

Timely Decisions with Big Data: Opportunities & Challenges



Anukool Lakhina
Guavus, Inc.

Double Clicking on Big Data



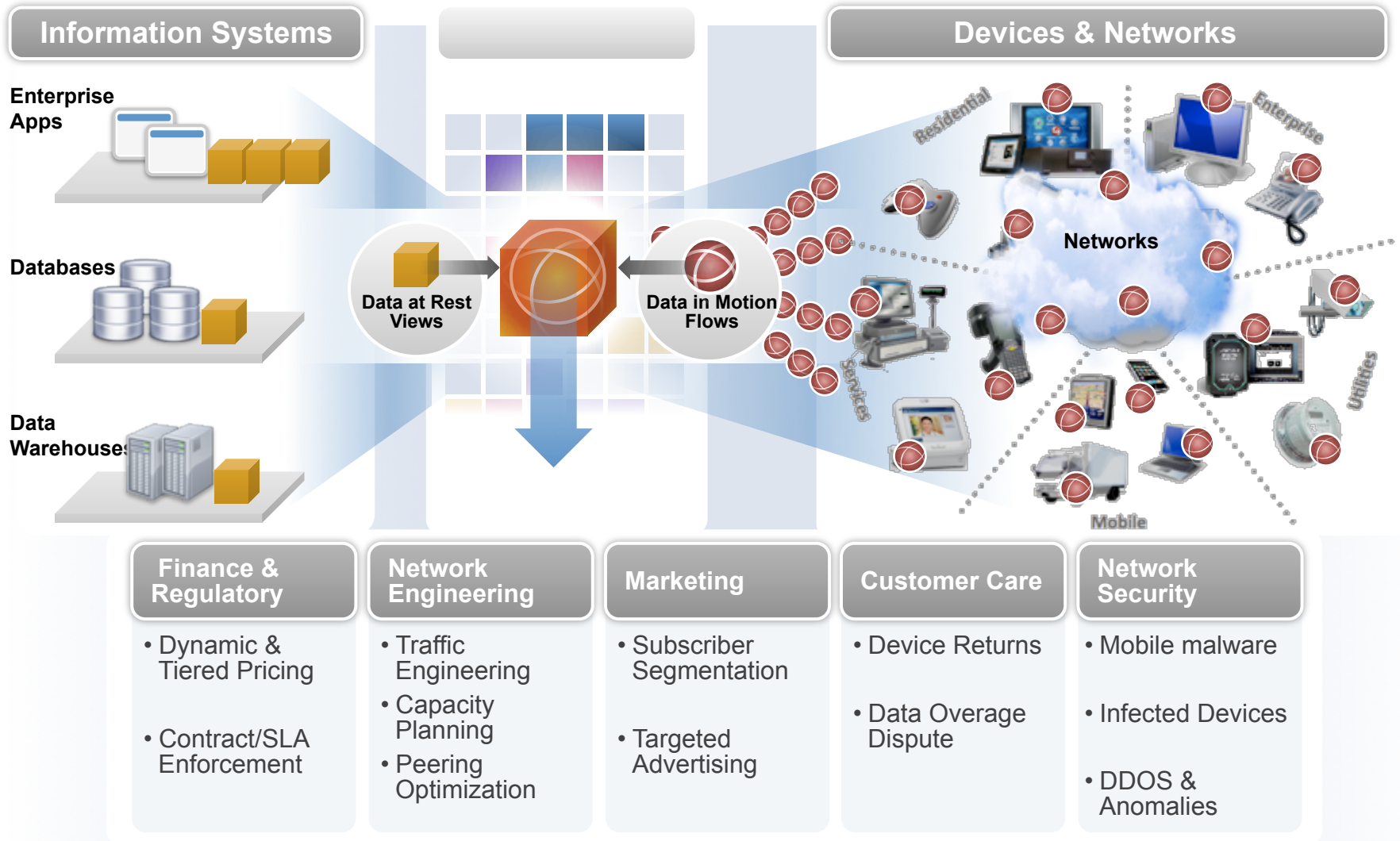
Sources of Big Data:

Data became big because networks of sensors started generating data

Defining Big Data:

Big is not just about **volume**, it is also **velocity** at which data is produced, the **variety** & # of sources which generate data, and the **cardinality** of the dimensions

Opportunities with Big Data



A Typical Scenario

Opportunities

- **Network:** How much demand is the new iPad putting on my network now?
- **Marketing:** How can I profile my subscribers, and target them with just-in-time tailored monetization offers, e.g. advertisements?
- **Care:** How do I respond to customers when they call to complain about data charges?
- **Security:** How can I detect & isolate devices that are using malware apps / URLs?

