What is data independence?

What is the overall architecture of a database management system?

What is a synonym for a table

What are synonyms for a columns?

What are synonyms of a primary key?

What is a synonym for a row?

What are the three types of ways to break integrity in tabular data?

What is tabular integrity

What is atomic integrity?

What is domain integrity

How does NoSQL relate to SQL?

What are the different relational queries that can be made on a database?

Name Set queries

Name some filter queries

What is a projection?

Name some renaming queries

What are synonyms for a columns?

Attribute, column, field, property

What is a synonym for a row?

A business object, an item, and entity, a document or record

What is domain integrity

Not all rows have the same type for the same column, e.g. some booleans are 0 1 or true false

Name some filter queries

Selection, projection

Name some renaming queries

Relation renaming, attribute renaming

What is a synonym for a table

A collection

What are synonyms of a primary key?

Row ID, Name

What is atomic integrity?

It's the fact that a table does not contain other tables, each cell is an atomic type.

Name Set queries

Union, substraction, intersection

What is a projection?

Selecting attributes

What is the overall architecture of a database management system?

It contains a language (e.g. SQL), a model (a table), a compute layer and a storage layer.

What is tabular integrity

Som rows are missing, some columns undefined for certain rows.

What are the different relational queries that can be made on a database?

Set queries, filter queries, renaming queries, binary queries.

What is data independence?

It refers to the immunity of user applications to changes made in the definition and organization of data.

What are the three types of ways to break integrity in tabular data?

Tabular, atomic and domain integrity.

How does NoSQL relate to SQL?

SQL does not allow any of the breaks in integrity, whereas NoSQL encompasses the rest.

Name some binary queries

What is the cartesian product?

What are the ways to break the consistency of a table?

What is database normalization?

What are the requirements for a 1NF?

What are the requirements for a table to be in second normal form?

What does dependency mean in the context of a table?

What is partial dependency?

What are the conditions for a table to be in 3rd normal form?

What is transitive dependency?

What is Boyce-Codd Normal Form (BCNF)?

What is the 4th normal form?

What is multi-valued dependencies?

What is the 5th normal form?

What is join dependency?

What is the difference between the data manipulation language and a data definition language?

What is database normalization?

Database normalization is the process of structuring a relational database in accordance with a series of so-called normal forms in order to reduce data redundancy and improve data integrity.

What are the requirements for a table to be in second normal form?

It should be in 1st normal form and it should not have any partial dependencies

What does dependency mean in the context of a table?

The value of a given row determines the value of the primary key

What are the conditions for a table to be in 3rd normal form?

It should be in 2nd normal form and not have transitive dependency

What is transitive dependency?

The attribute depends on some other non primary attribute.

What is the 4th normal form?

Must be in 3rd normal form and should not contain multi-valued dependencies

What is the 5th normal form?

Should be in 4th normal form and should not have a join dependency

What is join dependency?

It is a relationship that allows you reconstruct the data if the tables are separated. If a join dependency does not exist then either data is lost or new entries are created.

What are the ways to break the consistency of a table?

Update anomaly, delete anomaly, insert anomaly

What are the requirements for a 1NF?

4 rules:
1. Each column should contain atomic values.
2. A column should contain values that are of the same type.
3. Each column should have the same name
4. Order in which data is stored does not matter.

What is partial dependency?

It occurs when a primary key determines the outcome of another attribute or set of attributes. E.g. teacher id depends strictly on subject id.

What is Boyce-Codd Normal Form (BCNF)?

It must be in the third normal form and a prime attribute must not depend on a non-prime attribute

What is multi-valued dependencies?

* For A derives B, for a single value of A, more than one value of B exists.
* The table must have at least three columns
* For this table with A,B,C columns, B and C should be independent

What is the cartesian product?

It includes all possible combinations of rows and columns

Cartesian product

$$T = R \times S$$

R		
A	**B**	**C**
string	integer	boolean
foo	1	true
bar	2	false

S	
D	**E**
string	integer
foo	1
bar	2
foo	3

T		
A	**B**	**C**
string	integer	boolean
foo	1	true
foo	1	true
foo	1	true
bar	2	false
bar	2	false
bar	2	false

Name some binary queries

Cartesian product, natural product, theta join

What is the difference between the data manipulation language and a data definition language?

DML: query, insert & remove rows
DDL: create table/schema, drop it

What is a database schema?

What are the conditions that a transaction must fulfill if the database should remain healthy?

What is atomicity?

What defines the consistency of a transaction?

What is isolation in the context of database transactions?

