

NEERAJ VERMA
TYBCA
BMEF
2593



KHOJ

Enter Anything to Search

Search

THE CUSTOM DESIRED SEARCH ENGINE

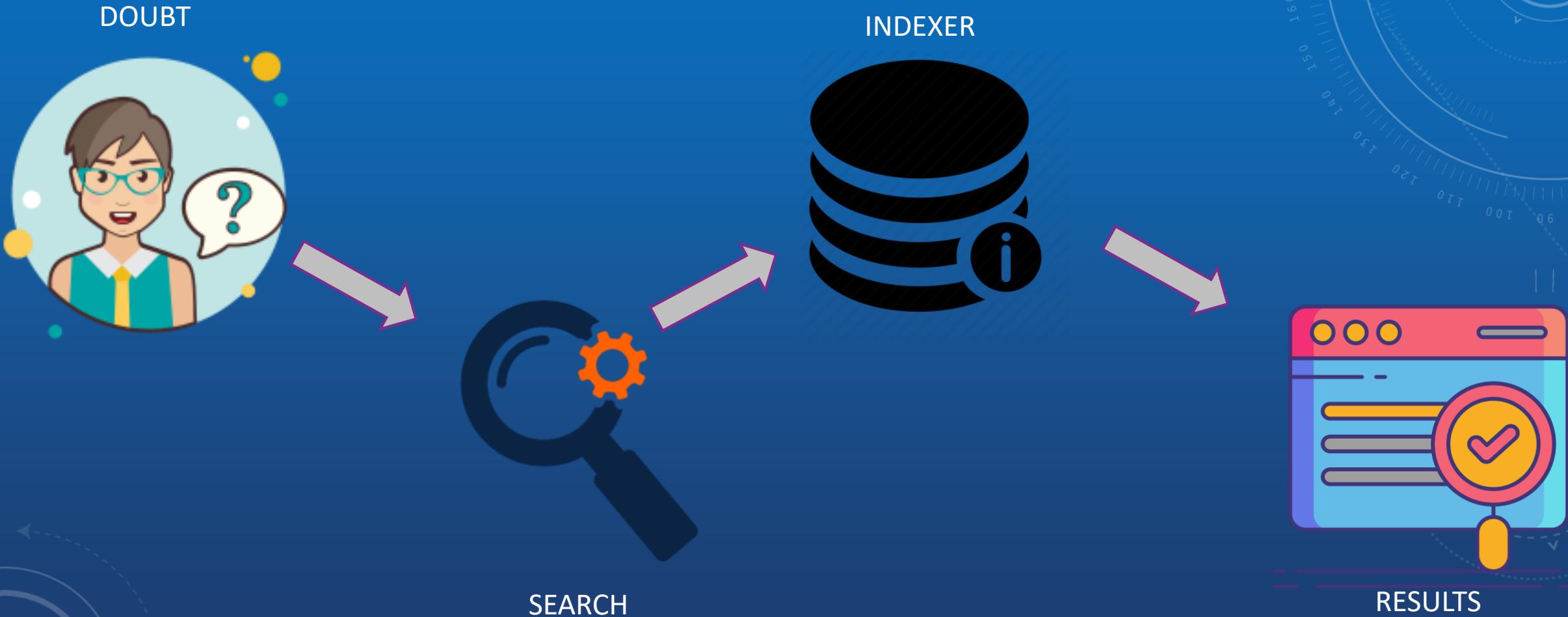
ABSTRACT

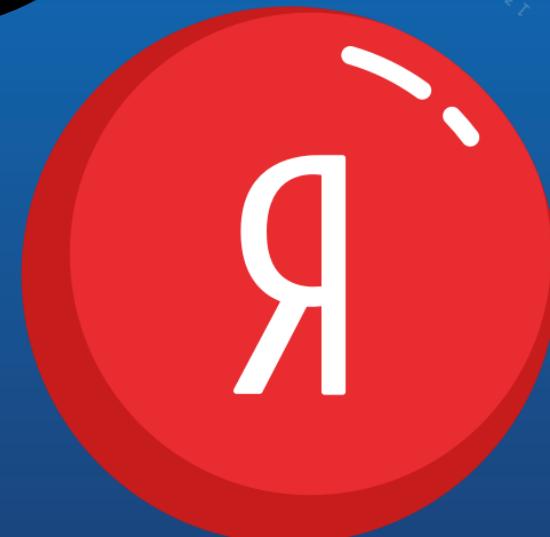
- In this report, I present the design and implementation of a search engine, named KHOJ. The search engine is designed to handle both text and image search, with a flexible architecture that will facilitate scaling up. Each module is independent of each other and therefore swappable to improve its own functionality without affecting the whole system. More importantly, a new algorithm, Click Rank, has been designed and fully implemented to achieve similar goal as that of the existing Page Rank [Sergey Brin, 1998] algorithm while providing similar performance and an additional feature for managing similar pages in search result.

CONTENT :

- **Search Engine**
- **How it works ?**
- **Mechanism**
- **Why Customize Search Engine ?**
- **Crawler**
- **How it works ?**
- **Features**
- **System Design**
- **Limitations**
- **Future Scope**

WHAT IS SEARCH ENGINE ?





approximately **5.6** billion searches are made
in a single day from search engine

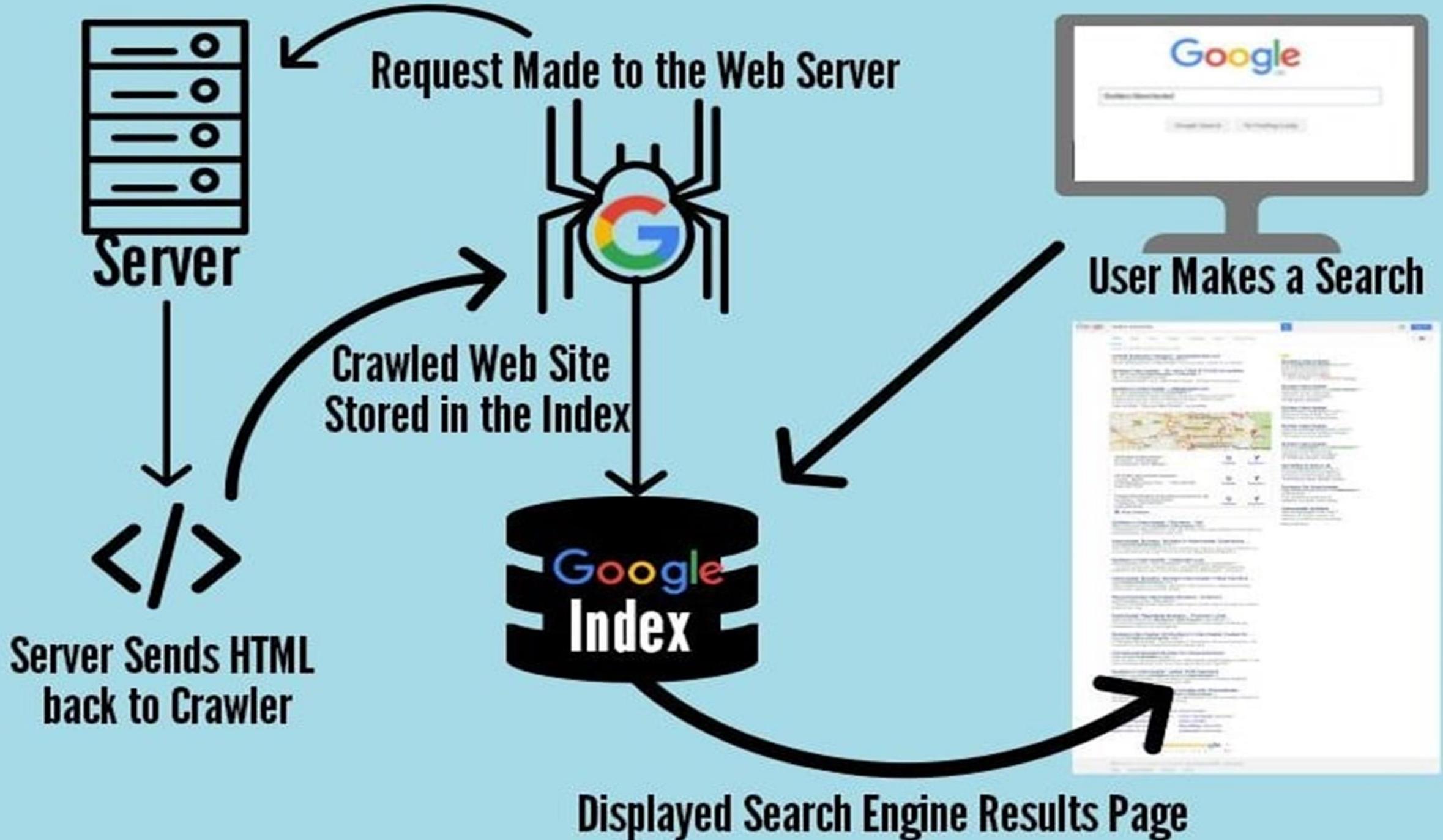


HOW SEARCH ENGINES WORK

HOW SEARCH ENGINE WORKS ?

There are differences in the ways various search engines work, but they all perform three basic tasks:





WHY CUSTOMISED SEARCH ENGINE ?



