Historical Data Dissemination of Capital Market Segment from NSE

From late 1995 onwards, NSE has been India's largest equity market. This has led to a series of requests for historical data, for use in research and analysis amongst market participants, researchers, and policy makers. In response to these requests, NSE has setup a formal framework for historical data dissemination.

Organization of Historical Data:

Now the data will be provided to clients only through website which will contain 6 directories:

Bhav copy: Summary information about each security for each trading day.

Index: Information about stock market indexes.

Masters: Database masters, listing out symbols, series, ISINs, etc.

Snapshots: Snapshots of the limit order book at many time points in the month.

Trades: A database of every single trade that took place.

Circulars: A database of all circulars issued by NSE or NSCCL. This serves as a formal documentation of the history of market design at NSE.

Bhav copy database:

The Bhav copy directory contains a directory structure organized as 1999/Mar/date.gz where the date is represented as yyyymmdd. Thus, the file for 5th March, 1999 would be located at Bhav copy/1999/Mar/19990305.gz

Sr No	Field	Description
1	Symbol	Symbol for each company given by NSE; e.g. "SBIN" for State Bank of India or "INFOSYS" for Infosys. A master table of all symbols is found in the Masters database.
2	Series	Series Symbol for each security given by NSE; e.g. "EQ" for common stock, "N1" for first debenture issue, "W" for warrants, etc. Once a symbol and a series have been specified, a security is uniquely known. A master table of all symbol + series combinations which are traded is in the Masters database.
3	Open price	The opening price of the day. On some days, when the pre-opening call auction has been used, and if the security trades in the call auction, the Open price is the (single price) from the call auction. Otherwise, the Open price is the price at the first trade of the day.
4	High price	The highest traded price of the day.
5	Low price	The lowest traded price of the day.
6	Closing price	This is the official closing price reported by NSE.
7	LTP	The last traded price of the day. In general, this need not be equal to the official closing price because the official closing price is calculated using a variety of rules (e.g. averaging of trades over the last 30 minutes), etc.
8	Traded quantity	The number of shares traded in the day.
9	Value of shares traded	The rupee value of all shares traded in the day. The volume weighted average (VWA) price is field 9 divided by field 8.
10	Number of trades	Number of trades which took place in the normal market (i.e. excluding trades in the auction market. The average trade size is field 8 divided by field 10 (in number of shares) or field 9 divided by field 10 (in rupees).
11	Corporate action flags	Ex-date indicators, e.g. XD for a dividend, etc.

Index database:

The Index directory contains databases connected with stock market indexes. Both end-of-day and intra-day information is available. Following indexes are covered: Nifty, CNX Midcap, these are found in directories called Nifty, CNX Midcap. Nifty is the main stock market index in India; it is composed of the top 50 highly liquid stocks in India which make up roughly half of the market capitalization of India.

The end-of-day data for the month of March '99 for Nifty is found in the file Index/Nifty/1999/Mar/summary.gz. Intra-day data for Nifty for the 5th of March, 1999 is found in the file Index/Nifty/1999/Mar/19990305.gz

Sr No	Field	Description
1	Name of Index	This identifies the index, e.g. Nifty.
2	Date	The date, formatted as yyyymmdd.
3	Open	The opening level of the index.
4	High	The highest level of the index in the day.
5	Low	The lowest level of the index in the day.
6	Close	The official closing price of the index: this is the reference rate that is used for measuring the expiration value of index futures or index options.
7	Number Of shares traded	The sum of the number of shares traded of each of the components of the index.
8	Value of shares traded	The sum of value of the shares traded of each of the components of the index.
9	Market capitalization	The sum of the market capitalization of all the components of the index.
10	Impact cost at a program trade of Rs.2.5 million	Average impact cost (measured in percent) faced when doing program trades on this index for a transaction size of Rs.2.5 million. Typically, the average is taken over the values seen in three snapshots of the limit order book on the day. Impact cost is provided at Rs.5 million, Rs.10 million and Rs.20 million*
11	Returns on TR index	Returns on the index, inclusive of dividends.

