DA 320 Assignment 1 M. Blanco

**Short Answers:**

**Problem 1:** Why accessibility to data is so vital in today’s computing environment?

Data accessibility is critical for the success of business in today’s competitive environment. Easy access to data enables companies to move quickly, be product focused, and foster data-driven culture to enable leaders to make better decisions based on the insight from data.

**Problem 2:** What’s the difference(s) between Mainframes and Supercomputers? List at least two main differences.

Supercomputers are intended to solve complex problems whose primary constraint is calculation speed. Supercomputers are ideal for performing complex calculations on a large data set. Mainframes are intended to solve problems constrained by input/output and which require high level of reliability. Mainframes are ideal for performing large number of concurrent transactions.

Mainframes:

• Run multiple programs concurrently

• Uninterrupted operation

• Perform tasks on large amounts of data

• Run many types of software applications and tackle extensive business tasks



Source IBM

Supercomputers:

• Processing power focus to execute a few programs or instructions as quickly as possible

• Speed and accelerated performance focus

• Push limits of what hardware and software can achieve



Source Foxconn

Good articles:

<https://www.forbes.com/sites/adrianbridgwater/2018/01/05/mainframes-for-a-new-generation-same-taste-new-can/#2322e40226bf>

<https://www.top500.org/news/foxconn-builds-taiwans-largest-supercomputer/>

**Problem 3:** Discuss the limitations of the “single-machine paradigm”.

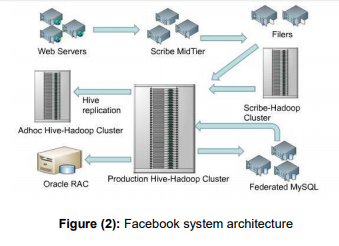
The single machine paradigm limitations become apparent when data volume increases. Impacts are not limited to reduced data availability/performance and analysis timeliness. Single machines are not easily scalable and are slow to perform data transformation tasks.

**Problem 4:**Define and explain the distributed computing model and give examples of said model.

A distributed computing model consists of networked computers that communicate, interact and coordinate their tasks through a network. The components interact to achieve a common goal. A key characteristic is that distributed computing systems appear as a single system to the end user.

The following are examples of distributed systems:

* Google Bots, Google Web Servers, Google Indexing Server
* Facebook (Social Media)
* Multiplayer online games
* Peer-to-peer applications



Source Erin O'Connor

**Problem 5:** Define and discuss the concept of data silos and the challenges of having data managed by different types of software that can’t easily communicate or share.

Data silos are often the result of the competing goals, priorities and responsibilities in large organizations. Data silo make managing business operations and organization more difficult. It is in the organization’s best interests to break down silos that hinder collaboration, productivity, efficiency and accessibility.

Data silos negatively impact data integrity. When data silos exist for the same data, the data contents often differ, which creates confusion as to which repository represents the most current version (source of truth). One of the challenges is the possibility of accidently overwriting data with outdated data. Additionally, integration between systems that were not intended to work together can be challenging. Data silos are often a result of actions from “knowledge workers” who refuse to share information. Overall, data silos hamper collaboration.



Source Forrester

Good articles:

<https://hbr.org/2016/12/breaking-down-data-silos>

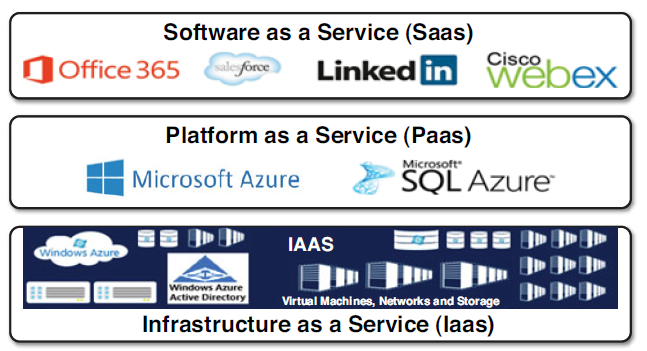
<https://www.forbes.com/sites/forbescommunicationscouncil/2017/06/09/contextual-marketing-and-the-power-of-removing-data-silos/#1bb24806121b>

**Problem 6:** Define and explain “infrastructure as a service.”

Infrastructure as a service (IaaS) is a type of cloud computing that provides virtualized computing resources over the internet. IaaS, is a self-service model for accessing, monitoring, and managing remote datacenter infrastructures, such as compute, storage, networking, and networking services. Users can purchase IaaS based on consumption instead of having to purchase hardware.

