



11° Jornadas Data Mining & Business Intelligence

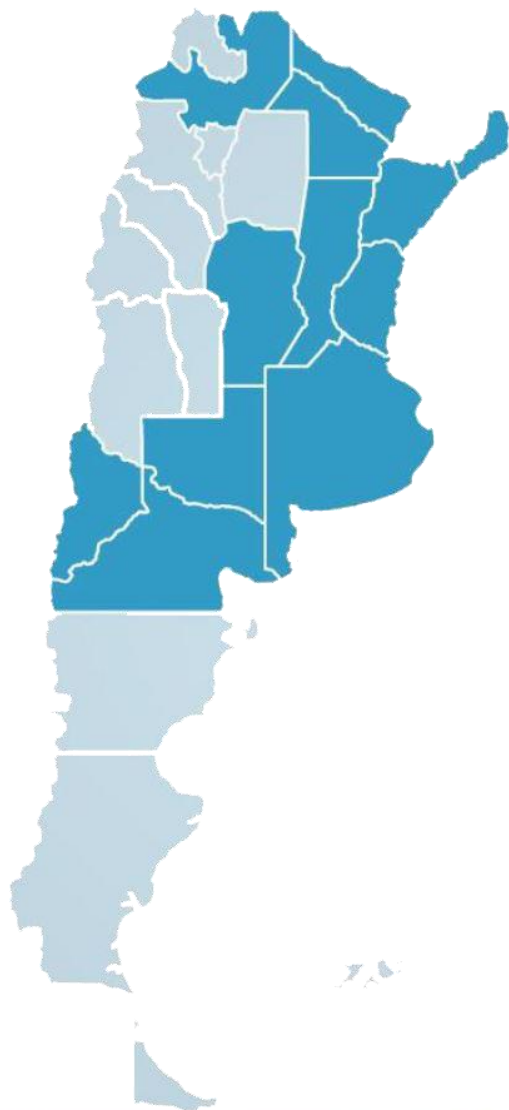
-
- ***Ecosistema Analítico en Cablevisión.***
Un caso de uso: Analytics de Redes
-

Alejandro Ciotti – *Jefe de BI & Analytics en Cablevisión - Fibertel*

Claudio Righetti – *Chief Scientist en Cablevisión – Fibertel*

Cablevisión

Fibertel



11MM Hogares País (Censo 2011 INDEC)

8MM Hogares bajo nuestra Red

4MM de Clientes Únicos

3,8MM TV Paga / 2MM Internet

¿Buscaron Big Data Landscape?



Big Data Landscape 2016



Buscar con Google

Me siento con suerte

Google.com.ar ofrecido en: [Español \(Latinoamérica\)](#)

Big Data Landscape 2016 (Version 3.0)

Infrastructure

Hadoop On-Premise
 cloudera, Hortonworks, MWRP, Pivotal, IBM InfoSphere, bluedata, jethro

Hadoop in the Cloud
 Amazon Web Services, Microsoft Azure, Google Cloud Platform, IBM InfoSphere, CAZENA, altiscale, Qubole

Spark
 databricks, GridGain, TACHYON NEXUS

Cluster Services
 Amazon Web Services, Kubernetes, HPCC SYSTEMS, docker, MESOSPHERE, CoreOS, pepperdata, StackIQ

Analytics

Analyst Platforms
 Palantir, AYASDI, Quid, enigma, Digital Reasoning, ORBITAL INSIGHT

Analytics Platforms
 Microsoft, GUAVUS, Datameer, Bottlenose, interlana

Data Science Platforms
 context relevant, CONTINUUM ANALYTICS, DataRobot, Alpine, MODE, Plotly, dataiku, tonian, DOMINO, sense, yhat, ALGORITHMIA

Visualization
 Google Cloud Platform, Tableau, Qlik, Looker, Roambi, Sisense, Qeodata, datarama, CHARTIO

Applications

Sales & Marketing
 RADIUS, Gainsight, bloomreach, Zeta, EVERSTRING, livefyre, blueyonder, Lattice, kahuna, infer, SAILTHRU, persado, AVISO, sense, QUANTIFIND, ACTIONIQ, fuse|machines, ENGA G I O

Customer Service
 MEDALLIA, ATTENITY, CLARIBRIDGE, CLEARFOX, STELLA Service, NGDATA, Preact, DigitalGenius, appuri, Wiseio

Human Capital
 gild, Connectifier, textic, entelo, hiQ

Legal
 RAVEL, JUDICATA, Everlaw, Brevia, PREMOTION

NoSQL Databases
 Amazon DynamoDB, Google Cloud Platform, Microsoft Azure, MarkLogic, mongoDB, DATASTAX, EROSPIKE, Couchbase, SequoiaDB, redislabs, influxdata

NewSQL Databases
 SAP HANA, Clustrix, Pivotal, paradigm4, nuODB, memsql, VOLTDB, splice MACHINE, MariaDB, citusdata, deepdb, Trafoion, Cockroach LABS

BI Platforms
 Power BI, Amazon Web Services, Domo, Wave Analytics, GoodData, birst, kyvos insights, platforma, atscale, ARCADIA, SISENSE

Statistical Computing
 SAS, SPSS, MATLAB

Log Analytics
 Splunk, sumologic, kibana, Cloud Physics, loggly

Social Analytics
 Hootsuite, Netbase, DataSift, tracr, bitly, synthesio, simplereach

Ad Optimization
 AppNexus, Criteo, MediaMath, OpenX, rocketfuel, Integral, theTradeDesk, Adgorithms, dstillery, Livelihood, TAPAD, DataXu, Cppier, MOAT

Security
 Cylance, CounterTack, cyberreason, ThreatMetrix, AREA 1 SECURITY, SentinelOne, Recorded Future, Guardian Analytics, FORTSCALE, siftscience, Keybase, feedzai, SICNIFYD

Vertical AI Applications
 Facebook, Clara, KASIST, lumiata

Graph Databases
 neo4j, Graph, OrientDB, InfiniteGraph

MPP Databases
 Teradata, Vertica, Netezza, Action, Kognitio, BASOL, dremio

Cloud EDW
 Amazon Web Services, Microsoft Azure, Pivotal, snowflake, WATERLINE DATA, Infoworks

Data Transformation
 Alteryx, TRIFACTA, tamr, StreamSets, Alation

Data Integration
 Informatica, MuleSoft, snapLogic, BedrockData, xplenty

Real-Time
 Amazon Web Services, METAMARKETS, striim, confluent, DATATORRENT, dataArtisans

Machine Learning
 Azure Machine Learning, H2O, Amazon Web Services, SKYTREE, Dato, rapidminer, DATARPM, deepsense.io, VISENZE, PredictionIO, glowfish

Speech & NLP
 NarrativeScience, NUANCE, WolframAlpha, semanticmachines, Gridspace, ARRIA, api.ai, cortico.io, malulba, MindMeld, IDIBON, YSEOP

Horizontal AI
 IBM Watson, Cortana, sentient, viv, nervana, vicarious, nara, Numenta, HyperSense, SI, Inferrence, Descartes Labs, clarifai, DEXTER, MetaMind

Publisher Tools
 Outbrain, Taboola, quantcast, Chartbeat, yieldbot, Yieldmo

Govt / Regulation
 Socrata, OPENGOV, FN, FiscalNote, enigma, PREPPO, mark43, OpenDataSoft

Finance
 Affirm, LendingClub, OnDeck, Kreditech, zest finance, LendUp, Kabbage, tdemark, Puff, INSIKT, ZUORA, Dataminr, Lenddo, KENSHO, AIDYIA, ISENTIUM, Quantopian, sentient

Management / Monitoring
 New Relic, APPDYNAMICS, Amazon Web Services, actifio, Numerify, splunk, DATADOG, DRIVEN, Anodot

Security
 Tanium, illumio, CODE42, DataGravity, CipherCloud, VECTRA, sqrrl, BlueTalon

Storage
 Amazon Web Services, Google Cloud Platform, Microsoft Azure, Panasas, nimblestorage, COHO DATA, Qumulo

App Dev
 Apigee, CASK, Typesafe, DRIVEN

Crowd-sourcing
 Amazon Mechanical Turk, CrowdFlower, WorkFusion

Search
 HP, Oracle, ENDECA, EXALEAD, Lucidworks, elastic, ThoughtSpot, MAANA, swifttype, Algolia, SINEQUA

Data Services
 OPERA, Mu Sigma, EXL, DATA SCIENCE, kaggle, datascopie, DataKind

For Business Analysts
 OrigamiLogic, ClearStory, CIRRO, import.io

Web / Mobile / Commerce
 Google Analytics, mixpanel, R.JMetrics, BLUECORE, AMPITUDE, granify, sumall, Airtable, retention, CUSTORA

Education / Learning
 Knewton, Clever, Declara, PANORAMA, knowre

Life Sciences
 23andMe, Counsyl, RECOMBINE, KYRUUS, FLATIRON, zymogen, HealthTap, METABIOTA, ZEPHYR HEALTH, ovia, Ginger.io, transcriptic, Glow, enlitic, AiCure, Atomwise

Industries
 OP@WER, eHarmony, RetailNext, STITCH FIX, duetto, WorkFusion, BLUE@RIVER, TACHYON, SwiftKey, Seeq, FarmLogs, HowGood, select, NIGHT MACHINE, statmuse, BOEYER

