

IBM InfoSphere DataStage und Quality Stage  
Version 11 Release 3

## *Zugriff auf unstrukturierte Daten*





IBM InfoSphere DataStage und Quality Stage  
Version 11 Release 3

## *Zugriff auf unstrukturierte Daten*



**Hinweis**

Vor Verwendung dieser Informationen und des darin beschriebenen Produkts sollten die Informationen unter „Bemerkungen und Marken“ auf Seite 67 gelesen werden.

Diese Veröffentlichung ist eine Übersetzung des Handbuchs  
*IBM InfoSphere DataStage and QualityStage, Version 11 Release 3, DS Guide to Extracting Unstructured Data*,  
IBM Form SC19-4330-00,  
herausgegeben von International Business Machines Corporation, USA

© Copyright International Business Machines Corporation 2013, 2014

Informationen, die nur für bestimmte Länder Gültigkeit haben und für Deutschland, Österreich und die Schweiz nicht zutreffen, wurden in dieser Veröffentlichung im Originaltext übernommen.

Möglicherweise sind nicht alle in dieser Übersetzung aufgeführten Produkte in Deutschland angekündigt und verfügbar; vor Entscheidungen empfiehlt sich der Kontakt mit der zuständigen IBM Geschäftsstelle.

Änderung des Textes bleibt vorbehalten.

Herausgegeben von:  
TSC Germany  
Kst. 2877  
September 2014

© Copyright IBM Corporation 2013, 2014.

---

# Inhaltsverzeichnis

## **Kapitel 1. Stage 'Unstructured Data'. . . 1**

Jobs mit Stages 'Unstructured Data' entwerfen . . . . .	1
Job definieren, der eine Stage 'Unstructured Data'	
enthält . . . . .	1
Daten aus Microsoft Excel extrahieren . . . . .	2
Daten in neue Microsoft Excel-Datei schreiben. . . . .	18
Daten in vorhandene Microsoft Excel-Dateien	
schreiben . . . . .	32

## **Kapitel 2. Referenz . . . . . 39**

Datentypkonvertierungen von Microsoft Excel in	
InfoSphere DataStage . . . . .	39
Datentypkonvertierungen von InfoSphere DataStage	
in Microsoft Excel . . . . .	45
Jobabbruchbedingungen in Microsoft Excel. . . . .	46

## **Kapitel 3. Fehlerbehebung . . . . . 49**

## **Kapitel 4. Umgebungsvariablen: Stage**

<b>'Unstructured Data' . . . . .</b>	<b>53</b>
CC_JNI_EXT_DIRS . . . . .	53
CC_JVM_OPTIONS. . . . .	53
CC_JVM_OVERRIDE_OPTIONS . . . . .	53
CC_IGNORE_TIME_LENGTH_AND_SCALE . . . . .	53

CC_MSG_LEVEL . . . . .	54
CC_UNST_JAVA_HEAP . . . . .	54

## **Anhang A. Eingabehilfen in den Produkten . . . . . 55**

## **Anhang B. Befehlszeilensyntax lesen 57**

## **Anhang C. Syntaxdiagramme lesen . . 59**

## **Anhang D. Kontaktaufnahme mit IBM 61**

## **Anhang E. Auf Produktdokumentation zugreifen. . . . . 63**

## **Anhang F. Feedback zur Produktdokumentation geben . . . . . 65**

## **Bemerkungen und Marken . . . . . 67**

## **Index . . . . . 73**



---

## Kapitel 1. Stage 'Unstructured Data'

Unstrukturierte Daten sind Informationen, die nicht über ein vordefiniertes Datenmodell verfügen oder nicht gut in relationale Tabellen passen. Unstrukturierte Daten können Text aus Handbüchern, Journals, Metadaten, Audio, Videodateien, der Hauptteil von Textverarbeitungsprogrammdokumenten, Webseiten und Präsentationsdiagramme sein. In diesem Release unterstützt die Stage **Unstructured Data** nur Microsoft Excel-Dateien als Datenquellen.

Über die Stage **Unstructured Data** können Sie die folgenden Operationen ausführen:

- Informationen aus unstrukturierten Datenquellen extrahieren und in Ihre Jobs integrieren. Daten aus Microsoft Excel-Arbeitsblättern, die unterschiedliche Spaltendefinitionen aufweisen, können über eine Stage **Unstructured Data** extrahiert werden.
- Daten in Microsoft Excel-Arbeitsblätter schreiben.

---

### Jobs mit Stages 'Unstructured Data' entwerfen

Sie können Stages **Unstructured Data** in Ihren Jobs verwenden, um im Kontext dieser Jobs Daten aus unstrukturierten Datenquellen zu lesen oder Daten in unstrukturierte Datenquellen zu schreiben.

#### Job definieren, der eine Stage 'Unstructured Data' enthält

Sie müssen einen Job erstellen, der die Stage **Unstructured Data** enthält, erforderliche weitere Stages hinzufügen und die notwendigen Links erstellen, bevor Sie Daten aus Microsoft Excel-Dateien lesen bzw. Daten in Microsoft Excel-Dateien schreiben können.

#### Vorgehensweise

1. Klicken Sie im Designer-Client auf **Datei > Neu**.
2. Klicken Sie im Fenster **Neu** auf das Symbol **Paralleler Job** und klicken Sie dann auf **OK**.
3. Klicken Sie in der Palette auf **Datei**.
4. Ziehen Sie das Symbol **Stage 'Unstructured Data'** auf den Entwicklungsbereich.
5. Erstellen Sie Stages für den Job.
6. Wählen Sie die Kategorie **Allgemein** im Palettenmenü auf der linken Seite des Designer-Clients aus und erstellen Sie dann die erforderlichen Links für den Job.
7. (Optional) Doppelklicken Sie auf das Symbol **Stage 'Unstructured Data'**, um die folgenden Attribute einzugeben oder zu modifizieren:
  - **Stage:** Modifizieren Sie den Standardnamen der Stage. Sie können bis zu 255 Zeichen eingeben. Sie können den Namen der Stage alternativ im Jobentwicklungsbereich modifizieren.
  - **Beschreibung:** Geben Sie eine Beschreibung der Stage ein.
8. Klicken Sie auf **Speichern**.

## Daten aus Microsoft Excel extrahieren

Über die Stage **Unstructured Data** können Sie mehrere Datentypen aus einer Microsoft Excel-Datei extrahieren.

### Voraussetzungen

Stellen Sie vor dem Start der Installation und Konfiguration der Stage **Unstructured Data** sicher, dass die System- und Softwarevoraussetzungen erfüllt werden.

### Vorbereitende Schritte

#### Vorgehensweise

1. Installieren Sie Information Server in der Sprache, in der die zu extrahierende Microsoft Excel-Datei vorliegt.
2. Stellen Sie sicher, dass die Microsoft Excel-Funktion zum Anzeigen des Inhalts von Microsoft Excel-Arbeitsblättern (z. B. Microsoft Excel, Microsoft Excel Viewer oder IBM Lotus Symphony) auf Ihrem Clientsystem installiert ist.
3. Stellen Sie sicher, dass die Dateierweiterungen `.xls` und `.xlsx` Ihrer Microsoft Excel-Anzeigefunktion ordnungsgemäß zugeordnet sind.

#### Unterstützte Datenquellen

Die Stage **Unstructured Data** unterstützt nur Microsoft Excel-Dateien als Quelldateien.

Die Stage **Unstructured Data** unterstützt die folgenden Dateiformate:

- Microsoft Excel 97-2003 OLE2 (`.xls`), einschließlich Unterstützung der mit Kennwort verschlüsselten Datei
- Microsoft Excel 2007-2010 OOXML (`.xlsx`), einschließlich Unterstützung der mit Kennwort verschlüsselten Datei

Die Stage **Unstructured Data** unterstützt Microsoft Excel-Dateien nicht, die von Microsoft Excel for Mac erstellt werden.

#### Datenbereiche

Über die Stage **Unstructured Data** können Sie Daten aus einem angegebenen Datenbereich in einem Microsoft Excel-Arbeitsblatt extrahieren.

Ein Datenbereich stellt eine Zelle, eine Zeile, eine Spalte oder eine Auswahl von Zellen dar, die mindestens einen kontinuierlichen Block von Zellen enthält. Ein Datenbereich wird durch den Bereichsausdruck angegeben. In der Stage **Unstructured Data** können Sie den zu extrahierenden Datenbereich über einen Bereichsausdruck angeben.

Beispielsweise wird über 'Employee\_Salary!A1:G8' ein Datenbereich beschrieben, in dem die erste Zelle A1 und die letzte Zelle G8 im Arbeitsblatt 'Employee\_Salary' ist.

Tabelle 1. Beispiel einer Microsoft Excel-Datei: Arbeitsblatt 'Employee\_Salary'

	A	B	C	D	E	F	G
1	EMPNO	FIRSTNAME	LASTNAME	DEPT	JOB	SALARY	BONUS
2	20	MICHAEL	THOMPSON	B01	MANAGER	94250	800
3	30	SALLY	KWAN	C01	MANAGER	98250	800
4	60	IRVING	STERN	D11	MANAGER	72250	500



Tabelle 1. Beispiel einer Microsoft Excel-Datei: Arbeitsblatt 'Employee\_Salary' (Forts.)

	A	B	C	D	E	F	G
1	EMPNO	FIRSTNAME	LASTNAME	DEPT	JOB	SALARY	BONUS
5	70	EVA	PULASKI	D21	MANAGER	96170	700
6	50	JOHN	GEYER	E01	MANAGER	80175	800
7	90	ELEEN	HENDERSON	E11	MANAGER	89750	600
8	100	THEODORE	SPENSER	E21	MANAGER	86150	500

Die Stage **Unstructured Data** ordnet die Microsoft Excel-Zeile und -Spalte im angegebenen Datenbereich der InfoSphere DataStage-Zeile und -Spalte zu und extrahiert die Datensätze.

In der folgenden Tabelle werden die von der Stage **Unstructured Data** extrahierten Datensätze beim Bereichsausdruck 'Employee\_Salary!A2:G8' beschrieben.

Tabelle 2. Beispiel für DataStage-Zeile und -Spalte

20	MICHAEL	THOMPSON	B01	MANAGER	94250	800
30	SALLY	KWAN	C01	MANAGER	98250	800
60	IRVING	STERN	D11	MANAGER	72250	500
70	EVA	PULASKI	D21	MANAGER	96170	700
50	JOHN	GEYER	E01	MANAGER	80175	800
90	ELEEN	HENDERSON	E11	MANAGER	89750	600
100	THEODORE	SPENSER	E21	MANAGER	86150	500

Über die Eigenschaft **Bereichsoption** der Stage **Unstructured Data** können Sie den Datenbereich durch die Auswahl der Option **Anfangszeile angeben** oder **Gesamten Datenbereich angeben** angeben. Wenn Sie die Option **Anfangszeile angeben** auswählen, müssen Sie die Anfangszeile angeben. Die Stage **Unstructured Data** ermittelt daraufhin die Endzeile des Datenbereichs. Wenn Sie die Option **Gesamten Datenbereich angeben** auswählen, müssen Sie die Anfangs- und Endzeile des zu extrahierenden Datenbereichs angeben.

Wenn Sie den Wert von Zellen in der ersten Zeile als Spaltenname für IBM® InfoSphere DataStage verwenden wollen, können Sie die Eigenschaft **Spaltenheader** konfigurieren. Wenn die Eigenschaft **Spaltenheader** auf **Erste Zeile von Datenbereichen** gesetzt ist und wenn Sie den Bereichsausdruck als 'Employee\_Salary!A1:G8' angeben, wird die erste Zeile als Header behandelt, und der Wert der Zellen in der ersten Zeile wird als DataStage-Standardspaltenname im Job verwendet. Sie können den Bereichsausdruck während der Entwicklung über die Stage **Unstructured Data** generieren.

## Aus Microsoft Excel extrahierbare Datentypen

Über die Stage **Unstructured Data** können Sie mehrere Datentypen aus einer Microsoft Excel-Datei extrahieren.

### Dateieigenschaften

In der folgenden Tabelle werden die Informationen aufgelistet, die als Dateieigenschaften extrahiert werden können:

Tabelle 3. Als Dateieigenschaften extrahierbare Daten

Daten	Beschreibung
Dateiname	Name der Datei. Beispiel: Workbook1.xls
Dateipfad	Pfad der Datei. Beispiel: C:\excel\Workbook1.xls
Dateigröße	Größe der Datei in Byte.
Datum der letzten Modifizierung	Datum und Zeit der letzten Dateiänderung.

### Dokumenteigenschaften

In der folgenden Tabelle werden die Informationen aufgelistet, die als Dokumenteigenschaften extrahiert werden können:

Tabelle 4. Als Dokumenteigenschaften extrahierbare Daten

Daten	Beschreibung
Autoren	Autoren des Dokuments.
Dokumentkommentare	Kommentare des Dokuments.
Datum der Inhaltserstellung	Datum und Zeit der Dokumenterstellung.
Schlüsselwörter	Schlüsselwörter des Dokuments.
Überarbeitungsnummer	Überarbeitungsnummer des Dokuments.
Betreff	Betreff des Dokuments.
Titel	Titel des Dokuments.
Unternehmen	Unternehmenseigentumswert des Dokuments.
Kategorie	Kategorie des Dokuments.
Manager	Manager des Dokuments.
Angepasste Eigenschaften	Angepasste Eigenschaften des Dokuments. Sie müssen den Namen der zu extrahierenden angepassten Eigenschaft angeben.

### Arbeitsblattinformationen

In der folgenden Tabelle werden die Informationen aufgelistet, die als Arbeitsblattinformationen extrahiert werden können:

Tabelle 5. Als Arbeitsblattinformationen extrahierbare Daten

Daten	Beschreibung
Arbeitsblattname	Name des Microsoft Excel-Arbeitsblatts.
Header (links, zentriert, rechts)	Header der angegebenen Position.
Fußzeile (links, zentriert, rechts)	Fußzeile der angegebenen Position.

### Zeileninformationen

In der folgenden Tabelle werden die Informationen aufgelistet, die als Zeileninformationen extrahiert werden können:

Tabelle 6. Als Zeileninformationen extrahierbare Daten

Daten	Beschreibung
Zeilennummer	Microsoft Excel-Zeilennummer im Arbeitsblatt. Die erste Zeilennummer ist 1.

Tabelle 6. Als Zeileninformationen extrahierbare Daten (Forts.)

Daten	Beschreibung
Ausgeblendet	Gibt an, ob die Zeile ausgeblendet ist. Wenn die Zeile oder das Arbeitsblatt, zu dem diese Zeile gehört, ausgeblendet ist, wird 'true' geschrieben.

### Zelleninformationen

Sie können die Zelleninformationen basierend auf der Microsoft Excel-Spalte oder -Zellenposition extrahieren. Sie können die Microsoft Excel-Quellenspalte basierend auf der relativen Position im Datenbereich angeben, wenn Sie die Zelleninformationen basierend auf der Microsoft Excel-Spalte extrahieren.

In der folgenden Tabelle werden die Informationen aufgelistet, die als Zelleninformationen extrahiert werden können:

Tabelle 7. Als Zelleninformationen extrahierbare Daten

Daten	Beschreibung
Wert	Wert einer Zelle. Wenn die Zelle eine Formel enthält, extrahiert die Stage den Wert aus dem Cache.
Kommentar	Kommentar einer Zelle.
Kommentarautor	Autor des Kommentars einer Zelle.
Formel	Formel einer Textzelle.
Hyperlinktyp	Typ von Hyperlink einer Zelle.
Hyperlinkadresse	Die Adresse, auf die dieser Hyperlink verweist. Das Format hängt vom Typ dieses Hyperlinks ab.
Hyperlinkbezeichnung	Textbezeichnung für diesen Hyperlink.

## Jobs entwerfen, die Daten aus Microsoft Excel-Datei extrahieren

Über die Stage **Unstructured Data** können Sie Jobs entwerfen, die unstrukturierte Daten aus Microsoft Excel-Dateien lesen.

### Informationen zu diesem Vorgang

Die Stage **Unstructured Data** liest Daten aus den Microsoft Excel-Dateien und übergibt die Zeilen an eine Stage **Transformer**. Die Stage **Transformer** setzt die Daten um und lädt sie dann in den ODBC-Connector. Wenn Sie die Stage **Unstructured Data** für das Lesen von Daten aus den Microsoft Excel-Dateien konfigurieren, erstellen Sie nur einen Ausgabelink.

### Vorgehensweise

1. Definieren Sie einen Job, der eine Stage 'Unstructured Data' enthält.
2. Wenn die Stage **Unstructured Data** als eine Quellenstage für das Lesen von unstrukturierten Daten aus Microsoft Excel-Dateien eingerichtet werden soll, führen Sie die folgenden Schritte aus:
  - a. Konfigurieren Sie die Stage 'Unstructured Data' als Quelle.
  - b. Modifizieren Sie die Spaltendefinition im Link.
3. Kompilieren Sie den Job und führen Sie ihn aus.

## Stage 'Unstructured Data' als Quelle konfigurieren:

Wenn Sie einen Job der Stage **Unstructured Data** erstellen, müssen Sie die Stage **Unstructured Data** so konfigurieren, dass sie die Daten extrahiert und die Ausgabe im vom Benutzer benötigten Datentyp generiert.

### Vorgehensweise

1. Doppelklicken Sie im Entwicklungsbereich für parallele Jobs auf die Stage **Unstructured Data**.
2. Wählen Sie in der Liste **Dokumenttyp** den Eintrag **Excel** aus.
3. Klicken Sie auf **Konfigurieren**, um weitere Eigenschaften zu konfigurieren und das Mapping zwischen Microsoft Excel-Elementen und DataStage-Spalten zu definieren.
4. Geben Sie die Dateinamendetails im Teilfenster **Datenquelle** an:
  - a. Geben Sie den Namen der Datei, aus der Sie die Daten lesen wollen, im Feld **Dateiname** an. Die Jobkompilierung schlägt fehl, wenn dieses Feld leer ist. Wenn die Datei kennwortgeschützt ist, geben Sie das Kennwort im Feld **Kennwort** an.
  - b. Optional: Wenn Sie im Dateinamen Platzhalterzeichen angeben, wählen Sie **Schablonendatei für Entwicklungszeit verwenden** aus und geben Sie einen Schablonendateinamen an. Die Schablonendatei wird für nachfolgende Konfigurationsschritte und nicht während der Ausführung verwendet. Geben Sie einen Wert für **Schablonenkennwort** an, wenn die angegebene Schablonendatei kennwortgeschützt ist.
  - c. Optional: Klicken Sie auf **Ansicht**, um die externe Microsoft Excel-Anzeige-funktion zu starten. Sie können den Inhalt der Microsoft Excel-Datei bestätigen, mit der Sie arbeiten.
  - d. Klicken Sie auf **Laden**.
5. Geben Sie die Datenbereichsdetails, die aus der Microsoft Excel-Datei gelesen werden sollen, im Teilfenster **Leseoptionen** an.
  - a. Optional: Geben Sie einen Wert für **Bereichsoption** an. Wenn Sie **Anfangs-zeile angeben** auswählen, brauchen Sie nur die erste Zeile anzugeben. Die Stage **Unstructured Data** sucht die letzte Zeile während der Ausführung. Wenn Sie **Gesamten Datenbereich angeben** auswählen, müssen Sie die Anfangs- und Endzeile angeben.
  - b. Optional: Geben Sie **Bereichsausdruck** an. **Bereichsausdruck** ist eine erforderliche Eigenschaft während der Ausführung. Sie kann jedoch leer sein, wenn Sie auf die Schaltfläche **Laden** klicken. Die Stage **Unstructured Data** durchsucht das gesamte Dokument und listet die Datenbereichskandidaten im Listenfeld **Datenbereich der Schablone** im Teilfenster **Importieren** auf. Die Eigenschaft **Bereichsausdruck** wird mit dem entsprechenden Wert festgelegt, wenn Sie auf **Importieren** im Teilfenster **Importieren** klicken.
  - c. Optional: Wenn Sie Arbeitsblattnamen im Bereichsausdruck überspringen wollen, geben Sie die entsprechenden Namen im Feld **Zu überspringende Arbeitsblattnamen** an. Verwenden Sie dieses Feld, wenn die Arbeitsblattnamen im Bereichsausdruck übergangen werden.
  - d. Optional: Geben Sie **Erste Zeile von Datenbereichen** an. Wenn Sie während der Entwicklung **Ohne** auswählen, werden Microsoft Excel-Spaltennamen im Teilfenster **Zuordnen** im Format "Column#Spaltennummer(ColumnExcel-Spaltenbeschriftung)" ausgedrückt. Wenn Sie **Erste Zeile ist Header** auswählen, wird der Wert der ersten Zeile im Teilfenster **Zuordnen** angezeigt.

Wenn Sie während der Ausführung **Ohne** auswählen, wird die erste Zeile extrahiert. Wenn Sie **Erste Zeile ist Header** auswählen, wird die erste Zeile übersprungen.

6. Geben Sie die Datenbereichsdetails, die importiert werden sollen, im Teilfenster **Importieren** an.
  - a. Wählen Sie in **Datenbereich der Schablone** einen Datenbereich aus.
  - b. Optional: Wenn Sie weitere Microsoft Excel-Elemente wie Dokumenteigenschaften extrahieren wollen, wählen Sie die Registerkarte **Eigenschaft** und die zu extrahierenden Elemente aus.
  - c. Klicken Sie auf **Importieren**.
7. Sie können das Mapping zwischen InfoSphere DataStage-Spalten und Microsoft Excel-Elementen im Teilfenster **Zuordnen** definieren.
  - a. Sie können das Mapping zwischen InfoSphere DataStage-Spalten und importierten Microsoft Excel-Elementen definieren. Sie können InfoSphere DataStage-Spaltenmappings hinzufügen oder die Spaltenreihenfolge ändern, indem Sie auf die Schaltfläche **Nach oben**, **Nach unten**, **Einfügen** oder **Löschen** klicken. Geben Sie in **InfoSphere DataStage-Spalte** den InfoSphere DataStage-Spaltennamen für jedes Microsoft Excel-Element an. In **Microsoft Excel-Element** können Sie das Element auswählen, das Sie der InfoSphere DataStage-Spalte zuordnen wollen. Alle Elemente, die im Teilfenster **Importieren** ausgewählt werden können, werden in jeder Zelle aufgelistet. In **Importoption** können Sie das Microsoft Excel-Element auswählen, wenn Importoptionen verfügbar sind. Wenn Sie z. B. **Microsoft Excel-Spalte** in **Excel-Element** auswählen, sind die Optionen **Wert**, **Kommentar**, **Autor des Kommentars**, **Formel**, **Hyperlinktyp** und **Hyperlinkadresse** verfügbar.
  - b. Klicken Sie auf **OK**.
8. Geben Sie die erforderlichen Details auf den Registerkarten **Eigenschaften** und **Erweitert** an.
9. Klicken Sie auf **OK**, um die angegebenen Einstellungen zu speichern.

