UNIT 5

Introduction to SEO, How to use internet &search engines, Search engine and its working pattern, SEO phases, On-page and off-page optimization, SEO Tactics, Introduction to Digital Analytics- Data collection of Web Analytics, Key metrics Impact matrix, Machine learning in google Analytics, Multichannel attribution

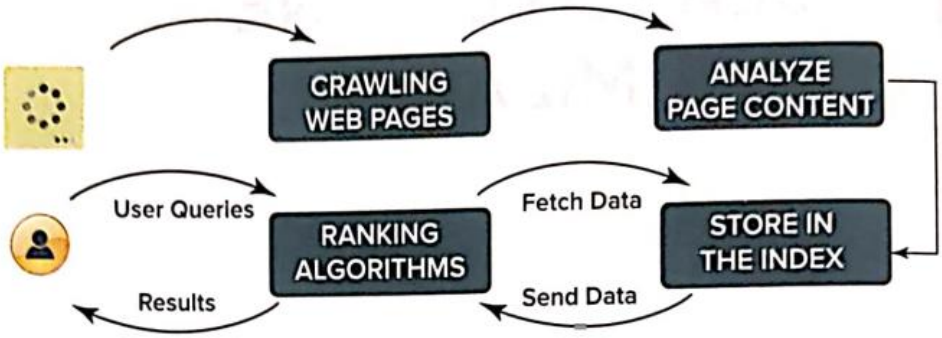
Introduction to Search Engine Optimization

Search engine optimization (SEO) is the process of enhancing the visibility of a website by improving the ranking in the Search Engine Results Page (SERP).

Search engine is web programme designed to retrive or search for information on the web. The search results are usually displayed in aline of results on pages known as the SERP. When auser enters a quesry,search engines dispaly both organic and paid search results. Organic results are natural and unpaid, whereas paid results are paid for - advertisers have to pay to get the sponsored web page link display for a search.

How Search Engines Work :

To show results matching user query, search engines perform many activities are



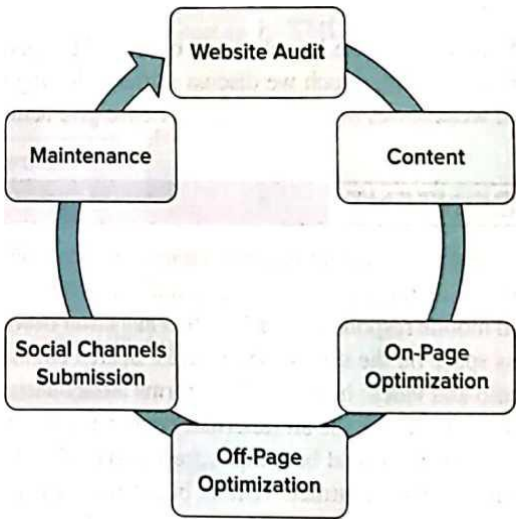
The process starts with web crawling, which is looking for the content available on the web.[1] Websites are crawled by automated bots or spiders or crawlers that are software programs that visit each web page. You may wonder how crawlers will know which domains to visit? Crawlers get information about registered domain names and their IP addresses from Internet Corporation for Assigned Names and Numbers (ICANN), which is a non-profit organization responsible for assigning unique identifiers such as domain names and IP addresses for the entire Internet. CrawOing is done periodically depending on the frequency that webmaster requests as websites keep updating their content.

Search engines then take all the data that has been crawled and place it in large data centres with thousands of petabytes worth of drives. After that, search engines index the data, which is a classification of pages into categories by identifying the keywords that best describe the page and assigning the page to keywords. Indexing involves many concepts from linguistics, cognitive psychology, mathematics and computer science. Using those concepts, search engines have developed capabilities to index media files such as video, audio and graphics along with the text.

When a search request comes, the search engine processes it, i.e. compare the search query with the indexed pages in the database. Since more than one page will contain the

search query, so the search engine starts calculating the relevancy of each of the pages in its index to the search query. The last step in search engines' processes is retrieving the pages with the highest relevance score on top of the search results and displaying them in the browser.

SEO PHASES :



The process of SEO for marketing involves six multiple steps phases. The starting point is audit so that the marketer knows where the company stands and identify the goals. After having the real- ity check done, marketers must put in efforts to follow best practices of search engines. The first step is to make sure original, relevant, high-quality content is there and is discoverable by the search engines using processes such as submitting a site map or a robot.txt file. Subsequently, on-page optimization has to be done, which is the easy part as it is to be done on the web pages and hence is within the control of the marketer. After that, off-page optimization, which involves backlinks, needs to be done. This is the difficult part of SEO. Subsequently, social sub- mission must be done across social media channels to increase the reach and get the interaction of users. After that, regular maintenance of the website must be done so that it does not fall behind others in SEO.

Website Audit (Phase -1):

The SEO process begins with an audit for a reality check so that we know where we stand. There are many freeresources available on the Internet fordoing an overall audit, such as seositecheckup.com, smallseotools. com, maiesticseo.com They give a score out of 100, which oilers a quick and easily understandable assess- ^nt of site performance on SEO. The target should be to get a score above 80.

