



Mastering JSON (Unstructured) Data Processing In Snowflake



Current Version:v2023-06-21 [find latest version [here](#)]



Processing JSON In Snowflake

Episode-01 is a 40-min long video explaining how to query JSON data with Snowflake, a simple process once you understand Snowflake functions and keywords. It covers inserting JSON data into Snowflake tables, querying and validating JSON data, parsing nested JSON, handling date and timestamp data, flattening JSON data, and joining JSON data with other tables in Snowflake.

- [Watch](#): Difference Between JSON Vs CSV Data Files
- [Watch](#): History Of JSON for Data Engineers
- [Watch](#): Syntax Rules for JSON (Valid vs Invalid JSON)
- [Watch](#): JSON Data Into CSV Format
- [Watch](#): JSON Support for Data Types
- [Watch](#): Snowflake Native Support For JSON
- [Watch](#): JSON Data & Variant Data Type in Snowflake
- [Watch](#): Parsing & Inserting JSON Data into a snowflake table
- [Watch](#): Parsing JSON in Snowflake
- [Watch](#): Date Type Casting in Snowflake JSON Processing
- [Watch](#): JSON Data Access Pattern in Snowflake Summary
- [Watch](#): Date & Timestamp Support in JSON & Snowflake
- [Watch](#): Flattening JSON in Snowflake & Inserting in tables.

Load S3 JSON Into Snowflake

Episode-02 is a 25-min long video that answers "How To Load JSON Data Into Snowflake" using AWS S3 Bucket. This lesson demonstrates how to configure AWS S3 bucket, apply a policy, create a JSON file format object in snowflake and then run a copy command to load JSON data into snowflake using 3 different approaches. Validate the performance of these 3 different approaches using a query profile screen and identify the best approach as per your need and JSON data file complexity.

- [Watch](#): AWS S3 Bucket JSON Data Flow In Snowflake
- [Watch](#): Create External Stage (AWS S3) In Snowflake
- [Watch](#): S3 Bucket Configuration For Snowflake
- [Watch](#): S3 JSON Data Loading Into Snowflake Table
- [Watch](#): JSON Loading for Variant Table Vs Std Table
- [Watch](#): Loading Complex JSON Data Into Snowflake From S3

Load JSON Via SnowSQL

Episode-03 is a 30+ min long tutorial and answers **How To Load JSON Data Into Snowflake** using Internal names and table stages. This tutorial demonstrates how to put the json file using SnowSQL CLI into table/named stages, with compression and without compression, analyze the data and then run a copy command. This will also help you to validate the performance benchmark for uncompressed and compressed data. Validate the performance of these 3 different approaches using a query profile screen and identify the best approach as per your need and JSON data complexity. Once you finish this tutorial, you will be able to answer the following questions:

- [Watch](#): Snowflake Internal Staes & JSON
- [Watch](#): Internal Stages & Loading JSON Data Files
- [Watch](#): Analyze & Load JSON Data From the Internal Stage
- [Watch](#): JSON Data & Internal Table Stage
- [Watch](#): Compressed JSON File Loading
- [Watch](#): JSON Files Loading Using Partition Folder

External Table For JSON Data

Episode-04 is a 20-min long tutorial providing a complete hands on guide to explain how to consume JSON data stored in S3 location and how an external table can be created with the help of an external stage. This tutorial covers stage creation to file format creation to external table DDL statements besides many important facts about JSON data handling in snowflake using external tables.

- [Watch](#): JSON Data, External Table & External Stage
- [Watch](#): External Table on JSON Data
- [Watch](#): CSV vs JSON Data External Tables
- [Watch](#): External Table Auto Refresh
- [Watch](#): External Table on Internal Table

Query Complex & Dynamic JSON

Episode-05 is a "Query Complex & Dynamic JSON Data In Snowflake" tutorial Chapter-05 dives deeper into the intricacies of working with JSON data. You will learn about the different types of JSON files and the best approach to design your queries to extract the data you need. The tutorial covers the use of colon, dot, and bracket notation, as well as the powerful built-in SQL functions such as `check_json()`, `parse_json()` and `json_extract_path_text()` to simplify the process. This chapter is essential for anyone who needs to work with complex and dynamic JSON data in Snowflake.

