1. 1. Student name :TARUN KUMAR Roll No:1314310112 Submitted To: Ms. Shivani Aggarwal SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF BACHELOR OF ENGINEERING (COMPUTER SCIENCE & ENGINEERING) IMS Engineering College, Ghaziabad (U.P.) India (Affiliated to Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India) NOV 2016
2. [2.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-2-638.jpg?cb=1480692221) ACKNOWLEDGEMENT I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them. I am highly indebted to IBM for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. I would like to express my gratitude towards my parents & member of IBMfor their kind co- operation and encouragement which help me in completion of this project. I would like to express my special gratitude and thanks to industry persons for giving me such attention and time. My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.
3. [3.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-3-638.jpg?cb=1480692221) COMPANY PROFILE International Business Machines Corporation (IBM), incorporated on June 16, 1911, is a technology company. The Company operates through five segments: Global Technology Services (GTS), Global Business Services (GBS), Software, Systems Hardware and Global Financing. Global Technology Services The Company's GTS segment offers services, including strategic outsourcing, integrated technology services, cloud and technology support services (maintenance services). The Company's strategic outsourcing delivers information technology (IT) outsourcing services focused on clients' enterprise IT infrastructure environments. Its integrated technology services deliver a portfolio of project-based and managed services. Its cloud delivers a set of hybrid cloud services, including assisting clients with building their own private clouds, building clouds, allowing clients to leverage cloud infrastructure services from the SoftLayer and cloud managed services offerings, and creating environments linking their private and public workloads together. Its technology support services deliver a line of support services from product maintenance through solution support. Global Business Services The Company's GBS segment provides consulting and systems integration, application management services and process services. The Company's consulting and systems integration delivers client value with solutions in strategy and transformation, application innovation services, enterprise applications and smarter analytics. Their application management service delivers application management, maintenance and support services for packaged software, as well as custom and legacy applications. Its global process services deliver a range of offerings consisting of transformational solutions, including processing platforms and business process outsourcing. Software
4. [4.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-4-638.jpg?cb=1480692221) DECLARATION My topic is Hadoop which is Analysis of Crime in India using Hadoop. Apache Hadoop is a software framework that supports data-intensive distributed applications under a free license. Hadoop was inspired by Google's MapReduce and Google File System (GFS) papers. Hadoop, however, was designed to solve a different problem: the fast, reliable analysis of both structured data and complex data. As a result, many enterprises deploy Hadoop alongside their legacy IT systems, which allows them to combine old data and new data sets in powerful new ways. The Hadoop framework is used by major players including Google, Yahoo and IBM, largely for applications involving search engines and advertising. I am going to represent the History, Development and Current Situation of this Technology. This technology is now under the Apache Software Foundaion via Clodera
5. [5.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-5-638.jpg?cb=1480692221) TECHNOLOGY Chapter 1: Introduction to HADOOP Today, we’re surrounded by data. People upload videos, take pictures on their cell phones, text friends, update their Facebook status, leave comments around the web, click on ads, and so forth. Machines, too, are generating and keeping more and more data. The exponential growth of data first presented challenges to cutting-edge businesses such as Google, Yahoo, Amazon, and Microsoft. They needed to go through terabytes and petabytes of data to figure out which websites were popular, what books were in demand, and what kinds of ads appealed to people. Existing tools were becoming inadequate to process such large data sets. Google was the first to publicize MapReduce—a system they had used to scale their data processing needs. This system aroused a lot of interest because many other businesses were facing similar scaling challenges, and it wasn’t feasible for everyone to reinvent their own proprietary tool. Doug Cutting saw an opportunity and led the charge to develop an open source version of this MapReduce system called Hadoop . Soon after, Yahoo and others rallied around to support this effort. Today, Hadoop is a core part of the computing infrastructure for many web companies, such as Yahoo , Facebook , LinkedIn , and Twitter. Many more traditional businesses, such as media and telecom, are beginning to adopt this system too. Hadoop is an open source framework for writing and running distributed applications that process large amounts of data. Distributed computing is a wide and varied field, but the key distinctions of Hadoop are that it is ■ Accessible—Hadoop runs on large clusters of commodity machines or on cloud computing services such as Amazon’s Elastic Compute Cloud (EC2 ). ■ Robust—Because it is intended to run on commodity hardware, Hadoop is architected with the assumption of frequent hardware malfunctions. It can gracefully handle most such failures. ■ Scalable—Hadoop scales linearly to handle larger data by adding more nodes to the cluster. ■ Simple—Hadoop allows users to quickly write efficient parallel code. Hadoop’s accessibility and simplicity give it an edge over writing and running large distributed programs. Even college students can quickly and cheaply create their own Hadoop cluster. On the other hand, its robustness and scalability make it suitable for even the most demanding jobs at Yahoo and Facebook. These features make Hadoop popular in both academia and industry.