Some of the main elements of the Audit are

- **Keyword Position -** For important keywords, what is the position of a website in SERP? https://smallseotools.com/keyword-position/ is an excellent free resource for finding out keyword positions.
- **Site Map** - Sitemap shows the architecture of the site to search engines such as category and deeper pages and hence facilitates crawling and indexing by search engines.

http://seositecheckup.com/tools/sitemap - text free tool. which helps us know whether sitemap exists for asite, or not. you can also check out your sitemap on www.example.com/sitemap.xml

- **Browser, Operating systems,Device Compatibility - It** is important to check if the website is compatible with different browsers, operating systems and screen sizes. It is possible that website functions very well on Chrome but not on Internet Explorer or vice versa, or works on desktop but not on mobile devices. There are hundreds of screen sizes available in the market, therefore websites should be checked for their responsiveness.
- **Backlink Checker -** Search engines use backlinks as an indicator of the authority of the site. Check out how many backlinks are coming from which domains and what is the authority of those domains.
- **Domain Authority -** Many free tools give domain authority of the site based on backlinks, which indicate the likelihood of website coming high in SERP.
- **Keyword Cloud -** Which keywords appear more often and have a greater density **on** the website? Are they the right keywords?
- **Speed Audit -** Website loading speed is one of the important aspects of user experience. A good benchmark is 2 seconds. Many users close the site if it takes more than 3 seconds to load. Two popular tools for measuring site speed are Google Page Insights and Pingdom. They give the score out of 100. A score of 85 and above indicates good performance.

Those are only the key elements of an audit. Apart from them, several other aspects are there in an overall SEO check-up, which we discuss in the following sections. This audit will help in identifying the strengths and weaknesses of the website and hence give actionable insights.

Content - (2nd Phase):

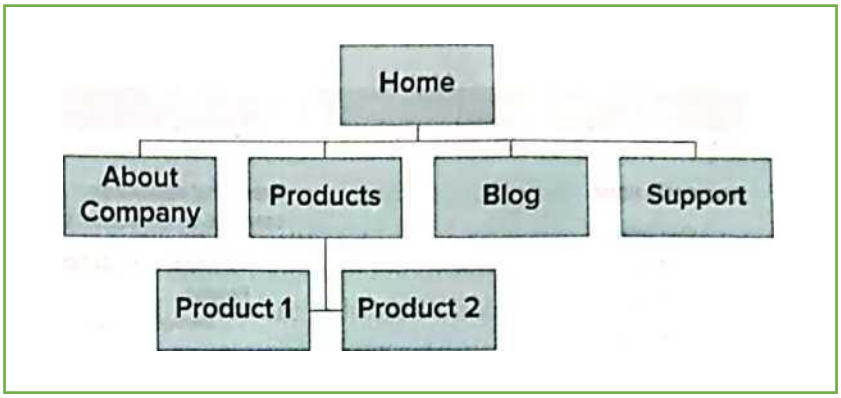
Content here refers to all the information contained on a web page. The page content can be displayed in the form of text, hyperlinks, images, audio, animation or videos. The text has advantages of speed, accessibility and mobile responsiveness. The text has faster download capabilities from the server than images as text takes less space on the server than images. Search engines have a limited ability to understand images, animation, audio and video; however, these forms attract users. In these cases, to determine page content search engine use file names or the alt (alternate) tag, which we will cover later in the chapter.

Content[3] should be unique, fresh, original and should add value to the target audience. Offering quality content not only attracts visitors but also attracts other websites to link to your site, thus enhancing your authority. Improving the content on your website should be a priority, regardless of the website type. Se\cral tools are available to check for plagiarism, www.duplichecker.com provides a plagiarism and quality check å web content, including proofreading and editing, smallseotools.com/plagiarism-checker/ is another plagiarism checker tool to make sure that content is original and unique.

ROBOTS.TXT :
After having agood content, it's important to ensure that your content is crawled and indexed. Robots.tct file is atext file that helps to regulate web robot behaviour and search engine indexing. It must be stored in root of any website.

Site Maps:

A Sitemap is an archive of every page on your website.can visualize your website as a trace with the home page as the trunk and category pages as branches and product pages as sub-branches. Crawlers may come and crawl only the home page (trunk) and a few category pages (branches) and go away as they may not know that deeper pages (branches and sub-branches) exist. To avoid this situation, it's best to create a sitemap and submit it to 'Search Console' so that search engines know all the URLs of the site. The sitemaps can be generated from tools in either XML or HTML format. The XML format is used for indexing by spiders. Apart from the list of URLs, it also has meta-data about the impor- tance of URL, frequency of changes and its relationship with other pages. HTML is for users, and usually y°u find a sitemap link in the footer part of the website. When it is clicked, it redirects to a page that has all URLs of web pages. For example, if you are unable to find a product page on the site, then you can locate it using the sitemap. If there are a lot of web pages on your website, then you can put links to major categories. An ideal way to put your sitemap is on the root location as www.example.com/sitemap.xml and not like www.example.com/..../sitemap.xml. An example of a sitemap in XML format is given in



ON-PAGE OPTIMIZATION (3RD PHASE) :

There are several on-page factors that affect search engine rankings.

Technical Elements : A good SEO roadmap is built on astrong technical foundation. Unleass the core technical components of the website are in place, all other SEO efforts will be in vain. Important elements are

1) Site Performance : Site performance is about the page speed, which is describe in terms of time aload. It is tough to hold a visitors on your website when the web page does not load with is afew seconds. The bench mark should be 2 seconds. Slow website tend to have higher bounce rates and lower average time per visit on the page. A slow webpage will require more time and hence will reduce your crawl budget. There are many tactics to improve site performance as follows

i) Enable Compressions by minifying HTML,CSS,JvaScript - minifying here refers to the

removal of all unnecessary characters from the source code without changing its actual functionality. Ideally HTML, CSS or JavaScript larger than 150 bytes need to be compressed on the server.

ii) Compress Images - generally, image uploads on the web are different from what you see when images are captured with a camera or created using image editor tools like photoshop. the captured images are usually large in size and high in resolution, and if they are uploaded on the server and added to any webpage in its original form, then the user may experience a long waiting time for the web page to laod. A good practice would be to compress images to reduce their size yet maintaining consistent quality so that the page loads faster and the image doesn't get blurred on the browser.

iii) **Reduce Redirects -** Redirect is a way to send users to a different URL from the one they initially the browser to open. It is also known as the URL forward. When visitor experiences redirections over the website, there is a waiting time for HTTP Request-Response cycle to finish. Reducing these redirects can help to improve site performance.