### Spaltendefinition im Link modifizieren:

Sie können Elemente der Spaltendefinition wie **SQL-Typ**, **Länge**, **Nachkommastellen** und **Kann Nullwerte enthalten** im Link modifizieren. Wenn Sie den vom Konfigurationsfenster importierten Spaltennamen ändern wollen, starten Sie das Konfigurationsfenster erneut, um den Namen anzugeben.

### Vorgehensweise

1. Doppelklicken Sie im Entwicklungsbereich für parallele Jobs auf die Stage **Unstructured Data**.
2. Wählen Sie die Registerkarte **Ausgabe** und dann den Ausgabelink in **Ausgabename (nachgeschaltete Stage)** aus.
3. Bearbeiten Sie **SQL-Typ**, **Länge** und **Nachkommastellen** für jede Spalte.
4. Klicken Sie auf **OK**, um die Änderungen zu speichern.

### Jobparameter verwenden

Die Stage **Unstructured Data** kann keine neuen Jobparameter im Konfigurationsfenster erstellen. Sie können jedoch die Jobparameter im Konfigurationsfenster verwenden. Sie müssen Jobparameter im Fenster **Jobeigenschaften** vor oder nach der Arbeit im Konfigurationsfenster erstellen, indem Sie **Bearbeiten** > **Jobeigenschaften** im Client von IBM InfoSphere DataStage und QualityStage Designer auswäh-

len. Weitere Informationen zum Erstellen von Jobparametern finden Sie in Lektion 2.4, Parameter hinzufügen, im IBM InfoSphere DataStage-Lernprogramm für parallele Jobs.

Ein Jobparameter wird im Konfigurationsfenster mit dem Zeichen # angegeben. Beispielsweise wird der Jobparameter *Dateiname* im Konfigurationsfenster als #**Dateiname** angegeben. Für ein Feld mit dem Datentyp 'string' wie die Eigenschaft **Dateiname** können Sie den Namen des Jobparameters von # umschlossen direkt eingeben.

Wenn Sie für die Eigenschaft **Listentyp** wie z. B. **Bereichsoption** Jobparameter verwenden wollen, müssen Sie einen Listentypparameter erstellen, der eine Liste von Zeichenfolgevariablen enthält. Die Zeichenfolgevariablen müssen mit dem Beschriftungstext der entsprechenden Eigenschaft im Konfigurationsfenster übereinstimmen. Wenn Sie z. B. Jobparameter für die Eigenschaft **Bereichsoption** verwenden wollen, müssen Sie einen Listentypparameter erstellen, der die Zeichenfolgevariablen **Anfangszeile angeben** und **Gesamten Datenbereich angeben** enthält. Wählen Sie nach der Erstellung eines Jobparameters <Parametrisieren...> im Konfigurationsfenster aus und geben Sie den Jobparameternamen vom Zeichen # umschlossen im Dialogfeld **Eingabeparameter** an. Klicken Sie auf **Laden**, um Variablen in der Anzeige **Jobparameter auflösen** zu bearbeiten oder auszuwählen.

### Optionen für das Lesen von Daten aus Microsoft Excel-Dateien

Über die folgenden Optionen können Sie modifizieren, wie die Stage **Unstructured Data** Daten liest.

#### Fehlerbehandlung:

Sie können angeben, ob eine Fehlernachricht protokolliert und der Job gestoppt werden soll, wenn beim Extrahieren von Daten aus der Datei ein Fehler auftritt.

Sie können die Eigenschaft **Fehleraktion** auf **Fehlschlagen** oder **Überspringen** setzen. Der Standardwert ist **Fehlschlagen**.

- Wenn Sie **Fehlschlagen** auswählen, wird von der **Unstructured Data** ein nicht behebbarer Fehler protokolliert und wird der Job gestoppt, wenn beim Extrahieren von Daten ein Fehler auftritt.
- Wenn Sie **Überspringen** auswählen, wird von der Stage **Unstructured Data** eine Warnung protokolliert und die Verarbeitung der verbleibenden Eingabefelder und Datensätze während der Datenextraktion fortgesetzt.

#### Nullzeilenbehandlung:

Sie können die Stage **Unstructured Data** so konfigurieren, dass Zeilen mit Nullwerten in ihren Zellen, die extrahiert werden, übersprungen werden.

Sie können die Eigenschaft **Nullzeilen überspringen** auf **Ja** oder **Nein** setzen. Der Standardwert ist **Nein**.

- Sollen Zeilen mit Nullwerten in ihren Zellen übersprungen werden, wählen Sie **Ja** aus.
- Soll die Verarbeitung von Zeilen mit Nullwerten in ihren Zellen fortgesetzt werden, wählen Sie **Nein** aus.

## Wert einer bestimmten Zelle oder angepasste Eigenschaften extrahieren:

Sie können den zu extrahierenden Wert einer Zelle oder die zu extrahierenden angepassten Eigenschaften angeben.

### Vorgehensweise

1. Doppelklicken Sie im Jobentwicklungsbereich auf das Symbol **Stage 'Unstructured Data'**.
2. Klicken Sie auf **Konfigurieren**.
3. Wählen Sie die Registerkarte **Erweitert** im Teilfenster **Importieren** aus.
4. Soll der Wert einer bestimmten Zelle importiert werden, wählen Sie **Bestimmte Zelle** in der Spalte **Typ** aus und geben Sie die Zellenposition in der Spalte **Wert** an. Wenn Sie z. B. den Wert in der Zelle 'A1' importieren wollen, geben Sie 'A1' in die Spalte **Wert** ein.
5. Soll eine angepasste Eigenschaft importiert werden, wählen Sie **Angepasste Eigenschaft** in der Spalte **Typ** aus und geben Sie den Eigenschaftswert im Feld **Wert** an. Wenn Sie z. B. die angepasste Eigenschaft namens 'Prop1' importieren wollen, geben Sie 'Prop1' im Feld **Wert** an.
6. Klicken Sie auf **Importieren**.
7. Definieren Sie das Mapping zwischen InfoSphere DataStage-Spalten und Microsoft Excel-Elementen im Teilfenster **Zuordnen** an.
8. Klicken Sie auf **OK**.

### Ausgeblendete Spalte weitergeben:

Sie können die Aktionen angeben, die für die ausgeblendeten Spalten bei der Spaltenweitergabe während der Ausführung ausgeführt werden sollen.

Sie können die Eigenschaft **Ausgeblendete Spalten** in der Kategorie **Spaltenweitergabe während der Ausführung** auf das Extrahieren oder Überspringen der ausgeblendeten Spalten für die Spaltenweitergabe während der Ausführung setzen.

- Wenn Sie **Extrahieren** auswählen, werden die ausgeblendeten Spalten extrahiert.
- Wenn Sie **Überspringen** auswählen, werden die ausgeblendeten Spalten übersprungen.

Der Standardwert ist **Extrahieren**.

### Spaltenweitergabe während der Ausführung

In InfoSphere DataStage können Sie einen Job konfigurieren, um zusätzliche Spalten, die nicht in den Metadaten definiert sind, im weiteren Verlauf des Jobs weiterzugeben. Dieser Prozess wird als Spaltenweitergabe während der Ausführung bezeichnet.

Wenn die Spaltenweitergabe während der Ausführung aktiviert ist, gibt die Stage **Unstructured Data** Microsoft Excel-Spalten auf der Basis des ersten Datenbereichs weiter. Wenn Platzhalterzeichen im Dateinamen verwendet werden, wird die erste Datei verwendet, die mit dem Ausdruck übereinstimmt. Die Einstellung der Eigenschaft **Ausgeblendete Spalten** legt fest, ob eine ausgeblendete Spalte weitergegeben wird. Für jede weitergegebene Microsoft Excel-Spalte werden nur Zellenwerte extrahiert. Sie können weitere Spalten im Konfigurationsfenster definieren, um Informationen wie den Dateinamen, den Arbeitsblattnamen oder die Zeilennummer zu extrahieren.

### Spaltenbenennungsregeln

InfoSphere DataStage-Spalten werden auf der Basis des Microsoft Excel-Spaltenbuchstabens des ersten Datenbereichs benannt. Dem Spaltennamen wird "**Column\_**" gefolgt vom Microsoft Excel-Spaltenbuchstaben vorangestellt. Beispiel: Column\_A, Column\_B, Column\_C usw.

Wenn im Job bereits eine Spalte mit diesem Namen vorhanden ist, wird der Job abgebrochen.

### Datentypen

Für alle von der Stage **Unstructured Data** hinzugefügten Spalten wird der Unicode-Typ **Varchar** mit nicht definierter Länge festgelegt.

### Beispiele für das Extrahieren von Daten aus Microsoft Excel-Dateien

Sie können Beispieljobs erzeugen, die Daten aus Microsoft Excel-Dateien extrahieren.

Extrahieren Sie die Datei *IS-Installation\Clients\Samples\Connectors\UnstructuredData\_Samples.zip*, um die Dateien für die Beispiele abzurufen.

#### Beispiel 1: Daten aus einem Bereich in einer Microsoft Excel-Datei extrahieren:

Erstellen Sie einen Job, der über die Stage **Unstructured Data** Daten aus einem Bereich in einem Microsoft Excel-Arbeitsblatt abrufen.

#### Informationen zu diesem Vorgang

In diesem Beispiel wird die Microsoft Excel-Beispieldatei *Employee1.xls* verwendet, die Details von Mitarbeitern in einem Unternehmen enthält. Diese Beispieldatei enthält drei Arbeitsblätter: Sheet1, Sheet2 und Sheet3. Sheet1 enthält Informationen zu den Mitarbeitern in jeder Abteilung im Unternehmen. Sheet2 und Sheet3 sind leer. In diesem Beispiel extrahieren Sie nur Geschäftsinformationen zu den Mitarbeitern, die in der Abteilung B01 arbeiten.

*Schritt 1: Job erstellen:*

Erstellen Sie einen Beispieljob, der eine Stage **Unstructured Data** und eine Stage **Sequential File** (Sequenzielle Datei) enthält.

#### Vorgehensweise

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.
2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und klicken Sie dann auf **Neu > Paralleler Job**.
3. Ziehen Sie die Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.
4. Ziehen Sie eine Stage **Sequential File** auf den Entwicklungsbereich und ordnen Sie diese Stage rechts von der Stage **Sequential File** an.
5. Erstellen Sie einen Link von der Stage **Unstructured Data** zur Stage **Sequential File**.
6. Benennen Sie die Stage und den Link um.
7. Wählen Sie **Datei > Speichern** aus und benennen Sie den Job.



### *Schritt 2: Stage 'Unstructured Data' konfigurieren:*

Konfigurieren Sie die Stage **Unstructured Data**, um Daten aus einem Bereich in einer Microsoft Excel-Datei zu extrahieren.

#### **Vorgehensweise**

1. Doppelklicken Sie auf die Stage **Unstructured Data**.
2. Klicken Sie auf **Konfigurieren**.

**Anmerkung:** Konfigurieren Sie in diesem Schritt keine Stage-Eigenschaften, weil Sie alle erforderlichen Konfigurationen im Konfigurationsfenster konfigurieren können.

3. Geben Sie den vollständigen Dateipfad der Microsoft Excel-Eingabedatei `Employee1.xls` im Konfigurationsfenster an.
4. Wählen Sie **Gesamten Datenbereich angeben** in der Liste **Bereichsoption** aus, um die Daten in einem bestimmten Bereich zu extrahieren.
5. Geben Sie im Feld **Bereichsausdruck** `Sheet1!A16:K28` an.
6. Wählen Sie **Erste Zeile von Datenbereichen** im Feld **Spaltenheader** aus. Wenn **Erste Zeile von Datenbereichen** ausgewählt ist, wird die erste Zeile als Header betrachtet und fängt die Stage **Unstructured Data** mit dem Extrahieren von Daten aus der zweiten Zeile an.
7. Klicken Sie auf **Laden** und stellen Sie dann sicher, dass die Kontrollkästchen neben den Microsoft Excel-Spalten, die mit den Jobs extrahiert werden sollen, ausgewählt sind. Die Stage **Unstructured Data** greift auf die angegebene Datei zu und listet die Microsoft Excel-Spalten im angegebenen Datenbereich im Teilfenster **Importieren** auf. Alle Microsoft Excel-Spalten sind standardmäßig ausgewählt.
8. Wählen Sie die Kontrollkästchen neben den Microsoft Excel-Spalten E (PHONE NO) und I (BIRTH DATE) ab.
9. Klicken Sie auf **Importieren**. Nach dem Klicken auf **Importieren** wird das Teilfenster **Zuordnen** in der unteren rechten Ecke des Konfigurationsfensters aktualisiert.
10. Klicken Sie auf **OK**.
11. Überprüfen Sie, ob die im Konfigurationsfenster angegebenen Werte auf der Registerkarte **Eigenschaft** des Stage-Editors gespeichert wurden.
12. Ändern Sie den Typ der Spalte **EMP\_NO** auf der Seite **Ausgabe > Spalte** in Integer und klicken Sie dann auf **OK**.

### *Schritt 3: Stage 'Sequential File' konfigurieren:*

Konfigurieren Sie die Stage **Sequential File**.

#### **Informationen zu diesem Vorgang**

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

#### **Vorgehensweise**

1. Doppelklicken Sie auf die Stage **Sequential File**.
2. Geben Sie den Pfad, in dem die Ausgabedatei erstellt werden soll, gefolgt vom Dateinamen `OutputOfExample1.txt` auf der Seite **Eigenschaften** an.
3. Klicken Sie auf **OK**.

Schritt 4: Ausgabe des Jobs anzeigen:

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

### Informationen zu diesem Vorgang

Wenn z. B. eine Microsoft Excel-Eingabedatei die Mitarbeiterinformationen für verschiedene Abteilungen in einem Unternehmen enthält, kann der Job Daten aus der angegebenen Abteilung extrahieren.

### Vorgehensweise

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.

In der folgenden Tabelle werden die Informationen in einer Microsoft Excel-Eingabedatei angezeigt, die die Mitarbeiterinformationen für verschiedene Abteilungen enthält.

Tabelle 8. Microsoft Excel-Beispieldatei mit Mitarbeiterdetails

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
Mitarbeiter in DEPT_A00								
10	CHRISTINE	I	HAAS	3978	1/1/1995	PRES	F	8/24/1963
20	MICHAEL	L	THOMSON	3476	10/10/2003	MANAGER	M	2/2/1976
30	SALLY	A	KWAN	4738	4/5/2005	MANAGER	F	5/11/1971
50	JOHN	B	GEYER	6789	8/17/1979	MANAGER	M	9/15/1955
Mitarbeiter in DEPT_B01								
60	IRVING	F	STERN	6423	9/14/2003	MANAGER	M	7/7/1975
70	EVA	D	PULASKI	7831	9/30/2003	MANAGER	F	5/26/2003
90	EILEEN	W	HENDERSON	5498	8/15/2000	MANAGER	F	5/15/1971
100	THEODORE	Q	SPENSER	742	6/19/2000	MANAGER	M	12/18/1980
110	VINCENZO	G	LUCCHESI	3490	5/16/1988	SALESREP	M	11/5/1958
120	SEAN		O'CONNELL	2167	12/5/1993	CLERK	M	10/18/1972
130	DELORES	M	QUINTANA	4578	7/28/2001	ANALYST	F	9/15/1955
140	HEATHER	A	NICHOLLS	1793	12/15/2006	ANALYST	F	1/19/1976
150	BRUCE		ADAMSON	4510	2/12/2002	DESIGNER	M	5/17/1972
160	ELIZABETH	R	PIANKA	3782	10/11/2006	DESIGNER	F	4/12/1980
1770	MASATOSHI	J	YOSHIMURA	2890	9/15/1999	DESIGNER	M	1/5/1981
180	MARILYN	S	SCOUTTEN	1682	7/7/2003	DESIGNER	F	2/21/1978

3. Öffnen Sie die Datei OutputOfExample1.txt nach der Jobausführung, um das Ergebnis anzuzeigen.

Die Datei OutputOfExample1.txt enthält Daten für DEPT\_B01 mit Ausnahme der Spalten PHONE NO und BIRTH DATE:

```
"60", "IRVING", "F", "STERN", "2003-09-14", "MANAGER ", "M", "72250", "500"
"70", "EVA", "D", "PULASKI", "2005-09-30", "MANAGER ", "F", "96170", "700"
"90", "EILEEN", "W", "HENDERSON", "2000-08-15", "MANAGER ", "F", "89750", "600"
"100", "THEODORE", "Q", "SPENSER", "2000-06-19", "MANAGER ", "M", "86150", "500"
"110", "VINCENZO", "G", "LUCCHESI", "1988-05-16", "SALESREP", "M", "66500", "900"
"120", "SEAN", " ", "O'CONNELL", "1993-12-05", "CLERK ", "M", "49250", "600"
"130", "DELORES", "M", "QUINTANA", "2001-07-28", "ANALYST ", "F", "73800", "500"
"140", "HEATHER", "A", "NICHOLLS", "2006-12-15", "ANALYST ", "F", "68420", "600"
"150", "BRUCE", " ", "ADAMSON", "2002-02-12", "DESIGNER", "M", "55280", "500"
"160", "ELIZABETH", "R", "PIANKA", "2006-10-11", "DESIGNER", "F", "62250", "400"
"170", "MASATOSHI", "J", "YOSHIMURA", "1999-09-15", "DESIGNER", "M", "44680", "500"
"180", "MARILYN", "S", "SCOUTTEN", "2003-07-07", "DESIGNER", "F", "51340", "500"
```

## Beispiel 2: Daten aus mehreren Microsoft Excel-Arbeitsblättern extrahieren:

Erstellen Sie einen Job, der mithilfe der Stage **Unstructured Data** Daten aus mehreren Microsoft Excel-Arbeitsblättern extrahiert.

### Informationen zu diesem Vorgang

In diesem Beispiel wird die Microsoft Excel-Beispieldatei Employee2.xls verwendet. Diese Beispieldatei enthält die folgenden Arbeitsblätter: DEPT A00, DEPT B01, DEPT C01 und DEPT D01. Jedes Arbeitsblatt enthält Informationen zu den Mitarbeitern in der Abteilung.

Die Datenstruktur der einzelnen Arbeitsblätter ist ähnlich. Jedes Arbeitsblatt enthält die Spalten EMP NO, FIRST NAME, MID INIT, LAST NAME, PHONE NO, HIRE DATE, JOB und ADDRESS und die dritte Zeile ist der Header. Jedes Arbeitsblatt weist jedoch eine unterschiedliche Anzahl Zeilen auf.

#### Schritt 1: Job erstellen:

Erstellen Sie einen Beispieljob, der eine Stage **Unstructured Data** und eine Stage **Sequential File** (Sequenzielle Datei) enthält.

#### Vorgehensweise

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.
2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und wählen Sie **Neu > Paralleler Job** aus.
3. Ziehen Sie eine Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.
4. Ziehen Sie eine Stage **Sequential File** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich. Ordnen Sie die Stage rechts neben der Stage **Unstructured Data** an.
5. Erstellen Sie einen Link von der Stage **Unstructured Data** zur Stage **Sequential File**.
6. Benennen Sie die Stages und Links um.
7. Wählen Sie **Datei > Speichern** aus und benennen Sie den Job.

#### Schritt 2: Stage 'Unstructured Data' konfigurieren:

Konfigurieren Sie die Stage **Unstructured Data**, um Daten aus mehreren Microsoft Excel-Arbeitsblättern zu extrahieren.

#### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Unstructured Data**, um die Stage-Eigenschaften zu öffnen.
2. Klicken Sie auf **Konfigurieren**.

**Anmerkung:** Konfigurieren Sie in diesem Schritt keine Stage-Eigenschaften, weil Sie alle erforderlichen Konfigurationen im Konfigurationsfenster konfigurieren können.

3. Geben Sie den vollständigen Dateipfad der Microsoft Excel-Eingabedatei Employee2.xls im Konfigurationsfenster an.

4. Wählen Sie in der Liste **Bereichsoption** den Eintrag **Anfangszeile angeben** aus.
5. Geben Sie **A3:H3** im Feld **Bereichsausdruck** an. Wenn die Stage mit der Option **Erste Zeile von Datenbereichen** ausgeführt und kein bestimmter Arbeitsblattname für **Bereichsausdruck** angegeben wird, sucht der Job die letzte Zeile dynamisch und extrahiert Zeilen bis zur letzten Zeile während der Ausführung.
6. Wählen Sie **Erste Zeile von Datenbereichen** im Feld **Spaltenheader** aus.
7. Klicken Sie auf **Laden**. Die im angegebenen Datenbereich vorhandenen Excel-Spalten werden im Teilfenster **Importieren** aufgelistet.
8. Wählen Sie auf der Registerkarte **Eigenschaft** das Kontrollkästchen neben der Eigenschaft aus, um den Eigenschaftswert zu extrahieren. Wählen Sie in diesem Beispiel den Arbeitsblattnamen als Eigenschaft aus.
9. Klicken Sie auf **Importieren**. Die Spaltenmappings werden von der Stage generiert.
10. Gehen Sie wie folgt vor, um die Spalte mit dem Arbeitsblattnamen zur ersten Spalte in der Liste zu machen:
  - a. Wählen Sie die Spalte mit dem Arbeitsblattnamen aus.
  - b. Klicken Sie auf **Nach oben**, bis die Spalte mit dem Arbeitsblattnamen die erste Spalte in der Liste ist.
11. Fügen Sie in der Mapping-Tabelle eine Zeile für die Spalte ADDRESS in die Eingabedatei mit dem Hyperlink ein.
  - a. Klicken Sie auf **Einfügen**.
  - b. Wählen Sie die Spalte **ADDRESS** für die Option **Excel-Element** aus.
  - c. Wählen Sie **Hyperlinkadresse** in der Optionszelle **Importieren** in der neuen Zeile aus.
  - d. Geben Sie den DataStage-Spaltennamen EMAIL\_ADDRESS für die neue Zeile an.
12. Klicken Sie auf **OK**.
13. Überprüfen Sie, ob die im Konfigurationsfenster eingegebenen Werte auf der Registerkarte **Eigenschaft** des Stage-Editors gespeichert wurden.
14. Klicken Sie auf die Registerkarte **Ausgabe > Spalte**, um den Datentyp oder andere Attribute zu ändern. Ändern Sie den Typ der Spalte **EMP\_NO** in Integer.
15. Klicken Sie auf **OK**.

