EDWA : Enterprise Data warehouse Analysis

1. Domain : Telecom

2. Team Size

Total 9

- Mgr-1
- TL/ Scrum Master – 1
- HE/Dev – 2
- MidE/Dev-4
- Infra-1

3. Project Duration

Development : 1.5-2 year

Support – 6 months

4. Bigdata used in

- Telecom
- Banking
- Automobile
- Pharma
- Security
- Ecom
- Transport
- Etc

4. Project Description:

BIG DATA ANALYTICS IMPACT ON TELECOM INDUSTRY

The rapid rise in the use of smartphones and other connected mobile devices has triggered a spurt in the volume of data flowing through the networks of telecom operators. It is necessary that the operators process, store, and extract

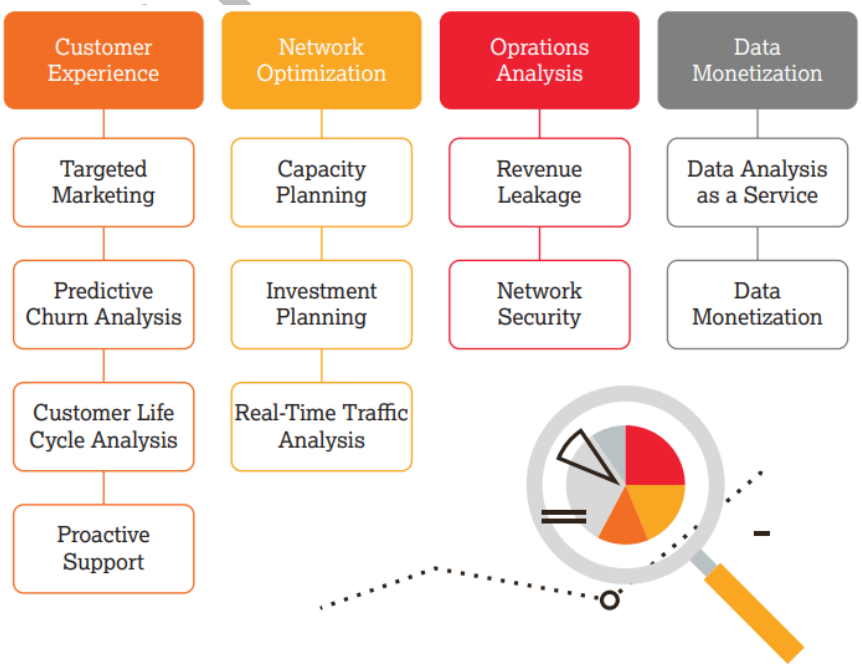
insights from the available data. Big Data analytics can help them increase profitability by helping optimize network usage and services, enhance customer experience, and improve security. Research has shown that the potential for telecom companies to benefit from Big Data analytics is substantial.

The potential of Big Data, however, poses a challenge: how can a company utilize data to increase revenues and profits across the value chain, spanning network operations, product development, marketing, sales, and customer service.

Big Data analytics, for instance, enables companies to predict peak network usage so that they can take measures to relieve congestion. It can also help identify customers who are most likely to have problems paying bills as well as those about to change operators, thus exacerbating churn.

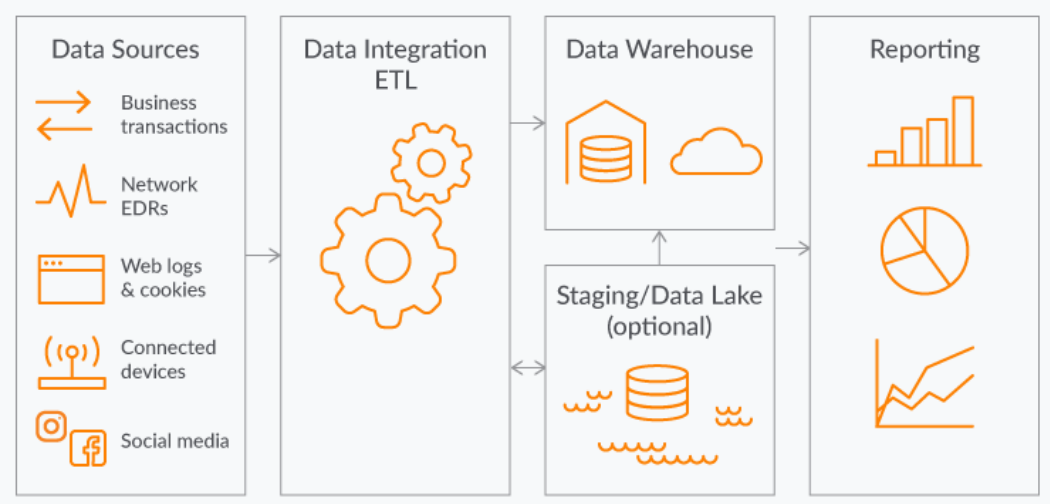
Operators are usually advised against taking the usual top-down approach when it comes to Big Data analytics, which marks out the problem to be solved and then seeks out the data that may help resolve it. Instead, the operators should focus on the data itself, using it to make correlations and connections. If done correctly, the data could reveal insights that could form the basis of more streamlined operations