Page load speed is one of the most important aspects of user experience. Page Speed Insights gives a detailed report about the time taken by different elements of the web page such as image, text, CSS to load. It gives a score ranging from 0 to 100 points. A high score is better, and a score of 85 or above shows that the web page is performing well. A Google page speed snippet.

Domains :

Domain names are Internet addresses of websites. Domains have extensions such as .com, .in, .org, etc.They are purchased from registrars such as GoDaddy or BigRock who get authorization of selling available domains by ICANN. There are a few points that one needs to keep in mind while registering any domain.

1. Domain name Memorability : They are many domains names availabile, but selecting one is a difficult task. A domain name should be short,catchy,easy to remember,spell and type. You can conduct an informal survey by giving a few options to people and later see which ones are the most memorable.

2. **Keyword-Rich Domains -** Having your keywords in your domain name can increase click-through rates. It also gives users an idea about your business. However, earlier keyword-rich domain name would contribute to SEO, but its importance for SEO has decreased over time.

3. **Subdomains and Subfolders -** Subdomains and subfolders are second-level parts of domains (top-level) that are free to be created under any domain that a webmaster could access. There is a separate section in this chapter later dealing with subdomains and subfolders and identifying the belter ones for your website.

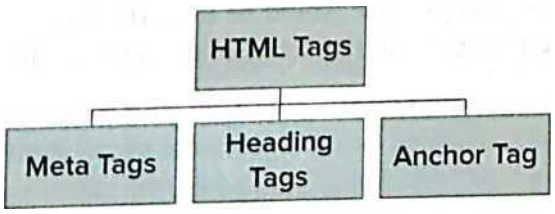
404 Error/500 Error

404 error is an error message that appears when the web page that the user is trying to reach could not be found on the server. When any web page that the user is trying to access is not available on that website server then the URL is redirected automatically to the 404 error page. It is important to set up the 404 error provide to give visitors navigational options to let them stay on to the website. This 404 error page ideally should have a link back to your root page and could also provide some popular content on your website. Typical 404 error page would contain: User notification that the page does not exist.

Search box
Homepage link

Another error that one can encounter is the '500 error'. These errors are internal server error shown when any unexpected conditions occur. The server here could not be more specific on why the problem has occurred and the solution to it. This error can occur due to server hardware or a software code issue publisher should avoid the 500 error as it gives a bad impression to the visitor as well as the search engine when 500 error occur the users can refresh the web page, clear the browser's cache or they can even try deleting browser's cookies .

HTML Tags : Hyper Text Markup Language is markup language commonly used to create web pages. To create webpages in HTML, we use different tags so that web browsers can read the code and process them to display on your screen It provides a means to create structures webpages that browsers can understand. Meta tags are written that describes the pages, count and does not appear on the front end to users. if only exists in the html and usually in the <head> scope.



Meta Tags

Meta Title - While creating any HTML document you often indicate page titles using title tag on your web page. This page title is visible on the browser tab. A title tag describes the topic of any web page. It is denoted by <title> and should be placed within the scope of the <head> tag of any HTML web page. Ideally, there should be a unique title for each page on any website. When the search engine displays any website or web page, it uses page titles in the snippet, as shown in Figure 10.11. Therefore, it is always recommended to use short and informative titles. If a title is too long, the search engine will show only a portion of it in the search result. The tliree dots at the end of the title indicates that the page title is longer than the space meant for the search result (60 characters) and search engines have clipped the title.

Meta Keywords - Meta keywords are used to define the content of a web page by providing a bunch of keywords or tags specific to that web page's content. Most

search engines (Google and Yahoo!) penalize user for abusing this function. The best practice is to use keywords in all HTML and meta tags such as title, description, alt tags, anchor text, URL.

Meta Description - Meta description is used to describe a web page that gives search engines a summary of that page. It can be written in a sentence or two or even in paragraphs if needed. The meta description is important as search engines may use them as snippets of your web pages on the search result page. Ideally it should be within a 150-character limit to fit in the snippet. Search engines may choose to use a relevant section of your web page's text if it matches the user's query. In case the search engine cannot find a good selection of text to use in the snippet, page description would be used.

Heading Tags:

Heading tags help to dfine page structure and allow users to scan a page to find what they are looking for. A good practice would be to start a page with H1 and then follow with other heading tags depending on the content. Atotal of six headings tags have been used.

Anchor Tag

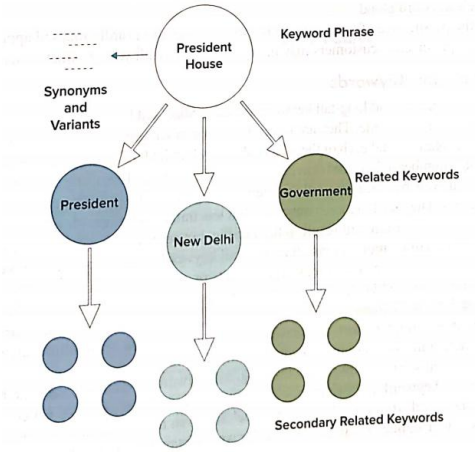
Anchor Text — Anchor text is highlighted hypertext link, which can be an internal website link or external source. Appropriate anchor text helps the reader to learn content associations. For search engines, link relevancy is one of the factors that determines web page rank.