What is durability in the context of database transactions?

What are the new sets of criteria that databases should comply to?

What is atomic consistency according to this new paradigm?

What does availability mean in the context of modern data transactions?

What does partition tolerance in the context of modern data transactions?

What are the three Vs in Big data?

What are the prefixes of the international systems of units?

What are the difference shapes data can adopt?

What are the key determinants of data velocity?

What is big data?

What is the main difference between SQL and NoSQL databases?

What defines the consistency of a transaction?

after a transaction, the database is in a consistent state again (i.e. any given database transaction must change affected data only in allowed ways)

What is atomicity?

Either the entire transaction is applied or none of it. (rollback)

What are the conditions that a transaction must fulfill if the database should remain healthy?

ACID: Atomicity, consistency, isolation and durability

What is a database schema?

It is its structure described in a formal language supported by the database management system (DBMS)

What is atomic consistency according to this new paradigm?

All nodes see the same data.

What are the new sets of criteria that databases should comply to?

The CAP theorem: consistency, availability and partition tolerance

What is durability in the context of database transactions?

The updates may mot dissapear again

What is isolation in the context of database transactions?

In database systems, **isolation** determines how transaction integrity is visible to other users and systems.

What are the three Vs in Big data?

Volume, variety and velocity

What does partition tolerance in the context of modern data transactions?

The database should continue to function even if the network gets partitioned.

What does availability mean in the context of modern data transactions?

It should be possible to query the database at all times.

What are the prefixes of the international systems of units?

kilo = 3
mega = 6
giga = 9
tera = 12
peta = 15
exa = 18
zetta = 21
yotta = 24

What is the main difference between SQL and NoSQL databases?

The first, second and third normal forms need not be respected. Heterogenous, nested and denormalized data are allowed.

What is big data?

It's a portfolio of technologies that were designed to store, manage and analyze data that is too large to fit on a single machine while accomodating for the issue of growing discrepancy between capacity, throughput and latency.

What are the key determinants of data velocity?

Capacity, throughput and latency

What are the difference shapes data can adopt?

Tables, trees, graphs, cubes, text,

What are the key differences in terms of the properties that need to be respected by transactions in SQL vs. NoSQL databases?

What is eventual consistency?

What are the elements included in "The stack"?

What are some storage systems?

What are some examples of encodings?

Name some examples of syntaxes.

What are some examples of data models?

What is a schema? and what is data validation in the stack?

Name some examples of processing layers in the stack.

What are ways to index data in the stack?

What are the difference data stores in the stack?

What are the ways to conduct queries on the stack?

Name some examples of interaces of the stack

What is a file made of?

How can distributed systems be scaled? What does this lead to?

What is object storage?

What are some storage systems?

Local filesystem, NFS, GFS, HDFS, S3, Azure Blob Storage

What are some examples of encodings?

ASCII, ISO-8859-1, UTF-8, BSON

Name some examples of syntaxes.

Text, CSV, XML, JSON, RDF/XML, Turtle, XRBL

What are some examples of data models?

Tables: relational model
Trees: XML Infoset, XDM
Graphs: RDF
Cubes: OLAP

What is a schema? and what is data validation in the stack?

The term "schema" refers to the organization of data as a blueprint of how the database is constructed (divided into database tables in the case of relational databases). The formal definition of a database schema is a set of formulas (sentences) called integrity constraints imposed on a database.

Name some examples of processing layers in the stack.

Two-phase processing: MapReduce
DAG-driven processing: Tez, Spark, Flink, Ray
Elastic computing: EC2

What are ways to index data in the stack?

Key-value stores, hash indices, B-trees, Geographical indices, spatial indices

What are the elements included in "The stack"?

| User interfaces |
| Querying |
| Data stores |
| Indexing |
| Processing |
| Validation |
| Data models |
| Syntax |
| Encoding |
| Storage |

What is eventual consistency?

Eventual consistency is a consistency model used in distributed computing to achieve high availability that informally guarantees that, if no new updates are made to a given data item, eventually all accesses to that item will return the last updated value.

What are the key differences in terms of the properties that need to be respected by transactions in SQL vs. NoSQL databases?

SQL: ACID (atomicity, consistency, isolation, durability).
NoSQL: CAP (atomic consistency, availability, partition tolerance and eventual consistency)

What is a file made of?

Content and metadata

Name some examples of interaces of the stack

Excel, Access, Tableau, Qlikview, BI tools

What is object storage?

It contains:
+ Black-box objects
+ Flat and global key-value model
+ Flexible metadata
+ Commodity hardware

What are the ways to conduct queries on the stack?

