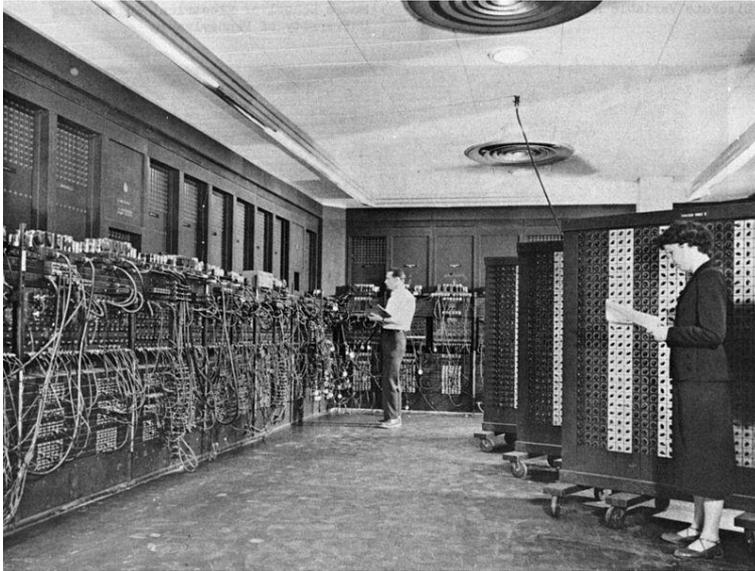
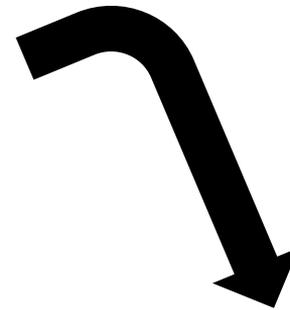


