Web: Inceptez.com Mail: info@inceptez.com Call: 7871299810, 7871299817

Competitive Learning - Part1

Marks – 20/100

Team, As a part of competitive learning, Based on the following business cases, Please create a case study article that covers synopsis, technology stacks you choose, standard architectures you use, data pipeline diagram, technical and business questions you raise with the client before start working, implementation steps, final business benefits etc.,. Please read the whole requirement fully and start create the synopsis document.

Note: You act as a solution architect for asking questions and answer the same yourself as client does in the way you feel better.

There are 2 use cases given below, please work on any one of it as per your wish…

Try to create the document/ppt based on the below key points

1. SDLC – Helps you to cover **Daily roles and responsibilities you do in your project.**
2. Why particular tools – **Will help you to justify the tools with your business**
3. **Challenges** you faces in each tools such as version issue, dependency issues, traditional systems to Bigdata migration issues etc.,

CUSTOMER CHURN PREDICTION & CALL VOLUME REDUCTION

High Level Business Requirement:

A large telecommunications provider analyzed call logs (CDR) and other complex data from multiple sources. The aim of this project is to analyze the data to predict the likelihood that any particular customer would leave based on the call type, call duration, agent response, customers responses in the social media, data captured from the activities performed by the call center agents in their desktop when they are engaging the customers.

The business challenge for this large mobile carrier that needs to analyze multiple data sources to understand how and why customers decided to terminate their service contracts. Were customers actually leaving, or were they merely trading one service plan for another? Were they leaving the company entirely and moving to a competitor? Were pricing, coverage gaps, or device issues a factor? What other issues were important, and how could the provider improve satisfaction and retain customers?

Understanding the above high level requirement from the customer, Please fulfil the following topics to provide technical solution using Bigdata eco systems.

1. Prepare a questionnaire that you ask from the customer in terms of data type, growth rate etc.,
2. Assume you are getting set of responses from the client for the questions you asked, for eg. You got the response as the feeds are going to be continuous or hourly or daily. Assume what ever it may be and create the following scenarios.
3. Following are the sample data sets.

CDR - Data Format

ID, CALLING_NUM, CALLED NUMBER, START TIME, END TIME, CALL TYPE, CHARGE, CALL RESULT

Sample Data:

```
f7c2aed3-8209-472f-afa6-6a1dc4a0e0b6|8131166797|3470914600|2016-03-04T00:10:05.681+05:30|2016-03-
04T00:10:05.681+05:30|SMS|0.8428941|ANSWERED
49dde45a-dd7e-4568-9432-549fe59cb658|9149565512|3539212588|2016-01-17T07:28:39.157+05:30|2016-01-
17T07:31:25.587+05:30|VOICE|0.52268726|ANSWERED
fca332c0-d0e7-43cd-8918-133de466c19d|3917407842|3976557336|2016-03-05T18:44:51.669+05:30|2016-03-
05T18:44:51.669+05:30|SMS|0.92834604|ANSWERED
c94c4ef7-128c-4d16-82db-bb70a05c4948|1190111696|4794533137|2016-02-02T09:30:57.015+05:30|2016-02-
02T09:33:30.020+05:30|VOICE|0.4371317|ANSWERED
d802f413-060c-4800-9319-2ae3ed9cb9b3|4947005879|5596524704|2016-01-27T01:56:35.448+05:30|2016-01-
27T01:59:23.904+05:30|VOICE|0.40826023|ANSWERED
ff51c7fb-48e6-4dc5-bd91-62352d07608b|2339563418|4229486404|2016-01-22T07:51:16.941+05:30|2016-01-
22T07:51:16.941+05:30|SMS|0.6248307|ANSWERED
d070d101-f383-43e3-ac70-e42ecb176235|1874571022|9163944393|2016-03-02T00:07:07.179+05:30|2016-03-
02T00:09:22.678+05:30|VOICE|0.09043157|ANSWERED
99102812-fe69-4c2c-930d-c27616f62915|1350057252|2751071043|2016-02-19T21:09:14.280+05:30|2016-02-
19T21:11:45.164+05:30|VOICE|0.19128877|ANSWERED
7b549f79-7bef-4a85-9788-c18a7d24b925|3701263930|8504596517|2016-01-29T01:26:11.234+05:30|2016-01-
29T01:28:23.702+05:30|VOICE|0.5474914|ANSWERED
04b3e83b-edbf-4bc0-8d14-0bf65a672159|2689744648|9413615188|2016-01-27T08:14:58.783+05:30|2016-01-
27T08:14:58.783+05:30|SMS|0.36705214|ANSWERED
c082cc74-707d-4a0e-8ddc-45b03a5b8447|3407615224|1071273179|2016-02-06T05:30:17.333+05:30|2016-02-
06T05:32:28.604+05:30|VOICE|0.672031|ANSWERED
```

```
df850964-dfd7-44c5-80c7-6400bc3f58bd|7615931747|5471126090|2016-03-01T02:06:18.860+05:30|2016-03-
01T02:06:18.860+05:30|SMS|0.22748524|ANSWERED
d41a1084-4ede-4772-ab5b-dbc3f0060651|1347582243|9889461894|2016-02-06T15:32:29.473+05:30|2016-02-
06T15:32:29.473+05:30|SMS|0.41556|ANSWERED
1fa9930e-6bff-4ff5-b472-49b794a336e2|3864774659|7654837485|2016-01-26T12:31:49.613+05:30|2016-01-
26T12:33:52.070+05:30|VOICE|0.15566134|ANSWERED
361cd018-77f6-44e3-9fcb-fe15f4d54d80|8119639213|1819076611|2016-01-30T23:23:11.493+05:30|2016-01-
30T23:26:08.366+05:30|VOICE|0.477566|ANSWERED
2b446baf-fdd9-4a63-8ba3-3aa23278bc70|1945030245|0273614418|2016-02-03T08:17:02.883+05:30|2016-02-
03T08:17:02.883+05:30|SMS|0.66082704|ANSWERED
89f5c75d-170c-4056-8d4e-6c72dd728018|3772584645|6955789493|2016-02-26T10:33:03.718+05:30|2016-02-
26T10:35:27.604+05:30|VOICE|0.3173859|ANSWERED
d114b7d6-2e32-4c75-be78-861be71cfedf|3527771817|2373124924|2016-03-03T06:46:52.406+05:30|2016-03-
03T06:49:01.864+05:30|VOICE|0.22446603|ANSWERED
3f9fea2e-3eb5-4a00-93ab-e65583ceaabe|9563569212|2227814067|2016-01-30T23:10:16.299+05:30|2016-01-
30T23:10:16.299+05:30|SMS|0.62429583|ANSWERED
217f7d54-6372-4430-bd21-d4321a224b97|2193898757|5666756896|2016-01-19T14:33:08.206+05:30|2016-01-
19T14:33:08.206+05:30|SMS|0.561812|ANSWERED
170e230d-a874-423a-af1f-16fce48cc715|2570224262|3802970224|2016-02-20T15:53:46.401+05:30|2016-02-
20T15:53:46.401+05:30|SMS|0.30017197|ANSWERED
```