*Schritt 3: Stage 'Sequential File' konfigurieren:*

Konfigurieren Sie die Stage **Sequential File**.

### Informationen zu diesem Vorgang

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Sequential File**.
2. Geben Sie den Pfad, in dem die Ausgabedatei erstellt werden soll, gefolgt vom Dateinamen OutputOfExample2.txt auf der Seite **Eigenschaften** an.
3. Klicken Sie auf **OK**.

Schritt 4: Ausgabe des Jobs anzeigen:

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

### Vorgehensweise

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.

Es folgt ein Beispiel für eine Microsoft Excel-Datei mit mehreren Arbeitsblättern als Eingabe. Der Job extrahiert die Mitarbeiterdaten aus allen Arbeitsblättern.

Tabelle 9. Informationen von Mitarbeitern in DEPT\_A00

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
10	CHRISTINE	I	HAAS	3978	1/1/1995	PRES	F	8/24/1963
20	MICHAEL	L	THOMSON	3476	10/10/2003	MANAGER	M	2/2/1976
30	SALLY	A	KWAN	4738	4/5/2005	MANAGER	F	5/11/1971
50	JOHN	B	GEYER	6789	8/17/1979	MANAGER	M	9/15/1955

Tabelle 10. Details von Mitarbeitern in DEPT\_B01

EMP NO	FIRST NAME	MIDI NIT	LAST NAME	PHONE NO	HIRE DATE	JOB	SEX	BIRTH DATE
60	IRVING	F	STERN	6423	9/14/2003	MANAGER	M	7/7/1975
70	EVA	D	PULASKI	7831	9/30/2003	MANAGER	F	5/26/2003
90	EILEEN	W	HENDERSON	5498	8/15/2000	MANAGER	F	5/15/1971
100	THEODORE	Q	SPENSER	742	6/19/2000	MANAGER	M	12/18/1980
110	VINCENZO	G	LUCCHESSI	3490	5/16/1988	SALESREP	M	11/5/1958
120	SEAN		O'CONNELL	2167	12/5/1993	CLERK	M	10/18/1972
130	DELORES	M	QUINTANA	4578	7/28/2001	ANALYST	F	9/15/1955
140	HEATHER	A	NICHOLLS	1793	12/15/2006	ANALYST	F	1/19/1976
150	BRUCE		ADAMSON	4510	2/12/2002	DESIGNER	M	5/17/1972
160	ELIZABETH	R	PIANKA	3782	10/11/2006	DESIGNER	F	4/12/1980
1770	MASATOSHI	J	YOSHIMURA	2890	9/15/1999	DESIGNER	M	1/5/1981
180	MARILYN	S	SCOUTTEN	1682	7/7/2003	DESIGNER	F	2/21/1978

3. Öffnen Sie nach der Jobausführung die Datei OutputOfExample2.txt, die das folgende Ergebnis enthält.

```
"DEPT A00", "10", "CHRISTINE", "I", "HAAS", "3978", "1995-01-01", "PRES  ",
"CHRISTINE HAAS", "mailto:CHRISTINE%20HAAS@abc.com"
"DEPT A00", "20", "MICHAEL", "L", "THOMPSON", "3476", "2003-10-10", "MANAGER ",
"MICHAEL THOMPSON", "mailto:MICHAEL%20THOMPSON@abc.com"
"DEPT A00", "30", "SALLY", "A", "KWAN", "4738", "2005-04-05", "MANAGER ",
"SALLY KWAN", "mailto:SALLY%20KWAN@abc.com"
"DEPT A00", "50", "JOHN", "B", "GEYER", "6789", "1979-08-17", "MANAGER ",
"JOHN GEYER", "mailto:JOHN%20GEYER@abc.com"
"DEPT B01", "60", "IRVING", "F", "STERN", "6423", "2003-09-14", "MANAGER ",
"IRVING STERN", "mailto:IRVING%20STERN@abc.com"
"DEPT B01", "70", "EVA", "D", "PULASKI", "7831", "2005-09-30", "MANAGER ",
"EVA PULASKI", "mailto:EVA%20PULASKI@abc.com"
"DEPT B01", "90", "EILEEN", "W", "HENDERSON", "5498", "2000-08-15", "MANAGER ",
"EILEEN HENDERSON", "mailto:EILEEN%20HENDERSON@abc.com"
"DEPT B01", "100", "THEODORE", "Q", "SPENSER", "972", "2000-06-19", "MANAGER ",
"THEODORE SPENSER", "mailto:THEODORE%20SPENSER@abc.com"
"DEPT B01", "110", "VINCENZO", "G", "LUCCHESSI", "3490", "1988-05-16", "SALESREP",
"VINCENZO LUCCHESSI", "mailto:VINCENZO%20LUCCHESSI@abc.com"
"DEPT B01", "120", "SEAN", " ", "O'CONNELL", "2167", "1993-12-05", "CLERK  ",
"SEAN O'CONNELL", "mailto:SEAN%20O'CONNELL@abc.com"
"DEPT B01", "130", "DELORES", "M", "QUINTANA", "4578", "2001-07-28", "ANALYST ",
"DELORES QUINTANA", "mailto:DELORES%20QUINTANA@abc.com"
"DEPT B01", "140", "HEATHER", "A", "NICHOLLS", "1793", "2006-12-15", "ANALYST ",
"HEATHER NICHOLLS", "mailto:HEATHER%20NICHOLLS@abc.com"
"DEPT B01", "150", "BRUCE", " ", "ADAMSON", "4510", "2002-02-12", "DESIGNER",
"BRUCE ADAMSON", "mailto:BRUCE%20ADAMSON@abc.com"
```

```

"DEPT B01","160","ELIZABETH","R","PIANKA","3782","2006-10-11","DESIGNER",
"ELIZABETH PIANKA","mailto:ELIZABETH%20PIANKA@abc.com"
"DEPT B01","170","MASATOSHI","J","YOSHIMURA","2890","1999-09-15","DESIGNER",
"MASATOSHI YOSHIMURA","mailto:MASATOSHI%20YOSHIMURA@abc.com"
"DEPT B01","180","MARILYN","S","SCOUTTEN","1682","2003-07-07","DESIGNER",
"MARILYN SCOUTTEN","mailto:MARILYN%20SCOUTTEN@abc.com"
"DEPT C01","190","JAMES","H","WALKER","2986","2004-07-26","DESIGNER",
"JAMES WALKER","mailto:JAMES%20WALKER@abc.com"
"DEPT C01","200","DAVID"," " ,"BROWN","4501","2002-03-03","DESIGNER",
"DAVID BROWN","mailto:DAVID%20BROWN@abc.com"
"DEPT C01","210","WILLIAM","T","JONES","942","1998-04-11","DESIGNER",
"WILLIAM JONES","mailto:WILLIAM%20JONES@abc.com"
"DEPT C01","220","JENNIFER","K","LUTZ","672","1998-08-29","DESIGNER",
"JENNIFER LUTZ","mailto:JENNIFER%20LUTZ@abc.com"
"DEPT C01","230","JAMES","J","JEFFERSON","2094","1996-11-21","CLERK  ",
"JAMES JEFFERSON","mailto:JAMES%20JEFFERSON@abc.com"
"DEPT C01","240","SALVATORE","M","MARINO","3780","2004-12-05","CLERK  ",
"SALVATORE MARINO","mailto:SALVATORE%20MARINO@abc.com"
"DEPT C01","250","DANIEL","S","SMITH","961","1999-10-30","CLERK  ",
"DANIEL SMITH","mailto:DANIEL%20SMITH@abc.com"
"DEPT C01","260","SYBIL","P","JOHNSON","8953","2005-09-11","CLERK  ",
"SYBIL JOHNSON","mailto:SYBIL%20JOHNSON@abc.com"
"DEPT D01","270","MARIA","L","PEREZ","9001","2006-09-30","CLERK  ",
"MARIA PEREZ","mailto:MARIA%20PEREZ@abc.com"
"DEPT D01","280","ETHEL","R","SCHNEIDER","8997","1997-03-24","OPERATOR",
"ETHEL SCHNEIDER","mailto:ETHEL%20SCHNEIDER@abc.com"
"DEPT D01","290","JOHN","R","PARKER","4502","2006-05-30","OPERATOR",
"JOHN PARKER","mailto:JOHN%20PARKER@abc.com"
"DEPT D01","300","PHILIP","X","SMITH","2095","2002-06-19","OPERATOR",
"PHILIP SMITH","mailto:PHILIP%20SMITH@abc.com"
"DEPT D01","310","MAUDE","F","SETRIGHT","3332","1994-09-12","OPERATOR",
"MAUDE SETRIGHT","mailto:MAUDE%20SETRIGHT@abc.com"
"DEPT D01","320","RAMLAL","V","MEHTA","9990","1995-07-07","FIELDREP",
"RAMLAL MEHTA","mailto:RAMLAL%20MEHTA@abc.com"
"DEPT D01","330","WING"," " ,"LEE","2103","2006-02-23","FIELDREP",
"WING LEE","mailto:WING%20LEE@abc.com"
"DEPT D01","340","JASON","R","GOUNOT","5698","1977-05-05","FIELDREP",
"JASON GOUNOT","mailto:JASON%20GOUNOT@abc.com"

```

### Beispiel 3: Daten aus mehreren Bereichen mit unterschiedlichen Datenstrukturen in eine Microsoft Excel-Datei extrahieren:

Sie können einen Job erstellen, der über die Stage **Unstructured Data** Daten aus mehreren Bereichen mit unterschiedlichen Datenstrukturen in eine Microsoft Excel-Datei extrahiert.

#### Informationen zu diesem Vorgang

In diesem Beispiel wird die Microsoft Excel-Beispieldatei Employee3.xls verwendet. Diese Beispieldatei enthält die beiden Arbeitsblätter Departments und Employees, die unterschiedliche Datenstrukturen aufweisen.

In diesem Beispiel hat die Stage **Unstructured Data** zwei Ausgabelinks. Sie extrahieren Daten aus dem Arbeitsblatt Departments in den ersten Link und Daten aus dem Arbeitsblatt Employees in den zweiten Link.

#### Schritt 1: Job erstellen:

Erstellen Sie einen Beispieljob, der eine Stage **Unstructured Data** und zwei Stages **Sequential File** enthält.

### Vorgehensweise

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.
2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und wählen Sie **Neu > Paralleler Job** aus.
3. Ziehen Sie eine Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.
4. Ziehen Sie zwei Stages **Sequential File** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich. Ordnen Sie die Stages rechts neben der Stage **Unstructured Data** an.
5. Benennen Sie die neuen Stages **Sequential File** in **Output\_1** und **Output\_2** um.
6. Erstellen Sie einen Link von der Stage **Unstructured Data** zu den Stages **Sequential File**.
7. Benennen Sie die Links in **Departments** und **Employees**.
8. Speichern Sie den Job.

*Schritt 2: Stage 'Unstructured Data' konfigurieren:*

Konfigurieren Sie die Stage **Unstructured Data**, um Daten aus mehreren Microsoft Excel-Arbeitsblättern zu extrahieren.

### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Unstructured Data**, um die Stage-Eigenschaften zu öffnen.
2. Klicken Sie auf **Konfigurieren**.
3. Geben Sie den vollständigen Dateipfad der Microsoft Excel-Eingabedatei **Employee3.xls** im Konfigurationsfenster an.
4. Geben Sie die Daten an, die aus dem Arbeitsblatt **Departments** extrahiert werden sollen, und führen Sie die folgenden Unterschritte aus, um die Spaltenmappings zu generieren.
  - a. Wählen Sie im Listenfeld **Link** den Eintrag **Departments** aus.
  - b. Wählen Sie in der Liste **Bereichsoption** den Eintrag **Gesamten Datenbereich angeben** aus.
  - c. Geben Sie **Departments!A2:C6** im Feld **Bereichsausdruck** an.
  - d. Wählen Sie **Erste Zeile von Datenbereichen** im Feld **Spaltenheader** aus.
  - e. Klicken Sie auf **Laden**. Die Excel-Spalten im angegebenen Datenbereich werden im Teilfenster **Importieren** aufgelistet.
5. Geben Sie die Daten an, die aus dem Arbeitsblatt **Employees** extrahiert werden sollen, und führen Sie die die folgenden Unterschritte aus, um die Spaltenmappings zu generieren.
  - a. Wählen Sie im Listenfeld **Link** den Eintrag **Employees** aus.
  - b. Wählen Sie in der Liste **Bereichsoption** den Eintrag **Gesamten Datenbereich angeben** aus.
  - c. Geben Sie **Employees!A2:L34** im Feld **Bereichsausdruck** an.
  - d. Wählen Sie **Erste Zeile von Datenbereichen** im Feld **Spaltenheader** aus.
  - e. Klicken Sie auf **Laden**. Die Excel-Spalten im angegebenen Datenbereich werden im Teilfenster **Importieren** aufgelistet.
  - f. Klicken Sie auf **Importieren** und dann auf **OK**. Die Stage ordnet Spalten zu.
6. Überprüfen Sie, ob die im Konfigurationsfenster eingegebenen Werte auf der Registerkarte **Eigenschaft** des Stage-Editors gespeichert wurden.

7. Wählen Sie auf der Seite **Ausgabe** den Link **Employees** als Ausgabenamen aus.
8. Ändern Sie auf der Seite **Spalten** den Datentyp der Spalte **EMP\_NO** in Integer und klicken Sie dann auf **OK**.

*Schritt 3: Stages 'Sequential File' konfigurieren:*

Konfigurieren Sie die Stages **Sequential File**. In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Stage verwenden.

### Informationen zu diesem Vorgang

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Sequential File** namens Output\_1.
2. Geben Sie auf der Seite **Eigenschaften** den Pfad an, in dem die Ausgabedatei erstellt werden soll, gefolgt von dem Dateinamen OutputOfExample3\_1.txt.
3. Klicken Sie auf **OK**.
4. Wiederholen Sie die Schritte 1 bis 3 für die zweite Stage **Sequential File** namens Output\_2 und geben Sie ihr den Namen OutputOfExample3\_2.txt.

*Schritt 4: Ausgabe des Jobs anzeigen:*

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

### Vorgehensweise

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.  
Ein Beispiel für eine Microsoft Excel-Eingabedatei Employee3.xls enthält Abteilungsinformationen im Arbeitsblatt **Departments** und Mitarbeiterinformationen im Arbeitsblatt **Employees**. Der Job extrahiert Abteilungsinformationen in die Datei OutputOfExample3\_1.txt und Mitarbeiterinformationen in die Datei OutputOfExample3\_2.txt. Jedes Arbeitsblatt enthält folgende Daten:
3. Öffnen Sie die Dateien OutputOfExample3\_1.txt und OutputOfExample3\_2.txt nach der Jobausführung. Die Datei OutputOfExample3\_1.txt sollte mit dem Arbeitsblatt **Departments** und die Datei OutputOfExample3\_2.txt mit dem Arbeitsblatt **Employees** aus der Datei Employee.xls übereinstimmen.

## Daten in neue Microsoft Excel-Datei schreiben

Sie können die Stage **Unstructured Data** in Jobs verwenden, die Daten in eine neue Microsoft Excel-Datei schreiben, indem Sie den vollständigen Pfadnamen angeben. Sie können auch mehrere Microsoft Excel-Dateien erstellen, indem Sie die Position angeben, an der die Dateien erstellt werden, sowie ein Präfix für die Dateinamen.

Jeder Eingabelink der Stage **Unstructured Data** wird einem separaten Microsoft Excel-Arbeitsblatt zugeordnet. Die Microsoft Excel-Arbeitsblätter werden entsprechend den jeweiligen Eingabelinks als **Sheet1**, **Sheet2**, **Sheet3** usw. benannt. Jede InfoSphere DataStage-Spalte des Eingabelinks wird einer Microsoft Excel-Spalte zugeordnet.



## Jobs entwerfen, die die Stage 'Unstructured Data' enthalten

Mithilfe einer Stage **Unstructured Data** können Sie Daten in unstrukturierte Datenquellen schreiben.

### Vorgehensweise

1. Definieren Sie einen Job, der eine Stage 'Unstructured Data' enthält.
2. Wenn die Stage **Unstructured Data** als eine Zielstage eingerichtet werden soll, um Daten in unstrukturierte Datenquellen zu schreiben, führen Sie die folgenden Schritte aus:
  - a. Konfigurieren Sie die Stage 'Unstructured Data' als Ziel.
  - b. Geben Sie die Spaltendefinition im Link an.
3. Kompilieren Sie den Job und führen Sie ihn aus.

### Stage 'Unstructured Data' als Ziel konfigurieren:

Sie können die Stage **Unstructured Data** konfigurieren, um eine Microsoft Excel-Datei zu generieren.

### Informationen zu diesem Vorgang

Die Stage **Unstructured Data** unterstützt nur das OOXML-Format (.xlsx) der Microsoft Excel-Dateien als Zieldatei.

Die Stage **Unstructured Data** unterstützt die Spaltenweitergabe während der Ausführung. Wenn die Spaltenweitergabe während der Ausführung im Ausgabelink einer vorgeschalteten Stage aktiviert ist, werden weitergegebene zusätzliche Spalten nach im InfoSphere Designer-Client definierten Spalten angehängt.

Die Stage **Unstructured Data** unterstützt nicht die Generierung von .xls-Dateien oder von mit einem Kennwort verschlüsselten Dateien.

### Vorgehensweise

1. Doppelklicken Sie im Entwicklungsbereich für parallele Jobs auf die Stage **Unstructured Data**.
2. Wählen Sie **Excel** in der Liste **Dokumenttyp** der Registerkarte **Stage** aus.
3. Wählen Sie **Datei erstellen** in der Liste **Schreibzugriff** aus.
4. Klicken Sie auf **Konfigurieren**, um die Eigenschaften für das Schreiben von Daten in eine Microsoft Excel-Datei zu konfigurieren.
5. Machen Sie im Teilfenster **Ausgabedatei** folgende Angaben:
  - a. Geben Sie im Feld **Dateiname** den vollständigen Pfadnamen der Microsoft Excel-Datei an, in die die Daten geschrieben werden sollen. Sie können beispielsweise C:\tmp\employee.xlsx angeben.
  - b. Optional: Machen Sie Angaben für **Dateiaktualisierungsmodus**. Wenn Sie **Erstellen (Fehler, wenn vorhanden)** angeben, schlägt Ihre Jobausführung fehl, wenn die Microsoft Excel-Zieldatei bereits vorhanden ist. Wenn Sie die Standardeinstellung **Überschreiben** auswählen, überschreibt die Stage **Unstructured Data** die vorhandene Datei.
  - c. Optional: Machen Sie Angaben für **Schreibmethode**. Wenn Sie **Mehrere Dateien generieren** auswählen, erstellt die Stage **Unstructured Data** mehrere Microsoft Excel-Dateien auf der Basis zusätzlicher Eigenschafteneinstellungen. Wenn Sie die Standardeinstellung **Bestimmte Datei** auswählen, erstellt die Stage **Unstructured Data** eine Microsoft Excel-Datei mit dem in der Eigenschaft **Dateiname** angegebenen Namen.

6. Machen Sie im Teilfenster **Eigenschaften** folgende Angaben:
  - a. Optional: Wählen Sie das Kontrollkästchen **Für alle Links definiert** ab, um die Eigenschaften für jeden Eingabelink anzugeben, der in der Liste **Link** ausgewählt ist. Wenn Sie das Kontrollkästchen **Für alle Links definiert** (Standardeinstellung) auswählen, werden die Eigenschafteneinstellungen auf alle Eingabelinks angewendet.
  - b. Optional: Machen Sie Angaben für **Spaltenheader**. Wenn Sie **Spaltennamen** auswählen, schreibt die Stage **Unstructured Data** InfoSphere DataStage Spaltennamen in die erste Zeile des Microsoft Excel-Arbeitsblatts. Wenn Sie die Standardeinstellung **Ohne** auswählen, schreibt die Stage **Unstructured Data** die Daten in die erste Zeile des Microsoft Excel-Arbeitsblatts.
  - c. Optional: Machen Sie Angaben für **Spaltenbreite anpassen**. Wenn Sie **Ja** auswählen, passt die Stage **Unstructured Data** die Breite der einzelnen Spalten im generierten Microsoft Excel-Arbeitsblatt an den Spalteninhalt an. Wenn Sie die Standardeinstellung **Nein** auswählen, passt die Stage **Unstructured Data** die Spaltenbreite nicht an.
7. Optional: Ändern Sie die Arbeitsblattreihenfolge im Teilfenster **Arbeitsblattreihenfolge** und modifizieren Sie die Arbeitsblattnamen.
8. Klicken Sie auf **OK**, um die angegebenen Einstellungen zu speichern.

#### Zugehörige Tasks:

„Daten in mehrere Microsoft Excel-Dateien schreiben“ auf Seite 21

Wenn Sie über große Datenvolumen verfügen, können Sie über die Stage **Unstructured Data** Daten in mehrere Microsoft Excel-Dateien schreiben.

#### Spaltendefinition im Link angeben:

Sie können Elemente der Spaltendefinition wie **SQL-Typ**, **Länge**, **Nachkommastellen** und **Kann Nullwerte enthalten** im Link angeben.