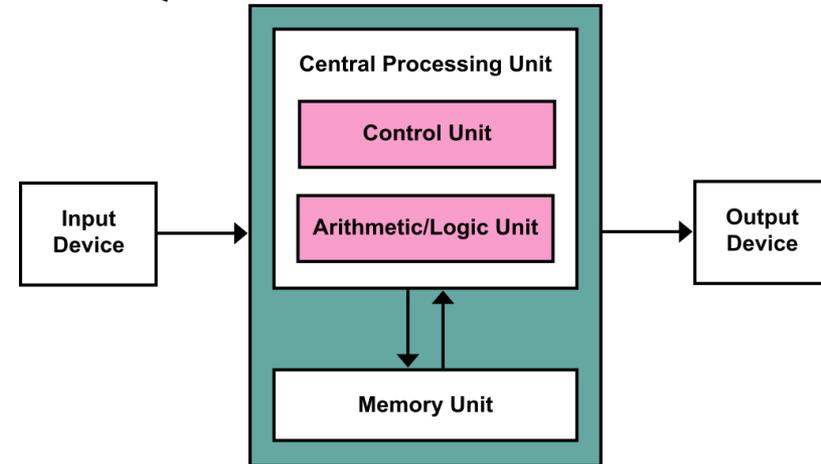
Evolution of Computers



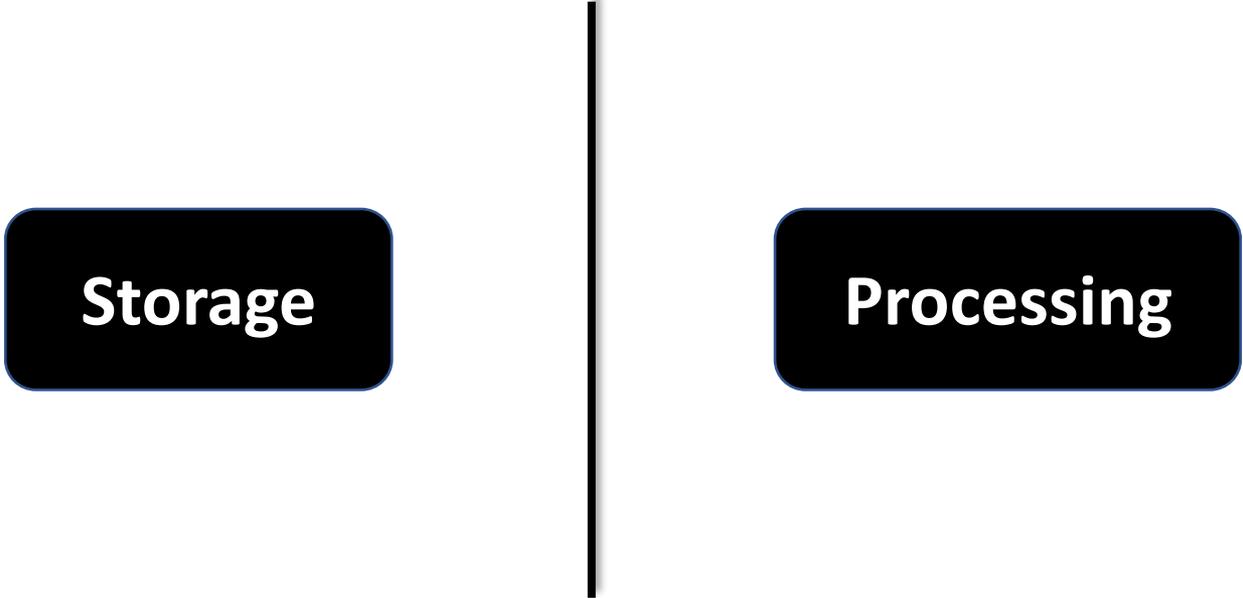
ENIAC
Early 1900s



**Stored-program
Von Neumann
Architecture
1940**



Two Kinds of Problems

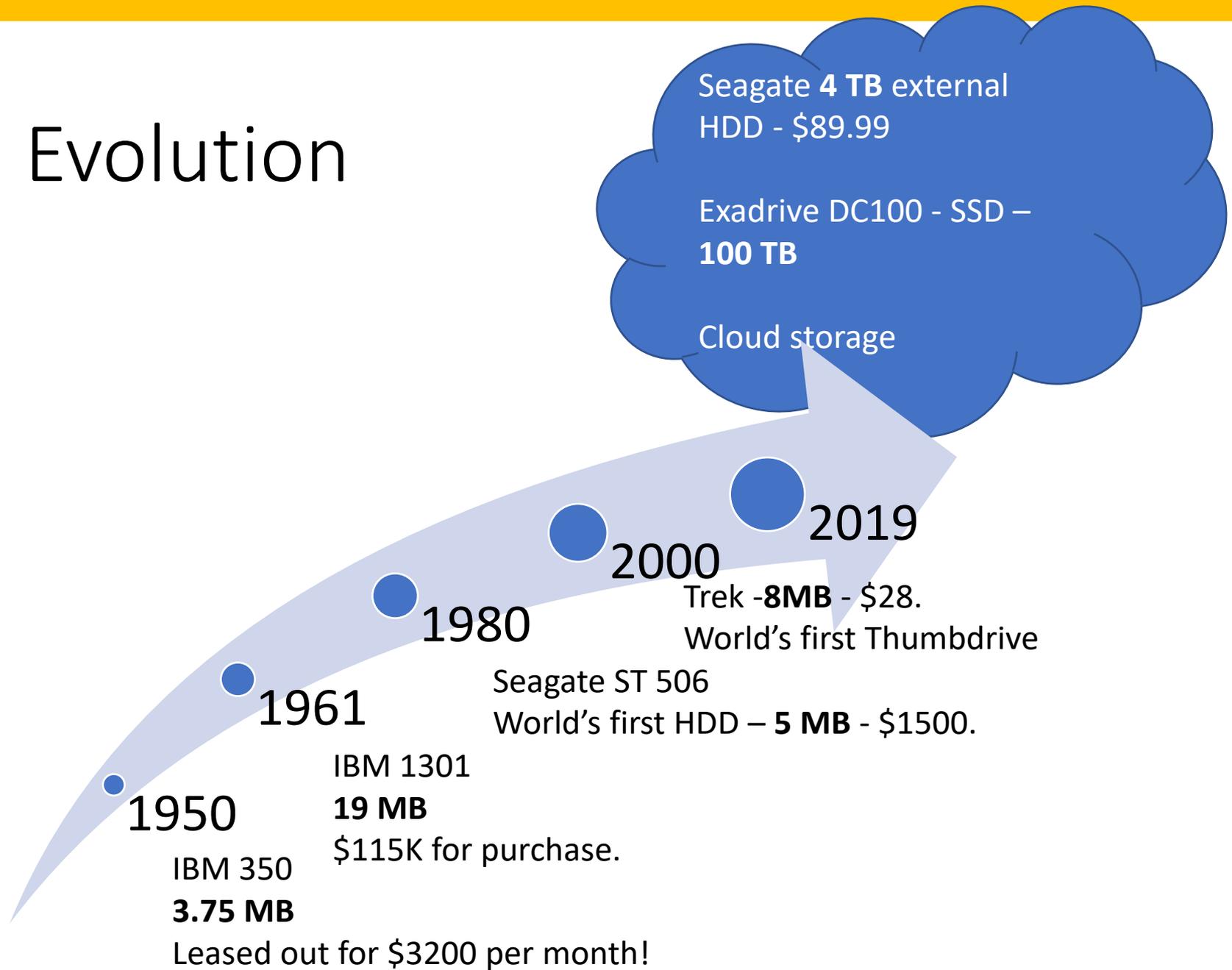


Storage

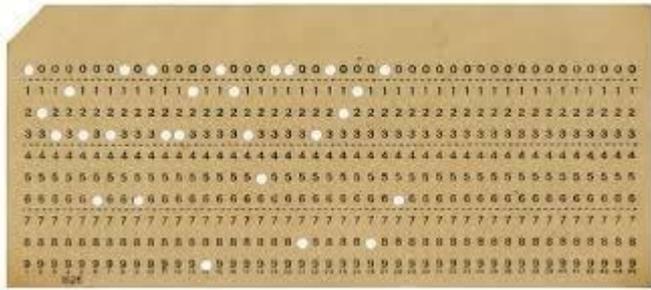
Processing

Data Storage

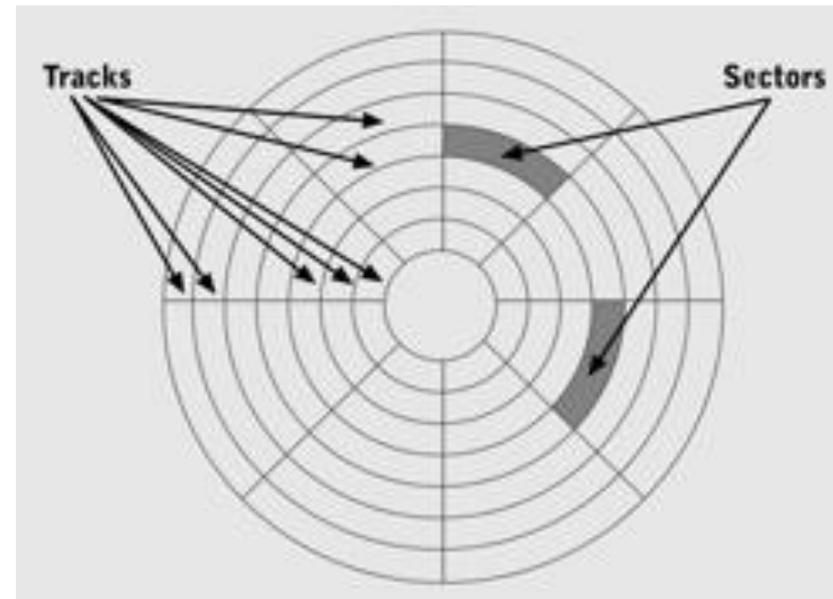
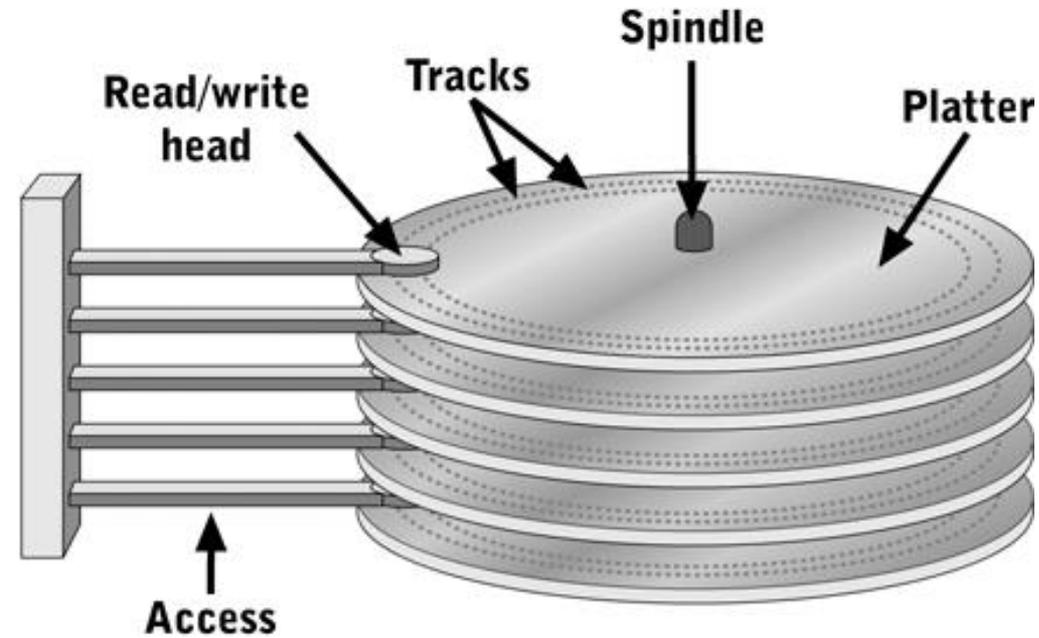
Evolution



(Secondary) Storage Technologies



Disk Drive and Access Time





Roll over image to zoom in

Seagate 500GB SATA Laptop Hard Disk

by [Seagate](#)

★★★★☆ 279 ratings | 493 answered questions

M.R.P.: ₹ 2,999.00

