

بسم الله الرحمن الرحيم

داشتن اطلاعات به معنای دانش نیست. تنها منبع دانش تجربه است . . .

پایگاه داده کاساندرا

دوره آموزش مقدماتی کاساندرا

www.bigdata-ir.com

Telegram: [@BigData Channel](https://t.me/BigData_Channel)

مدرس:

علیرضا حبیبی

فهرست مطالب

◎ پایگاه داده های غیر رابطه ای

- مورد استفاده پایگاه داده های غیر رابطه ای
- مقایسه با مدل رابطه ای
- انواع پایگاه داده های غیر رابطه ای و مقایسه آنها
- معرفی کسندرا و لایسنس کسندرا و مشتری های کسندرا

◎ معرفی امکانات و ویژگی های معماری کسندرا (تئوری)

- مقایسه نسخه ها کسندرا از لحاظ امکانات
- Timestamp, replication, Compaction

◎ مواجهه با داده در حوزه کلان داده

- فصل دوم کتاب داده کاوی

◎ نصب و راه اندازی کسندرا

- سخت افزار مورد نیاز و پرت های مورد نیاز برای تنظیم فایروال
- روش های نصب و تنظیمات نصب و اجرا

◎ مدل داده ای کسندرا

- نگاشت با مدل های رابطه ای
- نوع ها داده ای برای فیلد ها

◎ عملکرد کسندرا

- حافظه، نوشتن و خواندن، حذف، فشرده سازی

- ◎ Multiple Secondary Indexes Primary Indexes Secondary Indexes

◎ دستورات کسندرا

- خط فرمان کسندرا CQL3,CQL2,CLI
