

Good Morning!

I want to thank you all for getting up early after a night of guns and beer to attend this talk. I'll try to make it worth your while. Before I get started, I want to thank the fantastic volunteers that work so hard to make Southeast Linuxfest such an outstanding event!

This is the first time I've given this specific talk so I'm not completely sure about timing. My goal is to present the slides in about 40 minutes and then give us 10 minutes for Q&A

Presenter – Brad Whitehead

- 47 Years in IT Consulting and Operations
- Former Partner with Accenture
- Speciality in "Fire Fighting" and Large Scale Biometric Systems
 - Trusted Traveler, DHS IDENT Biometric System, TWIC Cards
- Presently Chief Scientist of Formularity
- Role in Addhaar "Fire Fighter/Chief Architect"

So, who am I?

My name is Brad Whitehead. I've been in professional IT consulting for 47 years, the majority of that time as a Partner with Accenture. My specialty is "fire fighting". I get called in when large scale systems don't perform as expected and customers are unhappy. When I'm not fire fighting, my actual specialty is large biometric systems. I have been Accenture's chief architect for several US Government biometric systems, including TSA Pre-Check.

- Presently, I'm Chief Scientist for Formularity, an electronic enrollment company.
- The following talk is based on my role as Accenture's fire fighter during the Republic of India's national biometric identification program rollout.



Incredible India - Circa 2005



3

- Second largest country, based on population
- Third "richest" country based on governmental per-capita spending
- Largest democratic government in the world
- 80% of the population are "BPL" (Below Poverty Line)
- 70% of food subsidies are stolen through fraud
- 90% of retirement benefits are "misdirected"
- Large scale government programs
 - Mahatma Gandhi National Rural Employment Guarantee Act (MG NREGA)
 - Public Distribution System (PDS) food and cooking/heating gas subsidies

Let's start this story in 2005.

- India advertises itself as "Incredible India". And it's a good description. The Republic of India is the largest democratic government in the world. It trails only behind China in population. It's the third richest country based on per-capita government spending.
- But it has some problems. 80% of the population falls what they call "Below the Poverty Line" (the BPL). This population qualifies for Government aid programs. And India has a number of large welfare programs, including guaranteed employment and reduced cost for basic foodstuffs and heating/cooking propane.
- Unfortunately, there is widespread corruption and graft in the welfare programs. In 2005, 70% of food and gas subsidies were stolen. 90% of Government retirement benefits were "misdirected"

Incredible India – Circa 2005



4

- Vast majority of residents have no documentation (birth certificates, proof of address, drivers license, travel documents, etc.)
- Strict government-enforced "Know Your Customer/Client" (KYC+) requirements for cellular communications, banking, heating fuels
- Several different "national identification card" programs but all are inadequate and fraught with fraud
- Poor national data distribution infrastructure (3G emerging, limited fiber)
- Poor power distribution infrastructure
- No Tier 4 data centers

Let's start off by understand the challenges of the people Below the Poverty Line. Because of the political situation in the geographic region, the Government has strong "Know Your Customer" or "KYC" requirements. Like the American Patriot Act, these regulations are in place to combat terrorist funding and money laundering. To meet these KYC requirements, a person has to be able to present Government documentation with a valid postal address. Unfortunately, a large number of the population does not have the basic documents and can't open a bank account. Accordingly, most of India is a cash economy and where there's cash without records, there is thief. Other challenges for the country as a whole is that in 2005, cellphones were primarily 2G with 3G just rolling out; there was a poor overall power distribution infrastructure, and this meant that at that time, there were no Tier 4 data centers

Rampant Corruption



5

- Mahatma Gandhi NREGA "fees"
- PDS ration books for sale
- No subsidized foodstuffs remaining
- Loan sharking
- Money transfer schemes

Some examples of the corruption that was widespread in the 2005 time-frame:

- While there was guaranteed employment through the Mahatma Gandhi NREGA program, the worker usually didn't have a bank account so they were paid in cash – AFTER the supervisor took his or her 'cut'. So 5 rupee became 2 rupee
- 2) ration books were for sale. When I started the project, I lived in a very nice gated community with a doctor on one side and a barrister on the other. First thing my driver said was "I'll get you a coupon ration book for the cooking gas so you won't have to pay full price!"
- 3) Merchants were only required to sell a percentage of their overall food, like rice or wheat, at the ration discount price. So, they would sell their discounted food to their cronies and then turn away legitimate ration book holders – "Sorry, All sold out"
- 4) India has a large migratory farm labor workforce. These are men and women who start in the south and move up the country with the growing season. They need to send their wages back to their families. Without bank accounts there are any number of money transfer schemes where you pay 20% of your money as a transfer fee and then hope the transfer is legitimate and the money makes it to your family.