The best practice is to use rich keywords in the anchor text, which is related to the content of the landing page so that the user can anticipate the nature of the landing page. The content around anchor text is also important and should signify lhetheme to which anchor text belongs naturalty. It ts not a good practice use anchor texts such as 'click here'.

IMAGE/VIDEO OPTIMIZATION

On the web page the 'Alt' attribute provides image related information. 'Alt' basically stands for alternate, where we describe an image in a textual form. Every image should have a distinct filename and associated text of the image in 'alt' attribute which would allow specifying which images is for what. It also helps visitors who cannot access the image. In the case of inaccesibility, screen reader whould be able to identify the corrosponding 'alt' text and speak the text mentioned in the 'alt' attribute. To help search engines to understand the context of the used image we moust use the 'alt' tag.

Keywords : Keywords are the words and pharses in the content that makes it posible for users to find websites by using the search engines. While writing any content digital marketers should focus on building a theme out of the content. Themes are formed through relationships between concepts and gropus of keywords. Closely related keyword pharses strengthen the topically of any web page.



Choice of Keywords : careful research needs to be done for choosing the right keywords for optimizing the website.

Long Tail and shor Tail keywords :
The web is made up of short-tail and long tail keyword searches. Short tail leywords are very few, and each one has millions of monthly search volume. The are typically generic or category keywords. Most of the keywords are the long tail. which is no longer phrses and each of them has only a few hundreds pf monthly search volume.

RSS FEEDS : RSS(Rich Site Summary or Really simple Syndication) os an XML code that contains recent information updates, RSS feeds give auser the ability t opt for asubscription very similar to newspaper subscription.

OFF-PAGE OPTIMIZATION (4^{TH} phase) :

Initially SEO was mainly on-page. But since it was under control of webmasters, some started absuing it by stuffing keywords. Hence search engines introduced the concept of off-page optimization.
The major part of off-page optimization includes

BACKLINK :
A backlink is the process of getting hyperlinks from external pages that are not owned by you to link to aweb page of your website. Backlinks will help in building the authority of the website. Each link to awebpage is counted as avote for that page and the page getting most votes win. The link represents an ‘editorial endorsment’ of a web document.Building backlinks in the important and challenging activity in SEO. Only quality backlink will help in search engine ranking.High quality backlink comes from high quality websites that are trustworthy and have a high reputation. Source Diversity and Source Indenpendence are have a backlinks.

Authority and HUBS : The most important obejctive of off-page actiivities is to build the authority of the website. the concept of authority has been borrowed from academia. The quality of a scholarly paper is judged by how many citations the paper is judges by how many citations tha paper has receipved.

Blog Post/Commentings
Press Release
Directories/Classifieds
Forums
Article promotion and syndication
Avoid unnatural links

Social media Reach :
Social media is becoming increasingly important in SEO. Search engine consider social signals such as like, share,retweet as user feedback. content that is engaging and getting organic traction is considered to be good quality by the search engine.

Video Creation and submission:

Marketers create videos as they are richer format . But most only upload the videos without optimizing them for search and discoverability you must do the following for video optimization.

Rich Snippets -
Video title
Optimize your description
Transcripts
Leangth
Embedding Options
Informational Options
Informational, Not Just Promotional
Taret-specific

Maintenance :

SEO is not a one-time task as search engine algorithms change constantly. Moreover, SEO must be done regularly for new content that is posted. Also, if SEO is stopped, then the website will start falling behind, and competitors will catch up. To maintain your web presence and stay at the top of SERPs, you must regularly do SEO.

SEO TACTICS

Black Hat SEO

Spamdexing, search engine poisoning, webspam are some commonly known names of black hat SEO. When someone deliberately manipulates indexes of a search engine to improve the ranking of web pages, then we can call that the usage of black hat SEO. The search engine discourages such practices; hence we should avoid using them. It involves several techniques as follows.

Keyword Stuffing - Keyword stuffing[15] is a technique in which web page is loaded with keywords unnecessarily in the meta tags or in content to obtain a rank on search engines. It may lead to a website being penalized by search engines.

Example of Keyword Stufiing:

If *Rajasthani turban* is what you're looking for; then you are definitely in the right place to buy Rajasthani turban. When it comes to *Rajasthani turban,* you won't find a higher-quality selection of *Rajasthani turban* anywhere! Our *Rajasthani turban* experts know how to pick only the best material from the bunch, and we sell this premium-limited stock *Rajasthani turban* right here for you to enjoy *Rajasthani turban.* We guarantee you'll come crawling back to buy our *Rajasthani turban.*

Cookie Stuffing - It comes under illegal affiliate marketing technique, which involves placing an af- filiate third-party tracking cookie from an entirely different website on visitors' browsers without their knowledge. If the user later visits the target website and makes a purchase, the cookie stuffer will be paid a commission. Because the stuffer has not driven traffic from his site to the target site, this technique is illegitimate and can even steal commissions of genuine affiliate marketers as fraudulent cookies write their cookies.

Hidden Text/Links - Text can be hidden in several ways such that it's visible to the search engines but not to the users. An example is a white text on a white background, using CSS to position text off-screen setting the font size to 0, hiding the link by linking only one small character such as a hyphen.Within the code, when someone uses comment tags to hide keywords, link or content, then it also comes under hidden text/link practice. Comments tags are used by developers to give some clues to other coders.

Cloaking - This technique is an attempt to mislead search engines regarding the content served. Delivering content based on IP address does this. When a user is identified as a search engine based on IP address, a different web page is served, and when the user is identified as a human, then a different page is served. Hence while the user may see pornographic content, search engines may see non-pornographic content.