SQL, XQuery, JSONiq, N1QL, MDX, SPARQL, REST APIs.

How can distributed systems be scaled?
What does this lead to?

1. The hierarchical filesystem is thrown away
2. The metadata is made flexible
3. The datamodel is made trivial
4. We use commodity hardware
This leads to Object storage

What are the difference data stores in the stack?

RDBMS: relational database management system
MongoDB
CouchBase
ElasticSearch
Hive
HBase
MarkLogi
Cassandra

What is scaling up vs scaling out?

How do prices compare in scale up compared to scale out?

What is the network bandwidth for a server?

What is the amazon S3 model?

What do the 9s in a service level agreement imply in terms of yearly downtime?

What are the differnet parts of a URI?

What are the methods in HTTP?

What are the HTTP protocol? Give an example

What is a feature of a GET request?

A PUT request is idempotent. What does it mean?

What is a request that is more generic and that can have wider side effects?

To what structures can PUT, DELETE and GET be applied through the Amazon S3 API?

What are the different types of faults that can happen?

What are the advantages of having different regions?

How do companies deal with availability?

How do Amazon and Microsoft compare in their storage offerings?

What is the amazon S3 model?

Each bucket is assigned an ID, and each object within that bucket is assigned an object ID. Each object can max be 5 TB.

What are the differnet parts of a URI?

http://www.mywebsite.ch/api/collection/foo/object/bar?id=foobar#head
In order: scheme, authority, path, query, fragment

What are the methods in HTTP?

GET
PUT
DELETE
POST

What is the network bandwidth for a server?

1-100 GB/s for a server

What do the 9s in a service level agreement imply in terms of yearly downtime?

SLA	Outage
99%	4 days/year
99.9%	9 hours/year
99.99%	53 minutes/year
99.999%	6 minutes/year
99.9999%	32 seconds/year
99.99999%	4 seconds/year

How do prices compare in scale up compared to scale out?

Scale up
Scale out

61

What is scaling up vs scaling out?

...chines
...chines

What is a request that is more generic and that can have wider side effects?

POST request.

To what structures can PUT, DELETE and GET be applied through the Amazon S3 API?

Buckets and objects

How do companies deal with availability?

price differently, lower availability also higher latency

A PUT request is idempotent. What does it mean?

An idempotent method means that the result of a successful performed request is independent of the number of times it is executed.
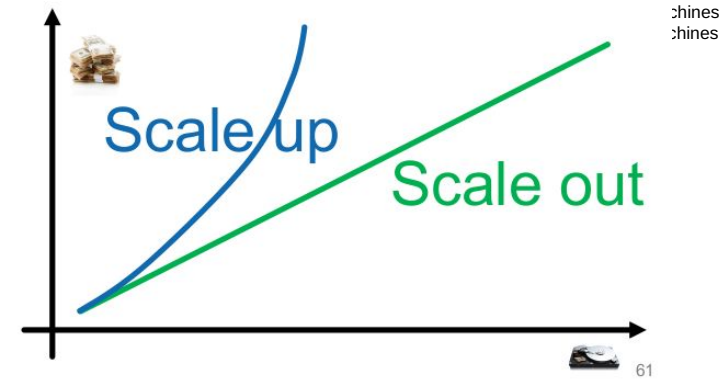
What are the advantages of having different regions?

Optimize latency and improve resilience to natural catastrophes

How do Amazon and Microsoft compare in their storage offerings?

	S3	Azure
Object ID	Bucket + Object	Account + Container + Blob
Object API	Blackbox	Block/Append/Page
Limit	5 TB	4.78 TB (block) 195 GB (append) 8 TB (page)

125

What is a feature of a GET request?

It is side-effect free, i.e. GET should not change the state of the server. In other words, they should not have side effects, beyond relatively harmless effects such as logging, web caching, the serving of banner advertisements or incrementing a web counter.

What are the different types of faults that can happen?