Cross-Infrastructure/Analytics

Amazon Web Services, Google, Microsoft, IBM, SAP, SAS, data, hp, Autonomy, VERTICA, vmware, TIBCO, TERADATA, ORACLE, NetApp

Open Source

Framework
 Hadoop, HADOOP MapReduce, YARN, Spark, MESOS, TEZ, Flink, CDAP

Query / Data Flow
 SLAMDATA, HIVE, APACHE DRILL, Google Cloud Dataflow

Data Access
 Apache, ACCUMULO, HBASE, mongoDB, cassandra, CouchDB, riak, SCiDB, OPENTSOB, nifi

Coordination
 talend, Apache Zookeeper, Apache Ambari

Real-Time
 STORM, Spark, APEX, Flink, TACHYON, druid

Stat Tools
 ScalaLab, NumPy, SciPy

Machine Learning
 mlilb, Aerosolve, Apache, SINGA, MADlib, Caffe, FeatureFu, VELES, WEKA, DIMSUM, jupyter, DL4J

Search
 elasticsearch, Solr, Lucene

Security
 Apache Ranger

Visualization
 Zepplin

Data Sources & APIs

Health
 Apple, JAWBONE, GARMIN, practice fusion, fitbit, Withings, VALIDIC, netatmo, kinsa, Human API

IOT
 UPTAKE, ThingWorx, helium, samsara, AUGURY, estimate

Financial & Economic Data
 Bloomberg, DOW JONES, THOMSON REUTERS, S&P CAPITAL IQ, YODLEE, PREMISE, S&P, quandl, xignite, CBINSIGHTS, mattermark, StockTwits, estimate, PLAID

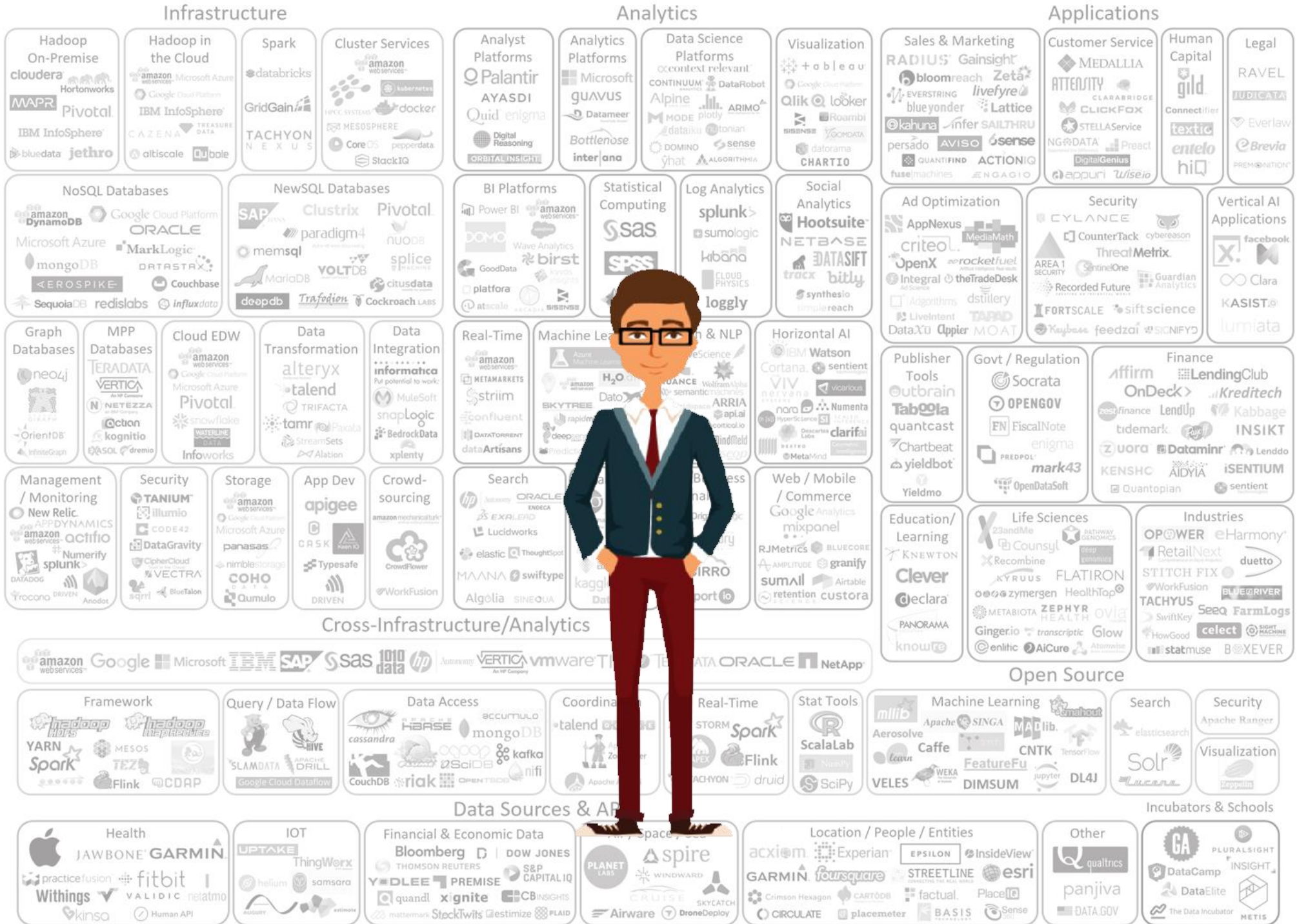
Air / Space / Sea
 PLANET LABS, spire, WINDWARD, CRUISE, SKY CATCH, Airware, DroneDeploy

Location / People / Entities
 acxiom, Experian, EPSILON, InsideView, GARMIN, foursquare, STREETLINE, esri, Crimson Hexagon, CARTODB, factual, PlaceIQ, CIRCULATE, placemeter, BASIS, Sense

Other
 qualtrics, panjiva, DATA.GOV

Incubators & Schools
 GA, PLURALSIGHT, DataCamp, INSIGHT, DataElite, The Data Incubator, METIS

Big Data Landscape 2016 (Version 3.0)



Big Data Landscape 2016 (Version 3.0)

Infrastructure

Analytics

Applications

Estadística

- Modelado predictivo
- Machine Learning
- Diseño de Experimento
- Bayes
- Supervisado / No supervisado
- Optimización

IT Skills

- Scripting (Python)
- Programación R
- SQL / No SQL
- Procesamiento paralelo
- MapReduce
- Hive / Pig



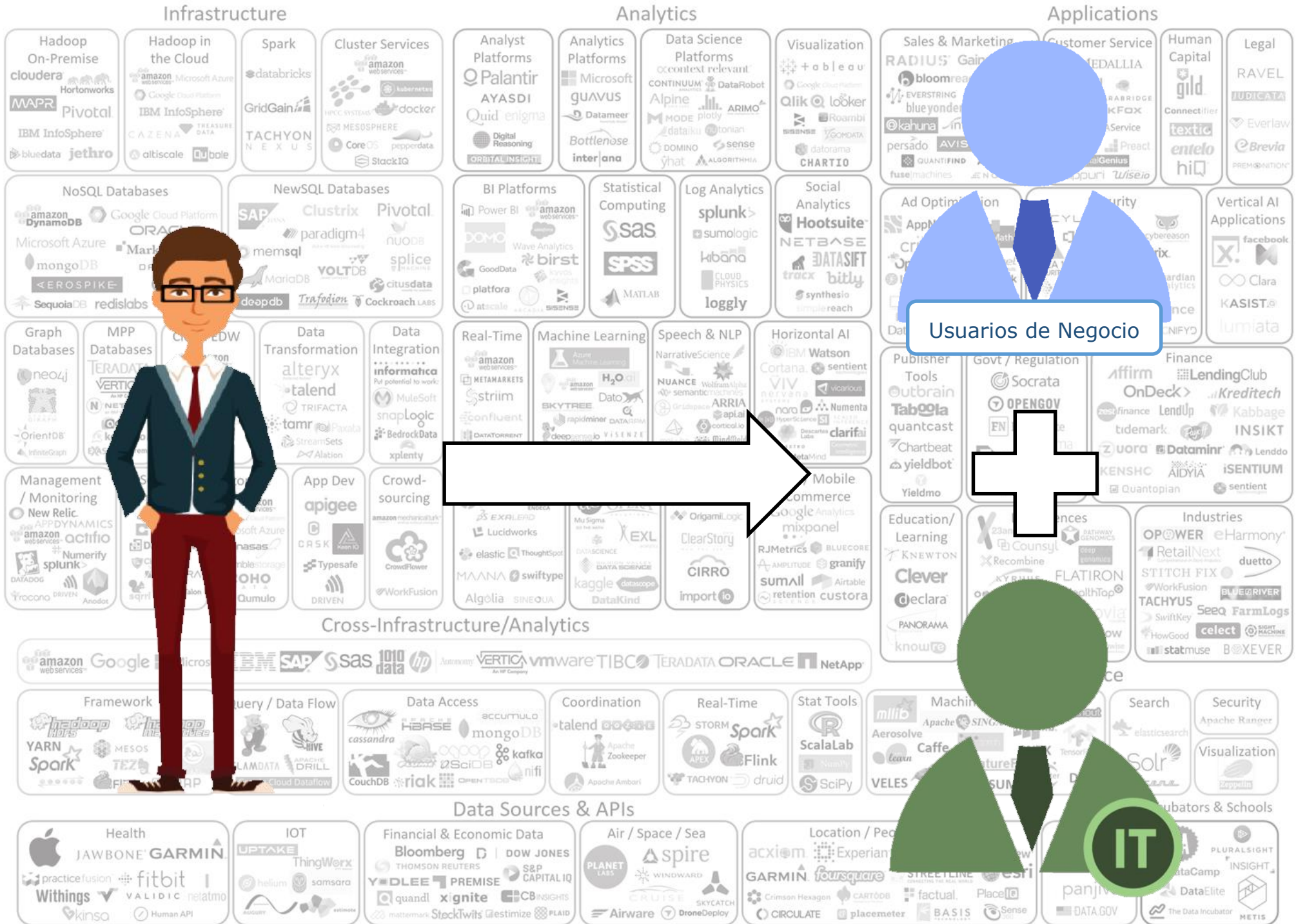
Negocio

- Estrategia de Negocio
- Valor en los Datos
- Resolver problemas
- Innovador
- Colaborador
- Liderazgo

