17th Int'l Workshop on High Performance Transaction Systems, Asilomar, 10 Oct 2017
Updated Slides from 37th IEEE International Conference on Distributed Computing Systems (ICDCS), Atlanta, 6 June 2017
& from 43rd International Conference on Very Large Data Bases (VLDB), Munich, 29 August 2017

Blockchains and Databases: A New Era in Distributed Computing

C. Mohan

IBM Fellow
Distinguished Visiting Professor
Tsinghua University, Beijing

IBM Almaden Research Center, San Jose, USA

@seemohan cmohan@us.ibm.com http://bit.ly/CMwIkP

Links to Videos, PPT Slides, Bibliography & Twitter Handles @ http://bit.ly/CMbcDB

Agenda (VLDB 2017 Tutorial Version)

Goal: Educate DB people about private/permissioned blockchains (BCs) to convince them to get more involved to improve them

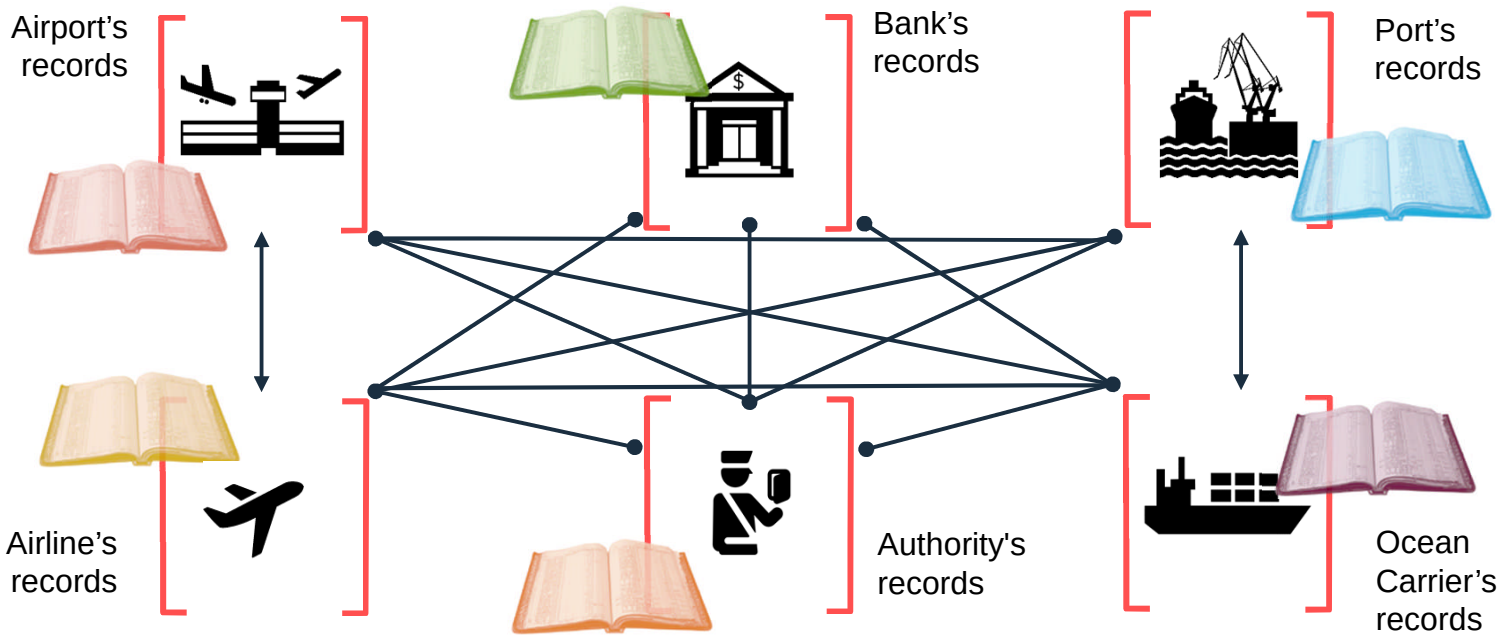
- Origin of blockchains
- Related distributed systems/databases topics
- Evolution: Private BCs, Smart Contracts, …
- Applications
- Market Scene
- Benchmarks
- Architectural Choices and Relationship to DB Replication
- Technical Details of Representative Systems:
  Enterprise Ethereum, Hyperledger Fabric, R3 Corda, BigchainDB, Sawtooth, Ripple
- Futuristic Topics

Blockchain (BC)