SECURE DATABASE



NO ADS POLICY



QUICK RESULTS

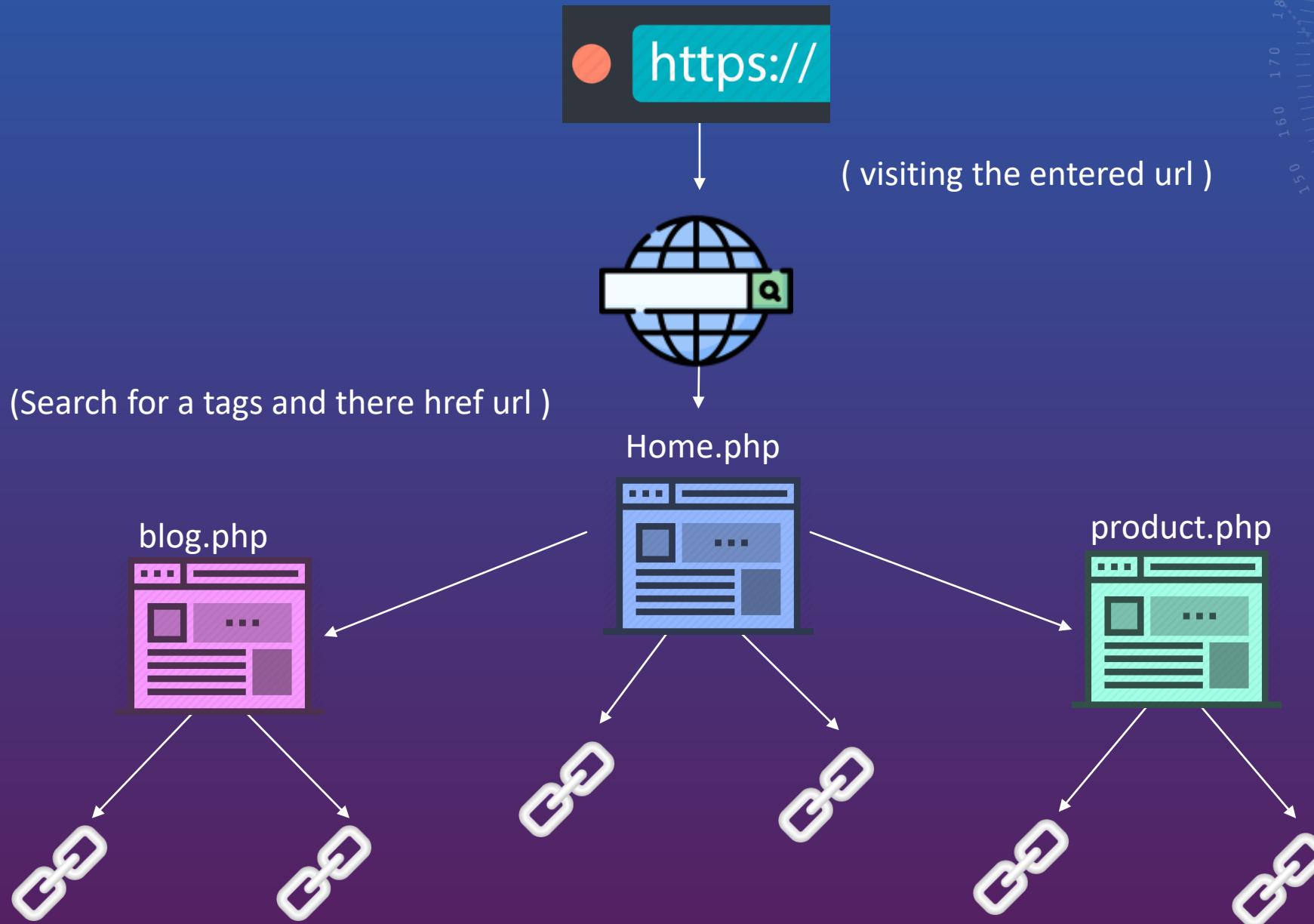


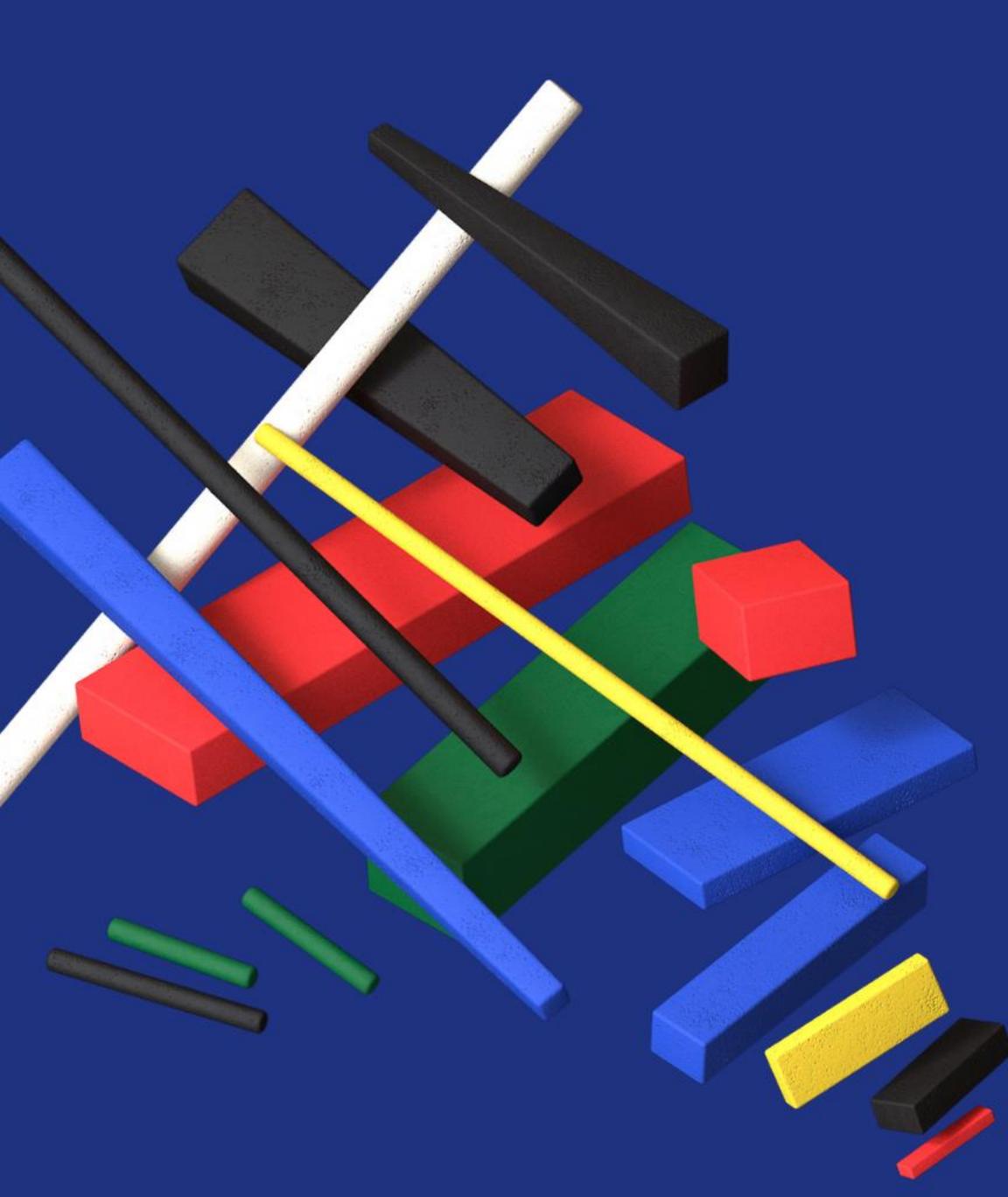
SECURE SEARCH

WORKING OF CRAWLER.

- A crawler is a computer program that automatically searches documents on the Web. Crawlers are primarily programmed for repetitive actions so that browsing is automated. Search engines use crawlers most frequently to browse the internet and build an index.
- Other crawlers search different types of information such as RSS feeds and email addresses. The term crawler comes from the first search engine on the Internet: the Web Crawler. Synonyms are also “Bot” or “Spider.” The most well known webcrawler is the Googlebot.
- The Gupt-Char is the name of KHOJ’s crawler and it is at beginning stage that fetches site links when it is triggers to add.

HOW GUPT-CHAR HAS BEEN DESIGNED ?



A cluster of 3D geometric shapes in various colors (red, green, blue, yellow, black, white) are scattered across a solid blue background. The shapes include cubes, rectangular prisms, and cylinders, some standing upright and others lying on their sides.

Features Of KHOJ

WHAT FEATURES HAS BEEN INCLUDED ?



