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
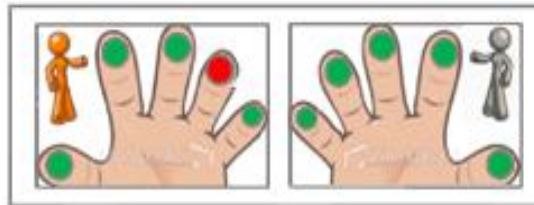
Enrollment

Demographic Data

- Compulsory data:
  - Name, Age/Date of Birth, Gender and
  - Address of the resident.
- Optional data:
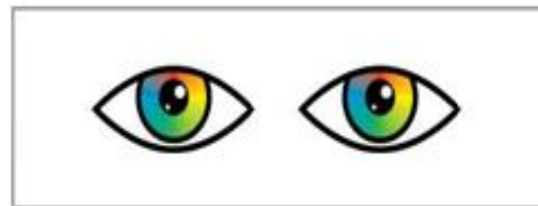 - Mobile number
  - Email address

Biometric Data
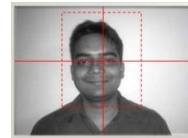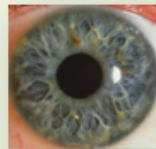

Photograph

All 10 Fingerprints

Both Iris

12-digit Aadhaar Number

Unique, lifetime, biometric based identity

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

Detail whitepaper at http://bit.ly/uid_tech

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

AADHAAR

Enrolment Architecture

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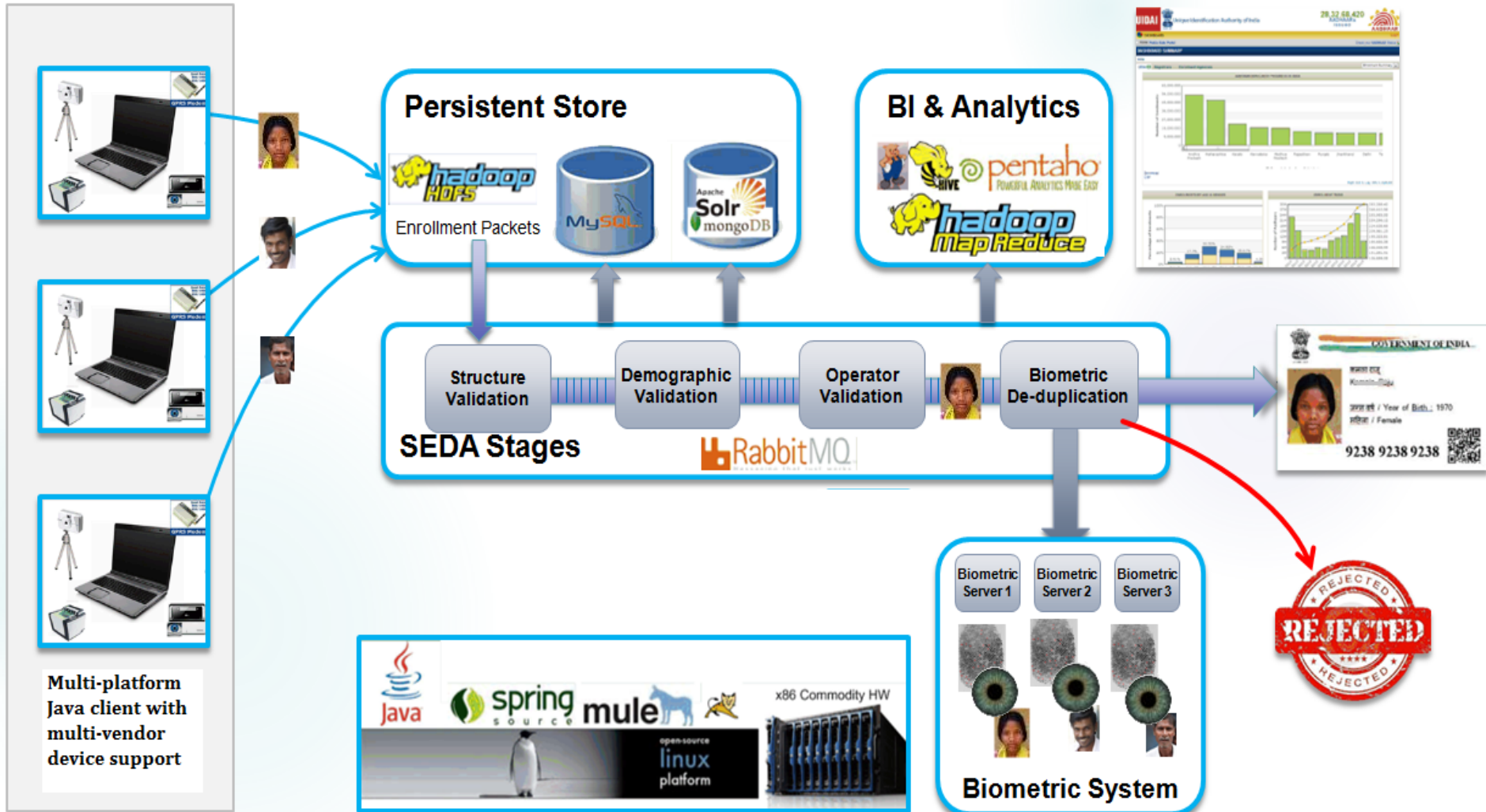
Authentication Scale