The following are a list of IaaS examples:

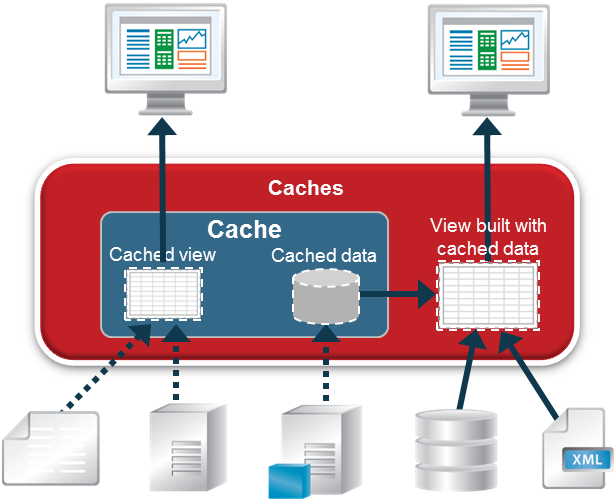
* Amazon Web Services (AWS
* Microsoft Azure
* Google Cloud



Source AIIM Community

**Problem 7:** Explain the theory behind having a data cache that provides data consistency.

Cached data is stored so future requests for data can be highly available by serving faster. Cached data stored in a cache might be derived from the output of an earlier computation. A subsequent request can be served faster by reading data from the cache rather than re-computing the output. In theory the system can perform faster when more requests can be served from the cache.



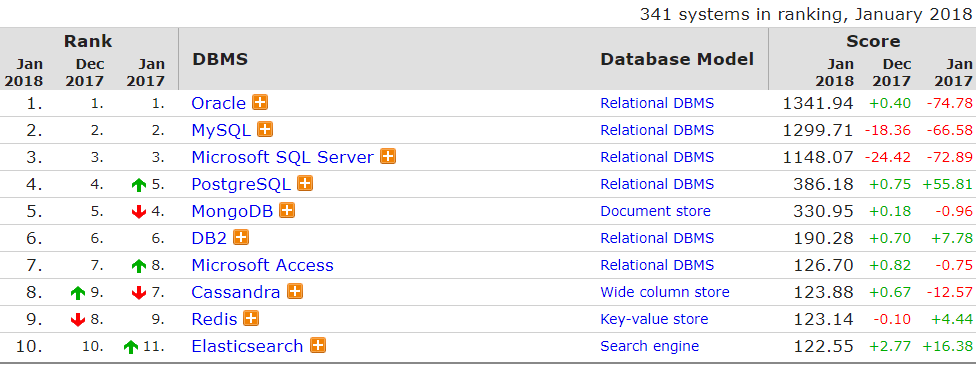
Source TIBCO Data Virtualization

**Problem 8:** Discuss where databases are today and the efforts of some companies to create such a magical database.

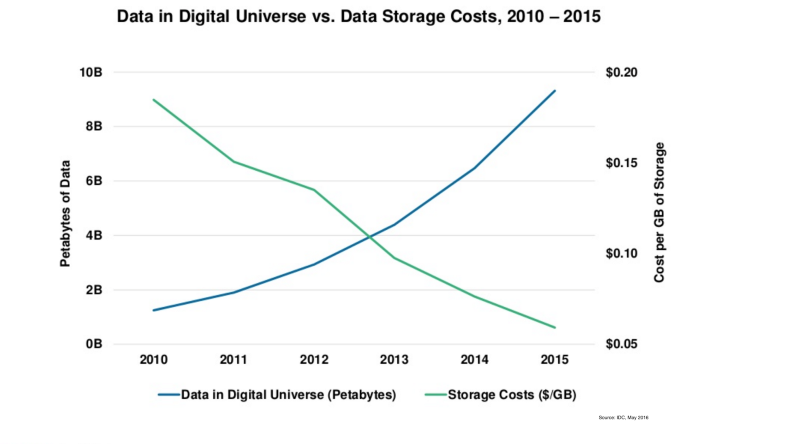
A recent study conducted by IDC estimated that the digital universe will grow to 44 zettabytes. The large growth of unstructured data (IoT and social media) has contributed to the increase complexity of data management systems. Unstructured data is estimated to account for 90 percent of digital information in the Big Data era. Organizations must have the ability to collect large amounts of data to improve business decision making.