Scale Challenges

- 100M+ Customers
- 250+ Billion records generated per day
- 200+ Terabytes per day
- High Dozens of Distribution Centers
- PB of stored data for compliance & forensics
- High Dozens of distinct feeds from network that need to be fused for e2e view
- Need for high availability, rapid recovery & fail-over

Challenge: Need for an End-To-End View

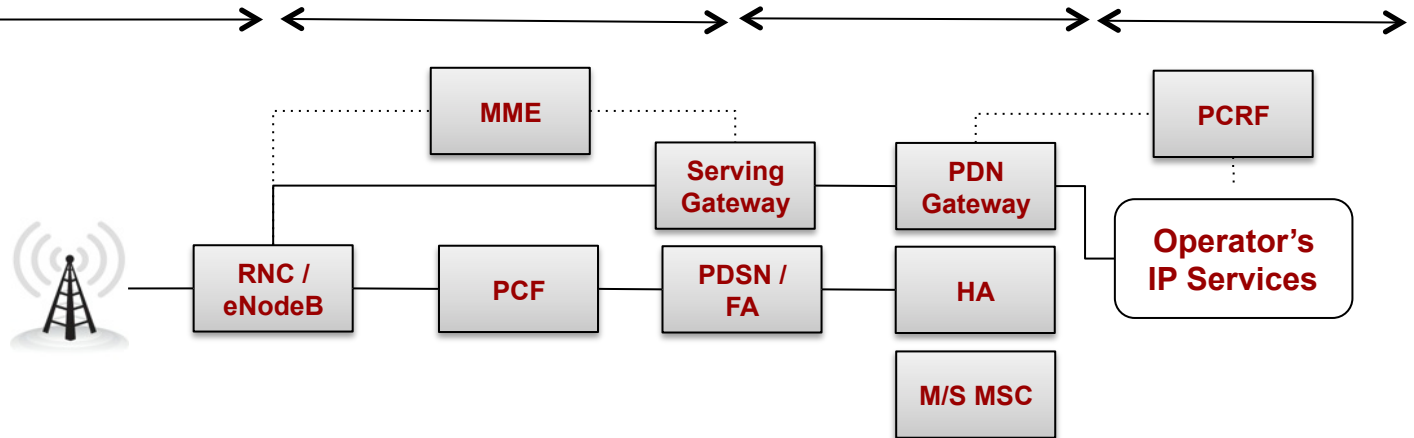
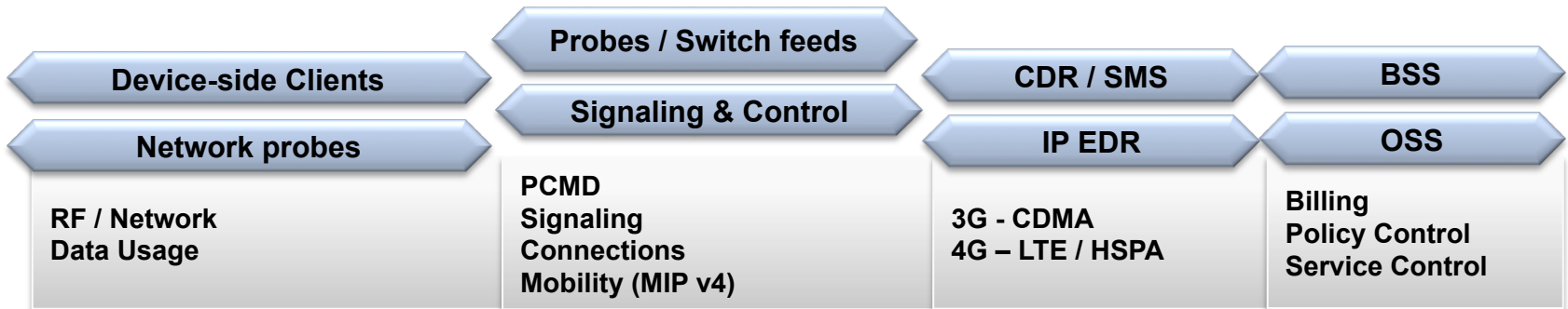
**Decisioning Apps
Analytics Platform**