6. [6.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-6-638.jpg?cb=1480692221) Chapter 2: History of HADOOP Hadoop was created by Doug Cutting, the creator of Apache Lucene, the widely used text search library. Hadoop has its origins in Apache Nutch, an open source web search engine, itself a part of the Lucene project. The Origin of the Name “Hadoop”: The name Hadoop is not an acronym; it’s a made-up name. The project’s creator, Doug Cutting, explains how the name came about: The name my kid gave a stuffed yellow elephant. Short, relatively easy to spell and pronounce, meaningless, and not used elsewhere: those are my naming criteria. Kids are good at generating such. Googol is a kid’s term. Subprojects and “contrib” modules in Hadoop also tend to have names that are unrelated to their function, often with an elephant or other animal theme (“Pig,” for example). Smaller components are given more descriptive (and therefore more mundane) names. This is a good principle, as it means you can generally work out what something does from its name. For example, the jobtracker keeps track of MapReduce jobs. Building a web search engine from scratch was an ambitious goal, for not only is the software required to crawl and index websites complex to write, but it is also a challenge to run without a dedicated operations team, since there are so many moving parts. It’s expensive too: Mike Cafarella and Doug Cutting estimated a system supporting a 1- billion-page index would cost around half a million dollars in hardware, with a monthly running cost of $30,000. Nevertheless, they believed it was a worthy goal, as it would open up and ultimately democratize search engine algorithms. Nutch was started in 2002, and a working crawler and search system quickly emerged. However, they realized that their architecture wouldn’t scale to the billions of pages on the Web. Help was at hand with the publication of a paper in 2003 that described the architecture of Google’s distributed filesystem, called GFS, which was being used in production at Google.# GFS, or something like it, would solve their storage needs for the very large files generated as a part of the web crawl and indexing process. In particular, GFS would free up time being spent on administrative tasks such as managing storage nodes. In 2004, they set about writing an open source implementation, the Nutch Distributed Filesystem (NDFS). In 2004, Google published the paper that introduced MapReduce to the world. Early in 2005, the Nutch developers had a working MapReduce implementation in Nutch, and by the middle of that year all the major Nutch algorithms had been ported to run using MapReduce and NDFS. NDFS and the MapReduce implementation in Nutch were applicable beyond the realm of search, and in February 2006 they moved out of Nutch to form an independent subproject of Lucene called Hadoop. At around the same time, Doug Cutting joined Yahoo!, which provided a dedicated team and the resources to turn Hadoop into a system that ran at web scale (see sidebar). This was demonstrated in February 2008 when Yahoo! announced that its production search index was being generated by a 10,000-core Hadoop cluster.† In January 2008, Hadoop was made its own top-level project at Apache, confirming its success and its diverse, active community. By this timem Hadoop was being used by many other companies besides Yahoo!, such as Last.fm, Facebook, and the New York Times.
7. [7.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-7-638.jpg?cb=1480692221) Chapter 3: Key Technology The key technology for Hadoop is the MapReduce programming model and Hadoop Distributed File System. The operation on large data is not possible in serial programming paradigm. MapReduce do task parallel to accomplish work in less time which is the main aim of this technology. MapReduce require special file system. In the real scenario , the data which are in terms on perabyte. To store and maintain this much data on distributed commodity hardware, Hadoop Distributed File System is invented. It is basically inspired by Google File System. 3.1 MapReduce MapReduce is a framework for processing highly distributable problems across huge datasets using a large number of computers (nodes), collectively referred to as a cluster (if all nodes use the same hardware) or a grid (if the nodes use different hardware). Computational processing can occur on data stored either in a filesystem (unstructured) or in a database (structured). Figure 3.1 MapReduce Programming Model "Map" step: The master node takes the input, partitions it up into smaller sub-problems, and distributes them to worker nodes. A worker node may do this again in turn, leading to a multi- level tree structure. The worker node processes the smaller problem, and passes the answer back to its master node. "Reduce" step: The master node then collects the answers to all the sub-problems and combines them in some way to form the output – the answer to the problem it was originally trying to solve. MapReduce allows for distributed processing of the map and reduction operations. Provided each mapping operation is independent of the others, all maps can be performed in parallel –
8. [8.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-8-638.jpg?cb=1480692221) though in practice it is limited by the number of independent data sources and/or the number of CPUs near each source. Similarly, a set of 'reducers' can perform the reduction phase - provided all outputs of the map operation that share the same key are presented to the same reducer at the same time. While this process can often appear inefficient compared to algorithms that are more sequential, MapReduce can be applied to significantly larger datasets than "commodity" servers can handle – a large server farm can use MapReduce to sort a petabyte of data in only a few hours. The parallelism also offers some possibility of recovering from partial failure of servers or storage during the operation: if one mapper or reducer fails, the work can be rescheduled – assuming the input data is still available. 3.2 HDFS (Hadoop Distributed File System) The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets. HDFS relaxes a few POSIX requirements to enable streaming access to file system data. HDFS was originally built as infrastructure for the Apache Nutch web search engine project. HDFS is now an Apache Hadoop subproject. Figure 3.2 HDFS Architecture
9. [9.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-9-638.jpg?cb=1480692221) HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system’s clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode. The NameNode and DataNode are pieces of software designed to run on commodity machines. These machines typically run a GNU/Linux operating system (OS). HDFS is built using the Java language; any machine that supports Java can run the NameNode or the DataNode software. Usage of the highly portable Java language means that HDFS can be deployed on a wide range of machines. A typical deployment has a dedicated machine that runs only the NameNode software. Each of the other machines in the cluster runs one instance of the DataNode software. The architecture does not preclude running multiple DataNodes on the same machine but in a real deployment that is rarely the case. The existence of a single NameNode in a cluster greatly simplifies the architecture of the system. The NameNode is the arbitrator and repository for all HDFS metadata. The system is designed in such a way that user data never flows through the NameNode.