#### Vorgehensweise

1. Doppelklicken Sie im Entwicklungsbereich für parallele Jobs auf die Stage **Unstructured Data**.
2. Wählen Sie die Registerkarte **Eingabe** und dann den Eingabelink in **Eingabename (vorgeschaltete Stage)** aus.
3. Bearbeiten Sie **SQL-Typ**, **Länge** und **Nachkommastellen** für jede Spalte.
4. Klicken Sie auf **OK**, um die Änderungen zu speichern.

#### Jobparameter beim Modifizieren einer vorhandenen Microsoft Excel-Datei verwenden:

Die Stage **Unstructured Data** kann keine neuen Jobparameter im Konfigurationsfenster erstellen. Sie können jedoch die Jobparameter im Konfigurationsfenster verwenden. Sie müssen Jobparameter im Fenster **Jobeigenschaften** vor oder nach der Arbeit im Konfigurationsfenster erstellen, indem Sie **Bearbeiten > Jobeigenschaften** im Client von IBM InfoSphere DataStage und QualityStage Designer auswählen. Weitere Informationen zum Erstellen von Jobparametern finden Sie in Lektion 2.4, Parameter hinzufügen, im IBM InfoSphere DataStage-Lernprogramm für parallele Jobs.

Ein Jobparameter wird im Konfigurationsfenster mit dem Zeichen # angegeben. Beispielsweise wird der Jobparameter *Dateiname* im Konfigurationsfenster als **#Da-**

**teiname#** angegeben. Für ein Feld mit dem Datentyp 'string' wie die Eigenschaft **Dateiname** können Sie den Namen des Jobparameters von # umschlossen direkt eingeben.

Wenn Sie für die Eigenschaft **Listentyp** wie z. B. **Aktion, wenn die Datei bereits vorhanden ist** Jobparameter verwenden wollen, müssen Sie einen Listentypparameter erstellen, der eine Liste von Zeichenfolgevariablen enthält. Die Zeichenfolgevariablen müssen mit dem Beschriftungstext der entsprechenden Eigenschaft im Konfigurationsfenster übereinstimmen. Wenn Sie z. B. Jobparameter für die Eigenschaft **Aktion, wenn die Datei bereits vorhanden ist** verwenden wollen, müssen Sie einen Listentypparameter erstellen, der die Zeichenfolgevariablen **Fehler** und **Überschreiben** enthält. Wählen Sie nach der Erstellung eines Jobparameters **<Parametrisieren...>** im Konfigurationsfenster aus und geben Sie den Jobparameternamen vom Zeichen # umschlossen im Dialogfeld **Eingabeparameter** an. Klicken Sie auf **Laden**, um Variablen in der Anzeige **Jobparameter auflösen** zu bearbeiten oder auszuwählen.

## Daten in mehrere Microsoft Excel-Dateien schreiben

Wenn Sie über große Datenvolumen verfügen, können Sie über die Stage **Unstructured Data** Daten in mehrere Microsoft Excel-Dateien schreiben.

### Informationen zu diesem Vorgang

Über die Stage **Unstructured Data** können Sie Jobs entwerfen, die Daten in mehrere Microsoft Excel-Dateien schreiben. Die vom OXML-Format von Microsoft Excel (.xlsx) unterstützte maximale Anzahl Datensätze ist 1.048.576. Wenn die Eingabelinks mehr als 1.048.576 Datensätze enthalten, müssen Sie die Datensätze auf mehrere Microsoft Excel-Dateien verteilen. Sie können die Links, auch wenn sie maximal 1.048.576 Datensätze enthalten, in mehrere Microsoft Excel-Dateien schreiben, weil das Öffnen einer großen Microsoft Excel-Datei eine große Speichermenge erfordert.

### Vorgehensweise

1. Definieren Sie einen Job, der eine Stage 'Unstructured Data' enthält.
2. Konfigurieren Sie die Stage 'Unstructured Data'.
3. Kompilieren Sie den Job und führen Sie ihn aus.

#### Zugehörige Tasks:

„Stage 'Unstructured Data' als Ziel konfigurieren“ auf Seite 19

Sie können die Stage **Unstructured Data** konfigurieren, um eine Microsoft Excel-Datei zu generieren.

#### Stage 'Unstructured Data' konfigurieren:

Sie können die Stage **Unstructured Data** konfigurieren, um mehrere Microsoft Excel-Dateien zu generieren.

### Vorgehensweise

1. Doppelklicken Sie im Entwicklungsbereich für parallele Jobs auf die Stage **Unstructured Data**.
2. Wählen Sie **Excel** in der Liste **Dokumenttyp** der Registerkarte **Stage** aus.
3. Wählen Sie **Datei erstellen** in der Liste **Schreibzugriff** aus.
4. Klicken Sie auf **Konfigurieren**, um die Eigenschaften für das Schreiben von Daten in eine Microsoft Excel-Datei zu konfigurieren.
5. Machen Sie im Teilfenster **Ausgabedatei** folgende Angaben:

- a. Geben Sie im Feld **Dateiname** die Position an, an der mehrere Microsoft Excel-Dateien erstellt werden sollen. Sie können auch das Präfix der Dateinamen nach der Dateiposition angeben. Wenn Sie beispielsweise Microsoft Excel-Dateien in C:\tmp erstellen wollen und Sample als Präfix der Dateinamen verwenden wollen, geben Sie C:\tmp\Sample im Feld **Dateiname** an. Die Stage **Unstructured Data** hängt eine dreistellige fortlaufende Zahl und eine Dateierweiterung (.xlsx) an das Präfix an. Wenn die Anzahl Dateien 999 übersteigt, enthält der Dateiname die erforderliche Anzahl Ziffern.
  - b. Optional: Machen Sie Angaben für **Dateiaktualisierungsmodus**. Wenn Sie **Erstellen (Fehler, wenn vorhanden)** angeben, schlägt Ihre Jobausführung fehl, wenn die Microsoft Excel-Zieldatei bereits vorhanden ist. Wenn Sie die Standardeinstellung **Überschreiben** auswählen, überschreibt die Stage **Unstructured Data** die vorhandene Datei.
  - c. Optional: Wählen Sie in der Liste **Schreibmethode** die Option **Mehrere Dateien generieren** aus. Die Stage **Unstructured Data** erstellt mehrere Microsoft Excel-Dateien an der im Feld **Dateiname** angegebenen Position.
6. Machen Sie im Teilfenster **Eigenschaften** folgende Angaben:
    - a. Optional: Wählen Sie das Kontrollkästchen **Für alle Links definiert** (Standardeinstellung) aus, damit die Eigenschafteneinstellungen auf alle Eingabelinks angewendet werden.
    - b. Optional: Machen Sie Angaben für **Spaltenheader**. Wenn Sie **Spaltennamen** auswählen, schreibt die Stage **Unstructured Data** InfoSphere DataStage-Spaltennamen in die erste Zeile des Microsoft Excel-Arbeitsblatts. Wenn Sie die Standardeinstellung **Ohne** auswählen, schreibt die Stage **Unstructured Data** die Daten in die erste Zeile des Microsoft Excel-Arbeitsblatts.
    - c. Optional: Machen Sie Angaben für **Spaltenbreite anpassen**. Wenn Sie **Ja** auswählen, passt die Stage **Unstructured Data** die Breite der einzelnen Spalten im generierten Microsoft Excel-Arbeitsblatt an den Spalteninhalt an. Wenn Sie die Standardeinstellung **Nein** auswählen, passt die Stage **Unstructured Data** die Spaltenbreite nicht an.
    - d. Machen Sie Angaben für **Maximale Anzahl Zeilen im Arbeitsblatt**. Der Standardwert ist 65536. Wenn Sie **Spaltennamen** aus der Liste **Spaltenheader** auswählen, enthält dieser Wert eine Spaltennamenzeile.
  7. Ändern Sie die Arbeitsblattreihenfolge im Teilfenster **Arbeitsblattreihenfolge** und modifizieren Sie die Arbeitsblattnamen.
  8. Klicken Sie auf **OK**, um die angegebenen Einstellungen zu speichern.

#### Hinweis zum Ende des Laufs:

In parallelen InfoSphere DataStage-Jobs können einige Stages eine Markierung für das Ende eines Laufs senden, die das Ende einer UOW oder einer Transaktion angibt. Wenn alle aus dem Eingabelink extrahierten Datensätze in einer UOW enthalten sind (einzelner Lauf genannt), generiert die Stage **Unstructured Data** Microsoft Excel-Arbeitsblätter, die die maximale Anzahl Datensätze enthalten, bis alle Datensätze geschrieben sind.

Angenommen, die Stage **Unstructured Data** hat zwei Eingabelinks, DSLink1 und DSLink2. DSLink1 ist Sheet1 zugeordnet und DSLink2 ist Sheet2 zugeordnet. Die maximale Anzahl Datensätze in einem Arbeitsblatt beträgt 65.536. DSLink1 hat 100.000 Datensätze und DSLink2 hat 150.000 Datensätze. Kein Arbeitsblatt enthält Spaltennamen in der ersten Zeile. In diesem Fall enthält jedes Arbeitsblatt jeder Datei die folgende Anzahl Datensätze.

Dateiname	Sheet1	Sheet2
Workbook001.xlsx	65.536	65.536
Workbook002.xlsx	34.464	65.536
Workbook003.xlsx	0	18.928

Wenn aus dem Eingabelink extrahierte Datensätze auf mindestens zwei UOWs verteilt werden (mehrere Läufe genannt), schreibt die Stage **Unstructured Data** keine Datensätze mehr in das Microsoft Excel-Arbeitsblatt und erstellt eine neue Microsoft Excel-Datei, wenn in einem Lauf in mindestens einem Arbeitsblatt die maximale Anzahl Datensätze überschritten wird. Selbst wenn ein Arbeitsblatt mehr Datensätze enthalten kann, schreibt die Stage **Unstructured Data** keine Datensätze des nächsten Laufs in die vorherige Datei. Angenommen, jeder Link enthält beispielsweise die folgende Anzahl Datensätze in jedem Lauf:

Laufnummer	DSLink1	DSLink2
1	90.000	50.000
2	5.000	90.000
3	5.000	10.000

Im ersten Lauf erstellt die Stage **Unstructured Data** eine Microsoft Excel-Datei namens `Workbook001.xlsx`, die die beiden Arbeitsblätter `Sheet1` und `Sheet2` enthält. Die Stage **Unstructured Data** schreibt aus `DSLink1` extrahierte Datensätze in `Sheet1`, bis sie die maximale Anzahl Datensätze (65.536) liest, und schreibt alle Datensätze (50.000) aus `DSLink2` in `Sheet2`. Danach erstellt die Stage **Unstructured Data** eine Microsoft Excel-Datei namens `Workbook002.xlsx` und schreibt die restlichen Datensätze (24.464) des ersten Laufs und alle Datensätze (5.000) des zweiten Laufs von `DSLink1` in `Sheet1`. Obwohl in `Sheet2` von `Workbook001.xlsx` die maximale Anzahl Datensätze nicht überschritten wird, schreibt die Stage **Unstructured Data** im zweiten Lauf aus `DSLink2` extrahierte Datensätze (65.536) in `Sheet2` von `Workbook002.xlsx` und nicht von `Workbook001.xlsx`. Wenn die Anzahl Datensätze im zweiten Lauf von `DSLink2` die maximale Anzahl überschreitet, erstellt die Stage **Unstructured Data** eine Microsoft Excel-Datei namens `Workbook003.xlsx` und schreibt die restlichen Datensätze (24.464) des zweiten Laufs von `DSLink2` in `Sheet2` von `Workbook003.xlsx`. Da in beiden Arbeitsblättern genügend Platz ist, schreibt die Stage **Unstructured Data** im dritten Lauf alle aus `DSLink1` extrahierten Datensätze (5.000) und alle aus `DSLink2` extrahierten Datensätze (10.000) in `Sheet1` und `Sheet2` der Datei `Workbook003.xlsx`.

Daher wird die folgende Anzahl Datensätze in jedes Arbeitsblatt jeder Datei geschrieben:

Dateiname	Sheet1	Sheet2
Workbook001.xlsx	65.536 aus 1. Lauf	50.000 aus 1. Lauf
Workbook002.xlsx	29.464 (= 24.464 aus 1. Lauf + 5.000 aus 2. Lauf)	65.536 aus 2. Lauf
Workbook003.xlsx	5.000 aus 3. Lauf	34.464 (= 24.464 aus 2. Lauf + 10.000 aus 3. Lauf)

## Beispiele für das Schreiben von Daten in Microsoft Excel-Dateien

Sie können Beispieljob erstellen, die Daten in Microsoft Excel-Dateien schreiben.

Extrahieren Sie die Datei *IS-Installation\Clients\Samples\Connectors\UnstructuredData\_Samples.zip*, um die Dateien für die Beispiele abzurufen.

### Beispiel 1: Daten in ein Microsoft Excel-Arbeitsblatt schreiben:

Erstellen Sie einen Job, der über die Stage **Unstructured Data** Daten in ein Microsoft Excel-Arbeitsblatt schreibt.

#### Informationen zu diesem Vorgang

In diesem Beispiel wird eine Textdatei, *Employee.txt*, als Quelldaten verwendet. Die Quelldatei enthält Informationen zu Mitarbeitern im CSV-Format. Sie schreiben diese Informationen in ein Microsoft Excel-Arbeitsblatt.

*Schritt 1: Job erstellen:*

Erstellen Sie einen Beispieljob, der eine Stage **Unstructured Data** und eine Stage **Sequential File** enthält.

#### Vorgehensweise

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.
2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und wählen Sie **Neu > Paralleler Job** aus.
3. Ziehen Sie eine Stage **Sequential File** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.
4. Ziehen Sie eine Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich. Ordnen Sie die Stage rechts neben der Stage **Sequential File** an.
5. Erstellen Sie einen Link von der Stage **Sequential File** zur Stage **Unstructured Data**.
6. Benennen Sie die Stage und den Link um.
7. Speichern Sie den Job.

*Schritt 2: Stage 'Unstructured Data' konfigurieren:*

Konfigurieren Sie die Stage **Unstructured Data**, um Daten aus einem Bereich in einer Microsoft Excel-Datei zu extrahieren.

#### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Unstructured Data**, um die Stage-Eigenschaften zu öffnen.
2. Wählen Sie **Excel** in der Liste **Dokumenttyp** der Registerkarte **Stage** aus.
3. Wählen Sie **Datei erstellen** in der Liste **Schreibzugriff** aus.
4. Klicken Sie auf **Konfigurieren**.
5. Geben Sie im Konfigurationsfenster den vollständigen Dateipfad an, in dem Sie die Ausgabedatei erstellen wollen, gefolgt von dem Dateinamen *OutputOfSample4.xls*.
6. Wählen Sie **Überschreiben** in der Liste **Dateiaktualisierungsmodus** aus.
7. Wählen Sie in **Schreibmethode** die Option **Bestimmte Datei** aus.
8. Wählen Sie **Spaltennamen** im Feld **Spaltenheader** aus.

9. Wählen Sie im Feld **Spaltenbreite anpassen** auf der Registerkarte **Eigenschaft** die Option **Ja** aus.
10. Geben Sie im Teilfenster **Arbeitsblattreihenfolge** als Arbeitsblattname **Employee** an.
11. Klicken Sie auf **OK**.
12. Überprüfen Sie, ob die im Konfigurationsfenster angegebenen Werte auf der Registerkarte **Eigenschaft** des Stage-Editors gespeichert wurden.
13. Klicken Sie auf **OK**.

*Schritt 3: Stage 'Sequential File' konfigurieren:*

Konfigurieren Sie die Stage **Sequential File**.

#### Informationen zu diesem Vorgang

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

#### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Sequential File**.
2. Geben Sie auf der Seite **Eigenschaften** den Pfad der Eingabedatei `Employee.txt` an.
3. Definieren Sie die Spalten auf der Seite **Spalten** wie in der nachfolgenden Abbildung dargestellt.

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	EMPNO	<input checked="" type="checkbox"/>	Char		6		No	
2	FIRSTNME	<input type="checkbox"/>	VarChar		12		No	
3	MIDINIT	<input type="checkbox"/>	Char		1		Yes	
4	LASTNAME	<input type="checkbox"/>	VarChar		15		No	
5	WORKDEPT	<input type="checkbox"/>	Char		3		Yes	
6	PHONENO	<input type="checkbox"/>	Char		4		Yes	
7	HIREDATE	<input type="checkbox"/>	Date		10		Yes	
8	JOB	<input type="checkbox"/>	Char		8		Yes	
9	EDLEVEL	<input type="checkbox"/>	SmallInt		2		No	
10	SEX	<input type="checkbox"/>	Char		1		Yes	
11	BIRTHDATE	<input type="checkbox"/>	Date		10		Yes	
12	SALARY	<input type="checkbox"/>	Decimal		9	2	Yes	
13	BONUS	<input type="checkbox"/>	Decimal		9	2	Yes	
14	COMM	<input type="checkbox"/>	Decimal		9	2	Yes	

4. Klicken Sie auf **OK**.

*Schritt 4: Ausgabe des Jobs anzeigen:*

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

#### Vorgehensweise

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.

3. Zeigen Sie die Quelldaten in der Datei **Employee.txt** an.

Im folgenden Beispiel werden die Daten oben in der Datei gezeigt:

```
"000010","CHRISTINE","I","HAAS","A00","3978","1995-01-01","PRES","18","F","1963-08-24","0152750.00",
"0001000.00","0004220.00"
"000020","MICHAEL","L","THOMPSON","B01","3476","2003-10-10","MANAGER","18","M","1978-02-02",
"0094250.00","0000800.00","0003300.00"
"000030","SALLY","A","KWAN","C01","4738","2005-04-05","MANAGER","20","F","1971-05-11","0098250.00",
"0000800.00","0003060.00"
"000050","JOHN","B","GEYER","E01","6789","1979-08-17","MANAGER","16","M","1955-09-15","0080175.00",
"0000800.00","0003214.00"
"000060","IRVING","F","STERN","D11","6423","2003-09-14","MANAGER","16","M","1975-07-07","0072250.00",
"0000500.00","0002580.00"
"000070","EVA","D","PULASKI","D21","7831","2005-09-30","MANAGER","16","F","2003-05-26","0096170.00",
"0000700.00","0002893.00"
```

4. Öffnen Sie die Datei **OutputOfExample4.xlsx** nach der Jobausführung, um das Ergebnis anzuzeigen. Die folgende Tabelle zeigt die Ausgabedaten in der Datei **OutputOfExample4.xlsx** an.

Tabelle 11. Details zu Ausgabedaten in einer Microsoft Excel-Datei

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	WORK DEPT	HIRE DATE	JOB	SEX	BIRTH DATE	SAL ARY	BON DE
10	CHRISTINE	I	HAAS	3978	A00	1/1/1995	PRES	F	8/24/1963	152750	1000
20	MICHAEL	L	THOMPSON	3476	A00	10/10/2003	MANAGER	M	2/2/1978	94250	800
30	SALLY	A	KWAN	4738	A00	4/5/2005	MANAGER	F	5/11/1971	98250	800
50	JOHN	B	GEYER	6789	A00	8/17/1979	MANAGER	M	9/15/1955	80175	800
60	IRVING	F	STERN	6423	B01	9/14/2003	MANAGER	M	7/7/1975	72250	500
70	EVA	D	PULASKI	7831	B01	9/30/2005	MANAGER	F	5/26/2003	96170	700
90	EILEEN	W	HENDERSON	5498	B01	8/15/2000	MANAGER	F	5/15/1971	89750	600
100	THEODORE	Q	SPENSER	972	B01	6/19/2000	MANAGER	M	12/18/1980	86150	500
110	VINCENZO	G	LUCCHESSI	3490	B01	5/16/1988	SALES REP	M	11/5/1959	665004	900
120	SEAN		O'CONNELL	2167	B01	12/5/1993	CLERK	M	10/18/1972	9250	600
130	DELORES	M	QUINTANA	4578	B01	7/28/2001	ANALYST	F	9/15/1955	73800	500
140	HEATHER	A	NICHOLLS	1793	B01	12/15/2006	ANALYST	F	1/19/1976	68420	600
150	BRUCE		ADAMSON	4510	B01	2/12/2002	DESIGNER	M	5/17/1977	55280	500
160	ELIZABETH	R	PIANKA	3782	B01	10/11/2006	DESIGNER	F	4/12/1980	62250	400
170	MASATOSHI	J	YOSHIMURA	2890	B01	9/15/1999	DESIGNER	M	1/5/1981	44680	500
180	MARILYN	S	SCOUTTEN	1682	B01	7/7/2003	DESIGNER	F	2/21/1979	51340	500
190	JAMES	H	WALKER	2986	C01	7/26/2004	DESIGNER	M	6/25/1982	50450	400
200	DAVID		BROWN	4501	C01	3/3/2002	DESIGNER	M	5/29/1971	57740	600



Tabelle 11. Details zu Ausgabedaten in einer Microsoft Excel-Datei (Forts.)

EMP NO	FIRST NAME	MID INIT	LAST NAME	PHONE NO	WORK DEPT	HIRE DATE	JOB	SEX	BIRTH DATE	SALARY	BONDE
210	WILLIAM	T	JONES	942	C01	4/11/1998	DESIGNER	M	2/23/2003	68270	400
220	JENNIFER	K	LUTZ	672	C01	8/29/1998	DESIGNER	F	3/19/1978	49840	600
230	JAMES	J	JEFFERSON	2094	C01	11/21/1996	CLERK	M	5/30/1980	42180	400
240	SALVATORE	M	MARINO	3780	C01	12/5/2004	CLERK	M	3/31/2002	48760	600
250	DANIEL	S	SMITH	961	C01	10/30/1999	CLERK	M	11/12/1969	49180	400
260	SYBIL	P	JOHNSON	8953	C01	9/11/2005	CLERK	F	10/5/1976	47250	300
270	MARIA	L	PEREZ	9001	D01	9/30/2006	CLERK	F	5/26/2003	37380	500
280	ETHEL	R	SCHNEIDER	8997	D01	3/24/1997	OPERATOR	F	3/28/1976	36250	500
290	JOHN	R	PARKER	4502	D01	5/30/2006	OPERATOR	M	7/9/1985	35340	300
300	PHILIP	X	SMITH	2095	D01	6/19/2002	OPERATOR	M	10/27/1976	37750	400
310	MAUDE	F	SETRIGHT	3332	D01	9/12/1994	OPERATOR	F	4/21/1961	35900	300
320	RAMLAL	V	MEHTA	9990	D01	7/7/1995	FIELD REP	M	8/11/1962	39950	400
330	WING		LEE	2103	D01	2/23/2006	FIELD REP	M	7/18/1971	45370	500
340	JASON	R	GOUNOT	5698	D01	5/5/1977	FIELD REP	M	5/17/1956	43840	500

**Beispiel 2: Daten in mehrere Arbeitsblätter einer Microsoft Excel-Datei schreiben:**

Erstellen Sie einen Job, der über die Stage **Unstructured Data** Daten in mehrere Microsoft Excel-Arbeitsblätter schreibt.