- Origin in digital currencies, in particular Bitcoin (Satoshi Nakamoto, 2008) – anonymity, **open/public/permissionless** environment
- Numerous organizations across the world working on various aspects of it: security, consensus, database, benchmarks, verification, …
- Banks, regulators, universities, startups, big technology companies, services companies, governments, … individually or as part of consortia
  - February 2017: First commercial deployment of BC technology by IBM and Guernsey's Northern Trust for admin of private equity fund managed by Unigestion
  - July 2017: **Hyperledger Fabric 1.0** Released
  - Hyperledger Fabric on IBM Cloud - IBM Blockchain Platform (formerly HSBN) on highly secure Linux on mainframes (System Z) with security hardware – announced August 2017
  - 10/2017: Oracle announced Blockchain Cloud Service (BCS) - **Fabric 1.0** based
- Grand View Research: Global BC Tech Market **$7.74B** by 2024
- **My focus**: **Private/Permissioned** BC Systems!

Problem Being Solved

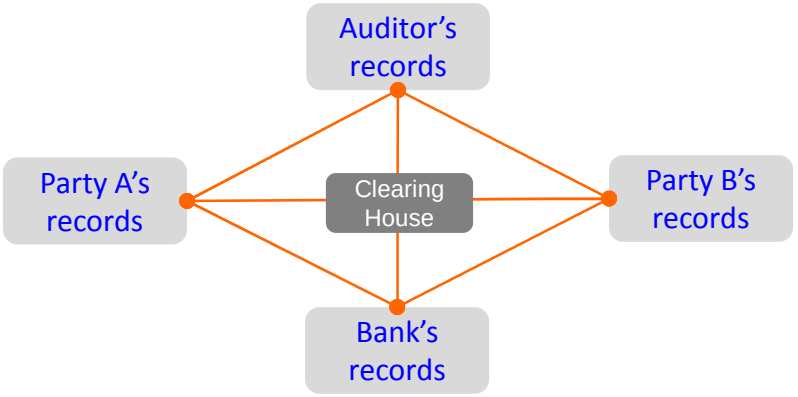
Recording of events is becoming much more complex…



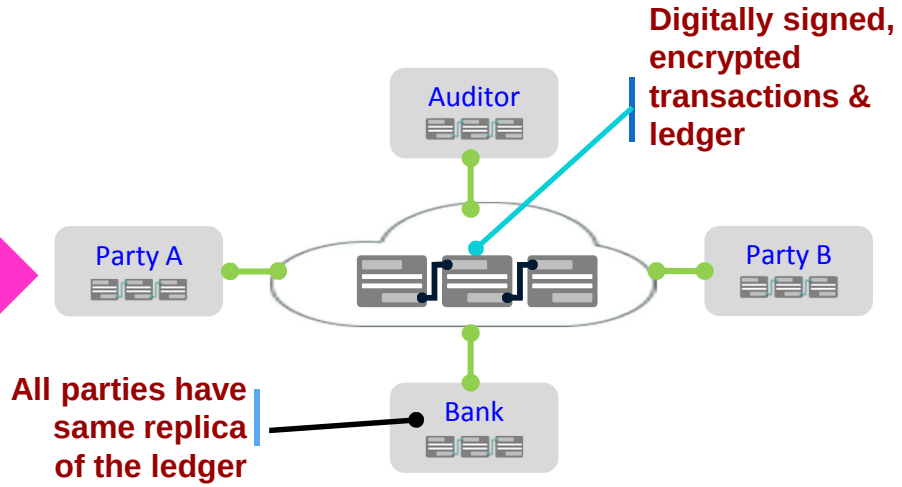
… Inefficient, expensive, vulnerable, lack of transparency

Basic Change to Business Processes

Traditional Way



Blockchain Way



… Inefficient, expensive, vulnerable

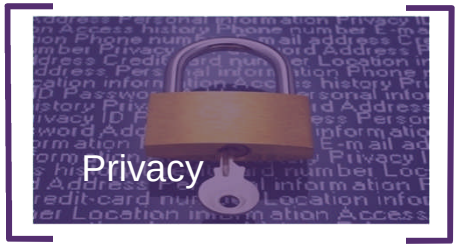
… Consensus, provenance, immutability, finality

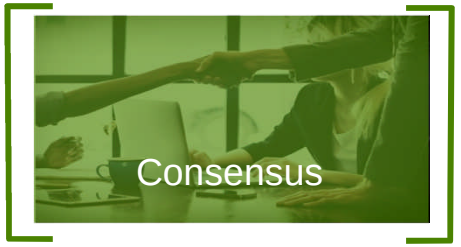
Blockchain for Business

Append-only distributed **system of record** shared across business network

Business terms embedded in transaction database & executed with transactions

Ensuring appropriate visibility; transactions are secure, authenticated & verifiable

