Architecting World's Largest Biometric Identity System: Aadhaar Experience

Strata + Hadoop World – New York, Oct 2014

@pramodkvarma
linkedin.com/in/pramodkvarma
India

• **1.2 billion** residents
  – 640,000 villages, ~60% under $2/day, ~75% literacy,
  – <3% pays Income Tax, <20% banking,
  – ~1 billion mobile connections
  – ~300-400m migrant workers

• **$50 billion** direct subsidies every year!
  – Residents have **no standard verifiable identity**
  – Most programs plagued with ghost and multiple identities causing **leakage of 20-40%**
Vision

• Create a digital “national identity”
  – Biometric backed “real” and “unique” identity
  – Purely a number (no cards!)
  – “Verifiable online” for portability
  – Built as online identity “platform”

• Applications ecosystem using open APIs
  – Aadhaar enabled bank account and payment platform
  – Aadhaar enabled electronic, paperless services
  – Many other applications being built
Enrollment

Demographic Data

• Compulsory data:
  – Name, Age/Date of Birth, Gender and
  – Address of the resident.

• Optional data:
  – Mobile number
  – Email address

12-digit Aadhaar Number
Unique, lifetime, biometric based identity
Any combination of attributes

Demographics
Name, gender, DoB, Age, Address, Mobile, Email, ...

PIN / OTP

CIDR
Authentication Servers

Service Provider (Authentication User Agency)

Applications deployed on any device form-factor, any OS, any programming language, any network
Work so far

• Program launched on the ground in Sept 2010
  – First 10 million enrolled by June 2011 (~270 days)
  – Second 10 million in 35 days and third in 24 days
  – Now 10 million IDs issued every 10 days!!
  – 690+ million Aadhaars (IDs) issued so far
  – On target to complete 1 billion by 2015

• Online services (APIs)
  – Authentication and related APIs launched in 2012
  – 100+ million authentications already done!
  – 100’s of developers in the developer community
Aadhaar Enabled Applications

• Number usage
  – As a “unique” ID to remove duplicates and ghosts
  – “Know Your Customer” during account opening
  – As “financial address” allowing money to be sent to the number!

• Authentication
  – Aadhaar in micro-ATMs for money transactions
  – Aadhaar for “person present” transactions

• Many more coming ...
  – Aadhaar for digital signature, electronic documents, and digital locker services
  – Aadhaar for school/college/university certificates
  – Aadhaar for credit registries
Technology

Architecture At a Glance

• Open architecture
  – Use of open standards to ensure interoperability
  – Allow the ecosystem to build on top of standard APIs
  – Vendor neutral, using of commodity computing and open source
  – Designed to work with any device, any form factor, any network

• Design for scale
  – Every component needs to scale to large volumes
  – 100’s of millions of transactions and billions of records
  – Fully multi location distributed architecture for horizontal scale

• Security & Privacy
  – End to end security of resident data (PKI, DSC, tamper detection)
  – Data privacy (Minimal data, API and data anonymization)
Enrolment Scale

• Original target - 600 million IDs in 4 years
  – 690 million IDs issued in 4 years!
  – Processing 1 million IDs every day
  – 600+ trillion biometric matches every day!!!

• ~3-5MB per resident
  – Maps to about 10-15 PB of raw data (2048-bit PKI encrypted!)
  – About 30 TB I/O every day
  – Replication and backup across DCs
  – Lifecycle updates and new enrolments will continue for ever

• Additional process data
  – Several million events on an average moving through async channels (some persistent and some transient)
  – Several billion records in analytics system already and is growing
  – Needing complete update and insert guarantees across data stores
Enrolment Architecture

- Persistent Store
  - Hadoop
  - Solr
  - MongoDB
  - MySQL
- BI & Analytics
  - Pentaho
- SEDA Stages
  - Structure Validation
  - Demographic Validation
  - Operator Validation
  - Biometric De-duplication
- Biometric System
  - Biometric Server 1
  - Biometric Server 2
  - Biometric Server 3

Multi-platform Java client with multi-vendor device support

Rejected
Biometric De-duplication Stage - 600+ trillion matches a day!

Operator Check

ABIS Accuracy

Tool + Manual Adjudication

Duplicate

Reject

Not-Duplicate

Aadhaar

Duplicate

Duplicate

Identify

50% 30% 20%

Duplicates

Identify

Identify

Identify

Identify

ABIS Solution 1

ABIS Solution 2

ABIS Solution 3

ABIS API

Insert 10cr

Resident Enrolment Database

2 cr M M M

E-Pkt 1

2 cr M M M

E-Pkt 2

2 cr M M M

E-Pkt 3

ABIS API

Insert 10cr

Resident Enrolment Database

2 cr M M M

E-Pkt 1

2 cr M M M

E-Pkt 2

2 cr M M M

E-Pkt 3

ABIS API

Insert 10cr

Resident Enrolment Database

2 cr M M M

E-Pkt 1

2 cr M M M

E-Pkt 2

2 cr M M M

E-Pkt 3
Authentication Scale

• API over HTTPS over private secure network
  – Supports fingerprint, iris, and OTP (single or multi factor)
  – Input PKI encrypted at source and digitally signed by partners
  – Front-end form factor and OS independent

• 100+ million authentications per day (10 hrs)
  – Possible high variance on peak and average
  – Sub second response and guaranteed audits
  – Multi-DC “active-active” architecture

• Authentication request is about 4 K
  – 100 million authentications a day
  – 1 billion audit records in 10 days (30+ billion a year)
  – 4 TB encrypted audit logs in 10 days
Authentication Architecture

- Prod Auth Server Cluster
- Production HSM
- MySQL Replica
- HBase
- HBase (Audit)
- API End-point (HTTPS) SSL Accelerator & LB
- Replication
- OTP and in-memory cache Cluster
- SMS/Email GW
- Production MySQL DB
- HBase
- HBase (Audit)
- Template Extraction Servers
- Data node
- Data node
- Data node
- Data node
- Data node
- Data node
- Data node
- Data node
- Data node
- Replication

DC 1

DC 2
System Stack

• Commodity hardware
  – Heterogeneous, multi-vendor servers, tiered storage

• Applications
  – Java, Spring, Mule, RabbitMQ
  – Running on Linux

• Data Sources & BI
  – Simple Distributed Archive Storage over XFS
  – HDFS & HBase (MapR), MySQL, SOLR, Hive, Pentaho

• User Interface
  – HTML 5, Liferay, Java Swing, Android
Aadhaar in Action
Payments through Micro-ATMs
Food Delivery
Cooking Gas Delivery
Old Age Pension Payments
Final Thoughts

• Aadhaar is the largest biometric identity system in the world already covering over 690 million people!
• Completely built on open standards and open source platforms, taking advantage of big data technologies
• Built truly as an “online identity platform” enabling many Government and private applications

A great example of using technology to make a difference in the lives of 1.2 billion people!!

Towards Digital India

Aadhaar identity platform and Aadhaar enabled applications Starting to help a billion people to participate in digital economy and avail services tailored to them!
Thank You!

@pramodkvarma
linkedin.com/in/pramodkvarma