10. [10.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-10-638.jpg?cb=1480692221) Chapter 4: Other Projects on HADOOP 4.1 Avro Apache Avro is a data serialization system. Avro provides: 1. Rich data structures. 2. A compact, fast, binary data format. 3. A container file, to store persistent data. 4. Simple integration with dynamic languages. Code generation is not required to read or write data files nor to use or implement RPC protocols. Code generation as an optional optimization, only worth implementing for statically typed languages. 4.2 Chukwa Chukwa is a Hadoop subproject devoted to large-scale log collection and analysis. Chukwa is built on top of the Hadoop distributed filesystem (HDFS) and MapReduce framework and inherits Hadoop’s scalability and robustness. Chukwa also includes a ﬂexible and powerful toolkit for displaying monitoring and analyzing results, in order to make the best use of this collected data. 4.3 HBase Just as Google's Bigtable leverages the distributed data storage provided by the Google File System, HBase provides Bigtable-like capabilities on top of Hadoop Core.
11. [11.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-11-638.jpg?cb=1480692221) 4.4 Hive Hive is a data warehouse system for Hadoop that facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in Hadoop compatible file systems. Hive provides a mechanism to project structure onto this data and query the data using a SQL-like language called HiveQL. At the same time this language also allows traditional map/reduce programmers to plug in their custom mappers and reducers when it is inconvenient or inefficient to express this logic in HiveQL. 4.5 Pig Apache Pig is a platform for analyzing large data sets that consists of a high-level language for expressing data analysis programs, coupled with infrastructure for evaluating these programs. The salient property of Pig programs is that their structure is amenable to substantial parallelization, which in turns enables them to handle very large data sets. 4.6 ZooKeeper ZooKeeper is a centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services. All of these kinds of services are used in some form or another by distributed applications. Each time they are implemented there is a lot of work that goes into fixing the bugs and race conditions that are inevitable. Because of the difficulty of implementing these kinds of services, applications initially usually skimp on them ,which make them brittle in the presence of change and difficult
12. [12.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-12-638.jpg?cb=1480692221) to manage. Even when done correctly, different implementations of these services lead to management complexity when the applications are deployed. Chapter 5: HADOOP Single Node Setup The steps involved in setting up a single node Hadoop cluster are as follow: 1. Download the Hadoop Software, the hadoop.tar.gz file using the ftp://hadoop.apche.org URL, and ensure that the software is installed on every node of the cluster. Installing the Hadoop Software on all the nodes require unpacking of the software, the hadoop.apache.org URL, on the nodes. 2. Create the keys on local machine such that ssh, required by Hadoop, does not need password. Use following command to create key on local machine: $ ssh-keygen -t rsa -P “ “ $ cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys 3. Modify the environment parameters in the hadoop-env.sh file. Use the following command to change the environment parameter: Export JAVA\_HOME=/path/to/jdk\_home\_dir 4. Modify the configuration parameters in files given below as shown below. Do the following changes to the configuration files under hadoop/conf 1) core-site.xml <configuration> <property> <name>hadoop.tmp.dir</name> <value>TEMPORARY-DIR-FOR-HADOOPDATASTORE</value> </property> <property> <name>fs.default.name</name> <value>hdfs://localhost:54310</value> </property> </configuration> 2) mapred-site.xml <configuration> <property> <name>mapred.job.tracker</name> <value>localhost:54311</value> </property>
13. [13.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-13-638.jpg?cb=1480692221) </configuration> 3) hdfs-site.xml <configuration> <property> <name>dfs.replication</name> <value>1</value> </property> </configuration> 5. Format the hadoop file system. From hadoop directory run the following bin/hadoop namenode –format
14. [14.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-14-638.jpg?cb=1480692221) MODULES OF TRAINING 1. Our project starts with loading of data first This is logical scheme of data stored in pig. Then after data is loaded
15. [15.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-15-638.jpg?cb=1480692221) Loaded data :- Main projects starts from here. All most populous crimes are being analyzed in these following screenshot. It includes major crime such as murder robbery , kidnapping etc of some populated cities over India.
16. [16.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-16-638.jpg?cb=1480692221) Grouping all state in a single group for making analysis easier:-
17. [17.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-17-638.jpg?cb=1480692221) Calculating average theft for all state:- 905.521 is what we got from our analysis:-
18. [18.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-18-638.jpg?cb=1480692221) Generate maximum crime and murder among all cities in year 2013:- Queries successfully running:-
19. [19.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-19-638.jpg?cb=1480692221) SUMMARY Big Data Analytics refers to the tools and practices that can be used for transforming this raw data into meaningful and crucial information which helps in forming a decision support system for the judiciary and legislature to take steps towards keeping crimes in check. With the ever increasing population and crime rates, certain trends must be discovered, studied and discussed to take well informed decisions so that law and order can be maintained properly. If the number of complaints from a particular state is found to be very high, extra security must be provided to the residents there by increasing police presence, quick redressal of complaints and strict vigilance. Crimes against women are becoming an increasingly worrying and disturbing problem for the government. The number of such crimes must be found, especially the ones against young women (age between 18-30 years). Extra security must be provided so that law and order can be maintained properly and there is a sense of safety and well-being among the citizens of the country Future work: 1. This analysis can be further carried out on Fully Distributed cluster mode that is hadoop 2. Similar analysis can be further carried out in different sector.
20. [20.](https://image.slidesharecdn.com/tarunf-161202152251/95/hadoop-project-report-20-638.jpg?cb=1480692221) BIBLIOGRAPHY [1] Jason Venner, Pro Hadoop, Apress [2] Tom White, Hadoop: The Definitive Guide , O’REILLY [3] Chuck Lam, Hadoop in Action, MANNING [4] Hadoop.apache.org [5] http://tutorialshadoop.com/pig-interview-questions-answers-part-3 [6] Lecture Notes in Computer Science, 2013.