- دستورات اتصال و ساخت و درخواست

◎ بهره گیری از کسندرا با جاوا

- راه اندازها و روش اتصال
- روش ارسال درخواست و روش دریافت پاسخ ها

◎ مانیتور کردن کسندرا

- نصب و راه اندازی آپاچی تام کت
- استفاده از helenos-1.4.war, Ops Center

◎ یک پیاده سازی عملی از کسندرا

- شرکت هواپیمایی
- انبار لاگ های سیستم ها

◎ مهاجرت به کسندرا از پایگاه داده های رابطه ای

- سرویس وبلاگ

◎ امکانات پیشرفته کسندرا به صورت عملی

- امکاناتی که به صورت خودکار فعال می شوند.
- فشرده سازی
- کلاسترها

- Replication, Partitioning

مقدمه ای بر پایگاه داده های NOSQL

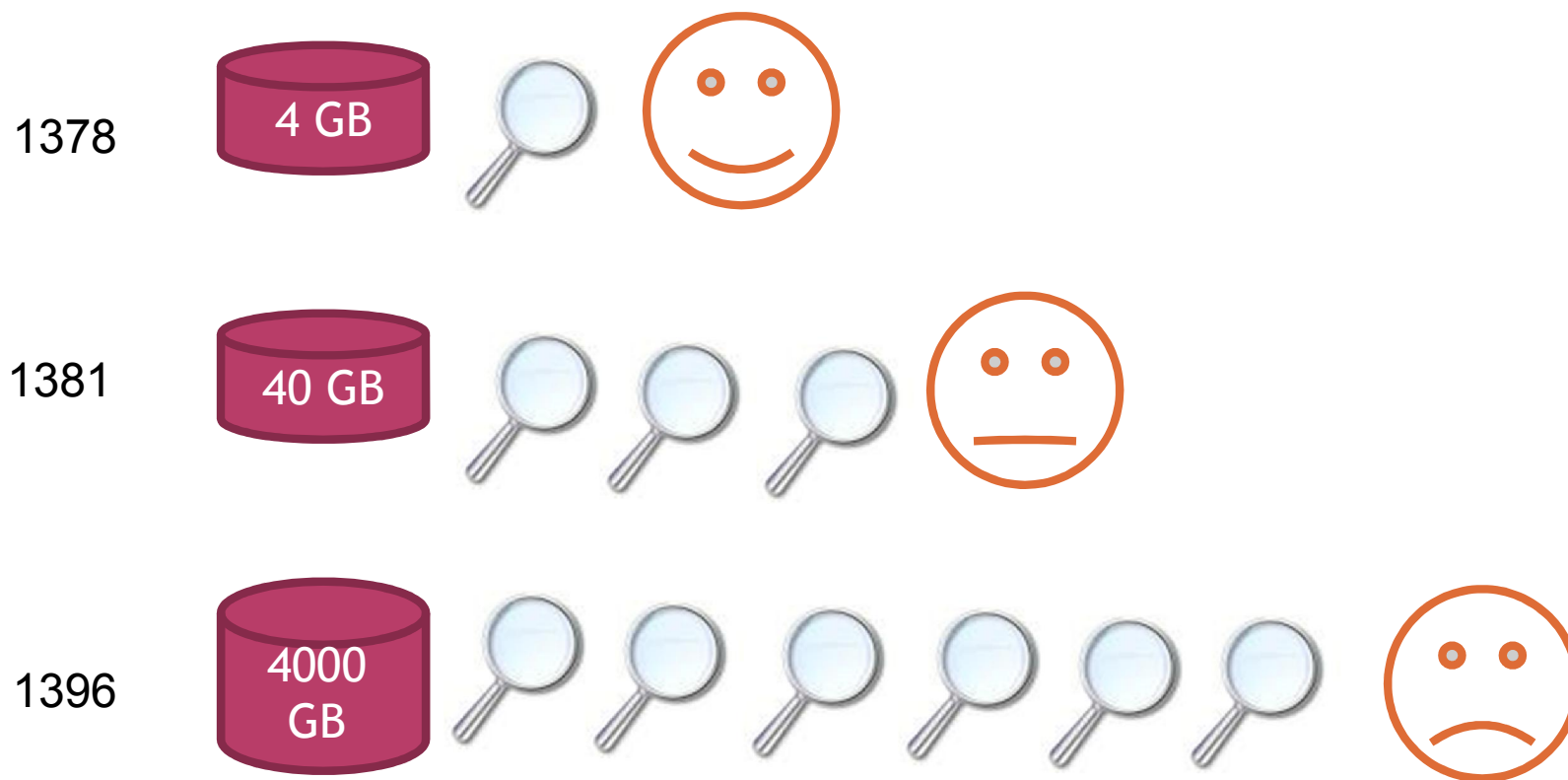
www.bigdata-ir.com

Telegram: [@BigData_Channel](https://t.me/BigData_Channel)

مدرس:

علیرضا حبیبی

صورت مسئله



راه حل های مواجهه با بیگ دیتا

◎ راه حل های سخت افزاری Big Data

■ DAS

■ SAN

■ NAS

◎ راه حل های نرم افزاری Big Data

■ سیستم فایل توزیع شده

■ پایگاه داده های توزیع شده

■ مدل برنامه نویسی توزیع شده

راه حل های سخت افزاری Big Data

◉ local

- DAS (Direct Attached Storage)



◉ Network

- NAS (Network Attached Storage)
- SAN (Storage Area Network)