Local (node failure)
Regional (natural catastrophe)

What are the HTTP protocol? Give an example

Two things:
Request - method/URI [Header] [Body]
Response - status code [Header] [Body]

```
GET /index.html HTTP/1.1
Host: www.example.com
```

```
HTTP/1.1 200 OK
Date: Tue, 25 Sep 2018 09:48:34 GMT
Content-Type: text/html; charset=UTF-8
Content-Length: 138
Last-Modified: Wed, 08 Jan 2003 23:11:55 GMT
Server: Apache/1.3.3.7 (Unix) (Red-Hat/Linux)
ETag: "3f80f-1b6-3e1cb03b"
Accept-Ranges: bytes
Connection: close

<html> <head> <title>An Example Page</title>
</head> <body> Hello World, this is a very si...
HTML document. </body> </html>
```
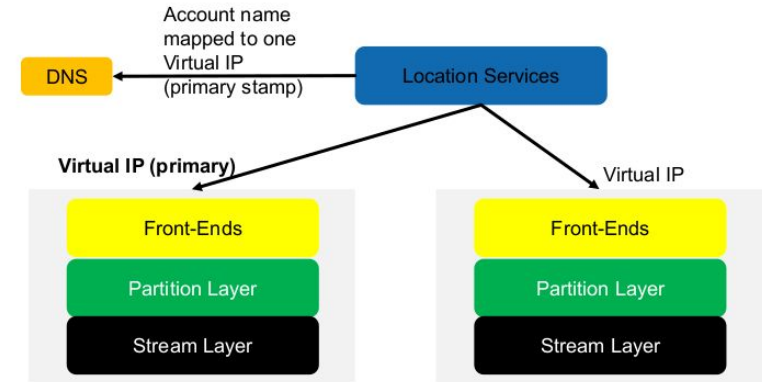
How does the storage stamp look like on Azure?

How much storage is contained in a storage stamp?

How is intra-stamp storage replicated? How is inter-stamp storage replicated?

What do location services provide? Give examples of regional location services

What goes to and from the DNS to the location services? What does to the stamp from the location services?

What is the Amazon mindset of building their AWS platform?

What is the Azure mindset to build their platform?

Overall, how do we scale out?

What are the two use cases of Big Data?

What are the technologies used for storing billions of sub tb files vs. millions of pb files?

What was the first distributed file system?

What is the inevitability to consider when designing a distributed file system?

What leads up to fault tolerance?

What are the different file read models of sub tb files vs. millions of pb files? How about the file update model?

What is the file update model suitable for?

What is the top requirement in a distributed file system?

What do location services provide? Give examples of regional location services

They map the account name to one virtual IP to the DNS. They redirect the requests to a stamp via a primary virtual IP and other stamps elsewhere via other virtual IPs to other regions.

Examples include north america, europe and asia

How is intra-stamp storage replicated? How is inter-stamp storage replicated?

intra - Synchronously
inter - Asynchronously

How much storage is contained in a storage stamp?

10-20 racks*18 storage nodes/rack (30PB) kept below 70/80% storage capacity

How does the storage stamp look like on Azure?

Overall, how do we scale out?

1. Simplify the model
2. Buy cheap hardware
3. Remove schemas

What are the two use cases of Big Data?

A huge amount of large files (which can fit on a single machine) or a large amount of huge files, which cannot fit on a single machine.

i.e.
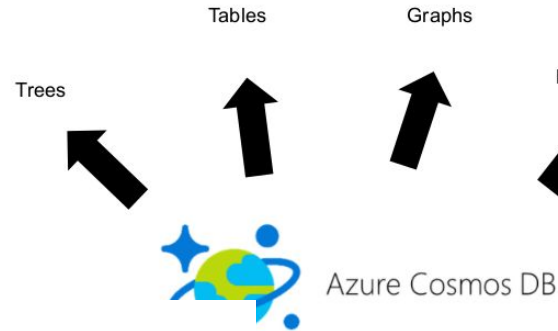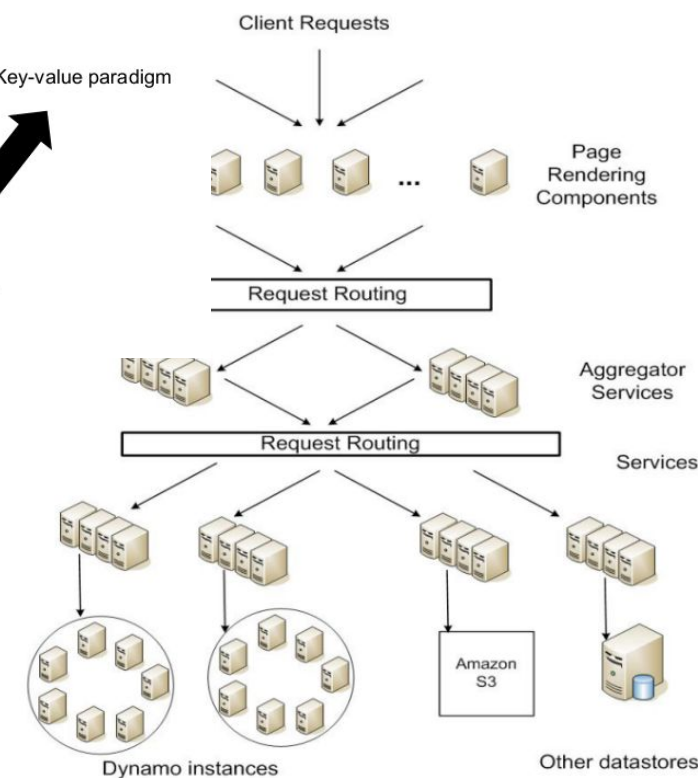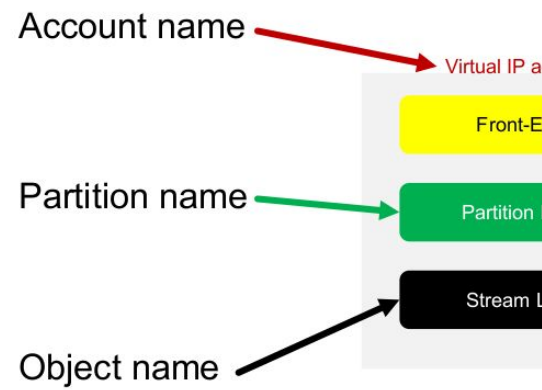
Billions of TB files (object storage)