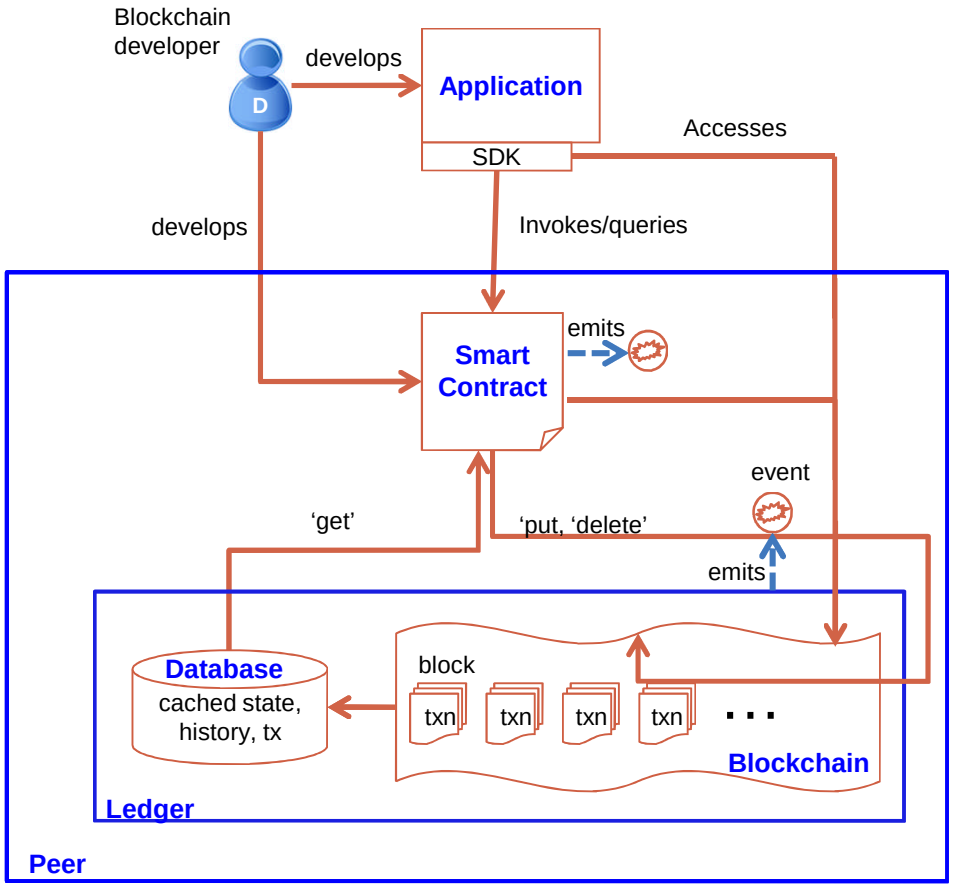




All parties agree to network verified transaction

… Broader participation, lower cost, increased efficiency

Overview of Application Flow



- Developers create application and smart contracts (chaincodes)
  - Chaincodes are deployed on the network and control the state of the ledger
  - Application handles user interface and submits transactions to the network which call chaincodes
- Network emits events on block of transactions allowing applications to integrate with other systems

Blockchain Applications

- Track provenance, ownership, relationships & lineage of assets
- Supply Chain – Food Safety (Walmart), Logistics (Maersk)
- Health Data Exchange (FDA)
- Know Your Customer
- Derivatives Processing
- Trade/Channel Finance (IGF)
- Trade Information Warehouse (DTCC)
- Post-Trade Reconciliation/Settlement
- Private Equity Fund Management (Unigestion)
- Syndicated Loans
- Diamond/Valuables Tracking and Protection – Provenance Management (Everledger)
- Cross-Border Payment, Payments for/by Unbanked Populations
- Low volume stock trading (JPX)

"Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World", Don Tapscott, Alex Tapscott, ISBN 978-1101980132.

Blockchain Companies/Consortia & Banks

New kid on the block

Enterprise Ethereum Alliance is the latest addition to a list of companies and consortiums focused on blockchain where banks serve as investors or partners

	No. of partners
Utility Settlement Coin	5
Global Payments Steering Group	6
Chain	9
Digital Asset Holdings	15
Enterprise Ethereum Alliance	30
R3	81
Ripple	90+
Hyperledger	122