NAS vs. SAN Storage

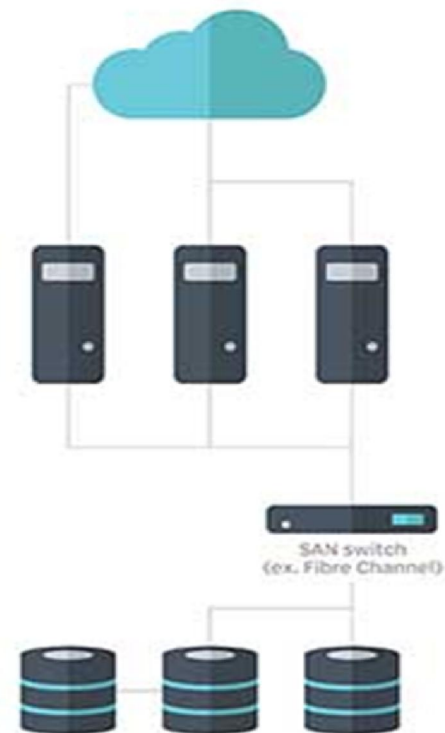
NETWORK-ATTACHED STORAGE

- Shared storage over shared network
- File system
- Easier management



STORAGE AREA NETWORK

- Shared storage over dedicated network
- Block storage
- Fast, but expensive



راه حل های نرم افزاری Big Data

○ سیستم فایل توزیع شده

■ GFS, HDFS, Colossus, Kosmos, TFS, FastDFS

○ پایگاه داده های توزیع شده

■ Relational

■ Non Relational

○ key/value

○ Column Oriented

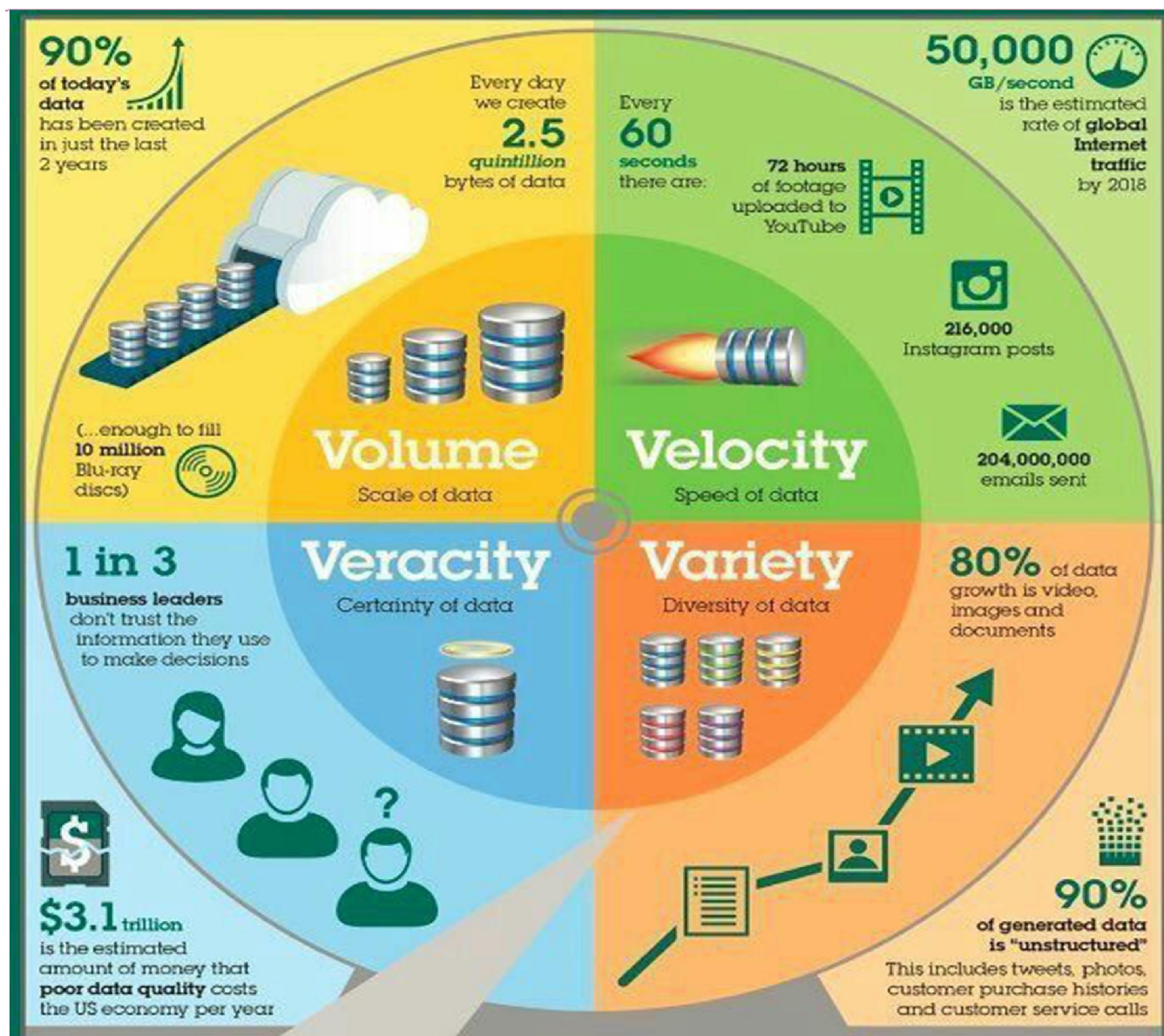
○ Document Store

○ Graph base

○ مدل برنامه نویسی توزیع شده

■ Daryad , Map Reduce, All-Paire, Pregel, MPI

مواجهه با انواع داده و گسترده اطلاعات



• نرخ ترافیک اینترنت در سال ۲۰۱۸ به ۵۰,۰۰۰ گیگا بایت در ثانیه خواهد رسید