Big Data includes trends and capabilities such as business intelligence, cloud computing, distributed data, NoSQL, Spark, Hadoop, sensors, streaming data, IoT and networked devices. The trend is for database administrators to manage multiple database such as Oracle, SQL Server and MongoDB.

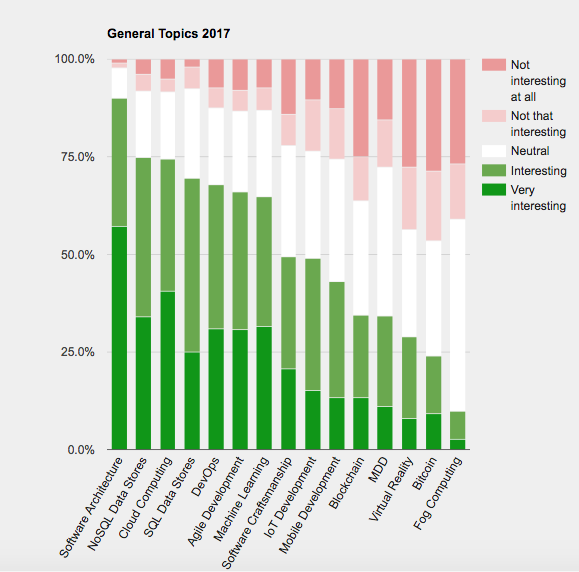
With the current data growth, organizations will find managing and analyzing data difficult using traditional systems such as data warehouse. The current trend is for organizations to move toward new database management strategies such as data lakes. Data warehouses are great for structured queries but fail when data is unstructured.



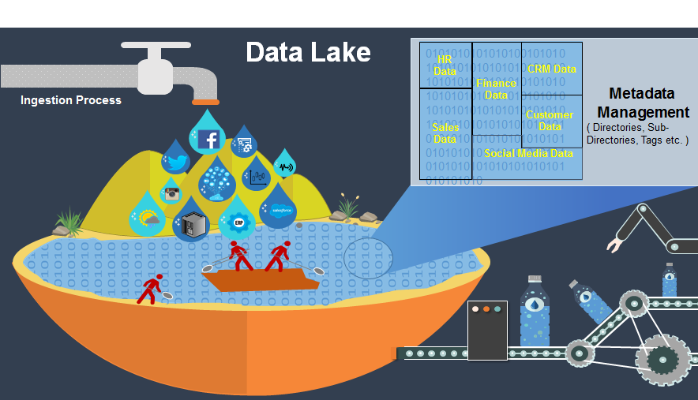
Source DB Engines



Source Medium



Source JAXenter technology trends survey 2017

Source Kiran Donepudi on LinkedIn

Good articles:

<https://jaxenter.com/top-databases-2017-132912.html>

<https://db-engines.com/en/ranking>

<https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>

[Source 1 IBM 1](#_Toc502933777)

[Source 2 Foxconn 2](#_Toc502933778)

[Source 3 Erin O'Connor 3](#_Toc502933779)

[Source 4 Forrester 4](#_Toc502933780)

[Source 5 AIIM Community 5](#_Toc502933781)

[Source 6 TIBCO Data Virtualization 5](#_Toc502933782)

[Source 7 DB Engines 6](#_Toc502933783)

[Source 8 Medium 7](#_Toc502933784)

[Source 9 JAXenter technology trends survey 2017 8](#_Toc502933785)

[Source 10 Kiran Donepudi on LinkedIn 9](#_Toc502933786)

# Bibliography

*IBM*. n.d. <https://www.ibm.com/us-en/marketplace/z14>.

O'Connor, Errin. *Understanding On-Premises, Cloud, & Hybrid Environments in SharePoint 2013 & Office 365*. 19 07 2014. <http://community.aiim.org/blogs/errin-oconnor/2014/07/19/understanding-on-premises-cloud--hybrid-environments-in-sharepoint-2013--office-365>.

Salem, Asma Mohammad. "Facebook Distributed System Case Study For Distributed System Inside Facebook Datacenters." *INTERNATIONAL JOURNAL OF TECHNOLOGY ENHANCEMENTS AND EMERGING ENGINEERING RESEARCH* 2.7 (2014). <http://www.ijteee.org/final-print/july2014/Facebook-Distributed-System-Case-Study-For-Distributed-System-Inside-Facebook-Datacenters.pdf>.

*TIBCO Data Virtualization*. n.d. <http://www.compositesw.com/data-virtualization/data-caches/>.