- 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

AADHAAR

Authentication Architecture
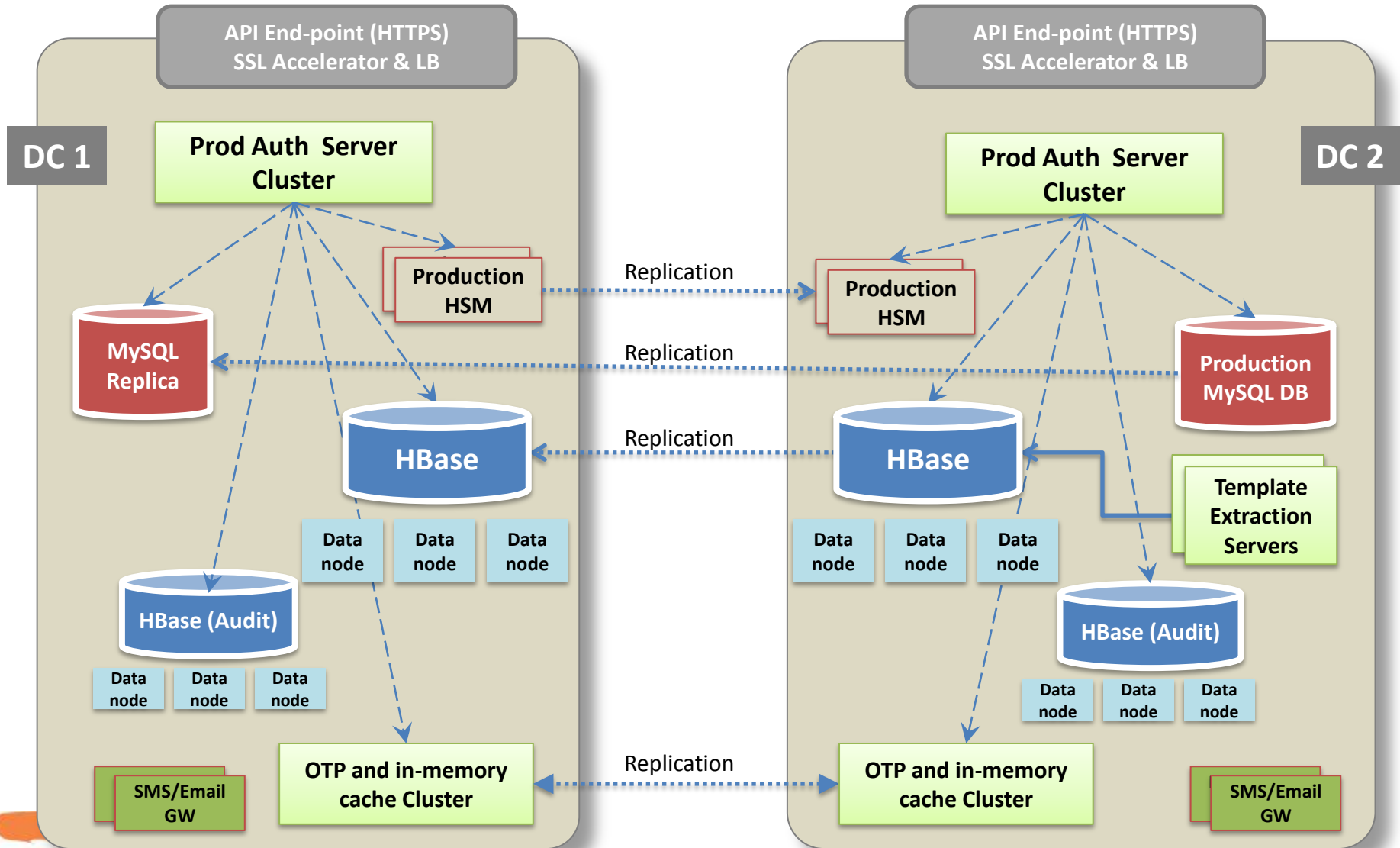
System Stack