[Recommended](https://www.slideshare.net/TARUNKUMAR362/hadoop-project-report#related-tab-content)

[[](https://www.linkedin.com/learning/teaching-techniques-classroom-cloud-strategy?trk=slideshare_sv_learning)](https://www.linkedin.com/learning/teaching-techniques-classroom-cloud-strategy?trk=slideshare_sv_learning" \o "Teaching Techniques: Classroom Cloud Strategy)

[Teaching Techniques: Classroom Cloud Strategy](https://www.linkedin.com/learning/teaching-techniques-classroom-cloud-strategy?trk=slideshare_sv_learning" \o "Teaching Techniques: Classroom Cloud Strategy)

[Online Course - LinkedIn Learning](https://www.linkedin.com/learning/teaching-techniques-classroom-cloud-strategy?trk=slideshare_sv_learning" \o "Teaching Techniques: Classroom Cloud Strategy)

[[](https://www.linkedin.com/learning/learning-everyday-math?trk=slideshare_sv_learning)](https://www.linkedin.com/learning/learning-everyday-math?trk=slideshare_sv_learning" \o "Learning Everyday Math)

[Learning Everyday Math](https://www.linkedin.com/learning/learning-everyday-math?trk=slideshare_sv_learning" \o "Learning Everyday Math)

[Online Course - LinkedIn Learning](https://www.linkedin.com/learning/learning-everyday-math?trk=slideshare_sv_learning" \o "Learning Everyday Math)

