



1.Problem statement :Huge Data of User, Comments ,Likes and Traffic which need special help for information extraction

2.Objective of the project: To democratize the Information retrieval for the Platform owners, Traffic analyst

3.Previous work Referred:

<https://github.com/johnthebrave/nliddb-datasets>

<https://github.com/ElleutherAI/stackexchange-dataset>

4. Future scope Envisioned: In a future scope for this project we want to build a NLIDB which can be configured to other type of Dataset related to User , comments, Traffics on your Different Blogs, Websites, YouTube Channels which you want to evaluate or monitor comparatively

5. Sample Questions & Queries:

The NLIDB endpoint should be able to answer questions and generate queries as given below-

NL> List all the comments that have 'hibernate' in the text.

*Q> select * from Comments where Text='*hibernate*'*

NL>select displayname, is, location reputation, number of userid from users and badges where reputation is >= 550000 grouped by displayname, id, location, reputation ordered by reputation desc"

Q>SELECT DisplayName, a.Id, Location, Reputation, count(b.userid) as BadgesCount From Users a, badges b where a.id=b.userid and Reputation>=550000 group by DisplayName, a.Id, Location, Reputation order by reputation desc

6. Sample Table schema & Dataset:

TABLE SCHEMA

Table(Links to Sample Data)	Field #	Column	type
Badges	1	Id	int(10)
Badges	2	UserId	int(10)
Badges	3	Name	nvarchar(50)
Badges	4	Date	datetime
Badges	5	Class	tinyint(3)
Badges	6	TagBased	bit
Comments	1	Id	int(10)
Comments	2	PostId	int(10)
Comments	3	Score	int(10)
Comments	4	Text	nvarchar(600)
Comments	5	CreationDate	datetime
Comments	6	UserDisplayName	nvarchar(40)
Comments	7	UserId	int(10)
Comments	8	ContentLicense	varchar(12)

SAMPLE DATASET

Id	PostId	Score	Text	CreationDate	UserId	ContentLicense	UserDisplayName
2	6	5	Awesome, me and my fiance have been debating this with people for a while.	11-01-2011 20:56	35	CC BY-SA 2.5	
4	14	3	Great book, but hellla depressing!	11-01-2011 21:08	54	CC BY-SA 2.5	
5	6	38	but without convenient plot holes, how does a story move forward?	11-01-2011 21:09	54	CC BY-SA 2.5	
7	4	2	I see you marked LOTR as not SciFi, but on a side note I wonder who is going to edit out all the fantasy-genre questions that will inevitably crop up. Plus, what about mixed-genres, like the "Shadowrun" series?	11-01-2011 21:26	9	CC BY-SA 2.5	
8	4	5	We will have to tolerate some chocolate in our peanut-butter.	11-01-2011 21:35	51	CC BY-SA 2.5	
10	1	0	Do the languages have to be alien? Must they have been created for the Novel/Show or could they just be the only place ever actually used. Like Lojban in The Moon is a Harsh Mistress?	11-01-2011 21:38	51	CC BY-SA 2.5	
11	38	1	Well, physically, Pluto remains exactly the same. The fact that they said "Planet" instead of "Planetoid" is minor. I have more trouble suspending disbelief when there are martians, like in Stranger in a Strange Land.	11-01-2011 21:42	55	CC BY-SA 2.5	
12	15	17	I couldn't help but think of [this scene](http://goo.gl/5Y8Nt). :-)	11-01-2011 21:45	42	CC BY-SA 2.5	



Voice driven chatbot to query on complex questions and insights from sales and HR functions , encouraging end-user to explore data by their own without dedicated Dashboards and bringing in a self-service platform for Data driven culture

1.Problem statement : Dependencies on dedicated and static report or dashboard without having direct access to data for questioning .