Price: **₹ 1,433.00** + ₹ 77.00 Delivery charge [Details](#)

You Save: **₹ 1,566.00 (52%)**

Inclusive of all taxes



Pay on
Delivery



10 Days
Replacement



Amazon
Delivered



1 Year
Warranty

In stock.

Delivery by: **Jan 8 - 10** [Details](#)

[📍 Deliver to Venkatesh - Chennai 600014](#)

Sold by [KCM_STORE](#) (3.9 out of 5 stars | 29 ratings).

New (20) from ₹ 1,510.00 + FREE Shipping

- 500 GB capacity
- 5400 RPM spin speed, 16 MB cache buffer
- Designed for durability and low-power consumption
- SATA 3GB interface with native command queuing
- Perpendicular recording technology for increased storage capacity
- Fast performance and whisper quiet acoustics

Average Access Time

- **Head switching time** is considered negligible (H)
- **Head seek time** (S)
- **Rotational delay** = Time taken for $\frac{1}{2}$ a rotation (average) (R)
- **Read time** = time to spin an entire sector (T)
- **Average Access Time** = $H + S + R + T$

*Sector is a minimum storage unit

Quiz

- If disk spins at 6000 RPM, compute the rotational delay.

Quiz

- If disk spins at 6000 RPM, compute the rotational delay.
 - One turn takes $1/6000$ min or $1/100$ sec = 10ms
 - $\frac{1}{2}$ a turn takes 5ms.

Read Time

- If the drive spins at 6000RPM and the disk has 20 sectors per track, what is the read time?

