Blockchain, Artificial Intelligence & Machine Learning Lecture Series

Feb 12 | Rise New York

Organized By

Blockchain NYC
http://blockchainNYC.io

Chainhaus
http://chainhaus.com

Blockchain, Artificial Intelligence & Data

Meetup Page: htttp://blockchainNYC.io

Upcoming Events

Feb 18 - Building a Crypto Price Prediction Web App with Phyton

Feb 25 - The Blockchain Masterclass

Feb 26 - Global Blockchain Healthcare

Mar 22 - Introduction to Python Coding

Mar 25 - The Python Masterclass

More Details: http://bit.ly/BlockchainNYCEvents

9:30 AM	Opening Remarks	Jamiel Sheikh
10:00 AM	Democratizing Data Science in Healthcare	Dr. Joel Park
11:00 AM	Adopting AI in healthcare - NLP and ML	Niteen Kumar
1:00 PM	Productionizing Tensorflow with Amazon Sagemaker	Harry Moreno
2:00 PM	Spark for Beginners: A deep dive and tutorial	Waseem Hussein
3:30 PM	Distributed word embedding with Spark	Naiem Yeganeh
5:00 PM	Building a large-scale AI platform	Waleed Nasir
6:00 PM	What is Data (Really) And What Can You Do (And Not to Do) with it	Walter Perry
6:45 PM	Data Science in Finance	Kayva Krishna
7:00 PM	Blockchain-based distributed shared-computing	Chong Li
8:00 PM	Data Analysis of Ethereum Chain	Jamiel Sheikh

Data Analysis of Ethereum Chain

Professor Jamiel Sheikh
jamiel@chainhaus.com

Bio

Jamiel is CEO of Chainhaus, an **advisory, software development, application studio and education** company focused on blockchain, artificial intelligence and machine learning. Jamiel has over 15 years of experience in technology, capital markets, real estate and management and is an **adjunct professor at Columbia Business School, NYU and CUNY teaching graduate-level blockchain, AI and data science** subjects.

He is currently **authoring a book on Corda with O'Reilly** and runs one of the largest blockchain, AI and data science Meetups in NYC. Jamiel is a licensed real estate agent in New York and New Jersey with Douglas Elliman.

Jamiel holds an MBA from Columbia University and BBA from Baruch College and is completing his second Masters in Artificial Intelligence from Georgia Institute of Technology.

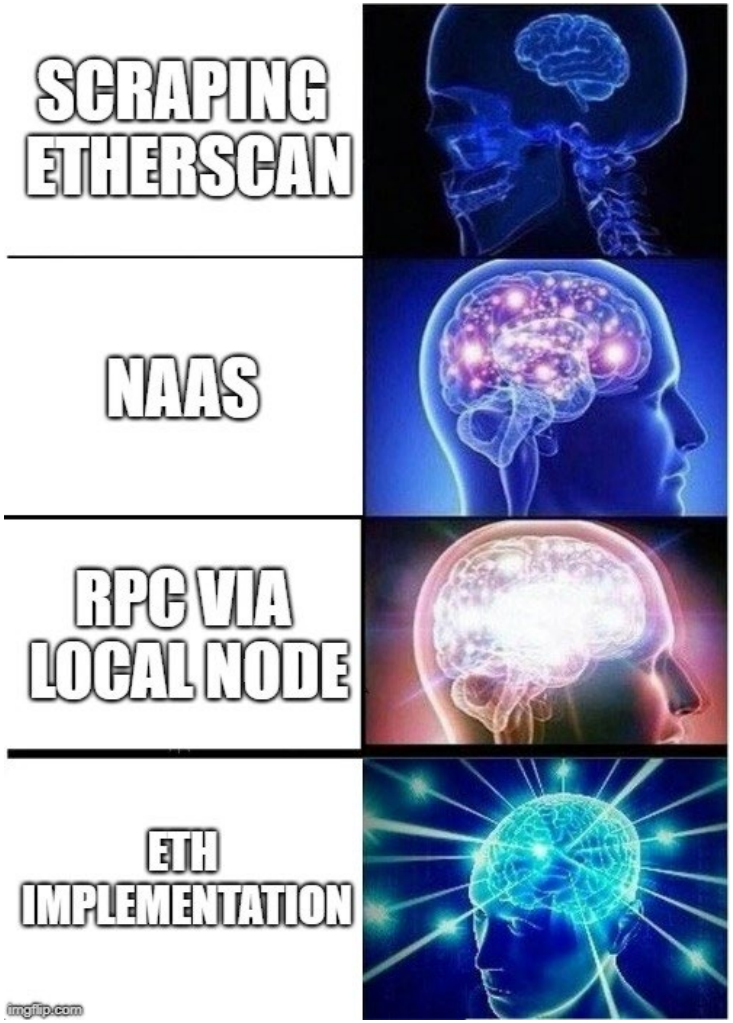
Jamiel enjoys coding in over 8 languages, travel and the elegant nuances of MMA.

Code

https://github.com/jamiels/pyethdata

Gimme Eth Data

- Web scrape - Scrape Etherscan.io
- Node as a Service - Infura.io / Quicknode
- Run your own Node - Node RPC
  - REST APIs
- LevelDB
- Roll your own Eth implementation


SCRAPING ETHERSCAN
NAAS
RPC VIA LOCAL NODE
ETH IMPLEMENTATION
imgflip.com

Web Scraping

Etherscan rip

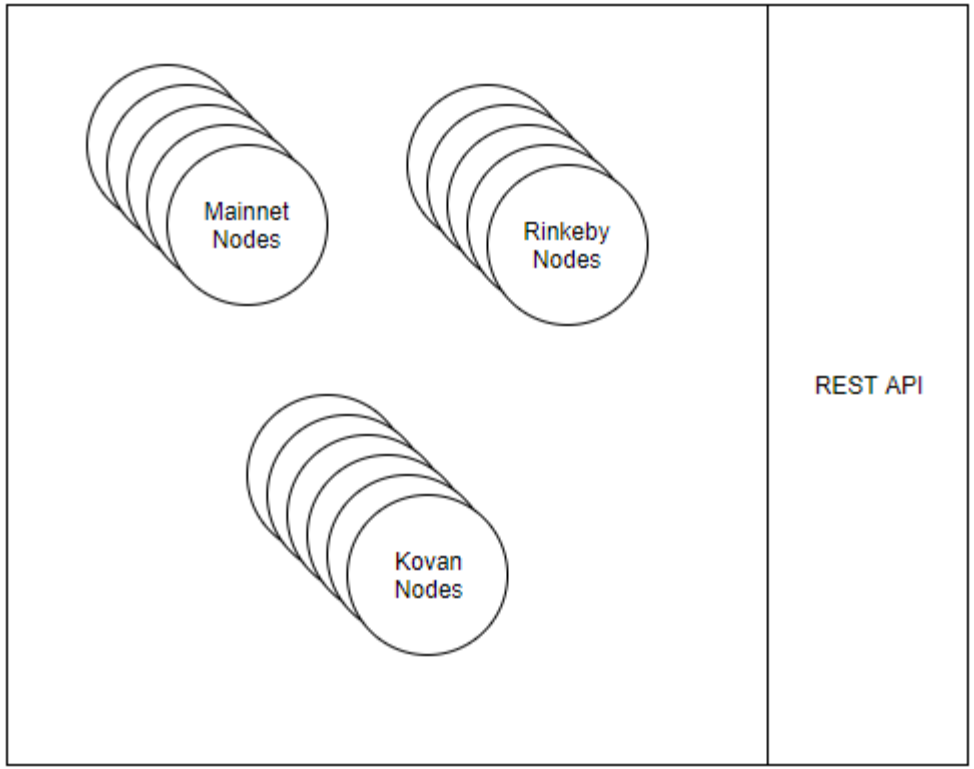
- Pros
  - Cloud-ish
- Cons
  - Brittle
  - Availability
  - Currency

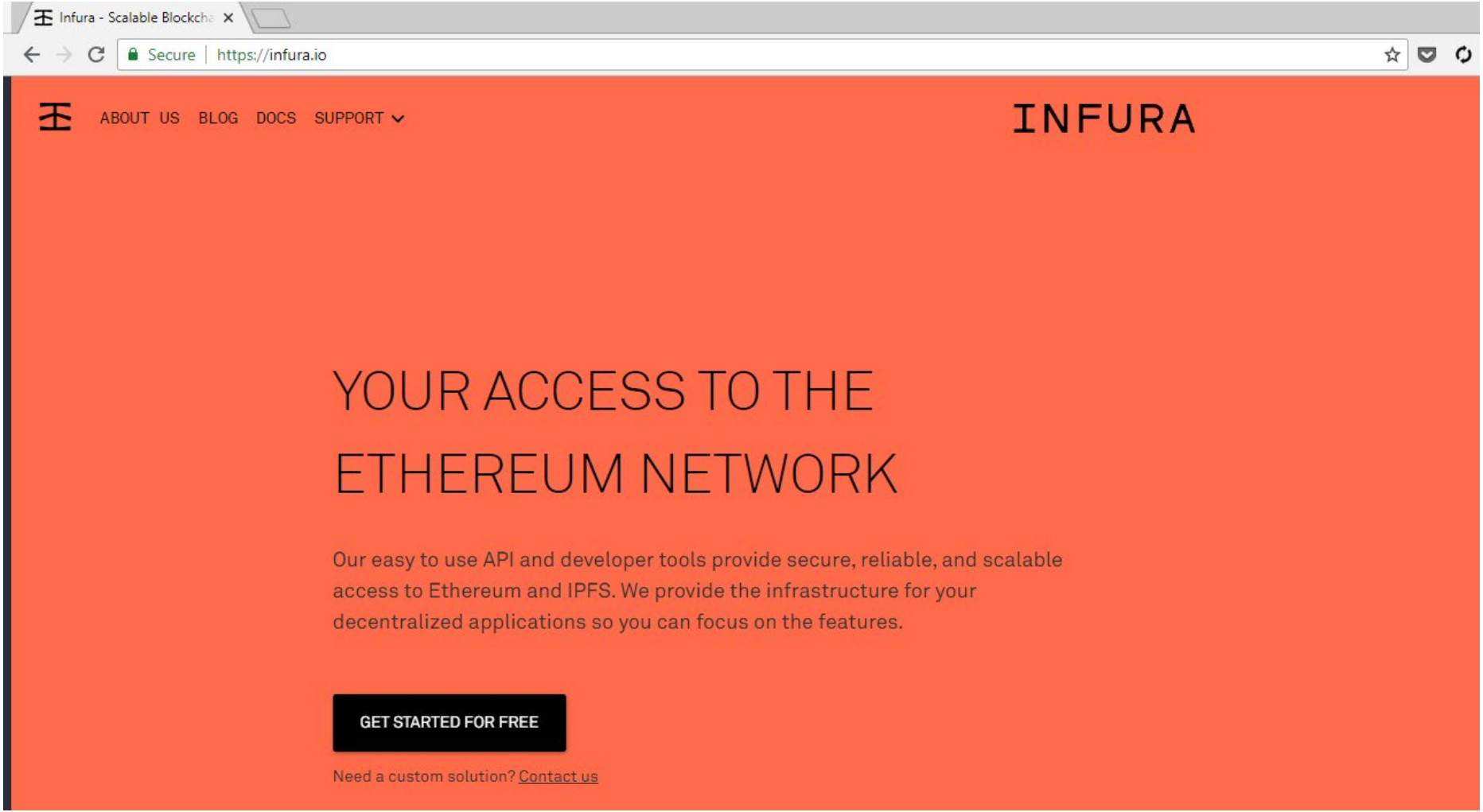
Demo

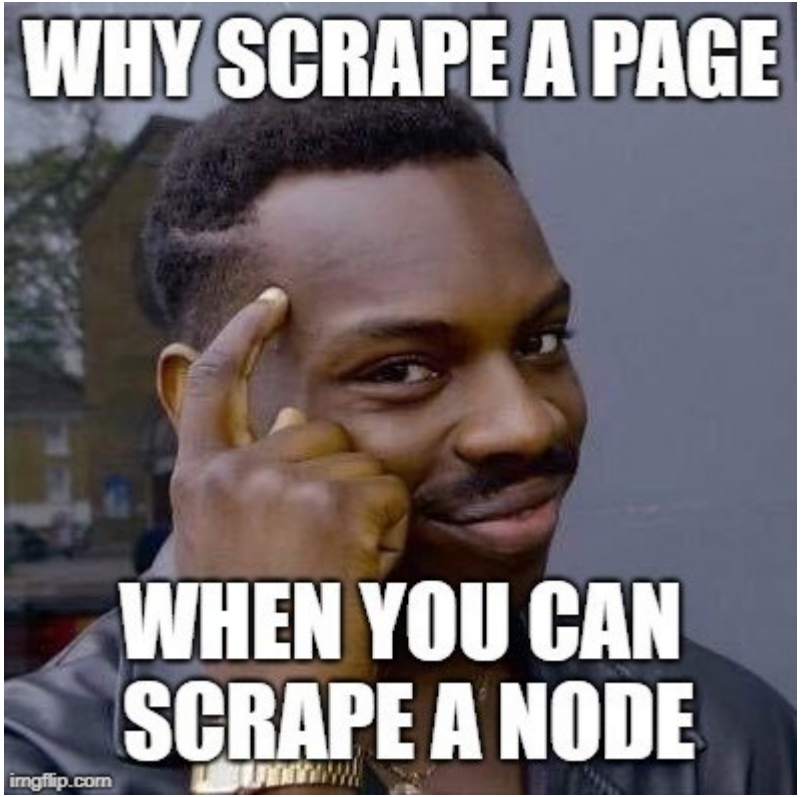
- `pip install bs4`

Node as a Service

Infura Infrastructure




Infura - Scalable Blockcha
Secure
https://infura.io
ABOUT US
BLOG
DOCS
SUPPORT
INFURA
YOUR ACCESS TO THE ETHEREUM NETWORK
Our easy to use API and developer tools provide secure, reliable, and scalable access to Ethereum and IPFS. We provide the infrastructure for your decentralized applications so you can focus on the features.
GET STARTED FOR FREE
Need a custom solution? Contact us

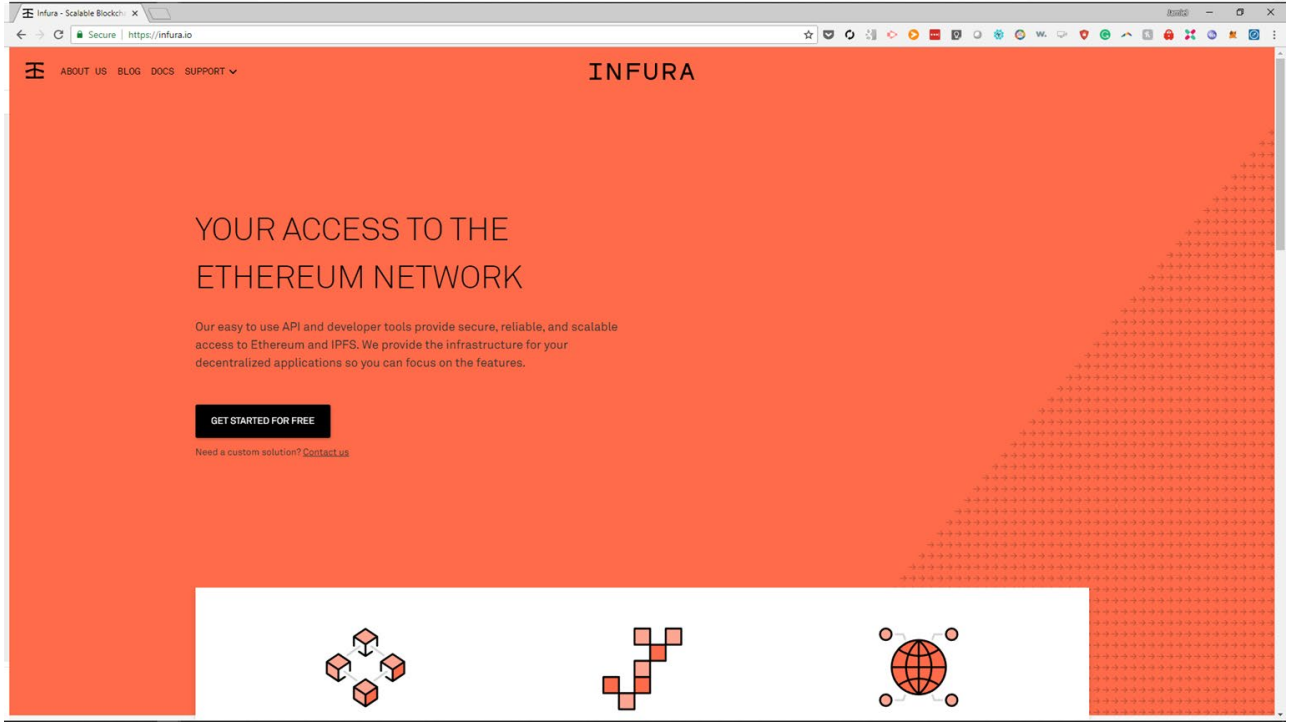

WHY SCRAPE A PAGE
WHEN YOU CAN
SCRAPE A NODE
imgflip.com

Infura

- Hosted Nodes
- API access to public Ethereum networks

Infura Signup

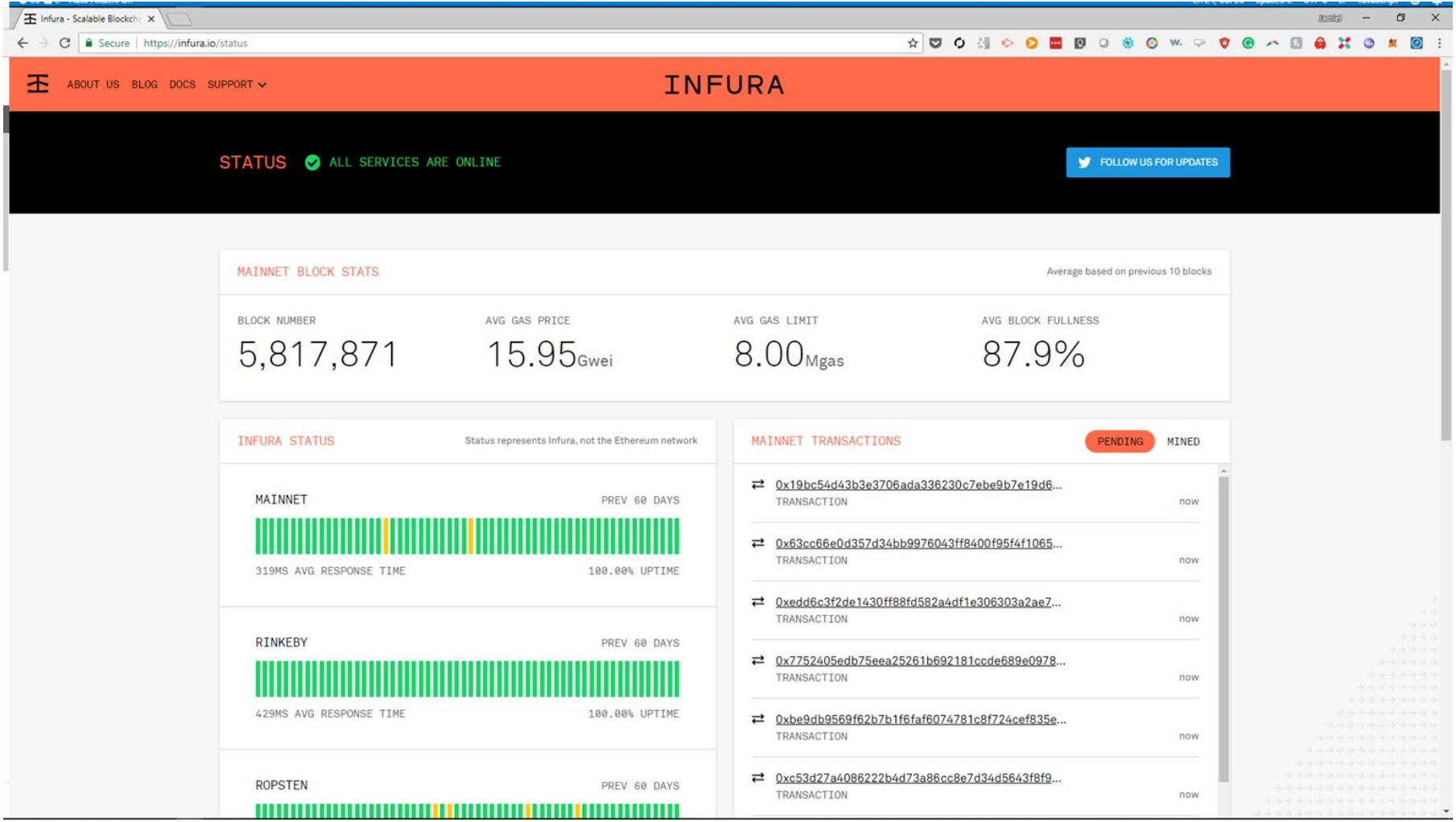
- Infura.io
- Keep your API key safe



Copy and save your API key for use within your app. We've also emailed it to you.

COPY

NETWORK	DESCRIPTION	URL
Mainnet	production network	https://mainnet.infura.io/iC
Ropsten	test network	https://ropsten.infura.io/iCx
INFURAnet	test network	https://infuranet.infura.io/iC
Kovan	test network	https://kovan.infura.io/iCxlS
Rinkeby	test network	https://rinkeby.infura.io/iCxl
IPFS	gateway	https://ipfs.infura.io


Infura - Scalable Blockch
Secure | https://infura.io/status
ABOUT US BLOG DOCS SUPPORT
INFURA
STATUS ALL SERVICES ARE ONLINE
FOLLOW US FOR UPDATES
MAINNET BLOCK STATS
Average based on previous 10 blocks
BLOCK NUMBER
5,817,871
AVG GAS PRICE
15.95Gwei
AVG GAS LIMIT
8.00Mgas
AVG BLOCK FULLNESS
87.9%
INFURA STATUS
Status represents Infura, not the Ethereum network
MAINNET
PREV 60 DAYS
319MS AVG RESPONSE TIME
100.00% UPTIME
RINKEBY
PREV 60 DAYS
429MS AVG RESPONSE TIME
100.00% UPTIME
ROPSTEN
PREV 60 DAYS
MAINNET TRANSACTIONS
PENDING
MINED
0x19bc54d43b3e3706ada336230c7ebe9b7e19d6...
TRANSACTION
now
0x63cc66e0d357d34bb9976043ff8400f95f4f1065...
TRANSACTION
now
0xedd6c3f2de1430ff88fd582a4df1e306303a2ae7...
TRANSACTION
now
0x7752405edb75eea25261b692181ccde689e0978...
TRANSACTION
now
0xbe9db9569f62b7b1f6faf6074781c8f724cef835e...
TRANSACTION
now
0xc53d27a4086222b4d73a86cc8e7d34d5643f8f9...
TRANSACTION
now

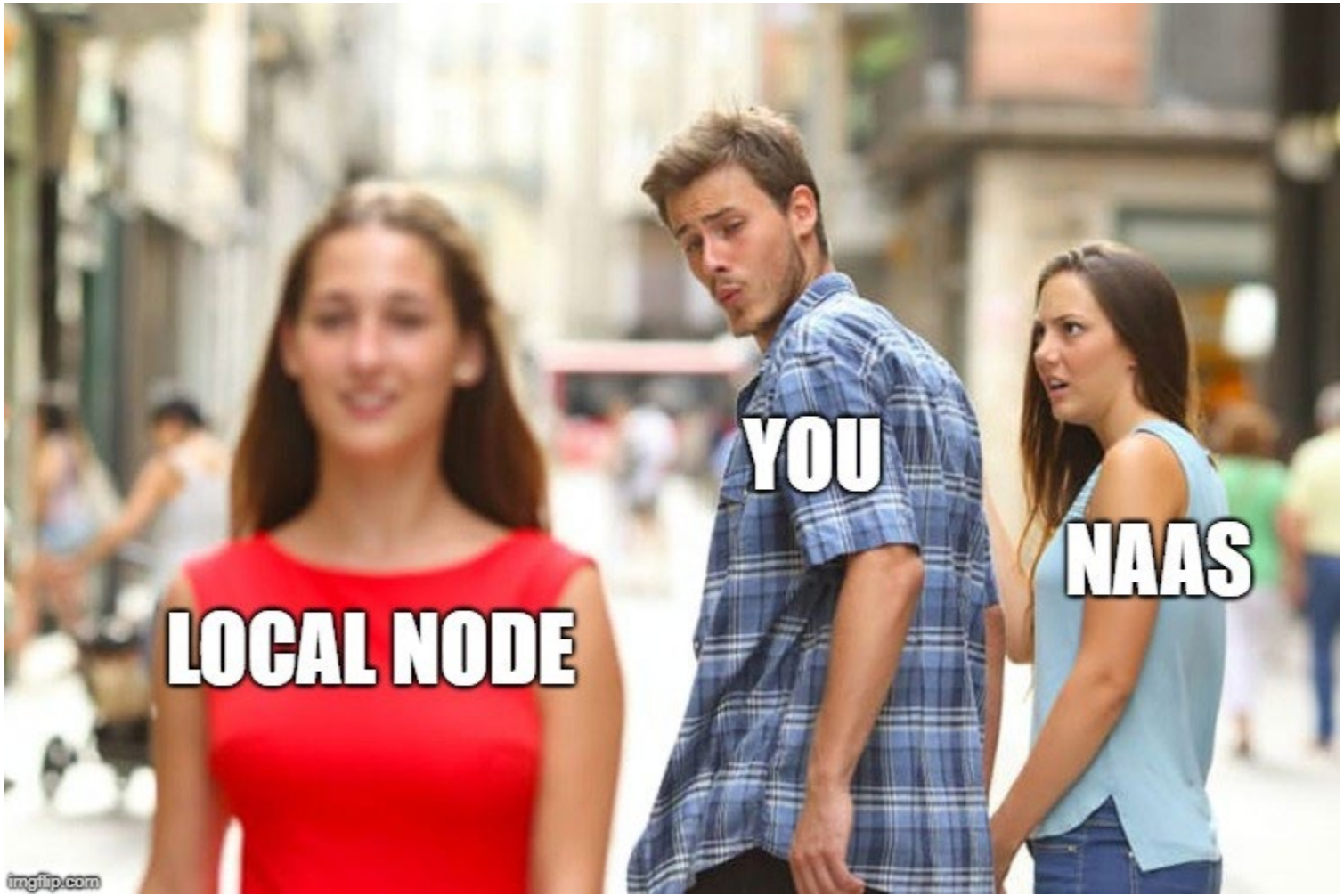
Cost/Benefit

- Pros
  - Free (for now)
  - Free (ops)
  - Free (scale)
- Cons
  - May not be free

Infura Demo

- `pip install infura`

Local Node


YOU
LOCAL NODE
NAAS
imgflip.com


local
eth node
root canal
JAKE-CLARK.TUMBLR
imgflip.com

Local node - Geth

- Install with development tools
- Parity
- Ethereumj
- Python implementation

Specific Versions

If you're looking for a specific release, operating system or architecture, below you will find:

- All stable and develop builds of Geth and tools
- Archives for non-primary processor architectures
- Android library archives and iOS XCode frameworks

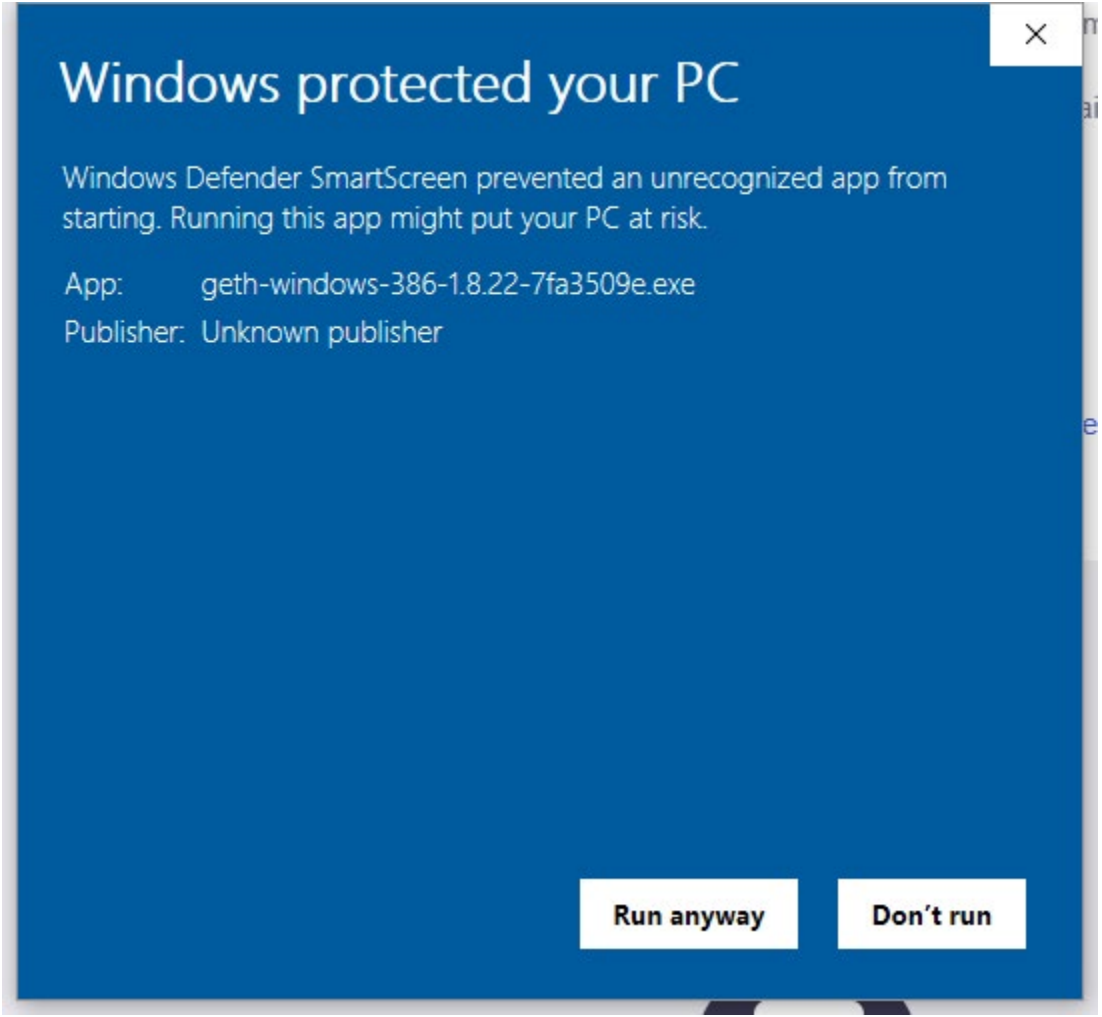
Please select your desired platform from the lists below and download your bundle of choice. Please be aware that the `MD5` checksums are provided by our binary hosting platform (Azure Blobstore) to help check for download errors. **For security guarantees please verify any downloads via the attached PGP signature files** (see OpenPGP Signatures for details).

