

**Untangling Block chain: A Data Processing View of Block chain Systems**

**Abstract:**

Blockchain technologies are gaining massive momentum in the last few years. Blockchains are distributed ledgers that enable parties who do not fully trust each other to maintain a set of global states. The parties agree on the existence, values, and histories of the states. As the technology landscape is expanding rapidly, it is both important and challenging to have a firm grasp of what the core technologies have to offer, especially with respect to their data processing capabilities. In this paper, we first survey the state of the art, focusing on private blockchains (in which parties are authenticated). We analyze both in-production and research systems in four dimensions: distributed ledger, cryptography, consensus protocol, and smart contract. We then present BLOCKBENCH, a benchmarking framework for understanding performance of private blockchains against data processing workloads. We conduct a comprehensive evaluation of three major blockchain systems based on BLOCKBENCH, namely Ethereum, Parity, and Hyperledger Fabric. The results demonstrate several trade-offs in the design space, as well as big performance gaps between blockchain and database systems. Drawing from design principles of database systems, we discuss several research directions for bringing blockchain performance closer to the realm of databases.