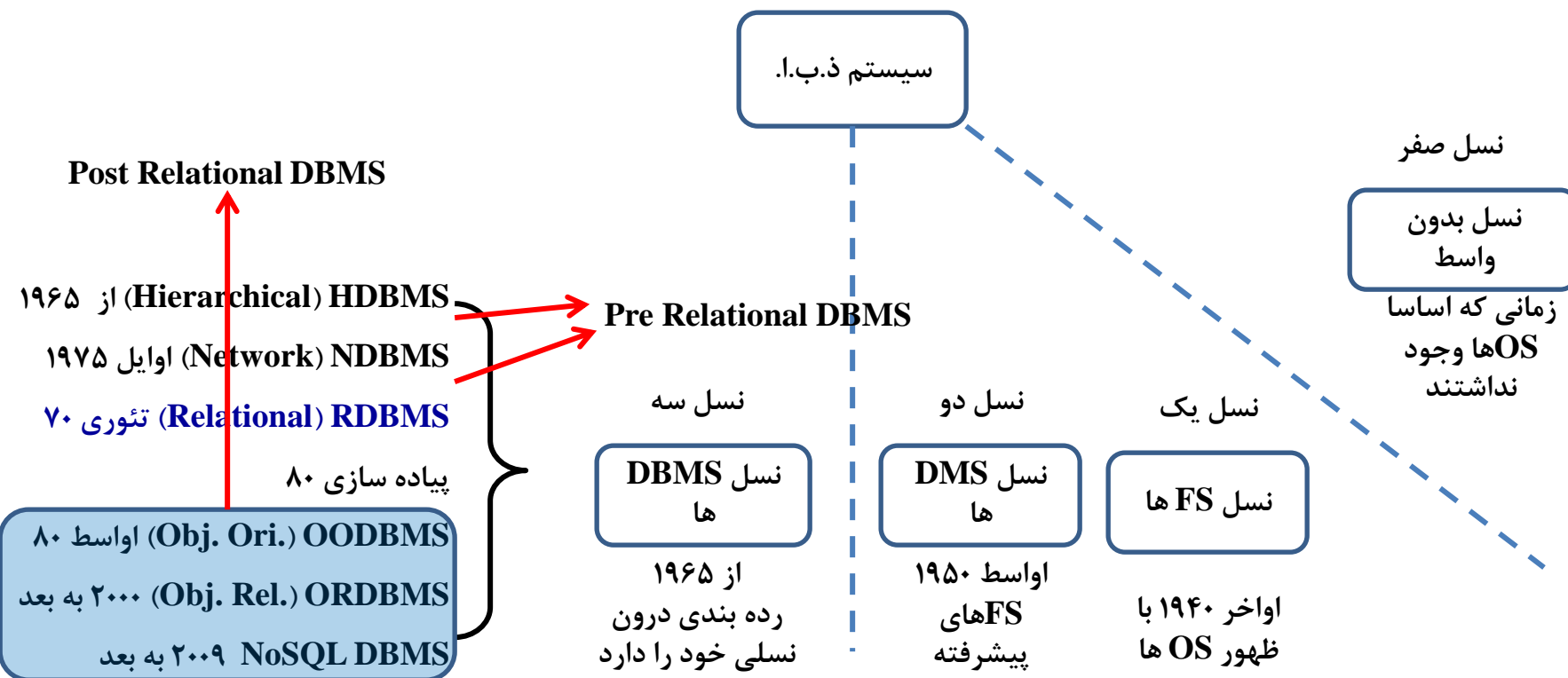


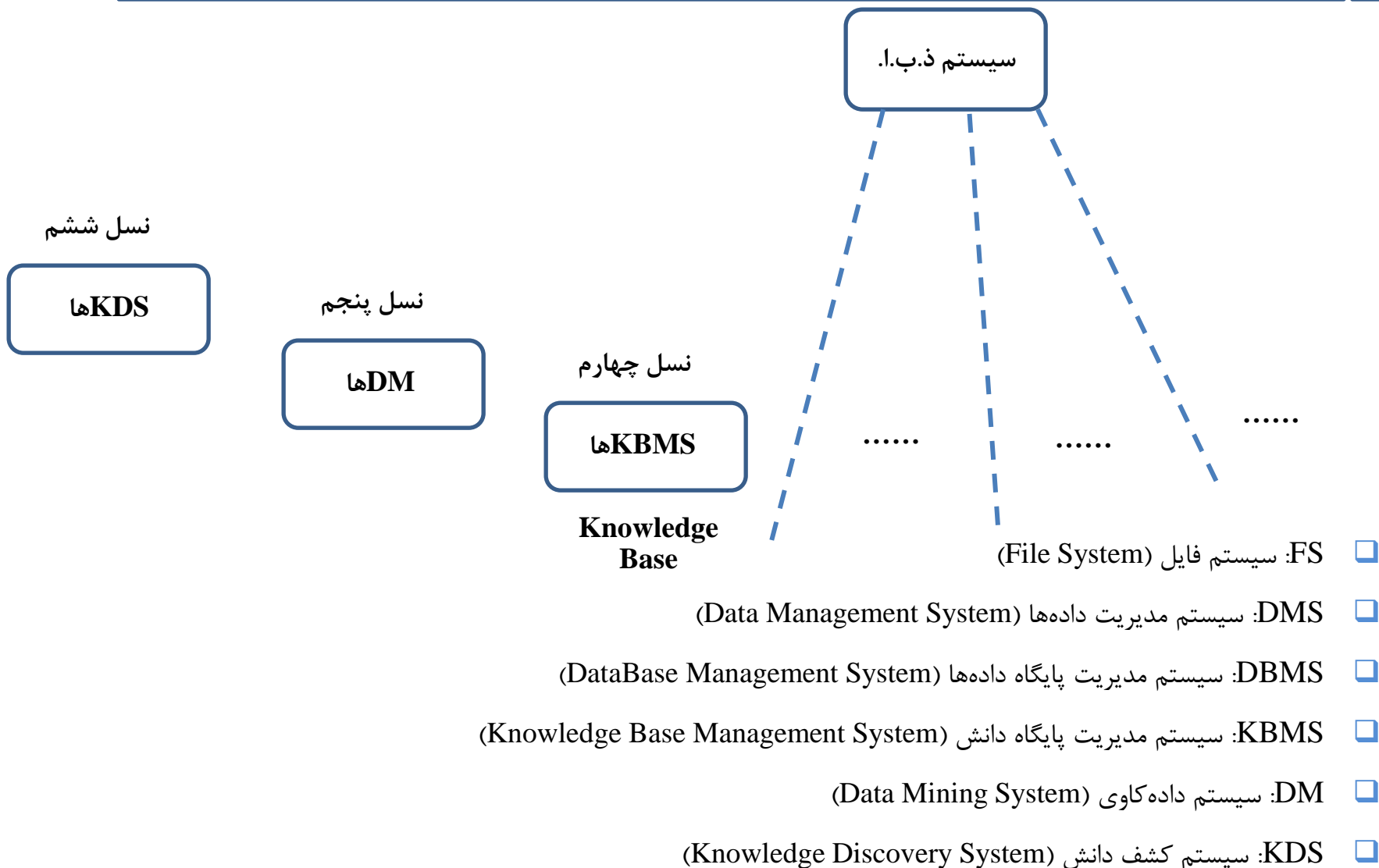


کنجکاوی: رده بندی از مفهوم کاربر ارایه کنید؟ به بیان دیگر گونه های دیگر کاربر کدامند؟

□ سیستم واسط "ISR" سیر تحول خاص خود را دارد :

□ ۶ نسل تکنولوژیک قابل بازیابی است (به طور کلی) [دیدگاه نرم‌افزاری]







□ در این نسل بندی، نسل بعدی نسل قبلی را منسوخ نمی کند. نسل بعدی نسل قبلی را تکمیل می کند و از آن استفاده می کند.

□ انواع نیازهای پردازشی، کنترلی، و عملیاتی سبب ایجاد نسل های سیستم «ذ.ب.ا.» شد.



داده (Data) □

□ تعریف اول ANSI: نمایش داده‌ها، پدیده‌ها، مفاهیم یا شناخته‌ها به طرزى 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)

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.



دانش (Knowledge) □

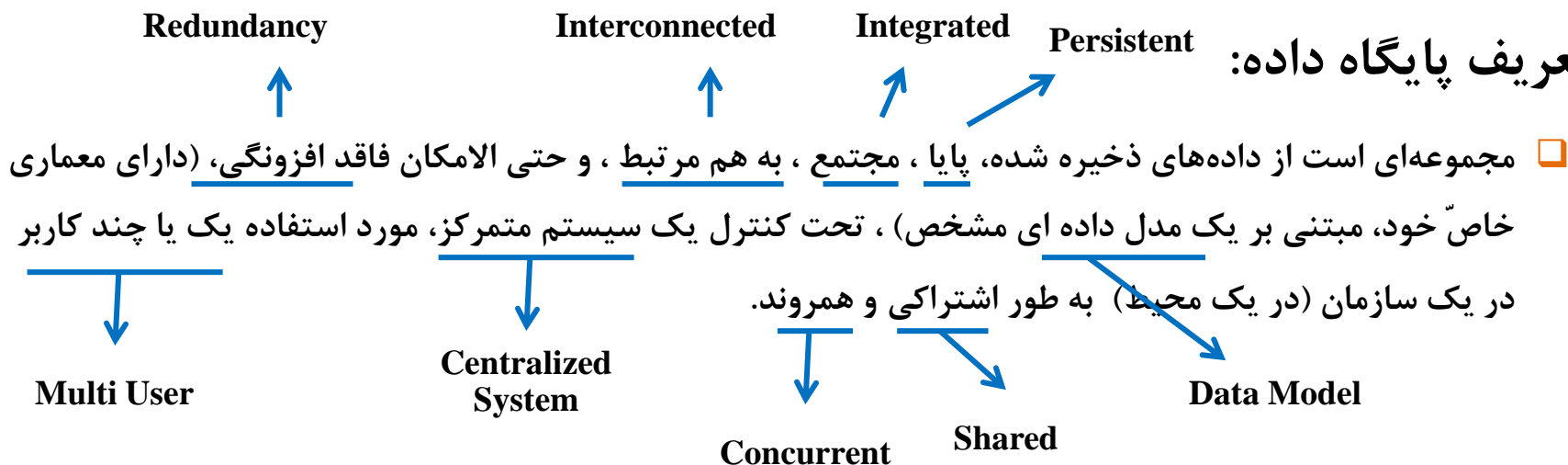