Benefits Big Data and Predictive Analytics can bring to telecoms:

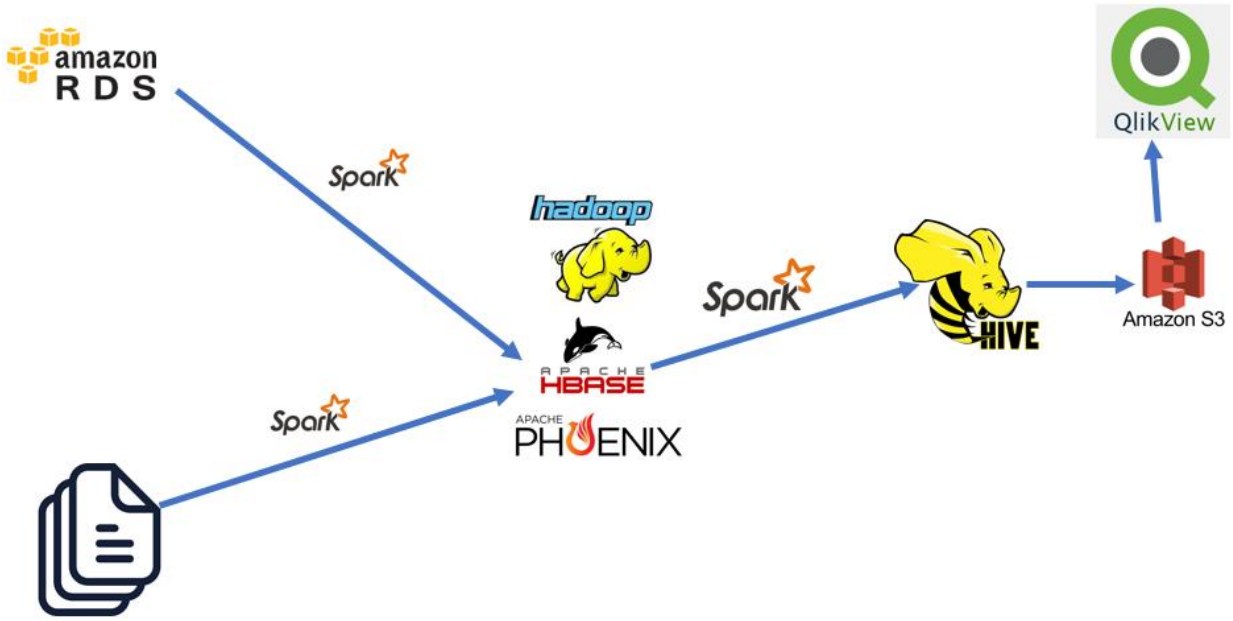


- Making smarter investment decisions
- Visibility into the profitability of different departments
- Reducing fraud
- Improved risk management
- Increased sales
- Smoother network operation
- Enhanced customer experience and reduced churn rate
- Cutting off operations that drain the budget
- Making the operations more efficient
- Increased average revenue per user

General Data Process Flow in Telecom:



Project Data Flow Diagram:



Tools Used with Versions :

1. RDS – Postgre	PostgreSQL 12.5-R1
2. EMR	5.33
3. Spark	2.4.7
4. Python	3.1.2
5. Hadoop	2.10.1
6. Hive	2.3.7
7. Hbase	1.4
8. Phoenix	phoenix-4.14.3
9. S3	NA
10. Linux	NA
11. Airflow	2.0
12. Workbench	Build 127
13. Java	1.8.0_282

Cluster Details :

Total Number of nodes : 9

Masters : 3 (ANN,PNN,RM)

Slaves : 6

Total RAM : 256 GB * 6 ~ = 1.35 TB

HDFS Size ~ = 120 TB (100 TB Process)

Total Cores = 64 * 6 = 350 cores

Cluster retention: 1 year

Replication Factor: 3

Block Size: 128 MB

Data in Depth :

Daily Data Size : ~200 GB

Input File Type : .csv

Delimiter : |

Daily 1 file for each transactional Table.