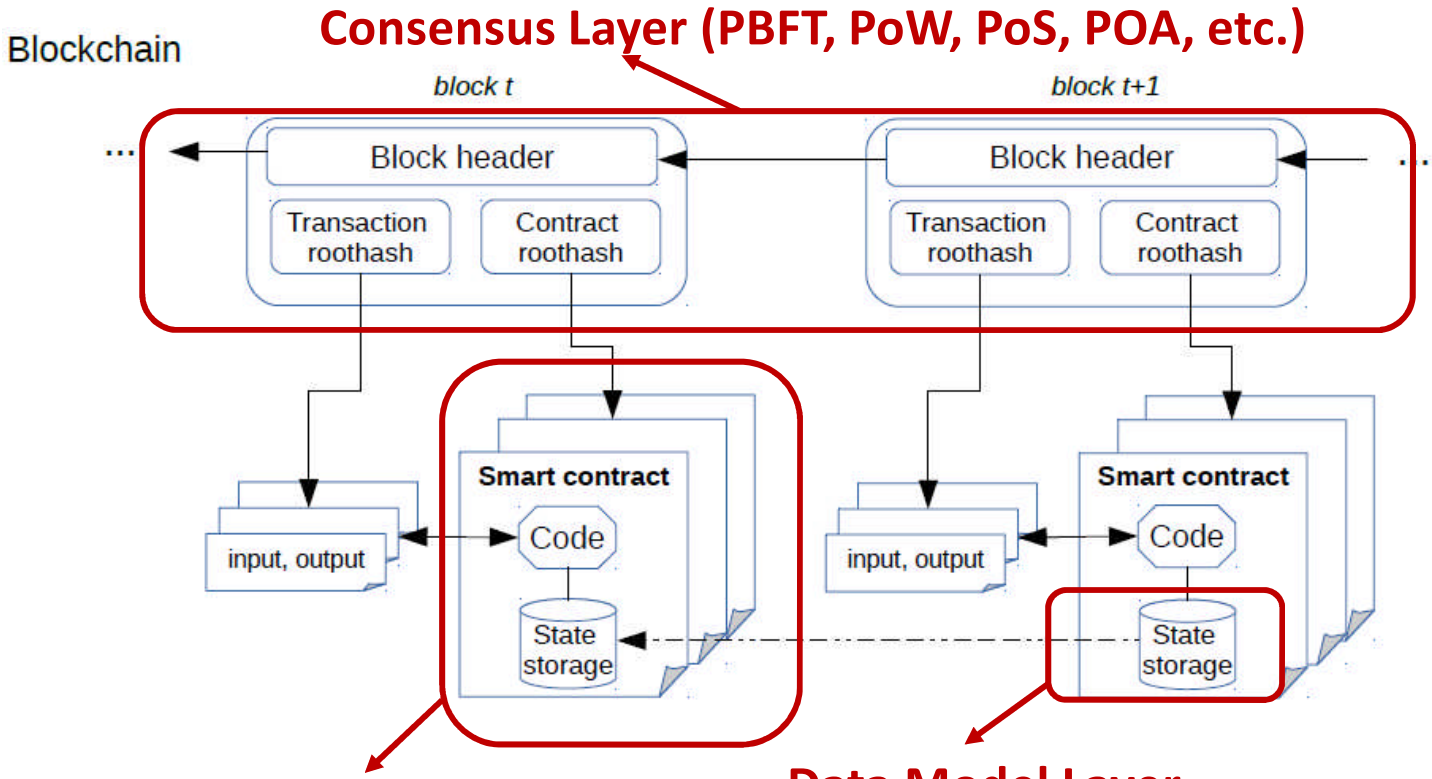
Source: Staff research

Source: American Banker 2-2017

Ongoing Industry Projects/Efforts

- 11/2016: R3 open sources Corda
- 2/2017: DTCC (Depository Trust & Clearing Corp) Selects IBM, AXONI and R3 to develop DTCC's distributed ledger solution for derivatives processing – expected to go live in early 2018
- 2/2017: Enterprise Ethereum Alliance launched with Quorum from JP Morgan being open sourced
- 3/2017: Fabric graduates, Incubation to Active
- 7/2017: V1 released
- BigchainDB (Berlin): Starts from DBMS end to add BC features
  - Uses a single RethinkDB Cluster
  - MongoDB support being added

BC Software Stack



Source: Anh Dinh, et al., SIGMOD 2017

Blockchain Architecture/Feature Choices

- Cryptocurrencies Vs Generalized Assets
- Permissionless/Public Vs Permissioned/Private
- Byzantine Vs Non-Byzantine fault model
- Consensus approach: PoW, PoA, PoET, PBFT, …
- SQL Vs NoSQL data stores
- Transactional stores Vs Non-transactional stores
- Versioned/Unversioned state database
- On-Chain Vs Off-Chain data
- Parallelism exploitation during different phases of transaction execution