- 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
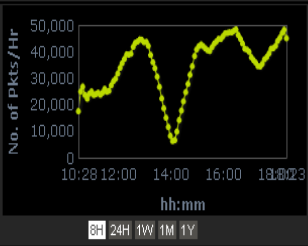- User Interface
  - HTML 5, Liferay, Java Swing, Android

AADHAAR

UIDAI

AADHAAR

Select View
NOC APP Metric

WELCOME APPMONUSER Help Logout

Select REG / EA All Registrar's All EA's ENTER
Select Date 22/12/2011 22/12/2011

INPUT RATE
0
packet per hr

THROUGHPUT RATE
44976
packet per hr

No. of Pkts/Hr
50,000
40,000
30,000
20,000
10,000
0
10:28 12:00 14:00 16:00 18:00 23
hh:mm
8H 24H 1W 1M 1Y

AVG LATENCY TIME
87197 Hr 24 min 0 sec

HOLD RATE
10805

COMPLETION
575,007
295,429
AADHAAR GENERATED COUNT
PKT HOLD COUNT
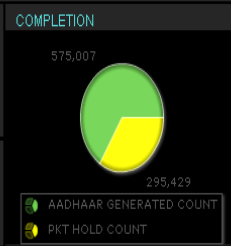
CUMULATIVE DETAILS
6,42,009 5,74,266 0
Packet(s) Received AADHAAR(s) Generated Packet(s) Rejected

5,74,266
AADHAAR(s) generated today

AADHAAR REJECTED
XML PRINT GENERATED
574667 1501

0 6,44,612 3,52,403 0
Packets Uploaded Today Packets Picked up Today HDFS writes Today Picked up from QA Out Today

SFTP PACKET EXTRACTION HDFS 1676 STRUCTURE VALIDATION AUTO DEMO CHECK AUTO OPERATOR CHECK AUTO SUPERVISOR CHECK AUTO INTRODUCER CHECK AUTO RESIDENT DEDUP

2776

MEDIA

0 PRE INBOX 2767 QA MANUAL 6,42,009

Morpho L1 Accenture

INSERT

QUEUE
No. of packets
5,000,000
4,000,000
3,000,000
2,000,000
1,000,000
10:29 12:00 14:00 16:00 18:00 06
hh:mm
8H 24H