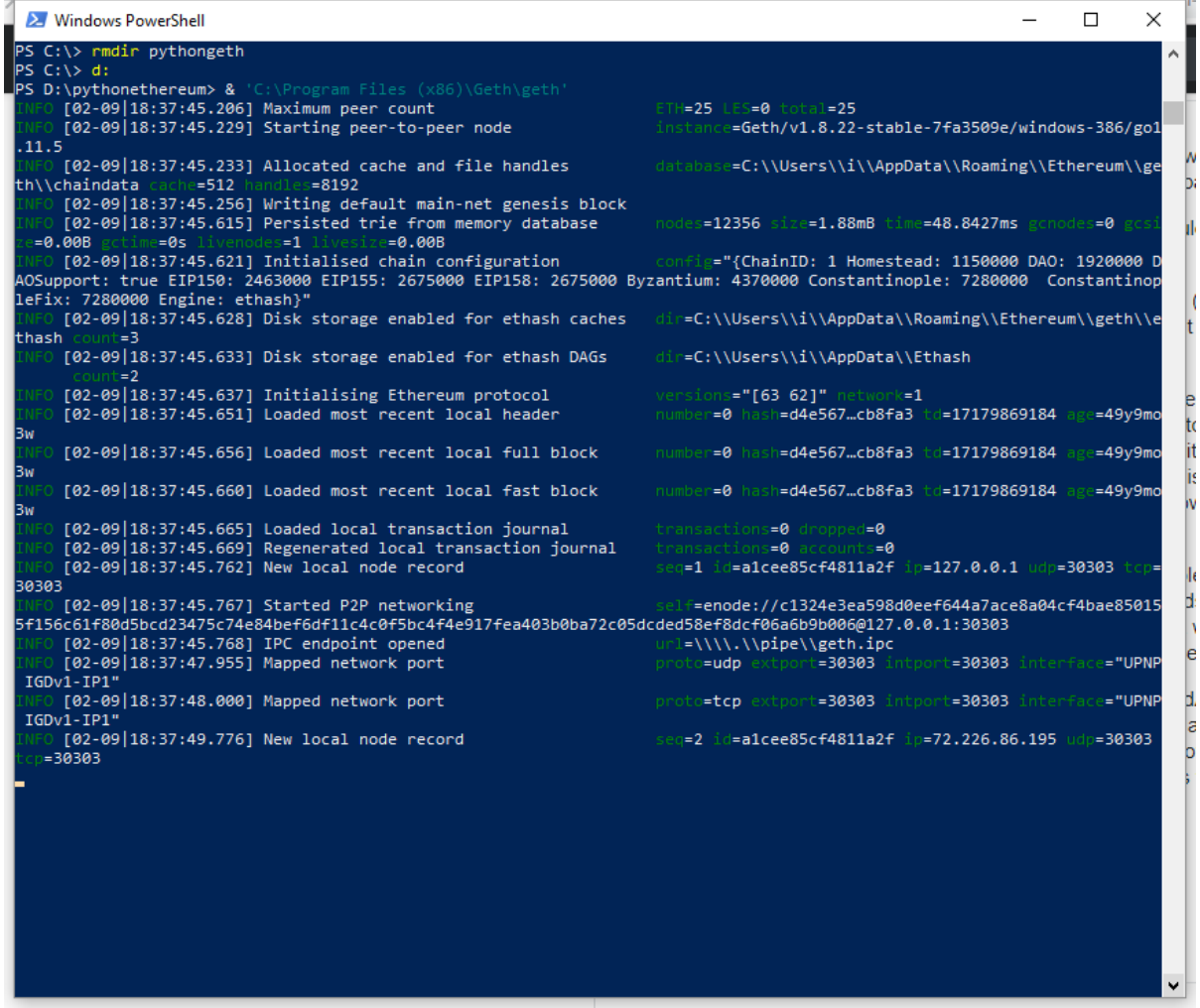
Stable releases

These are the current and previous stable releases of go-ethereum, updated automatically when a new version is tagged in our GitHub repository.

Android iOS Linux macOS Windows

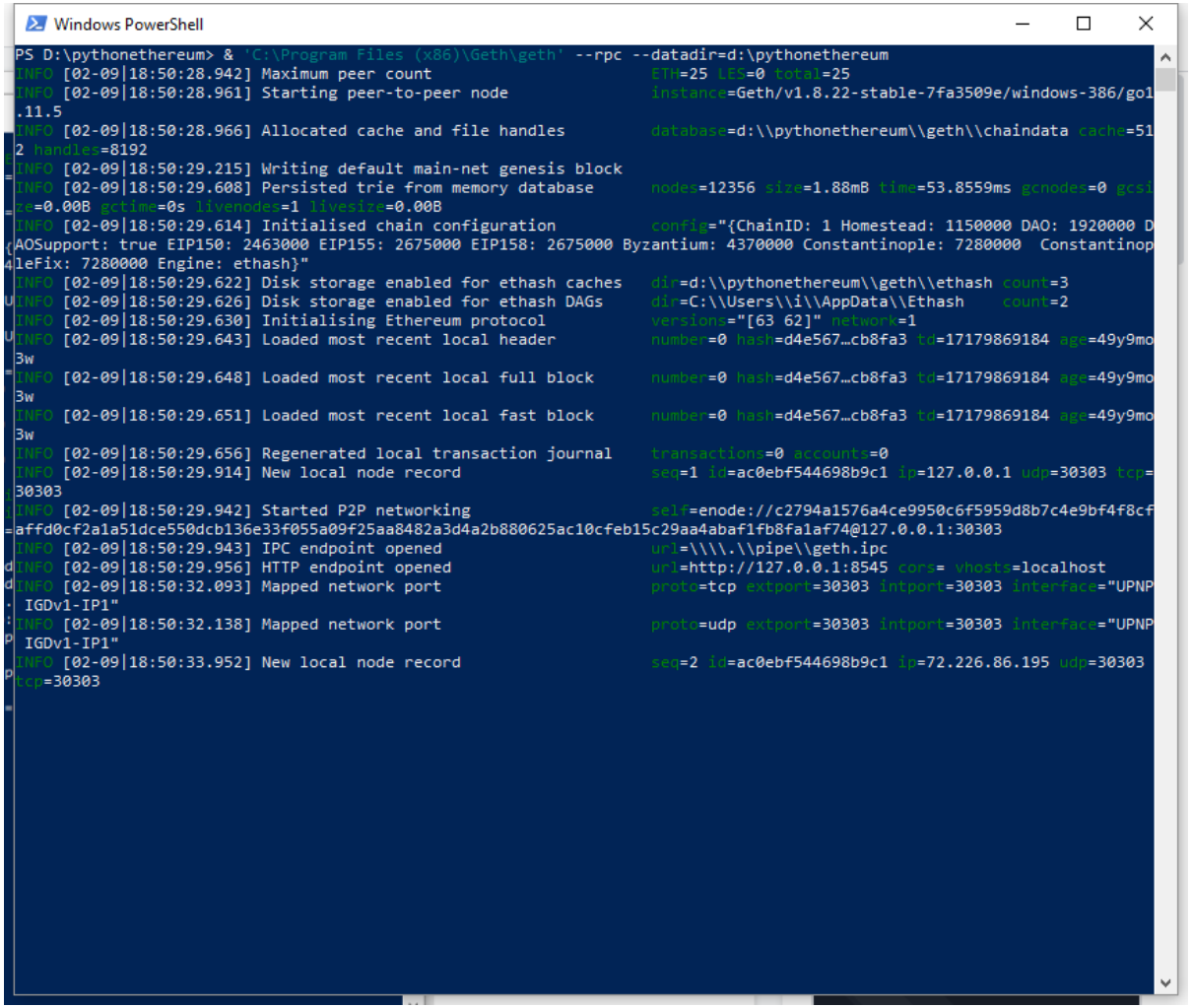
Release	Commit	Kind	Arch	Size	Published	Signature	Checksum (MD5)
Geth 1.8.22	7fa3509e...	Installer	32-bit	41.39 MB	01/31/2019	Signature	6b75335864f290fb94599d2611dd8151
Geth 1.8.22	7fa3509e...	Archive	32-bit	14.1 MB	01/31/2019	Signature	597137e95f3ca97579457043115dfc15
Geth & Tools 1.8.22	7fa3509e...	Archive	32-bit	55.54 MB	01/31/2019	Signature	45527dbf123b841ae57f1fa9c752824d
Geth & Tools 1.8.22	7fa3509e...	Archive	64-bit	58.04 MB	01/31/2019	Signature	dd6ef1ebc3c47ebb7ae59a5d944f6902
Geth 1.8.21	9dc5d1a9...	Installer	32-bit	41.39 MB	01/15/2019	Signature	3a4fe9fd9bccdfee19e4cddbf15c8185
Geth 1.8.21	9dc5d1a9...	Archive	32-bit	14.1 MB	01/15/2019	Signature	0c834c4b874854425ddc7dd157f327ce
Geth & Tools 1.8.21	9dc5d1a9...	Archive	32-bit	55.52 MB	01/15/2019	Signature	c2ada7c395e8552c654ea89dfaa20def

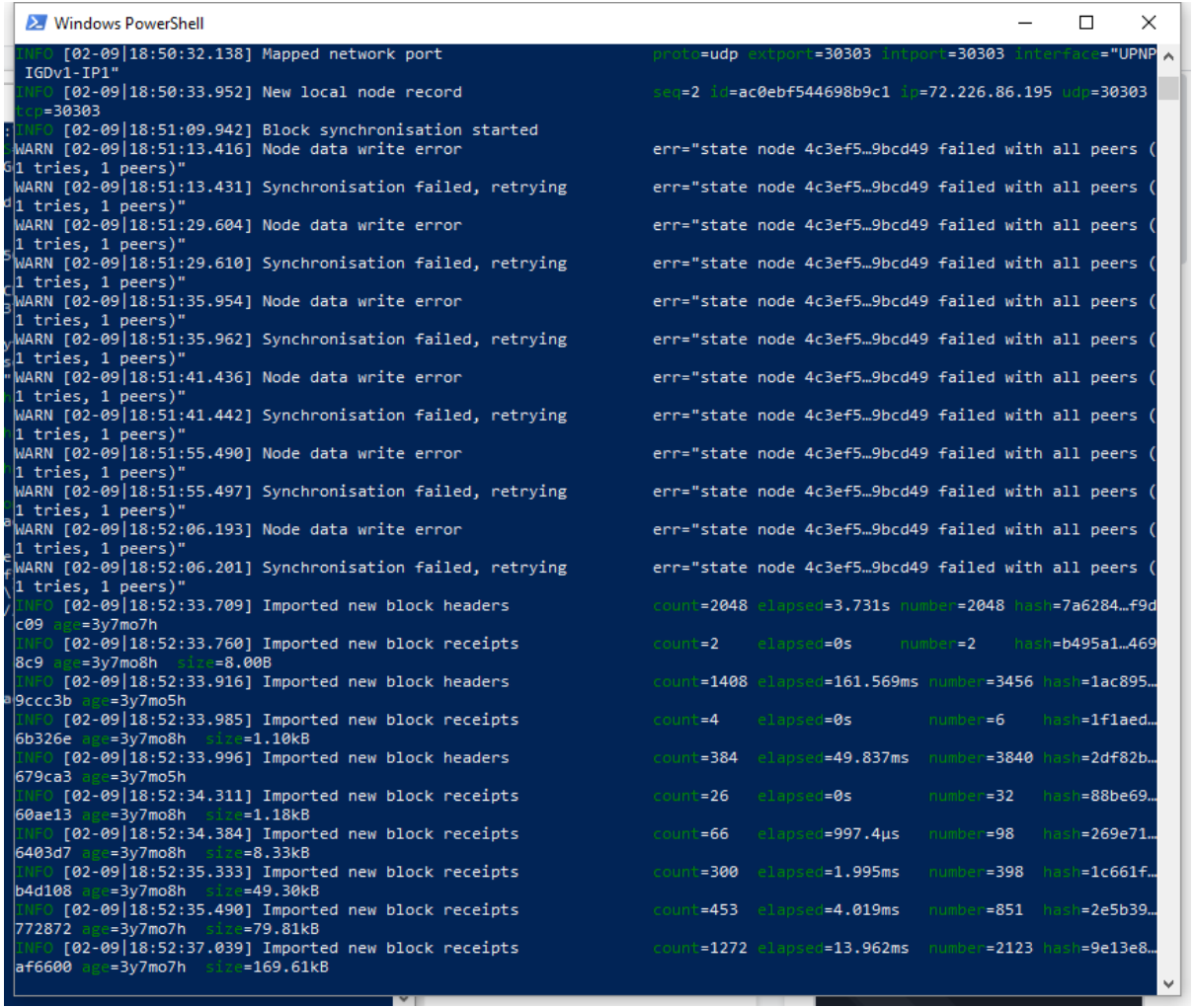

Windows protected your PC
Windows Defender SmartScreen prevented an unrecognized app from starting. Running this app might put your PC at risk.
App: geth-windows-386-1.8.22-7fa3509e.exe
Publisher: Unknown publisher
Run anyway
Don't run


Windows PowerShell
PS C:\> rmdir pythongeth
PS C:\> d:
PS D:\pythonethereum> & 'C:\Program Files (x86)\Geth\geth'
INFO [02-09|18:37:45.206] Maximum peer count ETH=25 LES=0 total=25
INFO [02-09|18:37:45.229] Starting peer-to-peer node instance=Geth/v1.8.22-stable-7fa3509e/windows-386/go1
.11.5
INFO [02-09|18:37:45.233] Allocated cache and file handles database=C:\\Users\\i\\AppData\\Roaming\\Ethereum\\ge
th\\chaindata cache=512 handles=8192
INFO [02-09|18:37:45.256] Writing default main-net genesis block
INFO [02-09|18:37:45.615] Persisted trie from memory database nodes=12356 size=1.88mB time=48.8427ms gcnodes=0 gcsi
ze=0.00B gctime=0s livenodes=1 livesize=0.00B
INFO [02-09|18:37:45.621] Initialised chain configuration config="{ChainID: 1 Homestead: 1150000 DAO: 1920000 D
AOSupport: true EIP150: 2463000 EIP155: 2675000 EIP158: 2675000 Byzantium: 4370000 Constantinople: 7280000 Constantinop
leFix: 7280000 Engine: ethash}"
INFO [02-09|18:37:45.628] Disk storage enabled for ethash caches dir=C:\\Users\\i\\AppData\\Roaming\\Ethereum\\geth\\e
thash count=3
INFO [02-09|18:37:45.633] Disk storage enabled for ethash DAGs dir=C:\\Users\\i\\AppData\\Ethash
count=2
INFO [02-09|18:37:45.637] Initialising Ethereum protocol versions="[63 62]" network=1
INFO [02-09|18:37:45.651] Loaded most recent local header number=0 hash=d4e567…cb8fa3 td=17179869184 age=49y9mo
3w
INFO [02-09|18:37:45.656] Loaded most recent local full block number=0 hash=d4e567…cb8fa3 td=17179869184 age=49y9mo
3w
INFO [02-09|18:37:45.660] Loaded most recent local fast block number=0 hash=d4e567…cb8fa3 td=17179869184 age=49y9mo
3w
INFO [02-09|18:37:45.665] Loaded local transaction journal transactions=0 dropped=0
INFO [02-09|18:37:45.669] Regenerated local transaction journal transactions=0 accounts=0
INFO [02-09|18:37:45.762] New local node record seq=1 id=a1cee85cf4811a2f ip=127.0.0.1 udp=30303 tcp=
30303
INFO [02-09|18:37:45.767] Started P2P networking self=enode://c1324e3ea598d0eef644a7ace8a04cf4bae85015
5f156c61f80d5bcd23475c74e84bef6df11c4c0f5bc4f4e917fea403b0ba72c05dcded58ef8dcf06a6b9b006@127.0.0.1:30303
INFO [02-09|18:37:45.768] IPC endpoint opened url=\\\\.\\pipe\\geth.ipc
INFO [02-09|18:37:47.955] Mapped network port proto=udp extport=30303 intport=30303 interface="UPNP
IGDv1-IP1"
INFO [02-09|18:37:48.000] Mapped network port proto=tcp extport=30303 intport=30303 interface="UPNP
IGDv1-IP1"
INFO [02-09|18:37:49.776] New local node record seq=2 id=a1cee85cf4811a2f ip=72.226.86.195 udp=30303
tcp=30303

Launching geth

- `geth --rpc`


Windows PowerShell
PS D:\pythonethereum> & 'C:\Program Files (x86)\Geth\geth' --rpc --datadir=d:\pythonethereum
INFO [02-09|18:50:28.942] Maximum peer count ETH=25 LES=0 total=25
INFO [02-09|18:50:28.961] Starting peer-to-peer node instance=Geth/v1.8.22-stable-7fa3509e/windows-386/go1.11.5
INFO [02-09|18:50:28.966] Allocated cache and file handles database=d:\\pythonethereum\\geth\\chaindata cache=512 handles=8192
INFO [02-09|18:50:29.215] Writing default main-net genesis block
INFO [02-09|18:50:29.608] Persisted trie from memory database nodes=12356 size=1.88mB time=53.8559ms gcnodes=0 gcsize=0.00B gctime=0s livenodes=1 livesize=0.00B
INFO [02-09|18:50:29.614] Initialised chain configuration config="{ChainID: 1 Homestead: 1150000 DAO: 1920000 DAOSupport: true EIP150: 2463000 EIP155: 2675000 EIP158: 2675000 Byzantium: 4370000 Constantinople: 7280000 Constantinople: 7280000 Engine: ethash}"
INFO [02-09|18:50:29.622] Disk storage enabled for ethash caches dir=d:\\pythonethereum\\geth\\ethash count=3
INFO [02-09|18:50:29.626] Disk storage enabled for ethash DAGs dir=C:\\Users\\i\\AppData\\Ethash count=2
INFO [02-09|18:50:29.630] Initialising Ethereum protocol versions="[63 62]" network=1
INFO [02-09|18:50:29.643] Loaded most recent local header number=0 hash=d4e567…cb8fa3 td=17179869184 age=49y9mo3w
INFO [02-09|18:50:29.648] Loaded most recent local full block number=0 hash=d4e567…cb8fa3 td=17179869184 age=49y9mo3w
INFO [02-09|18:50:29.651] Loaded most recent local fast block number=0 hash=d4e567…cb8fa3 td=17179869184 age=49y9mo3w
INFO [02-09|18:50:29.656] Regenerated local transaction journal transactions=0 accounts=0
INFO [02-09|18:50:29.914] New local node record seq=1 id=ac0ebf544698b9c1 ip=127.0.0.1 udp=30303 tcp=30303
INFO [02-09|18:50:29.942] Started P2P networking self=enode://c2794a1576a4ce9950c6f5959d8b7c4e9bf4f8cfaffd0cf2a1a51dce550dcb136e33f055a09f25aa8482a3d4a2b880625ac10cfeb15c29aa4abaf1fb8fa1af74@127.0.0.1:30303
INFO [02-09|18:50:29.943] IPC endpoint opened url=\\\\.\\pipe\\geth.ipc
INFO [02-09|18:50:29.956] HTTP endpoint opened url=http://127.0.0.1:8545 cors= vhosts=localhost
INFO [02-09|18:50:32.093] Mapped network port proto=tcp extport=30303 intport=30303 interface="UPNP IGDv1-IP1"
INFO [02-09|18:50:32.138] Mapped network port proto=udp extport=30303 intport=30303 interface="UPNP IGDv1-IP1"
INFO [02-09|18:50:33.952] New local node record seq=2 id=ac0ebf544698b9c1 ip=72.226.86.195 udp=30303 tcp=30303

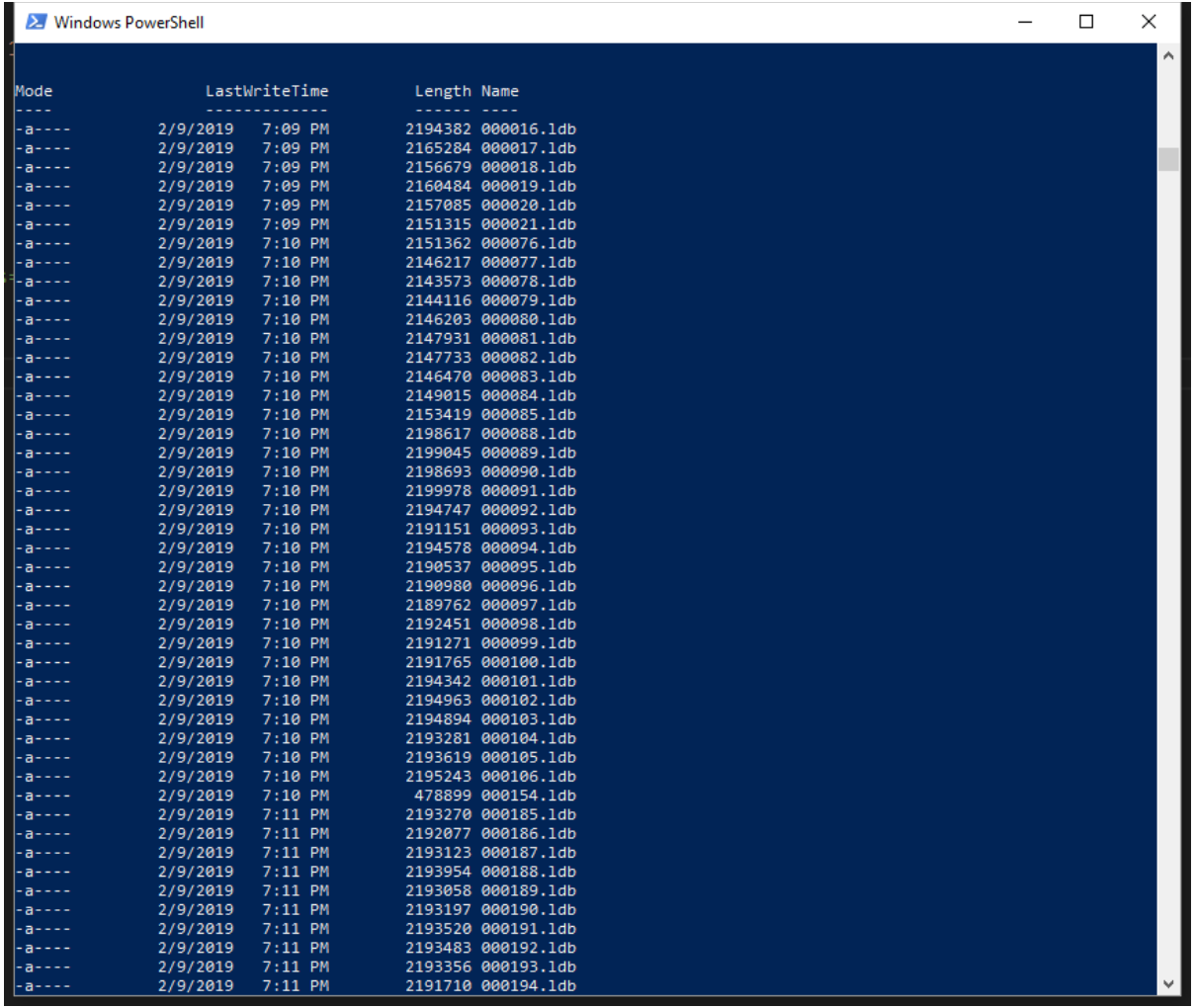

Windows PowerShell
INFO [02-09|18:50:32.138] Mapped network port proto=udp extport=30303 intport=30303 interface="UPNP IGDv1-IP1"
INFO [02-09|18:50:33.952] New local node record seq=2 id=ac0ebf544698b9c1 ip=72.226.86.195 udp=30303 tcp=30303
INFO [02-09|18:51:09.942] Block synchronisation started
WARN [02-09|18:51:13.416] Node data write error err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:13.431] Synchronisation failed, retrying err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:29.604] Node data write error err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:29.610] Synchronisation failed, retrying err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:35.954] Node data write error err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:35.962] Synchronisation failed, retrying err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:41.436] Node data write error err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:41.442] Synchronisation failed, retrying err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:55.490] Node data write error err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:51:55.497] Synchronisation failed, retrying err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:52:06.193] Node data write error err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
WARN [02-09|18:52:06.201] Synchronisation failed, retrying err="state node 4c3ef5…9bcd49 failed with all peers (1 tries, 1 peers)"
INFO [02-09|18:52:33.709] Imported new block headers count=2048 elapsed=3.731s number=2048 hash=7a6284…f9dc09 age=3y7mo7h
INFO [02-09|18:52:33.760] Imported new block receipts count=2 elapsed=0s number=2 hash=b495a1…4698c9 age=3y7mo8h size=8.00B
INFO [02-09|18:52:33.916] Imported new block headers count=1408 elapsed=161.569ms number=3456 hash=1ac895…9ccc3b age=3y7mo5h
INFO [02-09|18:52:33.985] Imported new block receipts count=4 elapsed=0s number=6 hash=1f1aed…6b326e age=3y7mo8h size=1.10kB
INFO [02-09|18:52:33.996] Imported new block headers count=384 elapsed=49.837ms number=3840 hash=2df82b…679ca3 age=3y7mo5h
INFO [02-09|18:52:34.311] Imported new block receipts count=26 elapsed=0s number=32 hash=88be69…60ae13 age=3y7mo8h size=1.18kB
INFO [02-09|18:52:34.384] Imported new block receipts count=66 elapsed=997.4µs number=98 hash=269e71…6403d7 age=3y7mo8h size=8.33kB
INFO [02-09|18:52:35.333] Imported new block receipts count=300 elapsed=1.995ms number=398 hash=1c661f…b4d108 age=3y7mo8h size=49.30kB
INFO [02-09|18:52:35.490] Imported new block receipts count=453 elapsed=4.019ms number=851 hash=2e5b39…772872 age=3y7mo7h size=79.81kB
INFO [02-09|18:52:37.039] Imported new block receipts count=1272 elapsed=13.962ms number=2123 hash=9e13e8…af6600 age=3y7mo7h size=169.61kB