□ **تعریف [FROS87]:** دانش عبارت است از نمایش نمادین جنبه‌هایی از بخشی از جهان واقع (جهان موردنظر یا محیط مطرح)

▪ 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، معمولاً با روش استقراء حاصل می‌شود. حصول این شناخت می‌تواند توسط انسان یا یک سیستم خودکار انجام شود.



تعریف پایگاه داده: □



□ مثال کاربردی

□ محیط عملیاتی: دانشگاه



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

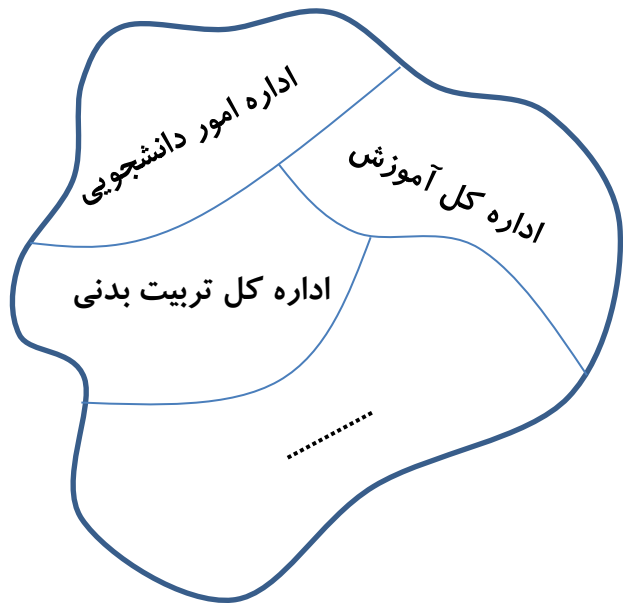


Micro Real World (خرد جهان واقع)
Mini World
Universe of Discourse (جهان مطرح)

□ نکته: هر محیط از تعدادی زیر محیط تشکیل شده است.

□ در هر محیط مجموعه‌ای از **نوع موجودیت‌ها** وجود دارند که نیازهای داده‌ای کاربران ناظر به آنهاست (یعنی به پردازی } داده‌ای

داده‌هایی در مورد آنها نیاز دارند).