THROUGHPUT RATE
No. of pkts/hr
130,000
104,000
78,000
52,000
26,000
10:29 12:00 14:00 16:00 18:00 06
hh:mm
8H 24H

IDENTIFY
15246
0
0

0 per hr 0 per hr 20744 per hr 13013 per hr 18109 per hr
0 743435 1336712 692556 1315239
0 0 0 82867 0

2944171 3918317 2453671

Estimated Completion Time Throughput Rate
51:34 57094
65:25 59900
46:36 52656

XML UPLOAD TO PRINTER
ACKNOWLEDGEMENT FROM PRINTER
DELIVERY ACKNOWLEDGEMENT
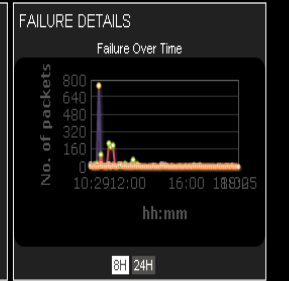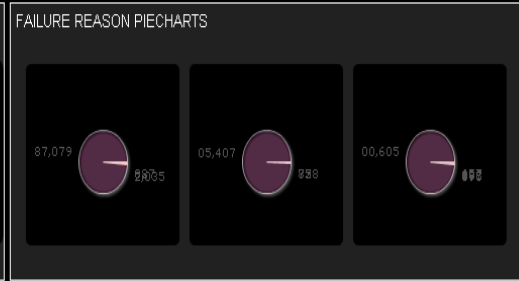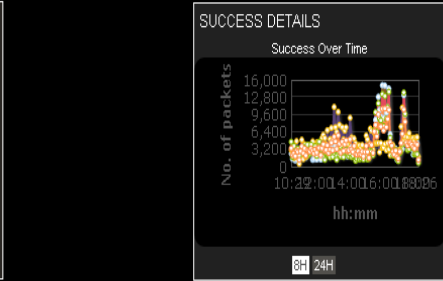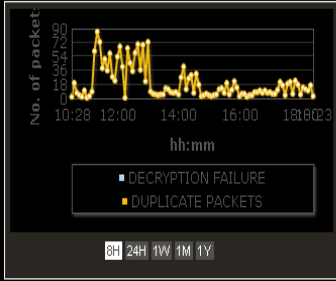
DUPLICATE PACKETS 2603 DECRYPTION FAILURE 0

H:0 R:0 H:1600 R:0 H:59246 R:0 H:2222 R:0 H:11 R:0

S:786988 S:751637 S:926449 F:2960 F:855 F:2440

2,95,184
on Hold today
0
Rejected today

No. of packets
90
72
54
36
18
0
10:28 12:00 14:00 16:00 18:00 23
hh:mm
DECRYPTION FAILURE
DUPLICATE PACKETS
8H 24H 1W 1M 1Y

SUCCESS DETAILS
Success Over Time
No. of packets
16,000
12,800
9,600
6,400
3,200
10:29 12:00 14:00 16:00 18:00 06
hh:mm
8H 24H

FAILURE REASON PIECHARTS
87,079 05,407 00,605

FAILURE DETAILS
Failure Over Time
No. of packets
800
640
480
320
160
0
10:29 12:00 16:00 18:00 05
hh:mm
8H 24H