- Reports
- BI Dashboards
- Decision Support & Planning systems
- External analysis workflows

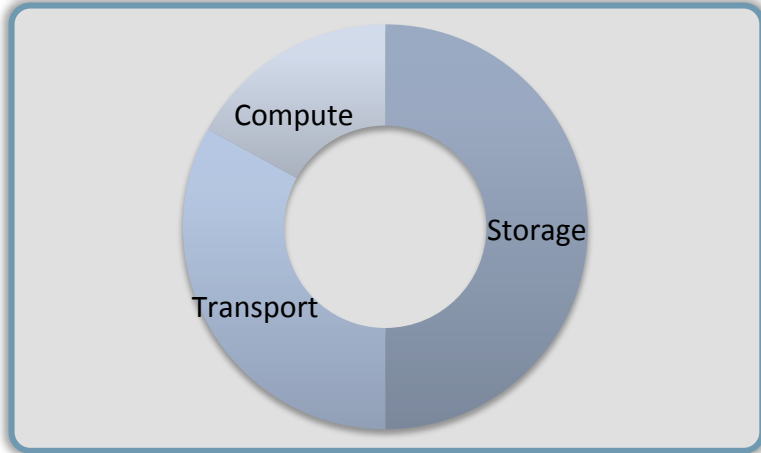
Analytics & Reporting Data Ingestion Fabric

Data Feeds



Challenge: Big Data Economics

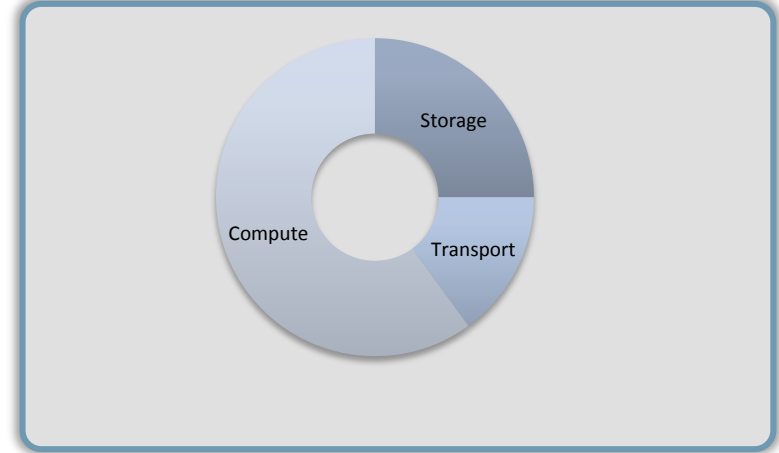
Traditional Solutions, when adopted for Big Network Data



Centralized, Store-First Architecture

- Consolidate data in a repository to analyze
- To answer a question, first need to transport and store the data
- Transport and storage costs alone may put it over budget
- Project may not even get started