❑ **نکته:** زیرمحیط های یک محیط معمولا با هم اشتراک دارند در نوع موجودیت ها (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 محیط ها

❑ برای این منظور در اساس دو مَشی-روش (approach) وجود دارد. $\left. \begin{array}{l} \text{مشی فایلینگ [سنتی یا کلاسیک] یا ناپایگاهی} \\ \text{مشی پایگاهی Database Approach} \end{array} \right\}$

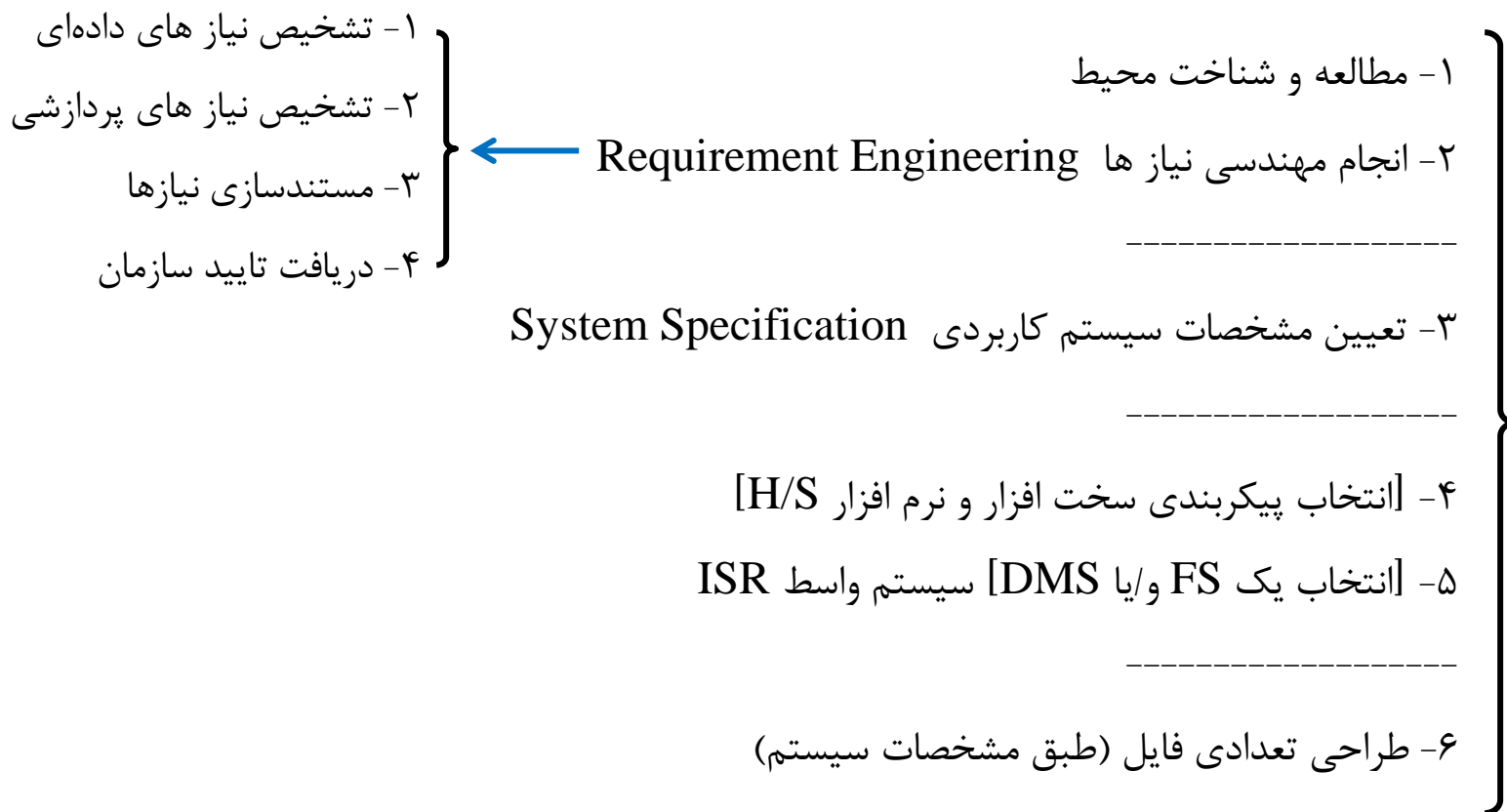
↓
یعنی ممکن است مشی های بینابینی نیز وجود داشته باشد.



کارهای لازم در مشی فایلینگ به طور خلاصه:

توجه: این کارها معمولاً برای هر زیرمحیط به طور جداگانه انجام می شود. ← تعدادی سیستم کاربردی جدا

(نامجتمع) و بی ارتباط در یک محیط ...





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 فایل ها + کاربران



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

۷- طراحی واسطه‌های کاربری (UFI)

