

Extraction Configuration

Last Modified on 03/11/2026 9:05 am EDT

Overview

In the extraction configuration, you decide which fields from the extracted table data in the DocIntel and Open AI GPT 4.0 are mapped to JSON. You can map fields from multiple tables by configuring them in the YAML files for backup and POD extraction. Only the fields defined in the configuration file are mapped and sent to the system.

There are two extraction YAML files configured in the system:

- backup-file-standardized-extraction-output-template
- pod-file-standardized-extraction-output-template

However, extraction at present happens only for POD. Extraction for backup is planned in the future.

Sample YAML settings

backup-file-standardized-extraction-output-template

```
kind: document
metadata:
  name: deduction/v1/documents/JsonTemplate/Extraction/backup-file-standardized-extraction-output-template
  description:
spec:
  templateAsJsonString: >
  {
    "fields": [
      {
        "name": "document_provider",
        "description": "The name of the entity or organization that issued this document. This is a critical field, so ensure to extract it accurately.",
        "type": "string"
      },
      {
        "name": "TotalDeduction",
        "description": "The total sum of all deducted amounts across the line items.",
        "type": "decimal"
      },
      {
        "name": "lines",
        "description": "An array containing all deduction claim line items.",
        "type": "array",
        "fields": [
          {
            "name": "description",
            "description": "The entire item description exactly as it appears in the document. take all the text from the table cell or field AS IS Including any text, numbers, or other characters.",
            "type": "string"
          },
          {
            "name": "deducted_amount",
            "description": "The total amount to be deducted for this line item. The value may sometimes appear in the summation column that totals the line items, sometimes in a separate field, and sometimes not appear at all.",
            "type": "decimal"
          },
          {
            "name": "Invoiced_qty",
            "description": "The quantity ordered and written in the invoice.",
            "type": "decimal"
          }
        ]
      }
    ]
  }
```

```

},
{
  "name": "received_qty",
  "description": "The quantity the customer actually received.",
  "type": "decimal"
},
{
  "name": "deducted_qty",
  "description": "The total quantity per line included in the deduction claim, whether due to items not received (shortage), damaged, or overages.",
  "type": "decimal"
},
{
  "name": "deducted_price_per_qty",
  "description": "The price per item that should have been paid by the customer.",
  "type": "decimal"
},
{
  "name": "deduction_reason",
  "description": "This field describes the reason for the deduction, which will be one of the following: Shortage, Damaged, or Overage.",
  "type": "string"
},
{
  "name": "invoice_number",
  "description": "The invoice number associated with this line item. In certain backup files, the invoice number and debit ID are combined in a single field as a numeric identifier that begins or ends with a letter. In these cases, the invoice number is the numeric portion only (for example, if the field contains 'Q9333541475', the invoice number is '9333541475' and the debit ID is 'Q9333541475').",
  "type": "string"
},
{
  "name": "purchaseOrderNumber",
  "description": "The purchase order number associated with this line item.",
  "type": "string"
},
{
  "name": "deduction_date",
  "description": "The deduction date. It Should be extracted in MM/DD/YYYY format",
  "type": "string"
},
{
  "name": "Debit",
  "description": "'Debit Memo' / 'Backup' ID number. In some backup files, the debit ID and invoice number appear together in a single field as a numeric identifier starting or ending with a character. In these cases, the debit ID is the full value (e.g., 'Q9110541475'), while the invoice number is just the numeric portion (e.g., '9110541475').",
  "type": "string"
},
{
  "name": "item_identifiers",
  "description": "A list of item identifiers and their types; typically, the identifier appears in a column cell and the type is indicated by the column header.",
  "type": "array",
  "fields": [
    {
      "name": "id",
      "description": "The item's identifier value.",
      "type": "string"
    },
    {
      "name": "type",
      "description": "The type or label of the identifier (e.g., UPC, SKU). Most of the time, the type of identifier is determined by either the column or field name, or by a label written near the identifier. For example, if there is a column named 'SKU #' and a cell contains the value '5235345', then '5235345' is the identifier and 'SKU #' is the type