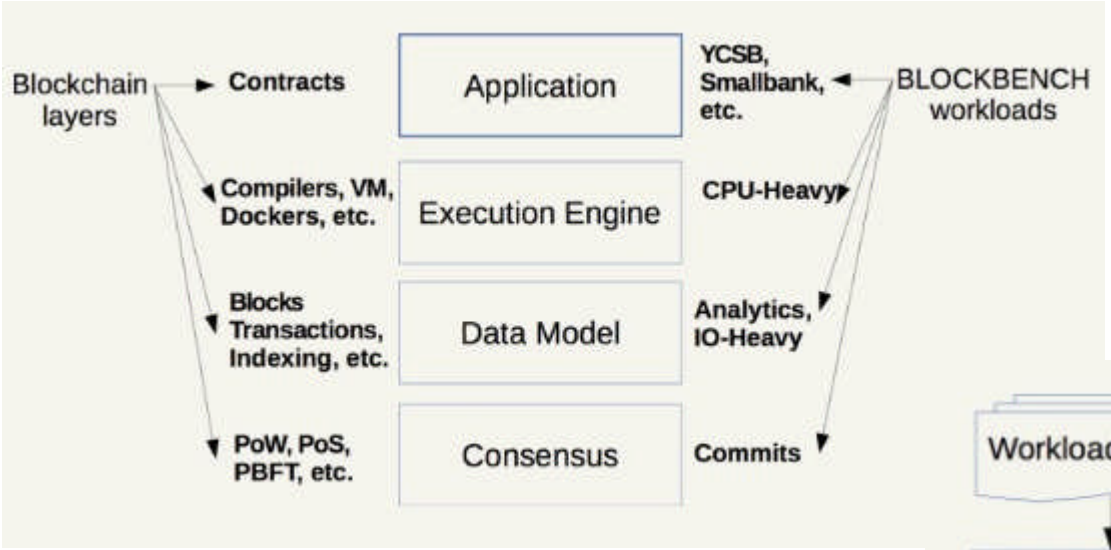
Good Survey Paper: Untangling Blockchain: A Data Processing View of Blockchain Systems, A. Dinh et al.

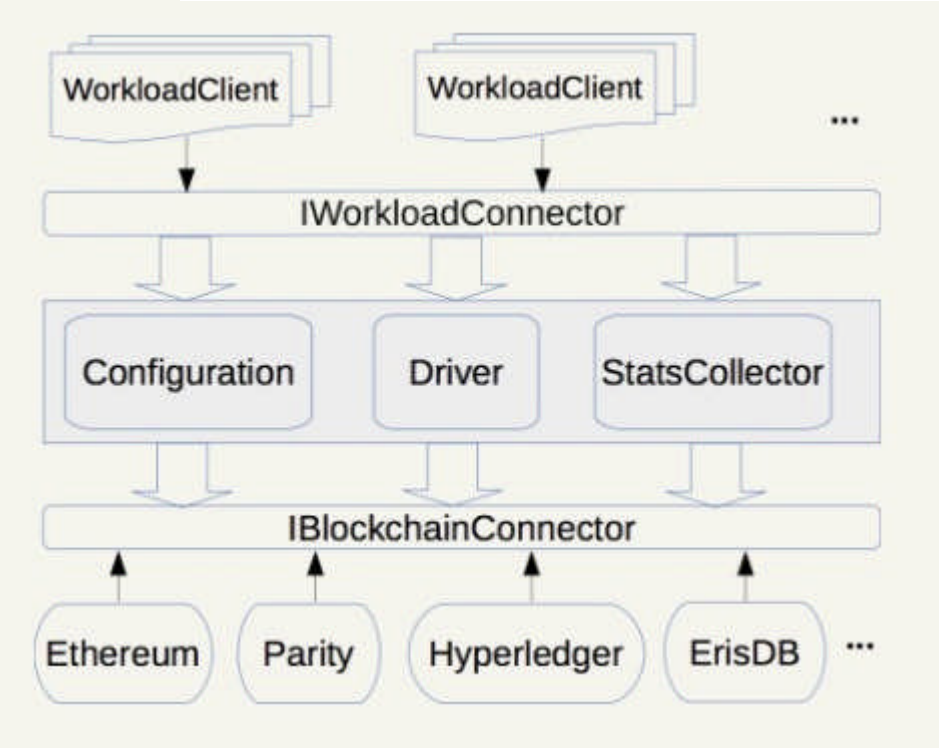
Database Replication

- Primary log replay at replica – homogeneous systems with full DB replicas, typically done for disaster recovery (DR) backup
- Log capture generates DML statements from what is logged and apply executes those statements (e.g., IBM Q Replication)
  - Can handle non-determinism and partial replicas
  - Requires dependency analysis to leverage parallelism at apply time
  - https://www.ibm.com/developerworks/data/roadmaps/qrepl-roadmap.html
- Capture DML statements as issued by application and re-execute them at replica (e.g., H-Store/VoltDB)
  - Cannot handle non-determinism
  - Typically, serial execution of transactions

Upfront (fairly random, unoptimized) ordering of transactions in blockchain systems – leads to all sorts of issues!