- Time for 1 full spin is $\frac{1}{6000} \text{ min} = \frac{1}{100} \text{ sec} = 10\text{ms}$

- Time for 1/20 of a spin is $10\text{ms} \times \frac{1}{20} = 0.5\text{ms}$

Average Access Time

- Drive spins at 7200RPM and has average seek time of 8ms. The disk has 24 sectors per track. What is the average access time?

Head seek time	0.008 sec (Given)
Rotational delay	$1/120 * (1/2) = 0.0042$ sec
Read time	0.0084 (full spin) / 24 sectors = 0.00035 sec
Avg Access Time	= $0.008 + 0.0042 + 0.00035$ = 0.01255 sec or 12.55 ms

Characteristics

Attribute	Description
Speed	Time to read/write
Volatility	Data persistence even when powered off
Access Method	Serial, Parallel
Portability	Internal, External
Capacity	Volume of data storage

Storing/Managing/Processing Data

- RDBMS
- ETL
- OLTP
- Data Warehouse
- Data Lake
- Cloud Storage
- STaaS

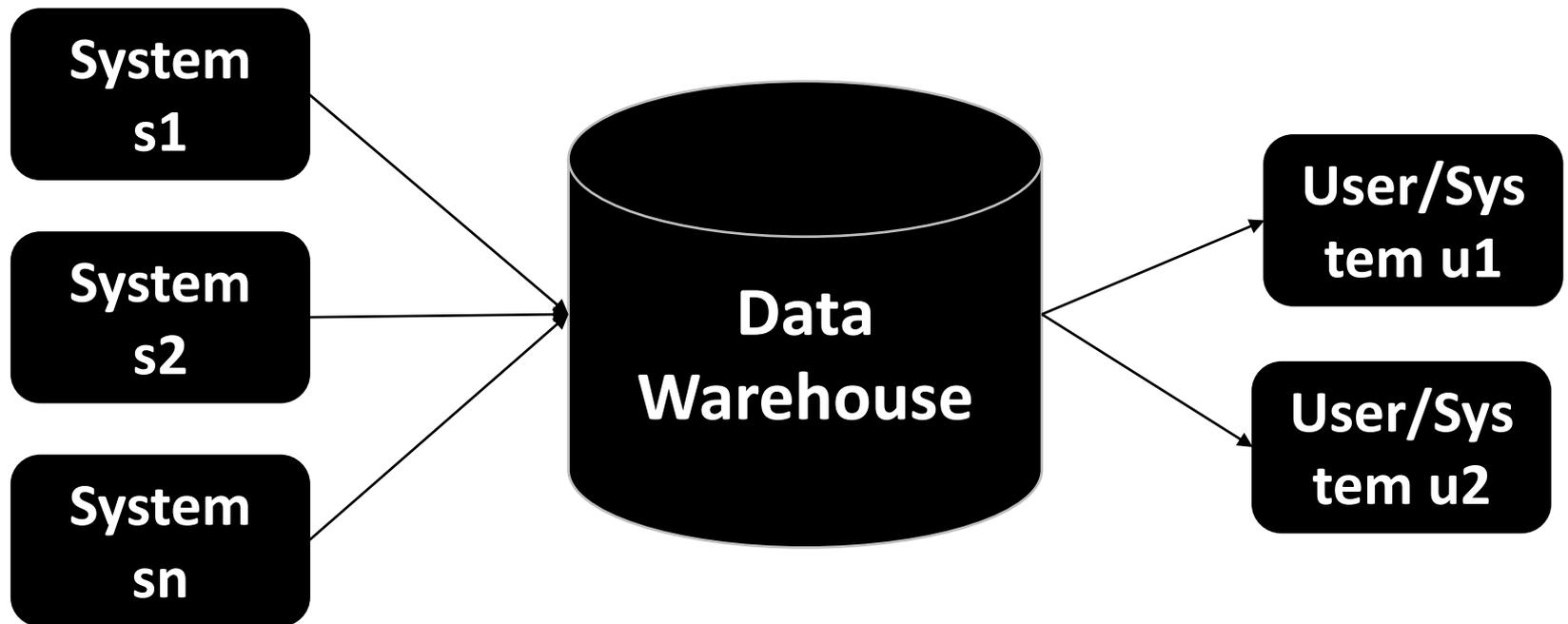
ETL

- Extract, Transform and Load (ETL)
 - general procedure followed to address data variety
- Variety of data sources
 - Tabs, Sensors, Desktops, Bots, Multiple databases, Files,...
- Variety of data formats
 - Text, PDFs, XML, JSON, Images, Videos, ...