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

۹- تولید برنامه‌های { ایجاد
کنترل
پردازش } فایل‌ها

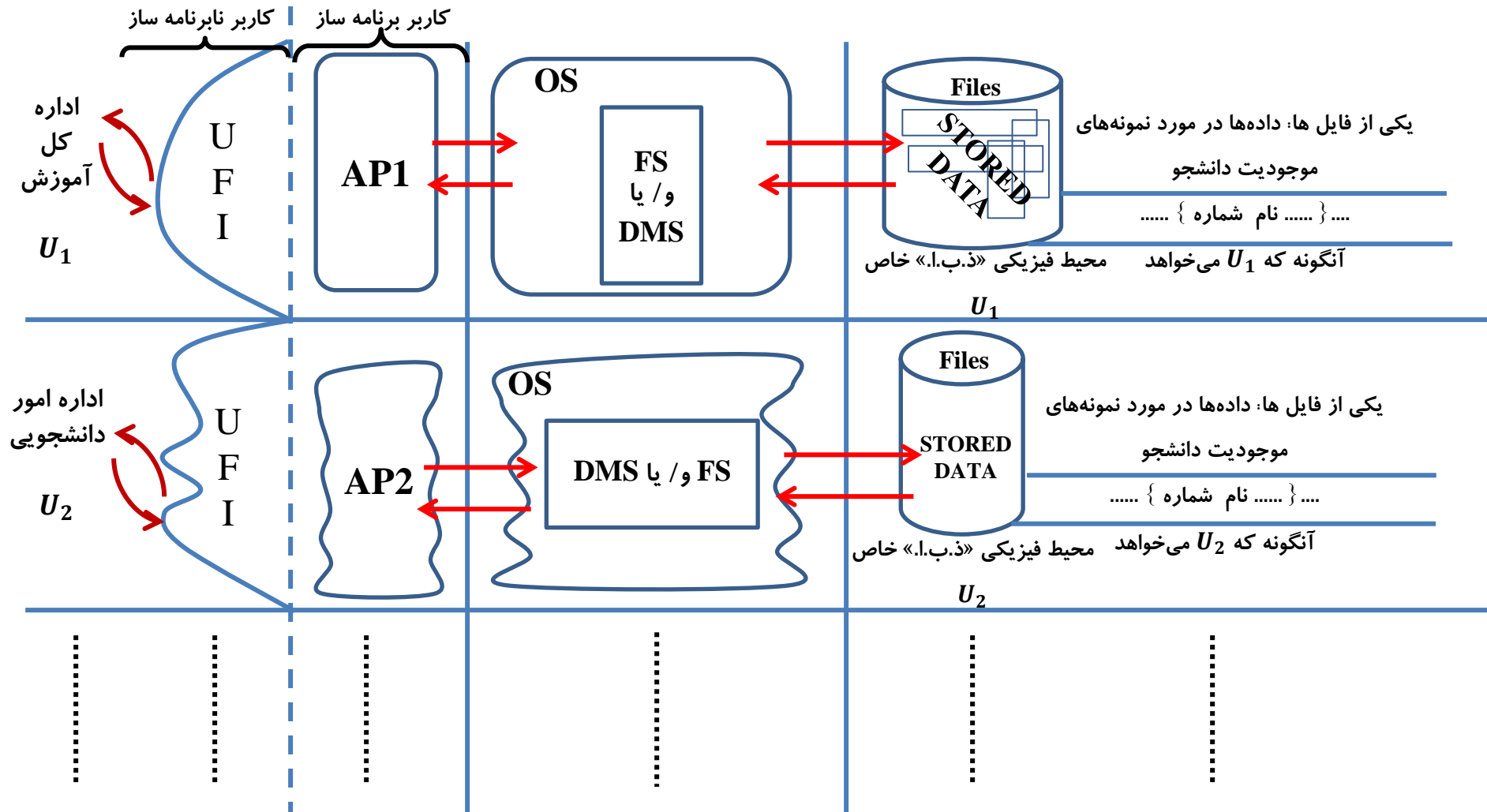
۱۰- ایجاد محیط فیزیکی «ذ.ب.ا.» و پیکربندی

۱۱- تست با داده‌های آزمایشی و واقعی

۱۲- ورود داده‌های اولیه (Data Entry)

۱۳- آغاز بهره‌برداری و نگهداری سیستم

۱۴- رفع معایب و بهینه‌سازی سیستم





❑ برخی از معایب مشی فایلینگ:

❑ وجود سیستم های نامجتمع در يك سازمان [محيط] و نامرتبط به هم

❑ عدم وجود يك سیستم کنترل متمرکز روی كل داده های سازمان

❑ وجود افزونگی زیاد

❑ خطر بروز ناسازگاری داده ها (Data Inconsistency) ← **کنجکاوی:** جنبه های بروز ناسازگاری کدامند؟

❑ عدم امکان اعمال ضوابط حفظ امنیت داده ها (Data Security)

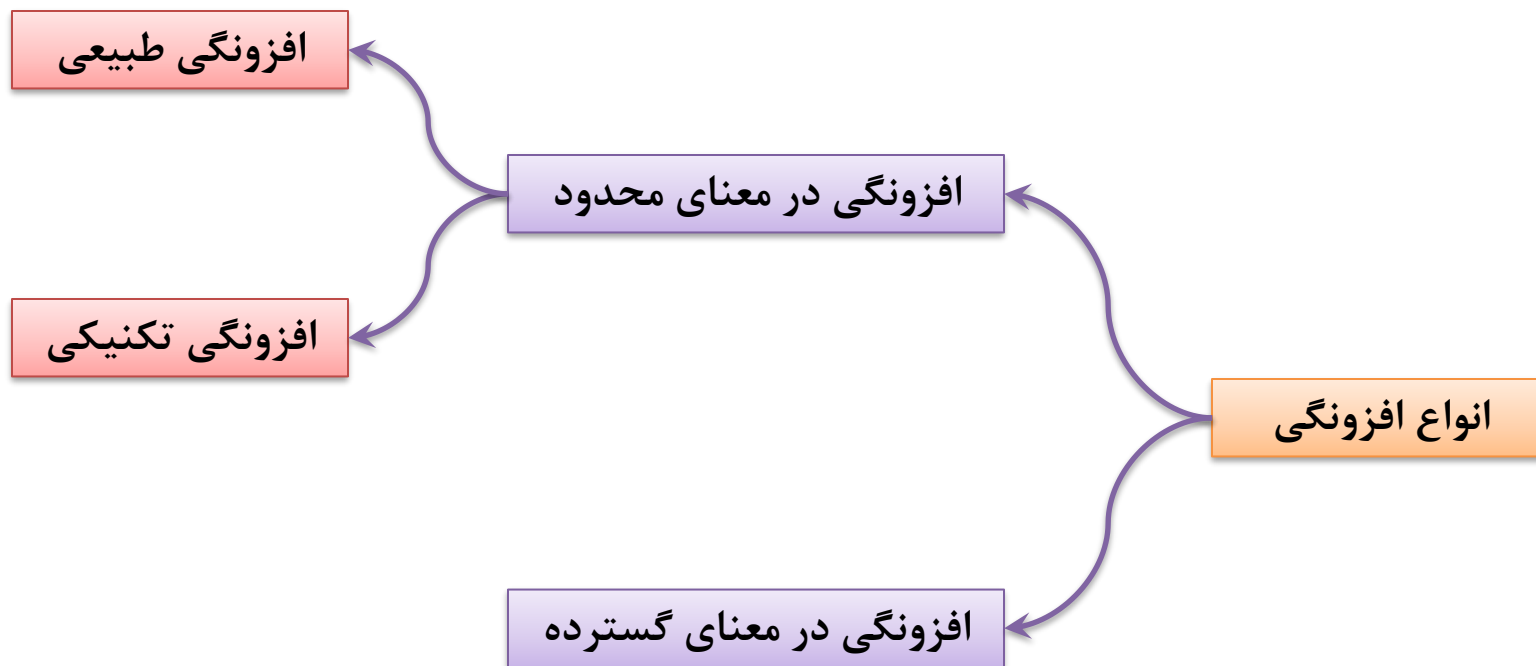
❑ عدم امکان اشتراکی شدن داده ها (Data Sharing) [یا در حداقل و یا با دشواری]