**Informationen zu diesem Vorgang**

In diesem Beispiel werden die folgenden 7 Textdateien als Quelldaten verwendet: DEPT\_A00.txt, DEPT\_B01.txt, DEPT\_C01.txt, DEPT\_D11.txt, DEPT\_D21.txt, DEPT\_E11.txt und DEPT\_E21.txt. Jede Quellendatei enthält Informationen zu Mitarbeitern der entsprechenden Abteilung im CSV-Format. Sie schreiben Informationen von jeder Quellendatei in jedes Arbeitsblatt. Die erstellte Microsoft Excel-Datei enthält mehrere Arbeitsblätter.

*Schritt 1: Job erstellen:*

Erstellen Sie einen Beispieljob, der sieben Stages **Sequential File** und eine Stage **Unstructured Data** enthält.

**Vorgehensweise**

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.

2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und wählen Sie **Neu > Paralleler Job** aus.
3. Ziehen Sie eine Stage **Sequential File** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.
4. Wiederholen Sie Schritt 3 noch sechsmal. Ordnen Sie sie in vertikaler Linie an.
5. Ziehen Sie eine Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich. Ordnen Sie die Stage rechts neben den Stages **Sequential File** an.
6. Erstellen Sie einen Link von einer Stage **Sequential File** zur Stage **Unstructured Data**.
7. Benennen Sie die Stages und Links um. Benennen Sie die Links, sodass sie mit dem entsprechenden Abteilungsnamen (A00, B01, C01, D11, D21, E11, E21) übereinstimmen.
8. Speichern Sie den Job.

*Schritt 2: Stage 'Unstructured Data' konfigurieren:*

Konfigurieren Sie die Stage **Unstructured Data**, um Daten aus einem Bereich in einer Microsoft Excel-Datei zu extrahieren.

#### **Vorgehensweise**

1. Doppelklicken Sie auf die Stage **Unstructured Data**, um die Stage-Eigenschaften zu öffnen.
2. Wählen Sie **Excel** in der Liste **Dokumenttyp** der Registerkarte **Stage** aus.
3. Wählen Sie **Datei erstellen** in der Liste **Schreibzugriff** aus.
4. Klicken Sie auf **Konfigurieren**.
5. Geben Sie im Konfigurationsfenster den vollständigen Dateipfad an, in dem Sie die Ausgabedatei erstellen wollen, gefolgt von dem Dateinamen `OutputOfSample5.xls`.
  - a. Wählen Sie **Überschreiben** in der Liste **Dateiaktualisierungsmodus** aus.
  - b. Wählen Sie in der Liste **Schreibmethode** die Option **Bestimmte Datei** aus. Stellen Sie sicher, dass die Option **Für alle Links definiert** ausgewählt ist.
  - c. Wählen Sie **Spaltennamen** im Feld **Spaltenheader** aus.
  - d. Wählen Sie im Feld **Spaltenbreite anpassen** auf der Registerkarte **Eigenschaft** die Option **Ja** aus.
  - e. Ordnen Sie im Teilfenster **Arbeitsblattreihenfolge** die Links zu A00, B01, C01, D11, D21, E11, E21 mithilfe der Schaltflächen **Nach oben** und **Nach unten** an.
  - f. Geben Sie im Teilfenster **Arbeitsblattreihenfolge** die Arbeitsblattnamen an, sodass sie mit den entsprechenden Linknamen übereinstimmen.
  - g. Klicken Sie auf **OK**.
6. Klicken Sie auf **OK**.

*Schritt 3: Stages 'Sequential File' konfigurieren:*

Konfigurieren Sie die Stages **Sequential File**.

#### **Informationen zu diesem Vorgang**

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Sequential File**.
2. Geben Sie auf der Seite **Eigenschaften** den Pfad zur Datei DEPT\_A00.txt an.
3. Definieren Sie die Spalten auf der Seite **Spalten** wie in der nachfolgenden Abbildung dargestellt.

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	EMPNO	<input checked="" type="checkbox"/>	Char		6		No	
2	FIRSTNME	<input type="checkbox"/>	VarChar		12		No	
3	MIDINIT	<input type="checkbox"/>	Char		1		Yes	
4	LASTNAME	<input type="checkbox"/>	VarChar		15		No	
5	WORKDEPT	<input type="checkbox"/>	Char		3		Yes	
6	PHONENO	<input type="checkbox"/>	Char		4		Yes	
7	HIREDATE	<input type="checkbox"/>	Date		10		Yes	
8	JOB	<input type="checkbox"/>	Char		8		Yes	
9	EDLEVEL	<input type="checkbox"/>	SmallInt		2		No	
10	SEX	<input type="checkbox"/>	Char		1		Yes	
11	BIRTHDATE	<input type="checkbox"/>	Date		10		Yes	
12	SALARY	<input type="checkbox"/>	Decimal		9	2	Yes	
13	BONUS	<input type="checkbox"/>	Decimal		9	2	Yes	
14	COMM	<input type="checkbox"/>	Decimal		9	2	Yes	

4. Klicken Sie auf **OK**.
5. Wiederholen Sie die Schritte 1 bis 4 sechsmal für die verbleibenden Quelldateien DEPT\_B01.txt, DEPT\_C01.txt, DEPT\_D11.txt, DEPT\_D21.txt, DEPT\_E11.txt und DEPT\_E21.txt. Alle Stages **Sequential File** haben dieselbe Spaltendefinition in ihrem Ausgabelink.

*Schritt 4: Ausgabe des Jobs anzeigen:*

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

### Vorgehensweise

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.
3. Öffnen Sie die Datei OutputOfExample5.xlsx nach der Jobausführung, um das Ergebnis anzuzeigen. Die Ausgabedaten in einem Microsoft Excel-Arbeitsblatt sollten mit den Informationen in der Quelltextdatei übereinstimmen. Die Microsoft Excel-Datei Sheet A00 sollte mit den Quelldaten in der Datei DEPT\_A00.txt übereinstimmen. Die Microsoft Excel-Datei Sheet B01 sollte mit den Quelldaten in der Datei DEPT\_B01.txt übereinstimmen. Die Microsoft Excel-Datei Sheet C01 sollte mit den Quelldaten in der Datei DEPT\_C01.txt übereinstimmen. Die Microsoft Excel-Datei Sheet D11 sollte mit den Quelldaten in der Datei DEPT\_D11.txt übereinstimmen. Die Microsoft Excel-Datei Sheet D21 sollte mit den Quelldaten in der Datei DEPT\_D21.txt übereinstimmen. Die Microsoft Excel-Datei Sheet E11 sollte mit den Quelldaten in der Datei DEPT\_E11.txt übereinstimmen. Die Microsoft Excel-Datei Sheet E21 sollte mit den Quelldaten in der Datei DEPT\_E21.txt übereinstimmen.

### Beispiel 3: Daten in mehrere Microsoft Excel-Dateien schreiben:

Sie können einen Job erstellen, der über die Stage **Unstructured Data** Daten in mehrere Microsoft Excel-Dateien schreibt.

#### Informationen zu diesem Vorgang

In diesem Beispiel wird eine Textdatei, `Employee.txt`, als Quelldaten verwendet. Die Quellendatei enthält Informationen zu 42 Mitarbeitern im CSV-Format. Diese Informationen werden durch den für die Option **Maximale Anzahl Zeilen im Arbeitsblatt** angegebenen Wert geteilt und in mehrere Microsoft Excel-Dateien geschrieben.

*Schritt 1: Job erstellen:*

Erstellen Sie einen Beispieljob, der eine Stage **Sequential File** und eine Stage **Unstructured Data** enthält.

#### Vorgehensweise

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.
2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und wählen Sie **Neu > Paralleler Job** aus.
3. Ziehen Sie eine Stage **Sequential File** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.
4. Ziehen Sie eine Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich. Ordnen Sie die Stage rechts neben der Stage **Sequential File** an.
5. Erstellen Sie einen Link von der Stage **Sequential File** zur Stage **Unstructured Data**.
6. Benennen Sie die Stage und den Link um.
7. Speichern Sie den Job.

*Schritt 2: Stage 'Unstructured Data' konfigurieren:*

Konfigurieren Sie die Stage **Unstructured Data**, um die Daten in ein Microsoft Excel-Arbeitsblatt zu schreiben.

#### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Unstructured Data**, um die Stage-Eigenschaften zu öffnen.
2. Wählen Sie **Excel** in der Liste **Dokumenttyp** der Registerkarte **Stage** aus.
3. Wählen Sie **Datei erstellen** in der Liste **Schreibzugriff** aus.
4. Klicken Sie auf **Konfigurieren**.
5. Geben Sie im Konfigurationsfenster den Pfad an, in dem Sie die Ausgabedatei erstellen wollen, gefolgt von dem Dateipräfix `OutputOfSample6_`. Wenn die Dateien generiert werden, werden diesem Präfix eine dreistellige fortlaufende Zahl und die Erweiterung `.xlsx` (beispielsweise `001.xlsx`, `002.xlsx`) hinzugefügt. Beispiel: `001.xlsx`, `002.xlsx`.
6. Wählen Sie **Überschreiben** in der Liste **Dateiaktualisierungsmodus** aus.
7. Wählen Sie in der Liste **Schreibmethode** die Option **Mehrere Dateien generieren** aus.

8. Wählen Sie **Spaltennamen** im Feld **Spaltenheader** aus.
9. Wählen Sie im Feld **Spaltenbreite anpassen** auf der Registerkarte **Eigenschaft** die Option **Ja** aus.
10. Geben Sie in **Maximale Anzahl Zeilen im Arbeitsblatt** den Wert **10** an.
11. Im Teilfenster **Arbeitsblattreihenfolge** geben Sie **Employee** als Arbeitsblattnamen an.
12. Klicken Sie auf **OK**.

*Schritt 3: Stages 'Sequential File' konfigurieren:*

Konfigurieren Sie die Stages **Sequential File**.

### Informationen zu diesem Vorgang

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Sequential File**.
2. Geben Sie auf der Seite **Eigenschaften** den Pfad zur Datei `Employee.txt` an.
3. Definieren Sie die Spalten auf der Seite **Spalten** wie in der nachfolgenden Abbildung dargestellt.

	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Data element
1	EMPNO	<input checked="" type="checkbox"/>	Char		6		No	
2	FIRSTNME	<input type="checkbox"/>	VarChar		12		No	
3	MIDINIT	<input type="checkbox"/>	Char		1		Yes	
4	LASTNAME	<input type="checkbox"/>	VarChar		15		No	
5	WORKDEPT	<input type="checkbox"/>	Char		3		Yes	
6	PHONENO	<input type="checkbox"/>	Char		4		Yes	
7	HIREDATE	<input type="checkbox"/>	Date		10		Yes	
8	JOB	<input type="checkbox"/>	Char		8		Yes	
9	EDLEVEL	<input type="checkbox"/>	SmallInt		2		No	
10	SEX	<input type="checkbox"/>	Char		1		Yes	
11	BIRTHDATE	<input type="checkbox"/>	Date		10		Yes	
12	SALARY	<input type="checkbox"/>	Decimal		9	2	Yes	
13	BONUS	<input type="checkbox"/>	Decimal		9	2	Yes	
14	COMM	<input type="checkbox"/>	Decimal		9	2	Yes	

4. Klicken Sie auf **OK**.

*Schritt 4: Ausgabe des Jobs anzeigen:*

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

### Vorgehensweise

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.
3. Öffnen Sie die Ausgabedatei nach der Jobausführung, um das Ergebnis anzuzeigen. Die Ausgabedaten in einem Microsoft Excel-Arbeitsblatt sollten mit den

Informationen in der Quellentextdatei übereinstimmen. Da die angegebene maximale Anzahl Zeilen in einem Arbeitsblatt 10 ist und 1 Zeile für den Spaltenheader enthalten ist, enthält jede Ausgabedatei maximal 9 Datensätze aus der Eingabedatei. Die Eingabedatei enthält 42 Datensätze, daher werden 5 Dateien generiert. Es werden die Dateien `OutputOfExample6_001.xlsx`, `OutputOfExample6_002.xlsx`, `OutputOfExample6_003.xlsx`, `OutputOfExample6_004.xlsx` und `OutputOfExample6_005.xlsx` generiert.

## Daten in vorhandene Microsoft Excel-Dateien schreiben

Sie können die Stage **Unstructured Data** in Jobs verwenden, die Daten in eine vorhandene Microsoft Excel-Datei schreiben. Sie können auch eine Microsoft Excel-Datei kopieren und Daten in die Kopie schreiben.

Wenn die Stage **Unstructured Data** Daten in eine vorhandene Microsoft Excel-Datei schreibt, schreibt sie nur die Zellendaten. Wenn eine Microsoft Excel-Zelle, in die die Stage schreibt, ein beliebiges Format hat, behält die Stage das vorhandene Format bei. Wenn die Microsoft Excel-Datei eine Formel oder ein Diagramm enthält, die bzw. das auf Zellen verweist, in die von der Stage **Unstructured Data** geschrieben wird, wird die Formel bzw. das Diagramm beim Öffnen durch Microsoft Excel neu berechnet.

Sie können von beliebigen InfoSphere DataStage-Spalten aus in Microsoft Excel-Spalten schreiben. Die Namen und die Reihenfolge von InfoSphere DataStage-Spalten und Microsoft Excel-Spalten müssen nicht übereinstimmen. Wenn das Microsoft Excel-Arbeitsblatt in der ersten Zeile einen Header enthält, können Sie die Stage **Unstructured Data** so konfigurieren, dass die Werte in der ersten Zeile zur Bestimmung der Spalte verwendet werden, in die Datensätze geschrieben werden. Sie können bis zu 1.048.576 Zeilen eines Microsoft Excel-Arbeitsblatts schreiben. Die InfoSphere DataStage-Quellenspalten müssen in der Entwicklungsphase definiert werden.

### Zugehörige Verweise:

„Datentypkonvertierungen von InfoSphere DataStage in Microsoft Excel“ auf Seite 45

Bevor die Stage **Unstructured Data** aus Eingabelinks extrahierte Daten in Microsoft Excel-Dateien schreibt, werden die Daten in Microsoft Excel-Datentypen konvertiert.

## Jobs entwerfen, die Daten in ein vorhandenes Microsoft Excel-Arbeitsblatt schreiben

Mithilfe einer Stage **Unstructured Data** können Sie Jobs entwerfen, die Daten in ein vorhandenes Microsoft Excel-Arbeitsblatt schreiben. Es kann immer nur eine Microsoft Excel-Datei von einer Stage **Unstructured Data** aktualisiert werden. Es kann nur eine Stage **Unstructured Data** dieselbe Microsoft Excel-Datei in einem Job aktualisieren.

### Vorbereitende Schritte

- Installieren Sie InfoSphere Information Server in der Sprache, in der die zu extrahierende Microsoft Excel-Datei vorliegt.
- Stellen Sie sicher, dass die Anwendung, die den Inhalt von Microsoft Excel-Arbeitsblättern anzeigt (z. B. Microsoft Excel, Microsoft Excel Viewer oder IBM Lotus Symphony), auf Ihrem Client-Computer installiert ist.
- Stellen Sie sicher, dass die Dateierweiterung `.xlsx` der Anwendung zugeordnet ist, mit der Sie Microsoft Excel-Arbeitsblätter anzeigen.

## Vorgehensweise

1. Definieren Sie einen Job, der eine Stage 'Unstructured Data' enthält.
2. Konfigurieren Sie die Stage 'Unstructured Data' als Ziel.
3. Kompilieren Sie den Job und führen Sie ihn aus.

### Job definieren, der eine Stage 'Unstructured Data' enthält:

Sie müssen einen Job erstellen, der die Stage **Unstructured Data** enthält, erforderliche weitere Stages hinzufügen und die notwendigen Links erstellen, bevor Sie Daten aus Microsoft Excel-Dateien lesen bzw. Daten in Microsoft Excel-Dateien schreiben können.

## Vorgehensweise

1. Klicken Sie im Designer-Client auf **Datei > Neu**.
2. Klicken Sie im Fenster **Neu** auf das Symbol **Paralleler Job** und klicken Sie dann auf **OK**.
3. Klicken Sie in der Palette auf **Datei**.
4. Ziehen Sie das Symbol **Stage 'Unstructured Data'** auf den Entwicklungsbereich.
5. Erstellen Sie Stages für den Job.
6. Wählen Sie die Kategorie **Allgemein** im Palettenmenü auf der linken Seite des Designer-Clients aus und erstellen Sie dann die erforderlichen Links für den Job.
7. (Optional) Doppelklicken Sie auf das Symbol **Stage 'Unstructured Data'**, um die folgenden Attribute einzugeben oder zu modifizieren:
  - **Stage:** Modifizieren Sie den Standardnamen der Stage. Sie können bis zu 255 Zeichen eingeben. Sie können den Namen der Stage alternativ im Jobentwicklungsbereich modifizieren.
  - **Beschreibung:** Geben Sie eine Beschreibung der Stage ein.
8. Klicken Sie auf **Speichern**.

### Stage 'Unstructured Data' als Ziel konfigurieren:

Sie können die Stage **Unstructured Data** konfigurieren, um eine vorhandene Microsoft Excel-Datei zu modifizieren. Die Ausführung im parallelen Modus wird beim Modifizieren einer vorhandenen Microsoft Excel-Datei nicht unterstützt. Sie müssen die Stage **Unstructured Data** in einem sequenziellen Modus konfigurieren.

## Informationen zu diesem Vorgang

Die Stage **Unstructured Data** unterstützt nur das OOXML-Format (.xlsx) von Microsoft Excel-Dateien als Zieldatei und Schablonendatei.

Die Stage **Unstructured Data** unterstützt nicht die Modifizierung von .xls-Dateien oder von mit Kennwort verschlüsselten Dateien.

Die Stage **Unstructured Data** unterstützt Microsoft Excel-Dateien nicht, die von Microsoft Excel for Mac erstellt werden.

Die Dateinamen, Arbeitsblattnamen und Headernamen müssen in der Standardcodepage von Microsoft Windows ausgedrückt werden können, in der der InfoSphere DataStage Designer-Client installiert ist. Wenn die Microsoft Excel-Datei, die Sie verarbeiten wollen, andere Zeichen enthält, müssen Sie die Microsoft Excel-Datei bearbeiten, sodass sie diese Zeichen nicht mehr enthält.

### Vorgehensweise

1. Doppelklicken Sie im Entwicklungsbereich für parallele Jobs auf die Stage **Unstructured Data**.
2. Wählen Sie auf der Registerkarte **Eingabe** den Eingabelink im Feld **Eingabename (vorgeschaltete Stage)** aus.
3. Stellen Sie auf der Seite **Spalten** sicher, dass die Spalten ordnungsgemäß definiert sind.
4. Wählen Sie auf der Registerkarte **Stage** als Dokumenttyp **Excel** auf der Seite **Eigenschaften** aus.
5. Wählen Sie **Vorhandene Datei modifizieren** in der Liste **Schreibzugriff** aus.
6. Klicken Sie auf **Konfigurieren**, um die Eigenschaften für das Schreiben von Daten in eine vorhandene Microsoft Excel-Datei zu konfigurieren.
7. Geben Sie die Details der Zieldatei an, in die die Daten geschrieben werden sollen.
  - a. Geben Sie im Feld **Dateiname** den Namen der Datei an, in die die Daten geschrieben werden sollen.
  - b. Wenn mehr als 32 Spalten im Microsoft Excel-Arbeitsblatt aktualisiert werden, geben Sie die Anzahl der Spalten im Feld **Anzahl der zu ladenden Spalten** an.
  - c. Klicken Sie auf **Laden**, um die Spalten in das Teilfenster **Importieren** zu laden.
8. Wählen Sie in der Liste **Link** einen zu konfigurierenden Eingabelink aus.
9. Geben Sie die Microsoft Excel-Details an, die in das Teilfenster **Importieren** importiert werden sollen.
  - a. Wählen Sie in der Liste **Arbeitsblatt** das zu aktualisierende Microsoft Excel-Arbeitsblatt aus.
  - b. Wählen Sie in der Liste der Spalten die zu aktualisierenden Microsoft Excel-Spalten aus.
  - c. Klicken Sie auf **Importieren**.
10. Ordnen Sie die importierten Microsoft Excel-Spalten den InfoSphere DataStage-Spalten zu, die im Eingabelink definiert sind. Wählen Sie in **InfoSphere DataStage-Spalte** eine InfoSphere DataStage-Spalte aus, in die die Daten geschrieben werden sollen.
11. Wiederholen Sie die Schritte 7 bis 9 für alle Eingabelinks.
12. Klicken Sie auf **OK**.

### Jobparameter beim Modifizieren einer vorhandenen Microsoft Excel-Datei verwenden:

Die Stage **Unstructured Data** kann keine neuen Jobparameter im Konfigurationsfenster erstellen. Sie können jedoch die Jobparameter im Konfigurationsfenster verwenden. Sie müssen Jobparameter im Fenster **Jobeigenschaften** vor oder nach der Arbeit im Konfigurationsfenster erstellen, indem Sie **Bearbeiten > Jobeigenschaften** im Client von IBM InfoSphere DataStage und QualityStage Designer auswählen. Weitere Informationen zum Erstellen von Jobparametern finden Sie in Lektion 2.4, Parameter hinzufügen, im IBM InfoSphere DataStage-Lernprogramm für parallele Jobs.