Local node RPC Demo

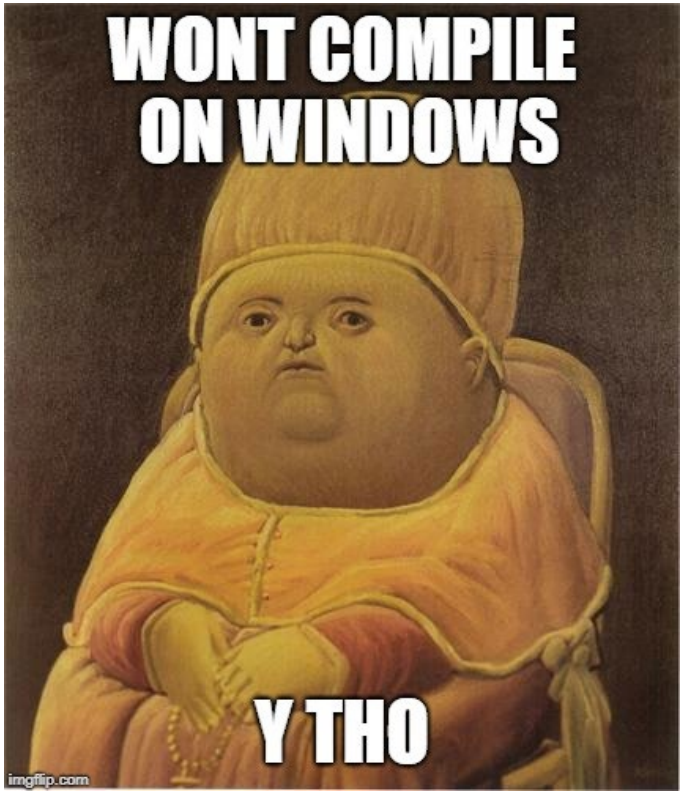
LevelDB

LevelDB

- Visual Studio 2015 Redistributable
- Pip install plyvel


Windows PowerShell
Mode LastWriteTime Length Name
---- ------------- ------ ----
-a---- 2/9/2019 7:09 PM 2194382 000016.ldb
-a---- 2/9/2019 7:09 PM 2165284 000017.ldb
-a---- 2/9/2019 7:09 PM 2156679 000018.ldb
-a---- 2/9/2019 7:09 PM 2160484 000019.ldb
-a---- 2/9/2019 7:09 PM 2157085 000020.ldb
-a---- 2/9/2019 7:09 PM 2151315 000021.ldb
-a---- 2/9/2019 7:10 PM 2151362 000076.ldb
-a---- 2/9/2019 7:10 PM 2146217 000077.ldb
-a---- 2/9/2019 7:10 PM 2143573 000078.ldb
-a---- 2/9/2019 7:10 PM 2144116 000079.ldb
-a---- 2/9/2019 7:10 PM 2146203 000080.ldb
-a---- 2/9/2019 7:10 PM 2147931 000081.ldb
-a---- 2/9/2019 7:10 PM 2147733 000082.ldb
-a---- 2/9/2019 7:10 PM 2146470 000083.ldb
-a---- 2/9/2019 7:10 PM 2149015 000084.ldb
-a---- 2/9/2019 7:10 PM 2153419 000085.ldb
-a---- 2/9/2019 7:10 PM 2198617 000088.ldb
-a---- 2/9/2019 7:10 PM 2199045 000089.ldb
-a---- 2/9/2019 7:10 PM 2198693 000090.ldb
-a---- 2/9/2019 7:10 PM 2199978 000091.ldb
-a---- 2/9/2019 7:10 PM 2194747 000092.ldb
-a---- 2/9/2019 7:10 PM 2191151 000093.ldb
-a---- 2/9/2019 7:10 PM 2194578 000094.ldb
-a---- 2/9/2019 7:10 PM 2190537 000095.ldb
-a---- 2/9/2019 7:10 PM 2190980 000096.ldb
-a---- 2/9/2019 7:10 PM 2189762 000097.ldb
-a---- 2/9/2019 7:10 PM 2192451 000098.ldb
-a---- 2/9/2019 7:10 PM 2191271 000099.ldb
-a---- 2/9/2019 7:10 PM 2191765 000100.ldb
-a---- 2/9/2019 7:10 PM 2194342 000101.ldb
-a---- 2/9/2019 7:10 PM 2194963 000102.ldb
-a---- 2/9/2019 7:10 PM 2194894 000103.ldb
-a---- 2/9/2019 7:10 PM 2193281 000104.ldb
-a---- 2/9/2019 7:10 PM 2193619 000105.ldb
-a---- 2/9/2019 7:10 PM 2195243 000106.ldb
-a---- 2/9/2019 7:10 PM 478899 000154.ldb
-a---- 2/9/2019 7:11 PM 2193270 000185.ldb
-a---- 2/9/2019 7:11 PM 2192077 000186.ldb
-a---- 2/9/2019 7:11 PM 2193123 000187.ldb
-a---- 2/9/2019 7:11 PM 2193954 000188.ldb
-a---- 2/9/2019 7:11 PM 2193058 000189.ldb
-a---- 2/9/2019 7:11 PM 2193197 000190.ldb
-a---- 2/9/2019 7:11 PM 2193520 000191.ldb
-a---- 2/9/2019 7:11 PM 2193483 000192.ldb
-a---- 2/9/2019 7:11 PM 2193356 000193.ldb
-a---- 2/9/2019 7:11 PM 2191710 000194.ldb

Windows?



LevelDB Demo

Roll your own

Eth

- Protocol
- Listeners

DEMOCRATIZING DATA SCIENCE IN HEALTHCARE

Joel Park, MD, FACEP

Clinician – Data Scientist

chainhaus

MIT Critical Data


POLIZIA
LOCALE


DOCTOR
TECH

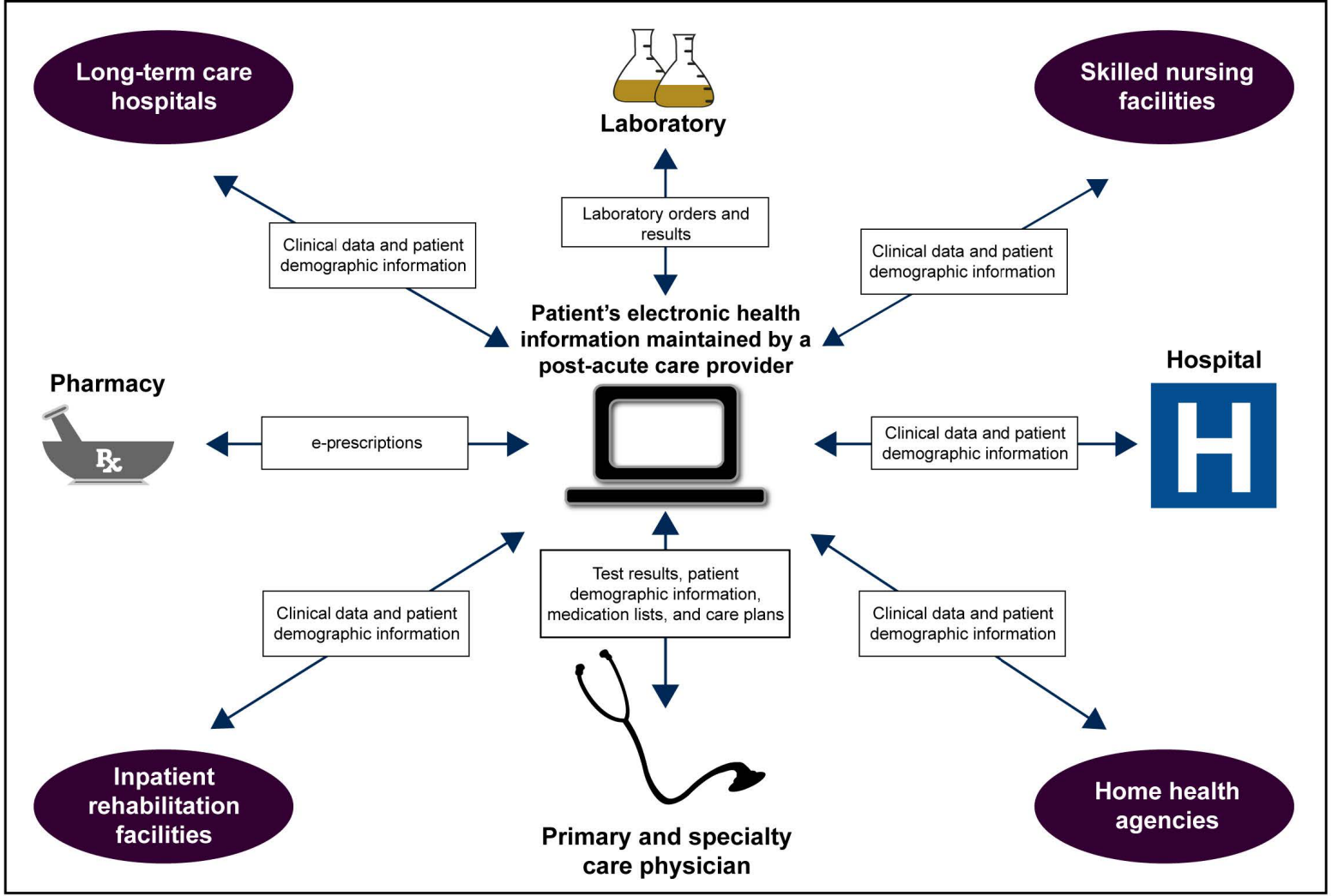

DAILY REPORT SCHEDULE
PATIENT NAME
TIME
BP
HR
PR
O2 SAT
TEMP
GLUCOSE
PAIN
PAIN IV MEDS CHECKS
MISC CHECKS
PATIENT MEDICATION SCHEDULE
PATIENT TIME
0700
0800
0900
1000
1200
1300
1400
MD
AM
08:15


Medical
Record
Name
Date

HIPAA

Medium.com

Names or part of names	Any other unique identifying characteristic
Geographical identifiers	Dates directly related to an individual
Phone numbers	Fax numbers
Email addresses	Social Security numbers
Medical record numbers	Health insurance beneficiary numbers
Account numbers	Certificate or license numbers
Vehicle license plate numbers	Device identifiers and serial numbers
Web URLs	IP addresses
Fingerprints, retinal and voice prints	Full face or any comparable photographic images



Sources: GAO analysis (data); Art Explosion (clip art). | GAO-17-184


BILLING
AND

KECM Keyword Code Helper: Connects term from indexes and KECM Common Word Dictionary to code

Results from KECM for fall (644)

ICD-10-CM

Exact Keyword Match

- D75 - Other and unspecified diseases of blood and blood-forming organs
- R29 - Other symptoms and signs involving the nervous and musculoskeletal systems
- R56 - Convulsions, not elsewhere classified
- V00 - Pedestrian conveyance accident
- V18 - Pedal cycle rider injured in noncollision transport accident
- V28 - Motorcycle rider injured in noncollision transport accident
- V38 - Occupant of three-wheeled motor vehicle injured in noncollision transport accident
- V80 - Animal-rider or occupant of animal-drawn vehicle injured in transport accident
- V81 - Occupant of railway train or railway vehicle injured in transport accident
- V82 - Occupant of powered streetcar injured in transport accident
- V91 - Other injury due to accident to watercraft
- V92 - Drowning and submersion due to accident on board watercraft, without accident to watercraft
- V93 - Other injury due to accident on board watercraft, without accident to watercraft
- V94 - Other and unspecified water transport accidents

INDEX

- CPT®
- HCPCS
- ICD-9-CM Diseases and Injuries
- ICD-9-CM External Causes of Injury
- ICD-10-CM Diseases and Injuries
- ICD-10-CM External Causes
- Index to ICD-10 PCS

TABLE

- HCPCS Drugs and Biologicals
- ICD-9-CM Drugs and Chemicals
- ICD-9-CM Hypertension
- ICD-9-CM Neoplasm
- ICD-10-CM Drug and Chemical
- ICD-10-CM Neoplasm

- W00 - Fall due to ice and snow
- W01 - Fall on same level from slipping, tripping and stumbling
- W03 - Other fall on same level due to collision with another person
- W04 - Fall while being carried or supported by other persons
  - W04.XXXA - Fall while being carried or supported by other persons, initial encounter
  - W04.XXXD - Fall while being carried or supported by other persons, subsequent encounter
  - W04.XXXS - Fall while being carried or supported by other persons, sequela
- W05 - Fall from non-moving wheelchair, nonmotorized scooter and motorized mobility scooter
  - W05.0 - Fall from non-moving wheelchair
  - W05.1 - Fall from non-moving nonmotorized scooter
  - W05.2 - Fall from non-moving motorized mobility scooter
- W06 - Fall from bed
  - W06.XXXA - Fall from bed, initial encounter
  - W06.XXXD - Fall from bed, subsequent encounter
  - W06.XXXS - Fall from bed, sequela
- W07 - Fall from chair
- W08 - Fall from other furniture

jet engine

Search

Narrow Your Search By

INDEX

- CPT® (0)
- HCPCS (0)
- ICD-9-CM Diseases and Injuries (0)
- ICD-9-CM External Causes(0)
- ICD-10-CM Diseases and Injuries (0)
- ICD-10-CM External Causes (0)
- ICD-10-CM PCS (0)

CODES

- All Results
- CPT® (0)
- HCPCS (0)
- ICD-9-CM (0)
- ICD-10-CM (3)
- Modifiers (0)
- PCS (0)

TABLE

CODES TOOLS PUBLICATIONS

All

No result found from indexes for jet engine

KECM Keyword Code Helper: Connects term from indexes and KECM Common Word Dictionary to code

Results from KECM for jet engine (3)

ICD-10-CM

Exact Keyword Match

V97 - Other specified air transport accidents

V97.3 - Person on ground injured in air transport accident

V97.33 - Sucked into jet engine

V97.33XA - Sucked into jet engine, initial encounter

V97.33XD - Sucked into jet engine, subsequent encounter

V97.33XS - Sucked into jet engine, sequela

No results found from tables for jet engine

CLINICAL IT

Survey: Physicians Cite EHRs as Biggest Contributor to Burnout

A recent survey on physician burnout and stress found that, perhaps unsurprisingly, physicians cited electronic health records (EHRs) as the top factor contributing to stress, followed by dealing with payers and pre-authorization and then regulatory compliance.

BY HEATHER LANDI — JULY 31, 2018

RELATED

Epic Earns Best in KLAS Recognition

Sepsis Core Measure Checklist

Date of Admission: _____ (Time Zero= Time at which infection is identified/documented + 2 SIRS present with 6 hours of one another)
ED Team_____ **ED Team**_____ **ED Team**_____

Inpt Team_____ **Inpt Team**_____ **Inpt Team**_____

☐ Infection identified/documented in ED with relevant Sepsis orders initiated.
☐ Lactate Result (not order) IF >2.0 mmol/l ☐ Documentation calling this Severe Sepsis ☐ Repeat Lactate result (order 2 hrs after prior draw time through "Infection" Order Set)
☐ Blood Cultures drawn (not ordered) prior to ATB ☐ Broad Spectrum (IV) ATB initiated (not ordered) within 3 hrs of Time Zero, *Selection from Empiric Broad Spectrum ATB List (on Green Sheet)*
☐ SIRS Template used in note: ☐ SIRS criteria indicated, ☐ Suspected Site(s) Indicated, ☐ In-hospital concurrent diagnosis indicated, ☐ Culture indicated, ☐ 30mL/kg Target documented, ☐ ATB/Medications indicated
☐ Assessment secondary to Organ Dysfunction indicating Severe Sepsis ***(****Lactate >2.0 mmol/l, INR >1.5, PTT > 60 sec, Platelet <100,000, Billirubin >2, Creatinine >2, Urine output < 0.5 mL/kg/hr for 2 hrs, SBP <90, MAP <65, SBP decrease by 40 from previous "normal"****)***-but not when chronic or due to medications
IF Severe Sepsis: ☐ Consider 30 mL/kg Crystalloid Fluid Bolus (0.9% NS or LR), ☐ Repeat Lactate result (order 2 hrs after prior draw time through "Infection" Order Set) which will order 2 additional Lactates.
IF Septic Shock: = Lactate **≥** 4.0 and/or Sepsis induced hypotension (SBP < 90 mmHg, MAP < 65 mmHg, or SBP decrease by 40 from previous "normal")-but not when chronic or due to medications
☐ Documentation calling this "Septic Shock with Severe Sepsis" ☐ 30 mL/kg Crystalloid Fluid Bolus (0.9% NS or LR) for hypotension or Lactate **≥** 4.0 > 125 mL hr, ☐ 30 mL/kg Target Achieved within 6 hrs of Time Zero of Lactate **≥** 4.0 and/or Sepsis induced hypotension
☐ **Vasopressors** (Norepinephrine 1st choice unless compelling reason for alternative) ☐ Within 6 hrs of Time Zero of Lactate **≥** 4.0 and/or Sepsis Induced hypotension ☐ Repeat Volume Status and Tissue Perfusion Assessment Note consisting of including Vital Signs, Cardiopulmonary, Capillary Refill, Pulse and Skin findings *(you may write the note after 6 hrs so long as you document the time you examined the patient which must be > 6 hrs)*
☐ Examination within 6 hrs of Time Zero of Lactate **≥** 4.0 and/or Sepsis Induced hypotension

Top Issues of Focus

☐ Broad Spectrum ATB AND Delivered within 3 hrs.	☐ ED Provider not thinking/documenting/acting upon Sepsis treatment plan.
☐ Infection/Sepsis Screen not suspected while in ED.	☐ 30 mL/kg ordered as one target volume based upon weight rather than small repeated boluses.
☐ Inpatient delay in timing of ATB administration from time ordered in Iatric.	☐ Communication from Inpatient provider to ED team on additional Sepsis orders on admission.
☐ Blood Cultures within 3 hrs.	☐ Lack of 6 hr Repeat Assessment note.

Reviewer Signature _____ **Date**_____**Time**_____

Reviewed With Signature_____ **Date**_____**Time**_____

INFECTION-SEPSIS SPECTRUM (ISS) CHECKLIST

AS DEFINED BY JOHNSON MEMORIAL HOSPITAL SEPSIS COMMITTEE:

Time Zero = *Time at which Infection is suspected/diagnosed + 2 or more SIRS present within 6 hours of one another*

SEPSIS = *Suspicion/diagnosis of infection + 2 or more SIRS (that cannot be excluded as due to the infection)*

SEVERE SEPSIS = *Suspicion/diagnosis of infection + 2 or more SIRS + organ dysfunction* ***(including Lactate >2.0)***

Date:_____TIME ZERO: _____

ALL of the following within (3) Hours of Time Zero				
☐ **Lactate result (not order)**	Draw Time:	Result Time:	Result:	Print Name
☐ **Blood Cultures drawn (prior to ATB) (not ordered)**		1st Set Time:	2nd Set Time:	Print Name
☐ **IV Antibiotic (ATB) initiated (not ordered)**		Time:		Print Name
AND within (3) Hours of Time Zero				
☐ **30 mL/kg Crystalloid Fluid Bolus (0.9% NS** *or* **LR) for Hypotension or Lactate ≥4 (consider for Severe Sepsis)** **Weight kg_____ X 30 = _____ mL predicted**		*Total volume given over 4-5 hours* **Target time to complete 30mL/kg:** **Amount infused in ED:**		Print Name
AND within (6) Hours of Time Zero				
☐ **Repeat Lactate result if initial is > 2.0 mmol/L (order 2hrs after prior draw time)**	Draw Time:	Result Time:	Result:	Print Name

SEVERE SEPSIS WITH SEPTIC SHOCK CHECKLIST

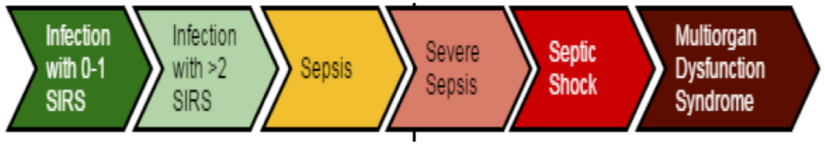
(all of the above measures plus the following)

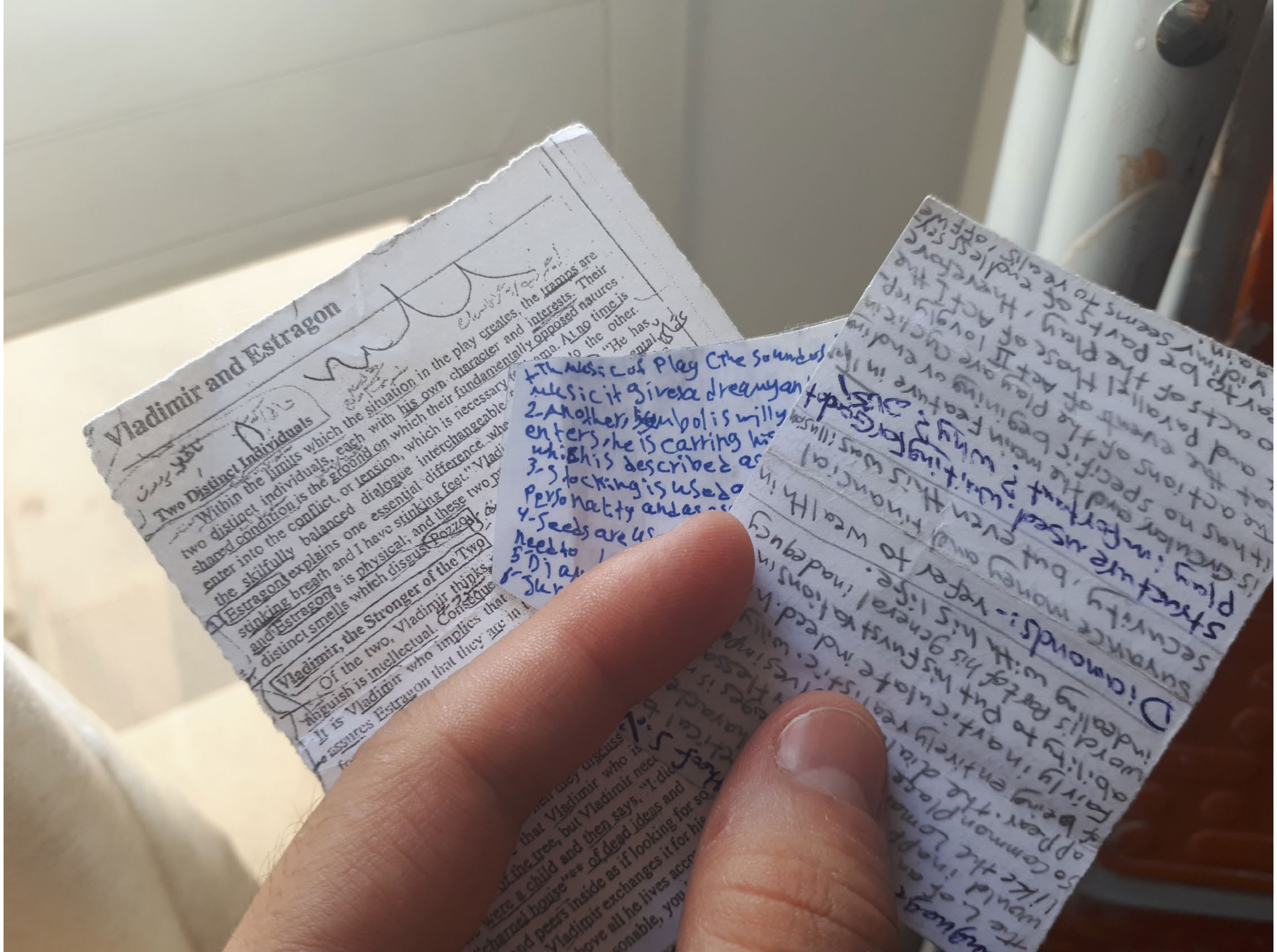
SEPTIC SHOCK = ***Lactate ≥ 4.0 and/or*** *Sepsis-induced* ***hypotension*** *(****SBP*** *less than* ***90*** *mmHg,* ***MAP*** *less than* ***65*** *mmHg, or* ***SBP*** *decrease greater than* ***40*** *mmHg from* ***baseline****) in the hour after fluid resuscitation (30mL/kg)* **for ≥ 2 consecutive BP readings**

Date:_____SEPTIC SHOCK CLOCK: _____

Within (6) Hours of Septic Shock Clock		
☐ **Vasopressors**	Time:	Print Name
Within (6) Hours of Septic Shock Clock		
☐ **Repeat Volume Status and Tissue Perfusion Assessment Note (written by NP/PA/MD/DO) consisting of including vital signs, cardiopulmonary, capillary refill, pulse, and skin findings**		

This form to remain in front of patient's chart until after six hour beyond time zero, and then forward it to Gina Croxford in the Quality Department. Not a part of the permanent medical record, DO NOT SCAN.