HOLD REASON
270,864
2,581
21,948
PACKETS OPERATOR BIO VAL FAIL
PACKETS INACTIVE

REJECTED REASON
619
421 157
307
PACKETS RDEMO DATA QUALITY REJECT
REASON - OTHERS
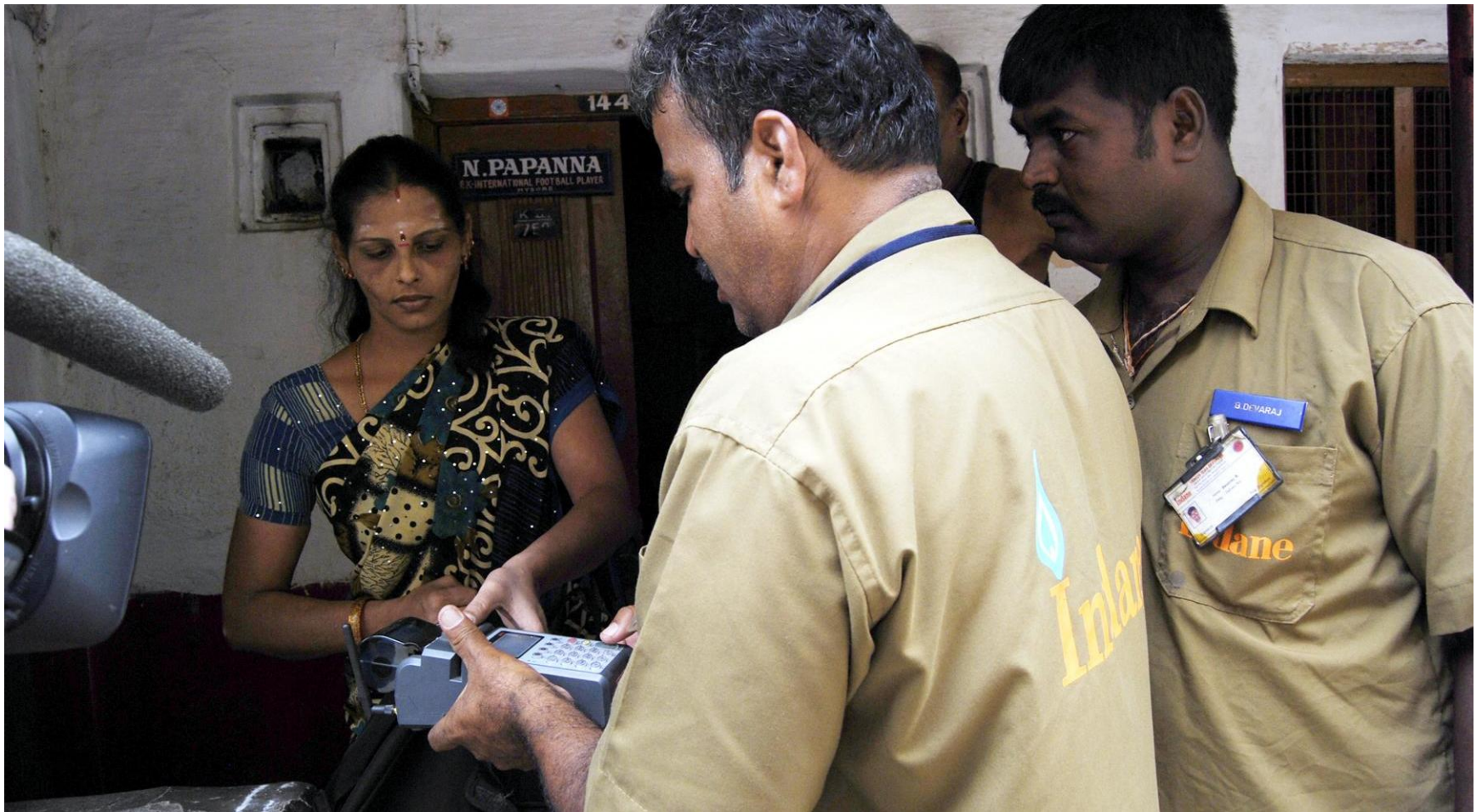
Aadhaar in Action

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!!

Detail whitepaper available at http://bit.ly/uid_tech

Towards Digital India

- Digital Signature & Electronic Documents
- Citizen Portals & Content
- Credit Registries
- Govt. Services
- Electronic Payment
- Insurance, banking
- Mobile Commerce

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

AADHAAR

Thank You!

@pramodkvarma

linkedin.com/in/pramodkvarma