Ein Jobparameter wird im Konfigurationsfenster mit dem Zeichen # angegeben. Beispielsweise wird der Jobparameter *Dateiname* im Konfigurationsfenster als **#Da-**



**teiname#** angegeben. Für ein Feld mit dem Datentyp 'string' wie die Eigenschaft **Dateiname** können Sie den Namen des Jobparameters von # umschlossen direkt eingeben.

Wenn Sie für die Eigenschaft **Listentyp** wie z. B. **Aktion, wenn die Datei bereits vorhanden ist** Jobparameter verwenden wollen, müssen Sie einen Listentypparameter erstellen, der eine Liste von Zeichenfolgevariablen enthält. Die Zeichenfolgevariablen müssen mit dem Beschriftungstext der entsprechenden Eigenschaft im Konfigurationsfenster übereinstimmen. Wenn Sie z. B. Jobparameter für die Eigenschaft **Aktion, wenn die Datei bereits vorhanden ist** verwenden wollen, müssen Sie einen Listentypparameter erstellen, der die Zeichenfolgevariablen **Fehler** und **Überschreiben** enthält. Wählen Sie nach der Erstellung eines Jobparameters **<Parametrisieren...>** im Konfigurationsfenster aus und geben Sie den Jobparameternamen vom Zeichen # umschlossen im Dialogfeld **Eingabeparameter** an. Klicken Sie auf **Laden**, um Variablen in der Anzeige **Jobparameter auflösen** zu bearbeiten oder auszuwählen.

### **Beispiel: Daten in vorhandene Microsoft Excel-Dateien schreiben**

Sie können einen Job erstellen, der über die Stage **Unstructured Data** Daten in vorhandene Microsoft Excel-Dateien schreibt.

#### **Informationen zu diesem Vorgang**

In diesem Beispiel wird eine Textdatei, *Employee.txt*, als Quelldaten verwendet. Die Quelldatei enthält Informationen zu Mitarbeitern im CSV-Format. Sie schreiben diese Informationen in die Microsoft Excel-Datei *ExcelModifySample1.xlsx*. Die Datei *ExcelModifySample1.xlsx* enthält ein Arbeitsblatt namens **Employee**, das wiederum folgende Spalten enthält: **EMP NO, FIRST NAME, MIDINIT, LAST NAME, HIRE DATE, JOB, SEX, SALARY, BONUS** und **TOTAL PAY**. In diesem Job wird in diese Spalten, mit Ausnahme der Spalte **TOTAL PAY**, geschrieben. Die Spalte **TOTAL PAY** enthält eine Formel zur Berechnung des Gehalts und des Bonus.

Sie können Beispieljob erstellen, die Daten in vorhandene Microsoft Excel-Dateien schreiben.

Extrahieren Sie die Datei *IS-Installation\Clients\Samples\Connectors\UnstructuredData\_Samples.zip*, um die Dateien für die Beispiele abzurufen.

#### **Schritt 1: Job erstellen:**

Erstellen Sie einen Beispieljob, der eine Stage **Sequential File** und eine Stage **Unstructured Data** enthält. In diesem Beispiel liest die Stage **Sequential File** Daten aus einer Microsoft Excel-Datei und dann schreibt die Stage **Unstructured Data** Daten in die unstrukturierte Datenquelle.

#### **Vorgehensweise**

1. Starten Sie den Client von IBM InfoSphere DataStage und QualityStage Designer.
2. Klicken Sie im Teilfenster **Repository** mit der rechten Maustaste auf den Ordner **Jobs** und wählen Sie **Neu > Paralleler Job** aus.
3. Ziehen Sie eine Stage **Sequential File** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich.

4. Ziehen Sie eine Stage **Unstructured Data** vom Abschnitt **Datei** der Palette auf den Entwicklungsbereich. Ordnen Sie die Stage rechts neben der Stage **Sequential File** an.
5. Erstellen Sie einen Link von der Stage **Sequential File** zur Stage **Unstructured Data**.
6. Benennen Sie die Stage und den Link um.
7. Speichern Sie den Job.

### Schritt 2: Stage 'Sequential File' konfigurieren:

Sie können die Stage **Sequential File** konfigurieren, um Daten aus der Quelldatei zu lesen. Sie müssen den Namen der Quelldatei angeben und die Spaltennamen und SQL-Eigenschaften definieren.

#### Vorbereitende Schritte

In diesem Beispiel wird die Stage **Sequential File** als Ausgabestage verwendet. Sie können zum Erstellen der Ausgabe eine beliebige andere Ausgabestage verwenden.

#### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Sequential File**.
2. Geben Sie auf der Seite **Eigenschaften** den Pfad zur Datei `Employee.txt` an.
3. Definieren Sie die Spalten auf der Seite **Spalten** wie in der nachfolgenden Tabelle dargestellt.

Spaltenname	Schlüssel	SQL-Typ	Erweitert	Länge	Nachkomma- stellen	Kann Nullwerte enthalten	Beschreibung
EMP_NO		Integer				Ja	
FIRST_NAME		VarChar				Ja	
MIDINIT		VarChar				Ja	
LAST_NAME		VarChar				Ja	
HIRE_DATE		Datum				Ja	
JOB		VarChar				Ja	
SEX		VarChar				Ja	
SALARY		Integer				Ja	
BONUS		Integer				Ja	

4. Klicken Sie auf **OK**.

### Schritt 3: Stage 'Unstructured Data' konfigurieren:

Konfigurieren Sie die Stage **Unstructured Data**, um die Daten in eine vorhandene Microsoft Excel-Datei zu schreiben.

#### Vorgehensweise

1. Doppelklicken Sie auf die Stage **Unstructured Data**, um die Stage-Eigenschaften zu öffnen.
2. Wählen Sie **Excel** in der Liste **Dokumenttyp** der Registerkarte **Stage** aus.
3. Wählen Sie **Datei erstellen** in der Liste **Schreibzugriff** aus.
4. Klicken Sie auf **Konfigurieren**.

5. Geben Sie im Konfigurationsfenster den Pfad an, in dem Sie die Ausgabedatei erstellen wollen, gefolgt von dem Präfix `ExcelModifySample1`.
6. Klicken Sie auf **Laden**.
7. Wählen Sie **Erste Zeile** im Feld **Spaltenheader** aus.
8. Geben Sie in **Ab dieser Zeile mit dem Schreiben beginnen** den Wert **2** an.
9. Wählen Sie in der Liste **Arbeitsblatt** den Eintrag **Employee** aus. Stellen Sie sicher, dass die Microsoft Excel-Spalten A bis I ausgewählt sind.
10. Klicken Sie auf **Importieren**.
11. Definieren Sie im Teilfenster **Zuordnen** das Mapping zwischen der Microsoft Excel-Spalte und der InfoSphere DataStage-Spalte.
12. Klicken Sie auf **OK**.

#### **Schritt 4: Ausgabe des Jobs anzeigen:**

Öffnen Sie die Datei nach der Jobausführung und prüfen Sie die Ausgabe.

#### **Vorgehensweise**

1. Speichern Sie den Job.
2. Kompilieren Sie den Job und führen Sie ihn aus.
3. Öffnen Sie die Ausgabedatei nach der Jobausführung, um das Ergebnis anzuzeigen. Die Ausgabedaten in einem Microsoft Excel-Arbeitsblatt sollten mit den Informationen in der Quelltextdatei übereinstimmen.



## Kapitel 2. Referenz

In den folgenden Abschnitten werden die unterstützten Microsoft Excel-Typen, Mappings zwischen Microsoft Excel-Typen und DataStage-Typen sowie Jobabbruchbedingungen beschrieben.

### Datentypkonvertierungen von Microsoft Excel in InfoSphere DataStage

Aus Microsoft Excel extrahierte Daten werden in InfoSphere DataStage-Datentypen konvertiert, bevor die Stage **Unstructured Data** sie in den Ausgabelink schreibt.

In der folgenden Tabelle wird das Mapping zwischen Microsoft Excel-Datentypen und InfoSphere DataStage-Datentypen gezeigt.

**Anmerkung:** Die Stage **Unstructured Data** ordnet die Datentypkonvertierungen von Microsoft Excel InfoSphere DataStage nur zu, wenn die Stage **Unstructured Data** Datensätze aus der Microsoft Excel-Datenquelle liest.

*Tabelle 12. Mapping zwischen Microsoft Excel-Zellenwertdatentypen und InfoSphere DataStage-Datentypen*

Microsoft Excel-Zellendatentyp	DataStage-Datentyp
Leer	<b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt <b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Bruchzahldatentypen</b> Double Float Real <b>Dezimaldatentypen</b> Decimal Numeric <b>Datum- und Zeitdatentypen</b> Date Time Timestamp
Boolesch	<b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt <b>Anmerkung:</b> Zuordnung: TRUE: 1, FALSE: 0 <b>Textdatentypen</b> Char VarChar LongVarChar <b>Anmerkung:</b> Zuordnung: TRUE: "true", FALSE: "false" <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Anmerkung:</b> Zuordnung: TRUE: "true", FALSE: "false"

Tabelle 12. Mapping zwischen Microsoft Excel-Zellenwertdatentypen und InfoSphere DataStage-Datentypen (Forts.)

Microsoft Excel-Zellendatentyp	DataStage-Datentyp
Fehler	<p><b>Textdatentypen</b> Char VarChar LongVarChar <b>Anmerkung:</b> Zeichenfolgeausdruck des Fehlers. Beispiel: #NAME?</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Anmerkung:</b> Zeichenfolgeausdruck des Fehlers. Beispiel: #NAME?</p>
Numerisch	<p><b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt</p> <p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p> <p><b>Bruchzahldatentypen</b> Double Float Real</p> <p><b>Dezimaldatentypen</b> Decimal Numeric</p> <p><b>Datum- und Zeitdatentypen</b> Date Time Timestamp</p>
Zeichenfolge	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p> <p><b>Datum- und Zeitdatentypen</b> Date Time Timestamp</p>

Tabelle 13. Sonstige Microsoft Excel-Zelleninformationsdatentypen und InfoSphere DataStage-Datentypen

Sonstige Microsoft Excel-Zelleninformationsdatentypen	InfoSphere DataStage-Datentypen
Formel	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p>
Kommentar	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p>
Kommentarautor	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p>

Tabelle 13. Sonstige Microsoft Excel-Zelleninformationsdatentypen und InfoSphere DataStage-Datentypen (Forts.)

Sonstige Microsoft Excel-Zelleninformationsdatentypen	InfoSphere DataStage-Datentypen
Hyperlinktyp	<b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt <b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Hyperlinkadresse	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Hyperlinkbezeichnung	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar

Tabelle 14. Mapping zwischen Microsoft Excel-Zellenwertdatentypen und InfoSphere DataStage-Datentypen

Microsoft Excel-Zellenwertdatentypen	InfoSphere DataStage-Datentypen
Dateiname	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Dateipfad	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Dateigröße	<b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt <b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Datum der letzten Modifizierung	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Anmerkung:</b> Zeichenfolgeausdruck im Format jjjj-mm-tt <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Anmerkung:</b> Zeichenfolgeausdruck im Format jjjj-mm-tt <b>Datum- und Zeitdatentypen</b> Date Time Timestamp

Tabelle 15. Mapping zwischen Microsoft Excel-Dokumenteigenschaften und InfoSphere DataStage-Datentypen

Microsoft Excel-Dokumenteigenschaften	InfoSphere DataStage-Datentypen
Autoren	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Dokumentkommentare	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Datum der Inhaltserstellung	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Anmerkung:</b> Zeichenfolgeausdruck im Format jjjj-mm-tt <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Anmerkung:</b> Zeichenfolgeausdruck im Format jjjj-mm-tt <b>Datum- und Zeitdatentypen</b> Date Time Timestamp
Schlüsselwörter	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Überarbeitungsnummer	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Betreff	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Titel	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Unternehmen	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Kategorie	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar



Tabelle 15. Mapping zwischen Microsoft Excel-Dokumenteigenschaften und InfoSphere DataStage-Datentypen (Forts.)

Microsoft Excel-Dokumenteigenschaften	InfoSphere DataStage-Datentypen
Manager	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar

Tabelle 16. Mapping zwischen angepasster Microsoft Excel-Eigenschaft und InfoSphere DataStage-Datentypen

Mapping zwischen angepasster Microsoft Excel-Eigenschaft	InfoSphere DataStage-Datentypen
Text	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar
Datum	<b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Datum- und Zeitdatentypen</b> Date Time Timestamp
Zahl	<b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt <b>Anmerkung:</b> Wenn der Wert ein Integer ist. <b>Textdatentypen</b> Char VarChar LongVarChar <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Bruchzahldatentypen</b> Double Float Real <b>Dezimaldatentypen</b> Decimal Numeric
Boolesch	<b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt <b>Anmerkung:</b> Zuordnung: TRUE: 1, FALSE: 0 <b>Textdatentypen</b> Char VarChar LongVarChar <b>Anmerkung:</b> Zuordnung: TRUE: "true", FALSE: "false" <b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar <b>Anmerkung:</b> Zuordnung: TRUE: "true", FALSE: "false"

Tabelle 17. Mapping zwischen Microsoft Excel-Arbeitsblattinformationen und InfoSphere DataStage-Datentypen

Microsoft Excel-Arbeitsblattinformationen	InfoSphere DataStage-Datentypen
Arbeitsblattname	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p>
Header	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p> <p><b>Anmerkung:</b> Microsoft Excel unterstützt für Textdatentypen und Datentypen für Text in Landessprachen Sonderbefehle, die in Microsoft Excel-Kopfzeilen und -Fußzeilen durch einen einzelnen Buchstaben mit einem führenden Et-Zeichen (&amp;) dargestellt werden. Die Stage <b>Unstructured Data</b> konvertiert diese Buchstaben nicht und bewahrt sie lediglich im extrahierten Text auf. Weitere Informationen zu den Sonderbefehlen finden Sie unter <a href="http://msdn.microsoft.com/en-us/library/dd773041%28v=office.12%29.aspx">http://msdn.microsoft.com/en-us/library/dd773041%28v=office.12%29.aspx</a>.</p>
Fußzeile	<p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p> <p><b>Anmerkung:</b> Microsoft Excel unterstützt für Textdatentypen und Datentypen für Text in Landessprachen Sonderbefehle, die in Microsoft Excel-Kopfzeilen und -Fußzeilen durch einen einzelnen Buchstaben mit einem führenden Et-Zeichen (&amp;) dargestellt werden. Die Stage <b>Unstructured Data</b> konvertiert diese Buchstaben nicht und bewahrt sie lediglich im extrahierten Text auf. Weitere Informationen zu den Sonderbefehlen finden Sie unter <a href="http://msdn.microsoft.com/en-us/library/dd773041%28v=office.12%29.aspx">http://msdn.microsoft.com/en-us/library/dd773041%28v=office.12%29.aspx</a>.</p>

Tabelle 18. Mapping zwischen Microsoft Excel-Zeileninformationen und ihren entsprechenden InfoSphere DataStage-Datentypen

Microsoft Excel-Zeileninformationen	InfoSphere DataStage-Datentypen
Zeilennummer	<p><b>Integerdatentypen</b> BigInt Integer SmallInt TinyInt</p> <p><b>Textdatentypen</b> Char VarChar LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b> NChar NVarChar LongNVarChar</p>

Tabelle 18. Mapping zwischen Microsoft Excel-Zeileninformationen und ihren entsprechenden InfoSphere DataStage-Datentypen (Forts.)

Microsoft Excel-Zeileninformationen	InfoSphere DataStage-Datentypen
Ausgeblendet	<p><b>Integerdatentypen</b>                      BigInt Integer SmallInt TinyInt  <b>Anmerkung:</b> Zuordnung: TRUE: 1, FALSE: 0</p> <p><b>Textdatentypen</b>                      Char VarChar LongVarChar  <b>Anmerkung:</b> Zuordnung: TRUE: "true", FALSE: "false"</p> <p><b>Datentyp für Text in Landessprachen</b>                      NChar NVarChar LongNVarChar  <b>Anmerkung:</b> Zuordnung: TRUE: "true", FALSE: "false"</p>

## Datentypkonvertierungen von InfoSphere DataStage in Microsoft Excel

Bevor die Stage **Unstructured Data** aus Eingabelinks extrahierte Daten in Microsoft Excel-Dateien schreibt, werden die Daten in Microsoft Excel-Datentypen konvertiert.

In der folgenden Tabelle wird das Mapping zwischen InfoSphere DataStage-Datentypen und Microsoft Excel-Datentypen gezeigt.

**Anmerkung:** Die Stage **Unstructured Data** ordnet die Datentypkonvertierungen von InfoSphere DataStage-Datentypen Microsoft Excel-Dateien nur zu, wenn die Stage **Unstructured Data** Datensätze in die Microsoft Excel-Datenquelle schreibt.

Tabelle 19. Mapping zwischen Microsoft Excel-Zellenwertdatentypen und InfoSphere DataStage-Datentypen

InfoSphere DataStage-Datentyp	Microsoft Excel-Zellendatentyp
<p><b>Textdatentypen</b>                      Char, VarChar, LongVarChar</p> <p><b>Datentyp für Text in Landessprachen</b>                      NChar, NVarChar, LongNVarChar</p>	Zeichenfolge
<p><b>Integerdatentypen</b>                      BigInt, Integer, SmallInt, TinyInt</p> <p><b>Bruchzahldatentypen</b>                      Double, Float, Real</p> <p><b>Dezimaldatentypen</b>                      Decimal, Numeric</p>	Numerisch

Tabelle 19. Mapping zwischen Microsoft Excel-Zellenwertdatentypen und InfoSphere DataStage-Datentypen (Forts.)

InfoSphere DataStage-Datentyp	Microsoft Excel-Zellendatentyp
<b>Datum- und Zeitdatentypen</b> Date, Time, Timestamp	Numerisch <b>Anmerkung:</b> Dieser Hinweis gilt nur, wenn Sie Daten in vorhandene Microsoft Excel-Dateien schreiben. <ul style="list-style-type: none"> <li>Die Stage <b>Unstructured Data</b> legt das Format für Datum- und Zeitdatentypen nicht fest. Es wird empfohlen, das richtige Format festzulegen, um für Ihre Microsoft Excel-Datei Datum, Zeit oder Zeitmarke auszudrücken.</li> </ul>
<b>Binärdatentypen</b> Binary, VarBinary, LongVarBinary, Bit	Nicht unterstützt.

**Zugehörige Konzepte:**

„Daten in vorhandene Microsoft Excel-Dateien schreiben“ auf Seite 32  
 Sie können die Stage **Unstructured Data** in Jobs verwenden, die Daten in eine vorhandene Microsoft Excel-Datei schreiben. Sie können auch eine Microsoft Excel-Datei kopieren und Daten in die Kopie schreiben.

---

## Jobabbruchbedingungen in Microsoft Excel

In den folgenden Tabellen werden die verschiedenen Jobabbruchbedingungen in Microsoft Excel-Dateien beschrieben.

**Anmerkung:** Die folgende Bedingung gilt nur, wenn die Stage **Unstructured Data** Datensätze aus den Microsoft Excel-Dateien liest.

**Dateiname (Platzhalterzeichen nicht verwendet)**

Wenn der Dateiname ohne Platzhalterzeichen verwendet wird, können die folgenden Jobabbruchbedingungen auftreten:

Tabelle 20. Dateiname (Platzhalterzeichen nicht verwendet)

Bedingung	Ergebnis
Die angegebene Datei ist nicht vorhanden.	Der Job wird abgebrochen.
Der Benutzer verfügt nicht über die Berechtigung zum Lesen der angegebenen Datei.	Der Job wird abgebrochen.
Die angegebene Datei ist keine gültige Microsoft Excel-Datei.	Der Job wird abgebrochen.
Die Datei kann nicht mit dem angegebenen Kennwort geöffnet werden.	Der Job wird abgebrochen.

**Dateiname (Platzhalterzeichen verwendet)**

Wenn der Dateiname mit Platzhalterzeichen verwendet wird, können die folgenden Jobabbruchbedingungen auftreten:

Tabelle 21. Dateiname (Platzhalterzeichen verwendet)

Bedingung	Ergebnis
Es gibt keine Datei mit dem angegebenen Namen.	Der Job wird mit einer Warnung fortgesetzt (keine Ausgabezeile).

Tabelle 21. Dateiname (Platzhalterzeichen verwendet) (Forts.)

Der Benutzer verfügt nicht über die Berechtigung zum Lesen einer abgeglichenen Datei.	Der Job wird mit einer Warnung fortgesetzt, wenn die Eigenschaft <b>Fehleraktion</b> auf Überspringen gesetzt ist. Andernfalls wird der Job abgebrochen.
Die abgeglichene Datei ist keine gültige Microsoft Excel-Datei.	Der Job wird mit einer Warnung fortgesetzt, wenn die Eigenschaft <b>Fehleraktion</b> auf Überspringen gesetzt ist. Andernfalls wird der Job abgebrochen.
Die abgeglichene Datei kann nicht mit dem angegebenen Kennwort geöffnet werden.	Der Job wird mit einer Warnung fortgesetzt, wenn die Eigenschaft <b>Fehleraktion</b> auf Überspringen gesetzt ist. Andernfalls wird der Job abgebrochen.

### Arbeitsblattname

Die folgende Jobabbruchbedingung kann für Arbeitsblattnamen auftreten:

Tabelle 22. Arbeitsblattname

Bedingung	Ergebnis
Der Arbeitsblattname ist im Datenbereich angegeben, das Arbeitsblatt ist jedoch nicht vorhanden.	Der Job wird mit einer Warnung fortgesetzt, wenn die Eigenschaft <b>Fehleraktion</b> auf Überspringen gesetzt ist. Andernfalls wird der Job abgebrochen.

### Spaltenheader

Die folgende Jobabbruchbedingung kann für Spaltenheader auftreten:

Tabelle 23. Spaltenheader

Bedingung	Ergebnis
Die erste Zeile ist der Spaltenheader und sein Wert stimmt nicht mit der ersten Zeile des Datenbereichs der Schablone überein.	Der Job wird fortgesetzt und eine Informationsnachricht wird protokolliert.