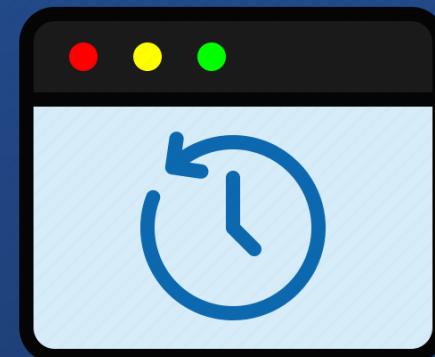
Login System



Multi User



Account Update



Search History

TYPES OF USERS



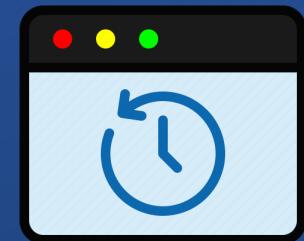
Explorer



Browse

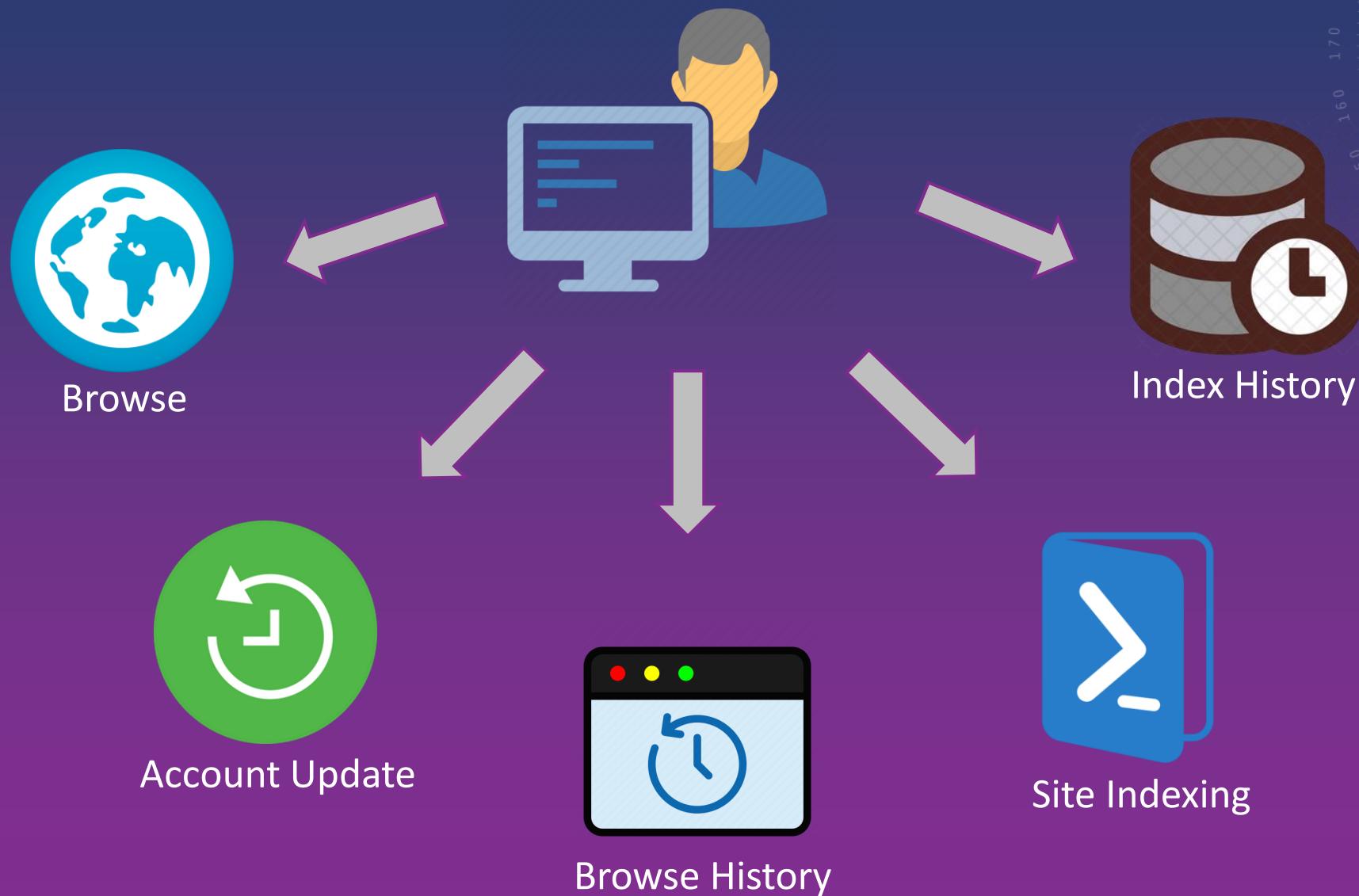


Account Update

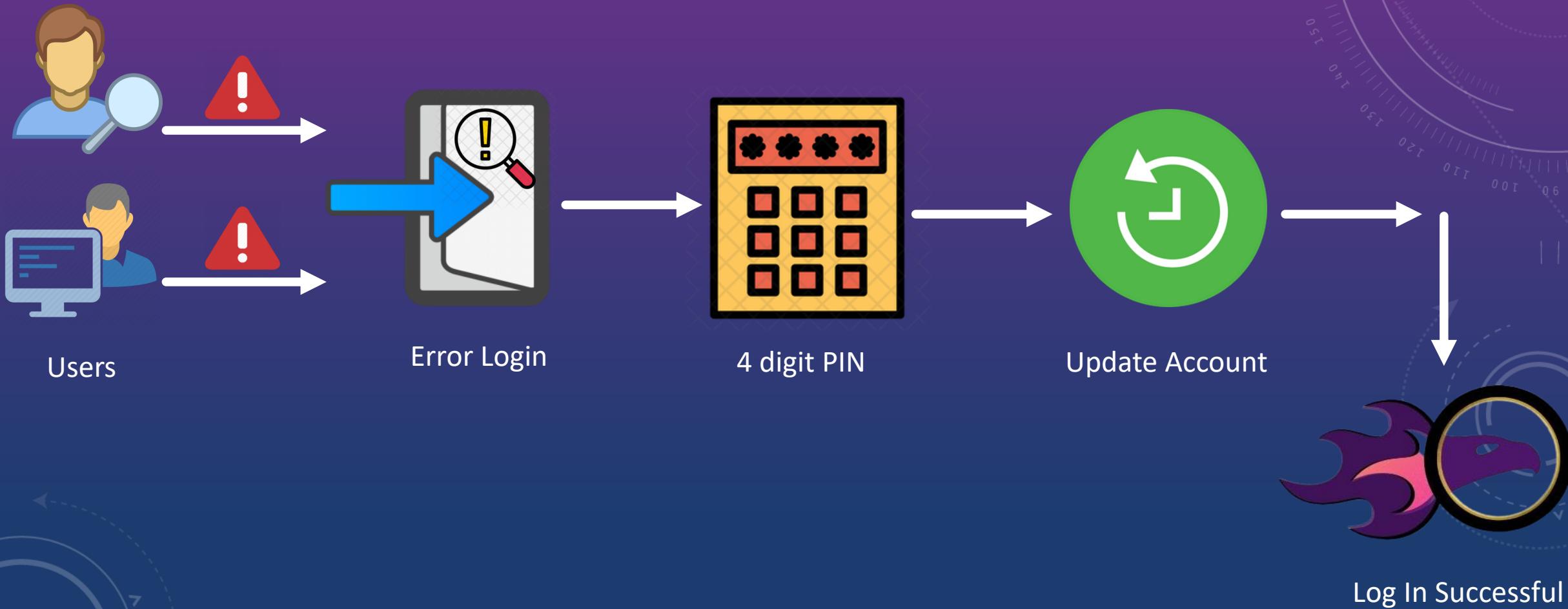


Searched History

Developer



PIN SECURITY & RECOVERY

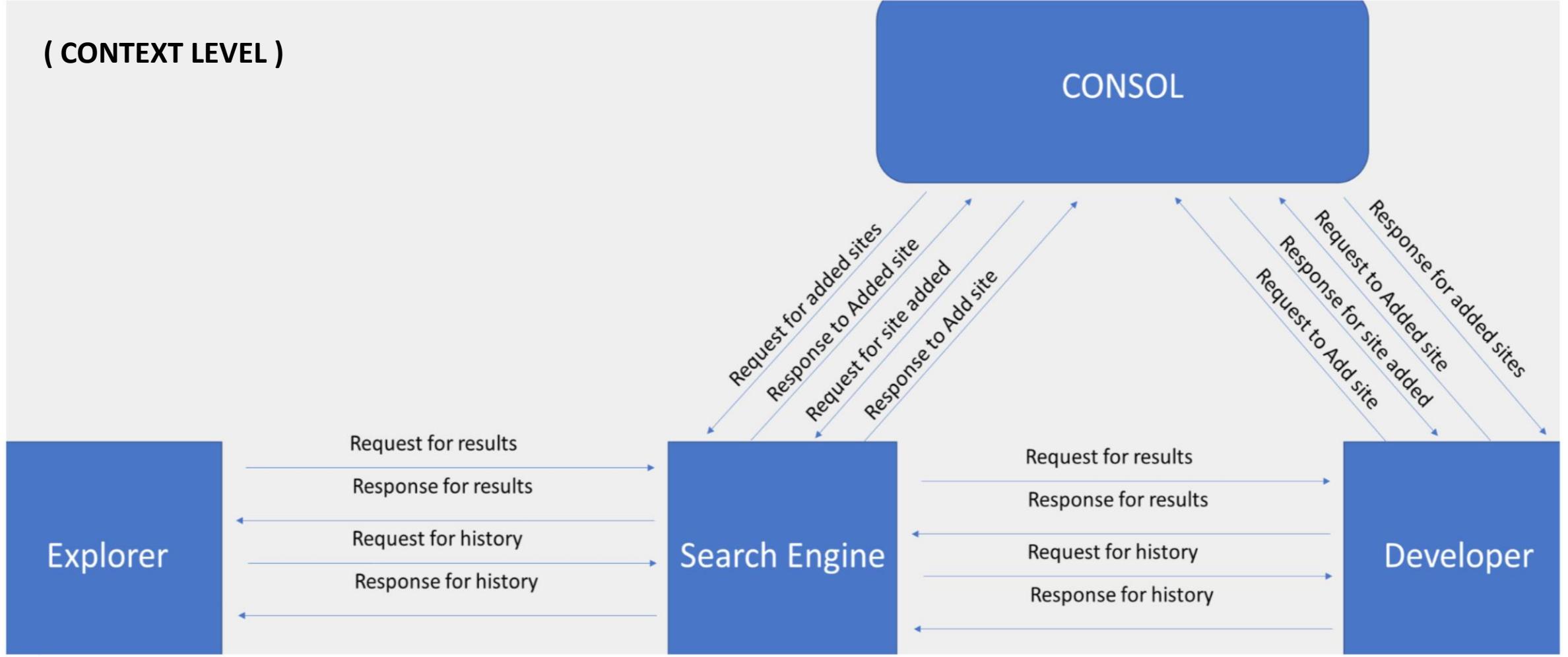


System Design



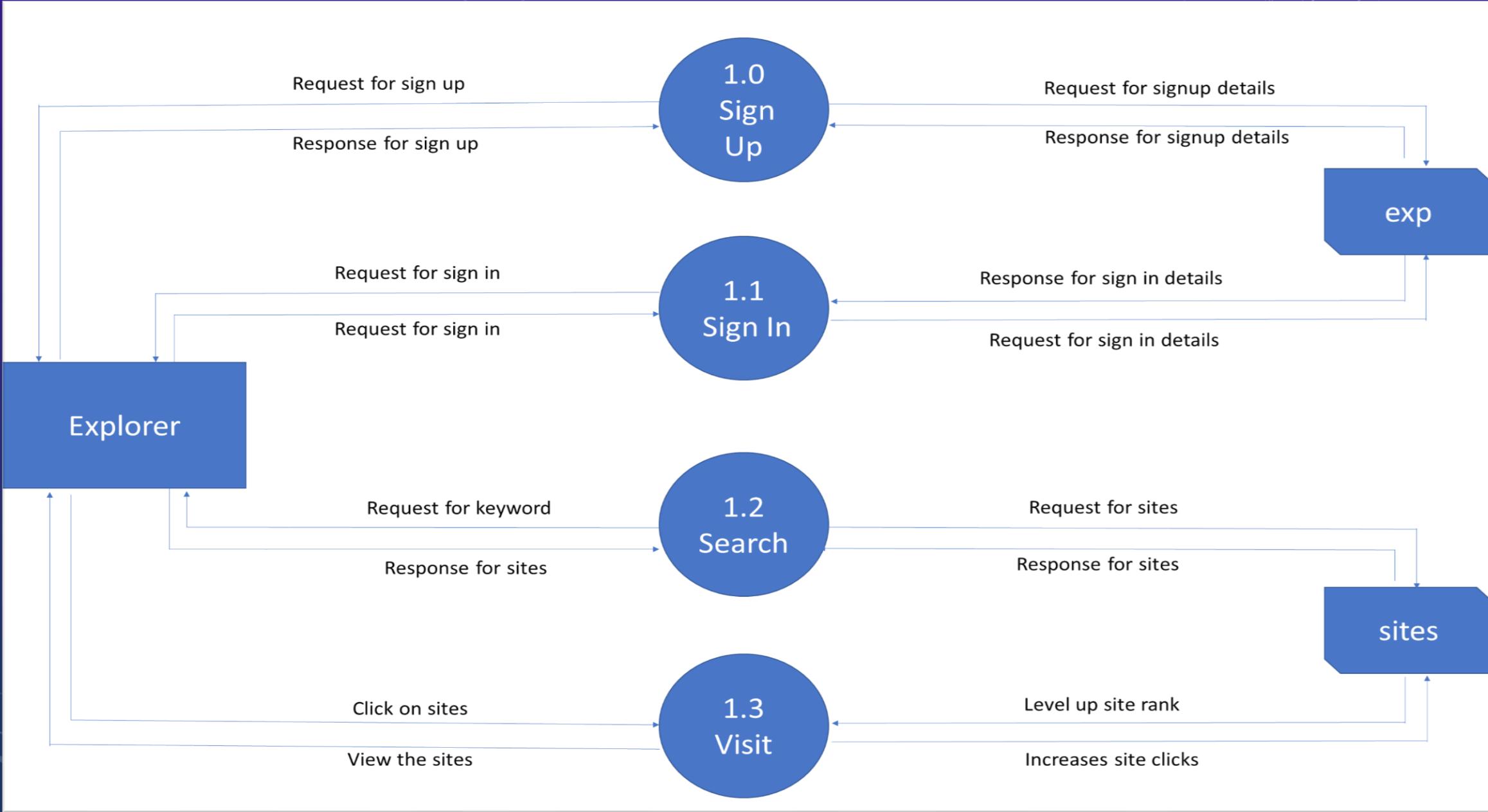
DATA FLOW DIAGRAM :