Gateway Pages — Also called doorway pages, designed to create fake pages that are stuffed with content and optumzed for one or two keywords that further link to another landing page. The end-user or visitor will never be able to see doorway pages as they are automatically redirected.

Mirror Site - Process of creating multiple websites with similar content and design hosted on different domains is called site mirroring. It is done to drive traffic to the main site and get backlinks. Search engines consider this as duplicate content and can penalize the site.

Blog Comment Spam - The spammer to get backlinks writes a certain script that targets certain websites. Because of the script, comments appear on multiple websites promoting some content with malicious URL that possibly contains a virus. Figure 10.35 demonstrates a WordPress blog comment spam. By looking at the email addresses of the commenter, we know that it's spam.

' S

Social Networking Spam - It is unwanted spam content appearing on social networking

sites that have user-generated content such as comments, updates, etc. It can be done using fake accounts to send bulk messages or hate speeches, fraudulent reviews, malicious links, etc. When there is a huge number of postings and messages on social networking websites, as shown in Figure 10.36, it is social spam.

Link Farms - Link farms are a group of websites that all hyperlink to each other and hence are formed with the sole objective of getting backlinks and thus improving search engine ranking.

Cybersquatting - Cybersquatting is an act of registering and using an Internet domain name, especially well- known company or brand names to earn profit from the goodwill of some other company.

WHITE HAT SEO

White hat SEO refers to following the search engine rules and policies for doing SEO and adopting optimisation strategies and tactics with a focus on the human audience and not search engines. A comparison between Black Hat and White Hat SEO technique is given below

	Black Hat SEO Technique	White Hat SEO Technique
On-page factor	Hidden text Duplicate content	Titles and meta-data Quality content
Off-page factor	Doorway pages or gateway pages	Guest blogging
Links	Page swapping Link farming	Link building Quality backlinks
Content	Keyword stuffing	Relevant keywords

Digital Analytics

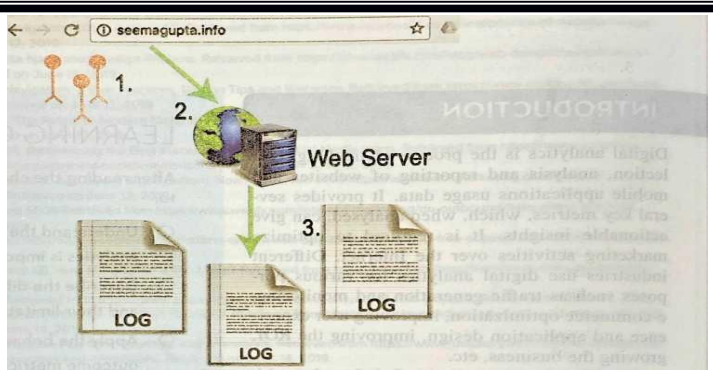
Digital Analytics is the process of tracking, colection, analysis and reporting of websites and mobile applications usage data. It provides several key metrics, which, when analysed, can give actionable insights. It is required to optimize marketing activities over the internet.

Data Collection :

WebLogs:

Weblogs or server logs are one of the oldset data collection techniques that were built for collecting information about server activity. It is automatically created and maintained by the website's server. the log consists of details such as visitor's IP address, data and time stamps, HTTP code,bytes served, referrer, user agent, etc. These weblog details are not publicly available and require admin's access to the server where the website is hosted.

The process of weblogs data collection, is as follows:



1. When the user enters a website URL in a browser, the request is made to the server where the website is hosted through one's Internet service provider (ISP). Hence, there are certain data that their ISP shares with the web servers.[2]
2. Then, the server sends the requested web page using a browser to the user.

At the backend, that web server builds an entry in the weblog for the request. These log files capture data such as page name, IP address, date and time stamps, and browser of the user.

Data in the log files are cluttered as they capture not just user behaviour but also search engine robots' behaviour. Usually, they are more useful for technical purposes (detecting technical glitches such as site down, tracing, etc.) and not for business decisions. Hence one can use parsers (a computer program) to clean the data and make it amenable for marketing purposes.

Benefits of Weblogs

Some of the benefits of using weblogs for marketing purposes are as follows:

1. Every web server has an inbuilt capability to create log files whether we want or not. If any other tool for web analytics such as Google Analytics has not been activated, one can use weblogs to analyse user behaviour on their website. The advantage of weblogs is that they belong to the business and is their own data.
2. Weblogs is a useful source for tracking the behaviour of search engine robots. Robots do not execute JavaScript tags, and thus they leave no trail in other JavaScript-based data capture methods such as Google Analytics. From weblogs, one can know the frequency with which robots are crawling and indexing your site.

Challenges with Weblogs

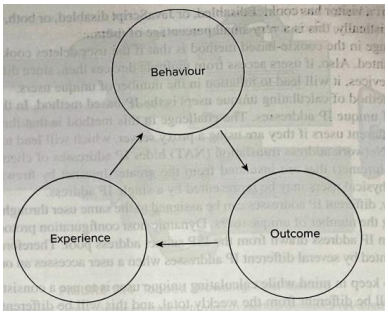
Weblogs face a few challenges, such as

1. **Page caching by ISP :** ISPs keep a temporary copy locally of the page served for a defined period so that when the next time request come on the same page, the request is met locally instead of being sent to the server.This helps in cutting down the timentaken in serving the page, and the page appears to the users to load faster thus enhancing the user experience.
2. **Dynamic IP Addresses:** With the increaseing number of users being assigned with dynamic IP addresses using dynamic Host Configuration Protocol (DHCP), it becomes difficult to identify unique users.
3. **Proxy servers -** It is a network server that behaves like an intermediate between the user's device and server on which the website is hosted. They help in improving the server's performance and its security. When anyone uses proxy servers to access websites, the request will not be sent to the main server.