Millions of PB files (file storage)

What is the Azure mindset to build their platform?

What is the Amazon mindset of building their AWS platform?

What goes to and from the DNS to the location services? What does to the stamp from the location services?

...ame
...y stamp's VIP
...VIP first): partition + object
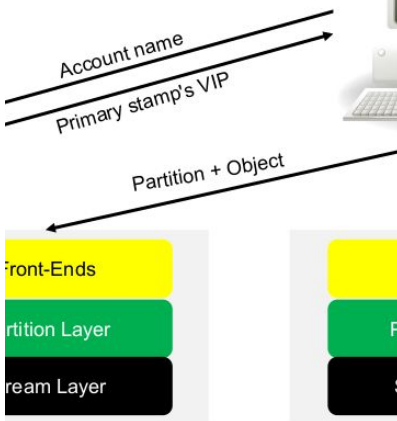
Billions of TB files

Object Storage

vs.

File Storage

Millions of PB files

What leads up to fault tolerance?

Monitoring, error detection, and automatic recovery

What are the different file read models of sub tb files vs. millions of pb files? How about the file update model?

random access vs. scanning of the file/appending to the file

What is the inevitability to consider when designing a distributed file system?

The fact that hardware will fail.

What was the first distributed file system?

Google FS

What is the top requirement in a distributed file system?

Throughput is the top priority, secondary is latency

What are the technologies used for storing billions of sub tb files vs. millions of pb files?

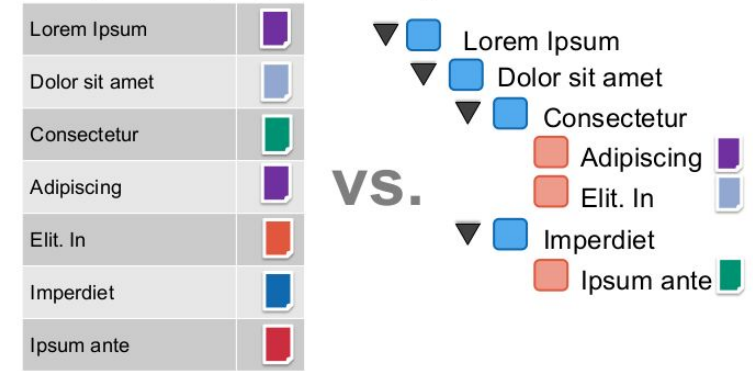Key-value model + object storage
File system + block storage

What is the file update model suitable for?

Sensor data, logs, and intermediate data

When was hadoop started?	What are the three core elements of Hadoop?	What was Hadoop inspired by?	What is the logical model of file systems vs. key-value model?
What is the difference in physical storage in file systems vs key-value models?	What is the elementary unit of HDFS vs. GFS?	How to the blocks and the hierarchy relate?	Why are blocks used in file systems? Give two reasons
What is involved in choosing the right block size?	How are computed related to each other in a HDFS?	What does this master-slave architecture provide?	What is in the namenode?
			What does a datanode contain?
What does a datanode provide as an advantage?	What are the different communication protocols in a HDFS?	What are the elements of the client protocol?	

What is the logical model of file systems vs. key-value model?

