Society of Cable Telecommunications Engineers
30TH ANNIVERSARY
SCTE CABLE-TEC EXPO '13
OCTOBER 21-24 / ATLANTA, GA

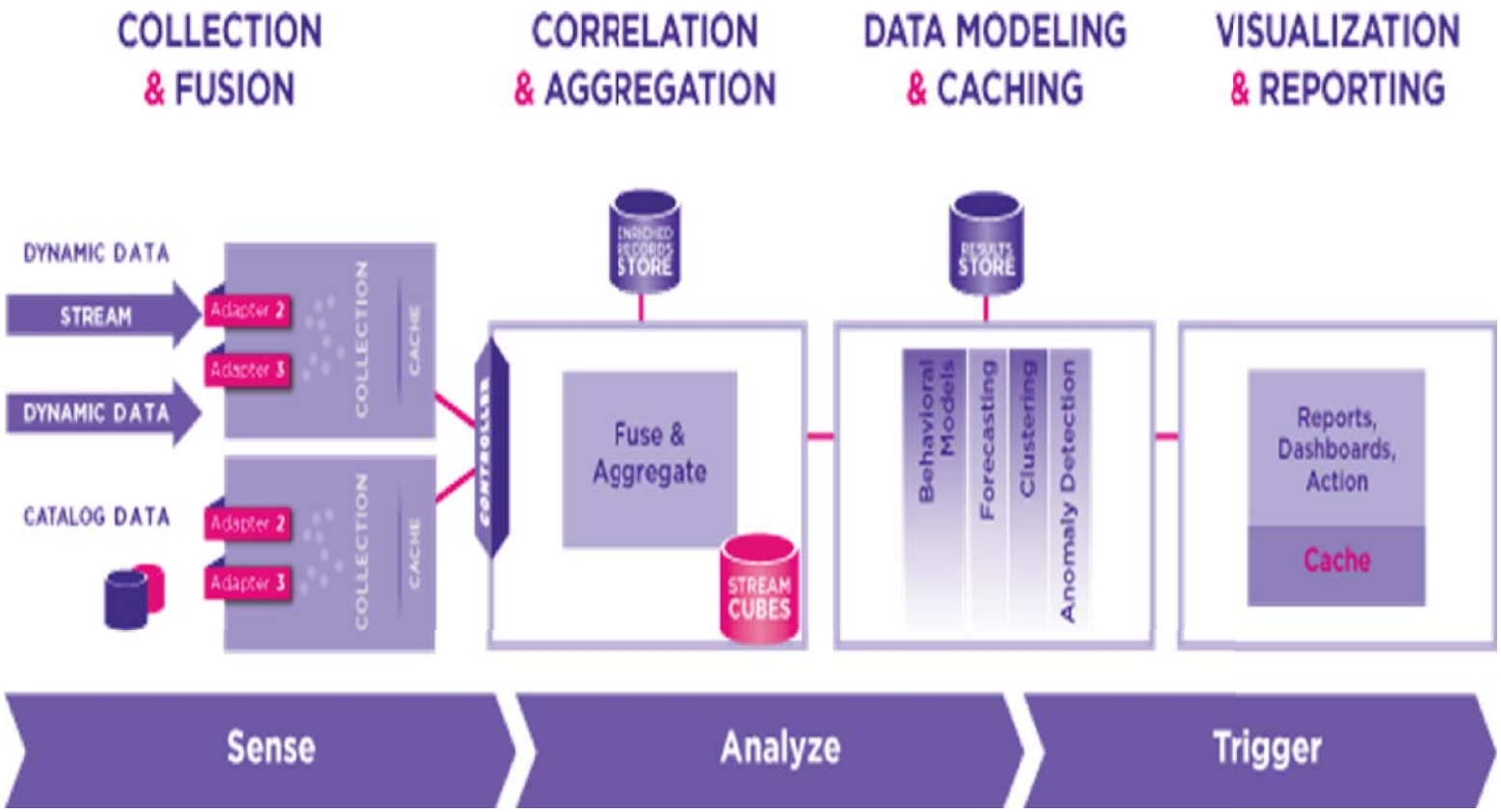
HOW DO YOU CONSUME A YOTTA DATA? ONE BYTE AT A TIME!