❑ مصرف نابهبینه سخت افزار (به ویژه سخت افزار ذخیره ساز)

❑ وابسته بودن برنامه ها به جنبه های فایلینگ محیط ذخیره سازی، به گونه ای که اگر قرار باشد در فایلینگ

تغییراتی ایجاد شود، برنامه ها هم متناسباً باید تغییر یابد. (به طور مثال فرمت ساختار یا نحوه دسترسی

(Access Strategy) را تغییر دهیم)





افزونگی در معنای محدود ☐

☐ عبارت است از تکرار ذخیره سازی مقادیر (value) یک صفت یا بیش از یک صفت یک نوع موجودیت 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)

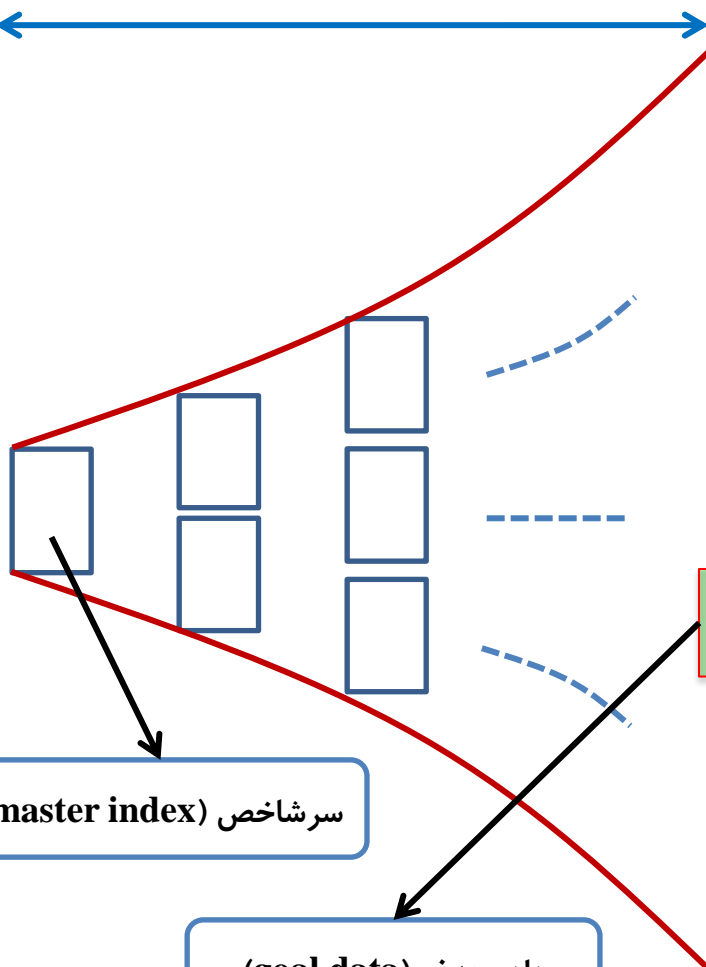
▪ **کنجکاوی**: برای کاهش مصرف حافظه در حالت افزونگی طبیعی چه باید کرد؟

۲- **تکنیکی**: ناشی از استفاده از یک تکنیک معمولا برای افزایش سرعت (مثل نمایه سازی [شاخص بندی
[Indexing])



(چندسطحی معمولاً با ساختار B-Tree یا B⁺-Tree)

مثالی از افزونگی تکنیکی: نمایه سازی



			شماره	نام	رشته	...
۱۰۰	۱	→	۱۰۰		نرم افزار	
۱۰۱	۲	→	۱۰۱		نرم افزار	
۱۰۲	۳	→	۱۰۲		سخت افزار	
۱۰۳	۴	→	۱۰۳		نرم افزار	
۱۰۴	۵	→	۱۰۴		سخت افزار	
۱۰۵	۶	→	۱۰۵		نرم افزار	
⋮			⋮			
k			k		نرم افزار	
⋮			⋮			
۹۹۷	۸۹۸	→	۹۹۷		سخت افزار	
۹۹۸	۸۹۹	→	۹۹۸		سخت افزار	
۹۹۹	۹۰۰	→	۹۹۹		نرم افزار	

افزونگی در معنای گسترده

عبارت است از تکرار ذخیره‌سازی داده‌ها در مورد نمونه‌های یک یا بیش از یک نوع موجودیت از یک محیط (تحت کنترل سیستم‌های مختلف).

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.