Comunicación

- Visualización
- Transmisión a C - Level Mng
- Comunicar una idea
- Convertir data-driven en acción
- Traducir estadística a resultados de negocio

Big Data Landscape 2016 (Version 3.0)



EQUIPOS ANALYTICS CABLEVISION



Datamining & BI

Comercial



Soporte Científico y Seguridad

Técnica



Consumer Insights

Comercial



Performance

Técnica



Control de Gestión

Administración & Finanzas



BI & Analytics

IT

- Modelado predictivo
- Machine Learning
- Diseño de Experimento
- Bayes
- Supervisado / No supervisado
- Optimización
- Scripting (Python)
- Programación R
- SQL / No SQL
- Procesamiento paralelo
- MapReduce
- Hive / Pig
- Estrategia de Negocio
- Valor en los Datos
- Resolver problemas
- Innovador
- Colaborador
- Liderazgo
- Visualización
- Transmisión a C – Level Mng
- Comunicar una idea
- Convertir data - driven en acción
- Estadística a resultados de negocio

ECOSISTEMA ANALÍTICO



APLICACIONES

Business Intelligence



Planning & Forecasting



Analytics & Datamining



INFRAESTRUCUTRA

Databases



Hadoop



ETL - Réplica



DATASOURCES



TERADATA.

EDWH; CRM; Billing;
Order Manager; GIS;
WorkForce

+ 700M Millones Registros
+ 3M Millones Diarios

 **mongoDB.**

Monitoreo de Red;
Eventos y Reclamos

+ 1 Millón Eventos Diarios
Real Time

 **neo4j**

Tipología de Red

Actualización Diaria
16 Instancias Concurrentes

 **hadoop**

Logs Redes;
Comportamiento; Pools

+ 700M Millones Registros
+ 3M Millones Diarios

ORACLE®

Recaudadoras; IVR; otras

+ 10GB Información
actualización Batch

ALGUNOS EJEMPLOS



```
{
  "name": "CM:DOWN",
  "eventId": "fd7ff1a2-b33e-4523-9c66-c14770afc203",
  "eventIdTrad": "100072807",
  "eventIdGen": "SiGMA",
  "deviceType": "CM",
  "deviceId": "00214315F1CA",
  "origin": "SiGMA",
  "originId": "SiGMA",
  "state": null,
  "cause": "",
  "creationDate": 1469548026000,
  "arrivalDate": 1469548099000,
  "processState": "Injected",
}
```

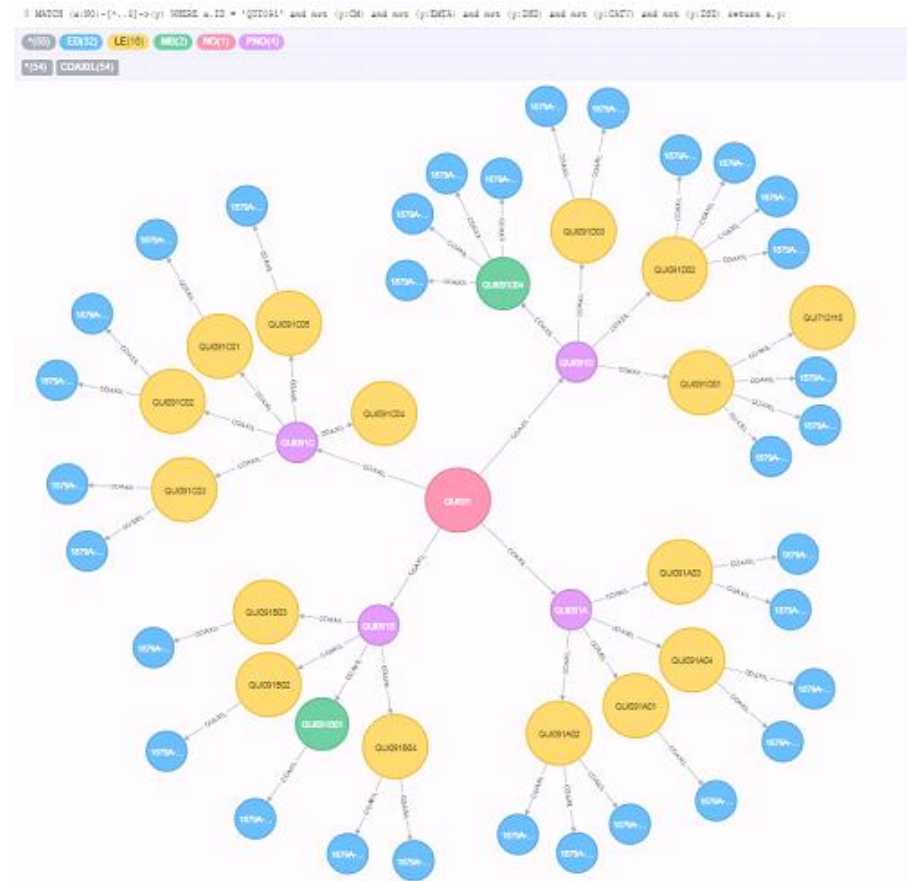
Caso

Manejo de Fallas

Se correlacionan los eventos ingestados en MongoDB sobre la Red diseñada en Neo4J.

Práctica Analítica

Se realiza Datamining sobre la información para detectar reglas.



ALGUNOS EJEMPLOS



TERADATA

```
PROC SQL NOERRORSTOP;
CONNECT TO TERADATA (USER="p_report_sas"
PASSWORD="p_report_sas"
SERVER=TDPROD DATABASE=P_BIDMT_V MODE=ANSI);
CREATE TABLE WORK.MPC_ANTICUACION_&VAR_NOMBRE. AS
SELECT *
FROM CONNECTION TO TERADATA
/*QUERY TERADATA*/
(SELECT
    CAST(SUBSTR(TRIM(DPP.ID_FECHA_CORTE),1,8) AS CHAR(8)) AS
ID_FECHA,
    DPP.ID_ANTICUACION_MORA,
    AM.ANTICUACION AS ANTICUACION_MORA,
    SUM(DPP.IMPORTE_SALDO) AS IMPORTE_SALDO,
    DPP.ID_SUSCRIPCION
FROM
    P_BIDMT_V.DM_ANTICUACIONES AM /* DM_ANTICUACIONES_MORA */
,
    P_BIDMT_V.FT_DETALLE_PAGOS_PENDIENTES DPP
WHERE
    DPP.ID_ANTICUACION_MORA = AM.ID_ANTICUACION
    AND DPP.ID_FECHA_CORTE =&VAR_IDCORTE /* FECHA CORTE SEGUN
VARIABLE */
GROUP BY
    DPP.ID_ANTICUACION_MORA;);
QUIT;
```



```
PROC SQL;
CONNECT TO hadoop
(authdomain="HadoopProd"
server='sr-XXX-XXX.XXXX.cablevision.com.ar'
port=9999 schema=sdb_datamining SUBPROTOCOL=hive2);
EXECUTE(drop table sdb_datamining.test2) by hadoop;
;
DISCONNECT FROM hadoop;
QUIT;
```

Caso

Analytics sobre diversos datasources.

Práctica Analítica

Los equipos hacen Discovery sobre los datos integrando distintas fuentes de distintas bases ejecutando consultas IN - DATABASE.

Armado de ABT / Variables.