### Datentyp

Die folgenden Jobabbruchbedingungen können für Datentypen auftreten:

Tabelle 24. Datentyp

Bedingung	Ergebnis
Der Datentyp wird nicht unterstützt, um den Microsoft Excel-Objektyp, der der DataStage-Spalte zugeordnet ist, zu extrahieren.	Der Job wird abgebrochen.
Der Datentyp wird unterstützt, um den Microsoft Excel-Objektyp, der der DataStage-Spalte zugeordnet ist, zu extrahieren, er stimmt jedoch nicht mit der Instanz überein. (Hierzu kommt es, wenn das Microsoft Excel-Objekt eine Zelle oder angepasste Eigenschaft ist.)	Der Job wird mit einer Warnung fortgesetzt, wenn die Eigenschaft <b>Fehleraktion</b> auf Überspringen gesetzt ist. Andernfalls wird der Job abgebrochen.

### Angepasste Eigenschaft

Die folgende Jobabbruchbedingung kann für eine angepasste Eigenschaft auftreten:

Tabelle 25. Angepasste Eigenschaft

Bedingung	Ergebnis
Die angegebene Eigenschaft ist nicht vorhanden.	Der Job wird mit einer Warnung fortgesetzt, wenn die Eigenschaft <b>Fehleraktion</b> auf Überspringen gesetzt ist. Andernfalls wird der Job abgebrochen.

---

## Kapitel 3. Fehlerbehebung

Mithilfe der Informationen in diesem Abschnitt können Sie Probleme mit der Stage **Unstructured Data** von InfoSphere DataStage verstehen, eingrenzen und beheben.

### Unten im Konfigurationsfenster angezeigte Nachrichten sind abgeschnitten

Wenn Sie feststellen, dass die unten im Konfigurationsfenster angezeigten Nachrichten abgeschnitten sind, bewegen Sie die Maus auf den Nachrichtenbereich, um die gesamte Nachricht anzuzeigen.

### Der erwartete Datenbereich der Schablone kann nicht abgerufen werden

Wenn Sie auf **Laden** klicken, ohne den Bereichsausdruck anzugeben, oder wenn Sie nur den Arbeitsblattnamen im Bereichsausdruck angeben, wird eine Liste von Datenbereichen der Schablone für Spaltenmapping angezeigt. Für einige Microsoft Excel-Dateien wird möglicherweise jedoch nicht der erforderliche Datenbereich der Schablone angezeigt.

Geben Sie den vollständigen Bereichsausdruck für den erwarteten Datenbereich der Schablone an und klicken Sie dann auf **Laden**, um eine vollständige Liste von Bereichsausdrücken für den Datenbereich der Schablone abzurufen. Wenn Sie die Startzelle für den erwarteten Datenbereich der Schablone angeben, können Sie den Datenbereich der Schablone ab der angegebenen Zelle abrufen.

### Eine Warnung wird angezeigt, wenn die Konfiguration für den Bereichsausdruck modifiziert wird

Das Feld **Bereichsausdruck** wird aktualisiert, wenn Sie einen der Datenbereiche der Schablone im Listenfeld auswählen und auf **Zuordnen** klicken. Sie können einen relevanten Bereichsausdruck anzeigen, der Ihrem ausgewählten Datenbereich der Schablone zugeordnet ist. Sie wollen den Bereichsausdruck möglicherweise jedoch ändern. Beispielsweise gibt die Stage **Unstructured Data** standardmäßig die Bereichsausdrucksinformationen einschließlich Arbeitsblattname zurück. Sie können jedoch die Konfiguration modifizieren, sodass der Bereichsausdruck ohne den Arbeitsblattnamen angezeigt wird. Wenn Sie auf **OK** klicken, um die modifizierte Konfiguration zu speichern, wird die folgende Warnung angezeigt:

Die Datenquelle wurde seit der Erstellung des Spaltenmappings geändert.  
Die Änderungen können einen Ausführungszeitfehler verursachen.  
Wollen Sie Ihre Änderungen speichern?

Wenn die Änderungen für das Spaltenmapping nicht erforderlich sind, klicken Sie auf **OK**, um die Konfiguration abzuschließen und zu bestätigen, dass der aktualisierte Bereichsausdruck mit dem Mapping konsistent ist.

### Zeitüberschreitung

Der folgende Fehler kann auftreten, wenn Sie versuchen, eine große Microsoft Excel-Datei in das Konfigurationsfenster zu laden, und wenn die Operation nicht im angegebenen Zeitraum für den CAS-Service abgeschlossen wurde:

Verarbeitung der Anforderung ist fehlgeschlagen:  
Empfangen der Antwort von der Behandlungsroutine ist fehlgeschlagen:  
Anforderungszeitlimit wurde überschritten.

Wenn dieser Fehler angezeigt wird, modifizieren Sie den für den Befehl **PropertyAdmin** angegebenen Zeitlimitwert.

Soll z. B. der Zeitlimitwert in 180 Sekunden geändert werden, geben Sie den folgenden Befehl in der Serviceschicht an:

```
InformationServer/ASBServer/bin/PropertyAdmin -set -key cas.agent.timeout  
-value 180
```

Dabei ist *InformationServer* das Installationsverzeichnis für InfoSphere Information Server.

## Nicht genügend Speicher beim Laden einer großen Microsoft Excel-Datei

Der folgende Fehler kann auftreten, wenn Sie versuchen, eine große Microsoft Excel-Datei in das Konfigurationsfenster zu laden:

Die Datei {0} kann nicht geladen werden, da nicht genügend Speicher verfügbar ist. Die Datei ist möglicherweise zu groß (sie ist {1} Byte groß).  
Geben Sie eine kleinere Datei als Schablone an.

Dabei gibt {0} den vom Benutzer angegebenen Dateinamen und {1} die Größe der angegebenen Datei an.

Wenn Sie eine große Microsoft Excel-Datei laden, wird viel Java-Heapspeicher belegt. Daher wird der von der Stage **Unstructured Data** verwendete Java-Prozess Connection Access Service möglicherweise blockiert. Wird der Prozess blockiert, wird die Anforderung abgebrochen und die Fehlernachricht angezeigt.

Führen Sie einen der folgenden Schritte aus, um den Fehler zu vermeiden:

- Geben Sie eine kleinere Microsoft Excel-Datei an.
- Erstellen Sie eine kleinere Datei, indem Sie die ursprüngliche Microsoft Excel-Datei kopieren. Löschen Sie die Zeilen und Arbeitsblätter, die nicht für Spaltenmapping verwendet werden.

## Fehler aufgrund der Änderung der Größe des Heapspeichers von ASBAgent

Der folgende ähnlich lautende Fehler kann auftreten, wenn Sie versuchen, eine große Microsoft Excel-Datei zu laden:

Die Größe des Heapspeichers von ASBAgent konnte nicht geändert werden.

Wenn Sie die Protokollansicht für die Kategorie **ISF-CAS-NATIVE** aktivieren, wird möglicherweise die folgende Nachricht protokolliert:

Warnung wurde von der nativen (C-)Schicht empfangen:  
Die Datei {0} kann nicht geladen werden, da nicht genügend Speicher verfügbar ist. Die Datei ist möglicherweise zu groß (sie ist {1} Byte groß).  
Die maximale Größe des JVM-Heapspeichers ist {2}.  
Die Größe des verarbeiteten Heapspeichers ist {3}. {4}

Dabei gilt Folgendes:

- {0} gibt den vom Benutzer angegebenen Dateinamen an.



- {1} gibt die Größe der angegebenen Datei an.
- {2} gibt die maximale Größe des Heapspeichers für JVM an.
- {3} gibt die zurzeit belegte Größe des Heapspeichers an.

Erhöhen Sie die Java-Heapspeicherkapazität von ASBAgent, um das Problem zu umgehen.

## Ausnahmebedingungsfehler bei Java-Ausführung

Der folgende fatale Fehler kann auftreten, wenn eine große Microsoft Excel-Datei verarbeitet wird:

```
Unstructured_Data_0,0: Java runtime exception occurred: java.lang.OutOfMemoryError
(java.util.Arrays::copyOfRange, file Arrays.java, line 4,138)
```

Der Fehler tritt auf, weil die Java-Heapspeicherkapazität nicht ausreicht.

Erhöhen Sie die verfügbare Größe des Java-Heapspeichers über die Umgebungsvariable `CC_UNST_JAVA_HEAP`, um das Problem zu umgehen. Der Wert der Umgebungsvariablen ist der ganzzahlige Wert der Größe des Java-Heapspeichers in MB. Soll die Größe des Java-Heapspeichers z. B. auf 512 MB gesetzt werden, legen Sie `CC_UNST_JAVA_HEAP=512` fest. Die Standardgröße des Heapspeichers beträgt 256 MB.

Microsoft Excel-Datei extrahieren	
Microsoft Excel 97-2003-Datei (.xls)	6-fache Dateigröße
Kennwortgeschützte Microsoft Excel 2007-2010-Datei (.xlsx)	30-fache Dateigröße
Ungeschützte Microsoft Excel 2007-2010-Datei (.xlsx)	Standardgröße des Heapspeichers
Microsoft Excel-Datei modifizieren	300-fache Größe der sich ergebenden Microsoft Excel-Datei
Microsoft Excel-Datei erstellen	Standardgröße des Heapspeichers

**Anmerkung:** Die tatsächliche erforderliche Größe des Heapspeichers hängt von dem in der Microsoft Excel-Datei enthaltenen Datentyp sowie von der Plattform ab und kann die in der obigen Tabelle angegebene Größe überschreiten.

## Ordnungsgemäßer Bereich für Angabe kann nicht gefunden werden

- Sie können möglicherweise keinen ordnungsgemäßen Bereichsausdruck für die Angabe finden, wenn Sie in der folgenden Situation Datenbereiche aus mehreren Arbeitsblättern oder mehreren Dateien extrahieren wollen.
  - Wenn jeder Datenbereich an einer anderen Position anfängt.
  - Wenn das Ende von Datenbereichen nicht die letzte Zeile in den Arbeitsblättern ist.

Definieren Sie in solchen Fällen den Namen von Datenbereichen in den Microsoft Excel-Quelldateien und geben Sie den Namen als Bereichsausdruck an.

## In älterer Version erstellter Job kann nicht kompiliert werden

Der folgende Fehler kann beim Kompilieren eines in InfoSphere DataStage Version 9.1 erstellten Jobs auftreten.

**Anmerkung:** Für die Eigenschaft **Bereichsausdruck** wurde kein Wert angegeben. Klicken Sie im Stage-Editor auf **Konfigurieren** und geben Sie dann einen Wert für die Eigenschaft an. Wenn dieser Job in InfoSphere DataStage Version 9.1 erstellt wurde, müssen Sie das Konfigurationsfenster öffnen und den Job speichern.

Dieser Fehler tritt auf, wenn Sie eine der folgenden Aktionen ausführen:

1. Öffnen Sie den Stage-Editor.
2. Klicken Sie im Stage-Editor auf **OK**, klicken Sie jedoch nicht auf **Konfigurieren**.

oder

1. Öffnen Sie den Stage-Editor.
2. Klicken Sie auf **Konfigurieren**.
3. Klicken Sie im Konfigurationsfenster auf **Abbrechen**.
4. Klicken Sie im Stage-Editor auf **OK**.

**Anmerkung:** Sie können einen in InfoSphere DataStage Version 9.1 erstellten Job nur ohne Aktionen ausführen, wenn Sie den Stage-Editor öffnen und im Stage-Editor auf **OK** klicken.

Führen Sie die folgenden Schritte aus, um diesen Fehler zu beheben.

1. Öffnen Sie den Stage-Editor.
2. Klicken Sie auf **Konfigurieren**.
3. Klicken Sie im Konfigurationsfenster auf **OK**.
4. Klicken Sie im Stage-Editor auf **OK**.

---

## Kapitel 4. Umgebungsvariablen: Stage 'Unstructured Data'

Die Stage **Unstructured Data** verwendet die folgenden Umgebungsvariablen.

---

### CC\_JNI\_EXT\_DIRS

Sie können diese Umgebungsvariable setzen, um dem Klassenpfad der Systemeigenschaft `java.ext.dirs` ein Präfix hinzuzufügen.

Wenn der Wert dieser Umgebungsvariablen festgelegt ist, wird dem Klassenpfad der Systemeigenschaft `java.ext.dirs` ein Präfix hinzugefügt.

---

### CC\_JVM\_OPTIONS

Sie können diese Umgebungsvariable setzen, um die bei der Ausführung eines Jobs verwendeten JVM-Argumente anzugeben.

Wenn diese Variable angegeben ist, hat sie Vorrang vor dem Wert der JVM-Standardargumente für die Java-basierten Connector. Wenn Sie **CC\_JVM\_OPTIONS** beispielsweise als `-Xmx512M` festlegen, wird die maximale Größe des Heapspeichers auf 512 MB gesetzt, wenn JVM-Instanzen für den Connector erstellt werden.

---

### CC\_JVM\_OVERRIDE\_OPTIONS

Sie können diese Umgebungsvariable setzen, um die JVM-Optionen für den Conductorknoten zu überschreiben, sodass ein möglicher Konflikt verhindert oder behoben werden kann.

In einem Conductorknoten in einem parallelen Job werden Java-Connector für die Schemaabstimmung initialisiert. Daher werden alle Java-Connector in einem Job in derselben Java Virtual Machine initialisiert. In einem Job können mehrere Stages in Java entwickelt werden. Jede dieser Stages kann JVM-Optionen definieren, beispielsweise Klassenpfad, Systemeigenschaft und Größe des Heapspeichers. Wenn zwei Stages in derselben physischen Java Virtual Machine ausgeführt werden, können die JVM-Optionen Konflikte untereinander verursachen.

---

### CC\_IGNORE\_TIME\_LENGTH\_AND\_SCALE

Sie können diese Umgebungsvariable setzen, um das Verhalten des Connectors im Entwicklungsbereich für parallele Jobs zu ändern.

Wenn diese Umgebungsvariable auf 1 gesetzt ist, ignoriert der mit der Parallelverarbeitungseingabe ausgeführte Connector die angegebene Länge und die angegebenen Nachkommastellen für die Zeitmarkenspalte. Wenn der Wert dieser Umgebungsvariablen beispielsweise nicht gesetzt ist und die Länge der Zeitmarkenspalte 26 und der Nachkommastellenwert 6 ist, geht der Connector im Entwicklungsbereich für parallele Jobs davon aus, dass die Zeitmarke eine Mikrosekundenauflösung hat. Wenn der Wert dieser Umgebungsvariablen auf 1 gesetzt ist, geht der Connector im Entwicklungsbereich für parallele Jobs nicht davon aus, dass die Zeitmarke eine Mikrosekundenauflösung hat, selbst wenn die Länge der Zeitmarkenspalte 26 und der Nachkommastellenwert 6 ist, außer die erweiterte Eigenschaft für Mikrosekunden ist gesetzt.

---

## CC\_MSG\_LEVEL

Sie können diese Umgebungsvariable setzen, um die Mindestbewertung der Nachrichten anzugeben, die der Connector in der Protokolldatei ausgibt.

Beim Standardwert 3 werden Informationsnachrichten und Nachrichten mit höherer Bewertung in der Protokolldatei ausgegeben.

Die folgende Liste enthält die gültigen Werte:

- 1 - Trace
- 2 - Debug
- 3 - Information
- 4 - Warnung
- 5 - Fehler
- 6 - Fatal

---

## CC\_UNST\_JAVA\_HEAP

Sie können diese Umgebungsvariable setzen, um die Größe des Java-Heapspeichers zu steuern, der von der Stage **Unstructured Data** verwendet werden kann.

Setzen Sie diese Variable auf einen ganzzahligen Wert, der die Größe des Java-Heapspeichers in MB darstellt. Soll die Größe des Java-Heapspeichers beispielsweise auf 512 MB gesetzt werden, setzen Sie **CC\_UNST\_JAVA\_HEAP** auf 512. Die Standardgröße des Java-Heapspeichers ist 256 MB.

---

## Anhang A. Eingabehilfen in den Produkten

Sie erhalten Informationen zum Status der Eingabehilfen in IBM Produkten.

Die Produktmodule und Benutzerschnittstellen von IBM InfoSphere Information Server sind nicht uneingeschränkt für behindertengerechte Bedienung geeignet.

Informationen zum Status der Eingabehilfen in IBM Produkten finden Sie unter [http://www.ibm.com/able/product\\_accessibility/index.html](http://www.ibm.com/able/product_accessibility/index.html).

### **Dokumentation im behindertengerechten Format**

Dokumentation im behindertengerechten Format für Produkt steht im IBM Knowledge Center zur Verfügung. Im IBM Knowledge Center wird zur Darstellung der Dokumentation das Format XHTML 1.0 verwendet, das mit den meisten Web-Browsern geöffnet werden kann. Da das IBM Knowledge Center XHTML verwendet, können Sie in Ihrem Browser Anzeigevorgaben festlegen. Darüber hinaus ist der Einsatz von Sprachausgabeprogrammen und anderen Unterstützungseinrichtungen für den Zugriff auf die Dokumentation möglich.

Die im IBM Knowledge Center zur Verfügung stehende Dokumentation wird auch in Form von PDF-Dateien bereitgestellt, die nicht uneingeschränkt für behindertengerechte Bedienung geeignet sind.

### **IBM und Eingabehilfen**

Weitere Informationen zum Engagement von IBM hinsichtlich der Eingabehilfen finden Sie im IBM Human Ability and Accessibility Center.



---

## Anhang B. Befehlszeilensyntax lesen

In dieser Dokumentation werden für die Befehlszeilensyntax Sonderzeichen eingesetzt.

Die folgenden Sonderzeichen werden zur Darstellung der Befehlszeilensyntax verwendet:

- [ ] Gibt ein optionales Argument an. Argumente, die nicht in eckige Klammern eingeschlossen sind, sind erforderlich.
- ... Gibt an, dass für das vorherige Argument mehrere Werte angegeben werden können.
- | Gibt an, dass sich Informationen gegenseitig ausschließen. Sie können entweder das Argument links des Trennzeichens oder das Argument rechts des Trennzeichens verwenden. Es ist nicht möglich, beide Argumente im selben Aufruf des Befehls zu verwenden.
- { } Begrenzt eine Gruppe sich gegenseitig ausschließender Argumente, wobei eines der Argumente erforderlich ist. Wenn die Argumente optional sind, sind sie in eckige Klammern eingeschlossen.

### Anmerkung:

- Die maximal zulässige Anzahl von Zeichen in einem Argument ist 256.
- Argumentwerte, die eingebettete Leerzeichen enthalten, müssen in einfache oder doppelte Anführungszeichen eingeschlossen werden.

Beispiel:

```
wsetsrc[-S Server] [-l Bezeichnung] [-n Name] Quelle
```

Das Argument *Quelle* ist das einzige erforderliche Argument für den Befehl **wsetsrc**. Die eckigen Klammern um die anderen Argumente geben an, dass diese Argumente optional sind.

```
wlsac [-l | -f Format] [Schlüssel...] Profil
```

In diesem Beispiel sind die Argumente *-l* und *-f Format* optional und schließen sich gegenseitig aus. Das Argument *Profil* ist erforderlich. Das Argument *Schlüssel* ist optional. Das Auslassungszeichen (...), das auf das Argument *Schlüssel* folgt, gibt an, dass mehrere Schlüsselnamen angegeben werden können.

```
wrb -import {Regelpack | Regelsatz}...
```

In diesem Beispiel schließen sich die Argumente *Regelpack* und *Regelsatz* gegenseitig aus, eines dieser Argumente muss jedoch angegeben werden. Außerdem zeigen die Auslassungspunkte (...) an, dass Sie mehrere Regelpacks oder Regelsätze angeben können.





---

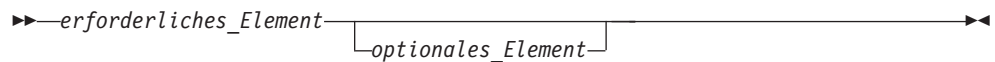
## Anhang C. Syntaxdiagramme lesen

Die folgenden Regeln gelten für die Syntaxdiagramme, die in den vorliegenden Informationen verwendet werden:

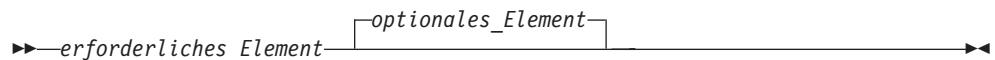
- Lesen Sie die Syntaxdiagramme von links nach rechts und von oben nach unten und folgen Sie dabei der Linie. Die folgenden Konventionen werden verwendet:
  - Das Symbol >>--- kennzeichnet den Anfang eines Syntaxdiagramms.
  - Das Symbol ---> gibt an, dass ein Syntaxdiagramm in der nächsten Zeile fortgesetzt wird.
  - Das Symbol >--- gibt an, dass ein Syntaxdiagramm von der vorhergehenden Zeile fortgesetzt wird.
  - Das Symbol --->< kennzeichnet das Ende eines Syntaxdiagramms.
- Erforderliche Elemente werden auf der horizontalen Linie (dem Hauptpfad) angezeigt.



- Optionale Elemente werden unterhalb des Hauptpfads angezeigt.

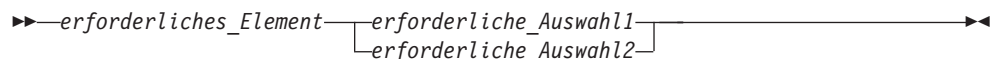


Wenn ein optionales Element oberhalb des Hauptpfads angezeigt wird, hat es keine Auswirkungen auf die Ausführung des Syntaxelements und wird nur zur besseren Lesbarkeit verwendet.

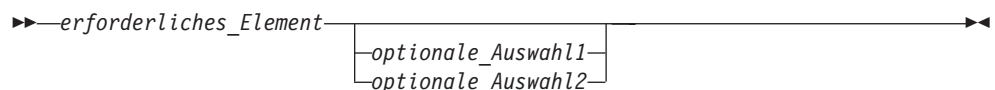


- Wenn Sie aus zwei oder mehr Elementen auswählen können, werden diese vertikal in einem Stapel angezeigt.