The intra-day files:

The intra-day files show a fresh calculation of the market index every time a trade takes place for an index component. Most of the time, more than one trade takes place in a given second, so multiple records are found for the same second. Hence, we often see days where there are more than 100,000 observations for Nifty. The records shown are in correct time-sorted order, even though it appears that they all have the same timestamp.

1. Index name: This is a string identifying the index, e.g. "Nifty".
2. Timestamp: This is formatted as hh:mm:ss.
3. Index level: The level of the index, rounded off to 10 decimal places.

Masters database:

The website shows a snapshot of the Masters table as of the end of the preceding month. Hence, the Master applicable for March 1999 is found in the file 199903/Masters/1999/Mar/19990228.gz

Sr No	Field	Description
1	ISIN	The International Security Identification Number (ISIN), if it has been allocated, e.g. INE117A01014.
2	Symbol	NSE's `symbol', e.g. ABB.
3	Series	NSE's `series', e.g. BE
4	Name	A descriptive string about the product.
5	Deleted	A flag which takes the value `N' if the security has not been deleted.

Order book snapshots database:

NSE is a limit order book market, also known to economists as the `open electronic limit order book market (OELOB)', or to practitioners as a market based on electronic order matching'. Liquidity on the OELOB market is embedded in the limit orders present at any point in time; these limit orders (the right to trade against them, without any obligation) are free options which others can exploit. Measurement of this liquidity is possible with high accuracy using "order book snapshots": pictures of the complete limit order book at a point in time. This is discrete, in only conveying the picture at a few time points in the day. However, at these time points, a variety of questions about liquidity can be accurately answered. The order book snapshot can yield the bid-ask spread, and it can be used to measure market impact cost for buying or selling any desired transaction (or for testing whether a desired transaction is feasible). The order book snapshots for 5 March, 1999 are stored in the directory 199903/Snapshots/19990305. The files that are found inside this have names of the form hhmmss.gz, to convey the time at which the snapshot was taken. For example, for 5 March, 1999 the website contains 110000.gz, 130000.gz and 140000.gz. These are order book snapshots at 11 AM, 12:00 PM, 1 PM and 2 PM. These files are databases with one record per line, and each record pertains to one limit order. The files are sorted by price. They have 14 fields per record:

Sr No	Field	Description
1	Order ID number	This is a field like 9801050079959 which is a unique code given to every limit order on NSE.
2	Symbol	The symbol for the security.
3	Series	The series for the security. Every security is uniquely defined once its symbol and series have been specified. A symbol of GLAXO and a series of EQ denotes common stock of Glaxo.
4	Quantity	The size of the limit order.
5	Price	The limit price on the limit order.
6	Timestamp	The time at which the order was placed (or last modified). This is formatted as hh:mm:ss.
7	Buy/Sell	This is B for buy limit orders and S for sell limit orders.
8	Day flags	This is a set of four Boolean flags: 1) Day order, 2) Good till date, 3) Cancel, 4) Immediate/Cancel. The commonest value found is ynnn.

9	Quantity flags	This is a set of three quantity flags: 1) Minimum fill, 2) All or none, 3) Disclosed quantity. The commonest value found is nnn.
10	Price flags	This is a set of three price flags: 1) At the open (ATO) price, 2) Market price, 3) Stop loss order. The commonest value is nnn.
11	Book type	There are two books which can be used: RL and SL. RL is the most common.
12	Minimum fill quantity	If the order specifies a minimum fill, then this field shows the minimum fill quantity specified.
13	Quantity disclosed	If the order discloses a smaller quantity as compared with the true order size, then this field shows the smaller quantity that is meant to be disclosed.
14	Date for GTD	GTD orders need to specify a date until which the order is good: that date is specified here.

Trades data:

This is a data about every trade that took place. If NSE does 400000 trades in a day, the dataset for that day would have 400000 trades.

