

Paper Reading 2019

File Systems

01. The Google File System.....	1
02. The Hadoop Distributed File System.....	16
03. Ceph: A Scalable, High-Performance Distributed File System.....	26
04. PolarFS: An Ultra-low Latency and Failure Resilient Distributed File System for Shared Storage Cloud Database.....	40
05. Tachyon: Reliable, Memory Speed Storage for Cluster Computing Frameworks.....	54

Timestamp & Transaction

06. Time, Clocks, and the Ordering of Events in a Distributed System.....	69
07. Logical Physical Clocks and Consistent Snapshots in Globally Distributed Databases...77	
08. Consensus on Transaction Commit.....	90
09. Calvin: Fast Distributed Transactions for Partitioned Database Systems.....	127
10. Large-scale Incremental Processing Using Distributed Transactions and Notifications..	139
12. A Critique of ANSI SQL Isolation Levels.....	153
13. Virtual Time and Global States of Distributed Systems.....	166

Databases

14. Spanner: Google's Globally-Distributed Database.....	181
15. Spanner: Becoming a SQL System.....	195
16. F1: A Distributed SQL Database That Scales.....	208
17. Online, Asynchronous Schema Change in F1.....	220
18. Bigtable: A Distributed Storage System for Structured Data.....	232
19. Mesa: GeoReplicated, Near RealTime, Scalable Data Warehousing.....	246
20. Dynamo: Amazon's Highly Available Key-value Store.....	258
21. Bitcoin: A Peer-to-Peer Electronic Cash System.....	274
22. Epidemic Algorithms for Replicated Database Maintenance.....	284
23. Cassandra - A Decentralized Structured Storage System.....	311
24. Amazon Aurora: Design Considerations for High Throughput Cloud-Native Relational Databases.....	317
25. Megastore: Providing Scalable, Highly Available Storage for Interactive Services.....	330
26. The VoltDB Main Memory DBMS.....	341
27. Anna: A KVS For Any Scale.....	348

Storage Engine

28. XORing Elephants: Novel Erasure Codes for Big Data.....	360
29. WiscKey: Separating Keys from Values in SSD-conscious Storage.....	372
30. An Efficient Design and Implementation of LSM-Tree based Key-Value Store on Open-Channel SSD.....	389
31. SuRF: Practical Range Query Filtering with Fast Succinct Tries.....	403
32. Optimizing Space Amplification in RocksDB.....	417
33. The Bw-Tree: A B-tree for New Hardware Platforms.....	426
34. B-trees, Shadowing, and Clones.....	438

Consistence

35. ZooKeeper: Wait-free coordination for Internet-scale systems.....	464
36. ZooKeeper's atomic broadcast protocol: Theory and practice.....	478
37. Zab: High-performance broadcast for primary-backup systems.....	497
38. The Chubby lock service for loosely-coupled distributed systems.....	509
39. In Search of an Understandable Consensus Algorithm.....	525
40. Paxos Made Simple.....	543
41. Paxos Made Live - An Engineering Perspective.....	557
42. The Part-Time Parliament.....	573
43. PaxosStore: High-availability Storage Made Practical in WeChat.....	606

Distribute Computation

44. MapReduce: Simplified Data Processing on Large Clusters.....	618
45. S4: Distributed Stream Computing Platform.....	631
46. Apache Hadoop YARN: Yet Another Resource Negotiator.....	639
47. Spark: Cluster Computing with Working Sets.....	655
48. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing.....	662
49. GraphX: Unifying Data-Parallel and Graph-Parallel Analytics.....	676
50. Apache Flink: Stream and Batch Processing in a Single Engine.....	688
51. Pregel: A System for Large-Scale Graph Processing.....	699
52. Dremel: Interactive Analysis of WebScale Datasets.....	710
53. Druid: A Real-time Analytical Data Store.....	720

CAP

54. Harvest, Yield, and Scalable Tolerant Systems.....	732
55. CAP Twelve years later: How the "Rules" have Changed.....	738

Garbage Collection

56. Very Concurrent Mark-&-Sweep Garbage Collection without Fine-Grain Synchronization.....	756
57. A Generational Mostly-concurrent Garbage Collector.....	755
58. Garbage-First Garbage Collection.....	767
59. The Pauseless GC Algorithm.....	779