• ۹۰٪ داده‌های امروز در طی دو سال قبل تولید شده است

• ۹۰٪ داده‌های امروز بودن ساختار هستند داده‌ها ساخت نیافته

• نا همگون
• دائم در حال تغییر
• داده‌ها بزرگ

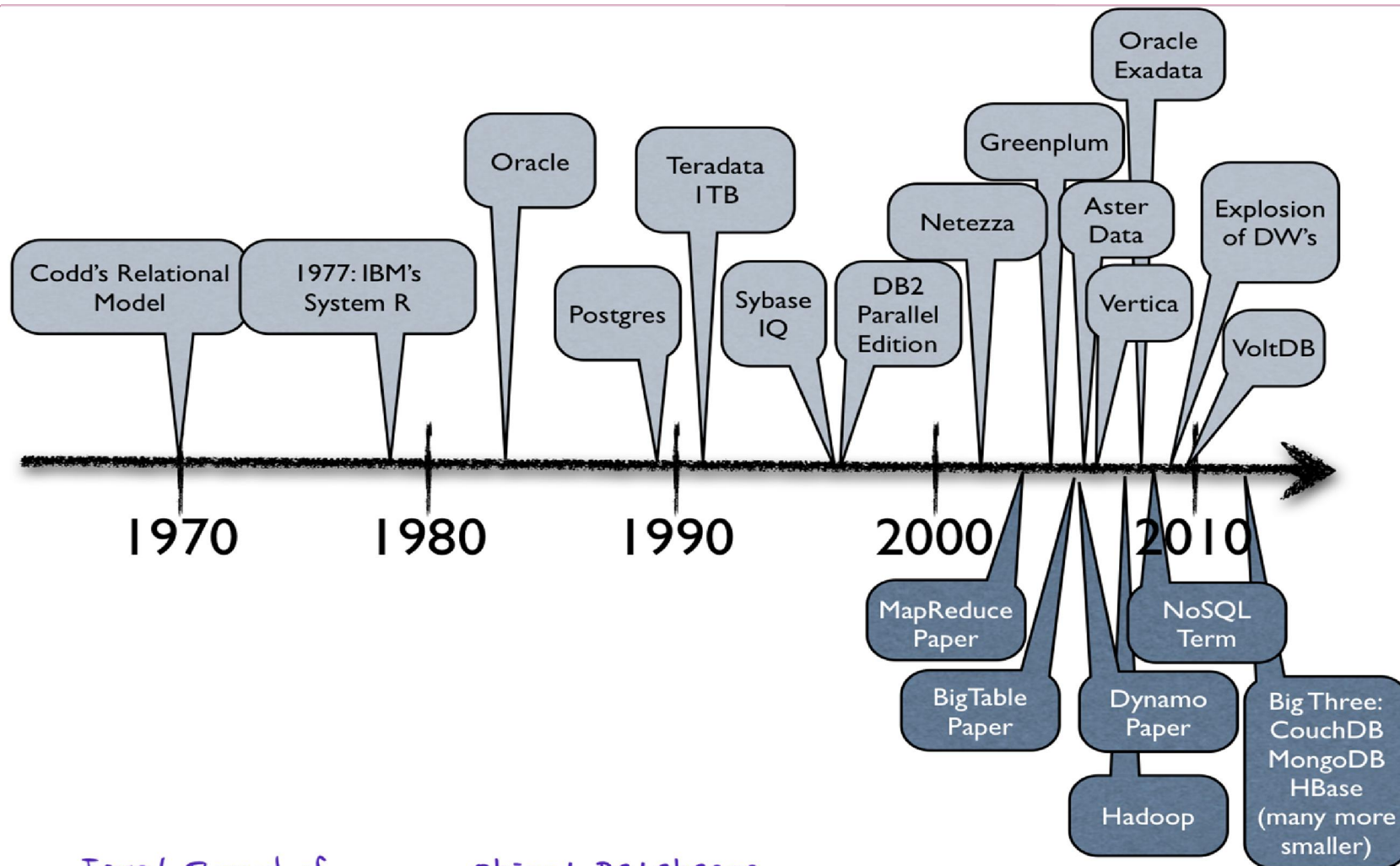
دوره آموزشی

چه وقت پایگاه داده های غیر رابطه ای؟

NOT ONLY SQL

- ◉ وقتی که داده های ساخت نیافته یا نیمه ساخت یافته
- ◉ داده ها ساخت یافته هستند ولی بسیار زیادند
- ◉ سرعت تولید داده زیاد (Velocity) یا به صورت سیل آسا می آیند
- ◉ داده ها زیاد نیست ولی رجوع به داده ها بسیار زیاده
- ◉ نیاز به سرعت بالا پردازشی (High Performance)
- ◉ داده ها فعلا کم هستند و پیوسته در حال رشد هستند. مقیاس پذیری (Scalability)
- ◉ نیاز به توزیع شدگی به هر دلیلی مثل Safety
- ◉ عدم سرویس دهی زیان آور هست. پایداری و دسترسی پذیری بالا (High Availability)
- ◉ شاخص گذاری بخش زیادی از اطلاعات