(CONTEXT LEVEL)



DATA FLOW DIAGRAM [EXPLORER]:



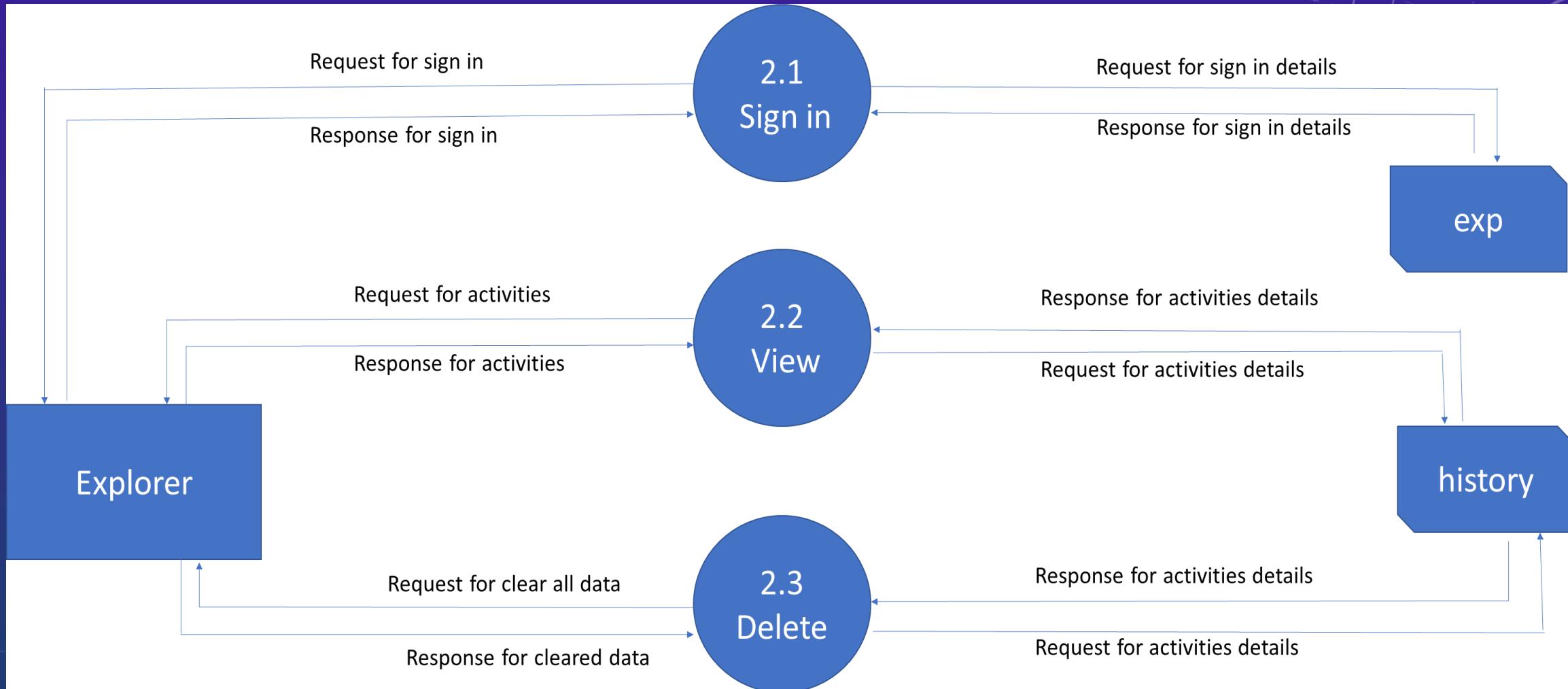


Explorer login details

1.0 Explorer SignUp

Explorer login details

In this process, Explorer user registers and inserts data in exp table

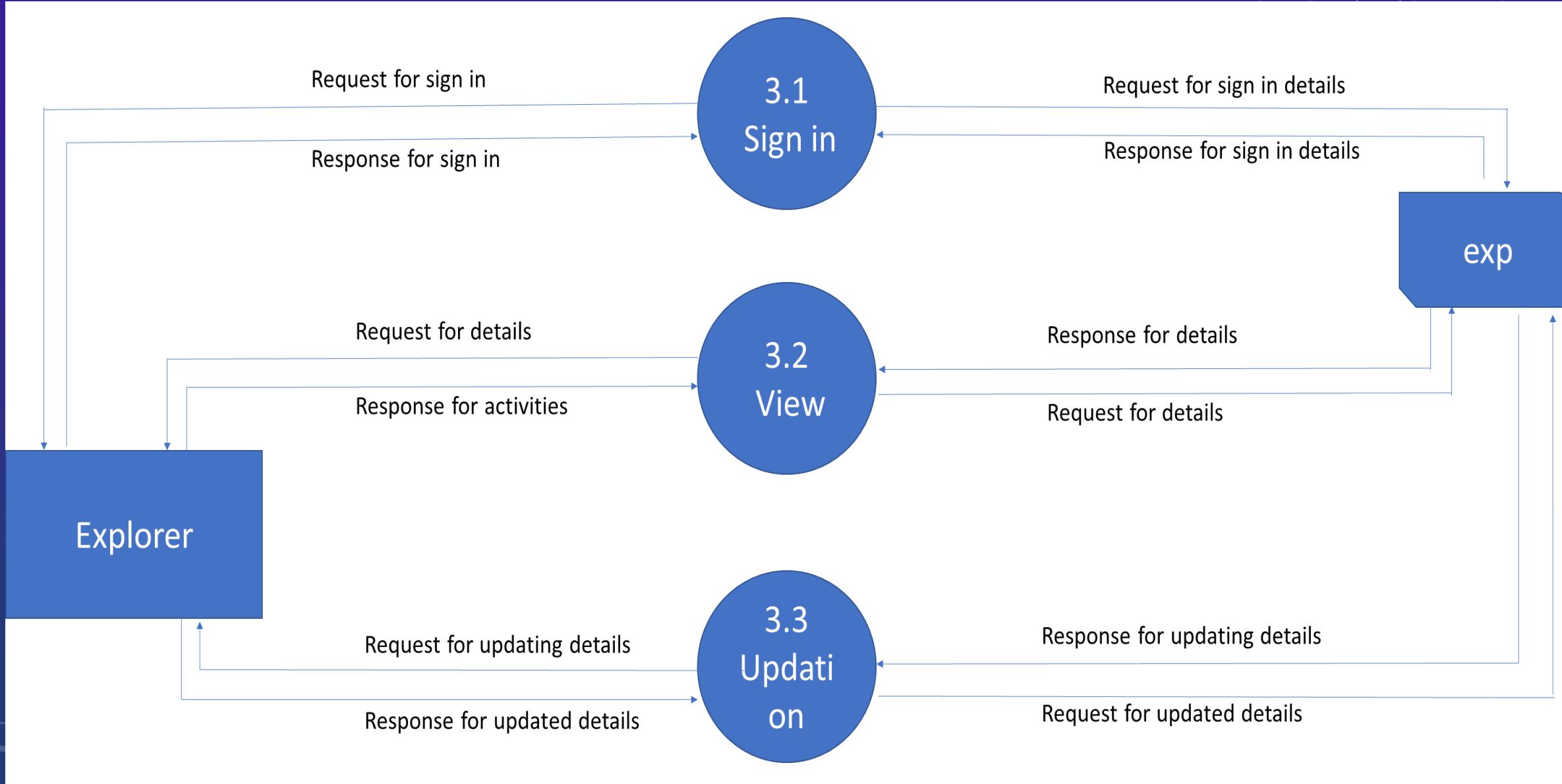


Explorer history details

2.0 Explorer History

Explorer history details

In this process, Explorer user view and clear the history data in history table



Explorer account details

3.0 Explorer Update

Explorer account details

In this process, Explorer user view and update the account details in exp table



DATA FLOW DIAGRAM [DEVELOPER]:

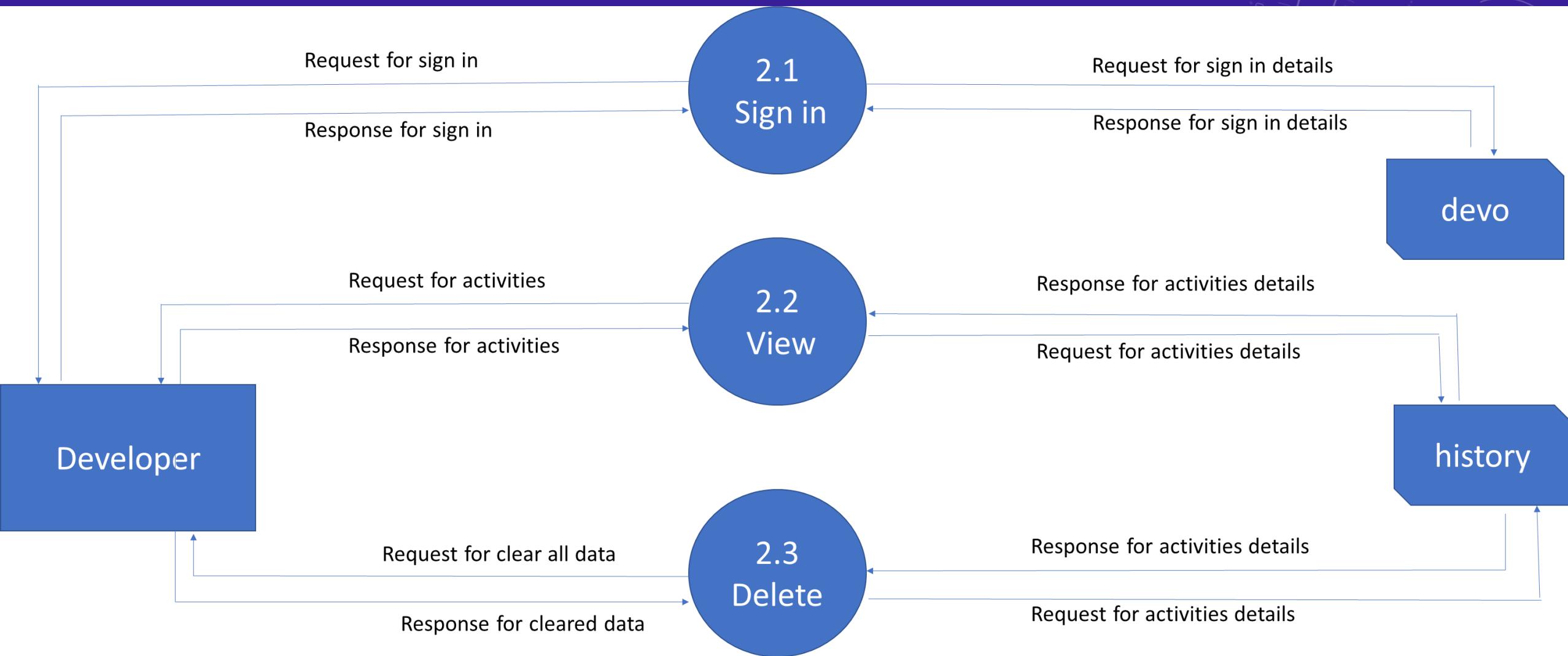


Developer login details

1.0
Develope
r
SignUp

Developer login details

In this process, Developer user registers and inserts data in devo table

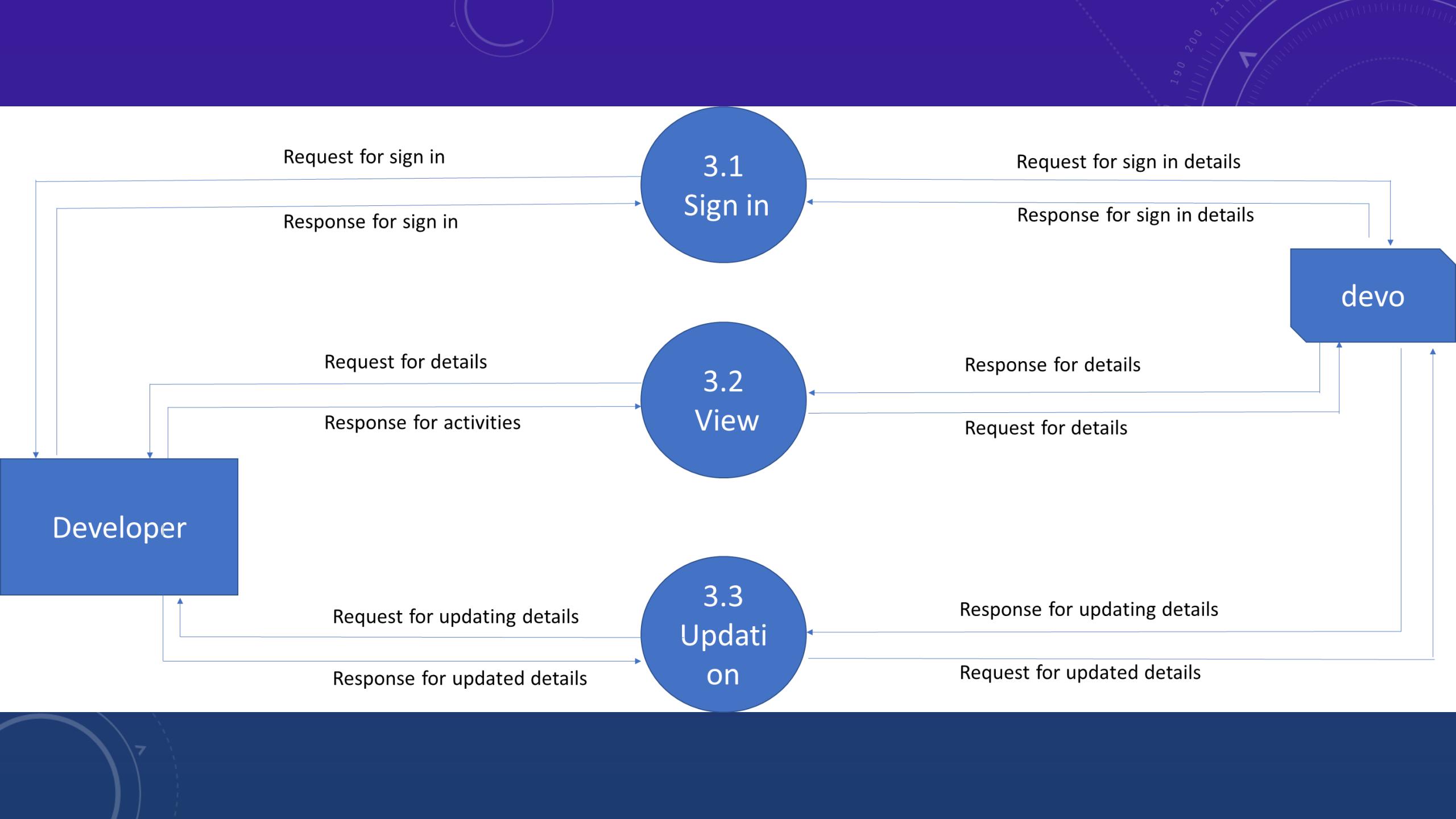


Developer history details

2.0 Develope r History

Developer history details

In this process, Developer user view and clear the history data in history table

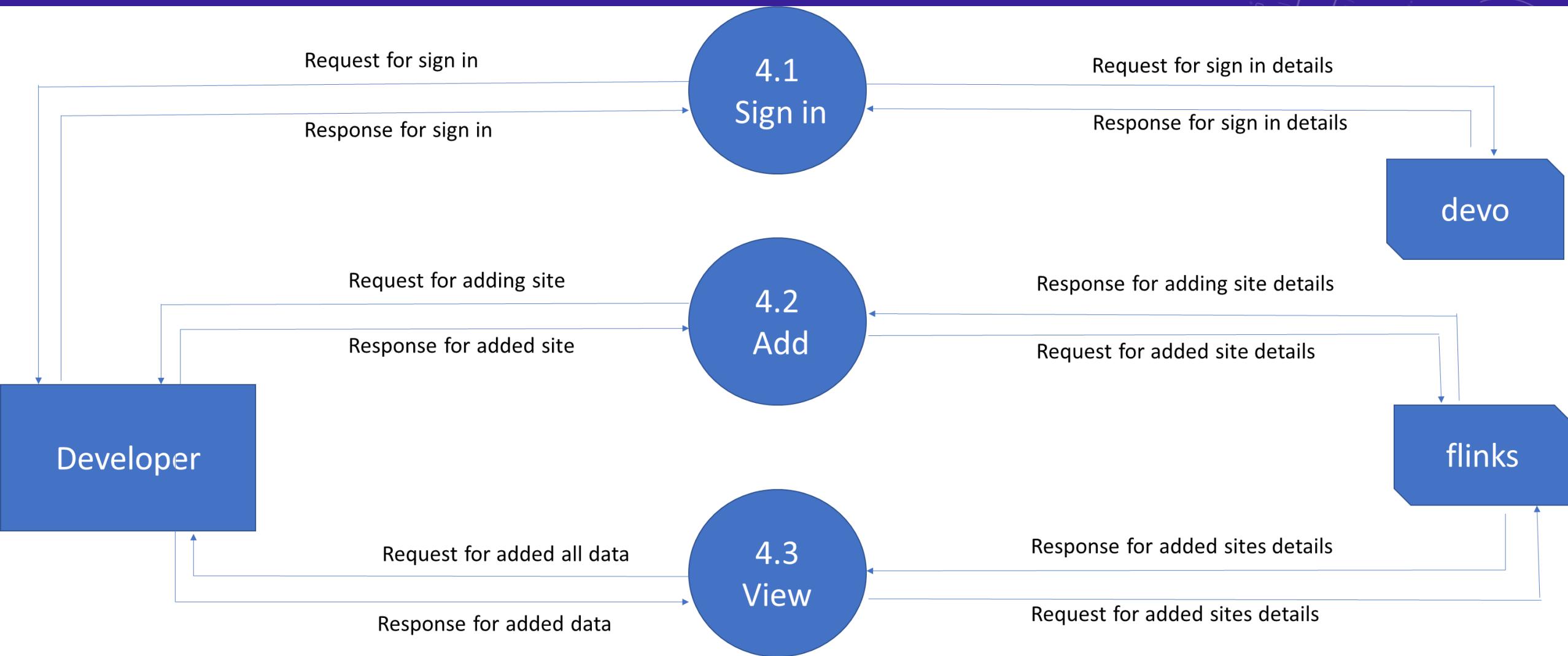


Developer account details

3.0 Develop er Update

Developer account details

In this process, Developer user view and update the account details in devo table