TIPS COLABORATIVOS



UNIVERSIDAD
AUSTRAL

INGENIERÍA



NEGOCIO

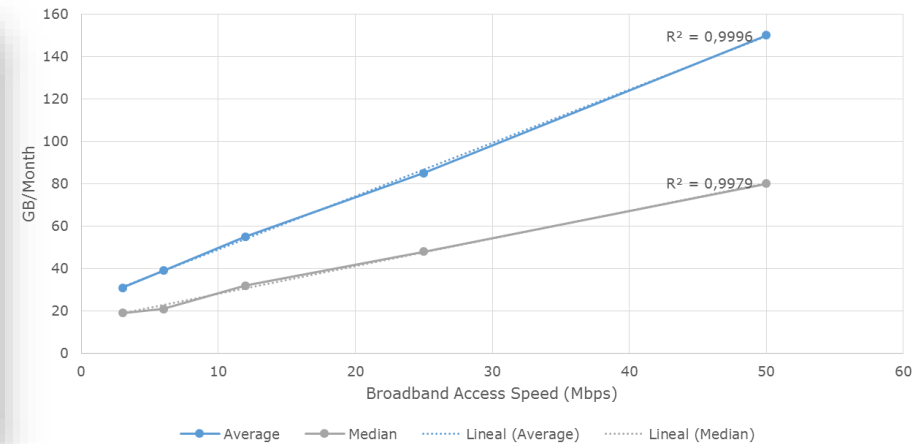
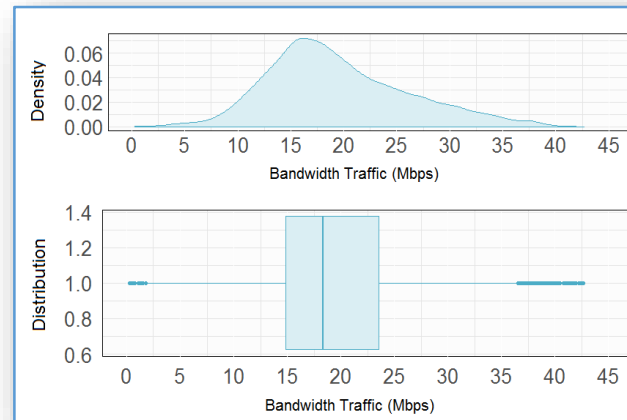
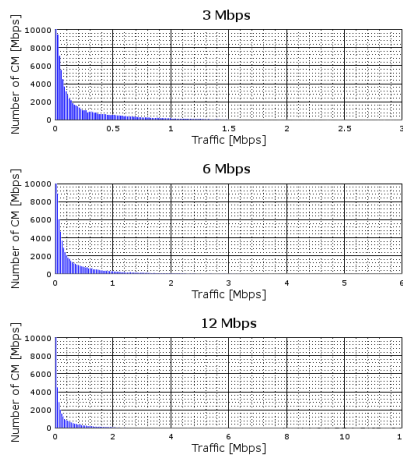
- Compartir desde el inicio
- Atender los lineamientos técnicos
- Definir el alcance
- Comunicar los resultados obtenidos

- Priorizar la agilidad
- Todo se puede
- Comprender el Por Qué
- Orientar los esfuerzos en el objetivo que se persigue



CASO DE USO

Analytics de Redes.

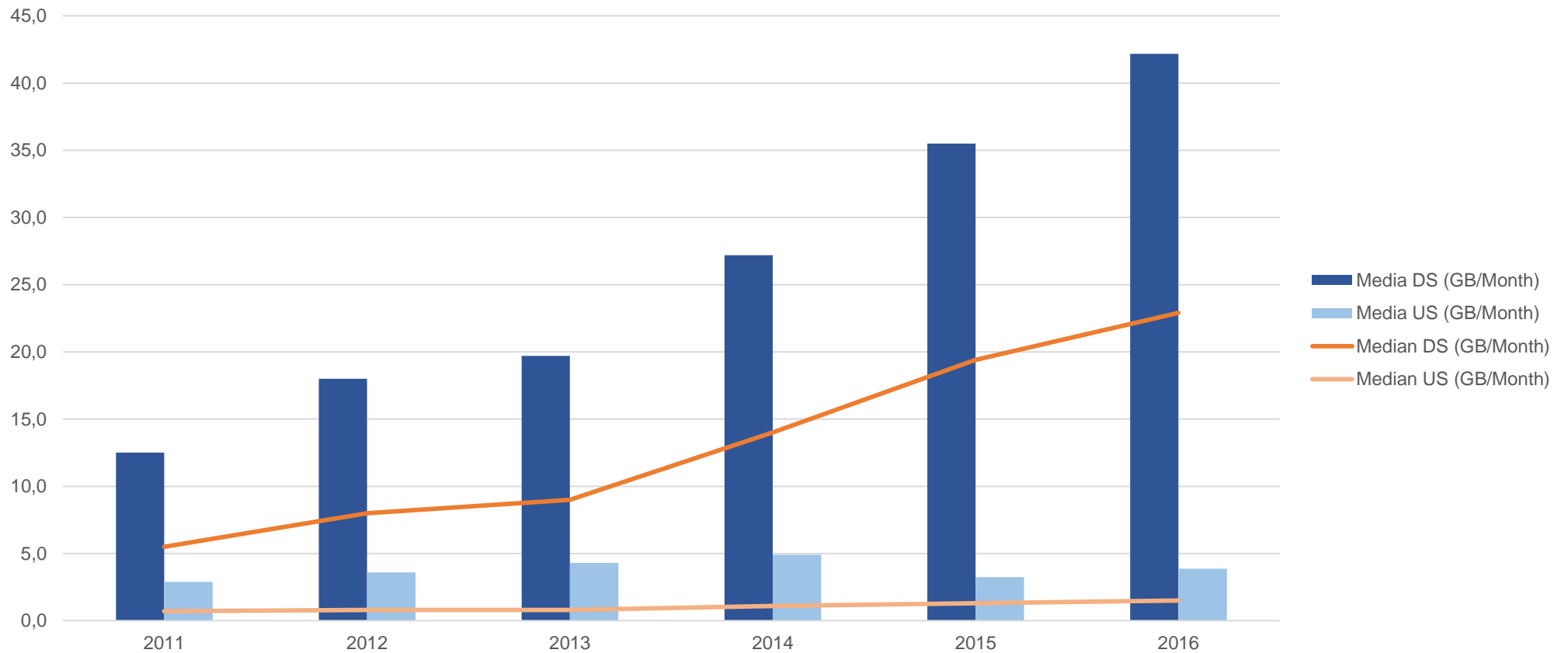


REALIZACION

Claudio Righetti – crighetti@cablevision.com.ar
Chief Scientist en Cablevisión – Fibertel