Vladimir and Estragon
Two Distinct Individuals
Vladimir, the Stronger of the Two


LAW
CASES
INDEX
LAW
CASES
LAW
CASES
Volume

If you use MIMIC data or code in your work, please cite the following publication:

MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman L, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. Scientific Data (2016). DOI: 10.1038/sdata.2016.35.
Available from: http://www.nature.com/articles/sdata201635

Requesting access

The latest version of MIMIC is MIMIC-III v1.4, which comprises over 58,000 hospital admissions for 38,645 adults and 7,875 neonates. The data spans June 2001 - October 2012. The database, although de-identified, still contains detailed information regarding the clinical care of patients, so must be treated with appropriate care and respect.

Researchers seeking to use the database must formally request access with the steps below.

Complete the required training course

Prior to requesting access to MIMIC, you will need to complete the CITI "Data or Specimens Only Research" course:

- First register on the CITI program website, selecting "Massachusetts Institute of Technology Affiliates" as your affiliation (***not*** *"independent learner"*): https://www.citiprogram.org/index.cfm?pageID=154&icat=0&ac=0
- Follow the links to add a Massachusetts Institute of

- CHARTEVENTS
- CPTEVENTS
- DATETIMEEVENTS
- D_CPT
- DIAGNOSES_ICD
- D_ICD_DIAGNOSES
- D_ICD_PROCEDURES
- D_ITEMS
- D_LABITEMS
- DRGCODES
- ICUSTAYS
- INPUTEVENTS_CV
- INPUTEVENTS_MV
- LABEVENTS
- MICROBIOLOGYEVENTS
- NOTEEVENTS
- OUTPUTEVENTS
- PATIENTS
- PRESCRIPTIONS

Query | NOTEEVENTS | OUTPUTEVENTS | PATIENTS

Description | Preview (100) rows

row_id	subject_id	gender	dob	dod	dod_hosp	dod_ssn	expire_flag
234	249	F	2075-03-13 00:00:00	None	None	None	0
235	250	F	2164-12-27 00:00:00	2188-11-22 00:00:00	2188-11-22 00:00:00	None	1
236	251	M	2090-03-15 00:00:00	None	None	None	0
237	252	M	2078-03-06 00:00:00	None	None	None	0
238	253	F	2089-11-26 00:00:00	None	None	None	0
239	255	M	2109-08-05 00:00:00	None	None	None	0
240	256	M	2086-07-31 00:00:00	None	None	None	0
241	257	F	2031-04-03 00:00:00	2121-07-08 00:00:00	2121-07-08 00:00:00	2121-07-08 00:00:00	1
242	258	F	2124-09-19 00:00:00	None	None	None	0
243	260	F	2105-03-23 00:00:00	None	None	None	0
244	261	M	2025-08-04 00:00:00	2102-06-29 00:00:00	2102-06-29 00:00:00	2102-06-29 00:00:00	1
245	262	M	2090-01-05 00:00:00	None	None	None	0
246	263	M	2104-06-18 00:00:00	2168-06-13 00:00:00	2168-06-13 00:00:00	None	1

MIMIC-III Query Builder
MIMIC-III

- ADMISSIONS
- CALLOUT
- CAREGIVERS
- CHARTEVENTS
- CPTEVENTS
- DATETIMEEVENTS
- D_CPT
- DIAGNOSES_ICD
- D_ICD_DIAGNOSES
- D_ICD_PROCEDURES
- D_ITEMS
- D_LABITEMS
- DRGCODES
- ICUSTAYS
- INPUTEVENTS_CV
- INPUTEVENTS_MV
- LABEVENTS
- MICROBIOLOGYEVENTS

Query

```
SELECT text FROM NOTEEVENTS
WHERE subject_id = 13702
```

Execute Query Showing only 187 results.

Export Results

text

[**2118-6-5**] 11:18 AM CHEST (PORTABLE AP) Clip # [**Clip Number (Radiology) 13147**] Reason: evaluate for PNA, consolidation, effusion. Admitting Diagnosis: CHRONIC OBSTRUCTIVE PULMONARY DISEASE _____ [**Hospital 2**] MEDICAL CONDITION: 81 year old woman with COPD flare not improved with BIPAP and steroid, has cough. REASON FOR THIS EXAMINATION: evaluate for PNA, consolidation, effusion. _____ FINAL REPORT HISTORY: 81 year old woman with COPD flair and cough. Please evaluate for pneumonia. AP UPRIGHT PORTABLE CHEST [**2118-6-5**] at 11:30 a.m.: No change from prior study dated [**2118-6-2**]. As before, there is retrocardiac density consistent with a hiatal hernia. No acute infiltrate or congestive failure. IMPRESSION: No acute cardiopulmonary disease. Hiatal hernia.

[**2118-6-10**] 6:01 AM CHEST (PORTABLE AP) Clip # [**Clip Number (Radiology) 13179**] Reason: eval for pulm effusions vs infiltrate Admitting Diagnosis: CHRONIC OBSTRUCTIVE PULMONARY DISEASE _____ [**Hospital 2**] MEDICAL CONDITION: 81 year old woman with COPD flare not improved with BIPAP and steroid. REASON FOR THIS EXAMINATION: eval for pulm effusions vs infiltrate _____ FINAL REPORT HISTORY: COPD flare, failing to improve. COMPARISON: [**2118-6-7**]. FINDINGS: AP portable supine view. The endotracheal tube remains in stable position. The previously noted coiled nasogastric tube is removed. There is a new feeding tube which extends below the left hemidiaphragm, and its tip is below the margin of the image. The heart, mediastinum and pulmonary vessels are within normal limits. The lung parenchyma appears stable, without opacities or nodules. There is no pleural effusion. There is bronchial wall thickening consistent with chronic bronchitis. A large hiatal hernia is again noted. IMPRESSION: Satisfactory position of the feeding tube. Otherwise, no interval change.

Selected publications

A data-driven approach to optimized medication dosing: a focus on heparin

— Ghassemi et al.
Intensive Care Medicine, 2015

Leveraging a critical care database: SSRI use prior to icu admission is associated with increased hospital mortality

— Ghassemi et al. Chest, 2013

Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study

— Pirracchio et al.
Lancet Respiratory Medicine, 2015

prev next

Selected publications

Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study

— Pirracchio et al.

Lancet Respiratory Medicine, 2015

A targeted real-time early warning score (TREWScore) for septic shock

— Henry et al.

Science Translational Medicine, 2015

Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension

— Mayaud et al.

Critical Care Medicine, 2013

prev next

AMIA Annual Symposium Proceedings Archive

AMIA Annu Symp Proc. 2017; 2017: 994–1003.
Published online 2018 Apr 16.

PMCID: PMC5977709
PMID: 29854167

Real-time mortality prediction in the Intensive Care Unit

Alistair E.W. Johnson, DPhil[1] and Roger G. Mark, MD PhD[1]

Author information ► Copyright and License information ► Disclaimer

Research | Open Access

An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study

Leo Anthony Celi ✉, L Hinske Christian, Gil Alterovitz and Peter Szolovits

Critical Care 2008 **12**:R151

https://doi.org/10.1186/cc7140

Received: 25 August 2008 | **Accepted:** 01 December 2008 | **Published:** 01 December 2008

The latest from MIT Critical Data

2017.HST.953: Collaborative Data Science in Medicine

on September 8, 2017

2017.HST.953: COLLABORATIVE DATA SCIENCE IN MEDICINE HST.953: Collaborative Data Science in Medicine, focuses on the secondary analysis of clinical data that is routinely collected in the process of care. In this course, students will work with Boston-area clinicians on research projects with the goal of a publication-ready manuscript at the end of the semester. Three of the 15

Critical Datathon 2015

on September 25, 2015

Datathon 2015 This weekend long (September 25-27) event brings together clinicians, data scientists and innovators in healthcare to address current problems in intensive care. The increasing wealth of patient data available through electronic health records has created a surge in research funding and industry interest in health data analytics. Challenges in extracting knowledge from health record databases, however, are significant. The

Critical Datathon Fall 2014

on September 5, 2014

Critical Datathon Fall 2014 Our 2nd Datathon brought together frontline healthcare providers (nurses, pharmacists, doctors) with data scientists to answer clinically-relevant questions over the course of a weekend. Participants will work with a large open-access ICU database called MIMIC, a creation of a public-private partnership between the Beth Israel Deaconess Medical Center (BIDMC), MIT, and Philips Healthcare. This weekend event ran in

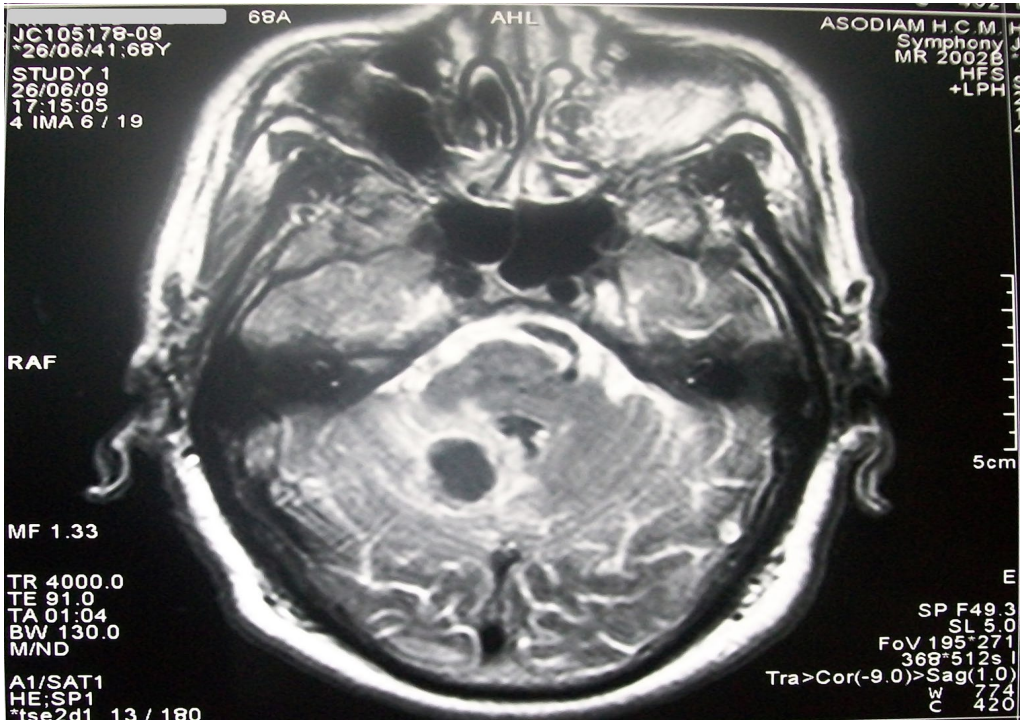
Critical Datathon Spring 2014

on January 3, 2014

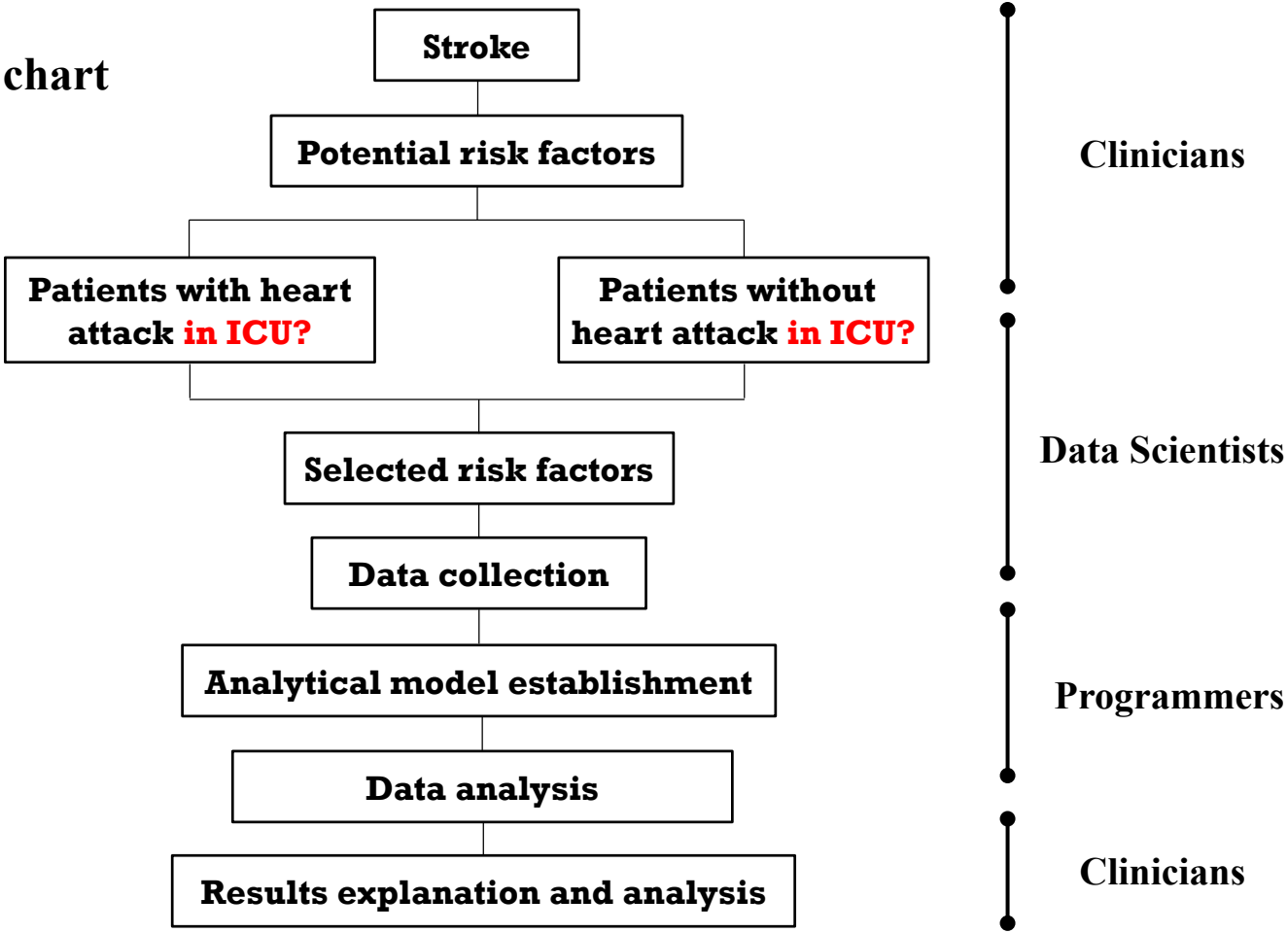
The inaugural Critical Data Marathon brought together various disciplines – computer science, medicine, nursing, pharmacy, biostatistics, epidemiology, informatics, business, health policy, and the social sciences – from both academia and industry. Watch the summary of the data marathon presented by Dr. Leo Celi at the Critical Data Conference. Stata Center, MIT, 3-5th January 2014 Check out the Google+


2018
医疗大数据与人工智能推动跨界融合
MEDICAL BIG DATA AND ARTIFICIAL INTELLIGENCE PROMOTE CROSS-DISCIPLINARY COLLABORATION
第三届解放军总医院-麻省理工学院
医疗大数据研讨及Datathon活动
THE 3RD PLAGH-MIT CRITICAL DATA CONFERENCE AND DATATHON
2018/11.30-12.2
中国 · 北京 2018 Nov.30-Dec.2 China Beijing
主办单位:
协办单位:
支持单位:
支持媒体:
中国报道、金融界、中国网、网易新闻、光明网、中国产经新闻网、健康时报、科学时报、中国日报、中国资讯网、央视网
会议日程
2018.11.30
内科大楼第一会议室
2018.12.1
内科大楼第十一会议室
2018.12.2
内科大楼第十一会议室
北京市科技委员会北京市自然科学基金
对外合作交流活动基金资助项目
www.plagh-mit-datathon.com

PREDICTION OF MYOCARDIAL INFARCTION IN ICU PATIENTS WITH ISCHEMIC STROKES

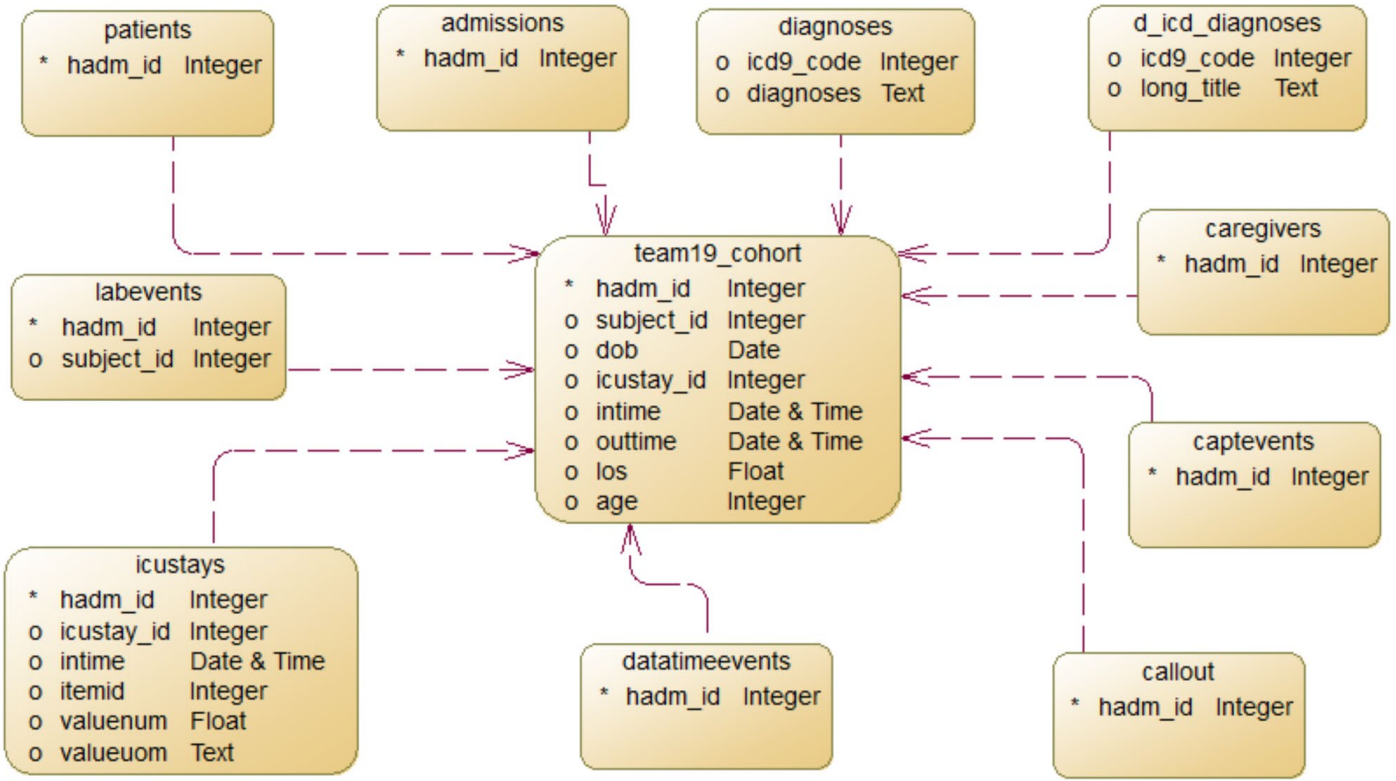


Flow chart


Stroke
Potential risk factors
Patients with heart attack in ICU?
Patients without heart attack in ICU?
Selected risk factors
Data collection
Analytical model establishment
Data analysis
Results explanation and analysis
Clinicians
Data Scientists
Programmers
Clinicians

BACKGROUND

- Which adult patients who recently were admitted to the ICU for a stroke will also have a heart attack?
- Using Risk Factors including:
  - Vital Signs: **BP, HR, RR, Oxygen Saturation**
  - Demographics: **Age, Gender, Ethnicity**
  - Laboratory: **Hgb, Hgb A1c, Glucose, Lactic Acid, Creatinine, Triglycerides**
- Investigated the **risk factors for the first 8 hour admission in the ICU**.
- Methods:
  - Extract data using SQL query
  - JOIN the relevant tables and perform statistical analysis / model building in Python
  - Analyze metrics including AUC in evaluating for the best model.
  - Deploy the model on the validation set and evaluate for effectiveness and accuracy.


patients
* hadm_id Integer
admissions
* hadm_id Integer
diagnoses
o icd9_code Integer
o diagnoses Text
d_icd_diagnoses
o icd9_code Integer
o long_title Text
team19_cohort
* hadm_id Integer
o subject_id Integer
o dob Date
o icustay_id Integer
o intime Date & Time
o outtime Date & Time
o los Float
o age Integer
caregivers
* hadm_id Integer
labevents
* hadm_id Integer
o subject_id Integer
captevents
* hadm_id Integer
icustays
* hadm_id Integer
o icustay_id Integer
o intime Date & Time
o itemid Integer
o valuenum Float
o valueuom Text
datatimeevents
* hadm_id Integer
callout
* hadm_id Integer

Project

sql code

- admissions.sql
- callout.sql
- captevents.sql
- caregivers.sql
- chatevents extract.sql
- create_table_team19.sql
- datatimeevents.sql
- diagnoses_icd.sql
- icustays.sql
- labevents.sql
- patients.sql

create_table_team19.sql

```sql
/* team_19*/
 --create table team19_cohort
create table mimiciii.team19_cohort as
with t1 as (
  select p.subject_id, p.dob, a.hadm_id,a.icustay_id, a.intime, a.outtime,a.los,
     EXTRACT(EPOCH from (a.intime-p.dob))/(365.25*24*3600) as age
from mimiciii.patients p
inner join mimiciii.icustays a
on p.subject_id=a.subject_id
where p.subject_id in
(select subject_id from mimiciii.diagnoses_icd
 where icd9_code in
 (select icd9_code from mimiciii.d_icd_diagnoses
  where lower(long_title) like '%cerebral infarction%'
  and icd9_code not like '3465%' and icd9_code != 'V1254')))
select * from t1 where age>18 and los*24>12
```

create_table_team19.sql 1:1 CRLF UTF-8 SQL GitHub Git (0)

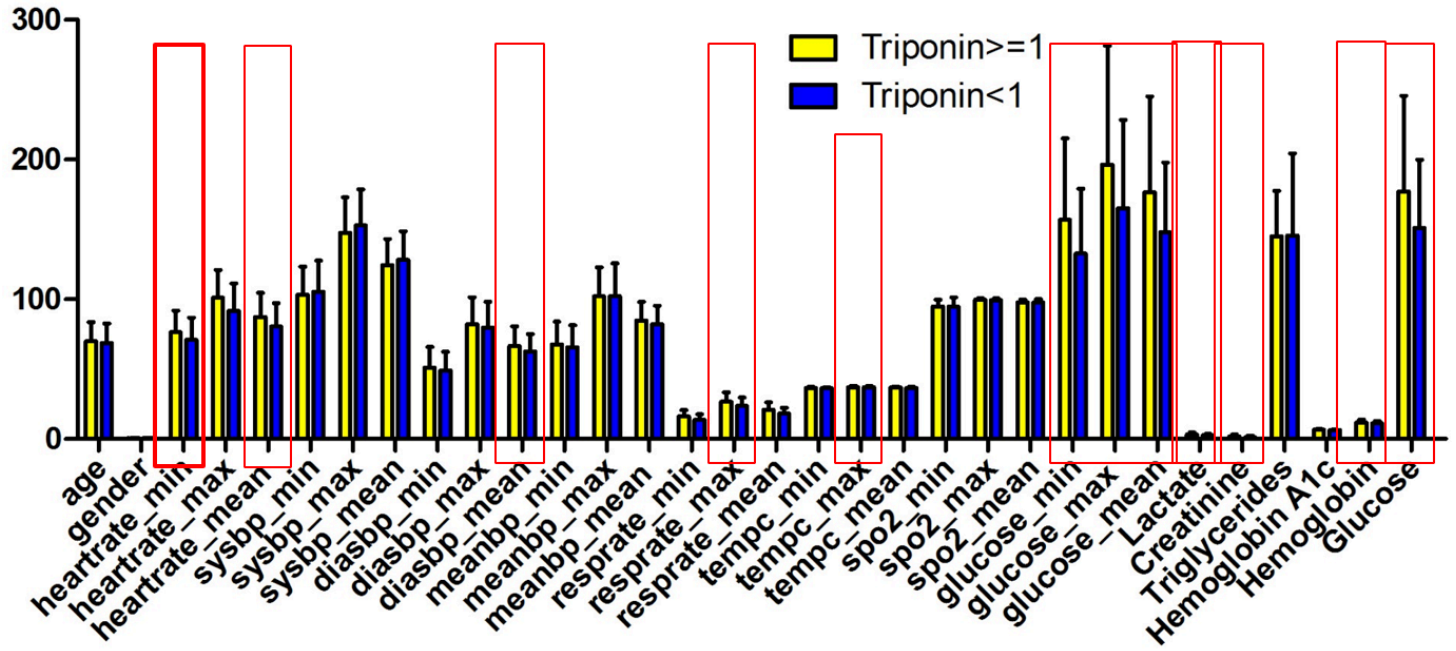
SQL QUERY

- Identified **2771** patients with ischemic stroke.
- **2662** patients had admission troponins.
- **2489** patients had all information including:
  - Demographics, Vital Signs, Labs
- **Troponin >= 1** considered positive for myocardial infarction.
- **249** patients had positive troponins (**7.4%**)

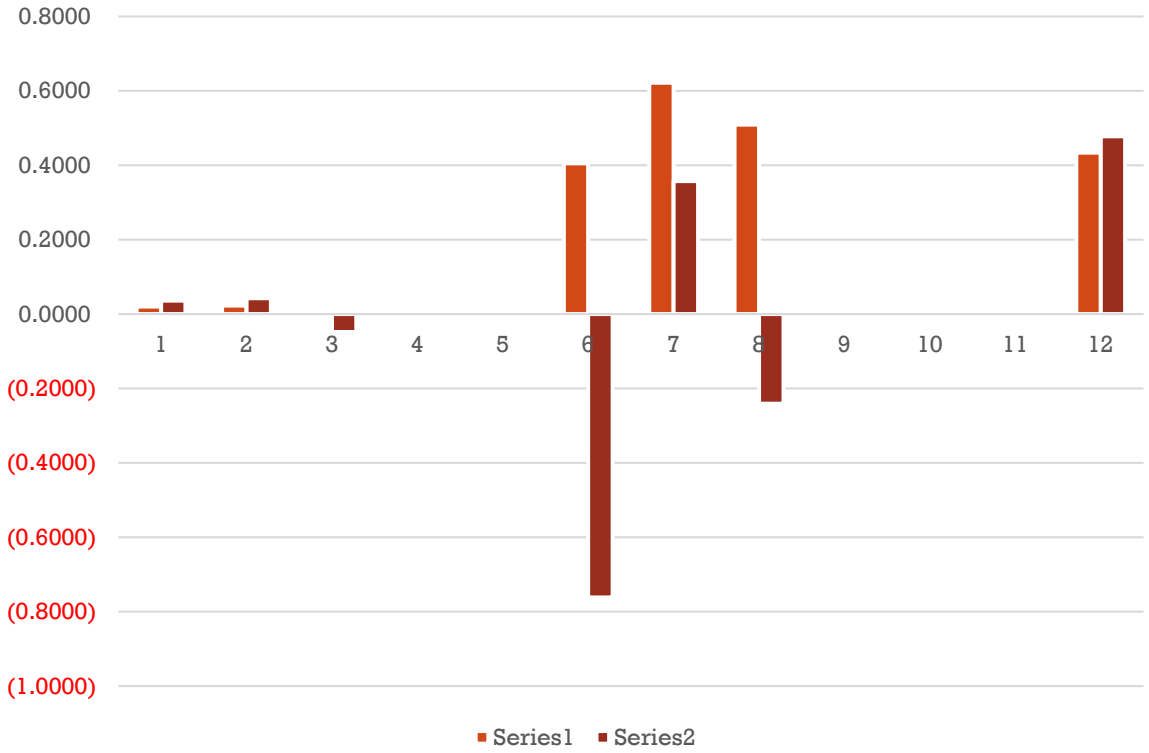
FEATURE SELECTION AND MODEL DEVELOPMENT

1. **T Test:** Used to find factors/features with significant statistical differences.
    1. **Reached statistical significance:** heartrate_min, heartrate_max, heartrate_mean, diasbp_mean, resprate_max, resprate_mean, tempc_max, glucose_min, glucose_max, glucose_mean, lactate, creatinine, hemoglobin A1c, hemoglobin, glucose
2. **Model Development:**
    1. **Principle Component Analysis:** PCA was used to reduced the dimension of all selected features to two principle components.
    2. **Comparison of several classifiers:** Logistic Regression, Naïve Bayes, Random Forest, Decision Tree, SVM, etc.