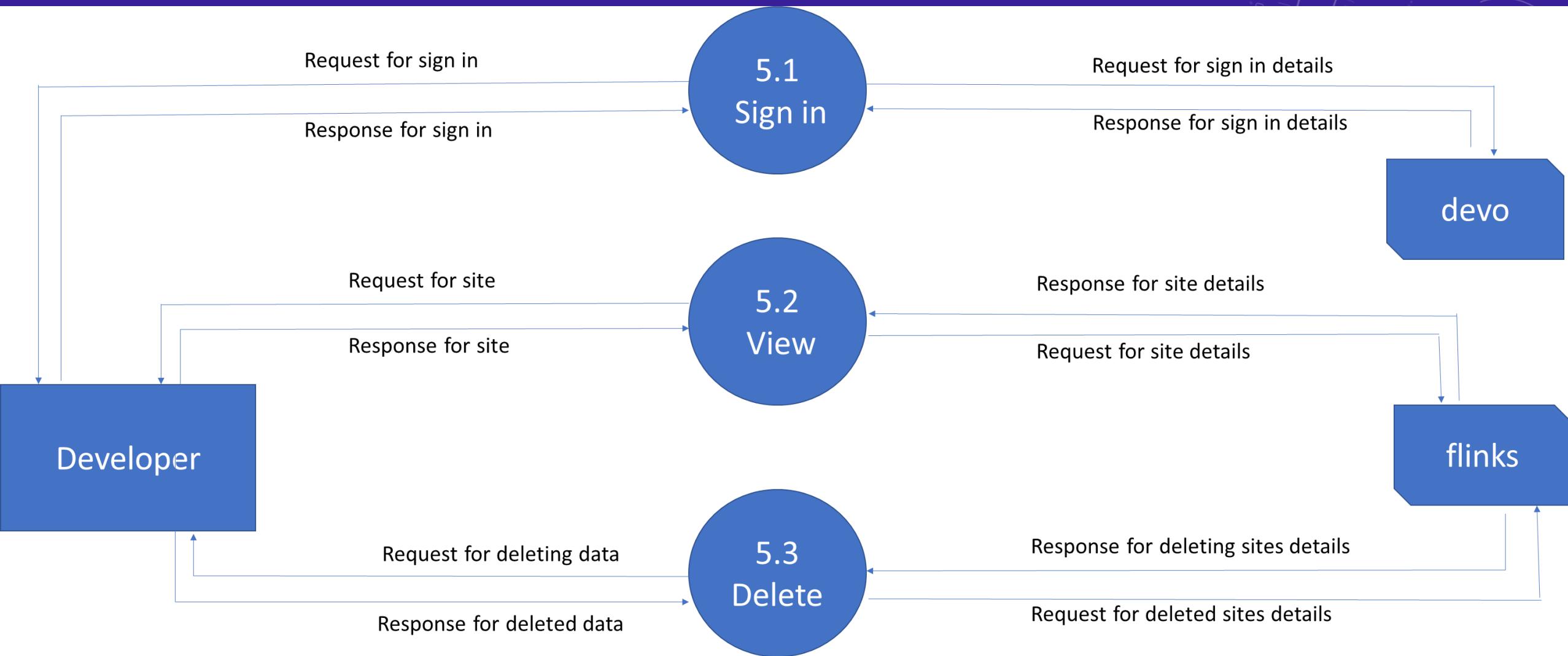


Developer site url

4.0
Develop
er
Add Site

Developer site added

In this process, Developer user adds sites to search engine in flinks table



Developer site history

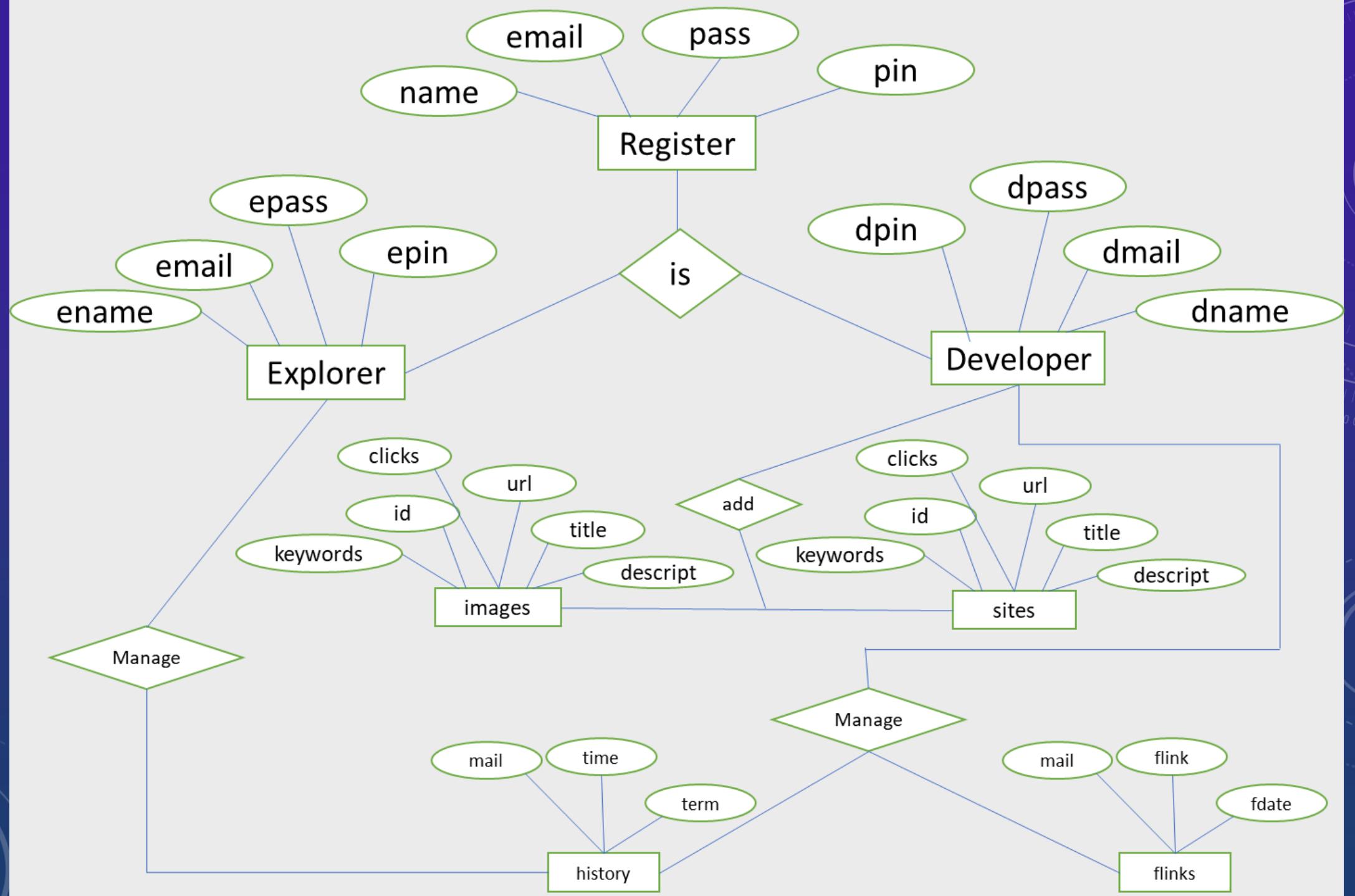
5.0
Develope
r
Add Site

Developer site history deleted

In this process, Developer user view and delete the added sites history in flinks table



ERP DIAGRAM





DATABASE DIAGRAM

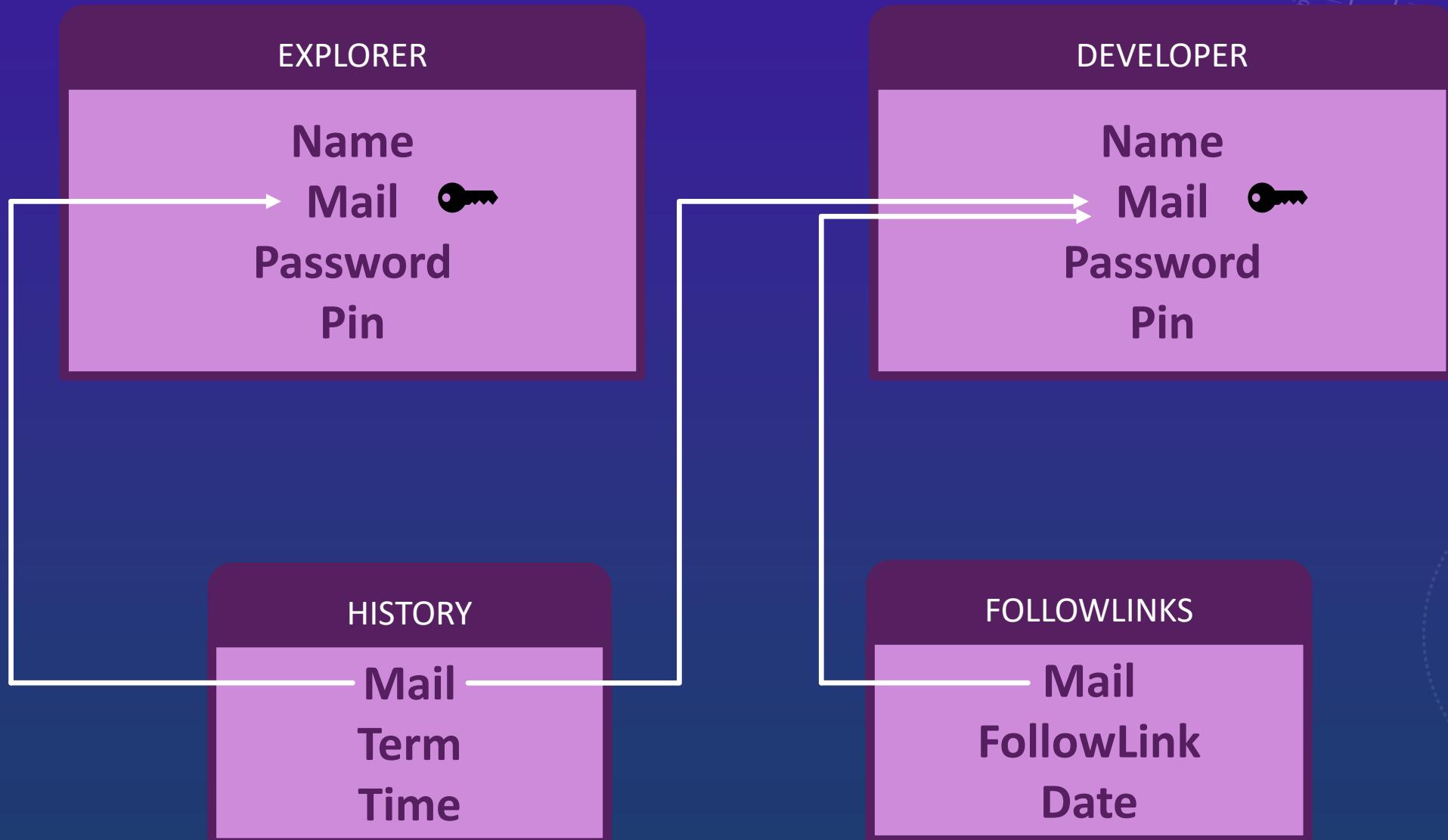
DESIGN VIEW OF TABLES

Images

ID 
SITEURL
IMAGEURL
ALT
TITLE
CLICKS
BROKEN

Sites

ID 
URL
TITLE
DESCRIPTION
KEYWORDS
CLICKS



DATA DESCRIPTION

EXP TABLE

Field No.	Field Name	Data Type	Size	Constraints	Description
1	ename	nvarchar	50	Not Null	Name of the user.
2	email	nvarchar	50	Primary Key	Email of the user.
3	epass	nvarchar	50	Not Null	Password for user.
4	epin	Int	4	Not Null	Recovery PIN for user

DEV TABLE

Field No.	Field Name	Data Type	Size	Constraints	Description
1	dname	nvarchar(512)	50	Not Null	Name of the user.
2	dmail	nvarchar(512)	50	Primary Key	Email of the user.
3	dpass	nvarchar(512)	50	Not Null	Password for user.
4	dpin	Int(4)	4	Not Null	Recovery PIN for user

IMAGES TABLE

Field No.	Field Name	Data Type	Size	Constraints	Description
1	id	int	11	Primary Key	Id for image.
2	siteUrl	nvarchar	512	Not Null	url of sites added by user.
3	Title	nvarchar	512	Not Null	Title of images.
4	imgUrl	nvarchar	512	Not Null	url of images added by user.
5	alt	nvarchar	512	Not Null	alternate used for images
6	clicks	int	11	Not Null	Number clicked on images.
7	broken	tinyint	4	Not Null	Value for broken images.