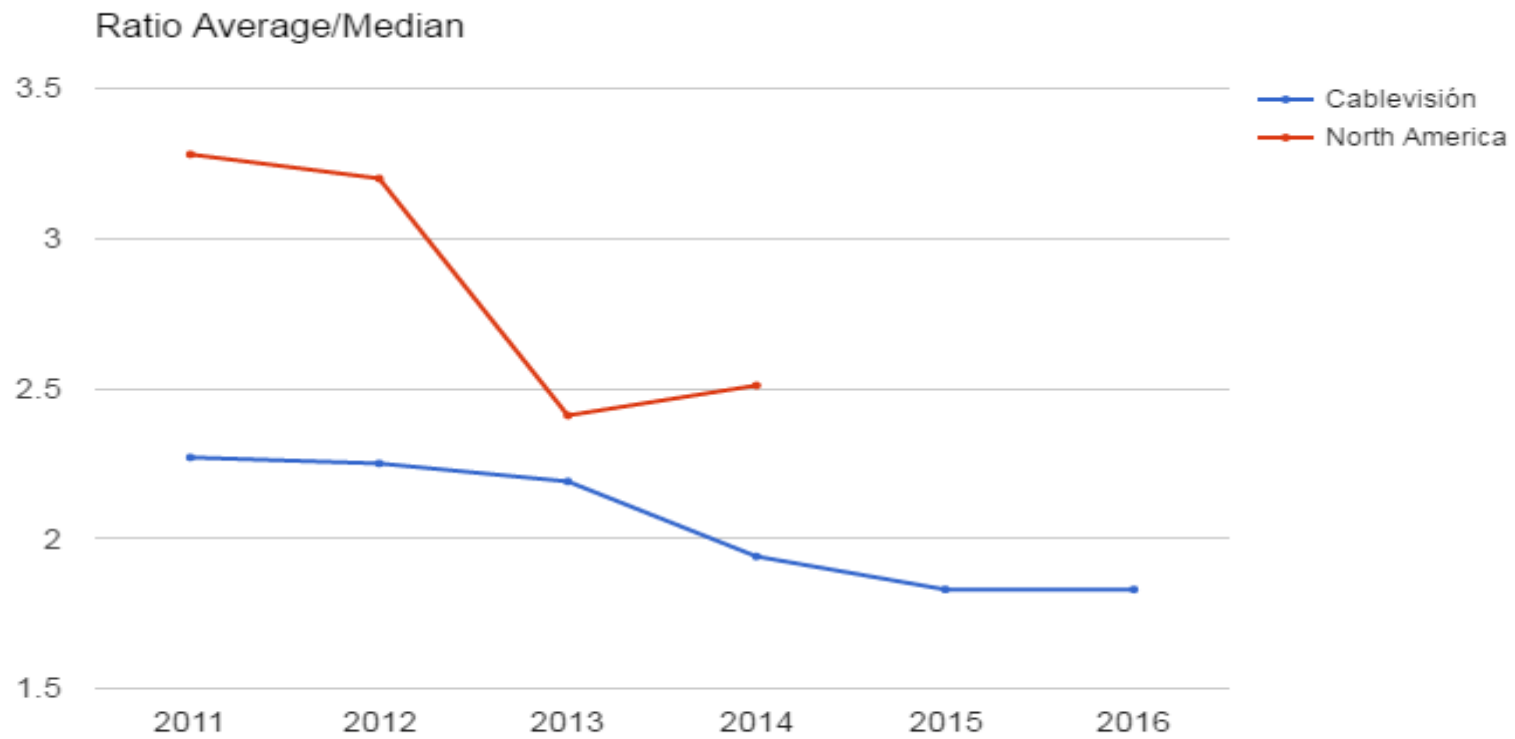


Downstream and Upstream Monthly Consumption Trends





Downstream : Average/Median Trend



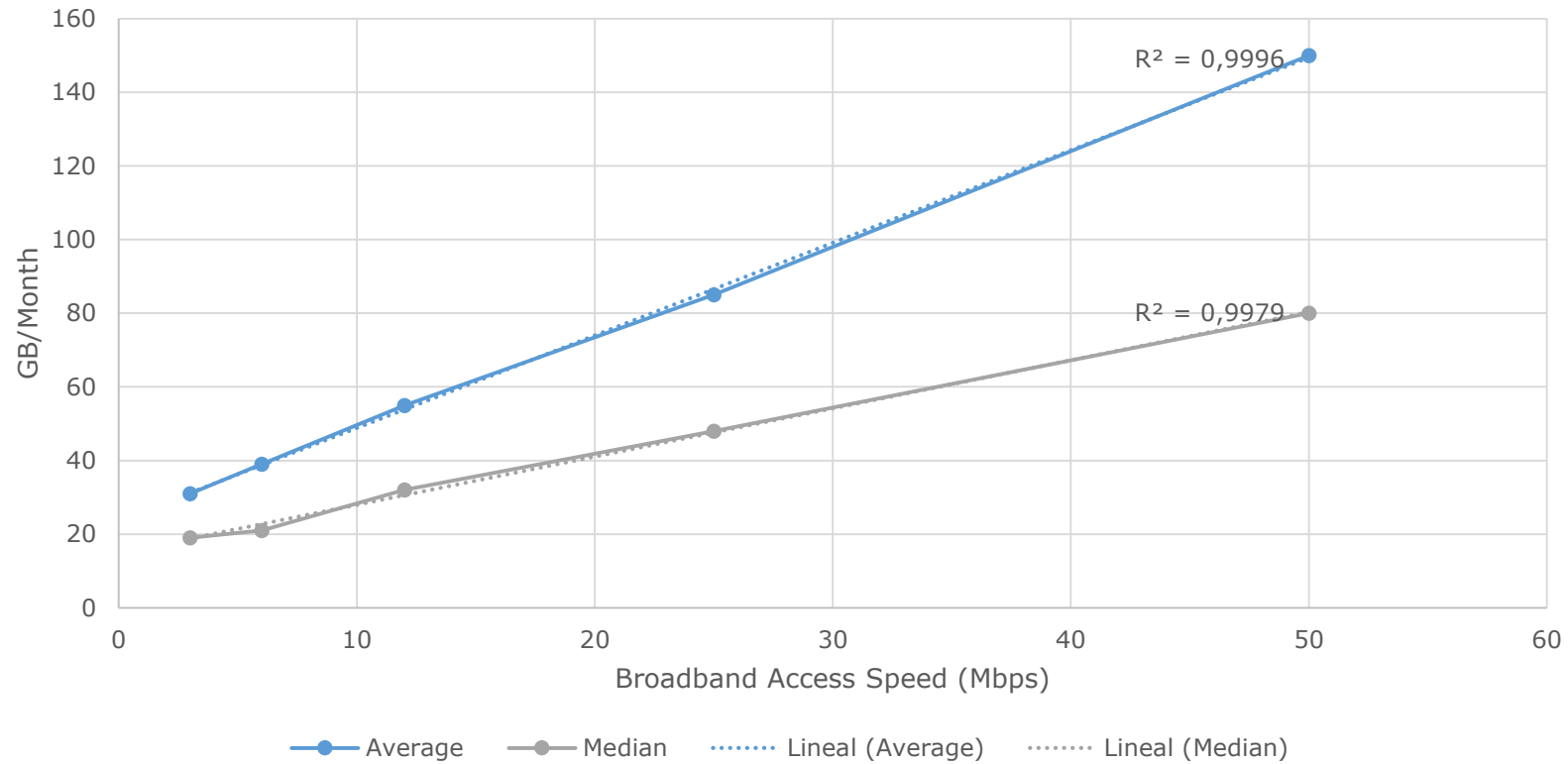
North America Source: Global Internet Phenomena Report - (2011 H1 to 2014 H1) SANDVINE

Downstream Consumption by Tier [GB/Mo]



Tier	Average	Median	10th Pctl	25th Pctl	75th Pctl	90th Pctl
3 Mbps	31	19	2	6	43	77
6 Mbps	38	21	2	7	50	94
12 Mbps	55	32	4	11	72	132
25 Mbps	72	36	2	10	93	180
50 Mbps	148	77	11	32	161	294