2.Objective of the project: To democratize the Information retrieval for different Analytical Insights directly based on voice chat

3.Previous work Referred:

<https://github.com/IBM/TempQ4NLIDB-dataset>

4. Future scope Envisioned: In a future scope of this project can be extended to other modules and cross module and can be made accessible from teams and other chat platforms

5. Sample Questions & Queries:

The NLIDB endpoint should be able to answer questions and generate queries as given below-

NL> how many products are sold with prices greater than 200 dollars between 2017-01-01 and 2017-12-31

Q> select distinct sum (T2.QUANTITY) from SALES AS T1 JOIN SALES_DETAILS AS T2 on T1.SALES_ID = T2.SALES_ID JOIN PRODUCTS AS T3 on T2.PRODUCT_ID = T3.PRODUCT_ID where T3.PRICE > 200 and T1.DATE >= '2017-01-01' and T1.DATE <= '2017-12-31'

NL>show price, type, address of the shop of the products that are in stock with a price higher than 300 in BESTBYU

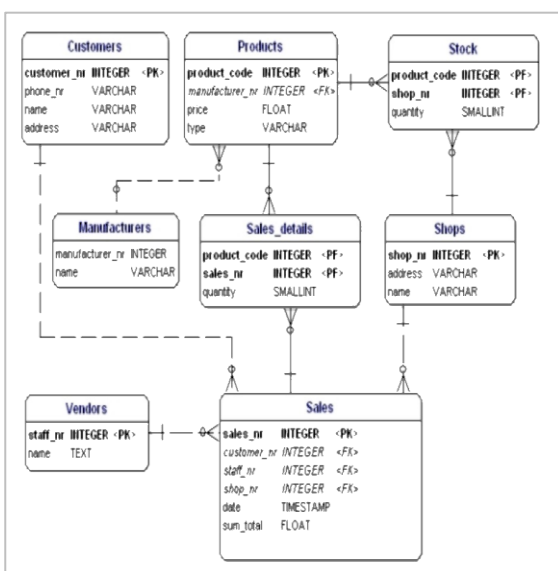
Q>SELECT T1.type , T1.price, T3.address FROM products AS T1 JOIN stock AS T2 ON T1.product_id = T2.product_id JOIN shops AS T3 ON T2.shop_id = T3.shop_id WHERE T1.price >= 300 AND T3.name = 'BESTBYU'

NL>who has been working in Marketing department since hired between 2015 and 2020

Q>select distinct EMPLOYEE.EMPNO from EMPLOYEE where EMPLOYEE.DPTNAME = 'Marketing' and EMPLOYEE.HIREDATE >= '2015-01-01' and EMPLOYEE.HIREDATE <= '2020-12-31'

6. Sample Table schema & Dataset:

SALES SCHEMA



HR SCHEMA

EMPLOYEE

EMPNO	INTEGER
EMPNAME	VARCHAR
MGRNAME	VARCHAR
BIRTHDATE	TIMESTAMP
HIREDATE	TIMESTAMP
LEAVEDATE	TIMESTAMP
SALARY	INTEGER
BONUS	INTEGER
DPTNAME	VARCHAR

SAMPLE TABLE DATA

PRODUCT_ID	MANUFACTURER_ID	TYPE	PRICE
5501	2001	PHONE	525
5502	2001	WACBLK	934
5503	2001	BOOKMARK	139
5504	2002	GALACTY	140.99
5505	2002	GALACTYTAB	159
5506	2003	BURRAS	200.99
5507	2003	WINDCOORS	119.99

EMPNO	EMPNAME	MGRNAME	BIRTHDATE	HIREDATE	LEAVEDATE	SALARY	BONUS	DPTNAME
337	Jordan	Mark Doe	1960-02-13	2016-11-01		58000	6000	Manufacturing
331	Talant	Mark Doe	1960-07-23	2015-10-07		62000	5700	Manufacturing
340	Samantha	Mark Doe	1961-05-17	2016-01-15		62000	5700	Manufacturing
320	Brian	Mark Doe	1955-08-05	2016-08-01		63000	5800	Manufacturing
324	Caitie	Mark Doe	1961-12-31	2016-08-01		64000	5700	Manufacturing
232	Mike	Robert Doe	1960-08-27	2016-12-01		66000	6000	Marketing
115	Matt	Peter Doe	1965-08-08	2016-08-01		67000	6000	Sales