Kalpa Subramanian & Dave Kowolenko
Director, NETO & Vice President, NED
Comcast Cable

Agenda

- Data Lifecycle
- The Big Data Problem
- Case Study
  - The “one”
  - Architecture
- Q&A

Data Lifecycle



Picture referenced above “Data Analytics Frameworks”, Guavas Inc., July 27th 2013, http://www.guavus-new.com/solutions/platform/

The Big Data Problem

New product deployments:
- Effectiveness of new product
- Customer satisfaction
- Usage, frequency of touch
- Example: Speed increase project

Routine maintenance:
- Planning & communication
- Pre and post validation of success of maintenance

Outage Management
- Immediate correlation
- Just-in-time communication to frontline
- Rapid resolution across entire outage footprint

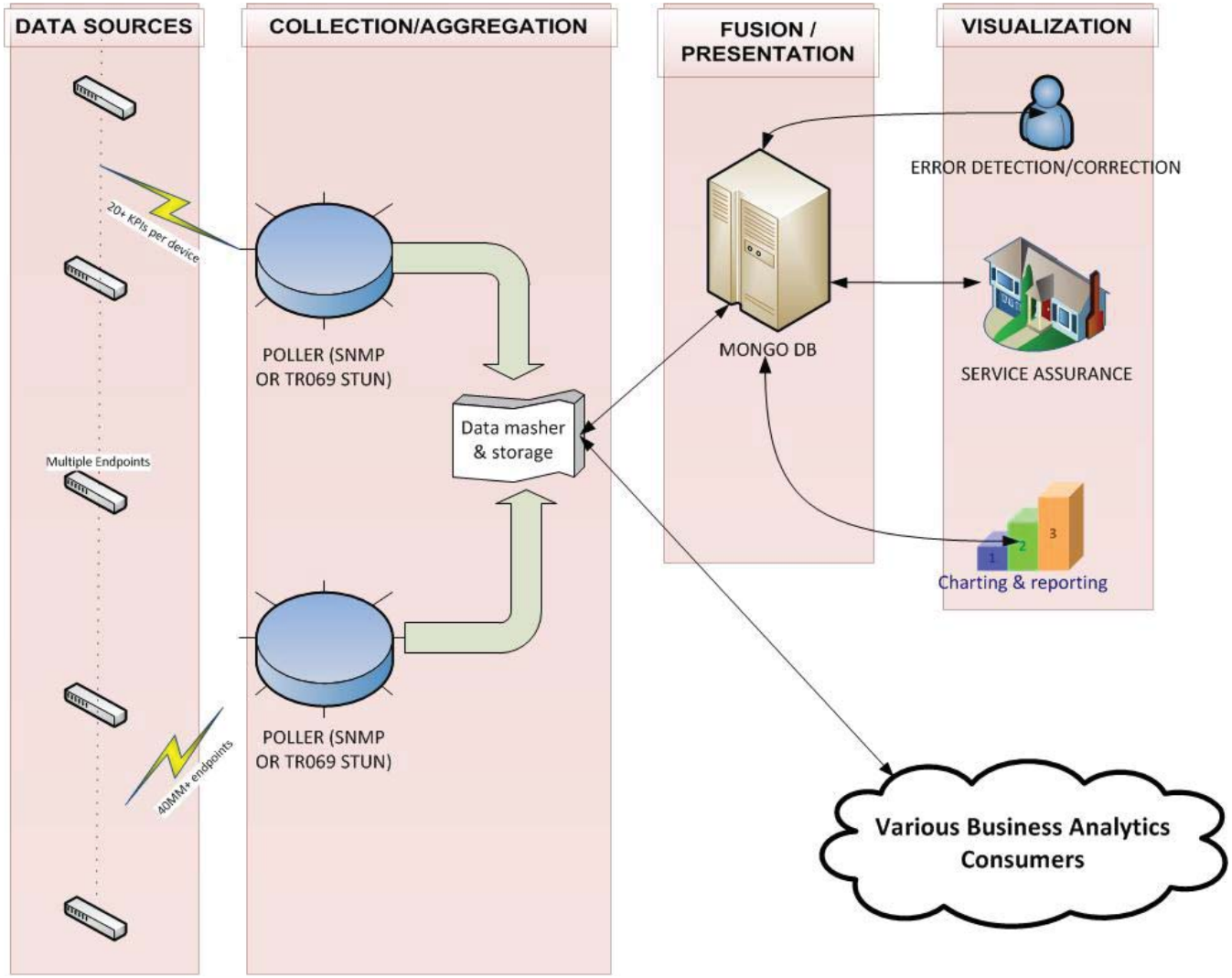
Demonstrate a potential solution

- No silver bullet!

Example of an real issue

- 40 million customer premise equipment (CPE) polled every 10 minutes
- Multiple key performance indicators are polled from each device during polling cycles
- Data collected via SNMP to feed various REAL TIME business analytics
- Need to make this data consumable

Case Study



- Data size = 18Gb every 10 minutes.
- 800 million KPI inserts theoretically
- Case study to evaluate how FAST the data can be evaluated and presented.

NEED SPEED!