SQL & NOSQL history



First Round of
Database Wars

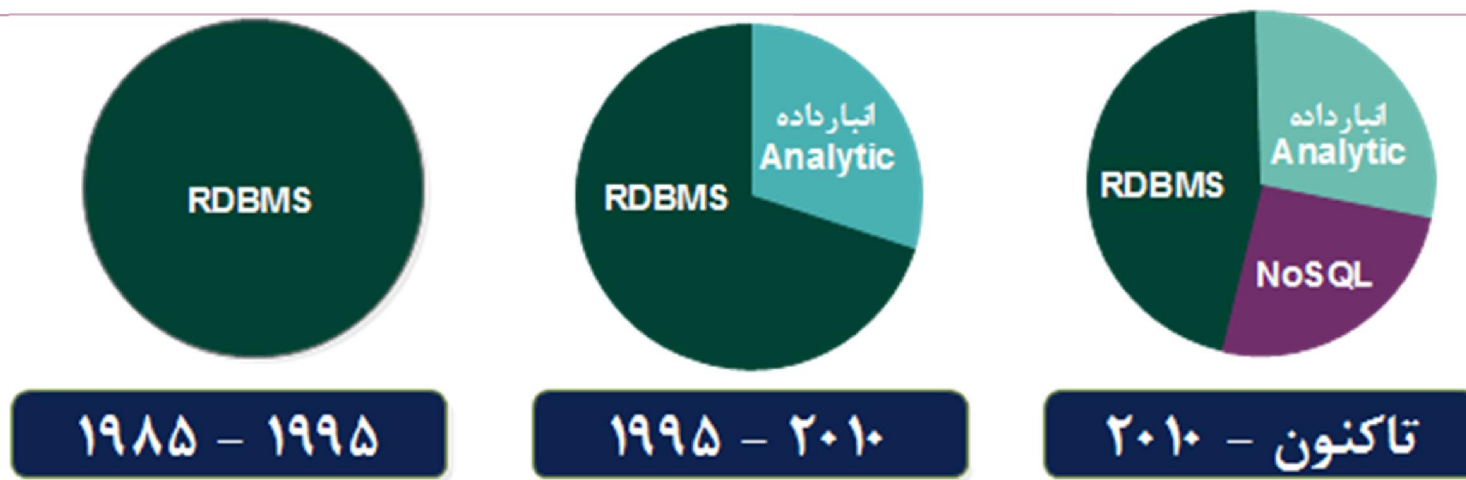
مدرس: علیرضا حبیبی

Object Databases
challenge

دوره آموزشی کاساندرا

Semi-structured

دوره های مختلف بانک های اطلاعاتی



وضعیت قبل از NoSQL ها

Relational Databases

- Relational databases have been around for 40+ years and will always have a place in technology. There are many situations where a relational database is the best choice
- **NoSQL** databases are **NOT** going to replace relational databases in all areas

مزایای استفاده از پایگاه داده‌های رابطه‌ای

- Better For Complex Data Select
- Better Support
- Better For “Relational” or “Object Orientated” Data
- Been Around Longer
- Better Data Analytics
- Full support ACID

مزایای استفاده از پایگاه دادهای غیر رابطه ای

- Handles Big Data Much Better Than RDBMS
- Cheaper Hardware & Software
- Easier Scaling (**Scale-Out**)
- Much More Flexible (Relatively **Schema-Free**)
- Map and Reduce Capability
- Less Management

Scale-Up vs. Scale-Out

- **Scale-Up Architecture (Relational)** – Storage is confined to a single form-factor which needs more resources to scale (CPU, Memory, etc) to a single node.
- **Scale-Out Architecture (NoSQL)** - The total amount of disk space can be easily expanded as needed. When a storage array reaches it's max, another will pick it up where the last left off. This makes scaling significantly easier and less expensive.

مشکلات پایگاه داده‌های رابطه

- SQL Queries can take a huge amount of **time** with large data volume
- Data can be **inconsistent**
- Require **expensive** proprietary hardware



Both Have Their Places

- Basically, it comes down to what type of **application** or data **structure** you need.
- **NoSQL** databases are better with handling **huge** amounts of data with **easy scaling**
- **Relational databases** are better with more **complex data** and queries.

ملاحظات سطح غیر عملیاتی

When choosing a system/database for your projects or company, there are many things to factor in...