Imagining India - Circa 2006



6

- Some of the World's Largest IT Firms are In India —"Indian Pure Plays"
 - Infosys
 - TCS
 - Wipro
- InfoSys CEO Nandan Nilekani Convenes Think Tank "Improving India Through Technology"
- Recruited wealthy, successful Indian ex-pats from Silicon Valley

 Looking to "repay their social debt" to India
 - -DotCom/Start-up founders (Tibco, WebMD, Intel, etc.)
 - -Working for "1 rupee per year"
 - -Start-up, Web 2.0 mentality (squared!)

India has some of the largest, most skilled technology organizations in the world. I'm not talking about the Indian branches of Western corporations. I'm not talking about the Indian division of Microsoft, Oracle, or Google. I mean what we call the Indian "Pure Plays", like InfoSys, TCS, and Wipro.

- In 2006, the CEO of InfoSys, Nandan Nilekani decided that Indian technology had a social debt to the country. Nandan is a Fabian Socialist. Like the original Fabian Society, he convened a think tank of Indian technology leaders. They also invited Silicon Valley Indians that had made their fortunes here and were now looking to give back to India. These were the entrepreneurs that started or lead such companies as Tibco, WebMD, Intel, Google, etc.
- These expats brought a strong startup, Web 2.0 mentality with them.

What Did They Envision?



7

- Government-Recognized, Immutable Identification for Everybody —Biometrics
- Government Acceptance of Identity for "KYC" Requirements
 - -Open Bank Accounts
 - -Transfer Funds
 - -Purchase Cellphones
- Incorporate This Identity into Subsidy Process
 - -Direct deposit of salaries/payments
 - -Proof of foodstuff and gas distribution
- This identification would be just that identify an individual –It does not signify citizenship, or any entitlements

What they came up with was the idea of a universal, immutable identity for everybody in India that would be recognized by the Government and that would satisfy Government KYC requirements.

- This immutable identification could then be used by the BPL people to open bank accounts, buy cellphones, and transfer funds. It could be used as proof of foodstuff and gas ration purchases.
- One of the biggest keys and differentiators was that the identity would not signify ANYTHING. It wouldn't say you were a citizen and it wouldn't mean that you were entitled to welfare. It would just uniquely identify a person.
- The best unique immutable identifier is a person's biometrics. Their fingerprints, retinas, irises, hand geometry, face, etc.
- Any number of unique biometric measures!

How?



8

- Create Government Agency for Identification
 Unique Identification Authority of India (UIDAI)
- Collect Biometric Characteristics from All Residents of India
- Establish a Nationwide "Identity Verification" Service
- Change "KYC" Requirements in Banking and Communications Acts
- Incorporate Biometric Verification into Subsidy Processes
- Strong Advertising and Commitment By Government Agency
 Aadhaar "New Dawn"

So, how did they envision this working?

- 1) First, create a Government-level agency responsible specifically for identity and the administration of this identity concept. This would later become the "Unique Identification Authority of India or "UIDAI"
- 2) Collect, process, and store the biometric identities of all applicants, with the goal of enrolling all residents in the country [remember, citizenship doesn't matter].
- 3) Start a national identity verification service.
- 4) Change the KYC requirements in the Banking Act and other laws to accept this unique biometric identification.
- 5) Incorporate the biometrics into the ration programs
- 6) And sixth, most importantly, shepherd a strong advertising campaign about the new program and make sure all Government agencies, at all levels, were committed to incorporating the program into their workflows.
- The UIDAI chose the Hindi word for "New Dawn" or Aadhaar as the name for the program and the advertising.



Here'a a commecial Accenture made in support of Aadhaar. It sums up the current problems and the future results.