The “one”

Relational SQL DB

ACID compliant

Rollback & recover

ANSI SQL out of the box

Mostly row based for relationship

No SQL DB

Generally open source w/ some commercial

Varying degrees of ACID

Reload & Rebuild

SQL compliancy on as needed basis

Requirements

BIG SPEED

Therefore

NoSQL

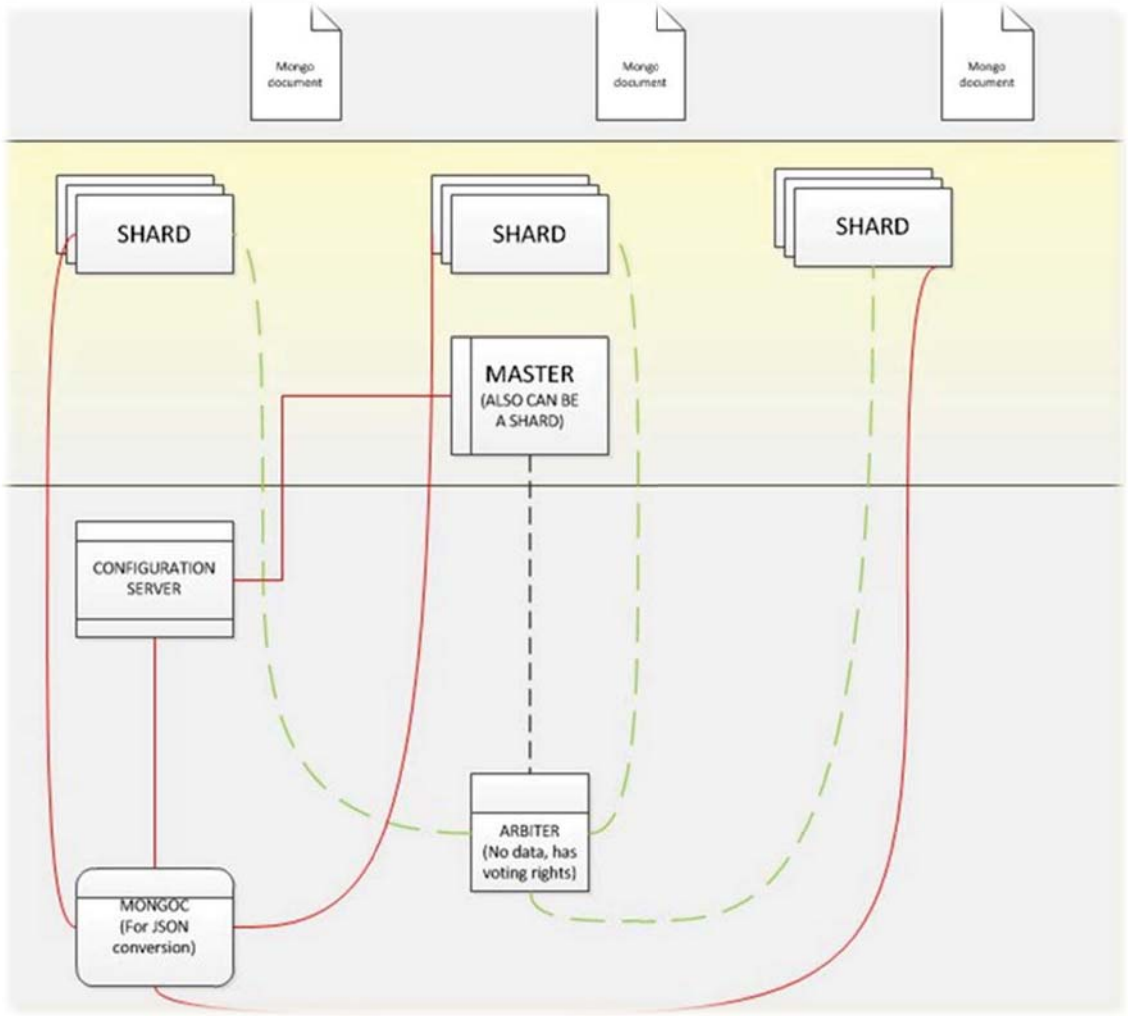
True: Hadoop, Riak, MongoDB, etc.

False: IBM, Vertica, Oracle

MONGO DB!

Schauer, Paul; 2013; “CUDA Data Store Overview”; Published in Comcast Users and Developers Association

Architecture



Example 1 - Count of modems with a specific vendor:
Postgresql
- Return time = 11.4s

MongoDB
- Return time = 4.4s

Example 2 – Single modem query for all KPIs collected:
Postgresql
- Return time = 9.2s

MongoDB
- Return time = 1s

Example 3 – Complete list of all modems in a CMTS
Postgresql
- Return time = 10.7s

MongoDB
- Return time = 4.6s

* Comparative test results was take at point in time

Society of Cable Telecommunications Engineers

30TH ANNIVERSARY

SCTE CABLE-TEC EXPO '13

OCTOBER 21-24 / ATLANTA, GA

Kalpa Subramanian

Director, National Engineering & Technical Operations,

COMCAST,

Kalpa _subramanian@cable.comcast.com

Dave Kowolenko

Vice President, Engineering
North East Division,
COMCAST,
Dave_kowolenko@cable.comcast.com

COMCAST