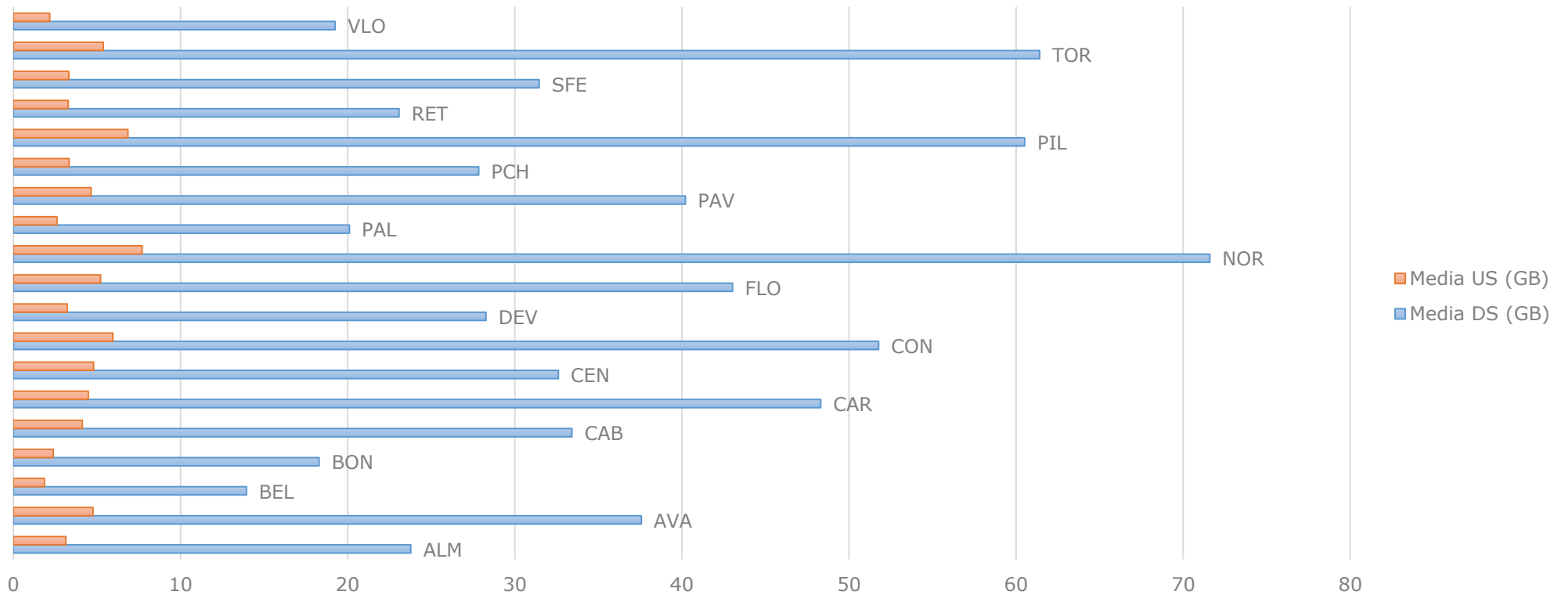
Downstream Consumption by Tier



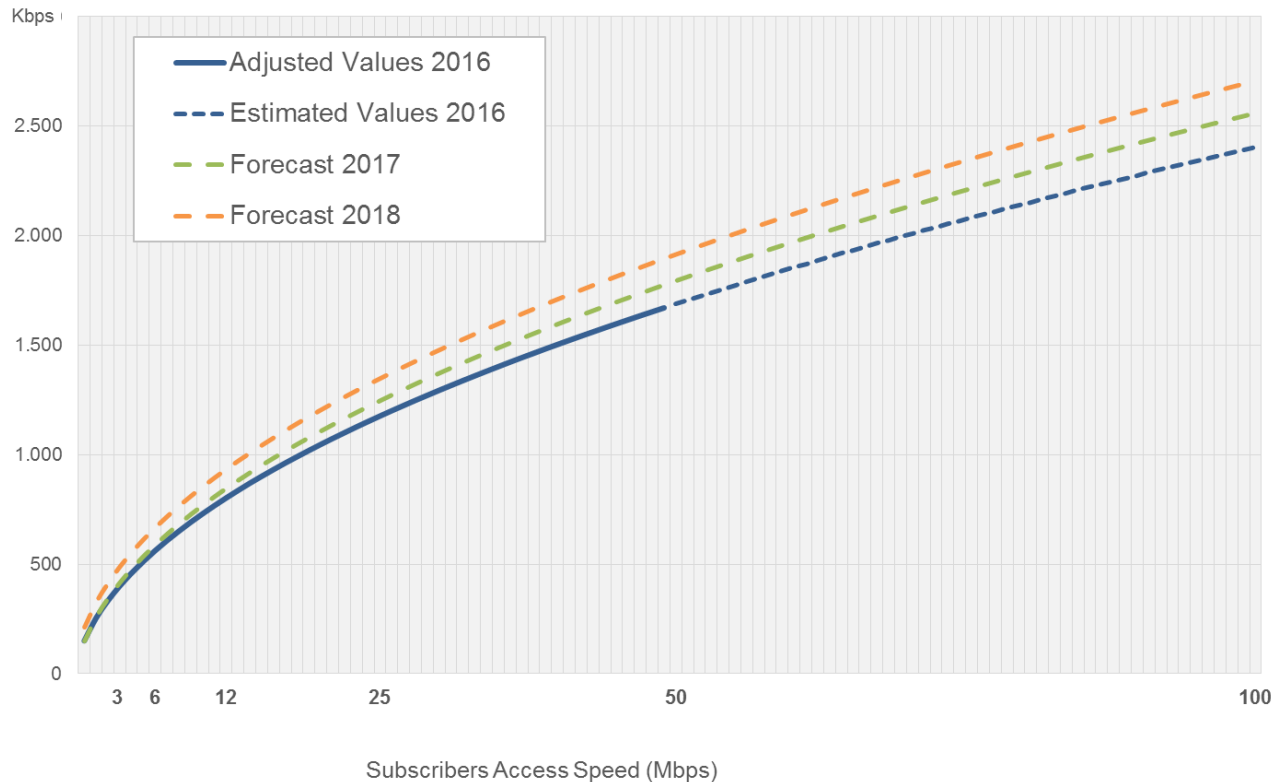


Average Consumption per HUB Site

Monthly Consumption - Buenos Aires



Downstream Peak Utilization per Tier



Objectives

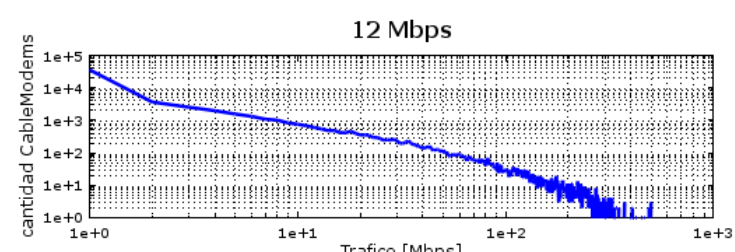
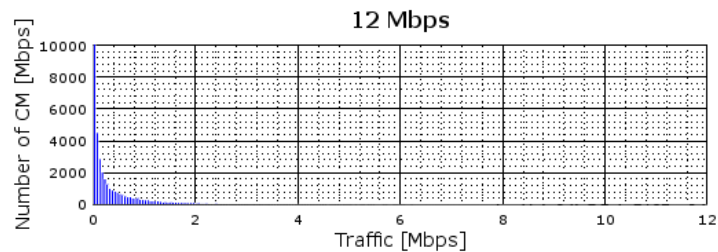
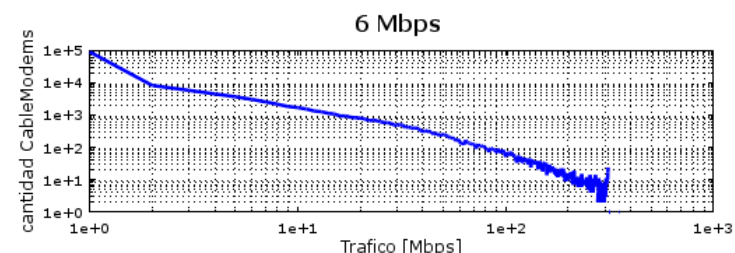
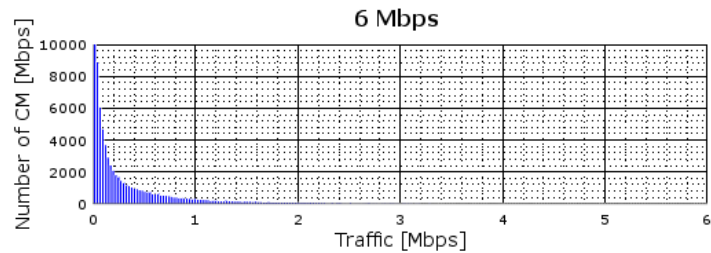
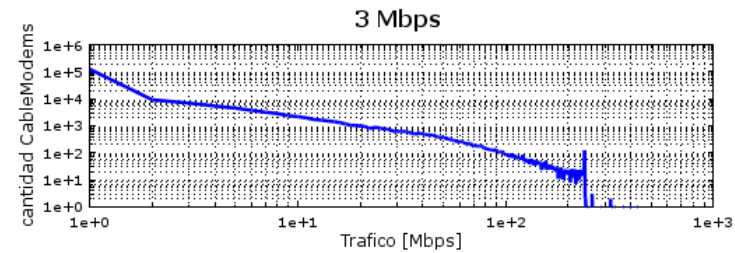
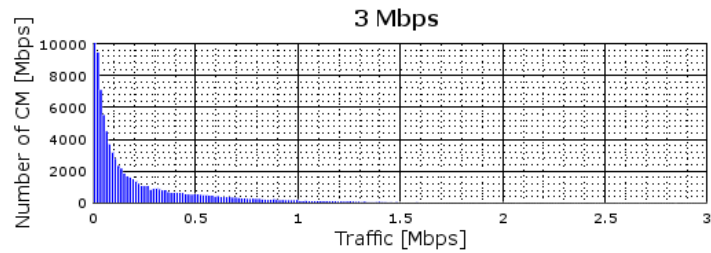
- Adjust current BW traffic per Tier.
- Estimate BW traffic for future access speed (100 Mbps and 500 Mbps).
- Forecast BW traffic.

- Análisis y Dimensionamiento de banda ancha : STEM



UNIVERSIDAD
AUSTRAL

INGENIERÍA

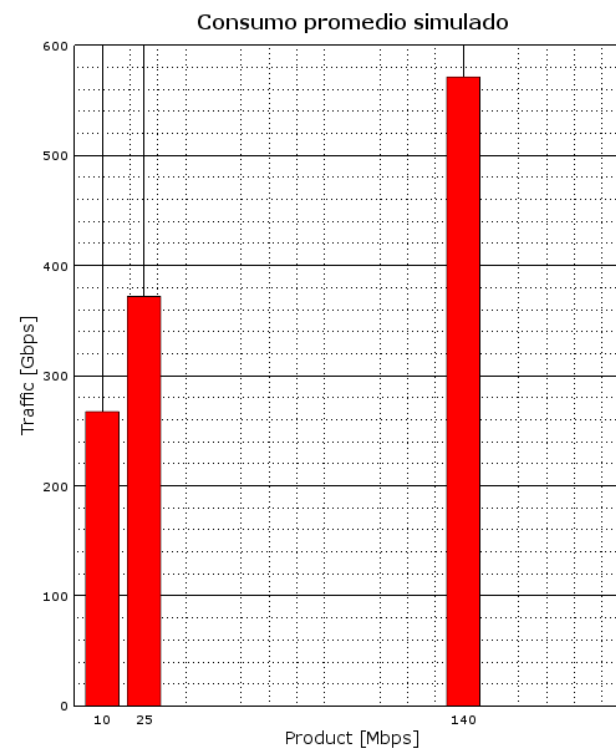
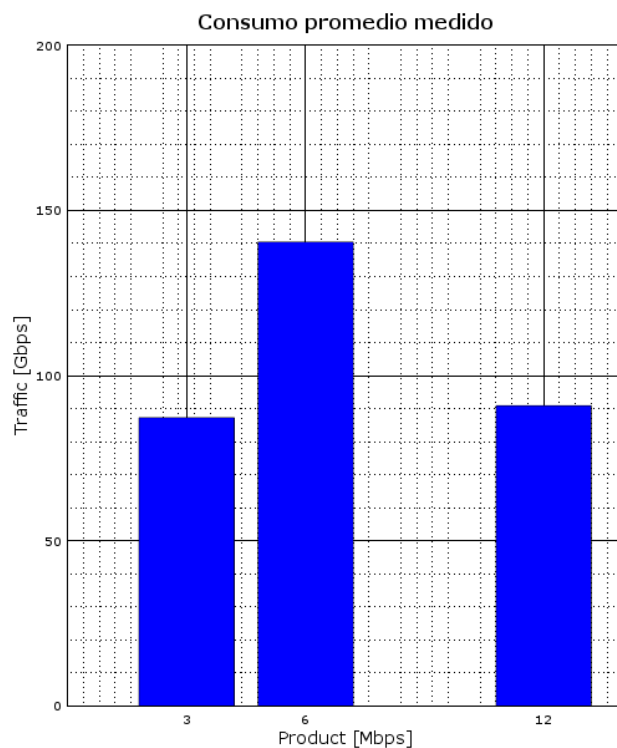