(Use laser pointer)

- Aadhaar starts with collecting biometrics and verifying they are not duplicates. In the end, the decision was made to collect all 10 fingerprints, both irises, and the face. Irises are as good as retinas but significantly easier to read. It doesn't require the user to squint into a laser and the iris is like a circular barcode,
- 2) After successful enrollment, the biometrics and demographic information is stored in a Central Database,
- Aadhaar also includes a national authentication service that matches the individual with their biometrics to confirm they are who they claim to be
- and finally, we have the Government and private sector corporations accepting this authenticated ID.

How Does This Help?



11

- Instantly satisfy "Know Your Customer" laws
 Aadhaar holders can open bank accounts
 - -Purchase cellphones
 - -Salary direct deposit
 - -Money transfers
- Proper allocation of food and gas subsidies
- Eliminate Identity Thief and Identity Fraud –School and exams

This biometric identification system solves a number of the problems I've highlighted. With an Aadhaar ID, residents can immediately open bank accounts and have their Government aid and salaries directly deposited into their accounts. No shrinkage!

- They can buy cellphones and use these cellphones like we use them here in the US for bank account management, payments, and funds transfer.
- Merchants can't sell subsidized food and gas without a corresponding Aadhaar ID.
- Even in schools students can no longer play games with college admission and testing systems by hiring somebody to take their place.





12

- Issue a unique identification number (12 digits) to every one of the 1.2 billion residents of India – Aadhaar Number
- Capture and associate an individual's biometrics and demographics with their Aadhaar number

 Ensure one-time enrollment (de-duplication) based on biometrics (10 fingers, 2 irises, facial image)

- Key Points
 - -Aadhaar number does not signify anything citizen, entitled, right-to-vote, etc.
 - -Patterned somewhat after US Social Security Number, with the addition of biometrics to prevent identity thief and fraud

The Aadhaar identification given to each successful enrollee is a 12digit, random number. As I've said, this number has no encoded meaning. It doesn't signify the enrollment station, the state of residence, race, citizenship, BPL status. It's a pure and simple 12 digit random number

First, Unique Identification Authority of India



13

- Created October 2009
- Cabinet-level Organization
- Charter: To Develop and Promote the Aadhaar Program
 - Enrollment
 - Database
 - Authentication
 - Spearhead Governmental and Private Sector Acceptance of Aadhaar

- So, the first step was to create the Cabinet-level agency, the Unique Identification Authority of India. It's charter was to enact the Aadhaar system, spearhead Government and private sector acceptance, and maintain a strong public advertising campaign.
- I spent some time at an enrollment center, observing. There was a long line of people waiting for their turn to enroll. I asked one gentleman who had been in line for a long time, "Why are you here?" This fellow told me "My wife told me to stand here. I don't know what I get with Aadhaar, but my wife told me it was wonderful" Talk about blind trust in advertising!

Nandan Nilekani – Imagining India



14



"The largest biometric database is 120 million in the US. We are doing 10X, that is 1.2 billion. That itself is unprecedented.

We are doing online authentication. Nobody is doing it online. We are doing cell phone authentication. Nobody has done that either. It's doable as technology is there, but nobody has put it all together.

It's a huge task and de-duplication is a big challenge"

"Q&A with Nandan Nilekani" Wordpress October '09

Here's Nandan Nilekani's statement about how big and difficult Aadhaar is. He was off by a factor of 2 when it came to our US biometrics database. At that point, we had 300 million individuals on file, not 120 million. Still that does pale in comparison to 1.2 billion! The UID will enable the common man to get access to a host of government services quickly and more efficiently





"This mammoth and unprecedented exercise will serve as a great enabler to improve targeting and delivery of major government welfare programmes and public services, especially to those who are poor and marginalised."

Address by the Hon'ble President of India, Shrimati Pratibha DeviSingh Patil, to Parliament on 22nd February, 2010.

And here is the President of India, selling the program to Parliament..



Here are the steps in the enrollment process. BTW, that's not a typo in the slide title. Indians spell enrollment with only one "L".

- We record both the biometric and demographic information from the resident. This information is digitally signed by the enrollment center
- 2) At the Central Database, the digital signature is checked
- 3) Next, we check the incoming biometrics against all the existing biometrics to confirm there are no duplicates.
- 4) The Aarhaar number is assigned and the record is entered into the Central Database.
- 5) The resident receives their Aadhaar number in the form of a cardboard card, from India Post.