```

```
e.",
    "type": "string"
  }
]
}
]
}
]
```

pod-file-standardized-extraction-output-template

```
kind: document
metadata:
  name: deduction/v1/documents/JsonTemplate/Extraction/pod-file-standardized-extraction-output-template
  description: spec
spec:
  templateAsString: |
    {
      "fields": [
        {
          "name": "documents",
          "description": "The POD/BOL PDF may contain one or more individual BOL or POD documents. Each document should be represented as a separate item in this 'documents' array.",
          "type": "array",
          "fields": [
            {
              "name": "document_provider",
              "description": "The name of the entity or organization that issued this document.",
              "type": "string"
            },
            {
              "name": "customer_sign",
              "description": "A handwritten signature of the customer. Usually, it is not in a specified field and can appear anywhere on the document. It may not exist. It is not the same as the carrier/shipper/driver signature.",
              "type": "bool"
            },
            {
              "name": "customer_sign_text",
              "description": "Extract the handwritten text from the customer_sign.",
              "type": "string"
            },
            {
              "name": "carrier_sign",
              "description": "The carrier's handwritten signature, taken from specific field designated for carrier signature . It is not the same as the customer signature.",
              "type": "bool"
            },
            {
              "name": "carrier_sign_text",
              "description": "Extract the handwritten text from the carrier_sign.",
              "type": "string"
            },
            {
              "name": "subject_to_count",
              "description": "Does the document contain a comment stating 'STC' (which stands for 'subject to count') or specifically the phrase 'subject to count'? Respond true if yes, false if not.",
              "type": "bool"
            },
            {
              "name": "bol",
              "description": "Does the document contain a comment stating 'BOL' (which stands for 'bill of lading') or specifically the phrase 'bill of lading'? Respond true if yes, false if not."
            }
          ]
        }
      ]
    }
```

```

      "description": "Provide the bill of lading (BOL) number for this document. Use the BOL number from the document header.",
      "type": "string"
    },
    {
      "name": "freight_charge_terms",
      "description": "Provide the freight charge terms for this document. The only acceptable values are: 'Prepaid', 'Collect', '3rdParty', or 'NotSpecified'.",
      "type": "string"
    },
    {
      "name": "total_packages_quantity",
      "description": "Specify the total number of packages for the goods described in the document.",
      "type": "decimal"
    },
    {
      "name": "total_lbs_quantity",
      "description": "Specify the total weight in pounds (LBS) for the goods described in the document.",
      "type": "decimal"
    },
    {
      "name": "lines",
      "description": "List all line items in the document. Each entry should represent a single line item.",
      "type": "array",
      "fields": [
        {
          "name": "invoice_number",
          "description": "Provide the invoice number associated with this line item.",
          "type": "string"
        },
        {
          "name": "description",
          "description": "The line item description (if exists).",
          "type": "string"
        },
        {
          "name": "item_identifier",
          "description": "Provide the item identifier value (such as SKU number or UPC code) for this line item. If the item identifier is not specified, or if the line item is at the purchase order level, this field can be left empty."
        },
        {
          "name": "item_identifier_type",
          "description": "Specify the type or label of the identifier (e.g., 'SKU', 'UPC'). Leave this field empty if the item identifier was not specified or if its type is unknown."
        },
        {
          "name": "order_number",
          "description": "Provide the purchase order number (often labeled as PO#, P.O#, or similar) for this line item. If the purchase order number is not specified at the line item level, use the purchase order number from the document header.",
          "type": "string"
        },
        {
          "name": "packages_quantity",
          "description": "Specify the total number of packages for this line item.",
          "type": "decimal"
        },
        {
          "name": "lbs_quantity",
          "description": "Specify the total weight in pounds (LBS) for this line item.",
          "type": "decimal"
        }
      ]
    }
  ],
},

```

```

{
  "name": "notations",
  "description": "Provide a list of customer notations regarding the actual goods received compared to the document content. Notations may be handwritten, printed, or stamped on the document. For each entry, specify as many details as possible, including: quantity, quantity type (unit of measurement), item identifier (if the note refers to a specific item), item identifier type, and the notation type.",
  "type": "array",
  "fields": [
    {
      "name": "notation_type",
      "description": "Specify the type of notation for this document. The only possible values are: 'Received', 'Shortage', 'Damaged', or 'Overage'. Use 'Received' when the customer confirms receipt of goods; use 'Shortage', 'Damaged', or 'Overage' if the customer is claiming a shortage, damage, or overage in the goods received.",
      "type": "string"
    },
    {
      "name": "quantity",
      "description": "The number of items, as a decimal value.",
      "type": "decimal"
    },
    {
      "name": "quantity_type",
      "description": "The unit of measurement for the quantity (for example: 'Pallets', 'Skids', 'PCS').",
      "type": "string"
    },
    {
      "name": "item_identifier",
      "description": "Item identifier value (e.g., SKU number, UPC code) for this entry. If empty, this entry applies to all items in the BOL/POD."
    },
    {
      "name": "item_identifier_type",
      "description": "The type or label of the identifier (e.g., 'SKU', 'UPC')."
    },
    {
      "name": "handwritten_source",
      "description": "Write the actual text as it appears in the original source (for example: 'Received 12 skids'). Do not summarize or interpret—copy the text exactly.",
      "type": "string"
    }
  ]
},
{
  "name": "handwritten",
  "description": "A list of all handwritten text or markings found anywhere in the document. Each entry should represent a separate handwritten section, determined by its physical separation in the document. If a handwritten signature is identified, label it explicitly as 'signature' or a similarly clear term, even if no characters are discernible.",
  "type": "array",
  "fields": [
    {
      "name": "handwritten",
      "description": "The specific handwritten text or marking from a single physically separated area (e.g., a character, word, line, or paragraph). If the handwritten content is a signature, label it as 'signature' or a similar clear term, even if the signature's characters cannot be identified.",
      "type": "string"
    },
    {
      "name": "handwritten_with_context",
      "description": "The specific handwritten text or marking from a single physically separated area, combined with nearby printed text to provide context. If the handwritten content is a signature, label it as 'signature' or a similar clear term, and include relevant printed text nearby for context.",
      "type": "string"
    }
  ]
}

```

```
]
}
]
}
]
}
```

Parameter	Description
Name	Determines the name of the field in JSON.
Description	Description of the field.
Type	Determines the datatype of the field.