- [Watch](#): json data structure classification
- [Watch](#): query json vs process json
- [Watch](#): query json using bracket notation
- [Watch](#): json sql query profile
- [Watch](#): query complex & nested json
- [Watch](#): use of `check_json()`
- [Watch](#): `check_json()` vs `parse_json()`
- [Watch](#): use of `json_extract_path_text` sql function
- [Watch](#): query dynamic json data

Flatten Function & JSON Processing

Episode-06 "Flatten Function & JSON Data Processing In Snowflake" explains the process of using the "flatten" table function in Snowflake to transform complex JSON, array, and object data structures into a flat, denormalized format that is easier to work with. This entire hands on tutorial will unleash the power of flattening table function to extract and query JSON in Snowflake with real life examples. The video covers the following topics:

- [Watch](#): what is flatten function in snowflake
- [Watch](#): flatten array example
- [Watch](#): flatten object example
- [Watch](#): flatten object & array together example
- [Watch](#): flatten & simple json example
- [Watch](#): flatten & nested json example
- [Watch](#): how to use outer parameter in flatten with json

Multiline JSON Processing

Episode-07 is a The video "Faster Multiline JSON Processing Using JSON File Format In Snowflake" provides an explanation of the important parameters required for

processing multiline JSON data files in Snowflake. The video covers topics such as the different file format parameters for JSON file processing, when to use the "strip outer array" and "strip null values" parameters, the difference between NDJSON and standard JSON, the role of date and timestamp parameters in the JSON file format, and the impact of different compression methods on JSON data loading performance.

- [Watch](#): file format parameters to process JSON files
- [Watch](#): NDJSON (New-line Delimiter JSON)
- [Watch](#): Strip Outer Array Param
- [Watch](#): Strip Null Values Param
- [Watch](#): Strip Null Values & Storage Impact
- [Watch](#): Date, Timestamp & Trim Space
- [Watch](#): Compression Parameter

Handle Duplicate Attribute in JSON

Episode-08 When duplicate attributes are encountered in a JSON file during data loading in Snowflake, the behavior depends on the file format attributes specified. To allow processing of duplicate attributes in the JSON file, you can modify the "ALLOW_DUPLICATE" attribute in the file format settings. However, it's important to note that in the case of a direct insert operation, Snowflake will treat the duplicate attribute as a syntax error. To debug and identify the duplicate attribute and this video will talk everything about duplicate handling in JSON.

- [Watch](#): File format parameters to process JSON files
- [Watch](#): How Duplicate Attributes Looks Like in JSON
- [Watch](#): Load JSON File with duplicate attributes
- [Watch](#): Multiple Duplicate Attributes & JSON Loading
- [Watch](#): Duplicate Attributes JSON String & Insert Operation

How To Use Parse_JSON, Try_Parse_JSON & Check_JSON

Episode-09 In this video, we explore Snowflake's JSON SQL functions: `Parse_JSON`, `Try_Parse_JSON`, and `Check_JSON`. It discusses their usage, parameters, limitations, and provides examples. Learn how to optimize performance and discover common use cases. Unlock the full potential of your data with these functions.



Mastering JSON (Unstructured) Data Processing In Snowflake



YouTube
Data Engineering Simplified [↗](#)

- [Watch](#):How To Use Parse_JSON SQL Function & Input/Output
- [Watch](#):Parse_JSON SQL Function Hands-on
- [Watch](#):How To Use Try_Parse_JSON SQL Function & Input/Output
- [Watch](#): Try_Parse_JSON SQL Function Hands-on
- [Watch](#):How To Use Check_JSON SQL Function & Input/Output
- [Watch](#):Check_JSON SQL Function Hands-on
- [Watch](#):Use Parse_JSON, Try_Parse_JSON & Check_JSON Together
- [Watch](#): Ingest Change In JSON Form & Duplicate Data
- [Watch](#): Ingest Change (JSON) & Stream Object
- [Watch](#): Different JSON Streaming Ingestion Approaches
- [Watch](#): Snowflake JSON Streaming Best Practices
- [Watch](#): Different Type Of Streams (Insert/append/all) + JSON
- [Watch](#): Best Practices - JSON & CDC

Continuous JSON File Ingestion

Episode-10 this video will be demonstrating how to ingest JSON data into snowflake continuously as real time streaming This hands-on comprehensive guide explains all the setup required to process streaming JSON data file into S3 bucket and then process it using Snowpipe object into snowflake table. Once you complete this Streaming JSON Data Into Snowflake video tutorial, you will be able to answer following question confidently

- [Watch](#): Streaming JSON To Snowflake Architecture
- [Watch](#): S3 & SQS Configuration for JSON Streaming
- [Watch](#): File Format, Stage & Pipe Configuration
- [Watch](#):Live Streaming of JSON Data to S3 to SnowPipe
- [Watch](#):Create View To Flatten JSON
- [Watch](#): Different JSON Streaming Ingestion Approaches
- [Watch](#):Snowflake JSON Streaming Best Practices

JSON & Change Data Capture

Episode-11 Change Data Capture (CDC) is a technique that allows for the capture and propagation of changes made to data in a database. By combining CDC and JSON processing in Snowflake using Stream and Task objects, users can easily and efficiently capture changes to JSON data and propagate those changes throughout their data pipelines. This can be particularly useful for real-time data streaming and analysis, as well as for building data integration and data transformation workflows.

- [Watch](#): JSON/CDC Overall Data Flow Diagram
- [Watch](#): How Changes Arrives (Timeline View)
- [Watch](#): Ingest JSON Data From Stage To Table