It is recommended to use weblogs to analyse search engine robots' behaviour to measure the success of your SEO (search engine optimization).

Key Metrics :
Three key metrics in Digital Analytics
1) Behaviour Analysis
2) Outcome Analysis
3) Experience Analysis

The aim of these metricsis to obtain actionable insights from the world of web analytics



Behaviour analysis
Behaviour analysis is what is traditionally called click stream data analysis. It is the process of collecting, analysing and reporting aggregate data about which pages a website visitor visits and in what order. Inferring about **the intent** of the visitors is the aim.

Visits/Sessions

A visit or 'session' is defined as a series of page requests with a gap of no more than 30 minutes between two page requests. When someone visits a web page, it is called a visit or session.

1. **Click versus Visit -** Click is when a person clicks on an ad or a link. Upon clicking, the user visits the web page. In such a scenario, a number of clicks and visits should be the same. But it is not so. Can you guess why? Users may click on the link, but before they land on the page, they may close the browser. Sometimes the page is too slow to load, and hence the user may close the browser. Hence, the click may be 1, but the visit is 0. Some clicks may be unintentional. Hence, the user may close the browser or tab upon realization. Hence there is always some difference between clicks and visits. Typically, there is a drop of 10-15% between clicks and visit. If clicks are 100, visits may be 85. But if the difference is too much, then one must check the load speed of the page.
2. **Unique Visitors -** Unique visitors are the number of different users requesting web pages from a website during a given period, regardless of how often they visit those web pages.

Methods to calculate Unique Visitors
There are primarily two methods for calculating unique users.

1. First is a cookie-based method.[3] A cookie is a small text file placed in the browser of a device. Each cookie has a unique cookie ID. Hence a count of all unique cookie IDs during a given time will be a website's unique visitors, These cookies should be persistent cookies and not session cookies. Persistent cookies have an expiration date and are stored in the computer. Session cookies do not have an expiration date and expire upon closing the session or browser. Google Analytics sets persistent cookies. The information in the cookie is passed on to Google via the JavaScript tracking code installed on the page. So, if a visitor has cookies disabled, or JavaScript disabled, or both, then they are not tr However, statistically, this is a very small percentage of users.

The challenge in the cookie-based method is that if the user deletes cookies, then the unique cannot be counted. Also, if users access from multiple devices then, since different cookies will be set on different devices, it will lead to inflation in the number of unique users.

2. The second method of calculating unique users is the IP-based method. In this case, unique users the number of unique IP addresses. The challenge in this method is that the same IP address c:: assigned to different users if they are using a proxy server, which will lead to deflating the number^ unique users. Network address translation (NAT) hides IP addresses of clients operating from a subnet (e.g. intranet) that are insulated from the greater Internet by firewall/gateway. Therefore multitude of physical users may be represented by a single IP address.

Alternatively, different IP addresses can be assigned to the same user through dynamic IPs, which will lead to inflating the number of unique users. Dynamic host configuration protocol (DHCP) dynamically assigns client an IP address drawn from the ISP subnet address pool. Therefore, the same physical user may be represented by several different IP addresses when a user accesses an online resource.

The important thing to keep in mind while calculating unique user is to use a consistent time period. Sum of daily unique visitor will be different from the weekly total, and this will be different from the monthly total just because of the calculation used.

According to ComScore, the deletion of cookies alone contributes to 2.5-fold inflation of unique visitor statistics.[4] Additional cookie inflation comes from the usage of multiple computers, devices and locations to access the Internet. Hence both unique IP address and cookies overestimate unique visitors.

、 Time on Site ,

■ The time on site metric indicates engagement of the visitor. The more is time spent, the higher is the stickiness of the site. But it should not be considered as a thumb rule for all types of websites. For some sites, less time may be an indicator of better user experience. For example, for an FAQ page, less time may be better; whereas for a blog, more time on site is better. Hence for interpreting time on site,

one should consider the nature of the web page.

Calculating time on site

Entries in the log files are:

Click 1: Home Page 0900 hrs.

Click 2: Product Page 0901 hrs.

In the given situation, time spent on the home page would be 0900-0901, hence 1 minute.

Since the method of calculating time spent is by subtraction, time spent on single-page visits is not includ ed as the time calculated for them is 0. Another interesting fact is that the last page time is also not calculatec

There are always hacks available to record the time spent on the last page.

a. Tabbed Browsing

All of us open multiple tabs while browsing. How is the time calculated in tabbed browsing? There are tv ways to calculate this, as follows.

The First method

Outcomes: Assuming one session for each of the two tabs in the browser,

Let's take Tp as Time of Page,
Hence the total time spent on two tabs is 5+ 6= 11 minutes according to this method.

The second method

The second method is called the linearization method of calculating time spent. These are visits organised by time stamps.
Outcomes: One session
Ts(session duration) = 7 minuits
Note : Google Analytics uses the second method called "linearization' Method.

Page Views : Page views are the number of pages viewed or requested by user. Every unique URL is apage. It is also refered to as 'Depth of Visit'. It is very applicable to content websites. More page views mean more engagement with the visitors. One cal calculate an average number of page views per visitors. The metric will become actionable when one segments the page views according to trafic sources.

Bounce Rate: Bounce rate is the percentage of single-page visits. They are visits in which the users leaves from the landing page without interacting with the page. if the user interacts by playing a video or answering a poll or using a flicker or closing a pop-up, then it will not be counted as a bounce.