There is a hierarchical component to the file system model

What was Hadoop inspired by?

GFS, MapReduce and BigTable

What are the three core elements of Hadoop?

A distributed file system (HDFS)
MapReduce
Wide column store (HBase)

When was hadoop started?

2006

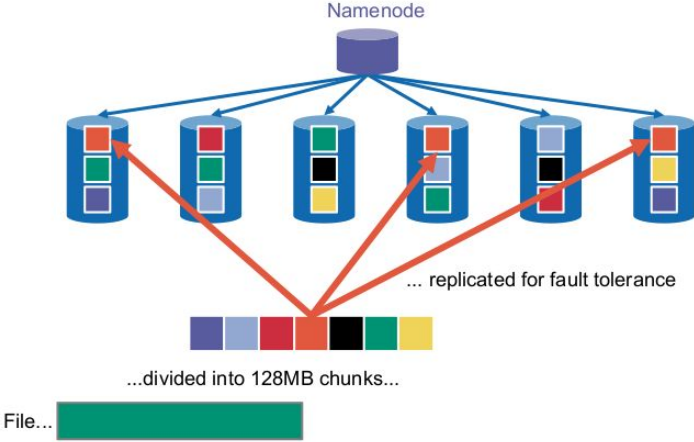**Why are blocks used in file systems? Give two reasons**

1. Files are larger than a disk (PBs)
2. Provide a simpler level of abstraction

What is involved in choosing the right block size?

Too small, annoying for large amounts of blocks. Too big: doesn't fit the data cluster's capacity.
For individual machines, one usually has 4kb of blocks for simple file systems. For relational databases, anywhere from 4kb to 32 kb and in distributed file systems we have anywhere between 64 and 128 MB blocks.

How are computed related to each other in a HDFS?

Master-slave architecture. With namenodes and datanodes.
The information in the datanode is replicated for fault tolerance.

How to the blocks and the hierarchy relate?

What is the elementary unit of HDFS vs. GFS?

unk for GFS

What is the difference in physical storage in file systems vs key-value models?

Object storage are used for sub tb case and block storage for sub pb

---

What does a datanode provide as an advantage?
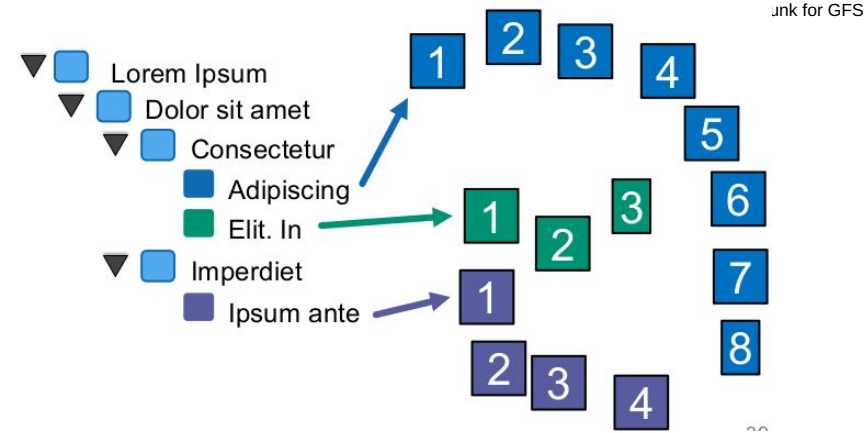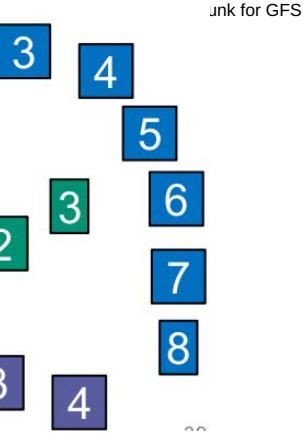
Close to the hardware, which facilitates disk failure detection

What are the elements of the client protocol?

What does a datanode contain?

Data, in the form of blocks, on the disk

What are the different communication protocols in a HDFS?

What is in the namenode?

File namespace + access control and file to block mapping and block locations

What does this master-slave architecture provide?

This provides the ability for the namenode to be accessed concurrently.

Client sends to namenode metadata operations. Namenode sends back datanode location block IDs

Client protocol between namenode and client. DataTransfer protocol between client and datanode and file writing pipeline and replication for communication between data nodes.

Who initiates the connection between a namenode and a datanode?

How does the data transfer protocol work?

What are some examples of metadata functionalities?

What happens if a client wants to read a file?