Benchmark Framework: BLOCKBENCH (NUS)





- OLTP & OLAP load measured
- Metrics: throughput, latency, scalability, fault tolerance, security
- Consensus methods: Ethereum (PoW), Fabric (PBFT), Parity (PoA)
- Old version of Fabric (pre-V1)
- Fabric performs better
- Fabric scales well up to 16 nodes

Source: Anh Dinh, et al., SIGMOD 2017

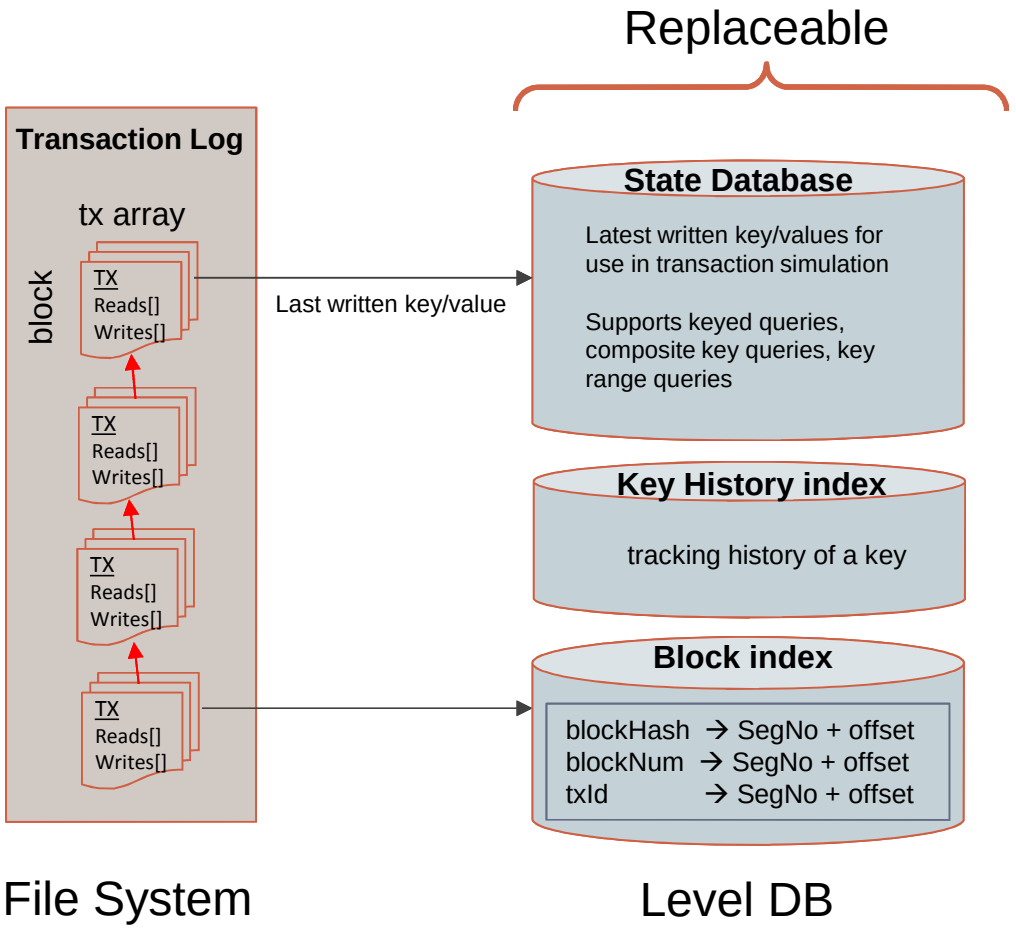
Hyperledger Fabric Project

- Initiated by IBM with IBM open source ledger contribution (Feb 2016)
  http://hyperledger-fabric.readthedocs.io/en/latest/
- Significant change in architecture from V0.6 to V1
  - Chaincode trust flexibility
  - Scalability
  - Confidentiality
  - Consensus modularity
- Used PBFT for consensus before V1 Miguel Castro, Barbara Liskov: Practical Byzantine Fault Tolerance and Proactive Recovery. ACM Trans. Comput. Syst. (TOCS), 20(4), 2002.
- Other Hyperledger Projects: Iroha, Sawtooth, Composer, …
  https://www.hyperledger.org/
- Customers doing trials: Bank of Tokyo, London Stock Exchange, Maersk, Northern Trust, Walmart