The visitors can possible add to the bounce rate by

- Clicking on the link to a different website
- choosing the 'Back' option to leave the website
- Closing an open window or tab
- Typing a new link
- Session timed out

Heat Map Analysis :

A heat map is the graphical representation of avisitor's engagement on the website. Heat map depicts the engagement section by using the colour spectrum. The Hot sections shows the higher attention of visitors and Cold sections show the lower attention of visitors.

Different types of Heat Map :

1. Scroll Map : It shows the percentage of visitors that scroll through each section of the webpage. The hottor section on the webpage shows visitors have viewed it.
2. Click Maps : It shows the page sections that have more visitors clicks.The section that is hotter shows the more frequest visitors click.
3. Hover Map - It shows the sections where visitors moved their cursor while reading web page. The hottor section shows that visitors hung their cursor over it for a longer period.

Exit Page: It is important to know from which pages users are existing the most. pages from where visitors are dropping off in the process of buying a product are called Exit Page.

Trafic source: This is one of the most important metrics and a very good segmentation variable. There are three kinds of traffic source.

1. Direct Visitors- users that visit a website by directly typing your URL in their browser address bar
2. Search Visitors : users that visit a website based on search query in the search engine
3. Referral : users that visit a website because it was mentioned on another blog or website.

if one gets many visitors directly, it indicates that the brand has a high pull or brand image. If a business's dominant traffic source is a search engine, then it means their SEO is good.

OUTCOME Analysis :
A website may get many sessions and visitors, but wht is more important for abusiness to track is the business outcome of visits and sessions. Businesses are intrested in knowing how much revenue was generated how many converions happend, etc.

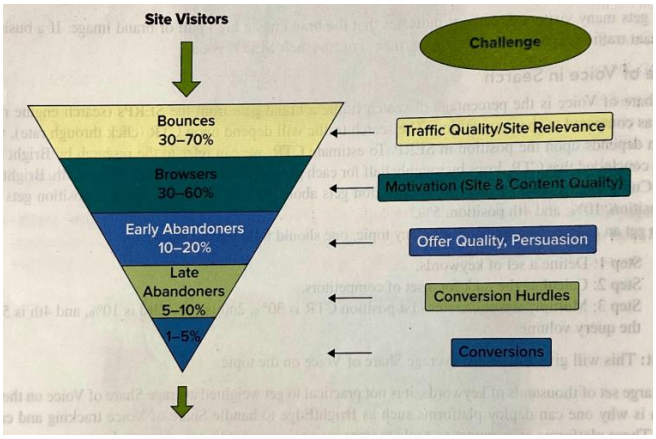
The Outcome metrics are

1. Conversion Rate : The converstion rate is the percentage of users who perfom an action that is desired by the website owner. The archetypical example of convertion rate would be the percentage of website visitors who do any online transaction on your website.

The conversion rate can also be:

- The user permitting the website to save his ATM card details for an easier payment in the future
- The user signs up for a subscription
- The user downloads your trial version software or a brochure, which may allow people to proceed in the sales funnel

- The user requests for more information
- Using a certain feature of an application (mainly new or advanced features)
- If the users have downloaded your mobile app and used it
- Have spent some time on your website or read some articles
- Returning to the site



AVERAGE ORDER VALUE

Average order value (AOV) can be formulated as the sum of revenue generated divided by the number of orders. This is one of the major business outcomes of an online business. One can segment visitors and marketing campaigns into high, medium and low AOV groups and identify where the best (e.g. high AOV) customers are coming from. One can also design their campaigns to increase the AOV.

Multichannel Funnel : Mutli-channel funnel reports enable understanding of the different users interact with on the path to convertions.

Topics converstion path- It shows the popular channel combinations users interact with before converstions. these channel includes

- paid and organic search
- referal sites
- affilates
- social networks
- email newsketters
- custom campaigns that you have created.

Assisted Converstion - the metrics shows the contribution of each channel towords converstion. there can be three ways in which a channel can contribute - i) last interaction ii) assist interaction and iii) the first interaction

Time Log - these reports can help one inderstand how many days users take to convert. Many days may indicte that users are finding challenging to take the final leap of faith in aproduct or service.

Path Log - Shows how long the sales cycle is. it show how many interactions users do before conventions

Visitors Frequency and Recency - how many user visit a website during the reporting time period.

New vs Return Visitors Conversition- how many new visitors they have in comparison to returing visitors.
Valu per visit - One must assign to every single visit on their website.

Micro Conversition - M?icro converstions are assist converstions, which include some intermediate step, shich might lead to a macro converstion.

Macro Converstion - This is ultimate sale or converstion.

Percentage of visitors who view product pages - the ultimate objective of the website is to get sales or converstions, and that will hapen only when the users visti the product pages.

Experience Analysis

The third category of web analytics is Experience Analysis - it is vital to do research on acontinious basis to know if visitors can find the information they are looking for andif the web page served their purpose. Experience of customer on awebsite can be gauged by various metrics and methods. there are many ways to know the experience of customers on awebsite

Research Data : one can carry out research using three methods i) site surveys ii) usability testing iii) site visits.

Site Surveys: survey questions can be asked for understanding the value of the web page.
Usability Testing - To ask the real users to test the functioning of the site to know how easy it is to navigate and also how intuitive it is.business should take their feedback about which aspects of the site they are not able to understand and are experiencing problems.
Site visits: - It is done by going to the customers' premises and observing how they accomplish tasks on websites amidst all distractions.

Website experimentation and Testing :

One must regularly experiment and test different things on the website to know what can be inproved.