[[](https://www.linkedin.com/learning/bruce-heavin-the-thinkable-presentation?trk=slideshare_sv_learning)](https://www.linkedin.com/learning/bruce-heavin-the-thinkable-presentation?trk=slideshare_sv_learning" \o "Bruce Heavin The Thinkable Presentation)

[Bruce Heavin The Thinkable Presentation](https://www.linkedin.com/learning/bruce-heavin-the-thinkable-presentation?trk=slideshare_sv_learning" \o "Bruce Heavin The Thinkable Presentation)

[Online Course - LinkedIn Learning](https://www.linkedin.com/learning/bruce-heavin-the-thinkable-presentation?trk=slideshare_sv_learning" \o "Bruce Heavin The Thinkable Presentation)

[[](https://www.slideshare.net/kaushudeshwal/nwr-summer-training-report)](https://www.slideshare.net/kaushudeshwal/nwr-summer-training-report" \o "NWR summer training report)

[NWR summer training report](https://www.slideshare.net/kaushudeshwal/nwr-summer-training-report" \o "NWR summer training report)

[Kaushu Deshwal](https://www.slideshare.net/kaushudeshwal/nwr-summer-training-report" \o "NWR summer training report)

[[](https://www.slideshare.net/plandeen/technology-presentation)](https://www.slideshare.net/plandeen/technology-presentation" \o "Technology Presentation)

[Technology Presentation](https://www.slideshare.net/plandeen/technology-presentation" \o "Technology Presentation)

[plandeen](https://www.slideshare.net/plandeen/technology-presentation" \o "Technology Presentation)

[[](https://www.slideshare.net/narangv43/seminar-presentation-hadoop)](https://www.slideshare.net/narangv43/seminar-presentation-hadoop" \o "Seminar Presentation Hadoop)

[Seminar Presentation Hadoop](https://www.slideshare.net/narangv43/seminar-presentation-hadoop" \o "Seminar Presentation Hadoop)

[Varun Narang](https://www.slideshare.net/narangv43/seminar-presentation-hadoop" \o "Seminar Presentation Hadoop)

[[](https://www.slideshare.net/TARUNKUMAR362/artificial-intelligence-notes)](https://www.slideshare.net/TARUNKUMAR362/artificial-intelligence-notes" \o "artificial intelligence notes)

[artificial intelligence notes](https://www.slideshare.net/TARUNKUMAR362/artificial-intelligence-notes" \o "artificial intelligence notes)

[TARUN KUMAR](https://www.slideshare.net/TARUNKUMAR362/artificial-intelligence-notes" \o "artificial intelligence notes)

[[](https://www.slideshare.net/TARUNKUMAR362/embedded-system-notes)](https://www.slideshare.net/TARUNKUMAR362/embedded-system-notes" \o "Embedded system notes)

[Embedded system notes](https://www.slideshare.net/TARUNKUMAR362/embedded-system-notes" \o "Embedded system notes)

[TARUN KUMAR](https://www.slideshare.net/TARUNKUMAR362/embedded-system-notes" \o "Embedded system notes)

[[](https://www.slideshare.net/TARUNKUMAR362/merging-in-parallel-algorithm)](https://www.slideshare.net/TARUNKUMAR362/merging-in-parallel-algorithm" \o "MERGING IN PARALLEL ALGORITHM)

[MERGING IN PARALLEL ALGORITHM](https://www.slideshare.net/TARUNKUMAR362/merging-in-parallel-algorithm" \o "MERGING IN PARALLEL ALGORITHM)

[TARUN KUMAR](https://www.slideshare.net/TARUNKUMAR362/merging-in-parallel-algorithm" \o "MERGING IN PARALLEL ALGORITHM)

[[](https://www.slideshare.net/TARUNKUMAR362/research-paperbig-data-analytict-of-various-growth-parameters-for-asian-subcontinents)](https://www.slideshare.net/TARUNKUMAR362/research-paperbig-data-analytict-of-various-growth-parameters-for-asian-subcontinents" \o "Research paper(Big Data Analytict of Various Growth Parameters for Asian Subcontinents))

[Research paper(Big Data Analytict of Various Growth Parameters for Asian Subc...](https://www.slideshare.net/TARUNKUMAR362/research-paperbig-data-analytict-of-various-growth-parameters-for-asian-subcontinents" \o "Research paper(Big Data Analytict of Various Growth Parameters for Asian Subcontinents))

[TARUN KUMAR](https://www.slideshare.net/TARUNKUMAR362/research-paperbig-data-analytict-of-various-growth-parameters-for-asian-subcontinents" \o "Research paper(Big Data Analytict of Various Growth Parameters for Asian Subcontinents))