Fabric V1 Architecture

- Elements of the Architecture
  - Chaincode: System and regular ones. Deploy and Invoke latter.
  - State: Versioned KV model - stored in Level DB or CouchDB
  - Ledger data: Blockchain has full state, including history
  - Transactions
- Kafka for Ordering:
  - No Byzantine fault handling
  - Done to improve performance
  - Pluggable consensus permits other methods

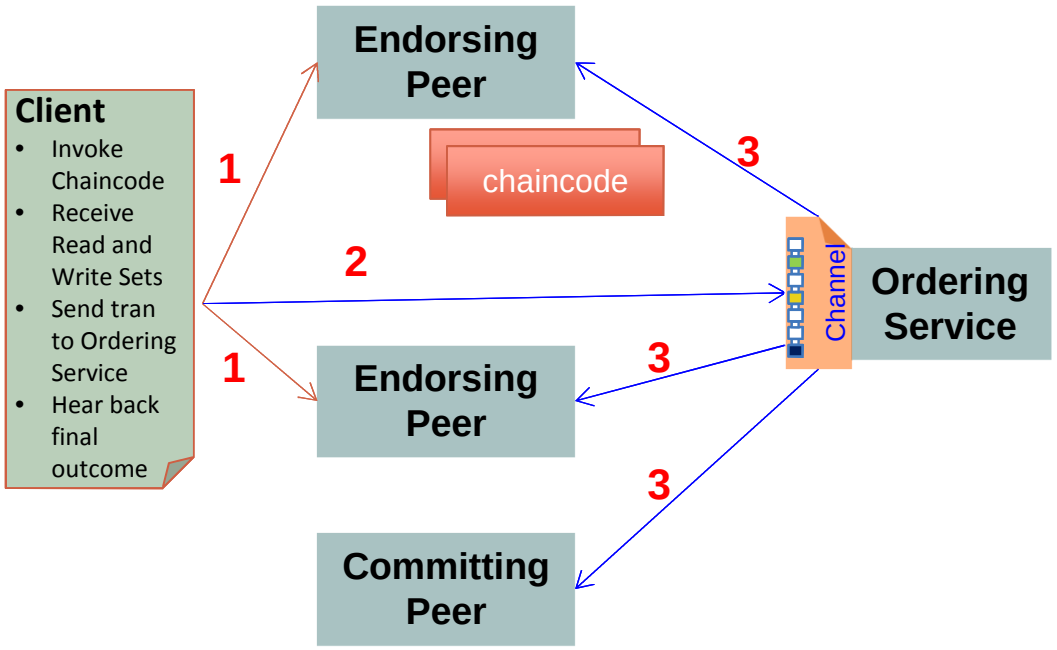
Fabric V1 Ledger



CouchDB (external option) supports keyed queries, composite key queries, key range queries, plus full data rich queries (beta in 1.0)