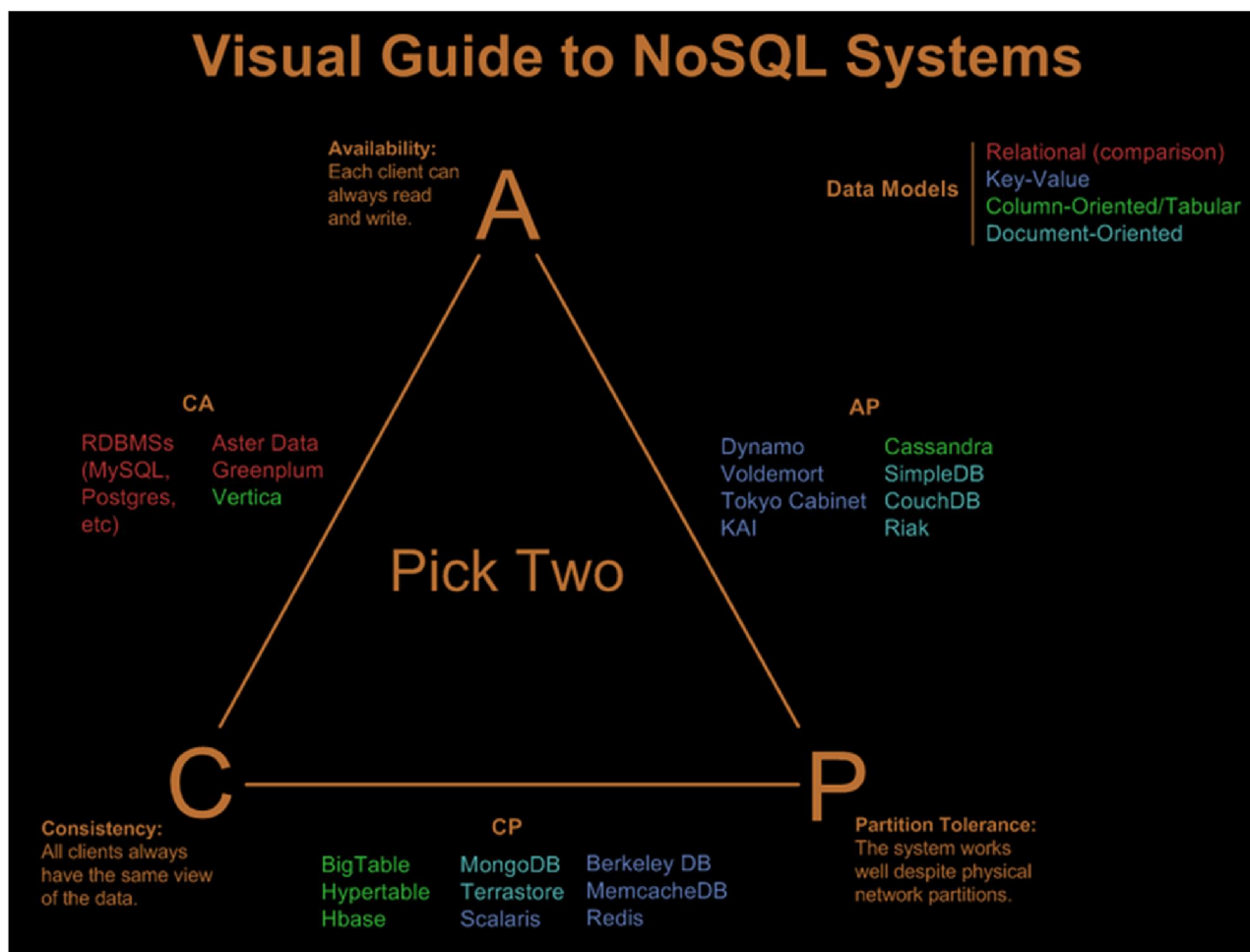
- Complexity
- Human Resource
- Budget
- Security & Safety
- Maintenance
- Programming Languages
- HardWare
-