[ ]:
```
#!usr/bin/env python
# -*- coding:utf-8 _*-
"""
@author:qzp
@file: classifier_sample.py
@time: 2017/10/{DAY}
"""
from sklearn import datasets
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
from sklearn.cross_validation import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.datasets import make_classification
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier, AdaBoostClassifier
import random

h = .02  # step size in the mesh

names = ["Linear SVM", "RBF SVM", "Random Forest", "AdaBoost", "XgBoost"]
classifiers = [
    SVC(kernel="linear", C=0.025),
    SVC(gamma=2, C=1),
    RandomForestClassifier(max_depth=5, n_estimators=10, max_features=1),
    AdaBoostClassifier()
]

X, y = make_classification(n_features=2, n_redundant=0, n_informative=2, random_state=1, n_clusters_per_cla
# car_data = pd.read_csv('./data/Car_Evaluation.csv')
# y2 = car_data[:, -1]
# car = car_data[:, :-2]
forest = datasets.fetch_covtype(data_home=None, download_if_missing=True, random_state=None, shuffle=False)
forest_data = forest.data[0:1000]
y2 = forest.target[0:1000]
rng = np.random.RandomState(2)
X += 2 * rng.uniform(size=X.shape)
```


300
200
100
0
Triponin>=1
Triponin<1
age
gender
heartrate_min
heartrate_max
heartrate_mean
sysbp_min
sysbp_max
sysbp_mean
diasbp_min
diasbp_max
diasbp_mean
meanbp_min
meanbp_max
meanbp_mean
resprate_min
resprate_max
resprate_mean
tempc_min
tempc_max
tempc_mean
spo2_min
spo2_max
spo2_mean
glucose_min
glucose_max
glucose_mean
Lactate
Creatinine
Triglycerides
Hemoglobin A1c
Hemoglobin
Glucose

PCA


0.8000
0.6000
0.4000
0.2000
0.0000
(0.2000)
(0.4000)
(0.6000)
(0.8000)
(1.0000)
1
2
3
4
5
6
7
8
9
10
11
12
Series1
Series2

MODEL PERFORMANCE

- K Nearest Neighbors: 0.9688755
- **Linear SVM**: 0.9738955
- SBF SVM: 0.97289156
- **Decision Tree**: 0.97389558
- Random Forest: 0.9688755
- Neural Net: 0.97188755
- **AdaBoost**: 0.97389558
- Naïve Bayes: 0.9698795
- QDA score: 0.963855

STUDY WEAKNESSES

- Did we choose the right features?
- Retrospective data in a single hospital site.
- Potential issues with imperfect data extraction.
- Unbalanced Data (~7% of positive troponin)
- Did not have time for cross-validation.
  - 80% Training, 20% Testing
- Does this ultimately reflect the real world?

DATATHON CONCLUSION

- The accuracy is exceedingly high.
  - However, with a severe imbalance of data (7%), we would automatically have a 93% accuracy if we had model that predicted only **No.**
- If we indeed are able to predict with 97% accuracy, this could potentially be helpful.
- However, we had not validated this model, and in all likelihood, will need more time to further develop the model.


QUESTIONS

Adopting AI in Healthcare – NLP and ML

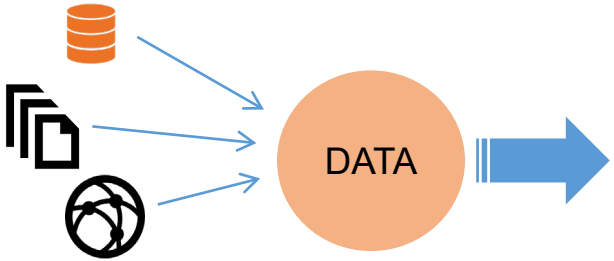
DRAFT

Niteen Kumar, Data Scientist

Feb 12, 2019

Natural Language Processing (NLP)

Natural language processing is an automated way to understand, analyze natural human languages and extract information from such data by applying machine algorithms



Automated

Machine algorithms
(mathematics and statistics)

Extract
information

Machine translation

Sentiment analysis

Insight to unstructured data

The Real Challenge

Unstructured text

Tons of data

Knowledge about languages

Knowledge about the world

Quantitative analysis on unstructured data such as texts, documents

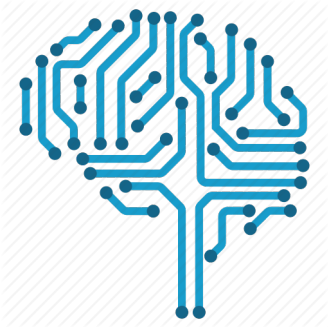
Ambiguity (context Vs. raw meaning)

Web streaming APIs

IOT data

The Solution - NLP

Full automation through modern software libraries



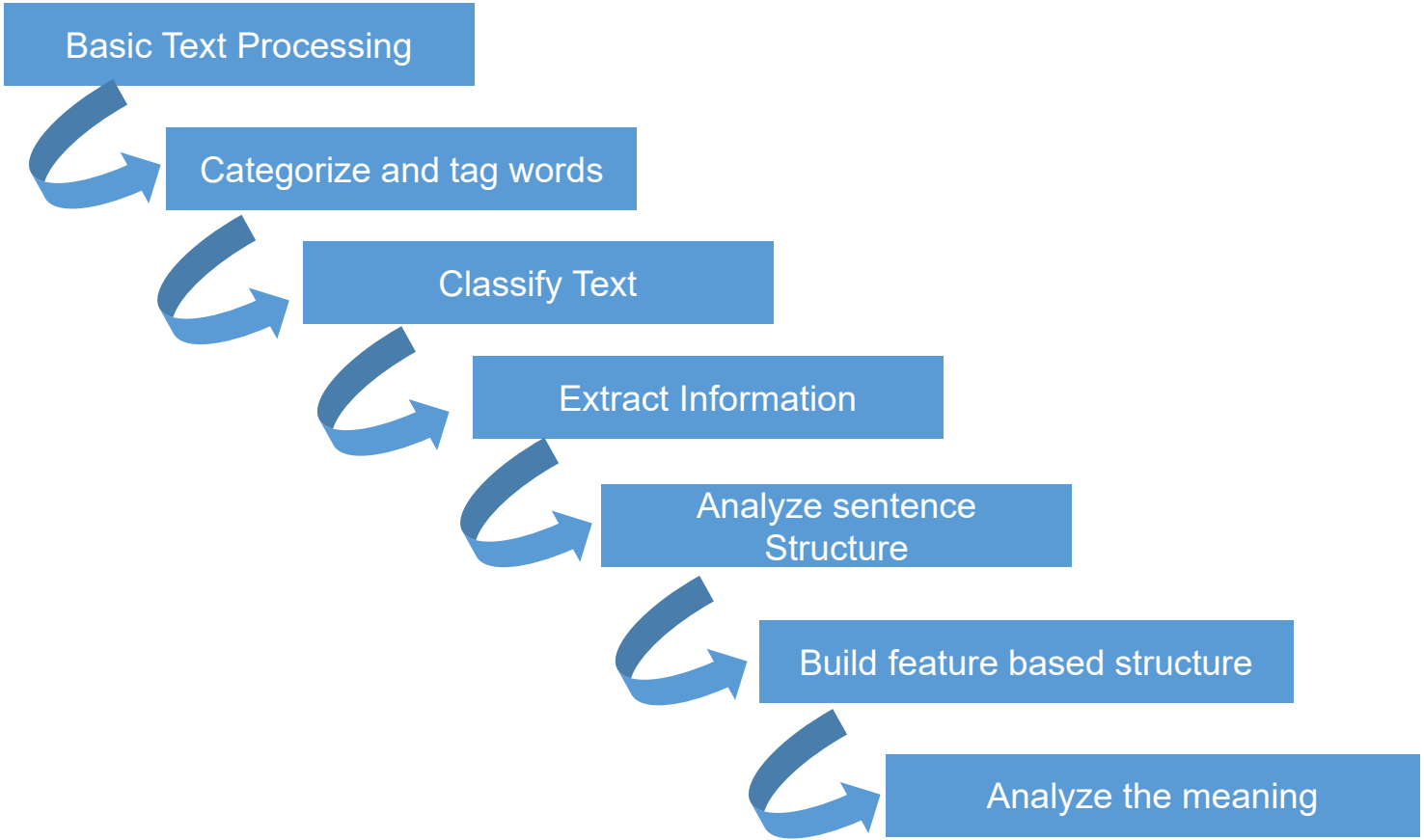
Intelligent processing through machine models

Knowledge about Languages and world (software libraries/ packages)

NLP Terminology

Word boundaries	Determine where one ends and other begins
Tokenization	Tokens are words, phrases , idioms
Stemming	Map to the valid root word
Tf-idf	Term frequency and inverse document frequency
Semantic analytics	Analyze relationship between set of documents
Disambiguation	Meaning and sense of word (context Vs. intent)
Topic models	Discover topics in collection of documents

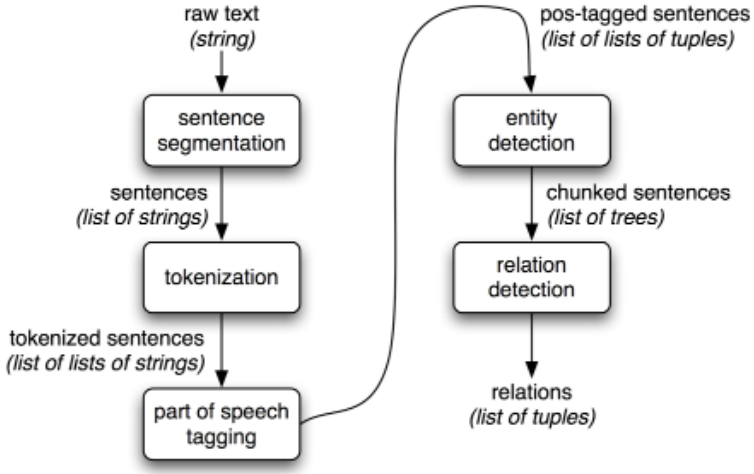
The NLP Approach Text data



The NLP POS Tagging

Pattern Module for POS tagging

Sentence and Word Tokenization Techniques



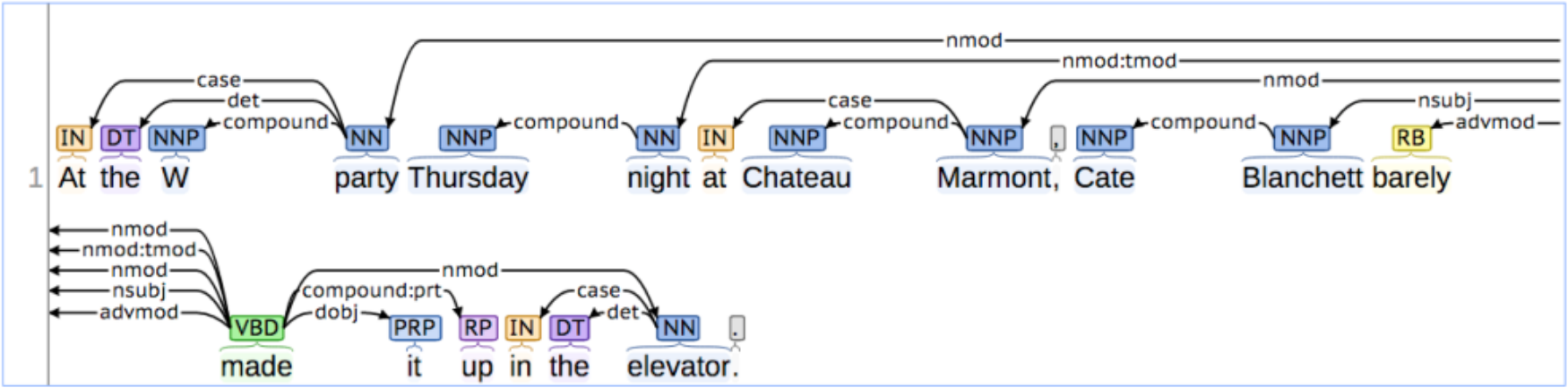
Stanford Core NLP NER Tagging

Named Entity Recognition:

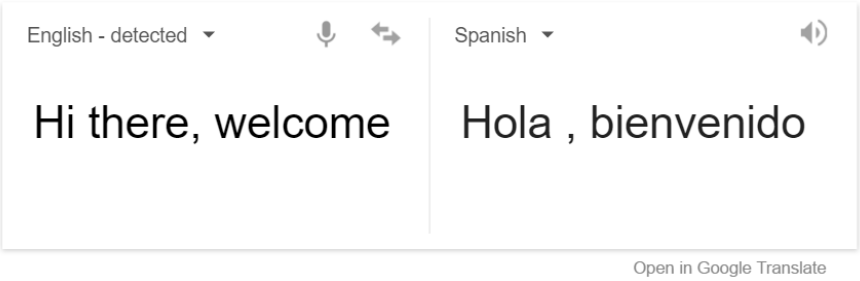
1 At the W party Thursday night at Chateau Marmont, Cate Blanchett barely made it up in the elevator.

(Date: Thursday; Time: night; Location: Chateau Marmont; Person: Cate Blanchett)

Basic Dependencies:



The NLP Applications



Machine translation

Speech recognition

Sentiment Analysis

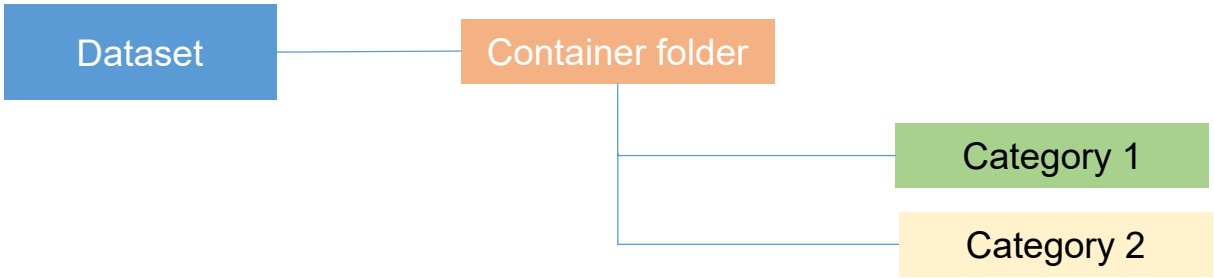
The SciKit Learn Approach

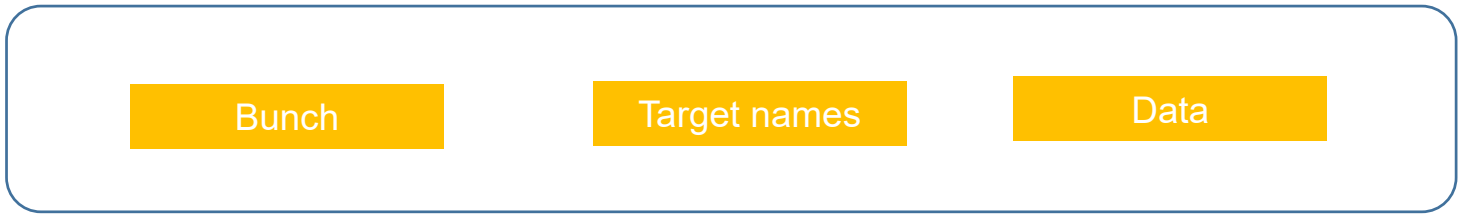
A very powerful library with set of modules to process and analyze natural language data such as texts and images and extract information using machine learning algorithms

Modules to load contents and categories	Feature extraction	Model training
Pipeline building mechanism	Performance optimization	Grid search for finding good parameters

Modules to load content and category

Built in modules for loading the dataset contents and categories





Feature Extraction

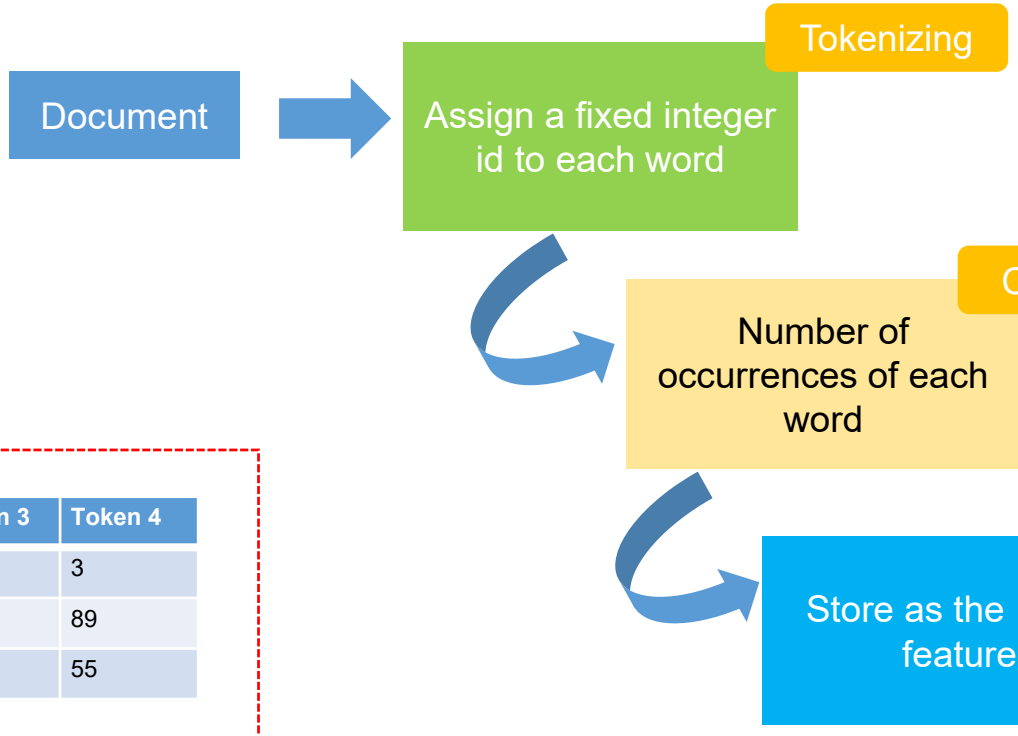
It’s a technique to convert the content into the numerical vectors to perform machine learning

Text feature extraction

Image feature extraction

Bag of words

Text data converted into numerical feature vectors with fixed size



Corpus of document

	Token 1	Token 2	Token 3	Token 4
Document 1	42	32	119	3
Document 2	1118	0	0	89
Document 3	0	0	0	55

Text Feature Extraction Considerations

Sparse	Utility to deal with sparse matrix while storing them in memory
Vectorizer	Implements tokenization and occurrence
Tf-idf	Term weighing utility for term frequency and inverse document frequency
Decoding	Utility to decode text files based on provided files encoding

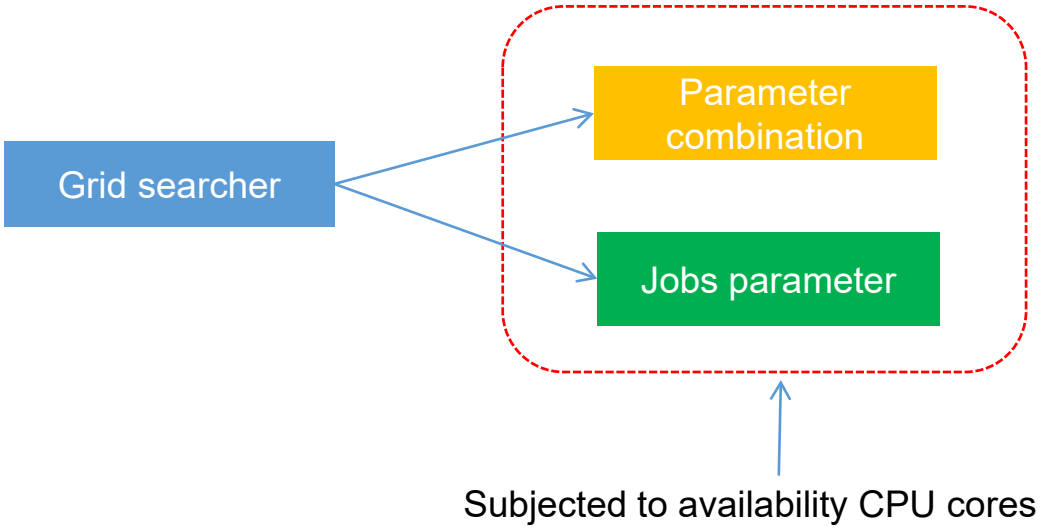
Model training

Predict the outcome using the feature extracting mechanism and train the model

Supervised	Models to train document classifiers Ex: classification of text documents using Naïve Bays, SVM, linear regression, KNN neighbors
Unsupervised	Group documents by applying clustering algorithms Ex: clustering text document using K means

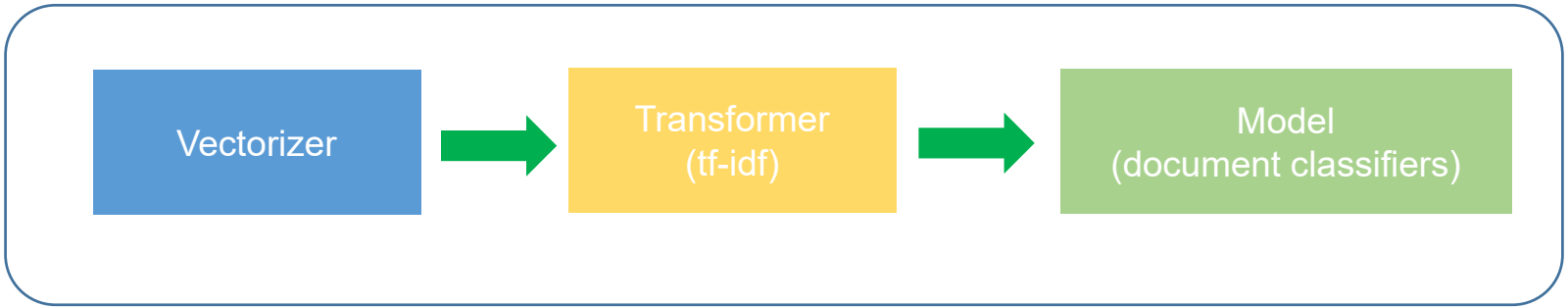
Grid search and multiple parameters

A document classifiers can have many parameters and a Grid approach helps to search the best parameters for model training and predicting the outcome accurately



Putting them together - Pipeline

Built in function to build pipeline and train the model



Tensorflow on Sagemaker

...