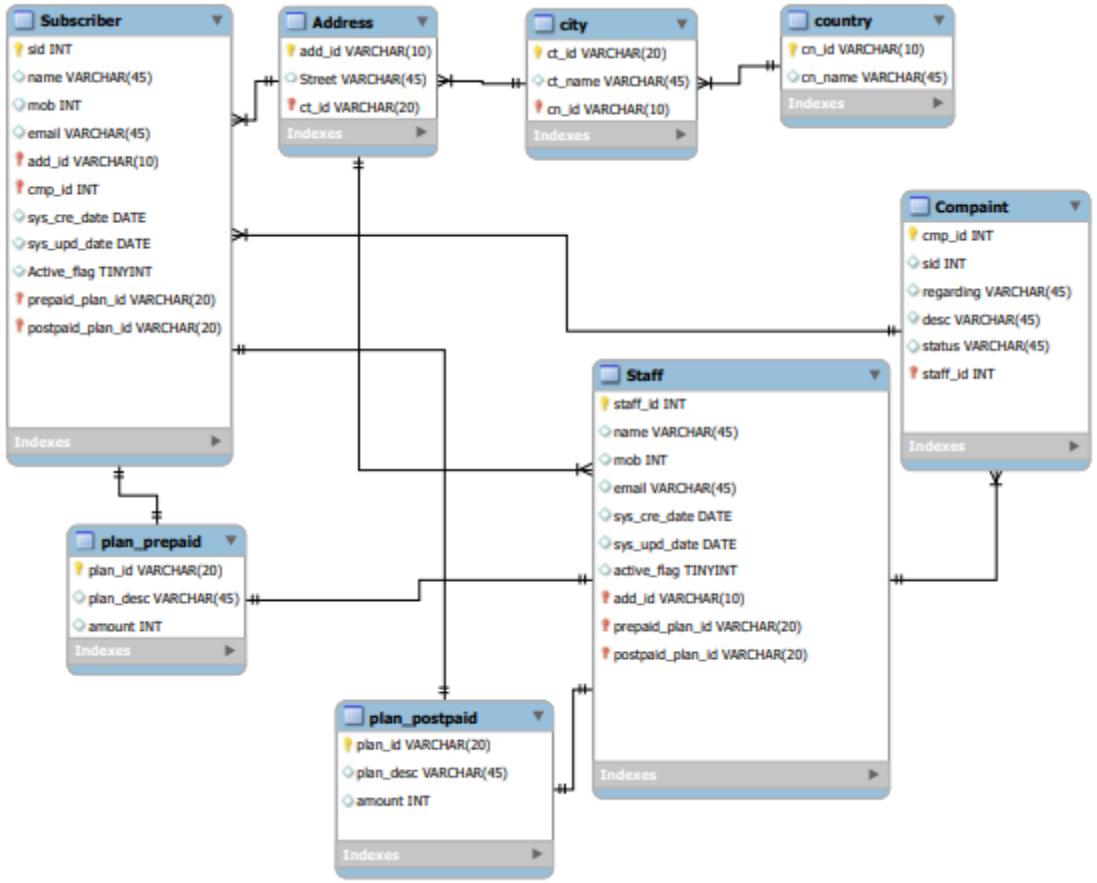
Total Number of tables :

~300 : (220 TX , 60 REF , 20 Hist.)

Total Number of Jobs : Around 340 jobs

Table Details :

1.Source data (Database) Design



Sample Tables

- Transactional Tables :
  - SUBSCRIBER
  - COMPLAINT
  - STAFF
  - ADDRESS
- Reference Tables
  - COUNTRY
  - CITY
  - PLAN_POSTPAID
  - PLAN_PREPAID

Sample Data in table :

Country :

```
1 SELECT * FROM country limit 5;
```

Result 1 | Messages

cn_id	cn_name
CN_1	Afghanistan
CN_2	Algeria
CN_3	American Samoa
CN_4	Angola
CN_5	Anguilla

City :

```
1 SELECT * FROM city limit 5;
```

Result 1 | Messages

ct_id	ct_name	cn_id
CT_1	A Corua	CN_87
CT_2	Abha	CN_82
CT_3	Abu Dhabi	CN_101
CT_4	Acua	CN_60
CT_5	Adana	CN_97