Transaction Execution Overview Fabric V1

3 Stage Execution: Endorsement, Ordering, Validation/Commit

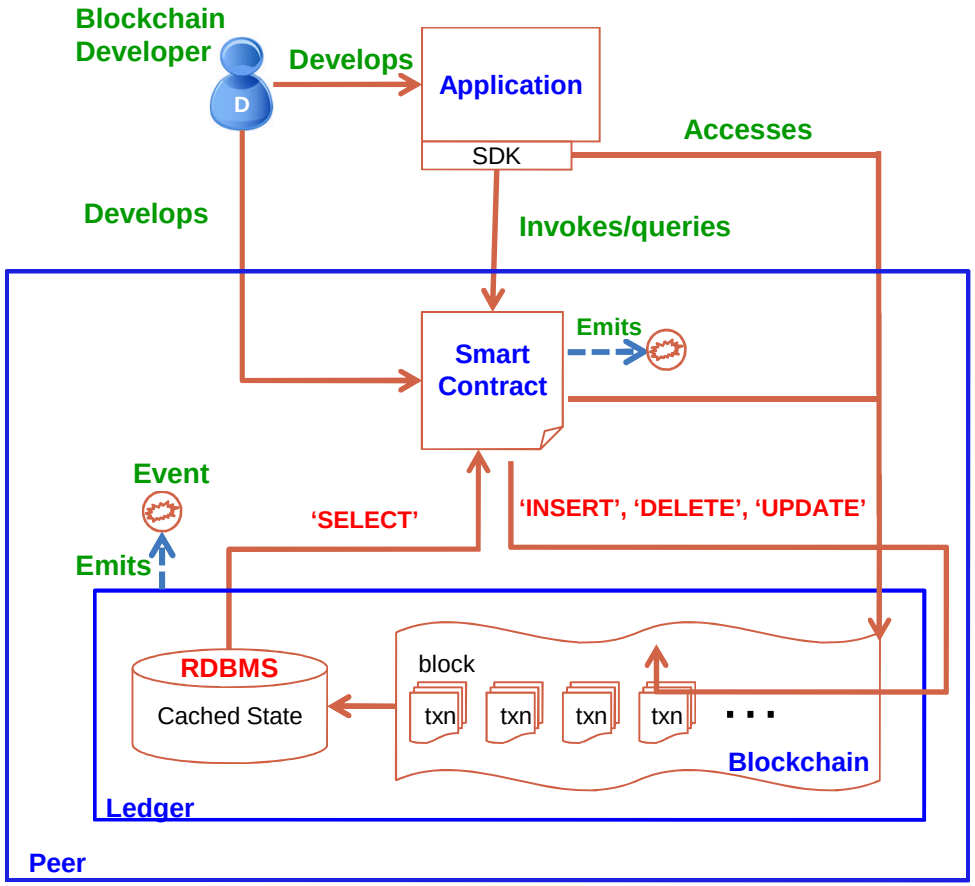


- Transaction is sent to the counter-parties represented by Endorsing Peers on their Channel
- Each Peer simulates transaction execution by calling specified Chaincode function(s) and signs result (Read-Write Sets)
- Each Peer may participate in multiple channels allowing concurrent execution
- Ordering Service accepts endorsed transactions and orders them according to the plug-in consensus algorithm then delivers them on the channel
- All (Committing) peers on channel receive transactions: on successful validation, commit to ledger. No chaincode execution.

DBMS Implications

- Simulation concept requires layer between chaincode and State DB having to take on analysis of DBMS calls
  - Update statements split into two: read part and write part
  - Read alone sent to DBMS with modifications to retrieve version #s for items read
  - Writes not sent to DBMS but processed and cached locally – doesn't allow for read your own write by chaincode transaction
- During Commit phase, read sets validated by retrieving each item's version # individually and then, if validation succeeds, writes also done one at a time
- Dealing with phantoms requires reexecution of query during commit phase to be sure simulation read set same as read set at Commit time
- Chaincode portability across different State DBMSs hard to do
- Lots of open questions and research issues in this area

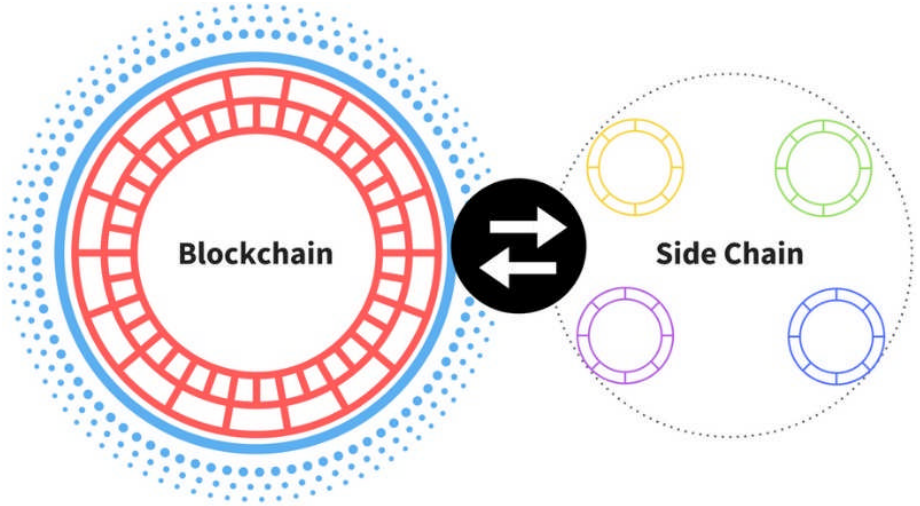
Application Flow with RDBMS (In Progress)



- Developers create application and smart contracts (chaincodes)
  - Chaincodes are deployed on the network and control the state of the ledger
  - Application handles user interface and submits transactions to the network which call chaincodes
- Network emits events on block of transactions allowing applications to integrate with other systems

Futuristic Topics

- Chaincode portability and power of data APIs
- DBMS enhancements to add BC features
- Standards across BC systems
- Cross channel transactions
- Non-deterministic actions
- Analytics on chaincode data
- Many app design issues
- Design tools for endorsement decisions



Numerous research possibilities for database and distributed systems people in this new era of distributed computing!

More Information

Links to Videos, Slides, Bibliography, Twitter Handles

http://bit.ly/CMbcDB

Follow me on

Twitter, WeChat: @seemohan

Facebook: http://www.facebook.com/cmohan

LinkedIn: http://www.linkedin.com/in/seemohan/

Talk at ACM Bay Area Chapter: 18 October 2017 (Wed)