February 12, 2019

Overview

- Tensorflow
- Sagemaker
- IAM
- Productionize a model

Tensorflow

Tensorflow

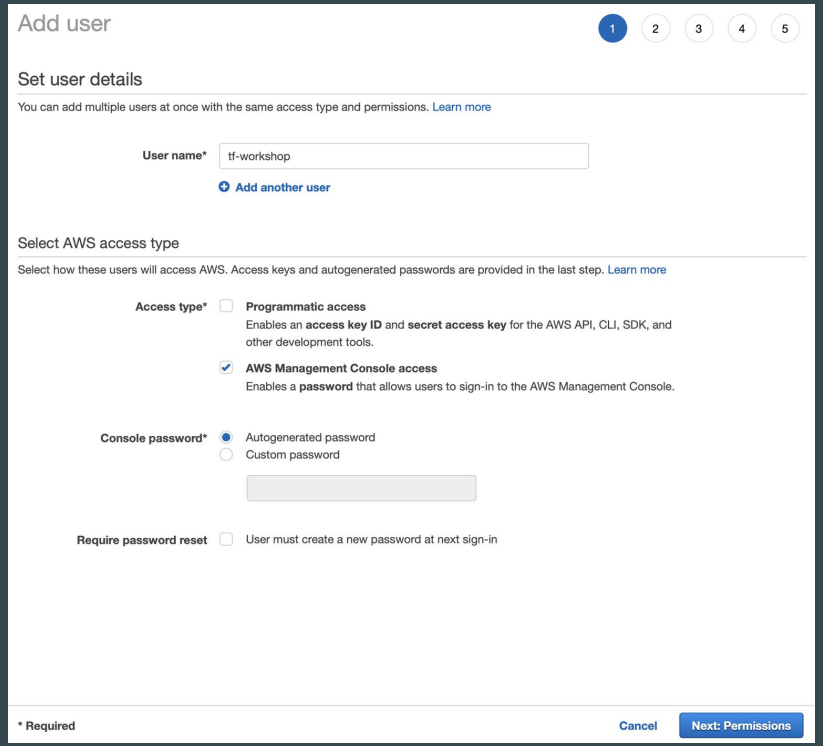
- foobar

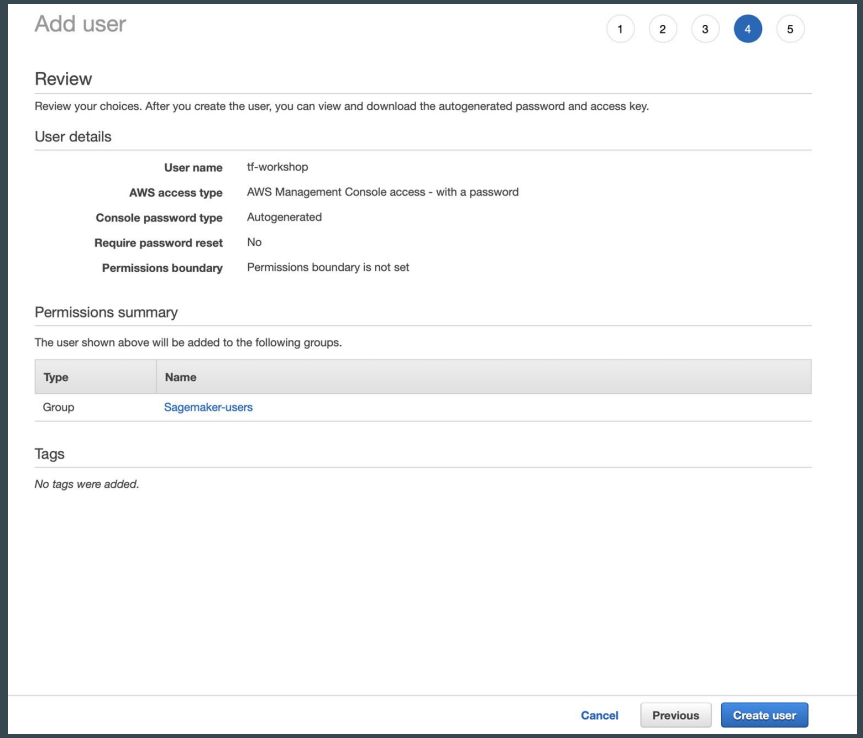
SageMaker

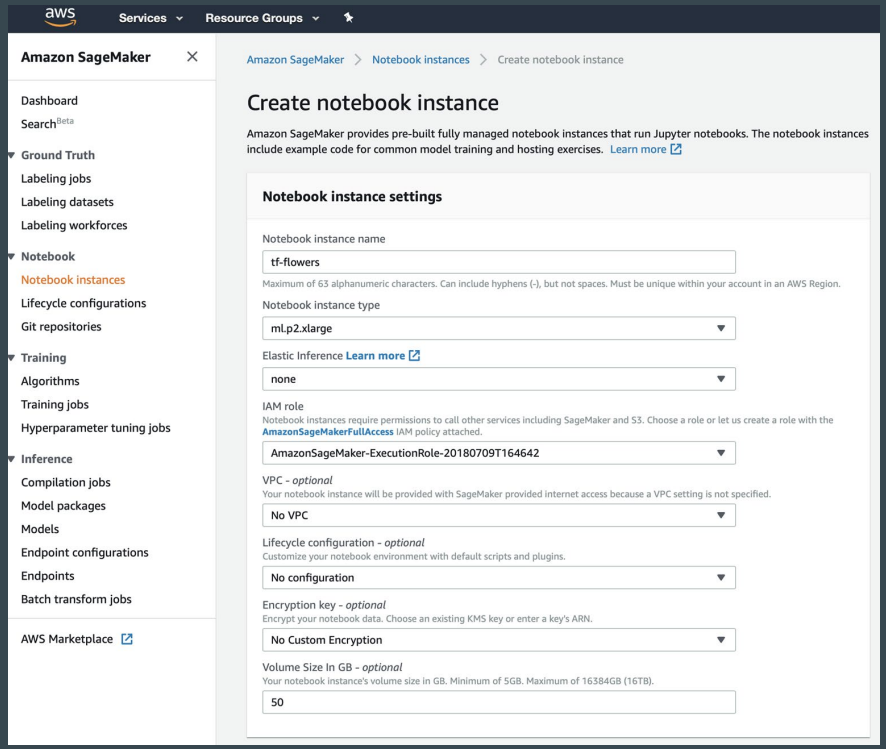
Overview

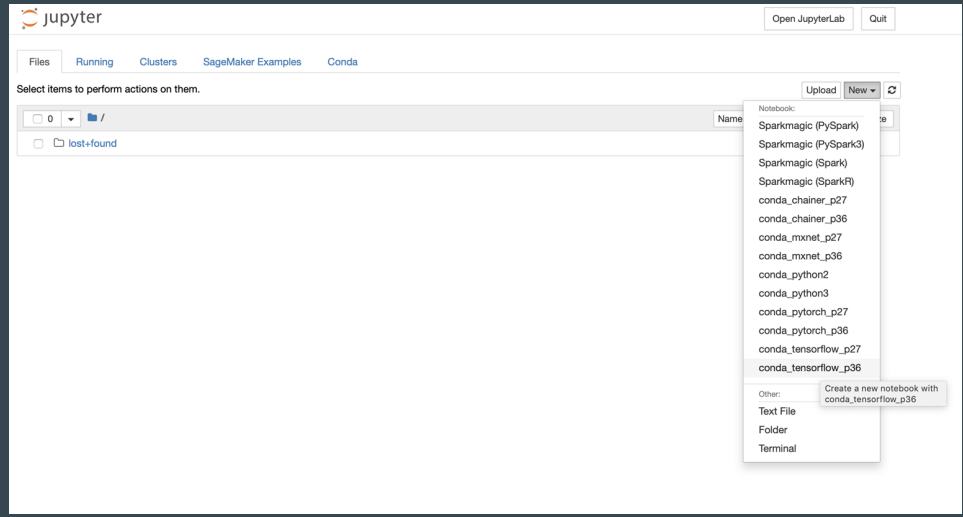
- Tensorflow
- Sagemaker
- Productionize a model

Let’s make something


Add user
1 2 3 4 5
Set user details
You can add multiple users at once with the same access type and permissions. Learn more
User name*
tf-workshop
Add another user
Select AWS access type
Select how these users will access AWS. Access keys and autogenerated passwords are provided in the last step. Learn more
Access type*
Programmatic access
Enables an access key ID and secret access key for the AWS API, CLI, SDK, and other development tools.
AWS Management Console access
Enables a password that allows users to sign-in to the AWS Management Console.
Console password*
Autogenerated password
Custom password
Require password reset
User must create a new password at next sign-in
* Required
Cancel
Next: Permissions


Add user
1 2 3 4 5
Review
Review your choices. After you create the user, you can view and download the autogenerated password and access key.
User details
User name
tf-workshop
AWS access type
AWS Management Console access - with a password
Console password type
Autogenerated
Require password reset
No
Permissions boundary
Permissions boundary is not set
Permissions summary
The user shown above will be added to the following groups.
Type
Name
Group
Sagemaker-users
Tags
No tags were added.
Cancel
Previous
Create user


aws
Services
Resource Groups
Amazon SageMaker
Dashboard
Search Beta
Ground Truth
Labeling jobs
Labeling datasets
Labeling workforces
Notebook
Notebook instances
Lifecycle configurations
Git repositories
Training
Algorithms
Training jobs
Hyperparameter tuning jobs
Inference
Compilation jobs
Model packages
Models
Endpoint configurations
Endpoints
Batch transform jobs
AWS Marketplace
Amazon SageMaker > Notebook instances > Create notebook instance
Create notebook instance
Amazon SageMaker provides pre-built fully managed notebook instances that run Jupyter notebooks. The notebook instances include example code for common model training and hosting exercises. Learn more
Notebook instance settings
Notebook instance name
tf-flowers
Maximum of 63 alphanumeric characters. Can include hyphens (-), but not spaces. Must be unique within your account in an AWS Region.
Notebook instance type
ml.p2.xlarge
Elastic Inference Learn more
none
IAM role
Notebook instances require permissions to call other services including SageMaker and S3. Choose a role or let us create a role with the AmazonSageMakerFullAccess IAM policy attached.
AmazonSageMaker-ExecutionRole-20180709T164642
VPC - optional
Your notebook instance will be provided with SageMaker provided internet access because a VPC setting is not specified.
No VPC
Lifecycle configuration - optional
Customize your notebook environment with default scripts and plugins.
No configuration
Encryption key - optional
Encrypt your notebook data. Choose an existing KMS key or enter a key's ARN.
No Custom Encryption
Volume Size In GB - optional
Your notebook instance's volume size in GB. Minimum of 5GB. Maximum of 16384GB (16TB).
50


jupyter
Open JupyterLab
Quit
Files
Running
Clusters
SageMaker Examples
Conda
Select items to perform actions on them.
Upload
New
lost+found
Notebook:
Sparkmagic (PySpark)
Sparkmagic (PySpark3)
Sparkmagic (Spark)
Sparkmagic (SparkR)
conda_chainer_p27
conda_chainer_p36
conda_mxnet_p27
conda_mxnet_p36
conda_python2
conda_python3
conda_pytorch_p27
conda_pytorch_p36
conda_tensorflow_p27
conda_tensorflow_p36
Other:
Text File
Folder
Terminal
Create a new notebook with conda_tensorflow_p36

Thanks!

- http://harrymoreno.com
- http://twitter.com/morenoh149

virtual force

TECHNOLOGY ADVISORY, INNOVATION, DESIGN & ENGINEERING

48 Wall Street, 5th Floor, Suite #9, New York, NY, 10005

+1 718 395 9793 | innovate@virtualforce.io | virtualforce.io

OUTLINE

- Introduction
- Overview of Virtual Force
- Product Development Best Practices
- Case Study: Arla Foods
- Case Study: Counseltyics
- Questions & Answers

Introduction

Overview of Virtual Force

Startup Agility with Enterprise Grade Quality

Virtual force brings in the concept of Lean Startup Methodology to help Enterprises transform their business.

For more than 7 years, Virtual Force has been serving as an innovation partner for small and large enterprises.

Our innovative agile process enables solving complex business challenges, experimenting with innovative ideas and rolling out reliable products at rapid speed.

Impact Thus Far

$ 150+ Million

Raised by our portfolio

1 Million+

Engineering Hours

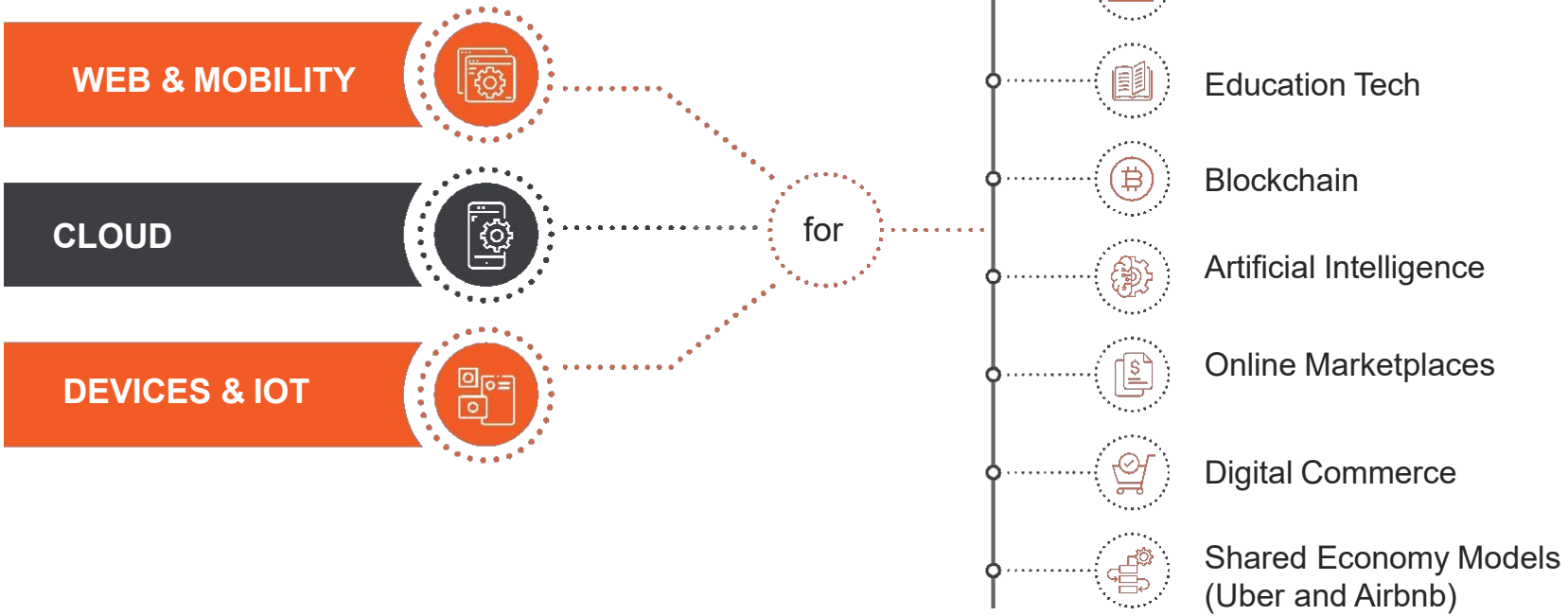
$200 Million+

Value Creation

100 Mil+

End User Touchpoints

Diversified Technology Portfolio



Our portfolio has GONE BIG!

Virtual force's supported enterprises have made headlines across the globe.

PRODUCEPAY

FUNDING

$90M+

Forbes 2018 most innovative AgTech companies

HEALTHRHYTHMS

GRANT MONEY

$2.1M

Top 20 Masschallenge HealthTech Companies

MOVING ANALYTICS

TOTAL FUNDING AMOUNT

$2.5M

Top 20 Masschallenge HealthTech Companies

Moodfit

TOTAL FUNDING AMOUNT

$400K

Top 50 startups in MENA

LoveStoriesTV

The place to watch wedding videos.

SEED FUNDING ROUND

$1.7M

Youtube for Wedding videos

GETMYCAR

TOTAL FUNDING AMOUNT

€2.5M

Europe based P2P car rental platform

BRANCHING MINDS

GRANT MONEY

$910K

Better learning through cognitive science.

vidcode

TOTAL FUNDING AMOUNT

$1.5M

Y-C backed EdTech startup

TOTAL FUNDING AMOUNT

$400K

Acquired by Shuttle

SWEETEN

TOTAL FUNDING AMOUNT

$9.8M

Match-making platform for Home Owners and Contractors

Acquisitions:

by SHUTTLE

StudyAbroad Apartments by Roomi

COUNSELYTICS by conga

Innovation Partners with Global Leaders

Virtual Force has been entrusted by leading Global Enterprises

Technology

Value-Added Solutions Partner since 2014

FMCG

Trusted Technology Partner since 2015

Real Estate

Enterprise Application Development Services and Testing Center of Excellence

Telco

Infrastructure Upgrade and Implementation Partner

Banking

Digital Transformation Consultancy and Implementation Partner

Value-Added Partners for One of the Largest:

- Real Estate Developer
- Oil Company
- Hedge Fund
- Media Agency
- Telco
- Information Technology Company

Selection of Work in AI and Blockchain

Retailytics

IBM

Efficiency through Retail Analytics

Retailytics is an IBM Watson-backed IoT-integrated digital solution that empowers retailers to gauge various metrics within their physical store or stock storage data. With the use of the Retailytics app, the retailer can calculate a wide number of datasets useful for flux management, stock replenishment solution, customer behavior and preference analysis, etc.

VF Role

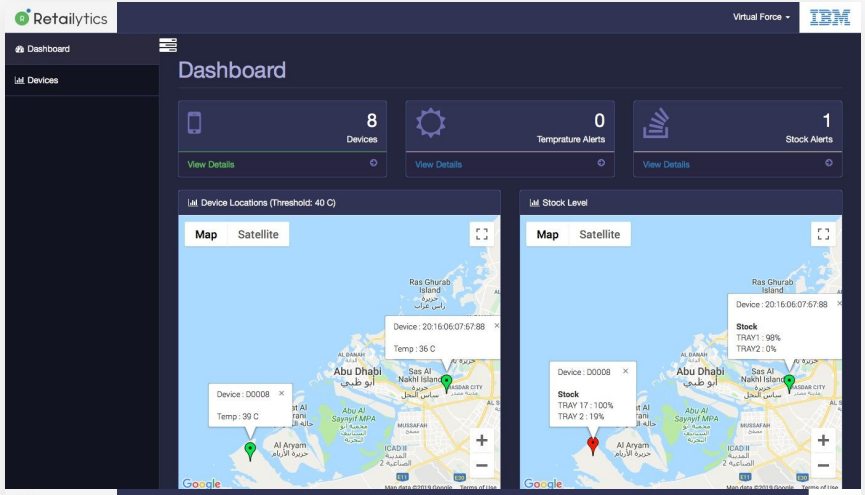
Ideation, Concept, Hardware Prototyping, Mockups, Research, Development, and Maintenance

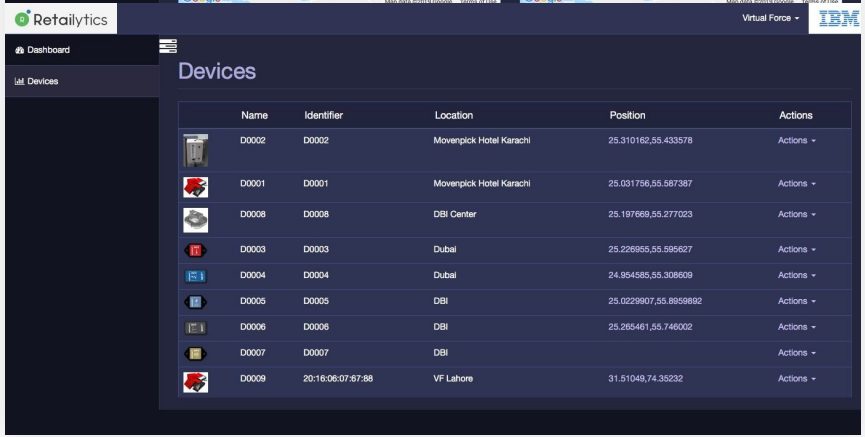
Tech

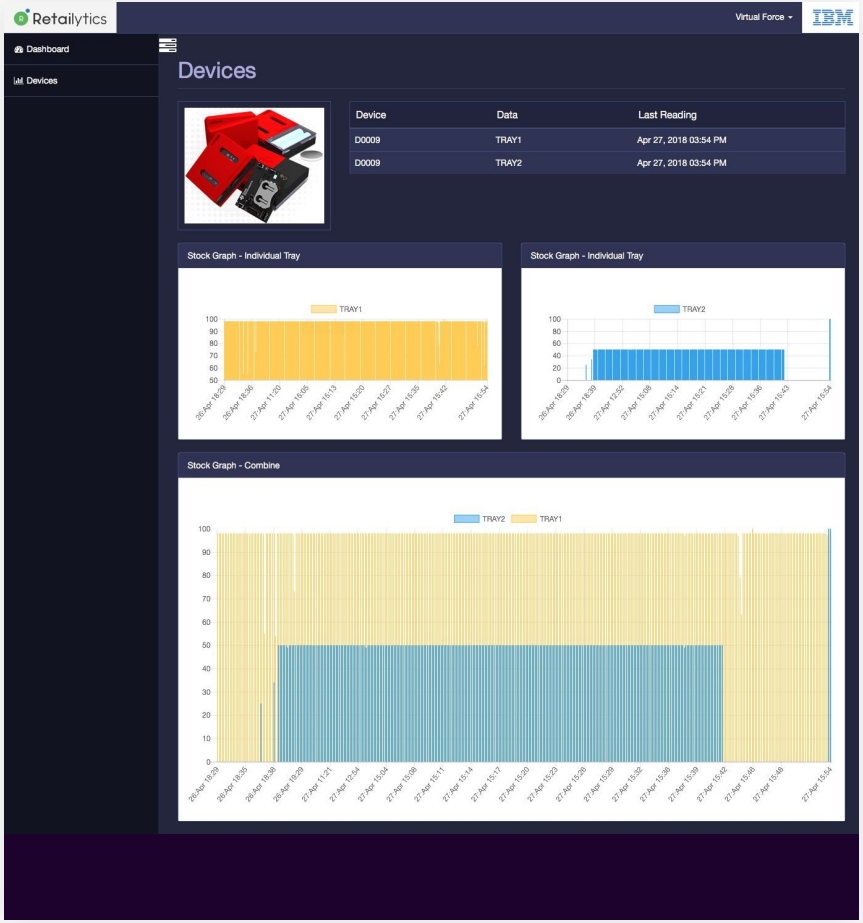
IBM IoT Platform, Node-RED, Twitter API, SendGrid, Twilio, PHP Laravel, Bootstrap, Google Maps, MySQL

Project Link

https://cloud.ibm.com/catalog?category=ai


Retailytics
Virtual Force
IBM
Dashboard
Devices
Dashboard
8
Devices
View Details
0
Temprature Alerts
View Details
1
Stock Alerts
View Details
Device Locations (Threshold: 40 C)
Map
Satellite
Device : 20:16:06:07:67:88
Temp : 36 C
Device : D0008
Temp : 39 C
Stock Level
Map
Satellite
Device : 20:16:06:07:67:88
Stock
TRAY1 : 99%
TRAY2 : 0%
Device : D0008
Stock
TRAY 17 : 100%
TRAY 2 : 19%


Retailytics
Virtual Force
IBM
Dashboard
Devices
Devices
Name
Identifier
Location
Position
Actions
D0002 D0002 Movenpick Hotel Karachi 25.310162,55.433578 Actions
D0001 D0001 Movenpick Hotel Karachi 25.031756,55.587387 Actions
D0008 D0008 DBI Center 25.197669,55.277023 Actions
D0003 D0003 Dubai 25.226955,55.595627 Actions
D0004 D0004 Dubai 24.954585,55.308609 Actions
D0005 D0005 DBI 25.0229907,55.8959892 Actions
D0006 D0006 DBI 25.265461,55.746002 Actions
D0007 D0007 DBI Actions
D0009 20:16:06:07:67:88 VF Lahore 31.51049,74.35232 Actions


Retailytics
Virtual Force
IBM
Dashboard
Devices
Devices
Device
Data
Last Reading
D0009 TRAY1 Apr 27, 2018 03:54 PM
D0009 TRAY2 Apr 27, 2018 03:54 PM
Stock Graph - Individual Tray
TRAY1
Stock Graph - Individual Tray
TRAY2
Stock Graph - Combine
TRAY2
TRAY1

Virtual Force Achievements

-Flux Management System
-Stock Replenishment Solution
-Data Tracking and Meaningful Insights
-Customer Behavior
-Customer Preference Trends
-Physical Value Alerts (Temperature, Low Stock, etc.)

Challenge

For retailers, there is a huge crisis and revenue loss in the form of inefficient merchandising practices, impractical and insufficient supply network, inventory and product waste, and, the resulting dissatisfaction. Virtual Force took it upon itself to utilize its resources to create a holistic automated and effective process to analyze consumer-generated and stock-related data. The challenge was to translate physical metrics on a digital platform furthering a physical response and outcome.

Solution

Flux Management System

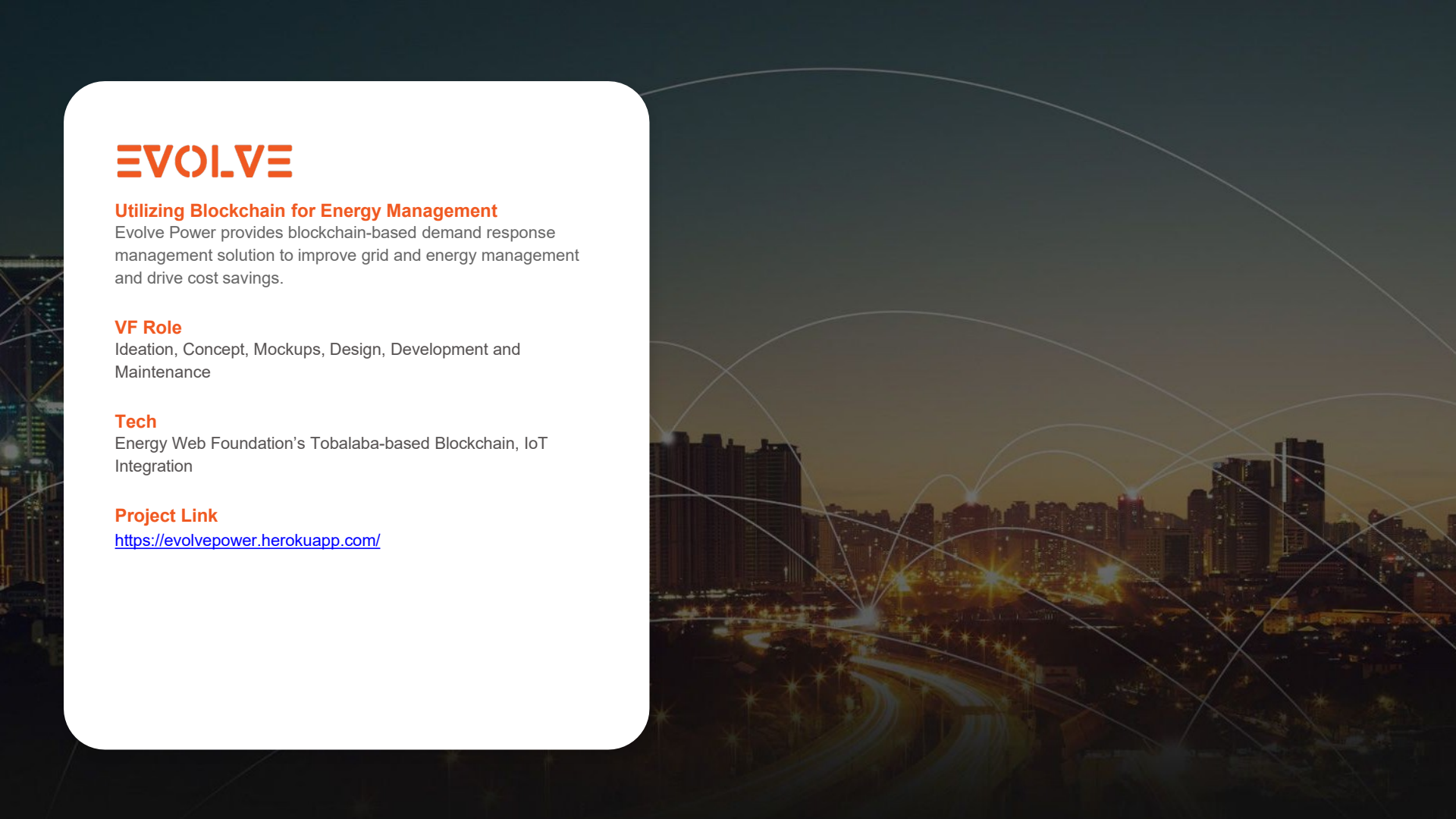
By utilizing IoT integrations with IBM Watson, Virtual Force created Retailytics Devices, strategically placed to collect insightful data such as customer preference, logistic flow, time mapping, counter checks, visitor count, etc. The collected data is visualized within Retailytics app to understand customers' shopping behavior and provide them with an optimal experience at all touchpoints. Retailytics enhances customer experience, improves operational performance, optimizes in-store logistics, reduces checkout times, and, improves conversion rates.

Stock Replenishment Solution

Virtual Force created a solution for stock replenishment via IBM Watson-backed IoT-integrated Retailytics devices. With a user dashboard accessible online and within an app, every retailer can stay updated about the status of their Smart Chillers. The user can view, and be alerted for temperature fluctuations, opening-closing cycles, GPS location of the supply network, as well as a Keep Alive Signal which reports chiller health. With these metrics being monitored in real-time by cognitive IoT, retailers get a more responsive, efficient and transparent supply network.

Features

Sensor-enabled Counter Checks
Low Stock Alerts
Temperature Alerts
Geo-tracking
Customized Alerting Rules
Historical and Trending Charts
Admin Panel



EVOLVE

Utilizing Blockchain for Energy Management

Evolve Power provides blockchain-based demand response management solution to improve grid and energy management and drive cost savings.