Biometrics are collected using an automated PC workstation.
This isn't the Aadhaar enrollment program. It's similar but I couldn't find an picture of the actual enrollment dialog



Supported Operating Systems: Windows XP/Vista/7, Ubuntu Desktop 9.04+ with Gnome Windowing Toolkit **Enrolment Client .NET Enrolment Client Multi-Platform** Short-term solution for Windows family Final solution for both Windows and Linux family of operating systems. of operating systems only. Microsoft technology stack. Java, open-source technology stack. • .NET Framework 3.5, WPF, Enterprise Java 6, AWT/Swing, Spring, iBatis, Library 4.1, SQL Server R2 2008 Derby, Apache Commons, POI, Hessian, Express BinaryNotes, jCIFS, JDBM, log4j, SuperCSV, BouncyCastle

The Key to Aadhaar and its success is one word – Java!

As you will see, most of the software in Aadhaar is written in Java.

- To jump start the enrollment process, the first set of PC workstations used Microsoft .NET, running on Windows. However, it was quickly replaced by a Java application that could run on any major operating system.
- I'll go over these components in greater detail in a minute...



20



- (Laser pointer) This is the Authentication architecture. There are two Data Centers in an active-active configuration; one in Mumbai (Bombay) and one in Delhi. There is a backup Data Center and Enrollment processing Center located in Bengaluru (Bangalore).
- Since most authentication transactions are localized, the key to fast authentication response are edge cache servers with a realtime cache invalidation.

Daily Identity Verification



21

- Bank Accounts
- Cellphones
- Food subsidies
- Gas rebates
- Mahatma Gandhi NREGA direct salary deposit
- School Enrollments and exams
- Money transfer ('Reliance Fresh')

I bring this slide up because the last item, Reliance Fresh, is a kinda of a clever marketing program. Reliance is one of the largest corporations in the country. They are involved in EVERYTHING. Their grocery store division is Reliance Fresh. They are offering in-store money transfers. So a migrant worker in the North can go to the local Reliance Fresh, buy their groceries and at the same time deposit money to be collected by their family in Southern India, using the wife's Aadhaar number. Wife collects the money in the local Reliance Fresh and then purchases the groceries she needs. Everybody is happy.

Identity Verification



22

- Aadhaar only provides a "YES" or "NO"
 - No Personally Identifiable Information
- Verification depends on "value" of transaction
 - Food subsidies Aadhaar only
 - Low value bank transactions biometrics
 - High value bank transfers "Live Detection" biometrics
 - Temperature
 - Multi-spectral light
 - Sweat gland analysis
 - Multiple images
 - Infrared mapping with pulse detection

Privacy is maintained in that Aadhaar authentication only says that the person either is who they say they are or they aren't. No PII is given out as part of the transaction.

The rigor of the authentication is based on the value of the transaction. For food subsidies, only an Aadhaar number is needed, no biometrics. For your routine bank transactions, biometrics are used. Large value transactions involve biometric readers that have "live detection" built in, to provide protection against gummy fingers...or Worse!

Some of the methods of live detection include temperature, analysis under multi-spectral light, sweat gland changes, multiple images, and IR mapping with pulse detection.

Keys to Enabling Aadhaar Success



23

- Long term commitment
- Government-wide commitment
 - KYC
 - Subsidies
- Competitive Services
- Complete transparency and visibility to contracting and processes
- No Vendor Lock-in
- Open source software
- Open APIs

Here are the keys to Aadhaar's success.

- 1) A long term commitment to the project. It takes time to involve 1.2 billion people and to change the status quo.
- 2) The entire Government embraced Aadhaar. There had been some small Government-based identification systems in the past, like the tax systems PAN card. But Aadhaar was a unified front
- 3) If you can't beat them, compete them. Competitive services provide the best service for a given cost point.
- 4) To go along with this competition, you need complete transparency in the contracting process. No cronyism!
- Of the 7 keys, we are most interested in the last three, no vendor lock in, open source software and open APIs

Open Source Software of Aadhaar (1 of 3)	
Functional Area	Open Source Solution
View Layer and Common User Interface	Java Swing, Spring Framework
Business Logic	Spring Framework
Relational Database	MySQL, Apache Derby
Data Access Layer/ORM	Apache iBatis, Apache Data Base Common Pooling (DBCP)
Remote Method Invocation	Hessian Binary Web Services (gRPC was still in the future)
Inter-Data Center Communication	Apache HTTPS Client

Here is the software of Aadhaar. As I said, the common thread throughout is Java. In fact, you could almost call it an Apache system, given the number of Apache projects used.