SITES TABLE

Field No.	Field Name	Data Type	Size	Constraints	Description
1	id	int	11	Primary Key	Id for links.
2	url	nvarchar	512	Not Null	Links added by user.
3	title	nvarchar	512	Not Null	Title of links
4	Description	nvarchar	512	Not Null	Description of links
5	keywords	nvarchar	512	Not Null	Keywords used for links
6	clicks	int	11	Not Null	Number clicked on links.

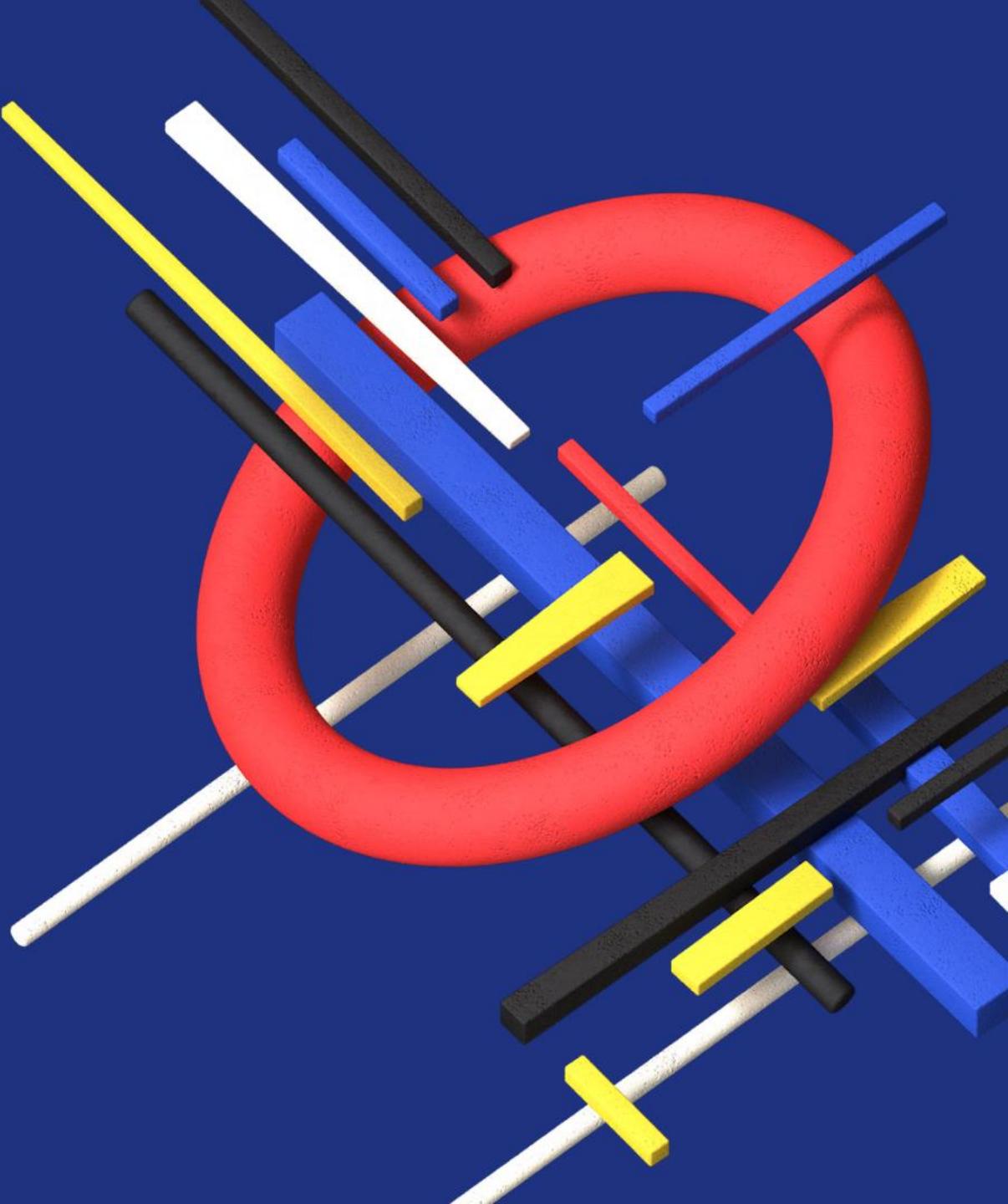
HISTORY TABLE

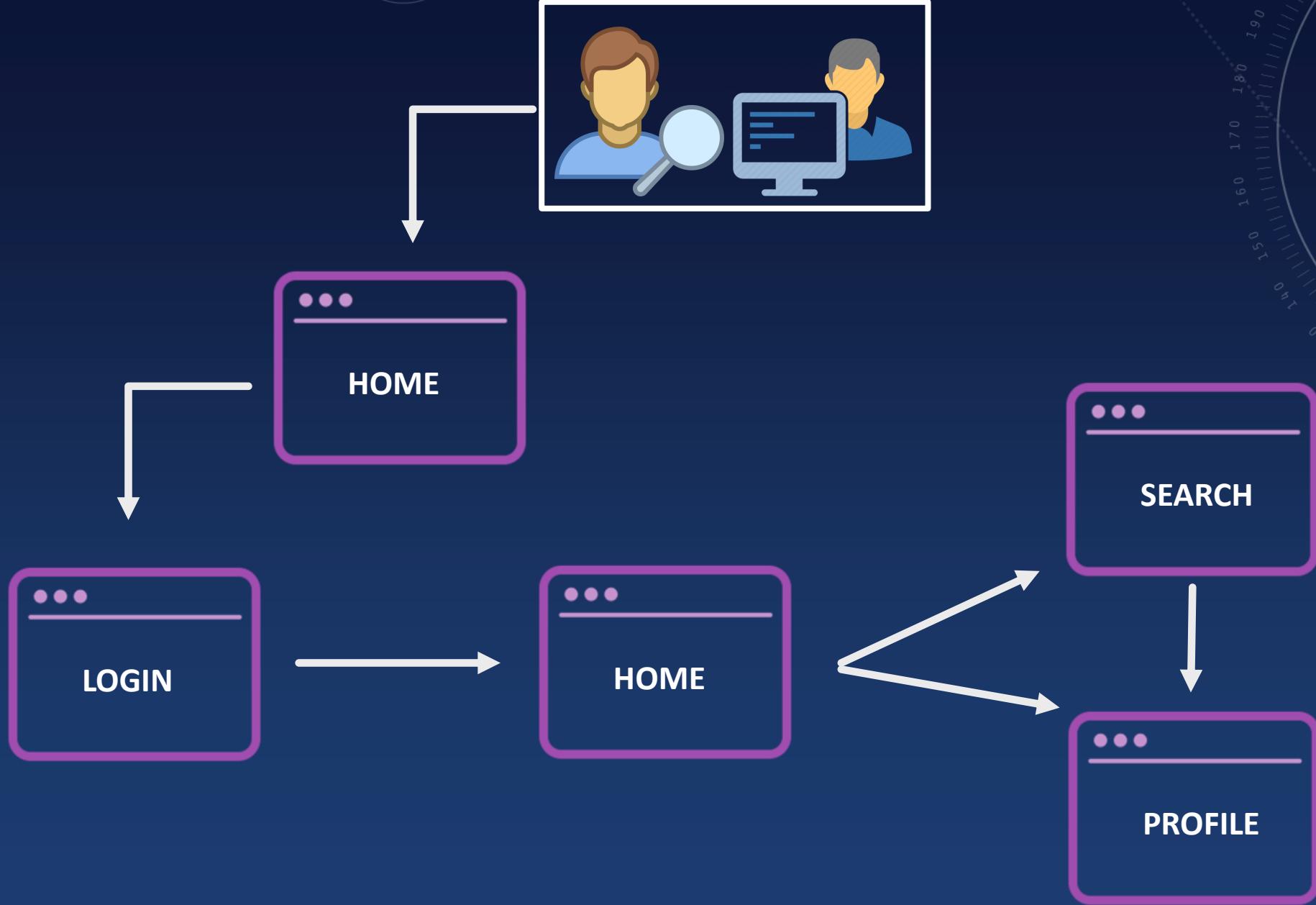
Field No.	Field Name	Data Type	Size	Constraints	Description
1	mail	nvarchar(512)	50	Not Null	Email of the user.
2	term	nvarchar(512)	50	Not Null	Search term by user.
3	time	datetime		Not Null	Timestamp at searched by user

FOLLOWLINKS TABLE

Field No.	Field Name	Data Type	Size	Constraints	Description
1	mail	nvarchar(512)	50	Not Null	Email of the user.
2	Flink	nvarchar(512)	50	Not Null	Links added by user.
3	fdate	datetime		Not Null	Timestamp at links are added by user

SYSTEM UI





REGISTRATION PAGE

New Member

Enter name

Enter email

Enter Password

Confirm Password

Enter PIN

Confirm PIN

Explorer

Sign Up

Existing User [Sign In](#)

Member Login

 Enter email Password Explorer ▾ Sign In[forgot password ?](#)[New User **Sign Up**](#)

HOME PAGE

Hi Raji



Enter Anything to Search

Search



india



Hi Login

[Sites](#) [Images](#)

144 results found

[India's 100 Richest 2020: Collective Net Worth Of Country's Richest Rose 14% Despite Pandemic](#)
<https://www.forbes.com/india-billionaires/>

[Delhi: Latest News & Updates | Get Breaking News Headlines & Exclusive Stories on Delhi at India.com](#)
<https://www.india.com/news/delhi/>

[Latest News, Breaking News Today - Bollywood, Cricket, Business, Politics - IndiaToday | IndiaToday](#)
<https://www.indiatoday.in>

[West Bengal](#)
<https://www.india.com/west-bengal/>

[Tamil Nadu](#)
<https://www.india.com/tamil-nadu/>

[Telangana](#)
<https://www.india.com/telangana/>

[Lifestyle: Latest Lifestyle News | Love, Sex, Relationship & Health Advice | Fashion Trends & Beauty Tips | Festivals, Food & Travel Articles | TV & Celeb Gossips](#)
| India.com News
<https://www.india.com/lifestyle/>

[Festivals & Events | Latest Articles, News, Photos & Videos on Festivals & Events in India | India.com News](#)
<https://www.india.com/festivals-events/>

KHOOOOOOOOJ

1 2 3 4 5 6 7 8 9 10

IMAGE RESULT PAGE

KHOJ

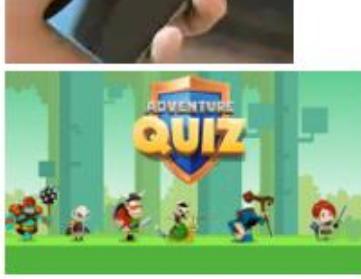
india



Hi Rajj

Sites Images

69 results found





Khoj

Profile

history

Logout

Update Account Details

Rajj

worldtechnics99@gmail.com

Enter Password or New

Enter PIN

Confirm PIN

Confirm Password

Explorer

Update

Only Email Can't Be Updated ! and
Password is required !