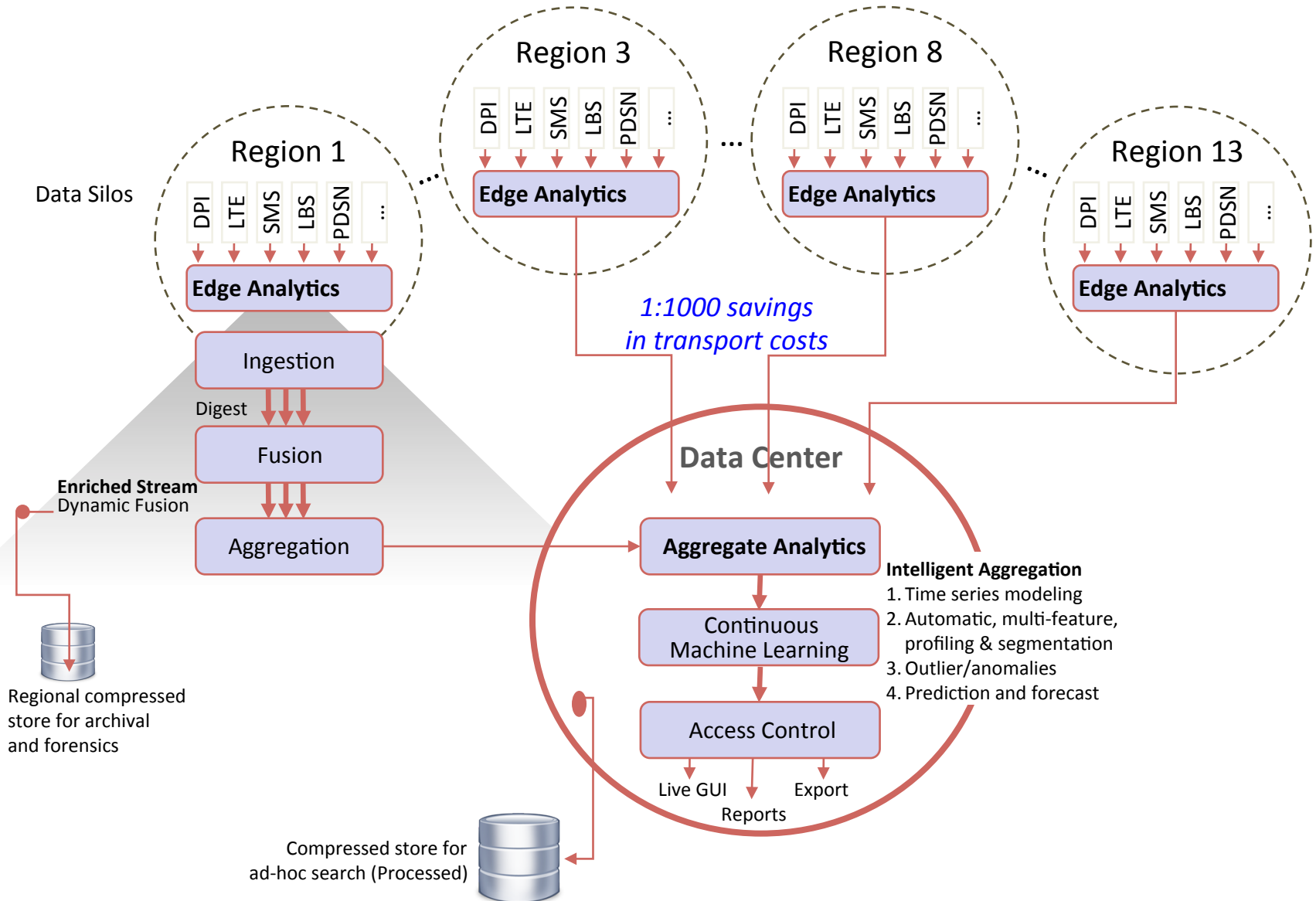
Solutions Purpose-Built for Big Network Data