This wasn't by chance. UIDAI, through their contractor, MindTree, chose proven open source software with the longevity of the Apache Project behind it. Additionally, there will always be a migration path from these projects to their replacements.

Java Swing and Spring Framework was used for the GUI layer.

Business logic was handled through Spring Framework.

Both MySQL and Apache's Derby were used when a relational database was required.

The data mapping between the Java objects and the relational databases was provided by Apache iBatis and Apache Data Base Common Pooling

Remote procedure calls were made using Hessian Binary Web Services (Google's gRPC was still in the future

Inter-Data Center Communications was via HTTPS

Open Source Software of Aadhaar (2 of 3)	
Functional Area / Module	Open Source Solution
Authentication	Java Common Internet File System(JCIFS) (Oauth was still very much in the future)
Preferences/Init Files	Java Commons Configurations
Object Pooling/Memory Management	Apache JDBM Hash Tables, Apache Commons Pool
Logging	Apache Commons Logging, Log4J
XML Marshaling/Parsing	Apache Digester, Java Architecture for XML Binding

- Authentication was through Java Common Internet File System (jCIFS). Oauth was still very much in the future.
- Init files were handled through Java Commons Configuration
- Memory Management and common data structures came from Apache JDBM Hash tables and Apache Commons Pool.
- Logging was Apache Commons Logging and our good friend Log4J.
- XML was still very much the dominate means of information exchange. XML Marshaling and Parsing were done with Apache Digester and the Java Architecture for XML Binding

Open Source Software of Aadhaar (3 of 3)	
Functional Area / Module	Open Source Solution
CSV Parsing	Super CSV
ASN.1 Parsing	Java BinaryNotes
Excel Format Import	Apache Poor Obfuscation Implementation (POI)
Transliteration (22 Indian Languages)	Quillpad
Queuing/Messaging	AMQP/RabbitMQ
File System	Hadoop Distributed File System (HDFS)

- Files of Comma Separated Values were parsed using Super CSV
- ASN.1 Parsing of certificates was done using Java BinaryNotes.
- A lot of information was transferred from Excel spreadsheets. For this, the Apache Poor Obfuscation Implementation was used (Love that name – in your face Microsoft!)
- Quillpad is an interesting product. There are 22 common languages throughout India; Hindi and English are the two languages set forth in the Constitution, while many of the 28 states and 8 territories have their own local language. Quillpad transliterates between these 22 languages.
- Queuing and interprocess messaging used the RabbitMQ implementation of the Advanced Message Queuing Protocol.
- Finally, the Hadoop Distributed File System was used for file storage, fault tolerance, and streaming.



Each authentication transaction is scanned for fraud. Here we can see that the fraud engine runs on the open source statistical package, Pentaho. This was during the early days of Aadhaar, it would surprise me if a open source neural network model isn't being used



This is not to say that everything was the right choice. Mistakes were made along the way...

- The Hadoop Distributed File System (HDFS) is a prime example. Hadoop Distributed File System is an open source file system designed to distribute data among commodity disk drives, allowing streaming and fault tolerance. HDFS can be considered an alternative to RAID. In the case of Aadhaar, Dell EMC donated a number of EMC storage arrays, hoping that these storage arrays would become the standard hardware for Aahaar. These used RAID. So Aadhaar put HDFS, intended to be a distributed file system, and centralized it on a storage array. Essentially, Aadhaar stacked one fault-tolerant protocol on top of another fault-tolerant protocol, with the expected loss of performance.
- It was also decided that MySQL was probably not a good fit. For performance, the database had to be sharded. This then required running multiple copies. A database with internal partitioning would be a better chose, with fewer "moving parts" and greater reliability.

Mistakes Along The Way (2 of 2)

- RabbitMQ (AMQP)
 - No experience or expertise
 - Single biggest source of problems
- No Monitoring and Management tools
 - Infrastructure monitoring/management (Xenoss, Puppet) only
 - No application level monitoring/management

Advanced Message Queuing Protocol (AMQP) was the key to moving data asynchronously through the enrollment and authentication processes. However, MindTree chose to use the RabbitMQ implementation. Nobody had any practical experience with RabbitMQ. In the end, it proved to be the single biggest source of problems during enrollment.