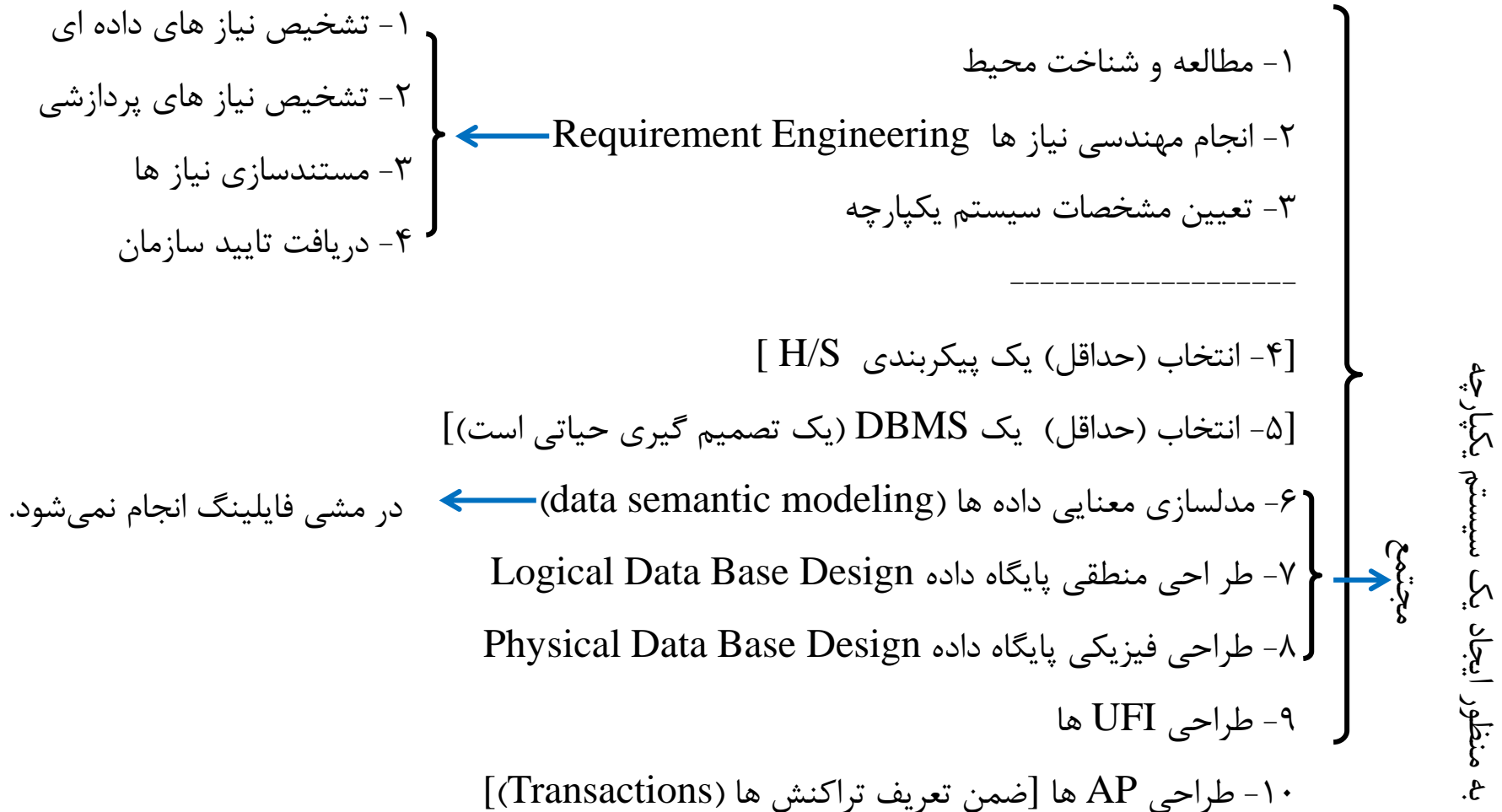
به طور مثال تکرار اطلاعات دانشجویان در دو زیرسیستم اداره کل آموزش و زیرسیستم اداره امور دانشجویی.

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

دلایل بروز افزونگی در سیستم های ISR به ویژه سیستم های پایگاهی کدامند؟



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



ادامه: ...

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

۱۱- تولید برنامه‌های تعریف (ایجاد) و کنترل DB

۱۲- تولید برنامه‌های عملیات در داده‌ها (پردازش داده‌ها)

۱۳- ایجاد محیط فیزیکی «ذ.ب.ا.» و پیکربندی

۱۴- تست با داده‌های آزمایشی و واقعی

به طور مثال به منظور افزایش کارایی

۱۵- تنظیم سیستم پایگاهی (Data Base System Tuning)

۱۶- آغاز بهره برداری و نگهداری از سیستم

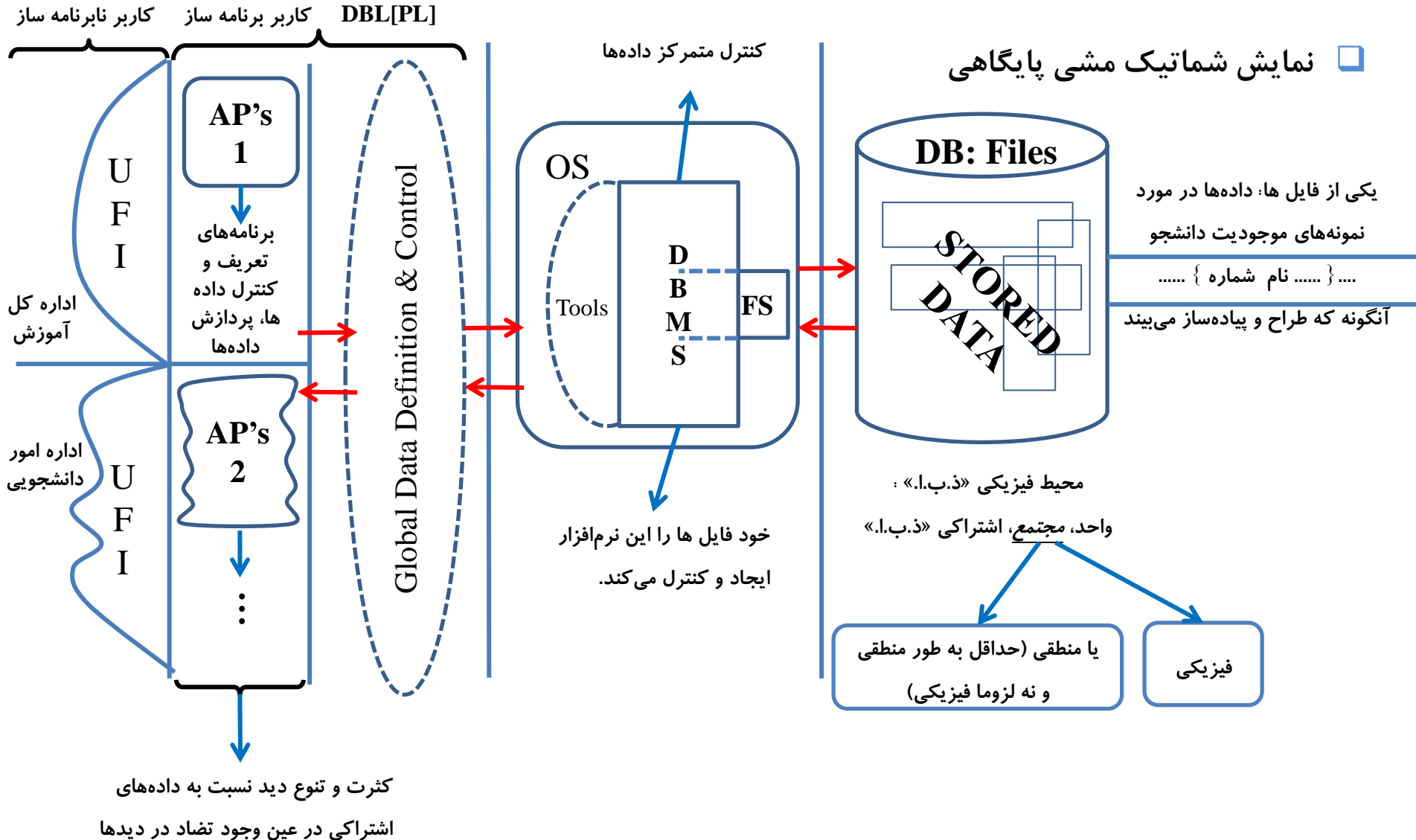
یکی از ویژگی‌های DBMS گسترش پذیری سیستم است.