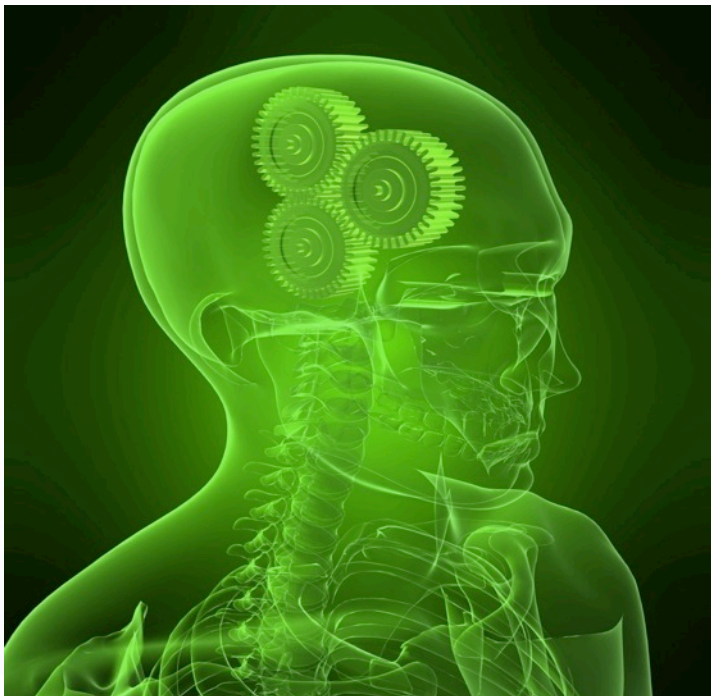
Distributed, Compute-First Architecture

- Move processing to data edge
- Spend on analytics first, then on others, by not transporting and storing data you do not need
- Continuous processing yields timely insights & actions
- Reduce overall spend per new analytics question

An Example Distributed Deployment



Decisioning Application: Network, Device, Apps & Subscribers Insights



What is the next iPhone going do to my network? What applications are causing most load? Which locations in network are most congested?

Business Problem

- How do I plan for a new smartphone or Tablet or M2M device being introduced?
- Which applications are using most bandwidth, tonnage or flows?
- What parts of my network are most congested?

Insights Provided

- Insights on Network, Device, Subscribers and applications
- Most popular devices or application in a location
- Traffic profile in busy hour
- Top N subscriber usage patterns

Benefits

- Better network planning
- Target offending users, devices and applications
- Promotions and campaigns based on usage data

- All
- Smart Phones
- Tablets
- iPad 2
- Blackberry
- Motorola
- RAZR

Devices - All

BY TONNAGE

HTC Incredible ADR 6300 LG CU915

Apple 3G iPhone

Samsung Rant SPH-M540

LG Dare VX97000 LG VX 9100

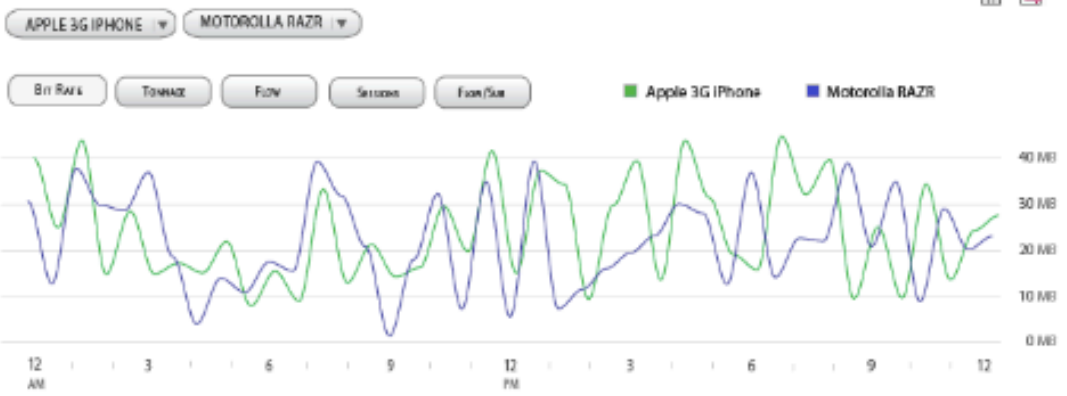
Motorola RAZR

LG Voyager VX HTC Whitestone

Top Categories - Apple 3G iPhone

Rank	Categories	Subscribers	Rate (bps)	Tonnage (MB)	Flows (fps)
1	Social Networking	531	31	1.19 MB	120 K
2	Information & Services Portal	519	19	1.02 MB	141 K
3	Adult Entertainment	458	15	960 K	122 K
4	Mobile Apps & Entertainment	417	17	922 K	109 K
5	Vides	386	16	915 K	99 K
6	Advertising	325	25	889 K	98 K
7	Travel & Weather	312	12	826 K	70 K
8	Business - Rendering Services	283	13	818 K	62 K
9	News & Entertainment	226	16	802 K	12 K
10	Commerce	192	12	790 K	12 K
TOTAL		4,333	173	32.1MB	845 K
OTHERS		1,562	562	3.7MB	1.34 M
GRAND TOTAL		5,895	1,735	35.8MB	2.19 M