Speaking of application problems, UIDAI did not pick any application observability products. They had Puppet for compliance management and Xenoss for infrastructure monitoring, but nothing at the application operations level. As Aadhaar started up operations, this caused a significant amount of finger pointing.



Lets Talk About Biometrics



- Irises are like circular barcodes
- Fingerprints are the challenge
 - No public means of indexing
 - Fingerprints are converted into "templates" pattern of topographic points
 - NIST provides "standard" fingerprint templating and matching algorithms
 - Fastest, lowest false-positive/false-negative statistics are proprietary algorithms and hardware (ASICs and neural networks)
- Aadhaar requires speed and accuracy of proprietary algorithms

30

Let's talk about biometrics....

As I said, irises are like barcodes. They are no problem to capture, template, and compare.

- Fingerprints are the challenge. There are no recognized means of indexing them. You have to do a print by print comparison. And of course you don't actually compare raw fingerprints. As each fingerprint is enrolled, it's converted to a topographic map called a template. Similarity between templates is determined based on the linear distances between the points on the topographic model.
- The National Institute of Standards and Technology (NIST) has released open source template and matching algorithm code. This reference code is good, but the fastest, more accurate fingerprint processing involves proprietary algorithms and in some cases, proprietary hardware (ASICs GPU neural networks, etc.).
- Aadhaar required the speed and accuracy of these proprietary matchers. Remember, when you are dealing with 12 billion fingerprints, an error rate of "1-in-a-million" is 1 million, 200 thousand, not exactly unique identification!

"... If you can't beat them, compete them!"



31

- Since open source fingerprint identification is not feasible
 - UIDAI issued 3 competitive "biometric service provider" contracts, each with a different algorithm
 - Remove dependency on one vendor/algorithm
 - Contract payments were based on speed and accuracy for competition and to yield best value for India
 - Raw fingerprints were retained in order to create a new database if one or more or the original vendors dropped out
- Result is three complete individual databases for increase security and accuracy

Despite having chosen a path of open source, nonproprietary software, UIDAI needed to use a proprietary fingerprint matching software. To prevent vendor lock-in, UIDAI chose three competing technologies. This way, they were not dependent on the longevity or product development of any one vendor. They also structured the contracts so that the three vendors were competing. This helped assure that India received the best pricing while the vendors stayed within the constantly tested quality standards.

- India also retained the raw fingerprint scans so that a new vendor could be added if any of the original three went out of business.
- In addition to mitigating the proprietary software risk, these approach yields three complete individual databases for increase security and accuracy.



The three Biometric Service Providers are L-1, Morpho, and Accenture, the company I was working for at the time. Unlike L-1 and Morpho, ours is a team of different specialists.





33

Authentication

-This is a "1:1" match

-Aadhaar number is used to index into the gallery to find the single set of biometrics that should be a match

-Consistent less-than 100 millisecond match times can sustained

EASY

Let's look at biometric matching for a minute... Authentication, checking somebody's prints against their prints on file is a one-to-one. This is EASY!

Biometric De-Duplication at Enrollment



34

- De-Duplication
 - As each new enrollment is processed, the biometrics are matched against all previously enrolled sets of biometrics – AKA "1:N" matching
 - -Time to compare two biometrics is essentially constant
 - -Therefore, time to do a 1:N de-duplication is governed by
 - The number of existing biometrics in the in-memory database [disk is too slow] (referred to as the "gallery") and
 - The number of threads or processors searching the gallery
 - -Biometric service providers "tune" their ability to meet 1:N matching SLAs by increasing the number of servers and by sharding the galleries

During enrollment, we check every new fingerprint against every other fingerprint on file. This is called de-duplication. This is a 1to-N problem. You can see how as the number of fingerprints on file increases, the de-duplication time increases. Biometric service providers control overall de-duplication times by increasing the number of servers. Aadhaar – The Last Day

AADHAAR

Comparing 1 Million New Enrollees...