Fast OLTP

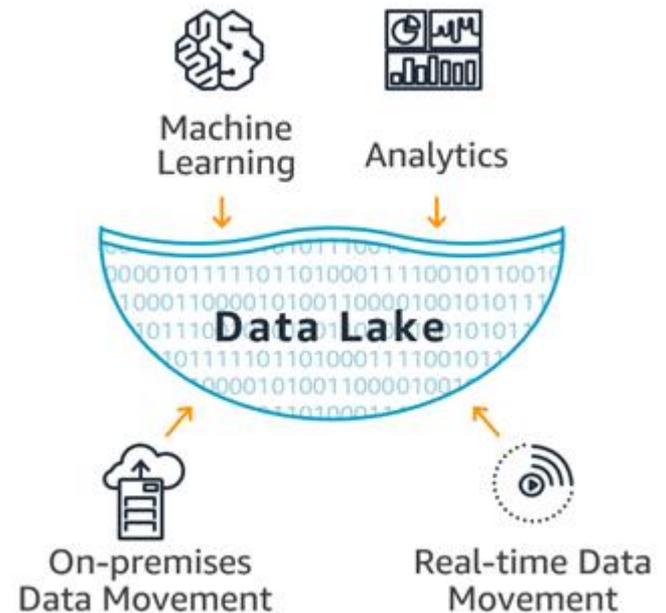
- Online Transaction Processing (OLTP)
- Real-time/Near Real-time Performance. Finds application in:
 - Banking
 - Railway Reservations
 - Stock Market Trading
 - Handle transactions in milliseconds.
 - VoltDB, SingleStoreDB(MemSQL), ...