ملاحظات در انتخاب پایگاه داده

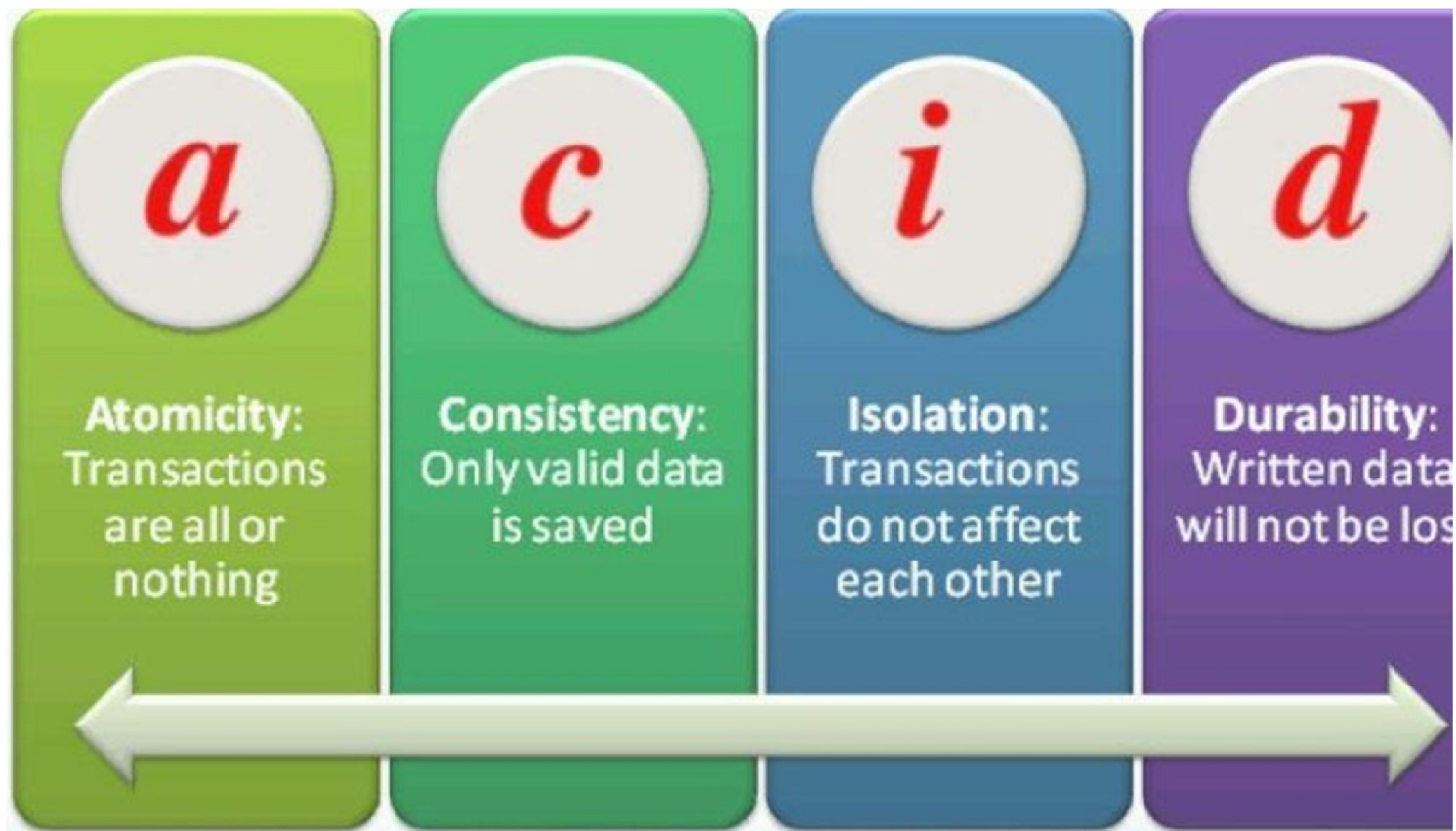
- قضیه CAP
- اصل ACID و BASE
- قانون 3V
- مدل ساختار داده در پایگاه داده های غیر رابطه ای
- سایر خصوصیات جزئی:
 - توان و روش ایندکس گذاری
 - روش های تکثیر و توزیع شدگی
 - روش های Sardining
 - میزان سادگی نگهداری و بازیابی
 - روش و متدولوژی استقرار (Local/geography)
 - ...

