algoseek

THE MARKET DATA COMPANY

US Equity Full Depth Guide

version 1.4 (Jul 2021)

algoseek | the market data company

We provide research market data for machine learning and quantitative trading

CONTACT US

We are here to help you do great things with our market data. For questions, feedback, and other concerns, you may reach our team of experts using the following contact information:

algoseek customer support

support@algoseek.com

(+1) 646 583 1832

algoseek sales

sales@algoseek.com

(+1) 646 583 1832

TABLE OF CONTENTS

INTRODUCTION

This document covers algoseek's normalized full depth direct feeds from the following US exchanges: NASDAQ, Direct Edge, NYSE ARCA, and BATS.

algoseek historical exchange data covers:

January 2013 to November 2017 for NASDAQ, Direct Edge, NYSE ARCA, and BATS

November 2017 to December 2018 for NASDAQ and NYSE ARCA

December 2018 to April 2019 for NASDAQ

All messages are consolidated and ordered by the millisecond timestamp. The data setup enables you to build a full depth of the book at any moment intraday. The data contains the entire trading session, including early and late hours from 04:00 to 20:00 EST.

DATA ORGANIZATION AND FILE FORMAT

algoseek provides Equity market data in plain-text CSV files. The first row of the CSV file is a fixed header, and then rows of data corresponding to individual events (see Table 1). By default, data is organized into one file per symbol per trading day. For example, all events for ticker AAPL on Mar 3, 2020, are stored in one CSV file.

Due to the large data size, CSV files are gzip-compressed (having a csv.gz extension) with a compression ratio of about 8:1.

Table 1: Sample Full Depth Data

Date	Timestamp	Order Number	Event Type	Ticker	Price	Quantity	MPID	Exchange
20190405	09:30:00.010	7394709	ADD BID	AAPL	164.99	100	CDRG	NASDAQ
20190405	09:30:00.010	7394789	ADD ASK	AAPL	227.94	100	CDRG	NASDAQ
20190405	09:30:00.051	7425621	ADD BID	AAPL	180.00	219		NASDAQ
20190405	09:30:00.052	7426017	ADD BID	AAPL	180.00	172		NASDAQ
20190405	09:30:00.054	7426417	ADD ASK	AAPL	200.00	50		NASDAQ
20190405	09:30:00.063	7430365	ADD ASK	AAPL	208.26	100		NASDAQ
20190405	09:30:00.063	7430373	ADD BID	AAPL	184.63	100		NASDAQ
20190405	09:30:00.182	7446757	ADD ASK	AAPL	198.46	100		NASDAQ
20190405	09:30:00.182	7446761	ADD BID	AAPL	194.45	100		NASDAQ

20190405	09:30:00.183	0	TRADE BID	AAPL	196.44	15		NASDAQ
20190405	09:30:00.185	109301	DELETE ASK	AAPL	0.00	0		NASDAQ
20190405	09:30:00.185	109269	DELETE BID	AAPL	0.00	0		NASDAQ

The table below provides the name, description, and data type for each data field (column) in an Equity Full Depth data file.

Timestamps in Excel. Excel fails when importing timestamp fields as Excel automatically tries to convert milliseconds and nanoseconds to Excel time format. When importing timestamp, you can import as Text fields instead.

Table 2: CSV File Fields Schema

Field	Type (Format)	Description
Date	string (yyyymmdd)	Trading date in yyyymmdd format
Timestamp	string (HH:MM:SS.mmm)	Event timestamp in milliseconds
OrderNumber	string	Unique order ID
EventType	string	The type of event
Ticker	string	Symbol name
Price	decimal	Order price, available for the ADD BID/ASK, CROSS, and TRADE BID/ASK order messages. Zero for cancellations and executions
Quantity	integer	Number of shares available for the ADD BID/ASK, EXECUTE BID/ASK, CROSS, CANCEL BID/ASK, TRADE BID/ASK messages. Zero for FILL BID/ASK and DELETE BID/ASK messages
MPID	string	Market Participant ID associated with the transaction (four characters) used to be available on specific exchanges but is no longer published in recent years
Exchange	string	The exchange or reporting venue

Time Range

The Full Depth dataset covers the entire trading day from the start of pre-market trading to the end of after-hours trading (EST time):

Pre-Market Hours: 04:00:00 to 09:29:59

Market Hours: 09:30:00 to 16:00:00

Post-Market Hours: 16:00:01 to 20:00:00

Note: Occasionally trade and quote events are recorded several minutes after 20:00.

Market Holidays and Early Closes

The stock market is closed for trading on most US holidays. For reference, algoseek publishes a list of historical holidays which is available at `s3://us-equity-market-holidays/holidays.csv` (direct download link: https://us-equity-market-holidays.s3.amazonaws.com/holidays.csv).

Markets sometimes close early at 13:00:00 on the day before holidays such as Independence Day and Thanksgiving. You can download algoseek's early close date and time list from AWS S3 storage at `s3://us-equity-market-holidays/earlycloses.csv` (or use a direct link us-equity-market-holidays.s3.amazonaws.com/earlycloses.csv).

Timestamp

The event timestamp has a millisecond resolution, and the time zone is EST. Timestamp field takes the format of HH:MM:SS.mmm, for example, 09:31:01.723, where

HH: Hour

MM: Minute

SS: Seconds

mmm: Milliseconds

Timestamps in Excel. Excel fails when importing timestamp fields as Excel automatically tries to convert milliseconds and nanoseconds to Excel time format. When importing timestamp, you can import as Text fields instead.

Event Types

Table 3 contains names and descriptions of event types found in the dataset.

Table 3: Event Types for Equity Full Depth Dataset

Event Type	Description
ADD BID	Add Bid order
ADD ASK	Add Ask order
CANCEL BID	Cancel outstanding Bid order in part
CANCEL ASK	Cancel outstanding Ask order in part
DELETE BID	Delete outstanding Bid order in full
DELETE ASK	Delete outstanding Ask order in full
EXECUTE BID	Execute outstanding Bid order in part
EXECUTE ASK	Execute outstanding Ask order in part
FILL BID	Execute outstanding Bid order in full
FILL ASK	Execute outstanding Ask order in full
TRADE BID	Execute non-displayed Bid order. The order number is then set to 0
TRADE ASK	Execute non-displayed Ask order. The order number is then set to 0
CROSS	Bulk volume for the opening or closing market cross. An exchange may do an opening or closing cross for a stock where multiple orders are aggregated into one price and block

Note: For the binary version of the dataset the last 4 digits in the “Price” field correspond to decimal digits. The decimal portion is padded on the right with zeros. The decimal point is implied by position; it does not appear inside the price field. Divide the Price by 10000 to convert into a currency value.

APPENDIX A. FREQUENTLY ASKED QUESTIONS

Why are time-based columns not recognized properly when I try importing data to Excel?

Older versions of Excel will automatically try to convert the “Timestamp” field into an Excel format timestamp but this fails when “Timestamp” is in HH:MM:SS.mmm (millisecond) or HH:MM:SS.mmmuuunnn (nanosecond) format. For timestamp with the nanosecond (millisecond) format, import the data using Excel “From Text” option and set the data type for column “Timestamp” to “Text,” so Excel does not automatically try to convert it.