... Against 1 Billion Existing Enrollees

$1 \times 10^{6} \times 1 \times 10^{9} = 1 \times 10^{15}$

1 Quadrillion Comparisons in a day!

If you are using a database for workflow, that's:

11 Billion Transactions/Second!

(It's Even More If You Individually Log Each of 10 Fingers, 2 Irises, and a Face)

So, let's look at Aadhaar on the hypothetical "last day", when enrolling the last million residents against the 1.2 billion residents already on file. That's 1 quadrillion comparisons in a day.
If you are using a database for workflow, which we were, that's 11 billion transactions per second. It's even more if you log each fingerprint and iris separately in the workflow

My Company's Problem



36

- Accuracy was outstanding
- Enrollment speed was miserable :-(
- UIDAI did a slow rollout of enrollment to work out kinks in the process
 - 500,000 enrollments per day
 - We were struggling
 - UIDAI proposed modifying the contract to 3 million enrollments per day to speed up rollout!
- What was the bottleneck?

I was called in in my fire fighter role when we ran into performance problems.

Our accuracy was great but our enrollment times were miserable.

UIDAI set the contract at 1 million enrollments per day, but they decided to start out slow at 500, 000 daily enrollments and let everybody work the kinks out.

We were struggling to make the 500,000. There was no way we could get to 1 million enrollments per day!

To make matters worse, UIDAI was talking about increasing the number of enrollments to 3 million per day!

Obviously, my first job was to find out where our bottleneck was.



Here is a high level view of our initial architecture. As I said, we were integrating the results of multiple vendors and we were using a very expensive, proprietary product called Daon for this integration and workflow. Daon was in turn using Oracle Real Application Clusters, pronounced "RAC" for workflow logging. RAC expands through multiple, parallel instances.



Here's the RAC architecture. It is supposedly infinitely expandable.

Now Oracle specifically prohibits customers from benchmarking their software. It's right in the license agreement



...but one Oracle partner has published data that shows that for the average real-world workload, most Oracle Real Application Cluster systems stop scaling after 5 nodes in the cluster. This was certainly my experience in India. We had 4 nodes in the cluster and we couldn't push more transactions through the system by adding more nodes or by tuning the nodes we had. And when I say "we", I mean Oracle themselves. We had Oracle's own developers from both Redwood Shores and Bangalore helping us.

Houston... We found our problem and it's Oracle!

My first job as the Accenture Chief Architect for Aadhaar was to replace the Oracle database-driven workflow system with a new workflow system, based on the Actor Model and queues. This workflow system now easily handles 1 million enrollments per day and we can dial in processing capacity by increasing the number of servers₃₉



- Here is our revised architecture. We used RabbitMQ to asynchronously publish fingerprint messages. Each fingerprint template or matcher, as it became free, grabbed the next message in the queue and processed it.
- We knew how many processes to run based on whether our queues were growing, shrinking, or remaining constant. This was autoscaling without worrying about how fast the system could spin up new processes! The queue just buffers...
- We then used an in-memory version of MySQL as a synchronization barrier and a MySQLbased program to route the end results correctly to UIDAI.

What Works

Asynchronous coupling of the processing pipeline

-Message queues

- Absorb differences in processing rates
- Fault tolerant
- High availability
- Geographical distribution
- No persistent "state"
 - -Databases too slow and compute intensive
 - -Rebuild state from redundant factors

This was an educational journey for me as well. I came away from this experience with new ideas of how to build large scale systems

What works...

- Asynchronous coupling in the processing pipeline. You do this using message queues. Some of the advantages are:
- The message queue buffers and absorbs differences in processing rates
- Using multiple, synchronized message queues, you can have fault tolerance and high availability. Put these synchronized queues in difference data centers and you have geographic distribution and local caching.
- Second, don't try to keep state if you can avoid it. Make every message self-contained, to the maximum extent possible.

41

In Conclusion...





Keys to Enabling Aadhaar Success



43

- Long term commitment
- Government-wide commitment
 - KYC
 - Subsidies
- Competitive Services
- Complete transparency and visibility to contracting and processes
- No vendor lock-in
- Open source software
 - Apache or other long term community project
- Open APIs

Here are the keys to Aadhaar's success.

Long term commitment Government wide commitment Competitive services Complete transparency in contracting No vendor lock-in Open source software And not just any open source software. The key is to select products had have a strong commitment from their projects or a recognized migration path And open APIs

What Brad Learned...