Data Warehouse



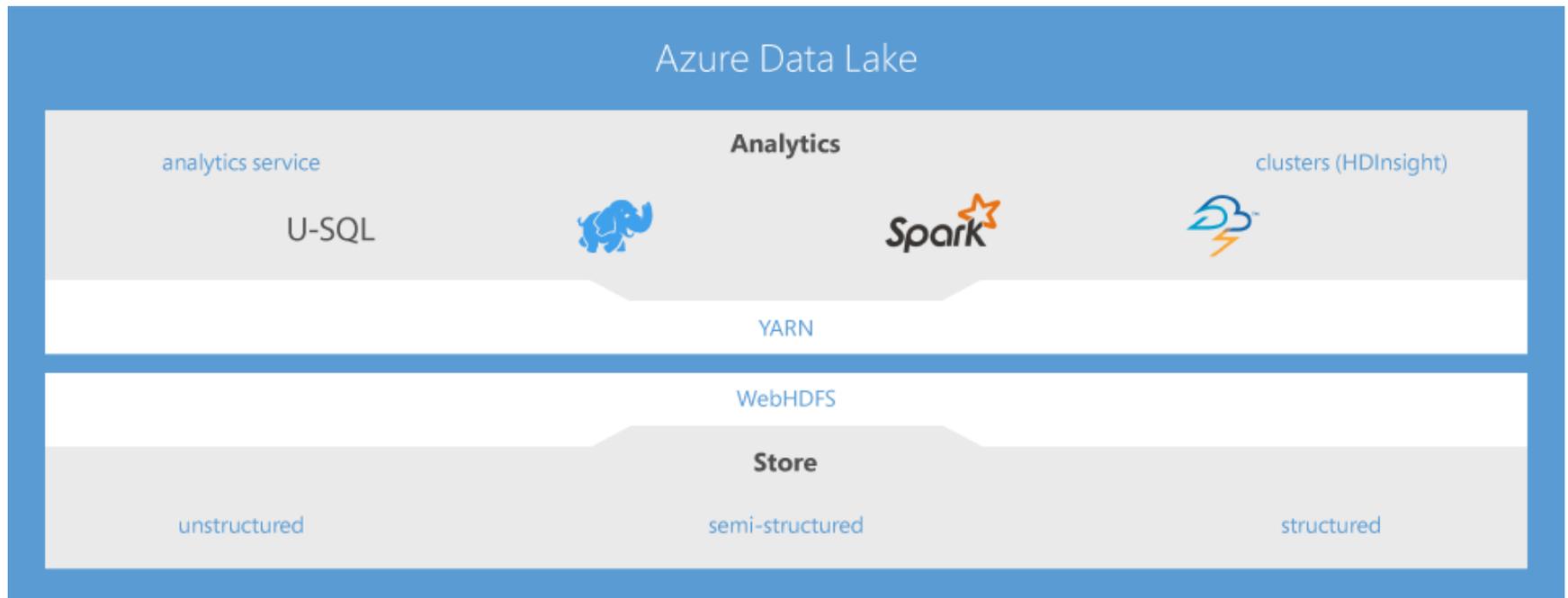
Data Lakes

- No schema definition.
- Store everything
 - often without or with very little pre-processing, /cleaning.
- Use ML, analytics to query, or gather insights.



Source: <https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>

Microsoft's Azure Data Lake



More details at <https://azure.microsoft.com/en-us/resources/videos/azure-data-lake-making-big-data-easy/>

Data Warehouse Vs. Data Lake

Characteristics	Data Warehouse	Data Lake
Data	Relational from transactional systems, operational databases, and line of business applications	Non-relational and relational from IoT devices, web sites, mobile apps, social media, and corporate applications
Schema	Designed prior to the DW implementation (schema-on-write)	Written at the time of analysis (schema-on-read)
Price/Performance	Fastest query results using higher cost storage	Query results getting faster using low-cost storage
Data Quality	Highly curated data that serves as the central version of the truth	Any data that may or may not be curated (ie. raw data)
Users	Business analysts	Data scientists , Data developers, and Business analysts (using curated data)
Analytics	Batch reporting, BI and visualizations	Machine Learning , Predictive analytics, data discovery and profiling

Source: <https://aws.amazon.com/big-data/datalakes-and-analytics/what-is-a-data-lake/>