۱۷- در صورت لزوم، گسترش سیستم

۱۸- رفع معایب و بهینه‌سازی سیستم



فصل اول - مقدمه





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

۲- معماری پایگاه داده

تمرین: مزایای مشی پایگاهی چیست؟ ← (طبق معلومات فعلی: عکس معایب مشی فایلینگ)

تمرین: چند سطح تعریف داده داریم؟



تراکنش Transaction:



□ دنباله ای از عملیات (در قالب یک قطعه برنامه) که معمولاً حداقل یک عمل تغییردهنده (درج، حذف، به روزرسانی) در محیط ذخیره سازی داده ها انجام می دهد و باید یا به تمامی اجرا شود و یا اجرا نشده تلقی شود.

□ دارای خواص ACID (Atomicity Consistency Isolation Durability)

دوام (پایداری) انفراد و جدایی سازگاری یا همه یا هیچ

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



❑ **اتمیک بودن (Atomicity):** یا کل تراکنش اجرا شود یا هیچ بخشی از آن اجرا نشود.

❑ اگر بخشی از یک تراکنش با شکست مواجه شود کل تراکنش شکست خورده تلقی شود و حالت پایگاه داده تغییری نکند.

❑ **سازگاری (Consistency):** هر تراکنش پایگاه داده را از یک حالت سازگار یا معتبر (منطبق با محدودیت‌های جامعیتی تعریف شده) به یک حالت سازگار یا معتبر دیگر ببرد.

❑ معتبر بودن حالات توسط تعدادی محدودیت جامعیتی تعریف می‌شود. مثلاً مانده حساب برابر مجموع واریزها منهای مجموع برداشتها است.

❑ **جدا بودن (Isolation):** اگر چند تراکنش به صورت هم‌روند اجرا شوند، حالت پایگاه داده مشابه حالتی باشد که این تراکنشها به صورت سریال (پشت سر هم) اجرا شده باشند.

❑ **مانایی (Durability):** هر گاه که یک تراکنش خاتمه یافته اعلام شود، اثر آن بر روی پایگاه داده مانا باقی بماند (حتی اگر سیستم دچار اختلال شود یا برق آن قطع شود یا هر مشکل دیگری پیش آید).



□ عناصر اصلی محیط پایگاهی:

- ۱- سخت افزار ←
 - ذخیره سازی
 - پردازشگر
 - ارتباطی (همرسانی) Data Communication
- ۲- نرم افزار
- ۳- کاربر
- ۴- داده



□ سخت افزار ذخیره سازی: {
- رسانه اصلی: دیسک ترجیحا با تکنولوژی RAID
(Redundant Array of Inexpensive Disk)

- رسانه فرعی: نوار مغناطیسی [از جمله برای تولید نسخه های پشتیبان]

□ اغلب DBMS های امروزی تکنیک های تولید Back up را دارا هستند.

تکنیک های تولید نسخه
پشتیبان؟



سطوح مختلف Back up

□ سخت افزار پردازشگر: {
- کامپیوترهای معمولی از هر رده [PC, main,...]

- اما ماشین های خاص DB هم داریم : DB Machines

□ سخت افزار ارتباطی (همرسانی): {
- امکانات محلی: برای ارتباط دستگاه های جانبی با پردازنده

- امکانات شبکه ای: برای ایجاد شبکه در سیستم پایگاهی نامتمرکز



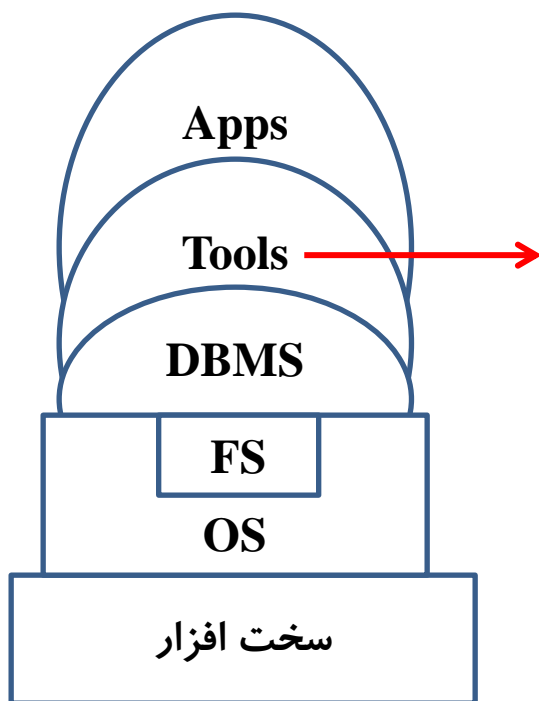
□ انواع نرم افزارهای مطرح در محیط پایگاهی:

□ سیستم عامل و سیستم فایل (OS و FS)

□ سیستم مدیریت پایگاه داده‌ها (DBMS)

□ ابزارها (Tools)