44

- Open source is generally better than proprietary software
- Pull is always better than push
 - Fudd's Law
- Queues are excellent connectors
- Stateless messaging has more advantages that trying to maintain state

I learned how to better architecture large scale systems.

- 1) open source is better than closed source
- It's always better to let your application pull work from a work queue than to pound it with procedure calls. Remember, if you push something hard enough, it will fall over
- 3) Queues are an excellent way of connecting applications
 - They are asynchronous
 - They aid fault-tolerance
 - They can be used for predictive autoscaling
- Finally, inclusive messaging allows for stateless processing and stateless processing scales better

Aadhaar Problems

- Over-zealous application (Supreme Court rulings)
 - Not required
 - Not applicable to voting
 - Can not restrict school entry
 - Not required for passport
- Resistance to law enforcement
- Data leaks
 - 3rd party storage of PII
 - "R.S. Sharma incident"
- Corruption within UIDAI
 - Unauthorized access

45

Needless to say, a national biometrics identification system has not been without its problems and critics

- While the Supreme Court of India has declared Aadhaar legal, it has also ruled in a number of cases where Aadhaar was being applied over-zealously.
- For example, while Aadhaar is sufficient to get an Indian visa, it is not a requirement.
- The court, in general, has said that Aadhaar should always expedite processes but it should never prevent them
- UIDAI has resisted attempts of law enforcement to use Aadhaar for tracking individuals
- Most of the information leaks associated with Aadhaar are actually leaks of third parties that have associated the individual's Aadhaar number with their PII. In one case, a former Director of UIDAI and a colleague of mine, Shri R.S. Sharma, to prove how "safe" Aadhaar was, published his Aadhaar number on Twitter. Within hours, virtually all his personal details and social media accounts had been published. None of this information came from Aadhaar. Rather it was third party databases, like social media, that, in some cases illegally, stored his Aadhaar number with his personal details.
- Aadhaar has never been hacked, but there has been corruption within UIDAI. Government managers have taken bribes for access to individuals records. In fact, there was a large purge of Aadhaar managers in direct response to corruption investigations.

Aadhaar Today

- This presentation deals with the initial roll-out of Aadhaar during the 2005-2010 time-frame
- How is Aadhaar today?
 - 1.5 billion Aadhaar Numbers issued
 - India has surpassed China in population (1.4 billion)
 - Residents, not Citizens
 - No duplication
 - 2.4 billion authentications per month
 - Fully integrated into the welfare programs, financial industry, Government services, cellphone management, etc.
- UIDAI publishes Aadhaar statistics at:
 - https://uidai.gov.in/aadhaar_dashboard/

This presentation, and my role in Aadhaar, covers the initial stand-up period of 2005 to 2010.

So, what's been happening in the past 14 years?

Well for one, India has become the most populated country in the world, having surpassed China. The current population is 1.4 billion

Aadhaar has enrolled 1.5 billion individuals. That breaks down to 1.33 billion citizens (95% enrollment rate) and 670 million residents.

UIDAI does 2.2 billion biometric authentications per month

Aadhaar has been fully integrated into both the Government and private sector.

UIDAI believes in total transparency and publishes all their statistics at: Here's where you either want to download my slides, or at least take a photo of this page for the URL to the Aadhaar dashboard.

Funny story about transparency. For a while Government employees are certain ministries were required to clock in and out using their biometrics. The Government then published the statistics on number of clocked in employees and their time at-desk. Let's just say that the statistics were embarrassing enough that the Government soon took that particular dashboard down! ;-)

46

Aadhaar as a Model



47

Being studied by multiple countries



ID4AfricaIndian ID Outsourcing as a Service (IIOaaS)

Aadhaar is a success. It has shown that a national biometric identification program can be used to solve a number of social welfare problems.

Accordingly, it is being studied by a number of countries and organization.

- Here is Shekanth Nanamuhdy, the Technical Director of Aadhaar demonstrating the system to the leader of a well-known first world country.
- In fact, because of the proven open source processes and the experience UIDAI has gained, I could see India becoming the Identity Leader, outsourcing Identity as a Service.



Thank you! Copies of these slides and my talking notes are available on the Formularity website, as well as through SouthEast Linux Fest. Are there any questions?...