Agent Desktop Events Data Format:

The below 2 activity of the agent shows what url the agent is visiting when engaged with the Customer.

DesktopId|EventStartTime|EventEndTime|EventType|Application|ApplicationAction|ApplicationURL|ActiveState

```
1|2016-02-20T15:53:46|2016-02-
20T15:54:40|Open|Chrome|Browsing|https://www.youtube.com/watch?v=wmcaTkI5Z2s|KeyboardActive
2|2016-02-20T15:54:46|2016-02-
20T15:56:00|Open|Chrome|Browsing|https://www.youtube.com/results?search_query=worldcup+football+2018+live|
MouseActive
```

Tweets and feedback from the Customer from twitter and from the feedback page of the website:

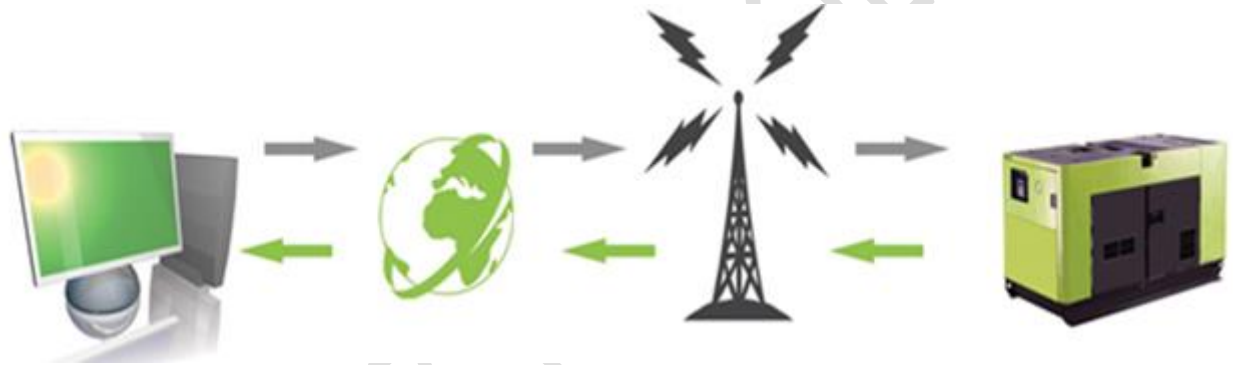
Freeform text data contains sentiment of the customer for example.

The product xyz is doing very great for the long time.

I got a bad experience when I call the call center of the particular carrier called Airtel.

4. Create the detailed description for the following topics: (In short explain whay, why, when, where we are going with the eco systems)
    a. Tools that you choose to create the end to end data pipelines either batch or realtime or both.
    b. What is the short description of each tools/eco systems you choose and why you choose those tools when other tools are available in the market. For eg. Open source, distributed, scalable, ease of code etc.,
    c. Create a data pipeline flow diagram (Batch, or realtime or both flows)– You can even draw with the free hand and take a snap and attach.
    d. Explain about how do you integrate the components, schedule, deploy in prod.
    e. What are the business benefit finally you may get with the data sample provided above.

