

Темы к зачету

1. ACID - properties of database transactions
2. Transaction isolation levels
3. Pessimistic vs optimistic locking. Lost update problem
4. SQL vs NoSQL. RDBMS problems. ORM (Object-relational mapping)
5. NoSQL properties (schemaless, aggregate orientation, transactions, ...)
6. Types of NoSQL databases
7. Column-family store
8. Cassandra data model: primary key, partition and clustering keys
9. Cassandra: Bloom filter. Queries
10. Distribution Models. Consistency problem
11. Distribution Models. Sharding
12. Distribution Models. Replication
13. Distributed systems. Consistency models: eventual consistency, monotonic reads, read your writes, strong consistency
14. Distributed systems. Consistency guarantee
15. Distributed systems. MongoDB and Cassandra parameters for consistency guarantee
16. Consistency Models: Strict consistency, Linearizability, Sequential consistency, Causal consistency, Serializable consistency
17. Distributed systems. CAP theorem. BASE.
18. CAP theorem with SQL and NoSQL DBs
19. Polyglot Persistence
20. MapReduce. Data locality. Phases
21. MapReduce. Standard algorithm (Word count, Inverted index, Top N)
22. MapReduce. Map/Reduce/Combine functions requirements
23. MapReduce alternatives
24. RDBMS vs NoSQL vs MapReduce
25. Hadoop architecture and ecosystem
26. Hadoop. HDFS
27. Hadoop. Yarn
28. Hadoop Application lifecycle
29. Hadoop (MR) limitations. Apache Spark
30. Apache Spark DAG
31. Apache Spark RDD

32. Apache Spark RDD transformations and actions: filter, map, flatMap, distinct, reduce, groupBy, collect
33. Spark Architecture

Lit:

- NoSQL Distilled - <https://martinfowler.com/books/nosql.html>
- NoSql, MapReduce, CAP - <https://drive.google.com/open?id=0B6VW3ge7g2HXR0VORFRpaFZiTDA>
- Untangling Apache Hadoop YARN - <http://blog.cloudera.com/blog/2015/09/untangling-apache-hadoop-yarn-part-1/>
- Apache Spark: core concepts, architecture and internals - <http://datastrophic.io/core-concepts-architecture-and-internals-of-apache-spar>
[k](#)