Information for the trades of each day is kept in a distinct file. The information for 5 March, 1999 is found in 199903/Trades/1999/Mar/19990305.gz

The files are organized as follows:

Sr.No	Field	Description
1	Trade ID number	A unique number for each trade, the files are sorted by this trade ID.
2	Symbol	The symbol of the security traded.
3	Series	The series of the security traded.
4	Timestamp	The time at which the trade took place, formatted as hh:mm:ss Many times, many trades are matched within the same second, in which case we see multiple records with the same timestamp. The pattern of seeing a large number of trades for the same stock towards the start or the end of the market is owing to the uniform-price order matching at the end of the call auctions.
5	Price	The price at this trade.
6	Quantity traded	The number of shares transacted in this trade.

Database of circulars:

Circulars are a formal method of communication between NSE and its member brokerage firms. The website will contain a comprehensive set of circulars issued in the month. Every development on the market in terms of market design is documented in these circulars.

All the circulars for a month can be accessed by pointing your web browser to the file index.html. For example, the circulars for March, 1999 can be read by loading up 199903/Circulars/1999/Mar/Index.html into your browser.

Historical Data Dissemination of Future and Options Segment

The derivatives trading on the exchange commenced with S&P CNX Nifty Index futures on June 12, 2000. The trading in index options commenced on June 4, 2001 and trading in options on individual securities commenced on July 2, 2001. Single stock futures were launched on November 9, 2001. From then till now NSE evolved into the largest market in Derivatives with a daily turnover of overRs.30,000 crores.

This has led to a series of requests for historical data, for use in research and analysis amongst market participants, researchers, and policy makers. In response to these requests, NSE has setup a formal framework for historical data dissemination of the Future and Options segment of the Exchange.

Data Organization:

When you visit the website the starting directory is 200301. This name is in yyyymm format and indicates that the website contains data pertaining to January 2003. Underneath this starting directory, there are 5 directories:

Bhav copy: Summary information about each security for each trading day.

Masters: Database masters, listing out symbols, series, ISINs, etc.

Snapshots: Snapshots of the limit order book at many time points in the month.

Trades: A database of every single trade that took place.

Circulars: A database of all circulars issued by NSE or NSCCL. This serves as a formal documentation of the history of market design at NSE.

Bhav copy Directory:

The Bhav copy directory contains a directory structure organized as 2003\Jan\date.gz, where the date is represented as yyyymmdd.gz. Thus, the file for 3rd Jan, 2003 would be located at Bhav copy\2003\Jan\20030103.gz. This directory structure is used to make it convenient for users to merge data from the website (for multiple months) into one coherent file system on their hard disk.

This gives us one file for each trading day. The lines in this file have 16 fields per line (each line is one observation), delimited by the pipe "|" character. These fields are:

Sr No	Fields	Description
1	Date	This gives the Trade date.
2	Symbol	This gives the underlying index or stock. e.g. NIFTY ACC, etc.
3	Instrument	This gives the contract descriptor for the various instruments available in the derivatives segment. e.g. FUTSTK, OPTIDX, etc.
4	Expiry date	The date on which the contract expires.
5	Option Type	This gives the type of option for the contract which is either call or put. e.g. CE- Call European, PE- Put European, CA- Call American, PA- Put American
6	Corporate Action level.	This is the Corporate Action Flag. This flag changes when there is a corporate action on a contract, which could either be a symbol change or a dividend announced by the company.
7	Strike Price	This gives the Strike Price of the contract.
8	Opening price	This gives the price at which the contract opened for the day.
9	High price	This gives the highest price at which the contract was traded during the day.
10	Low price	This gives the lowest price at which the contract was traded during the day.
11	Closing price	This gives the price of the contract at the end of the day.
12	Last traded price	This gives the price of the contract on its last trade.
13	Open Interest	For futures contracts open interest is equivalent to the open positions in that futures contract multiplied by its last available closing price. For option contracts, open interest is equivalent to the open positions multiplied by the notional value. Notional value with respect to an option contract is computed as the product of the open position in that option contract multiplied by the last available closing price of the underlying.