IOT - Failure Prediction of Devices



Business Summary

A very large public power company combined sensor data from the smart grid with a map of the network to predict which generators in the grid were likely to fail, and how that failure would affect the network as a whole.

Business Challenge

Power Utilities that runs big, expensive and complicated systems to generate power. Each of the generators includes sophisticated sensors that monitor voltage, current, frequency and other important operating characteristics. Operating a single generator means paying careful attention to all of the data streaming off of the sensors attached to it, these generators spread across multiple locations. The locations are connected to one another, and then each utility is connected to the public power grid. Monitoring the health of the entire grid requires capture and analysis of data from every utility, and even from every generator, in the grid. The volume of data is enormous. A clear picture of the health of the grid depends on both real-time and after-the-fact forensic analysis of all of it. Spotting facilities at risk of failure early, and doing preventive maintenance or separating them from the grid, is critical to preventing costly outages.

Another business challenge in IT Industry –

While this was a highly specialized project, it has an analog in data centers managing IT infrastructure grids. In a large data center with thousands of servers, understanding what the systems and applications are actually doing is difficult. Existing tools often don't scale. IT infrastructure can capture system-level logs that describe the behavior of individual servers, routers, storage systems and more. Higher-level applications generally produce logs that describe the health and activity of application servers, web servers, databases and other services. Large data centers produce an enormous amount of this data. Understanding the relationships among applications and devices is hard. Combining all of that data into a single repository, and analyzing it together, can help IT organizations better understand their infrastructure and improve efficiencies across the network. We need to store and analyze log data, and builds a higher-level picture of the health of the data center as a whole.

Data Sample:

SensorID|Date|Pressure|active
1|2018-01-05 23:45:30|90PSI|yes
2|2018-01-06 23:45:30|95PSI|yes

Date~Time~Global_active_power~Global_reactive_power~Voltage~Global_intensity~Sub_metering_1~Sub_metering_2~Sub_metering_3
1/1/2007~0:00:00~2.58~0.136~241.97~10.6~0~0~0
1/1/2007~0:01:00~2.552~0.1~241.75~10.4~0~0~0
1/1/2007~0:02:00~2.55~0.1~241.64~10.4~0~0~0
1/1/2007~0:03:00~2.55~0.1~241.71~10.4~0~0~0
1/1/2007~0:04:00~2.554~0.1~241.98~10.4~0~0~0
1/1/2007~0:05:00~2.55~0.1~241.83~10.4~0~0~0
1/1/2007~0:06:00~2.534~0.096~241.07~10.4~0~0~0
1/1/2007~0:07:00~2.484~0~241.29~10.2~0~0~0
1/1/2007~0:08:00~2.468~0~241.23~10.2~0~0~0
1/1/2007~0:09:00~2.486~0~242.18~10.2~0~0~0
1/1/2007~0:10:00~2.492~0~242.46~10.2~0~0~0
1/1/2007~0:11:00~2.5~0~242.88~10.2~0~0~0
1/1/2007~0:12:00~2.494~0~242.57~10.2~0~0~0
1/1/2007~0:13:00~2.492~0~242.41~10.2~0~0~0
1/1/2007~0:14:00~2.48~0~241.81~10.2~0~0~0
1/1/2007~0:15:00~2.478~0~241.73~10.2~0~0~0
1/1/2007~0:16:00~2.47~0~241.29~10.2~0~0~0
1/1/2007~0:17:00~2.466~0~241.11~10.2~0~0~0
1/1/2007~0:18:00~2.456~0~240.59~10.2~0~0~0
1/1/2007~0:19:00~2.46~0~240.83~10.2~0~0~0
1/1/2007~0:20:00~2.544~0.092~240.9~10.6~0~0~0
1/1/2007~0:21:00~2.55~0.116~241.15~10.4~0~1~0
1/1/2007~0:22:00~2.554~0.118~241.55~10.6~0~1~0
1/1/2007~0:23:00~2.65~0.218~241.67~11~0~2~0
1/1/2007~0:24:00~2.682~0.258~242.45~11~0~1~0
1/1/2007~0:25:00~2.66~0.252~241.6~11~0~1~0

Create the detailed description for the following topics : (In short explain whay, why, when, where we are going with the eco systems)

a. Tools that you choose to create the end to end data pipelines either batch or real time or both.
b. What is the short description of each tools/eco systems you choose and why you choose those tools when other tools are available in the market. For eg. Open source, distributed, scalable, ease of code etc.,
c. Create a data pipeline flow diagram (Batch, or realtime or both flows)– You can even draw with the free hand and take a snap and attach.
d. Explain about how do you integrate the components, schedule, deploy in prod.
e. What are the business benefit finally you may get with the data sample provided above.