Plan_Prepaid :

```
1 SELECT * FROM plan_prepaid limit 5;
```

Result 1 | Messages

plan_id	plan_desc	amount
PR_1	unlimited call 1.5GB/Day 28 days	199
PR_2	unlimited call 2GB/Day	299
PR_3	unlimited call 1.5GB/Day 84 Days	399
PR_4	unlimited call 3GB/Day 28 Days	499
PR_5	unlimited call 4GB/Day 56 Days	749

Plan Postpaid :

```
1 SELECT * FROM plan_postpaid limit 5;
```

Result 1 | Messages

plan_id	plan_desc	amount
PO_1	unlimited call 40 GB 100 SMS	199
PO_2	unlimited call 75 GB	299
PO_3	unlimited call 100GB	399
PO_4	unlimited call 150 GB 1000 SMS	499
PO_5	unlimited call 200 GB 100 SMS	749

Address :

```
1 SELECT * FROM address limit 5;
```

Result 1 | Messages

add_id	street	ct_id
AD_1	47 MySakila Drive	CT_300
AD_2	28 MySQL Boulevard	CT_576
AD_3	23 Workhaven Lane	CT_300
AD_4	1411 Lillydale Drive	CT_576
AD_5	1913 Hanoi Way	CT_463

Subscriber :

```
1 SELECT * FROM subscriber limit 5;
```

Result 1 | Messages

sid	name	mob	email	add_id	sys_cre_date	sys_upd_date	active_flag	prepaid_plan_id	postpaid_plan_id
1	MARY SMITH	54321	MARY.SMITH@sakilacustomer.org	AD_5	2021-01-01 00:00:00	2021-01-01 00:00:00	A	null	PO_1
2	PATRICIA JOHNSON	54322	PATRICIA.JOHNSON@sakilacustomer.org	AD_6	2021-01-01 00:00:00	2021-01-01 00:00:00	A	null	PO_2
3	LINDA WILLIAMS	54323	LINDA.WILLIAMS@sakilacustomer.org	AD_7	2021-01-01 00:00:00	2021-01-01 00:00:00	A	null	PO_3
4	BARBARA JONES	54324	BARBARA.JONES@sakilacustomer.org	AD_8	2021-01-01 00:00:00	2021-01-01 00:00:00	A	null	PO_4
5	ELIZABETH BROWN	54325	ELIZABETH.BROWN@sakilacustomer.org	AD_9	2021-01-01 00:00:00	2021-01-01 00:00:00	A	null	PO_5

Staff :

```
1 SELECT * FROM staff limit 5;
```

Result 1 | Messages

staff_id	name	mob	email	sys_cre_date	sys_upd_date	active_flag	add_id	prepaid_plan_id	postpaid_plan_id
10001	A	12345	A@A.com	2021-01-01 00:00:00	2021-01-01 00:00:00	A	AD_101	PR_1	PO_1
10002	B	12346	B@B.com	2021-01-01 00:00:00	2021-01-01 00:00:00	A	AD_102	PR_2	PO_2
10003	C	12347	C@C.com	2021-01-01 00:00:00	2021-01-01 00:00:00	A	AD_103	PR_3	PO_3
10004	D	12348	D@D.com	2021-01-01 00:00:00	2021-01-01 00:00:00	A	AD_104	PR_4	PO_4
10005	E	12349	E@E.com	2021-01-01 00:00:00	2021-01-01 00:00:00	A	AD_105	PR_5	PO_5

Complaint :

```
1 SELECT * FROM complaint limit 5;
```

Result 1 | Messages

cmp_id	sid	regarding	descr	status	staff_id	sys_cre_date	sys_upd_date
11111	1	Balance	Balance related Query	Open	10001	2021-01-01 00:00:00	2021-01-01 00:00:00
11112	101	Balance	Balance related Query	Closed	10001	2021-01-01 00:00:00	2021-01-01 00:00:00
11113	112	Recharge	Recharge Related Query	Checking	10002	2021-01-01 00:00:00	2021-01-01 00:00:00
11114	23	Recharge	Recharge Related Query	Open	10002	2021-01-01 00:00:00	2021-01-01 00:00:00
11115	11	Balance	Balance related Query	Closed	10003	2021-01-01 00:00:00	2021-01-01 00:00:00

#Historical Data and Delta

#Processed Sample Data in final Table

#Sample Extracted Report