A/B Testing : It is also called split testing or bucket testing. where one compares two verstions of a concept to see which one performs better. The concept that you wish to test could be an ad,price,page,call-to-action,product, etc.,

steps to perform A/B Testing :
The systematic steps that must be followed to perform A/B Tests

1. Conducting research : Research can be conducted to collect data on visitors behaviour through google analytics, heat maps and survey. Observations can be made to identify the customer conversion obstacles.
2. Generate hypothesis : Based on the data collected through research, formulate the

hypothesis aimed at increasing the conversion rate.
3. **Forming variation —** Creating a variation based on the hypothesis and A/B test it against the current version,. based on monthly visitors, current conversion rate and expected conversion rate, and initiate the test
4. **Testing -** After initiating the test, wait for the required time to achieve a significant result.
5. Analyse the achievement and failure

- **Pros of doing A/B testing -** A/B testing is a cost-effective way of testing cutting-edge ideas and yet exercising control on them by. testing on only a few users. One can be ahead of their competition if one constantly tests new ideas. This energizes the organization, and one can have some fun at the workplace. Google Analytics comes with content exper- iments integrated. One can click on the 'Experiments in Behaviour' tab on Google Analytics and set up the experiment (Figure 11.21). One must choose the objective of the experiment, percentage of traffic to experiment, set a minimum time for the experiment to run, and choose a confidence threshold.

  Most testing tools come with built-in capabilities for regression analysis, reports and a multivariate analysis.

- **Cons of doing A/B testing -** It is difficult to control all the external factors such as campaigns, search traffic, press releases, etc. Thus, one will not be 100% confident of the results (have around 70% confidence in the results, and decide accordingly) on one's A/B testing. The kinds of elements that can be tested on the website are also limited.

In A/B testing, one may know which page variant leads to more conversions, but one may not be able to discern which elements of the page contribute the most.

Multivariate Testing :

It is a technique that allows the test of multiple variants concurrently. for example, we wish to test three elements highlighted in the box - the heading, call-to-action and form. If each has two variants, there will be six different possible versions of the content that will compete to be the winning variation. The total number of variations in a multivariate test can be calculated by the following formula:

Total Number of Variations = [Number of Variations in Element A] x [Number of Variations in Element B] x [Number of Variations in Element C]

The challenge in MVT is that because of the factorial nature of the tests, the number of versions can quickly multiply requiring the large sample size to be distributed across different combinations. JavaScript tags must be implemented around the three identified elements.

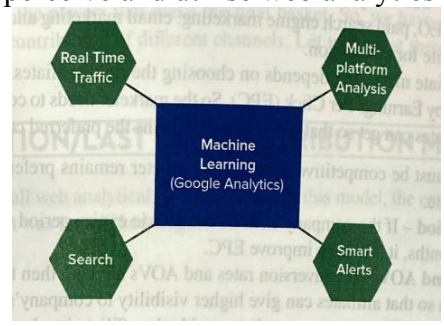
The Impact Matrix : The imact matric helps to evaluate business impact and time-to-useful.

How to use the matrix : The y-axis denotes business impact on an exponential scale

(from supertactical to super strategies)
The X-axis denotes time-touseful from real-time,weekly,monthly and quartely or 6-monthly.

Real Time		Weekly		Monthly		
		micro outcomes			24	super strategic
		conversions				
	page value					
outbound clicks	cost per Acquistion	Retail store visit	checkout abandon rates		10	
		click-2-delivery rate				
click-thru rate		page-depth	purchase intent			
AVOC	unique page views	Bounce rate	% completed videos		4	
viewability		GRPs	consideration			
	% new visits	applause rate	awareness			
Impressions	visits					super Tactical

Machine Learning in Google Analytics :
In web analytics, machine learning is the future. Considering the progress of technology like natural language processing, machine learning is surely going to revolutionise the way we perceive and utilise web analytics at present.



Here's how machine learning is incorporated in the Google Analytics system.

A. Real-Time Traffic Monitoring
Through this feature, we can monitor the impact of marketing campaigns and content/design modifications on our digital properties.

B. Search Feature
Through this feature, we can raise specific queries about a property, metric, duration, etc. For instance, we can get quick reports, data and insights by posting simple queries like "last month's acquisition from Mumbai or "Safari browser users in chennai last week"

C. Mulit Platform Analysis

this feature has the ability to provide combined information of app and web users. through ML technology, google analytics 4 gets empowered to extend instant data visualisation of metrics,dimentions etc. based on any platform that we select.

D. Smart Alerts

Google nalytics 4 uses ML to send alerts if your website or app experiences drastics trend changes such as significant rise or fall in number of users. The smart alert sid in doing root-cause analysis of unexpected spike/decline and in takingquick actions to stabilise the situation.

Multi Channel Attribution :

User are exposed to multi channels before they purchase a product from any websitte. Suppose one is selling aproduct; user saw the product advertisment on Facebook but did not buy it at that instance.Now the user happens to see the same advertisement again on the search network but does not buy this time as well. The user finally buys typing URL (direct). The question then aries - how much contribution will FB or search have in that purchase? this is called multi channel attribution. So we have different models that can be deployed to measure the contribution of different channels. Let us look at them.

1. Last interaction/Last click attributation Model
2. First interaction/First click attribution model
3. Linear Attribution model
4. Time decay attribution model
5. Position-based attribution model

Important Questions :

1. What is search engine Optimazattion and Explain the different phases of SEO.
2. Explain the Search working Process
3. Explain the process of data collection through Web logs and its benefits
4. Explain the Key metrics of web analytics
5. Explain the machine learning google analytics
6. Explaint Multi-Channel Atrribution