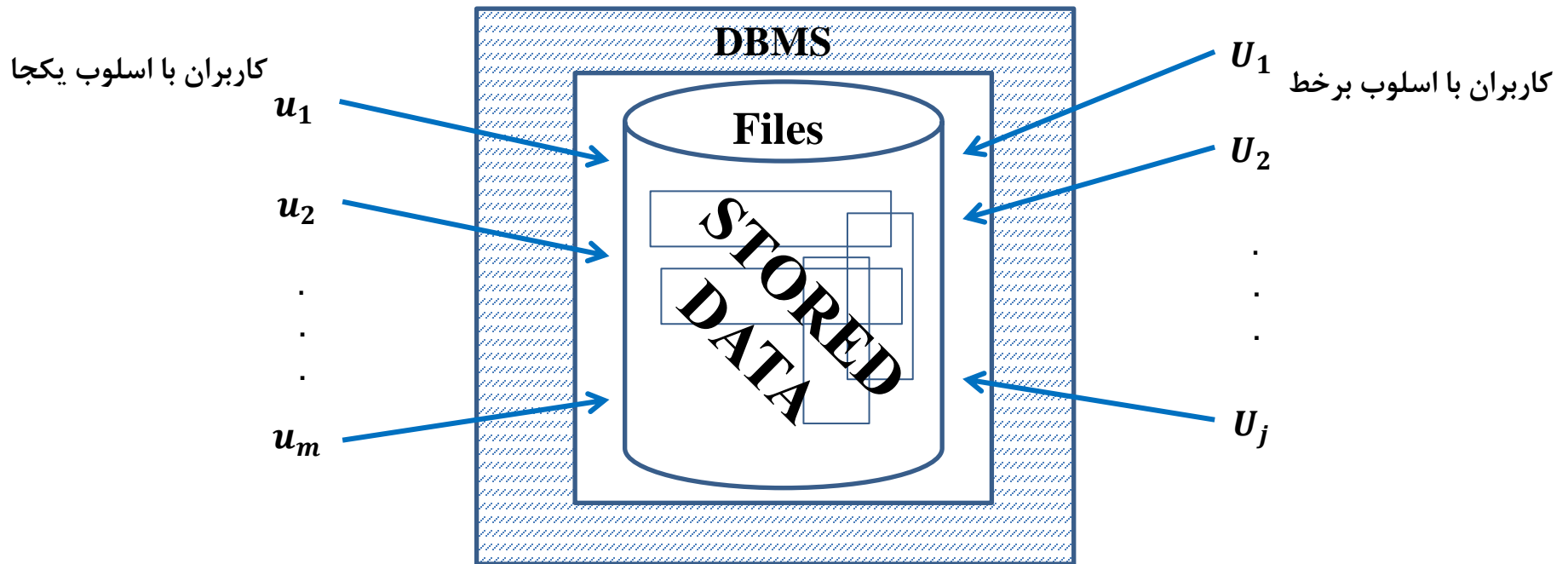
□ برنامه‌های کاربردی (Apps)



یا با خود DBMS می فروشند،
یا جداگانه خریداری می شود و به
امکانات آن اضافه می شود.

→ تسهیلات نرم افزار

□ در معنای عام هر استفاده‌کننده از سیستم پایگاهی را **کاربر** گوئیم که انواع مختلفی دارد. کاربر می‌تواند انسانی و یا ماشینی باشد.





□ انواع کاربر از نظر اسلوب عملیاتی:

□ **Batch – یکجا** (تعدادی برنامه یا پرس و جو جمع آوری می شود و به صورت یکجا به سیستم داده می شود و جواب آن بر می گردد).

□ **Online – برخط – پیوسته** (یک برنامه یا پرس و جو به سیستم داده می شود، اجرا می شود، و جوابش برمی گردد).

□ **Interactive – تعاملی** – بسته به اینکه چه جوابی داده شود عمل دیگری از کامپیوتر درخواست می شود.

▪ Online لزوماً Interactive نیست اما Interactive لزوماً Online است.

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



□ 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.



پرسش و پاسخ ...

amini@sharif.edu



تشکیل شده از تعدادی درایه (مدخل-entry)

نمایه متراکم dense

یک

یا

گروهی (به صورت

هر مدخل اشاره دارد به { از رکورد ها

چند سطحی)

نمایه نامتراکم Non-dense

مقدار

آدرس

تکیه گاه (Anchor point)

مقدار یک صفت (معمولا کلید)



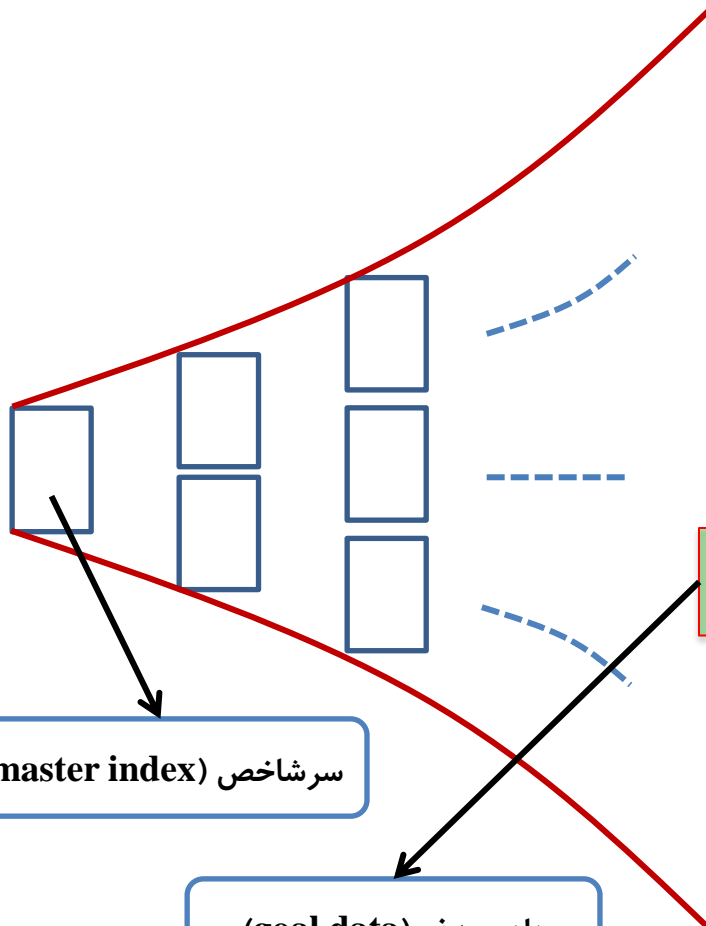
نمایه نامتراکم

نمایه متراکم

فایل نمایه سازی شده

(چندسطحی معمولاً با ساختار B-Tree یا B⁺-Tree)

(زمان جستجو بالاست)



سرشاخص (master index)

داده هدف (goal data)

شماره	نام	رشته	...
۱۰۰	۱	نرم افزار	
۱۰۱	۲	نرم افزار	
۱۰۲	۳	سخت افزار	
۱۰۳	۴	نرم افزار	
۱۰۴	۵	سخت افزار	
۱۰۵	۶	نرم افزار	
⋮			
k		نرم افزار	
⋮			
۹۹۷	۸۹۸	سخت افزار	
۹۹۸	۸۹۹	سخت افزار	
۹۹۹	۹۰۰	نرم افزار	