What happens if a client wants to write a file?

What is te default number of replicas per block?

What is often the minimum distanc between two blocks?

How are replicas placed?

If rep 1 and 2 were on same rack what would happen?

Why 2 and 3 on same rack?

What is the problem with this slave master architecture?

What do you want to have persist of the namenode?

Why is the persistence of the block locations not required?

What, in addition to the namespace file, do you also want to have persist while master is down?

How is the namenode restored?

How long does the restoring of a namenode require in naive conditions?

What happens if a client wants to read a file?

Client asks for the file to the namenode
The client gets block locations, multiple datanodes for each block sorted by distance.
The datanodes send their inputstream to the client

What are some examples of metadata functionalities?

Create directory, delete directory, write file, append to file, read file, delete file.

How does the data transfer protocol work?

Client connects directly to the datanodes and datablocks are send to a datanode which then replicates through pipelining to multiple other datanodes.

Who initiates the connection between a namenode and a datanode?

The datanode, needs to send registration, heartbeat, blockreport and blockreceived every interval of time.

How are replicas placed?

1. Same node as client or random, rack A
2. A node in a different rack B
3. A node in the same rack B
4. Random, at most one replica per node and two per rack

What is often the minimum distanc between two blocks?

4

What is te default number of replicas per block?

3

What happens if a client wants to write a file?

Client sends create instruction.
Datanodes for the first block is received.
The writing pipeline is organized
Send the data over
Ack signal is sent once done (acknowledgment)
The namenode sends the datanodes for the second block
The writing pipeline is organized
Send the data over
Ack signal is sent once done (acknowledgment)

What is the problem with this slave master architecture?

Master represents a single point of failure

Why 2 and 3 on same rack?

Better throughput

If rep 1 and 2 were on same rack what would happen?

If failure, only one copy left.

What, in addition to the namespace file, do you also want to have persist while master is down?

The edit log to play through once restored.

Why is the persistence of the block locations not required?

Can be provided via heartbeat protocol to master and data reconstituted this way.

What do you want to have persist of the namenode?

The file namespace and the block id mapping, not the locations.

How long does the restoring of a namenode require in naive conditions?

30 minutes

How is the namenode restored?

After 30 minutes. Namespace file restored, then edit log is played through, then reconstitute locations for blocks.

What are the tricks to reduce downtime after namenode knockout?

How is HDFS used?

Hoe can the populating of the HFDS be achieved?

How are objects identified in azure blob storage?

How are objects identified in S3?

What kind of objects can you create in Azure Blob Storage?

What kind of objects can you create in Amazon S3?

What changes when doing NoSQL?

When is normalization useful and when is denormalization useful?

What is data denormalized to first normal form?

What property do tuples have when nestedness is allowed?

What does a homogeneous collection of flat items in a relational database become in the NoSQL world?

What are semi structured documents?

Who issues the standards for XML and JSON

What is the advantage of JSON and XML?

What is well-formedness?

How are objects identified in azure blob storage?

Account ID + Container ID + Blob ID

How are objects identified in S3?

Bucket ID + Object ID

When is normalization useful and when is denormalization useful?

When data is write intensive, normalized data is useful to avoid update anomalies. When the data is read-intensive, the data needs to be highly denormalized because joins need to be avoided (intensive operation, requires a lot of processing and introduces latency).

What are semi structured documents?
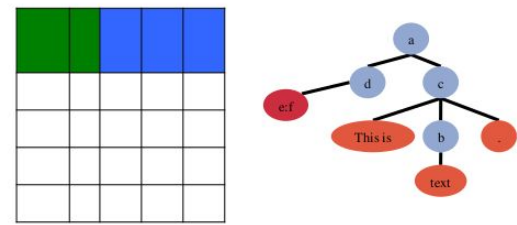

Structured   Semi-structured   Unstructured

Hoe can the populating of the HFDS be achieved?

Through collection, aggregation and moving of log data into HDFS using Apache Flume, Sqoop

What changes when doing NoSQL?

The data is denormalized: 1st, 2nd, 3rd, B-C normal form etc is discarded

What does a homogeneous collection of flat items in a relational database become in the NoSQL world?

collection of arborescent items in a
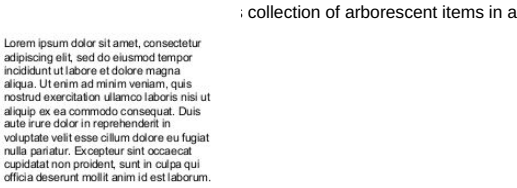
What is well-formedness?