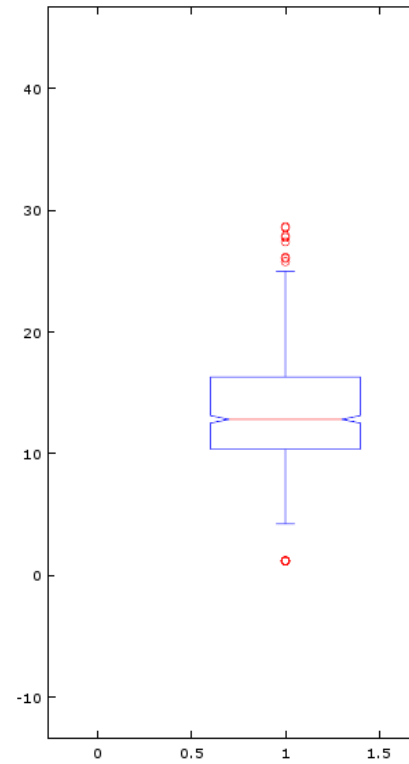
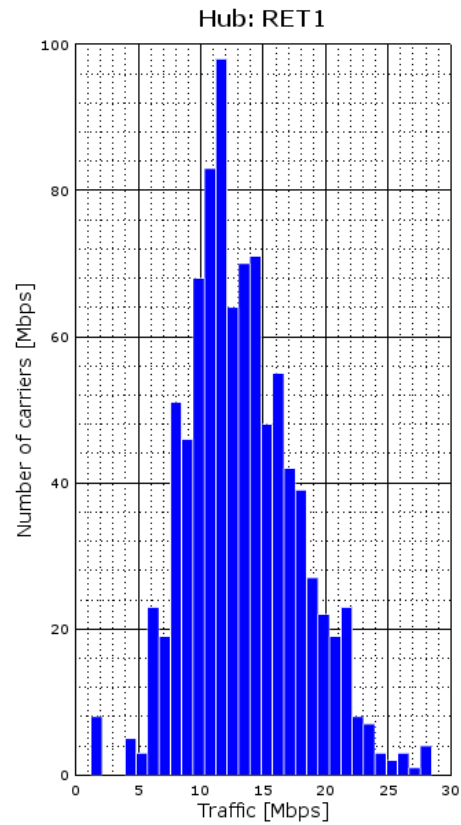
STEM: Simulación Tráfico generado por 10, 25 y 100 Mbps



UNIVERSIDAD
AUSTRAL

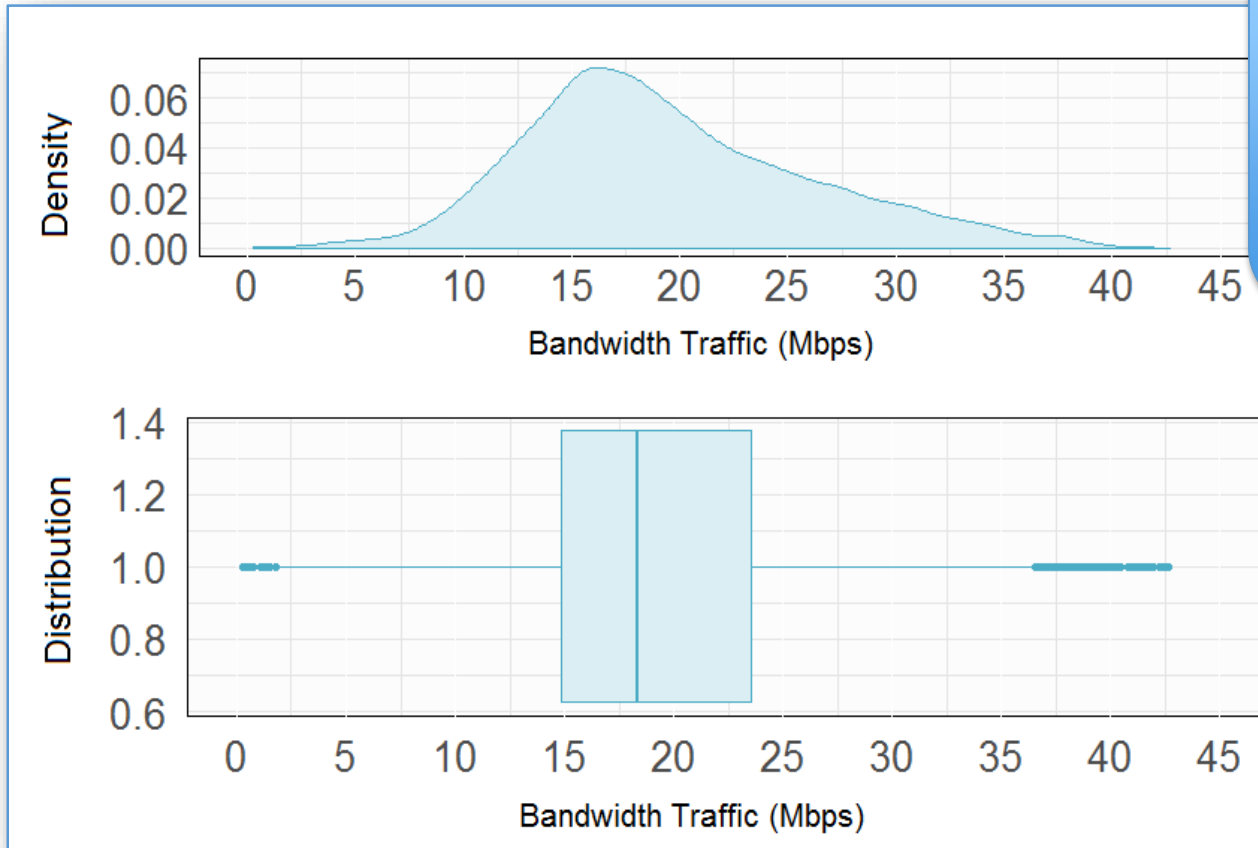
INGENIERÍA





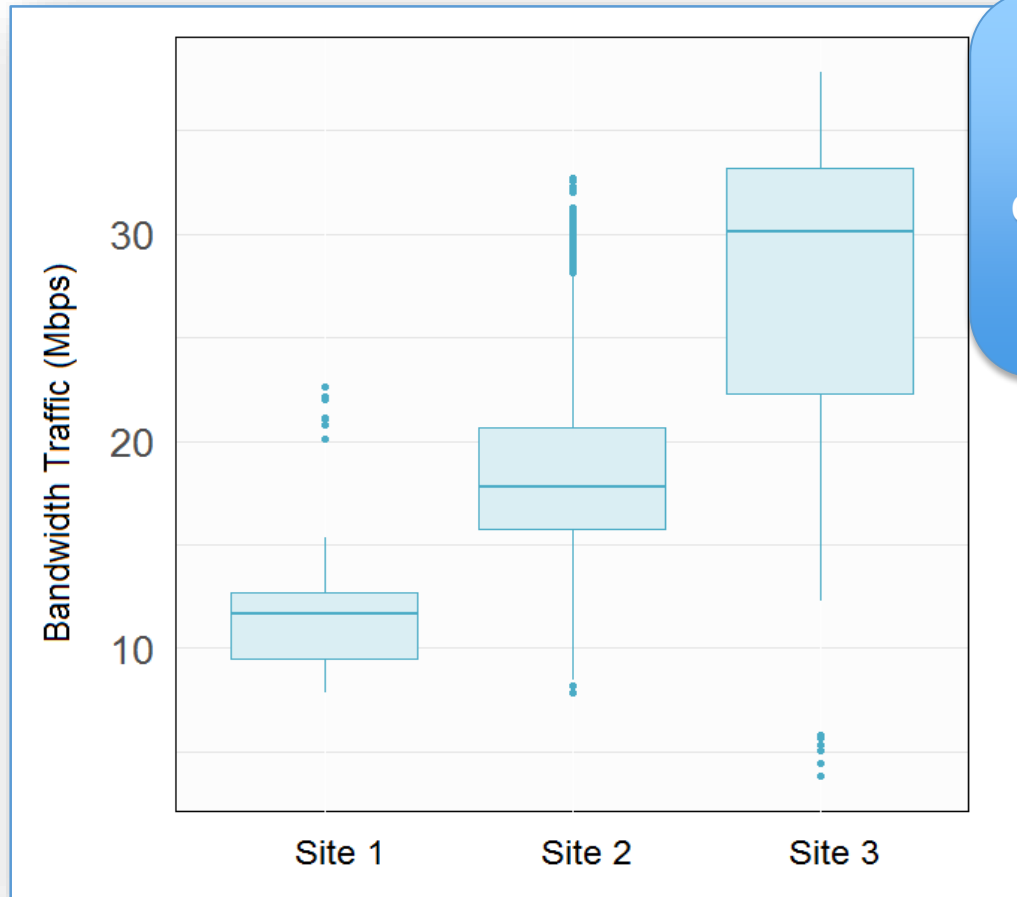
- IP Traffic Analysis: a Tool for Network Dimensioning
- Development overview of a network simulator