[History page for Explorer User to View or Clear Searched Activity]

[Khoj](#)[Profile](#)[history](#)[Logout](#)

SR.	Searched Activity	Time
1	cricket	2021-04-18 09:47:00
2	bitcoin	2021-04-18 09:47:07
3	coin	2021-04-18 09:47:12
4	medium	2021-04-18 09:47:17
5	medium	2021-04-18 09:47:24
6	job	2021-04-18 09:49:55
7	job	2021-04-18 09:49:58
8	score	2021-04-18 09:53:49
9	live score	2021-04-18 09:54:12
10	live score	2021-04-18 09:54:15
11	news	2021-04-18 09:54:32
12	news	2021-04-18 09:54:34
13	india	2021-04-20 05:38:45
14	india	2021-04-20 05:38:47
15	india	2021-04-20 05:39:02
16	india	2021-04-20 05:39:14
17	news	2021-04-20 05:59:32
18	news	2021-04-20 05:59:35
19	india	2021-04-20 05:59:42
20	india	2021-04-20 05:59:55
21	india	2021-04-20 06:00:24
22	india	2021-04-27 08:50:13
23	india	2021-04-27 08:50:21

[Clear History](#)

Khoj

Profile

Consol

Sites

history

Logout

Update Account Details

abc

abc@gmail.com

Enter Password or New

Confirm Password

Enter PIN

Confirm PIN

Developer

Update

Only Email Cant Be Updated ! and
Password is required !

[History page for Developer User to View or Clear Searched Activity]

[Khoj](#)[Profile](#)[Consol](#)[Sites](#)[history](#)[Logout](#)

SR.	Searched Activity	Time
1	hello	2021-04-18 08:50:06
2	is	2021-04-18 08:50:18
3	is	2021-04-18 08:50:23
4	cricket	2021-04-18 09:44:49
5	cricket	2021-04-18 09:45:00
6	cricket	2021-04-18 09:45:08
7	cricket	2021-04-18 09:45:12

[Clear History](#)

Khoj

Profile

Consol

Sites

history

Logout

Add Website To Khoj

<https://www.xyz.com>

Start Crawl

Be Patience, wait till the next page got loaded, it will takes a while.

Khoj

Profile

Consol

Sites

history

Logout

SR.	Crawled Activity	Time
1	https://www.cricbuzz.com	2021-04-18 09:41:35
2	https://news.google.com	2021-04-24 08:05:17
3	https://www.aajtak.in	2021-04-24 01:16:47
4	https://www.youtube.com	2021-04-25 06:29:32
5	https://www.cricbuzz.com	2021-04-25 06:32:51
6	https://www.cricbuzz.com	2021-04-25 06:33:27
7	https://www.cricbuzz.com	2021-04-25 06:37:14

Clear Consol History

LIMITATIONS :



Internet Required



Static Indexer



Simple Ranking



Not easy to Update



Future Scopes & Chellanges

FUTURE ENHANCEMENT :



Voice Recognition



Keywords Suggestion

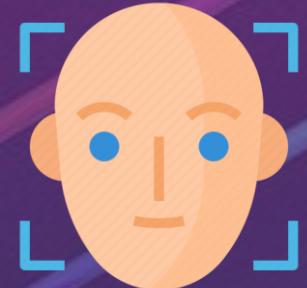


Image Recognition



Local Results



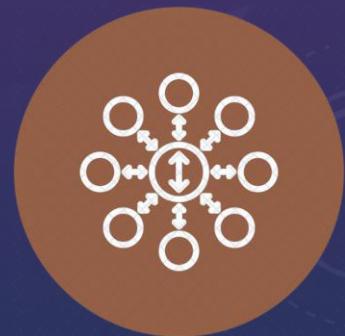
Page Ranking



Search Results



Bounce Timing



Auto Index Update

Références



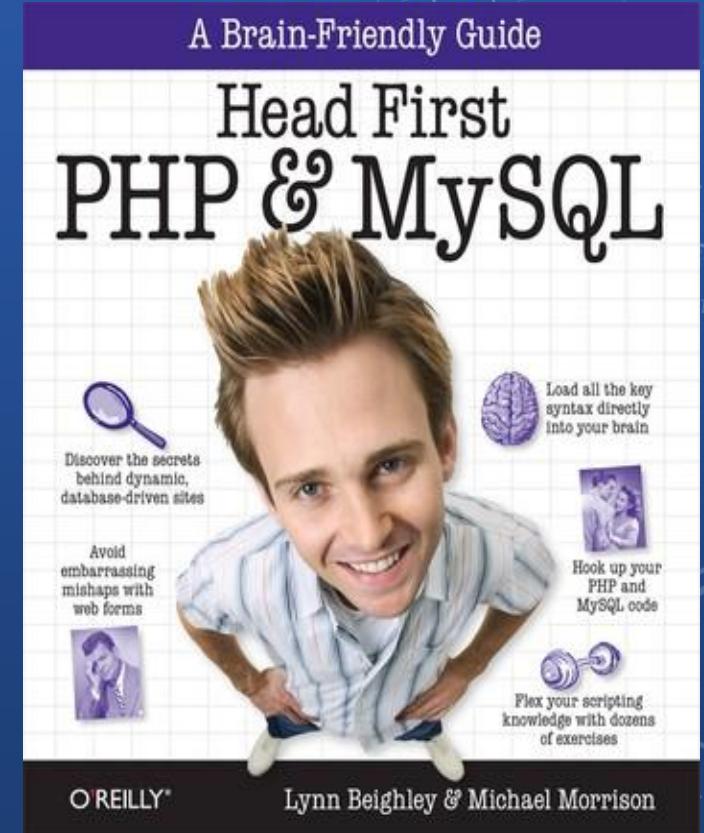
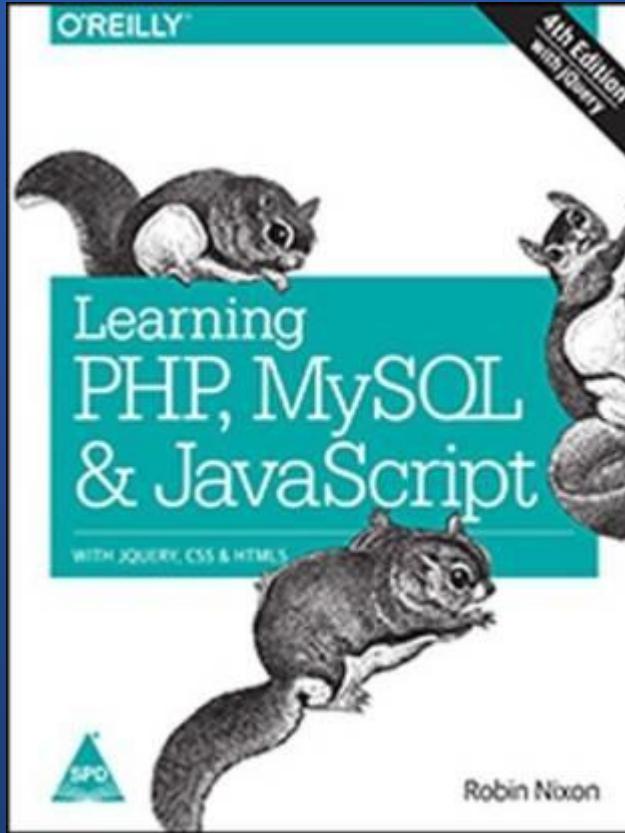
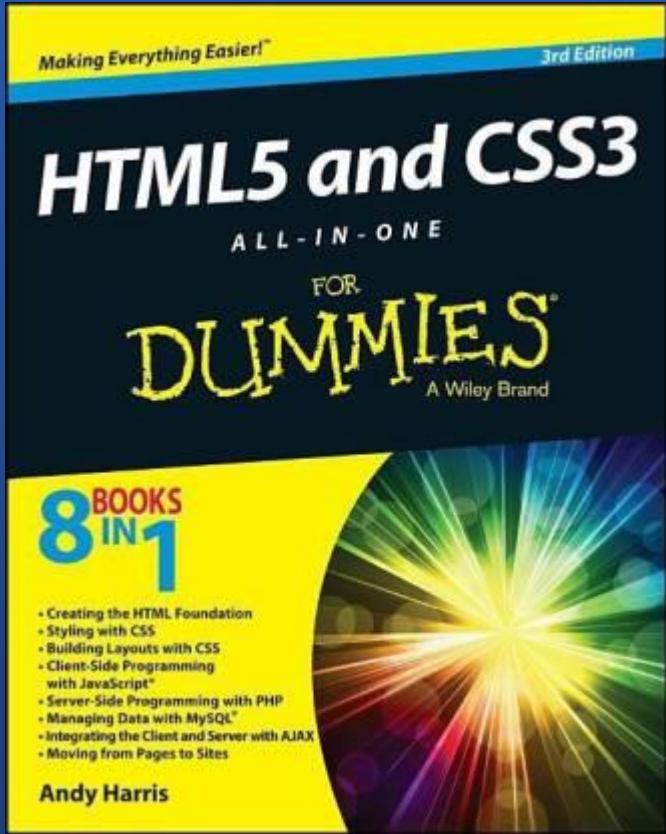
WEBOGRAPHY :



WIKIPEDIA



BIBLIOGRAPHY :



**THANKS YOU
FUTURE IS AWESOME**