VF Role

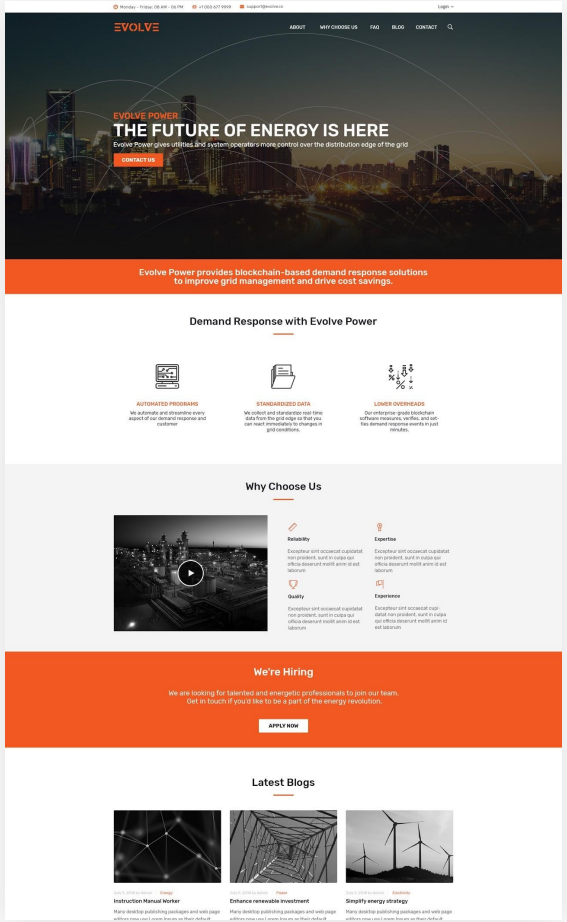
Ideation, Concept, Mockups, Design, Development and Maintenance

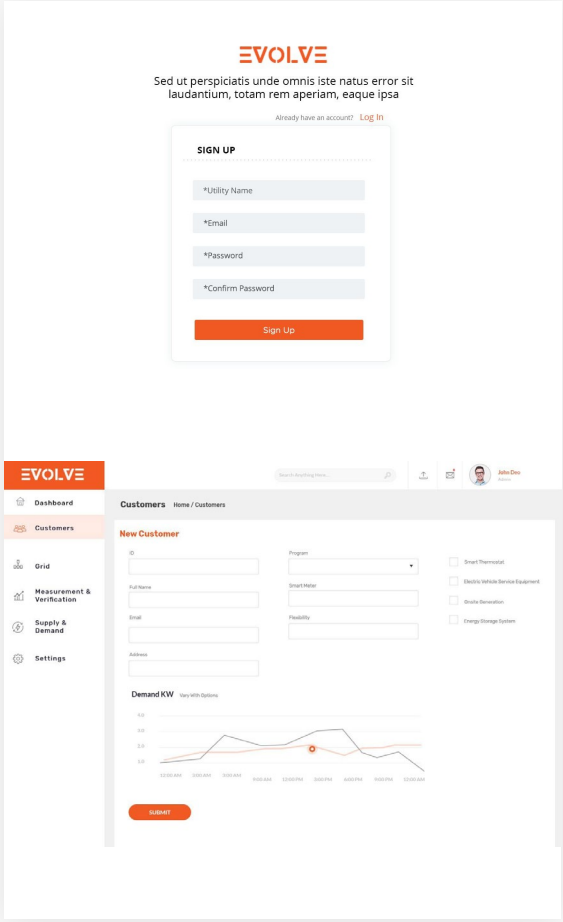
Tech

Energy Web Foundation’s Tobalaba-based Blockchain, IoT Integration

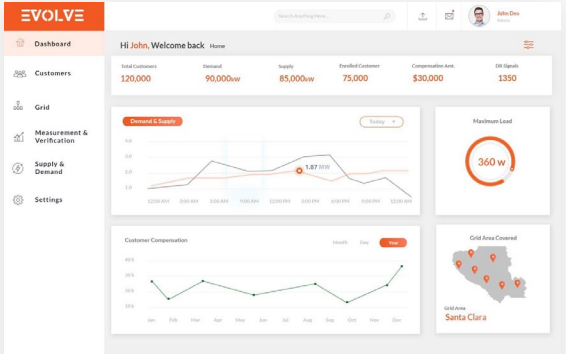
Project Link

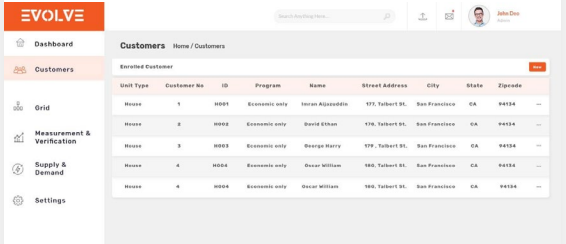
https://evolvepower.herokuapp.com/

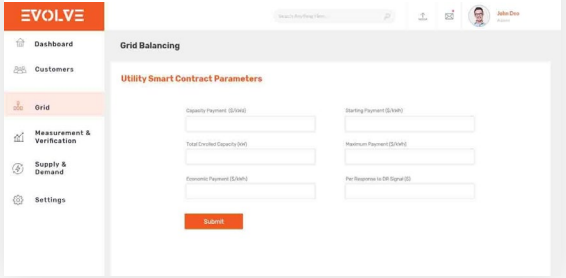

EVOLVE
ABOUT
WHY CHOOSE US
FAQ
BLOG
CONTACT
EVOLVE POWER
THE FUTURE OF ENERGY IS HERE
Evolve Power gives utilities and system operators more control over the distribution edge of the grid
CONTACT US
Evolve Power provides blockchain-based demand response solutions to improve grid management and drive cost savings.
Demand Response with Evolve Power
AUTOMATED PROGRAMS
STANDARDIZED DATA
LOWER OVERHEADS
Why Choose Us
Reliability
Expertise
Quality
Experience
We're Hiring
We are looking for talented and energetic professionals to join our team.
Get in touch if you'd like to be a part of the energy revolution.
APPLY NOW
Latest Blogs
Instruction Manual Worker
Enhance renewable investment
Simplify energy strategy


EVOLVE
Sed ut perspiciatis unde omnis iste natus error sit laudantium, totam rem aperiam, eaque ipsa
Already have an account? Log In
SIGN UP
*Utility Name
*Email
*Password
*Confirm Password
Sign Up

EVOLVE
Dashboard
Customers
Grid
Measurement & Verification
Supply & Demand
Settings
Customers
Home / Customers
New Customer
Demand KW
SUBMIT


EVOLVE
Dashboard
Customers
Grid
Measurement & Verification
Supply & Demand
Settings
Hi John, Welcome back
Home
Total Customers
120,000
Demand
90,000kW
Supply
85,000kW
Enrolled Customer
75,000
Compensation Amt.
$30,000
DR Signals
1350
Demand & Supply
Maximum Load
360 w
Customer Compensation
Grid Area Covered
Grid Area
Santa Clara


EVOLVE
Dashboard
Customers
Grid
Measurement & Verification
Supply & Demand
Settings
Customers
Home / Customers
Enrolled Customer
Unit Type
Customer No
ID
Program
Name
Street Address
City
State
Zipcode


EVOLVE
Dashboard
Customers
Grid
Measurement & Verification
Supply & Demand
Settings
Grid Balancing
Utility Smart Contract Parameters
Submit

Virtual Force Achievements

- Distributed Ledger Blockchain Technology Implementation
- Blockchain-based Decentralized App (dApp) Development
- Automated Demand Response Management System
- P2P Smart Contracts Self-Enforcement
- IoT Integration for DRMS via dApp
- Utility/Electricity Provider & Consumer Dashboards
- Real-Time Measurement & Verification
- Blockchain Recordkeeping, Monitoring & Network Security

Challenge

Electric utilities & providers have a hard time meeting peak demand. A great number of their consumers are willing to reduce their electricity consumption at the utility's request (demand response). Their demand response is manually administered, inefficient & time-consuming. The challenge faced by Evolve Power is to automate demand response between utilities & consumers to make this exchange of information seamless and near error-free. Using blockchain to disperse, manage & incentivize this utility-consumer network is critical to the core of DRMS.

Solution

- Conception and Execution of Utility/Consumer Dashboard UI and UX
- Ethereum-based Tobalaba by Energy Web Foundation
- Easy Accessibility for Utility and Consumers
- Real-time Monitoring, Verification, Execution and Incentivization

Features

- Reliability via self-enforcing digital Smart Contracts automated demand response process for better predictable outcomes
- Visibility via unparalleled visibility and control for utilities/grid operators over distribution edge of the grid
- Security through a decentralized infrastructure improves cybersecurity & data integrity among all parties
- Expertise in demonstrated track record of helping utilities and grid operators furthers clean energy initiatives

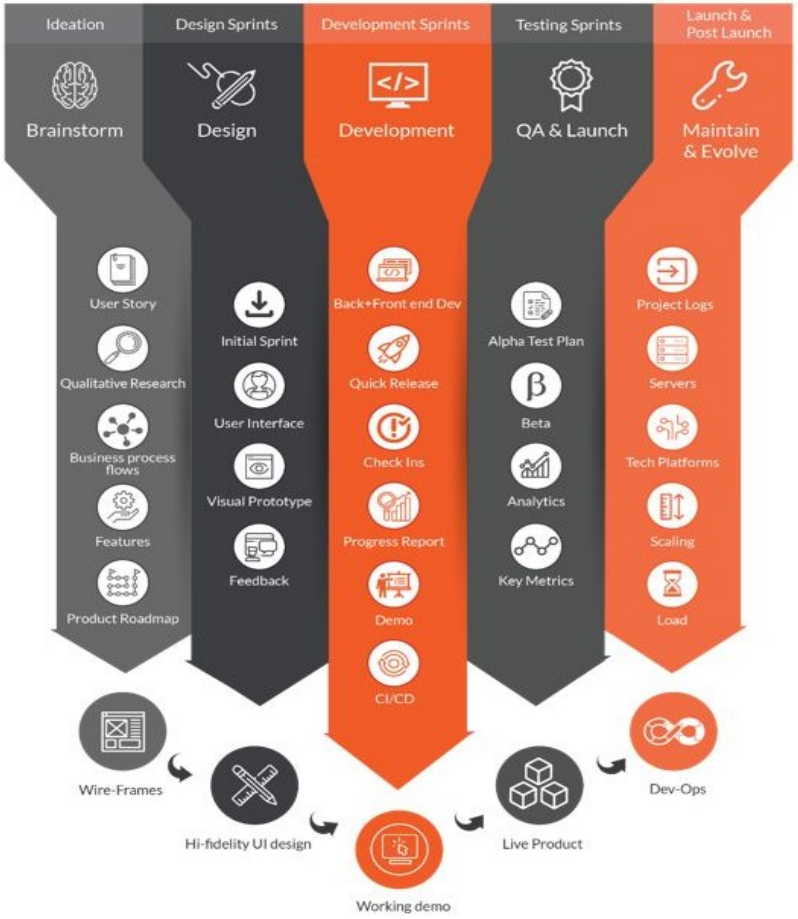
Product Development Manifesto


Stay
Tech-Stack
Agnostic
Lumen
MySQL
AJAX
Native Java
Kotlin
Android
Redis
InVision
Balsamiq
Material Design
Swift
Objective-C
Laravel
Creative Cloud
Photoshop
Illustrator
Symfony
Bootstrap
Teamwork
Jira
Stack
IBM Watson
TensorFlow
CSS
Ionic
Meteor
React Native
HTML
JavaScript
Angular
Ethereum
Hyperledger
React
Backbone
Vue.Js
Tobalaba Blockchain
Solidity
Ruby on Rails
PHP
.Net
Node.js
Fython
GitLab
GitBucket
Az
Digital Ocean
GitHub
Travis
Jenkins

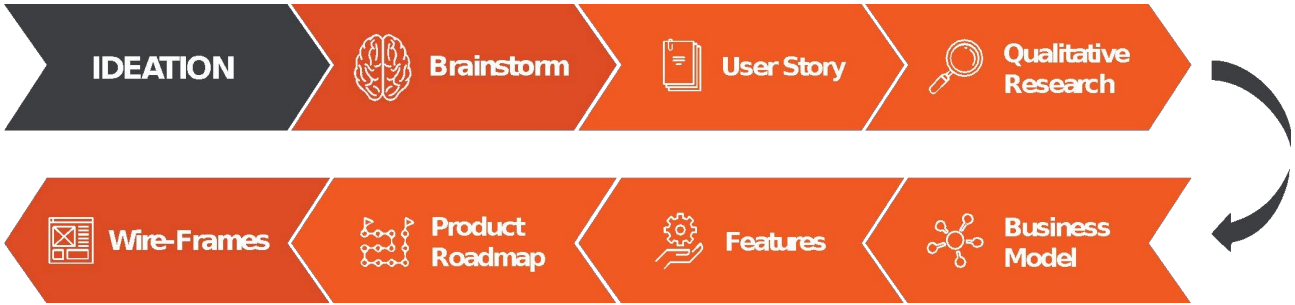
User and Problem Centric Development Approach

An enterprise should aim to rollout fast product iterations while getting feedback and input from the end users at various intervals. We recommend breaking down development process into multiple phases working in parallel:

- Ideation Sprints (**1-2 weeks**)
- Design Sprints (**1 week**)
- Development Sprints (**2 weeks**)
- Testing Sprints (**1 week**)
- Post Launch DevOps (as needed)



Ideation



Enterprise should formulate a **customer-focused product strategy**. This phase includes prioritization of use cases that bring in the most value for the product; creating storyboards, user flows and mockups of the product before they get into development.

Key Outputs:

- Prioritized user story document
- Business / User flows
- Wireframes
- Product Roadmap

Key inputs:

- Customer interviews
- Signoff on business flows and Product Roadmap.

Design Sprints

Create a Proof of Concepts through **highly-collaborative** design sprints. The design sprints enable you to hash out the user interface and user experience of the product. The deliverable of this phase is a **Visual Prototype** of the product that can be shown to prospective customers for feedback.

Deliverable:

- Hi-Fidelity Designs
- Visual Prototype

Client's input:

- Customer validation on designs

Development Sprints

A **development sprint** spans a duration of **2 weeks.** Overall product roadmap is covered through multiple sprints in an Agile development manner.

The outcome of this phase is a part of the overall application in a working condition

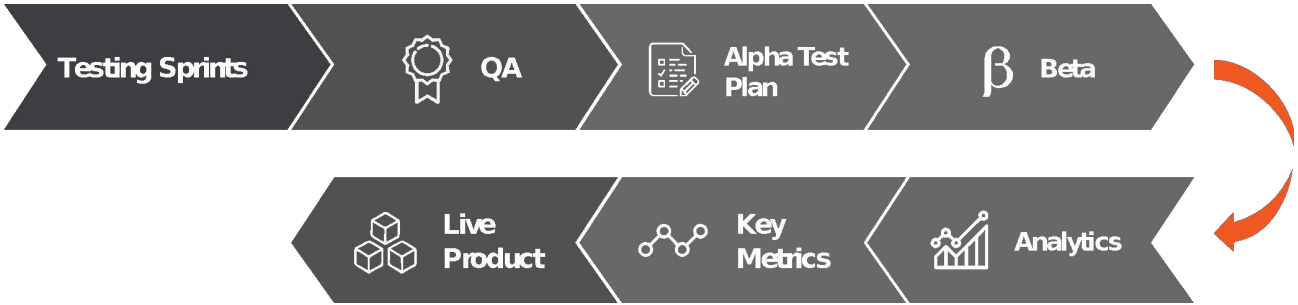
Deliverable:

- Demo-able features
- Parts of overall application in working condition

Key Rituals:

- Sprint planning meeting
- Backlog refinement meeting
- Sign-off on Sprint Deliverables

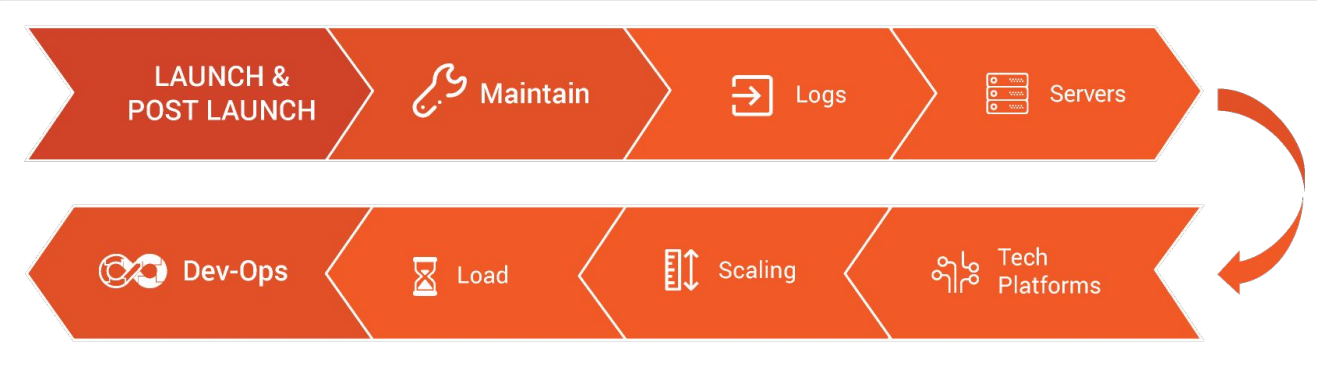
Testing Sprints



While there is continuous testing and QA that should be embedded in Development sprints a detailed QA and bug fixing phase that should be done after the development phase. We recommend use **Continuous Testing** and **Integration Testing** to ensure the end product does not pose any functional or integration issues. We also recommend using **Functional, UI, Regression, Load and Penetration** testing rounds on products as per their needs.

Once the in-house testing and **User Acceptance Testing** is completed , the product can be pushed to the **live server**

Launch & Post Launch



Make sure required server and production configurations are in place for a smooth **project launch.** Once the project is live, the devops team maintains the live server ensuring **load management** and **smooth server functioning**. This will ensure the product is ready to scale from **performance, stability and security standpoint.**

CASE STUDIES

Mother’s Day Initiative by Arla Foods

COUNSELYTICS

Smart NLP Engine that Tackles Entire Contractual Genome

Counselytics enable enterprises to process and manage large amount of legal content in fast, efficient and cost effective ways. Most of the data in a contract is in unstructured form, rendering it unproductive for further data analysis. Counselytics goes through legal documents and analyzes up to 120 key legal and business terms for data analysis. This gives you the ability to customize term extraction.

VF Role

Ideation, Mockups, Design, Development, NLP Engine Development

Challenge

Counselytics needed to build a software where documents from every aspect of business would be managed with ease.This would include invoices, employee contracts, business transactions and correspondence, legal and financial contracts. Around 90% of the data is in an unstructured form in contemporary contracts. Counselytics needed to change that by leveraging Natural Language Processing (NLP), sorting data and analyzing it.

Solution

Creatively brainstorming around business flows, we identified several area to target. We managed to classify all of the data into the following categories: Suppliers (orders, invoices, materials & returns), Employees Data (recruiting, retention/advancement, retirement), Customer Data (correspondence, history/transactions, install base/revenue), Contracts (legal & templates, terms and entitlement, renewals and revenue), Financial Data (audit, compliance regulatory, fraud & collections).

Features

File Transmission

Users can upload documents into the Counselytics applications through the easy to use interface which provides options for one or multiple document upload.

Data Security

The system ensures the highest security and integrity of customer data, and protects against security threats or data breaches.

Data Processing and NLP

Counselytics is cognitive augmentation. Its proprietary algorithms can extract and analyze up to 120 key legal and business terms related to contractual & legal material impact. Additionally, it provides users with the ability to create their own business terms.

Reports, Dashboard & Integration

Counselytics provides a snapshot view across an entire document and contract repository. It can perform search or apply filters to identify expiring contracts or commonly used clauses. It can also identify most and least favorable terms across a contract type, based on user preference.

Highlights

Team

1 UI/UX Designer
1 Backend Engineer
1 Project Manager
1 QA Engineer
1 ML Engineer

Duration

MVP - 8 Months
Evolution - 1 Year

Tech Stack

Fluxx
Angular JS
Node JS
ROR
PostgreSQL
NLP Techniques

Outcome


conga
Contact 303.465.1616
COUNSELYTICS
POWERED BY conga
CONGA ACQUIRES COUNSELYTICS


Q&A

Contact


48 Wall Street 5th Floor, Suite #9
New York, NY, 10005
waleed.nasir@virtualforce.io
virtual force
www.virtualforce.io
+1 203 606 2501

What Is Data (Really)?

RiseNY
12 February 2019

Data is Transactional

Data is Property

Data is NOT Metadata

And Data Science . . .?

“A good rule of thumb to keep in mind is that anything that calls itself a science probably isn’t.”

– John Searle

Professor Emeritus of the Philosophy of Mind and Language
UC Berkeley

Peter Wegner and Peter Denning
(separately) identified three paradigms
which define Computer 'Science':

Theorem and Proof
Abstraction (Modeling)
Design

[only the first is strictly speaking Science]

'Data Science' is Practiced in All Three Paradigms

And So Data Science Is . . .?

At Least Six Different Things:

– AI (but not General AI)

– Neural Networks (ANNs, GANs and others)

– Classification Engines

– Machine Learning

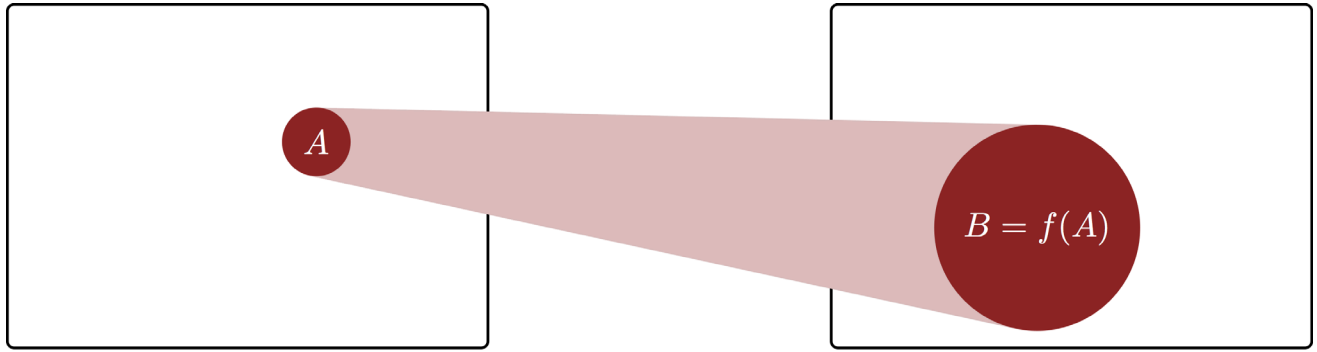
– Deep Learning
(recently re-branded Differentiable Programming)

– Plain Old Statistics

“When you’re fundraising, it’s AI. When you’re hiring, it’s ML. When you’re implementing, it’s logistic regression.”

Statistical Functions Are Metadata Operations

At the Level of the Data
They Are NOT Transactional


A
B = f(A)
X
f : X → Y
Y

Data is Transactional

Data is Property

Data is NOT Metadata

Data is Transactional

Data is Transactional

The whole process of applying this complex geometric transformation to the input data can be visualized in 3D by imagining a person trying to uncrumple a paper ball: the crumpled paper ball is the manifold of the input data that the model starts with. Each movement operated by the person on the paper ball is similar to a simple geometric transformation operated by one layer. The full uncrumpling gesture sequence is the complex transformation of the entire model. Deep learning models are mathematical machines for uncrumpling complicated manifolds of high-dimensional data.

—

François Chollet

The 'Natural' Primitive Data Model is
The Transaction

(NOT the Document)

http, the WWW and the Semantic Web
are all inadequate because
they are based on a Document Model

Documents Do Not Reify

(the converse of the Map/Territory Problem)

There is no Abstract Distance
Between a Transaction and the
Record of That Transaction

Transactions Manifest As Both Nouns and Verbs

A Transaction-as-a-Noun
is the Set of Instructions for the
Execution of that Transaction

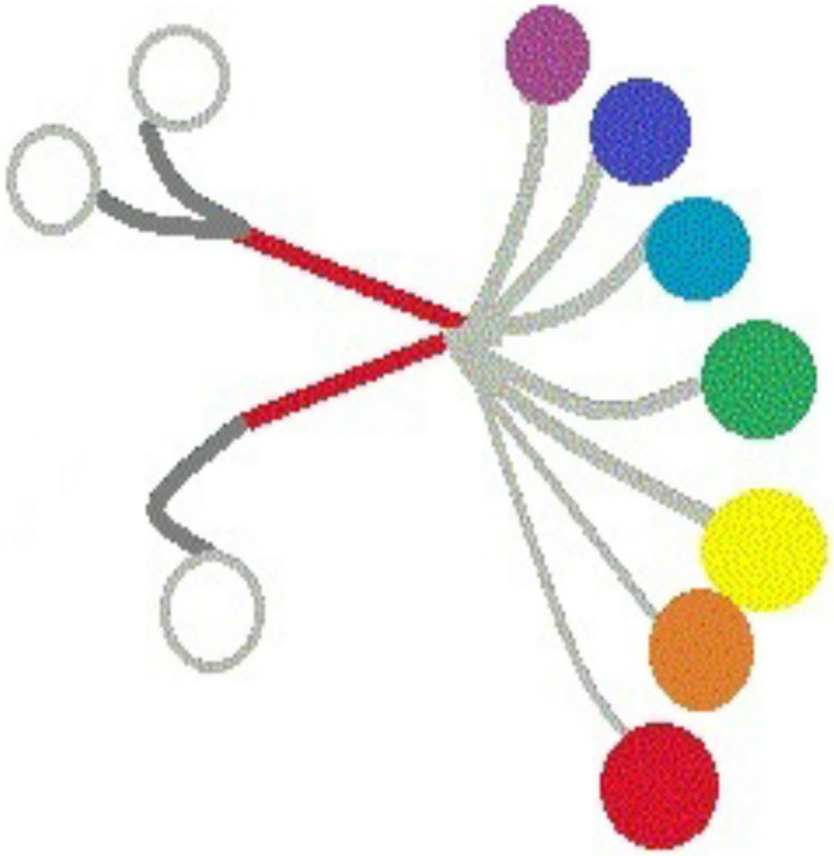
A Transaction-as-a-Verb
Is a Concrete Instance of the
Execution of that Transaction

Transactions both as Nouns and Verbs
Are Composable From Simpler Transactions

Transactions are Composed by Executing
Transactions-as-Verbs
Against Successive Versions of a
Transaction-as-a-Noun

(and at some point executing an instance of the
resulting complex transaction)

This Transaction-Based Data Model
is expressed most naturally in a
Graph Data Structure



Data is Property

A unit of data is a unit of value, in which rights inhere for the benefit of data producers, investors, validators and other processors

As property, each and every data record bundles the rights of the beneficial owners of that property

The assertion and the definition of these rights can and should be encoded in each unit of the data, and then preserved through transactional data transformations as an immutable component of the provenance of every data record

The Data Record is in fact Property: it can be spent or sold ('alienated' in the legal language of property), licensed, mortgaged ('hypothecated'), and otherwise transacted upon by its owner in contract with a consenting counterparty, and the use of it can be denied to any other party (right of exclusion)—all without requiring the validation of that transaction by anyone or any mechanism outside the principal parties to that transaction.