STEM for Analytics



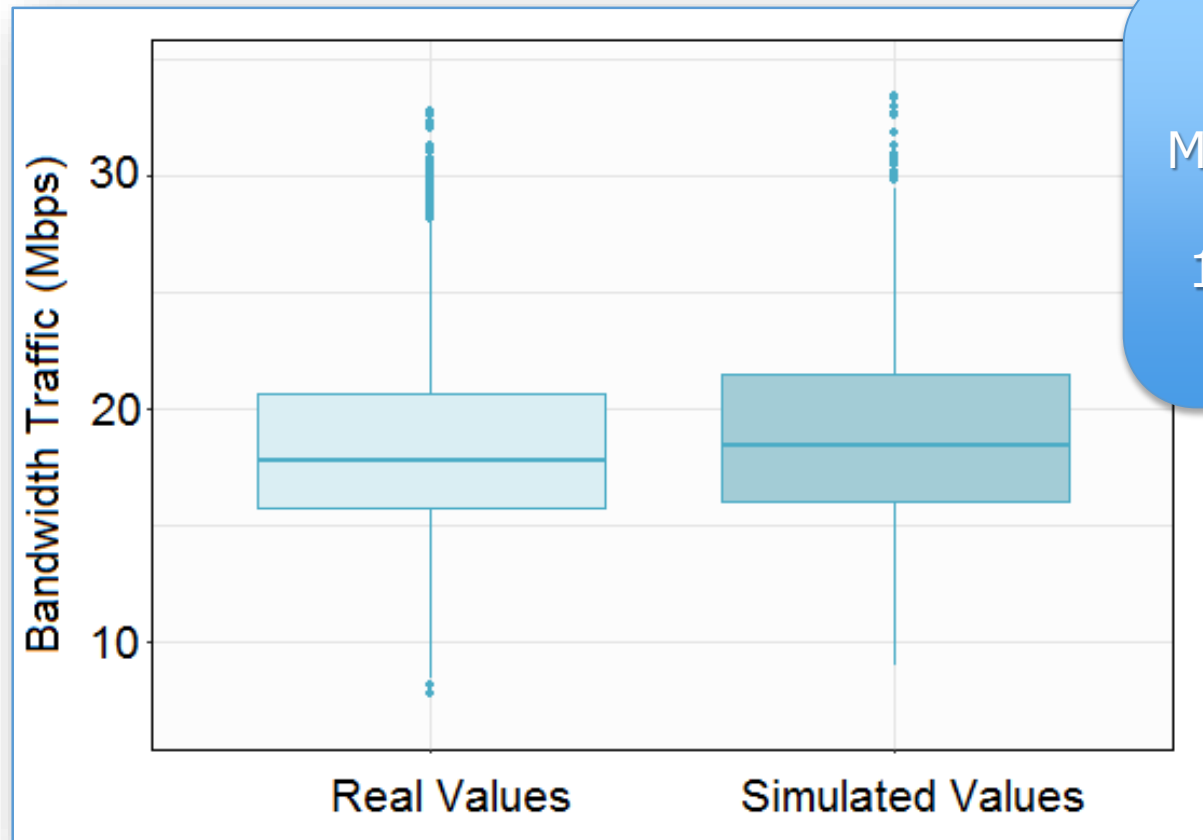
95th percentile of carrier's monthly traffic

STEM for Site Comparison



A site's characteristics makes the difference

STEM for Simulation



Max 95th percentile of 15 days vs 15 times 24 simulations

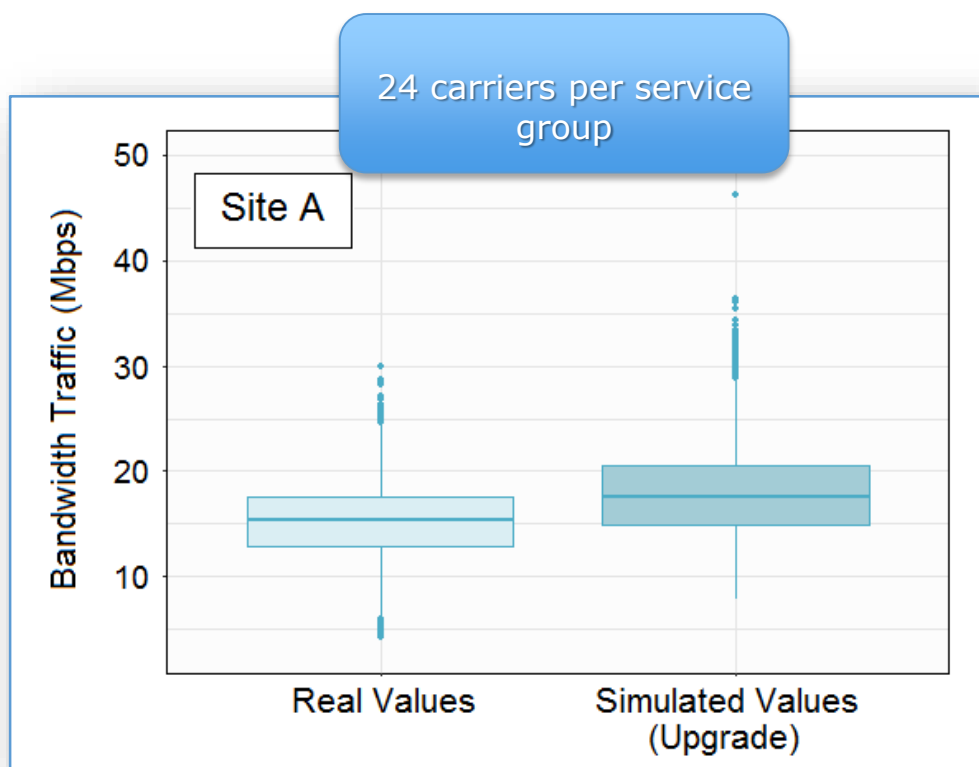
STEM for a Client's Portfolio Upgrade Scenarios



UNIVERSIDAD
AUSTRAL

INGENIERÍA

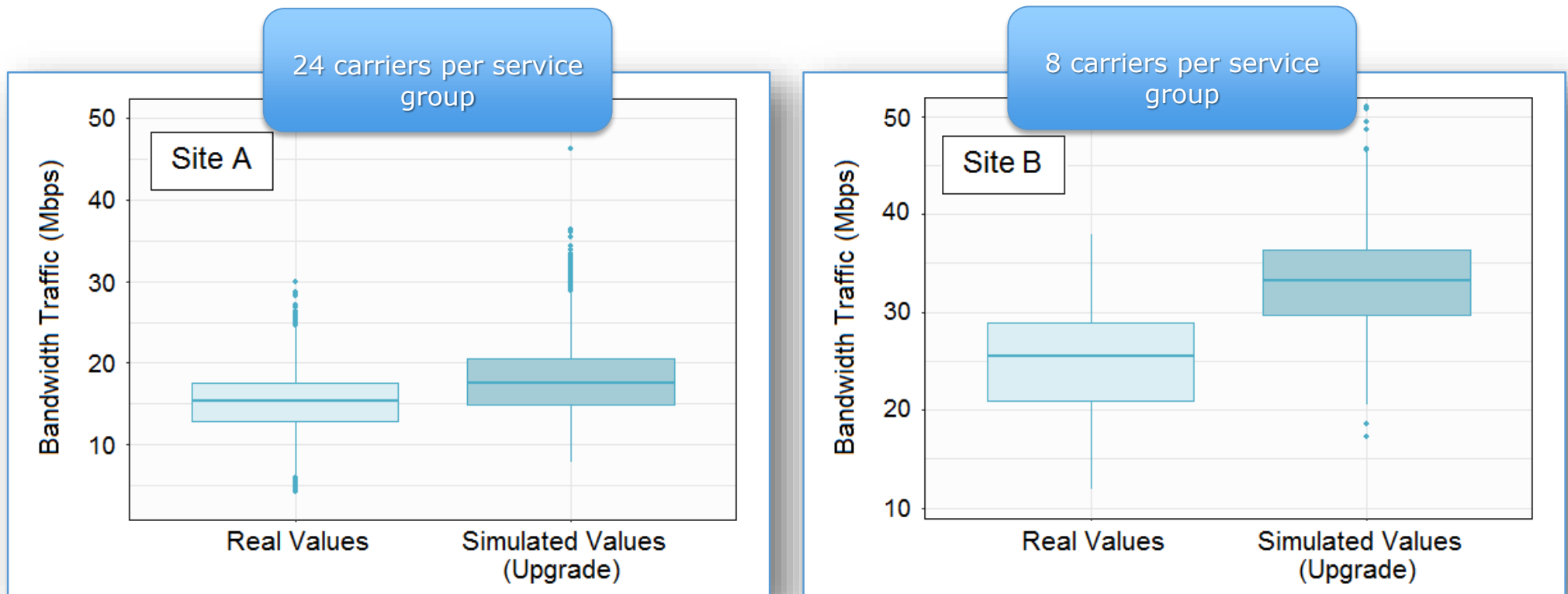
Economy and Popular Tiers → Performance Tier



STEM for a Client's Portfolio Upgrade Scenarios



Economy and Popular Tiers → Performance Tier



..

MUCHAS GRACIAS

Ecosistema Analítico en Cablevisión.
Un caso de uso: Analytics de Redes.

REALIZACION

Alejandro Ciotti – aciotti@cablevisión.com.ar

Jefe de BI & Analytics en Cablevisión - Fibertel

Claudio Righetti – crighetti@cablevisión.com.ar

Chief Scientist en Cablevisión – Fibertel