Comparison of Usage Segments



Usage Statistics

Apple 3G iPhone Motorola RAZR

Subscribers
7531 5919

Rate (bps)
31 35

Duration (seconds)
51 63

Concurrent Flows (fps)
7 8

Tonnage (MB)
1.6 1.7

Flow (Bytes)
82 91

DOWNLOAD UPLINK

Avg File Trend

Decisioning Application: Visibility for Customer Care



Customer calls have changed from focused on Voice to Data usage issues—No visibility end-2-end on user data sessions

Business Problem

- Customer care is blind to data usage
- Amount of information from network is overwhelming
- Customer need better explanation on how their usage limits were used
- No co-relation across multiple silos

Insights Provided

- Customer interaction with SP call center
- Subscribers usage patterns

Benefits

- Self service portal for users to help them understand the usage in a limited data plans.
- Pro-active identification of (data) network performance issues and changes
- Customer retention

Time Range: 1 Feb, 2011, 0000 hrs > 1 Feb, 2011, 0100 hrs

quick time ranges: Last 30 Days | Last 7 Days | Yesterday | Today | Most Recent Hour | Custom

Search Subscriber

SUBSCRIBER: 9825738056@mytelco.com	UPLOAD	5.2M	DOWNLOAD	20.6M
	TRANSACTIONS	190	DEVICE	iPhone 3GS
	HTTP ERRORS	6.4%	TOP DOMAIN	facebook.com

Search URL History

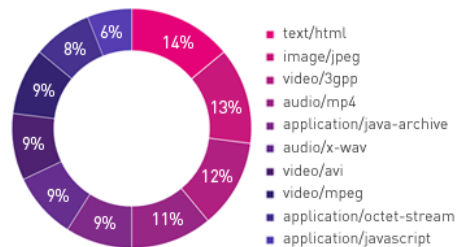
Search URL	Start Time	End Time	Response Code	MIME-Type	BYTES
URL	TIME STAMP		RESPONSE CODE	MIME-TYPE	
facebook.com/profile.php	Tue 01 Feb 2011 00:03:29 GMT		200	text/html	50
facebook.com/home.php	Tue 01 Feb 2011 00:04:19 GMT		200	text/html	49
google.com	Tue 01 Feb 2011 00:05:29 GMT		200	text/html	49
mail.google.com	Tue 01 Feb 2011 00:10:10 GMT		200	application/octet-stream	48
twitter.com/home	Tue 01 Feb 2011 00:10:49 GMT		404	application/octet-stream	1
cnn.com/news	Tue 01 Feb 2011 00:14:20 GMT		200	application/octet-stream	36
facebook.com/notifications.php	Tue 01 Feb 2011 00:15:29 GMT		200	text/html	27
maps.google.com	Tue 01 Feb 2011 00:16:10 GMT		200	application/octet-stream	45

Page 1 of 2 >>

Top Domains

Rank	Service Provider	Transactions	Volumes
1	facebook.com	9086	125.2M
2	gmail.com	8326	124.1M
3	google.com	7650	113.6M
4	twitter.com	7034	114.3M
5	cnn.com	6321	103.9M
6	yahoo.com	5686	93.2M
7	huffingtonpost.com	4945	98.5M
8	shopping.com	4204	67.1M
9	digg.com	3448	74.9M
10	nytimes.com	2825	73.2M

Top MIME-Types



HTTP Attributes

HTTP Attributes	Counter
HTTP Requests	59972
HTTP Responses	59586
HTTP 2XX	93.84
HTTP 3XX	3.18
HTTP 4XX	2.58
HTTP 5XX	0.4

Customer Care
Detailed Usage Troubleshooting

Opportunities with Big Data