14	Total Traded Quantity	This is the total no. of contracts on which business took place during the day.
15	Total Traded Value	The total money value of the business which took place on the contract during the day.
16	No. of Trades	The total no. of trades which took place on the instrument during the day.

Masters Directory:

This directory contains all the contracts as on the month end including the contracts that expired on the last Thursday of the month. The Master applicable for January 2003 is found in the file 200301\Masters\2003\Jan\20030131.gz

The 12 fields in this file are:

Sr.No.	Fields	Description
1	Symbol	This gives the underlying index or stock. e.g. NIFTY, ACC, etc
2	Instrument type	This gives the contract descriptor for the various instruments available in the derivatives segment. e.g. FUTSTK, OPTIDX, etc.
3	Expiry Date	The date on which the contract expires.
4	Option Type	This gives the type of option for the contract which is either call or put. e.g. CE- Call European, PE- Put European, CA- Call American, PA- Put American
5	Corporate Action Level	This is the Corporate Action Flag. This flag changes when there is a corporate action on a contract, which could either be a symbol change or a dividend announced by the company.
6	Strike Price	This gives the Strike Price of the contract.
7	Contract Regular Lot	This is the market lot of the contract available for trading.
8	Token Number	A unique number assigned to the contract by the system.
9	Issue Start Date	This indicates the date from which the contract is available for trading.
10	Issue Maturity Date	This indicates date on which the contract will mature.
11		

	Exercise Start Date (NULL in case of Futures.)	The date from which the contract can be exercised after introduction into the system.
12	Exercise End Date (NULL in case of Futures.)	The date till which the contract can be exercised after introduction into the system.

Snapshots Directory:

NSE is a limit order book market, also known to economists as the 'Open Electronic Limit Order Book Market (OELOB)', or to practitioners as a market based on 'electronic order matching'. Liquidity on the OELOB market is embedded in the limit orders present at any point in time; these limit orders (the right to trade to trade against them, without any obligation) are free options which others can exploit.

Measurement of this liquidity is possible with high accuracy using "order book snapshots"- pictures of the complete limit order book at a point in time. This is discrete, in only conveying the picture at a few time points in the day. However, at these time points, a variety of questions about liquidity can be accurately answered. The order book snapshot can yield the bid-ask spread, and it can be used to measure market impact cost for buying or selling any desired transaction (or for testing whether a desired transaction is feasible).

The order book snapshots for 3rd January, 2003 are stored in the directory 200301/Snapshots/20030103. The files that are found inside this have names of the form hhmmss.gz, to convey the time at which the snapshot was taken. For example, for 3rd January, 2003 the website contains 110000.gz, 120000.gz, 130000.gz, 140000.gz. And 150000.gz. These are the order book snapshots at 11 am, 12 noon, 1 pm, 2 pm and 3 pm.

The 18 fields in this file are:

Sr No	Field	Description
1	Order number	As and when valid orders are entered or received by the system, they are first numbered, time stamped and then scanned for a potential match. This means that each order has a distinct order number.
2	Symbol	This gives the underlying index or stock. e.g. NIFTY, ACC, etc