Storing on the Cloud

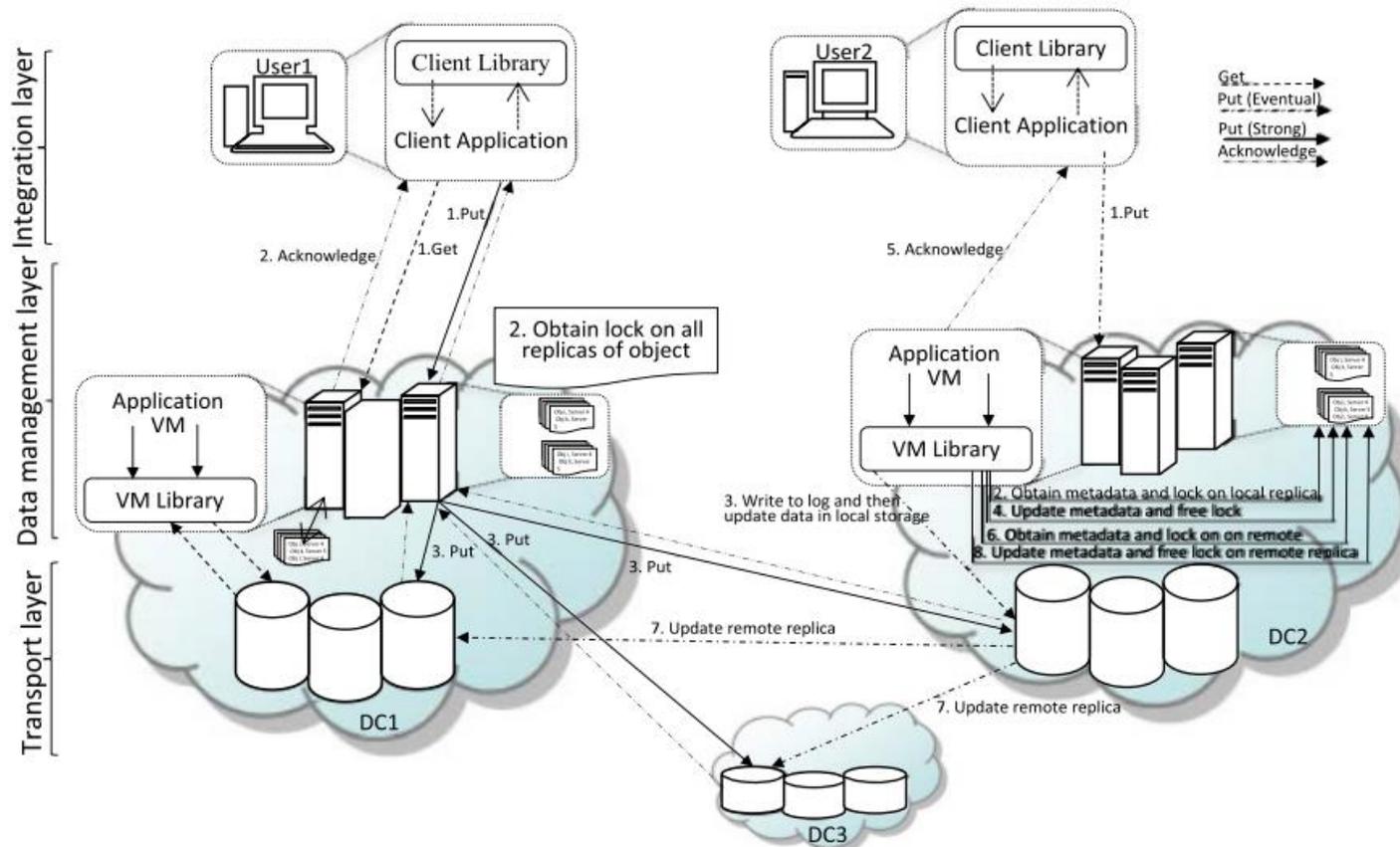
- Gmail: Gives 15 GB of free storage (as of 2023)
- Several online sites for storing images, apps, files, ...
 - Security
 - Ease of sharing
 - Backups
 - Availability



Storage as a Service (STaaS)

- What is it?
 - A business model in which a company rents space in their storage infrastructure to another company or individual.
- How does it work?
 - STaaS provider rents space
 - cost-per-gigabyte-stored and cost-per-data-transfer basis.
- Benefits
 - Shifting from Capital Expenditure to Operational Expenditure
 - Scale up/down at will (temporarily)

Cloud Storage



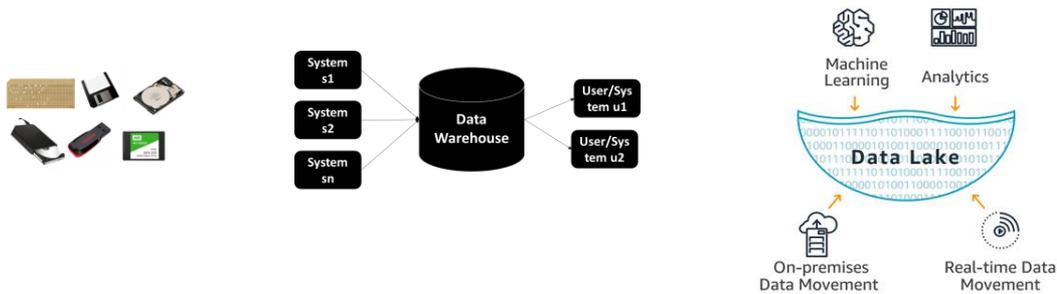
Mansouri et al., **Data Storage Management in Cloud Environments**,
ACM Computing Survey (CSUR) 2018.

A Data Center



Source: https://resources.sei.cmu.edu/asset_files/WhitePaper/2010_019_001_28877.pdf

Summary



Data Storage - Summary