You check whether a statement belongs to the language, as each syntax has its own language.

How is HDFS used?

Via shell, just like posix

What kind of objects can you create in Amazon S3?

Blackbox objects

What property do tuples have when nestedness is allowed?

Avoid repetition and data takes up less space.

What is the advantage of JSON and XML?

They are understood both by humans and machines.

What are the tricks to reduce downtime after namenode knockout?

1. Make checkpoints to avoid having to play through the edit log unecessarily
2. Have standby namenodes
3. Have federated DFS to have multiple namenodes that have all the information necessary

What kind of objects can you create in Azure Blob Storage?

3 types of blobs: BlockBlob, PageBlob, AppendBlob

What is data denormalized to first normal form?

A collection tuples.

Who issues the standards for XML and JSON

XML: W3C
JSON: ECMA

What are the six types of elements in a JSON string?

Is the string { "foo" : "bar", "foo" : "bar2" } well formed?

Is the string { [1] : "bar", 2 : "bar2" } well-formed?

What are tags in XML? Given examples of three different types of tags

What is an XML attribute? Give an example

How is text inserted in XML?

What are the remaining elements that are not tags, attributes or texts?

What is the first line of an XML file?

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>

Is this statement well-formed?

Is this statement well formed?
<?xml version="1.0" encoding="UTF-8"?>

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>
text
text

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>
<foo>
text text
</foo>

This this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>
<foo >
</foo>

How does a document type declaration look like?

What are tags in XML? Given examples of three different types of tags

Opening tag: <foo>
Closing tag: </foo>
Empty tag: <foo/>=

Is the string { [1] : "bar", 2 : "bar2" } well-formed?

No, keys are not strings

Is the string { "foo" : "bar", "foo" : "bar2" } well formed?

No, twice the same key

What are the six types of elements in a JSON string?

Strings, numbers, booleans, null, arrays and objects.

What is the first line of an XML file?

Text declaration:
<?xml version="1.0" encoding="UTF-8"?>

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>

Yes

What are the remaining elements that are not tags, attributes or texts?

Comments and processing instructution
<!-- This is a comment -->
<?myapp do whatever ?>

How is text inserted in XML?

<a>This is text

What is an XML attribute? Give an example

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>
text
text

No

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>

Is this statement well formed?
<?xml version="1.0" encoding="UTF-8"?>

Yes

Is this statement well-formed?

Yes

How does a document type declaration look like?

<?xml version="1.0"?>
<!DOCTYPE document>
<document>
Lorem ipsum dolor sit amet, consectetur adipiscing
elit, sed do eiusmod tempor incididunt ut labore et
dolore magna aliqua.
</document>

This this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>
<foo >
</foo>

No

Is this statement well-formed?
<?xml version="1.0" encoding="UTF-8"?>
<foo>
text text
</foo>

Yes

What appears where in an XML files?

Is this statement well-formed?

Is this statement well-formed?

Is this statement well-formed?
<a>

Is this statement well-formed?
<a>

Is this statement well-formed?
<a>1 < 2

Is this statement well-formed?
<a>1 < 2

What are entity references? Name all of them.

What are character references? name an example

What are some invalid XML names?

What are some valid XML names?

Is whitespace important in XML?

What are the ASCII characters allowed in XML names?

What is an example of a namespace?

Is this statement well-formed?
<a>

No

Is this statement well-formed?
<a>1 < 2

Yes

What are some invalid XML names?

<1234/>
<a

Is whitespace important in XML?

Only for readability (can be used for pretty print, no functional significance.)

Is this statement well-formed?

Yes

Is this statement well-formed?
<a>1 < 2

No

What are character references? name an example

Character references refer to characters not usually found on keyboards:
π = π

What are some valid XML names?

Is this statement well-formed?

No

Is this statement well-formed?
<a>

Yes.

What are entity references? Name all of them.

< = <
> = >
' = '
" = "
& = &

What is an example of a namespace?

MathML

What appears where in an XML files?

	Top-Level	Between Element Tags	
Elements	once		
Attributes			
Text			

What are the ASCII characters allowed in XML names?

Control characters											
Control characters											
SP	!	"	#	$	%	&	'	(	)	*	+
0	1	2	3	4	5	6	7	8	9	.	;
@	A	B	C	D	E	F	G	H	I	J	K
P	Q	R	S	T	U	V	W	X	Y	Z	[
`	a	b	c	d	e	f	g	h	i	j	k
p	q	r	s	t	u	v	w	x	y	z	{

Allowed anywhere in name
Allowed but not at start
not allowed