This Absence of Validation Does Not Mean That There Are Not Rules of Data Governance

The Set of Transaction Instructions
Which Make Up the Transaction-as-a-Noun
Are the Specific Rules of Data Governance
For the Execution of that Transaction-as-a-Verb

The Validation that a Transaction-as-a-Verb
Has Been Executed In Accordance With Those Rules
Is the Consent of the Principal Parties
To Accept the Outcomes
Of the Execution of that Transaction

Because the Transaction is the
Sole First-Class Citizen of the Data Model,
The Broadest Scope Bounding Any Data Entity
Is a Single Particular Transaction

The Data Records Output From the
Execution of a Transaction
Embed the Consent of the Beneficial Owners
Of Property Rights
To the Existence Of and
To the Potential For Future Uses
Of Those Data Records/That Property

Every Further Use of Those Output Records
Must Be Consented To By
The Beneficial Owners/Rights Holders
Of That Property

Consider Chollet's Crumpled Ball of Paper:

There Is Only One Set Of 'Uncrumplings'
Which Can Be Successfully Applied
Because They Undo The 'Crumpling' Transactions
Which Have In Fact Been Previously Applied

A Sequence of Transactions-as-Nouns Executed As Transactions-as-Verbs In 'Either Direction'

Data Is Not Metadata

Data Is Transactable
Metadata Is Not

Metadata Is Viewed From
Outside the Transactional Scope of Data

Data Is Viewed From Inside Transactions
By Principal Parties
('Privity of Contract')

Metadata is a Third Party Classification System

Some Metadata Can Be Reified As Data

Data Must Be Forked From Existing Data

(Bringing Along the Transactional History,
The Kinetics of Previously Executed Functions
And the Rights of Principal Parties
To those Previous Transactions
So That They Now Enjoy *Executable* Rights of
Beneficial Ownership in the Data Records
Which Were Output From Those Transactions)

That Fork is a Transaction
Applied Against a Transaction-as-a-Noun,
Executed as a Transaction-as-a-Verb,
And Resulting in the Output of Modified
Transactions-as-Nouns

Such Transactions Can Be Executed 'Anywhere',
Subject to Securing the Rights to the Use of
That Data for the Purposes of that Transaction

The Execution of that Transaction Will Proceed
By Embedded Rules of Governance
Which Will Ensure the Consent of Rights Holders
That Authorizes the Output Data Records
Which are Produced

Following Embedded Rules of Data Governance
Is Replaying the Kinetics of Earlier Transactions

('uncrumpling the ball of paper')

Metadata Does None Of This

(It Is Not Transactional, But Ontological)

The Semantic Web Fails Because Its Document Data Model Lacks the Transactional Nature Necessary to Enable the Kinetics of Functions and the Rights to Use and Transform Data Which the Elaboration of Semantics Requires

Metadata Makes Assertions About Data
To Which It Has No Privity

Metadata Has No Functional Capacity
To Transform Data
Because the Scope of its Data Model
Is Documentary, Not Transactional

Every Reaction of Data With Data is Transactional

Reactions with Metadata are Solely Observational

(Data Is Not Transformed By Metadata)

A Necessary Capability in the IoT,
Edge Computing, Fog Computing,
Decentralized World

(there is no perimeter, only rules of data governance enforced upon execution by the data itself)

Blockchain-Based Distributed Shared Computing

Chong Li

NAKAMOTO &TURING LABS

Who We Are

Nakamoto & Turing Labs (N&T Labs) is a NYC-based research lab

N&T is engaged in scientific and engineering research in the fields of blockchain and AI technologies

Ongoing Projects:

1. CanonChain: public blockchain for intelligent IoT
2. Pekka: blockchain-based shared-computing platform

Why Distributed Shared-Computing?

NAKAMOTO &TURING LABS

- For high performance computing jobs, cloud is the most popular choice

BUT…

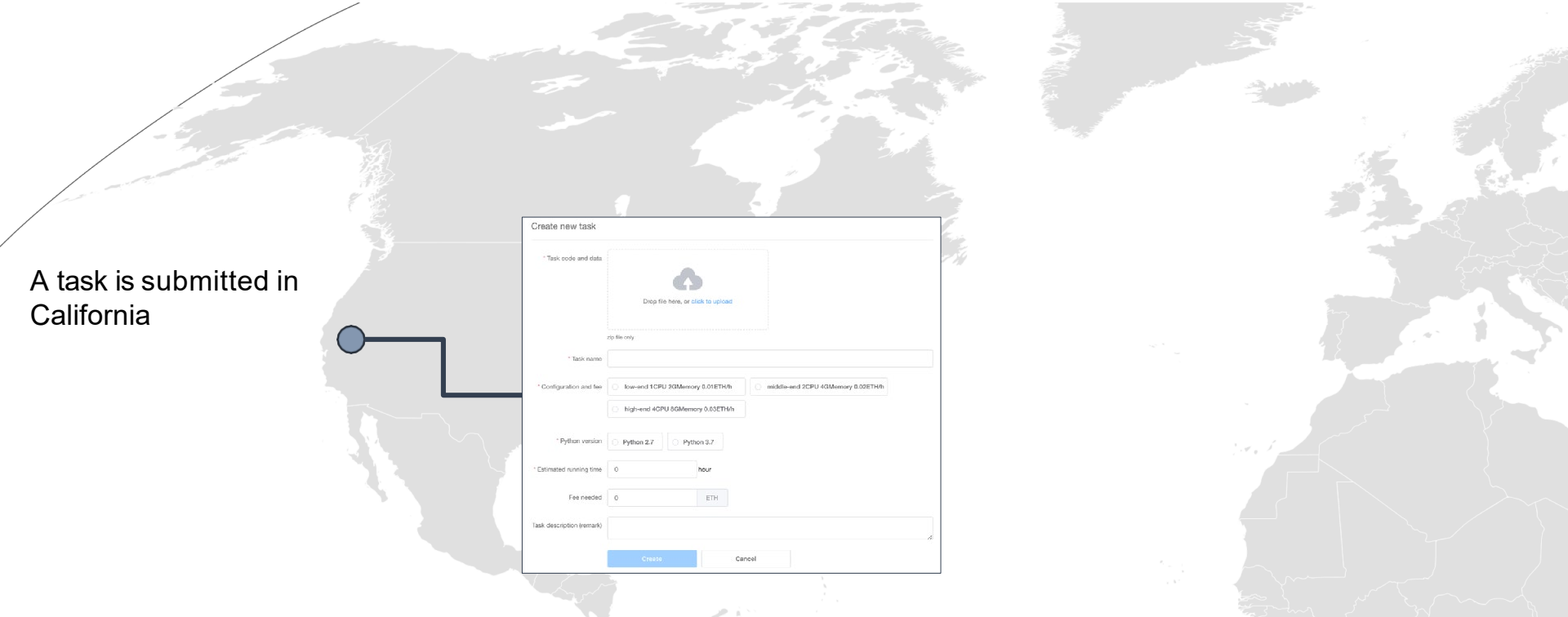
- Using cloud service means all data, even sensitive information, has to be shared with the cloud service provider
- Pay-as-you-go cloud service always has an overall price tag that ends up being higher than expected

Shared-Computing Platform

NAKAMOTO &TURING LABS

At every single moment, tens of millions of small computing devices such as laptops and desktops are idle

A task is submitted in California

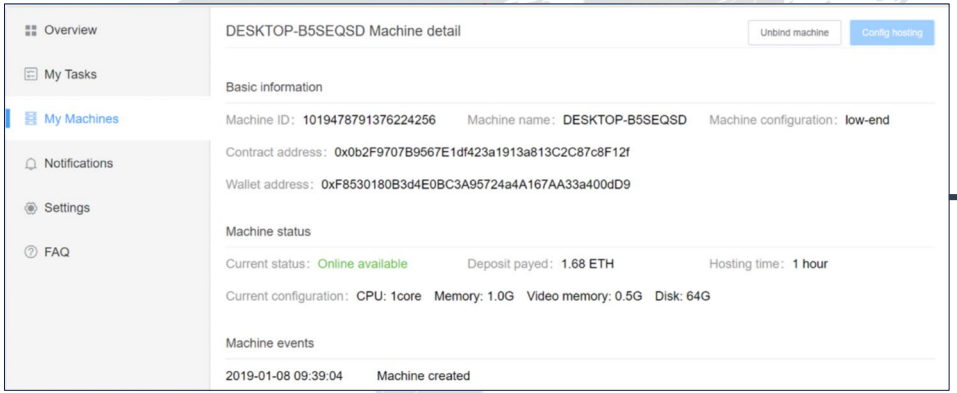


Shared-Computing Platform

NAKAMOTO &TURING LABS

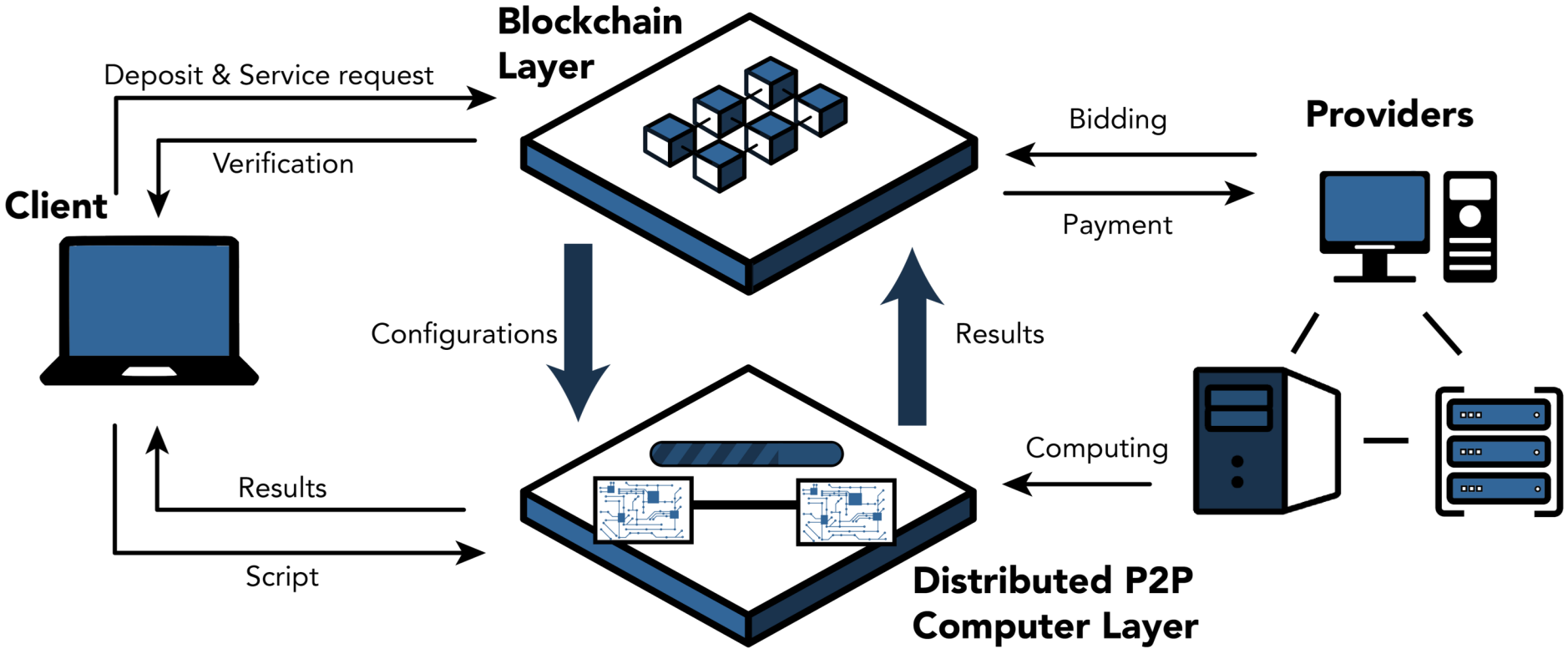
At every single moment, tens of millions of small computing devices such as laptops and desktops are idle

An idle machine in Paris computes the task



System Architecture

NAKAMOTO &TURING LABS



Blockchain Layer

Blockchain Layer Functionalities

NAKAMOTO &TURING LABS

- Payment: fast global p2p transactions at almost no cost
- Marketplace : fair service price via real-time bidding system
- Privacy: personal data and scripts protection
- Verification: guaranteed correctness of computing results

Blockchain Layer Functionalities: Payment

NAKAMOTO &TURING LABS

- Support cross border transactions
- Require low or no transaction fees
- Need novel payment mechanisms for abnormal service termination such as machine power cutoff and internet connection loss

Blockchain Layer Functionalities: Marketplace

NAKAMOTO &TURING LABS

- E-auction (eBay, Yahoo) is popular since its convenient and efficient.
- The main roles during E-auction include bidders, auctioneers, and the centralized third-party.
- Weakness:
  - The charge fees for the centralized third-party increases the transaction cost.
  - Personal data and transaction records stored in database might cause privacy leakage.

Blockchain Layer Functionalities: Privacy

NAKAMOTO &TURING LABS

- Facebook, the largest online social-network, collected 300 petabytes of personal data since its inception – a hundred times the amount the Library of Congress has collected in over 200 years
- Individuals have little or no control over the data that is stored about them and how it is used.
- In distributed shared computing platform, how to guarantee privacy for both client and provider?
  - Client: Multi-sig smart contract
  - Provider: record of docker access

Blockchain Layer Functionalities: Verification

NAKAMOTO &TURING LABS

- How to verify the result efficiently without re-executing the task by the client?
- Providers do not necessarily have strong incentives to ensure correctness.
- Complex and largescale providers (cloud servers) are unlikely to guarantee that the execution is always correct due to mis-configurations, randomness in hardware and more.

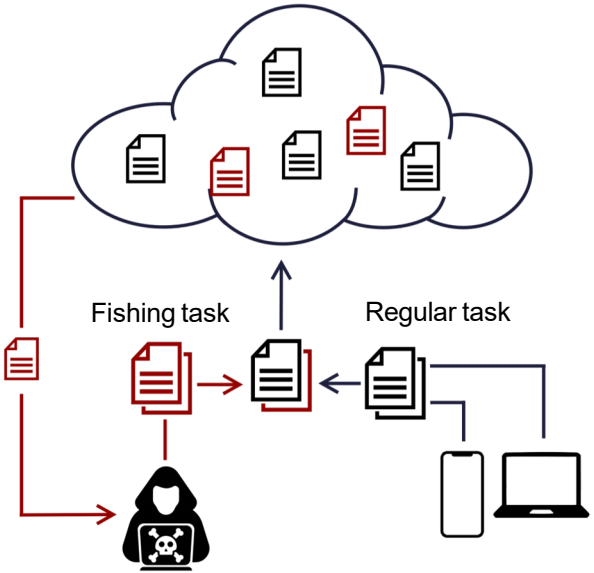
Blockchain Layer Functionalities: Verification

NAKAMOTO &TURING LABS

- Conventional approach: Verifiable Computing
  - Enabling a computer to offload the computation of some function, to other perhaps untrusted clients, while maintaining verifiable results
  - Well-established theory using “abstract algebra”, but only near practical *
- Blockchain-based approach: TrueBit, EntrapNet
- What is EntrapNet?

* M. Walfish and A. J. Blumberg, “Verifying Computations Without Reexecuting Them”, Communications of the ACM, 2015

EntrapNet



EntrapNet

Idea

- Borrows the idea from the practice of entrapment in criminal law to reduce the possibility of receiving incorrect computing results from trustless service providers
- Incentive to volunteer clients who wish to submit an fishing job. The outcome of fishing jobs are known in prior
- The deposit of a subverter, if caught, will be forfeited

Analysis

- Performance tradeoff:
  - More fishing jobs -> more trustable network
  - More fishing jobs -> more likely waste of network resource
- The real-time optimal rate of submitting fishing jobs to the network*?

* C. Li, L. Zhang and S. Yang, “EntrapNet: a Blockchain-Based Verification Protocol for Trustless Computing” to appear

Distributed Computing Layer

Motivation & Design Goal

NAKAMOTO &TURING LABS

- On the shared-computing platform, a large ML task needs to be executed by one or more geo-distributed computing devices
- Use data parallelism
- Need to develop a geo-distributed ML system that
  - Minimizes communication over WANs; and
  - Is applicable to a wide variety of ML algorithms

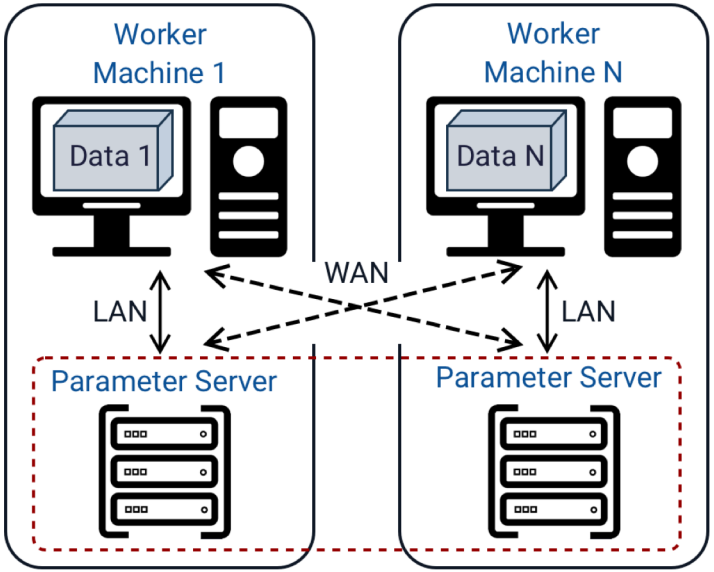
Geo-distributed Data Centers

NAKAMOTO &TURING LABS

* Figure source: DataCenter Knowledge

Atomia

Parameter Server Architecture



Parameter server (PS) architecture

- Each parameter server keeps a shard of the global model parameters
- Each worker machine communicates with the parameter servers to READ and UPDATE the corresponding parameters

Why this architecture?

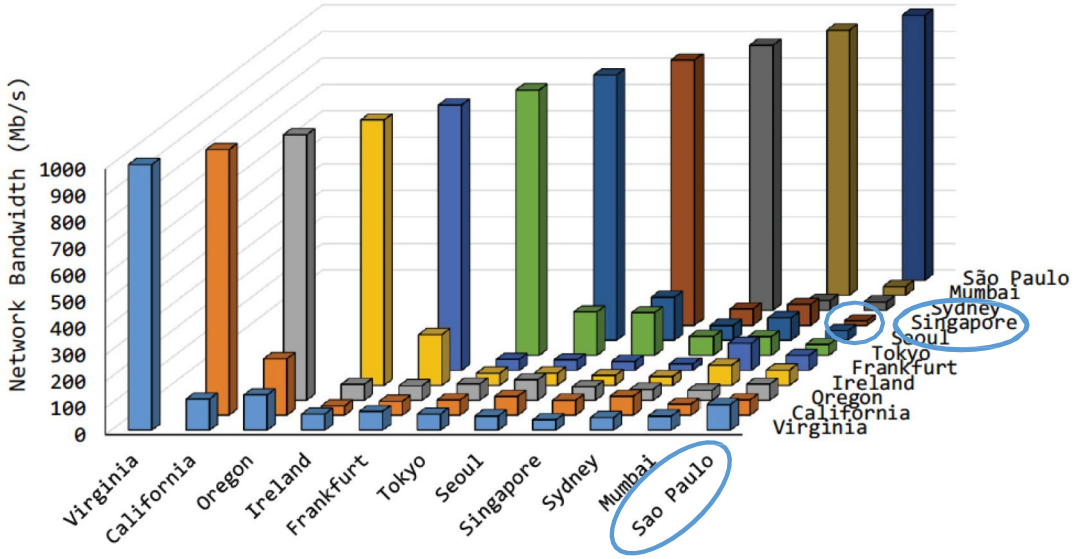
- ML programmers can view all model parameters as a global shared memory, and leave the parameter servers to handle the synchronization
- However, when ML algorithms iteratively refine the ML model until it converges to fit the data, WAN limits the performance.

* Source of this section: K. Hsieh, at el, “Gaia: Geo-distributed machine learning approaching LAN speeds”, 2017

WAN Bandwidth Constraints

NAKAMOTO &TURING LABS

- WAN BW is 15X slower than LAN on average and 60X slower in the worst case (Singapore <-> Sao Paulo)

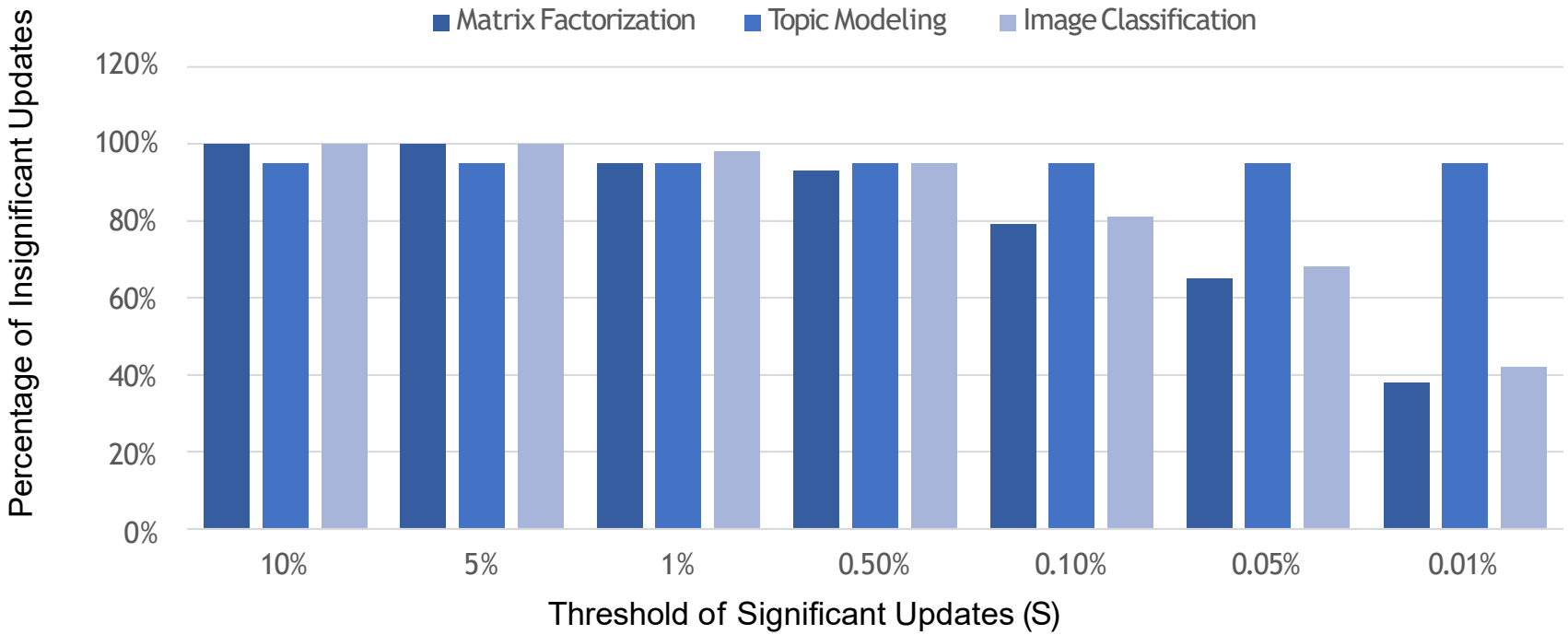


WAN Bandwidth Measurements

Key Observation

NAKAMOTO &TURING LABS

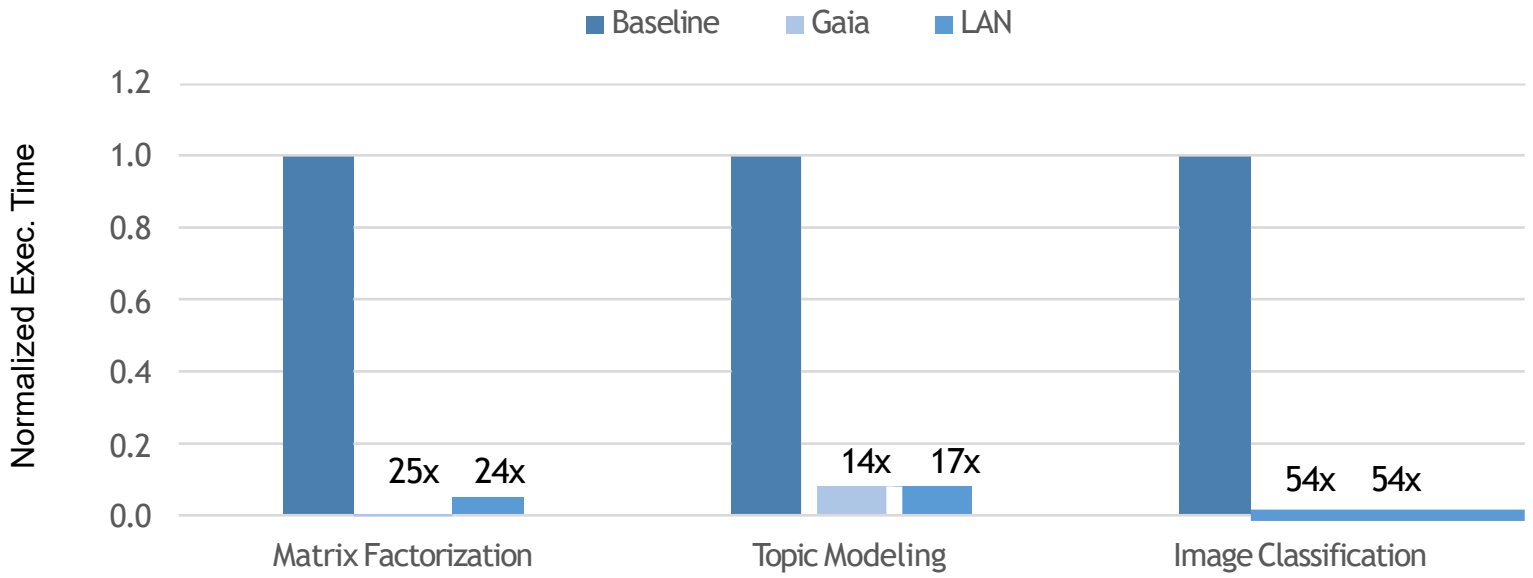
- Most of the updates on the ML model state are only very slightly



Solution

NAKAMOTO &TURING LABS

- Communicate over WANs only significant updates



Normalized execution time until convergence with the WAN bandwidth between Singapore ad Sao Paulo

Key Takeaways

NAKAMOTO &TURING LABS

- Blockchain-based Distributed Shared-Computing resolves security and cost issues of cloud service
- The proposed architecture consists of **blockchain** and **distributed computing** layers
- Blockchain layer provides solutions to payment, marketplace, privacy and verification
- Distributed computing layer handles the WAN constraint of geo-distributed computing network.


NAKAMOTO
&TURING
LABS