فارق از مدل NOSQL قابلیت آنها اهمیت دارد

انتخاب دو قابلیت از ACP



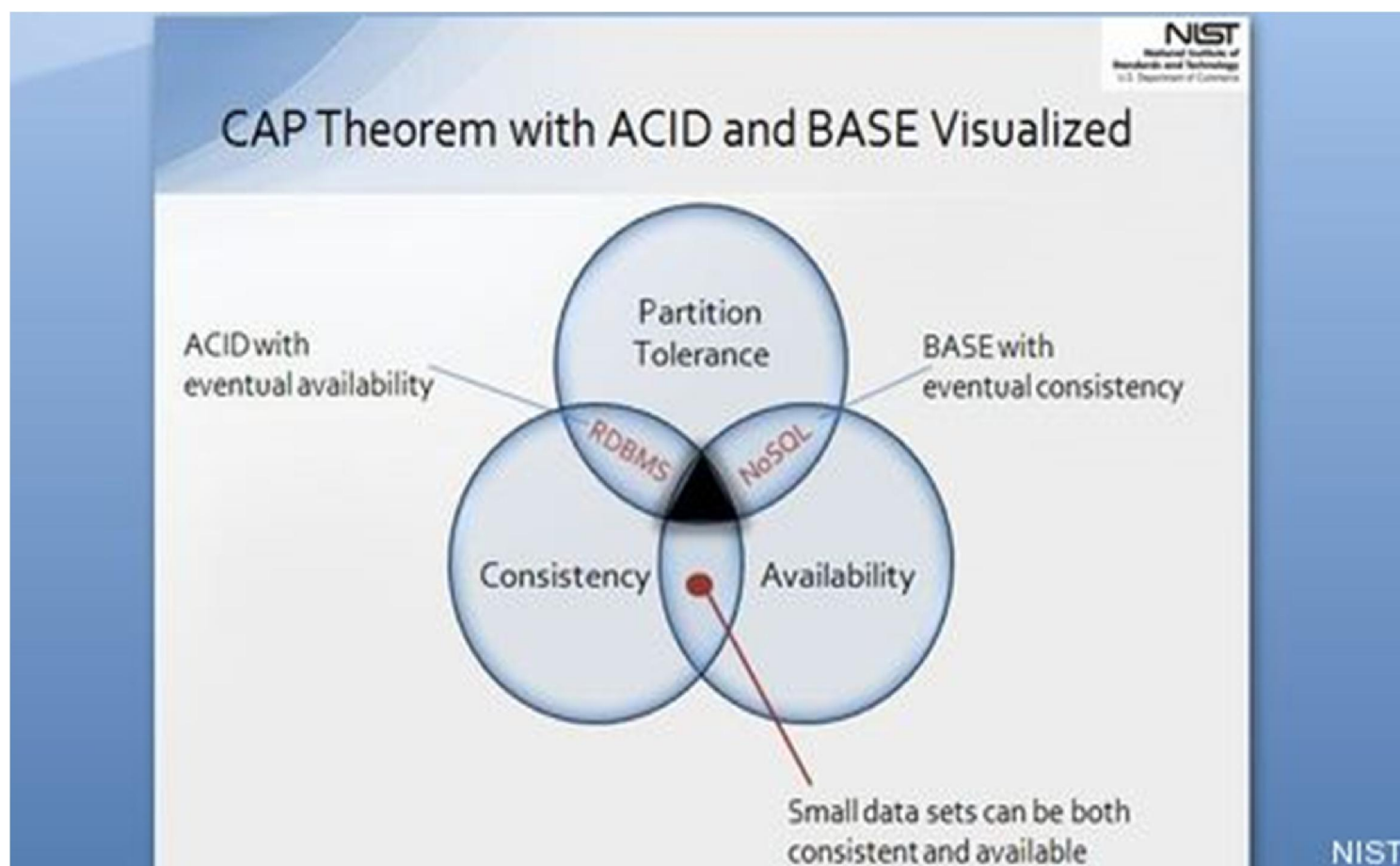
ACID



ACID

- **Atomicity یکپارچگی:** این خاصیت به « همه یا هیچ » موسوم می باشد . منظور این است که تمام دستورالعمل های یک تراکنش باید اجرا شود و یا هیچ کدام از آنها
- **Consistency همخوانی یا سازگاری داده:** این خاصیت به این منظور است که هر تراکنش باید تمام قوانین **جامعیت بانک اطلاعات** را رعایت کند. تراکنش ممکن است دو نوع پایان داشته باشد: پایان ناموفق که آنرا “برگشت” Rollback می نامند، پایان موفق که آنرا “انجام” Commit می نامند.
- **Isolation انزوا:** بر طبق این خاصیت اثر تراکنش های همروند روی یکدیگر چنان می باشد که گویا هر کدام در انزوا انجام می شود. در بانک اطلاعات تراکنش های همروند وجود دارند ولی همروندی آنها کنترل می شود تا **اثر مخرب روی هم نداشته** باشند.
- **Durability پایدایی:** براساس این خاصیت تراکنش هایی که به مرحله انجام Commit برسند اثرشان ماندنی است و هرگز به طور تصادفی از بین نمی رود.

BASE



BASE

◉ در اصل پایگاه داده های غیر رابطه ای خصوصی به نام **BASE** را دارا هستند که همه خاصیت های **ACID** را ندارد ولی در عوض قابلیت توزیع شدگی و تحمل سربار زیاد را ارائه می دهند.

■ Basic Availability: پایداری بنیادی

■ Soft-state: وضعیت غیر پایدار در داده ها و عدم اجبار در سازگاری هنگام عملیات نوشتن

■ Eventual consistency: در نهایت سازگار

MicroStrategy's View of the 3 V's + Value

VOLUME

- Petabytes generated by a surveillance camera
- Thousands of call records per second



Large Scale of Data
ranging in petabytes
or more

VARIETY

- Social Media
- Clickstream data
- Web logs
- Machine data



Structured, Semi-
structured and
Unstructured
Data

VELOCITY

Data in Motion
Time sensitive
streaming data



Data in Motion
Time sensitive
streaming data

VALUE

How fast/best the
data can be analyzed
and acted upon to
provide business
VALUE?