<table>
<tr><td>3</td><td>Instrument type.</td><td>This gives the contract descriptor for the various instruments available in the derivatives segment. e.g. FUTSTK, OPTIDX, etc.</td></tr>
<tr><td>4</td><td>Expiry date</td><td>The date on which the contract expires.</td></tr>
<tr><td>5</td><td>Strike price</td><td>This gives the Strike Price of the contract.</td></tr>
<tr><td>6</td><td>Option type</td><td>This gives the type of option for the contract which is either call or put. e.g. CE- Call European, PE- Put European, CA- Call American, PA- Put American</td></tr>
<tr><td>7</td><td>Corporate action level</td><td>This is the Corporate Action Flag. This flag changes when there is a corporate action on a contract, which could either be a symbol change or a dividend announced by the company.</td></tr>
<tr><td>8</td><td>Quantity</td><td>This gives the quantity remaining, i.e., if the order has been partly traded, the balance untraded quantity is indicated in this field.</td></tr>
<tr><td>9</td><td>Price</td><td>This is the limit price.</td></tr>
<tr><td>10</td><td>Time stamp</td><td>As and when valid orders are entered or received by the system, they are first numbered, time stamped and then scanned for a potential match. This means that each order has a unique time stamp on it.</td></tr>
<tr><td>11</td><td>Buy/Sell</td><td>It indicates whether the order placed is a buy order or a sell order. B: stands for a buy order S: stands for a sell order</td></tr>
<tr><td>12</td><td>Day flags.</td><td>These are called the day flags indicating whether it is a DAY order or GTD order or GTC order or FILL/ KILL order, in this respective sequence. The flag is 'y' accordingly at the relevant place. For e.g. if this field shows 'nynn', then it is a GTD order since the second place is 'y'.</td></tr>
<tr><td>13</td><td>Quantity flags.</td><td>This is a sequence of three flags in which the first flag indicates MF (min fill), second AON (all or none) and the third DQ (disclosed qty). Accordingly it will be 'y' or 'n'.</td></tr>
<tr><td>14</td><td>Price flags</td><td>This is again a sequence of three flags in which the first stands for ATO , the second for MKT, and the third for ONSTOP</td></tr>
<tr><td>15</td><td>Book Type</td><td>The various book types that are used and their corresponding values are:
<table>
<tr><td>Book Type</td><td>Value entered by trading member through the front end through the front end</td></tr>
<tr><td>Regular Book</td><td>RL</td></tr>
<tr><td>Special Terms Book</td><td>ST</td></tr>
<tr><td>Stop Loss Book</td><td>SL</td></tr>
<tr><td>Negotiated Trade Book</td><td>NT</td></tr>
</table>
</td></tr>
</table>

		Odd Lot Book: OL Spot Order Book: SP Auction Order Book: AU
16	Minimum fill quantity	It denotes the minimum fill quantity.
17	Quantity disclosed	It denotes the disclosed quantity in case of a special terms order. For all orders where the disclosed quantity has not been specified it has a value zero.
18	Date for GTD	It denotes the date up to which the order should stay in the system. If the instrument is not GTD then this field is left blank.

Trades Directory:

This directory contains details about every trade that took place. If NSE does 400000 trades in a day, the dataset for that day would have 400000 trades.

Information for the trades of each day is kept in a distinct file. The information for 3 January, 2003 is found in 200301\Trades\2003\Jan\20030103.gz

The files are organized as follows:

Sr No	Field	Description
1	Trade number	A number of each trade is given.
2	Symbol	This gives the underlying index or stock. e.g. NIFTY, ACC, etc.
3	Instrument type	This gives the contract descriptor for the various instruments available in the derivatives segment. e.g. FUTSTK, OPTIDX, etc.
4	Expiry date	The date on which the contract expires.
5	Option type	This gives the type of option for the contract which is either call or put. e.g. CE- Call European, PE- Put European, CA- Call American, PA- Put American

6	Corporate action level	This is the Corporate Action Flag. This flag changes when there is a corporate action on a contract, which could either be a symbol change or a dividend announced by the company.
7	Strike price	This gives the Strike Price of the contract.
8	Trade time.	Time at which the trade took place. The format is hh:mm:ss.
9	Traded price	The price at which the contract was traded.
10	Traded quantity.	The number of contracts traded for the instrument in this trade.

Circulars Directory:

Circulars are a formal method of communication between NSE and its member brokerage firms. This website contains a comprehensive set of circulars issued in the month of January, 2003. Every development on the market in terms of market design is documented in these circulars.

All the circulars for January can be accessed by pointing your web browser to the file index.html. For example, the circulars for January, 2003 can be read by loading up 200301/Circulars/2003/Jan/Index.html into your browser.