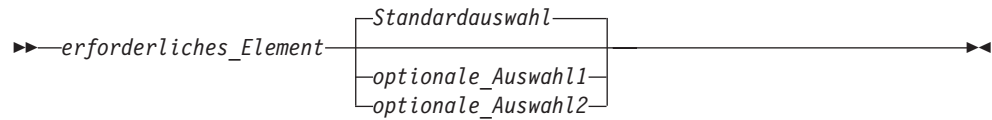
Wenn eines dieser Elemente ausgewählt werden muss, steht ein Element des Stapels auf dem Hauptpfad.



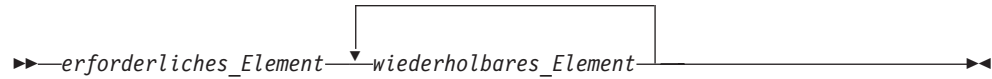
Wenn die Auswahl eines der Elemente optional ist, wird der gesamte Stapel unterhalb des Hauptpfads angezeigt.



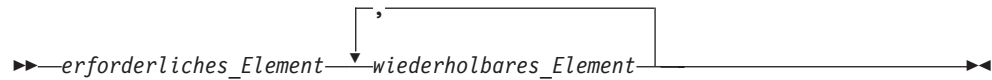
Wenn es sich bei einem der Elemente um den Standardwert handelt, wird es oberhalb des Hauptpfads angezeigt, während die übrigen Auswahlmöglichkeiten unterhalb des Hauptpfads stehen.



- Ein Pfeil zurück nach links, der oberhalb des Hauptpfads angezeigt wird, gibt ein wiederholbares Element an.

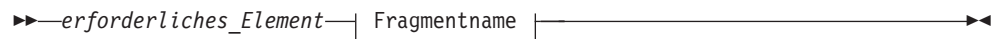


Enthält der Wiederholungspfeil ein Komma, müssen wiederholte Elemente durch ein Komma getrennt werden.

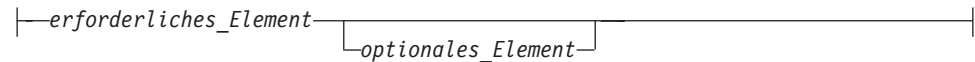


Ein Wiederholungspfeil oberhalb eines Stapels gibt an, dass die Elemente des Stapels wiederholt werden können.

- In manchen Fällen muss ein Diagramm in Fragmente aufgeteilt werden. Das Syntaxfragment wird separat vom eigentlichen Syntaxdiagramm angezeigt, der Inhalt des Fragments muss jedoch so gelesen werden, als sei er Teil des Hauptpfads des Diagramms.



### Fragmentname:



- Schlüsselwörter und die zugehörigen Mindestabkürzungen, falls zutreffend, werden in Großbuchstaben angezeigt. Sie müssen exakt in der angezeigten Schreibweise eingegeben werden.
- Variablen werden kursiv in gemischter Groß-/Kleinschreibung dargestellt (zum Beispiel **Name\_der\_Spalte**). Sie stellen Namen oder Werte dar, die der Benutzer angibt.
- Schlüsselwörter und Parameter müssen durch mindestens ein Leerzeichen voneinander getrennt werden, falls im Diagramm keine entsprechende Interpunktion als Trennzeichen angegeben ist.
- Geben Sie Interpunktionszeichen, runde Klammern, arithmetische Operatoren und andere Symbole genau so ein wie im Diagramm dargestellt.
- Fußnoten werden durch eine Zahl in runden Klammern angegeben, zum Beispiel (1).

---

## Anhang D. Kontaktaufnahme mit IBM

Sie können sich an IBM wenden, um Unterstützung, Informationen zu Software-Services, Produktinformationen sowie allgemeine Informationen zu erhalten. Darüber hinaus können Sie Feedback zu den Produkten und zur Dokumentation an IBM senden.

In der folgenden Tabelle sind Ressourcen für die Kundenunterstützung, für Software-Services, für Schulungen sowie für Produkt- und Lösungsinformationen aufgeführt.

*Tabelle 26. IBM Ressourcen*

<b>Ressource</b>	<b>Beschreibung und Position</b>
IBM Support-Portal	Sie können die Unterstützungsinformationen je nach Bedarf anpassen, indem Sie die Produkte und Themen, die für Sie von Interesse sind, unter <a href="http://www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server">www.ibm.com/support/entry/portal/Software/Information_Management/InfoSphere_Information_Server</a> auswählen.
Software-Services	Informationen zu Software-, IT- und Unternehmensberatungsservices erhalten Sie auf der Site 'Lösungen' unter <a href="http://www.ibm.com/businesssolutions/de">www.ibm.com/businesssolutions/de</a> .
Meine IBM	Auf der Site 'Meine IBM' unter <a href="http://www.ibm.com/account/de/de/">www.ibm.com/account/de/de/</a> können Sie ein Konto einrichten und so Links auf IBM Websites und Informationen Ihren speziellen Anforderungen an die technische Unterstützung entsprechend verwalten.
Schulung und Zertifizierung	Informationen zu technischen Schulungs- und Ausbildungsservices, mit deren Hilfe Einzelpersonen sowie Mitarbeiter von Unternehmen und öffentlichen Organisationen IT-Kenntnisse erwerben, optimieren und auf dem neuesten Stand halten können, finden Sie unter <a href="http://www.ibm.com/training">http://www.ibm.com/training</a> .
IBM Ansprechpartner	Einen IBM Ansprechpartner, bei dem Sie Informationen zu Lösungen erhalten, finden Sie unter <a href="http://www.ibm.com/connect/ibm/us/en/">www.ibm.com/connect/ibm/us/en/</a> bzw. <a href="http://www.ibm.com/contact/de/de/">www.ibm.com/contact/de/de/</a> .



---

## Anhang E. Auf Produktdokumentation zugreifen

Die Dokumentation wird in einer Vielzahl von Formaten bereitgestellt: online im IBM Knowledge Center, optional in einem lokal installierten Information Center sowie in Form von Handbüchern im PDF-Format. Sie können direkt über die Produktclientschnittstelle auf die Onlinehilfe oder die lokal installierte Hilfe zugreifen.

IBM Knowledge Center ist die beste Methode, um aktuelle Informationen zu InfoSphere Information Server zu suchen. Das IBM Knowledge Center enthält Hilfe für die meisten Produktschnittstellen sowie die gesamte Dokumentation für alle Produktmodule der Suite. Sie können das IBM Knowledge Center über das installierte Produkt oder über einen Web-Browser öffnen.

### Auf das IBM Knowledge Center zugreifen

Es gibt verschiedene Möglichkeiten, auf die Onlinedokumentation zuzugreifen:

- Klicken Sie auf den Link **Hilfe** rechts oben in der Clientschnittstelle.
- Drücken Sie die Taste F1. Mit der Taste F1 wird normalerweise das Thema aufgerufen, das eine Beschreibung des aktuellen Kontexts der Clientschnittstelle enthält.

**Anmerkung:** In Web-Clients kann die Taste F1 nicht verwendet werden.

- Geben Sie die Adresse in einem Web-Browser ein, beispielsweise, wenn Sie nicht am Produkt angemeldet sind.

Geben Sie die folgende Adresse ein, um auf alle Versionen der Dokumentation zu InfoSphere Information Server zuzugreifen:

<http://www.ibm.com/support/knowledgecenter/SSZJPZ/>

Wenn Sie auf ein bestimmtes Thema zugreifen wollen, geben Sie die Versionsnummer zusammen mit der Produkt-ID, dem Namen des Dokumentations-Plugins und dem Themenpfad in der URL an. Die URL für Version 11.3 dieses Themas lautet beispielsweise wie folgt, wobei das Symbol  $\Rightarrow$  eine Zeilenfortsetzung angibt:

[http://www.ibm.com/support/knowledgecenter/SSZJPZ\\_11.3.0/com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html](http://www.ibm.com/support/knowledgecenter/SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html)

#### **Tipp:**

Für das Knowledge Center gibt es auch eine Kurz-URL:

<http://ibm.biz/knowctr>

Zur Angabe einer Kurz-URL zu einer bestimmten Produktseite oder Version oder zu einem bestimmten Thema geben Sie zwischen der Kurz-URL und der Produkt-ID ein Hashzeichen (#) an. Die Kurz-URL für die gesamte Dokumentation zu InfoSphere Information Server lautet beispielsweise wie folgt:

<http://ibm.biz/knowctr#SSZJPZ/>

Die folgende URL ist die leicht verkürzte URL zum obigen Thema (das Symbol  $\Rightarrow$  gibt eine Zeilenfortsetzung an):

[http://ibm.biz/knowctr#SSZJPZ\\_11.3.0/com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html](http://ibm.biz/knowctr#SSZJPZ_11.3.0/com.ibm.swg.im.iis.common.doc/common/accessingiidoc.html)

## Hilfelinks so ändern, dass sie auf lokal installierte Dokumentation verweisen

Das IBM Knowledge Center enthält die aktuelle Version der Dokumentation. Sie können jedoch eine lokale Version der Dokumentation in Form eines Information Center installieren und Ihre Hilfelinks so konfigurieren, dass sie auf dieses Information Center verweisen. Ein lokales Information Center ist sinnvoll, wenn Ihr Unternehmen keinen Zugriff auf das Internet bereitstellt.

Befolgen Sie die Installationsanweisungen im Installationspaket für das Information Center, um das Information Center auf einem Computer Ihrer Wahl zu installieren. Nach der Installation und dem Start des Information Center können Sie die von der Taste F1 der Produkthilfe und den Hilfelinks verwendete Speicherposition für die Dokumentation mithilfe des Befehls **iisAdmin** auf der Serviceschicht ändern (das Symbol  $\Rightarrow$  gibt eine Zeilenfortsetzung an):

### Windows

```
IS-Installationspfad\ASBServer\bin\iisAdmin.bat -set -key  $\Rightarrow$   
com.ibm.iis.infocenter.url -value http://<Host>:<Port>/help/topic/
```

### AIX Linux

```
IS-Installationspfad/ASBServer/bin/iisAdmin.sh -set -key  $\Rightarrow$   
com.ibm.iis.infocenter.url -value http://<Host>:<Port>/help/topic/
```

Dabei ist <Host> der Name des Computers, auf dem das Information Center installiert ist, und <Port> ist die Portnummer für das Information Center. Die Standardportnummer lautet 8888. Für einen Computer mit dem Namen `server1.example.com`, der den Standardport verwendet, lautet der URL-Wert beispielsweise `http://server1.example.com:8888/help/topic/`.

## PDF- und Hardcopy-Dokumentation abrufen

- Die PDF-Versionen der Handbücher sind online verfügbar und können über <https://www.ibm.com/support/docview.wss?uid=swg27008803&wv=1> aufgerufen werden.
- Sie können IBM Veröffentlichungen auch im Hardcopy-Format online oder über den zuständigen IBM Ansprechpartner bestellen. Wenn Sie Veröffentlichungen online bestellen möchten, rufen Sie das IBM Publications Center unter <http://www.ibm.com/e-business/linkweb/publications/servlet/pbi.wss> auf.

---

## Anhang F. Feedback zur Produktdokumentation geben

Sie können hilfreiches Feedback zur IBM Dokumentation geben.

Ihr Feedback hilft IBM, Informationen von hoher Qualität bereitzustellen. Sie haben die folgenden Möglichkeiten, Kommentare zu senden:

- Wenn Sie einen Kommentar zu einem Thema im IBM Knowledge Center abgeben wollen, das sich auf der IBM Website befindet, melden Sie sich an und klicken Sie auf die Schaltfläche **Kommentar hinzufügen** am Ende des Themas. Auf diese Weise eingegebene Kommentare werden für alle Benutzer angezeigt.
- Wenn Sie einen Kommentar zu einem Thema im IBM Knowledge Center an IBM senden wollen, der nicht für alle Benutzer angezeigt werden soll, melden Sie sich an und klicken auf den Link **Feedback** unten im IBM Knowledge Center.
- Senden Sie Ihre Kommentare mithilfe des Onlineformulars für Leserkommentare unter [www.ibm.com/software/awdtools/rcf/](http://www.ibm.com/software/awdtools/rcf/).
- Senden Sie Ihre Kommentare per E-Mail an [comments@us.ibm.com](mailto:comments@us.ibm.com). Geben Sie dabei den Namen des Produkts, die Versionsnummer des Produkts sowie den Namen und die Teilenummer der Informationen (falls vorhanden) an. Wenn sich Ihr Kommentar auf einen bestimmten Text bezieht, geben Sie die Position des Texts an (z. B. eine Überschrift, eine Tabellenummer oder eine Seitenzahl).





---

## Bemerkungen und Marken

Die vorliegenden Informationen wurden für Produkte und Services entwickelt, die auf dem deutschen Markt angeboten werden. IBM stellt dieses Material möglicherweise auch in anderen Sprachen zur Verfügung. Für den Zugriff auf das Material in einer anderen Sprache ist eine Kopie des Produkts oder der Produktversion in der jeweiligen Sprache erforderlich.

### Bemerkungen

Möglicherweise bietet IBM die in dieser Dokumentation beschriebenen Produkte, Services oder Funktionen in anderen Ländern nicht an. Informationen über die gegenwärtig im jeweiligen Land verfügbaren Produkte und Services sind beim zuständigen IBM Ansprechpartner erhältlich. Hinweise auf IBM Lizenzprogramme oder andere IBM Produkte bedeuten nicht, dass nur Programme, Produkte oder Services von IBM verwendet werden können. Anstelle der IBM Produkte, Programme oder Services können auch andere, ihnen äquivalente Produkte, Programme oder Services verwendet werden, solange diese keine gewerblichen oder anderen Schutzrechte von IBM verletzen. Die Verantwortung für den Betrieb von Produkten, Programmen und Services anderer Anbieter liegt beim Kunden.

Für in diesem Handbuch beschriebene Erzeugnisse und Verfahren kann es IBM Patente oder Patentanmeldungen geben. Mit der Auslieferung dieses Handbuchs ist keine Lizenzierung dieser Patente verbunden. Lizenzanforderungen sind schriftlich an folgende Adresse zu richten (Anfragen an diese Adresse müssen auf Englisch formuliert werden):

IBM Director of Licensing  
IBM Europe, Middle East & Africa  
Tour Descartes  
2, avenue Gambetta  
92066 Paris La Defense  
France

Trotz sorgfältiger Bearbeitung können technische Ungenauigkeiten oder Druckfehler in dieser Veröffentlichung nicht ausgeschlossen werden. Die hier enthaltenen Informationen werden in regelmäßigen Zeitabständen aktualisiert und als Neuausgabe veröffentlicht. IBM kann ohne weitere Mitteilung jederzeit Verbesserungen und/oder Änderungen an den in dieser Veröffentlichung beschriebenen Produkten und/oder Programmen vornehmen.

Verweise in diesen Informationen auf Websites anderer Anbieter werden lediglich als Service für den Kunden bereitgestellt und stellen keinerlei Billigung des Inhalts dieser Websites dar. Das über diese Websites verfügbare Material ist nicht Bestandteil des Materials für dieses IBM Produkt. Die Verwendung dieser Websites geschieht auf eigene Verantwortung.

Werden an IBM Informationen eingesandt, können diese beliebig verwendet werden, ohne dass eine Verpflichtung gegenüber dem Einsender entsteht.

Lizenznehmer des Programms, die Informationen zu diesem Produkt wünschen mit der Zielsetzung: (i) den Austausch von Informationen zwischen unabhängig voneinander erstellten Programmen und anderen Programmen (einschließlich des

vorliegenden Programms) sowie (ii) die gemeinsame Nutzung der ausgetauschten Informationen zu ermöglichen, wenden sich an folgende Adresse:

IBM Corporation  
J46A/G4  
555 Bailey Avenue  
San Jose, CA 95141-1003 USA

Die Bereitstellung dieser Informationen kann unter Umständen von bestimmten Bedingungen - in einigen Fällen auch von der Zahlung einer Gebühr - abhängig sein.

Die Lieferung des in diesem Dokument beschriebenen Lizenzprogramms sowie des zugehörigen Lizenzmaterials erfolgt auf der Basis der IBM Rahmenvereinbarung bzw. der Allgemeinen Geschäftsbedingungen von IBM, der IBM Internationalen Nutzungsbedingungen für Programmpakete oder einer äquivalenten Vereinbarung.

Alle in diesem Dokument enthaltenen Leistungsdaten stammen aus einer kontrollierten Umgebung. Die Ergebnisse, die in anderen Betriebsumgebungen erzielt werden, können daher erheblich von den hier erzielten Ergebnissen abweichen. Einige Daten stammen möglicherweise von Systemen, deren Entwicklung noch nicht abgeschlossen ist. Eine Gewährleistung, dass diese Daten auch in allgemein verfügbaren Systemen erzielt werden, kann nicht gegeben werden. Darüber hinaus wurden einige Daten unter Umständen durch Extrapolation berechnet. Die tatsächlichen Ergebnisse können davon abweichen. Benutzer dieses Dokuments sollten die entsprechenden Daten in ihrer spezifischen Umgebung prüfen.

Alle Informationen zu Produkten anderer Anbieter stammen von den Anbietern der aufgeführten Produkte, deren veröffentlichten Ankündigungen oder anderen allgemein verfügbaren Quellen. IBM hat diese Produkte nicht getestet und kann daher keine Aussagen zu Leistung, Kompatibilität oder anderen Merkmalen machen. Fragen zu den Leistungsmerkmalen von Produkten anderer Anbieter sind an den jeweiligen Anbieter zu richten.

Aussagen über Pläne und Absichten von IBM unterliegen Änderungen oder können zurückgenommen werden und repräsentieren nur die Ziele von IBM.

Diese Veröffentlichung dient nur zu Planungszwecken. Die in dieser Veröffentlichung enthaltenen Informationen können geändert werden, bevor die beschriebenen Produkte verfügbar sind.

Diese Veröffentlichung enthält Beispiele für Daten und Berichte des alltäglichen Geschäftsablaufs. Sie sollen nur die Funktionen des Lizenzprogramms illustrieren und können Namen von Personen, Firmen, Marken oder Produkten enthalten. Alle diese Namen sind frei erfunden; Ähnlichkeiten mit tatsächlichen Namen und Adressen sind rein zufällig.

#### COPYRIGHTLIZENZ:

Diese Veröffentlichung enthält Beispielanwendungsprogramme, die in Quellsprache geschrieben sind und Programmier Techniken in verschiedenen Betriebsumgebungen veranschaulichen. Sie dürfen diese Beispielprogramme kostenlos kopieren, ändern und verteilen, wenn dies zu dem Zweck geschieht, Anwendungsprogramme zu entwickeln, zu verwenden, zu vermarkten oder zu verteilen, die mit der Anwendungsprogrammierschnittstelle für die Betriebsumgebung konform sind, für die diese Beispielprogramme geschrieben werden. Diese Beispiele wurden nicht

unter allen denkbaren Bedingungen getestet. Daher kann IBM die Zuverlässigkeit, Wartungsfreundlichkeit oder Funktion dieser Programme weder zusagen noch gewährleisten. Die Beispielprogramme werden ohne Wartung (auf "as-is"-Basis) und ohne jegliche Gewährleistung zur Verfügung gestellt. IBM übernimmt keine Haftung für Schäden, die durch die Verwendung der Beispielprogramme entstehen.

Kopien oder Teile der Beispielprogramme bzw. daraus abgeleiteter Code müssen folgenden Copyrightvermerk beinhalten:

© (Name Ihrer Firma) (Jahr). Teile des vorliegenden Codes wurden aus Beispielprogrammen der IBM Corporation abgeleitet. © Copyright IBM Corp. \_Jahr/Jahre angeben\_. Alle Rechte vorbehalten.

## Hinweise zur Datenschutzrichtlinie

IBM Softwareprodukte, einschließlich Software as a Service-Lösungen ("Softwareangebote"), können Cookies oder andere Technologien verwenden, um Informationen zur Produktnutzung zu erfassen, die Endbenutzererfahrung zu verbessern und Interaktionen mit dem Endbenutzer anzupassen oder zu anderen Zwecken. In vielen Fällen werden von den Softwareangeboten keine personenbezogenen Daten erfasst. Einige der IBM Softwareangebote können Sie jedoch bei der Erfassung personenbezogener Daten unterstützen. Wenn dieses Softwareangebot Cookies zur Erfassung personenbezogener Daten verwendet, sind nachfolgend nähere Informationen über die Verwendung von Cookies durch dieses Angebot zu finden.

Abhängig von den implementierten Konfigurationen kann dieses Softwareangebot Sitzungscookies oder persistente Cookies verwenden. Wird ein Produkt oder eine Komponente nicht aufgelistet, verwendet dieses Produkt bzw. diese Komponente keine Cookies.

Tabelle 27. Verwendung von Cookies durch Produkte und Komponenten von InfoSphere Information Server/InfoSphere Information Server

Produktmodul	Komponente oder Feature	Typ des verwendeten Cookies	Erfasste Daten	Zweck der Daten	Inaktivierung des Cookies
Beliebig (Bestandteil der InfoSphere Information Server-Installation)	InfoSphere Information Server-Webkonsole	<ul style="list-style-type: none"> <li>• Sitzung</li> <li>• Persistent</li> </ul>	Benutzername	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> </ul>	Kann nicht inaktiviert werden
Beliebig (Bestandteil der InfoSphere Information Server-Installation)	InfoSphere Metadata Asset Manager	<ul style="list-style-type: none"> <li>• Sitzung</li> <li>• Persistent</li> </ul>	Keine personenbezogenen Daten	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> <li>• Besserer Bedienungskomfort</li> <li>• SSO-Konfiguration</li> </ul>	Kann nicht inaktiviert werden
InfoSphere DataStage	Stage 'Big Data File'	<ul style="list-style-type: none"> <li>• Sitzung</li> <li>• Persistent</li> </ul>	<ul style="list-style-type: none"> <li>• Benutzername</li> <li>• Digitale Signatur</li> <li>• Sitzungs-ID</li> </ul>	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> <li>• SSO-Konfiguration</li> </ul>	Kann nicht inaktiviert werden
InfoSphere DataStage	Stage 'XML'	Sitzung	Interne IDs	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> </ul>	Kann nicht inaktiviert werden

Tabella 27. Verwendung von Cookies durch Produkte und Komponenten von InfoSphere Information Server/InfoSphere Information Server (Forts.)

Produktmodul	Komponente oder Feature	Typ des verwendeten Cookies	Erfasste Daten	Zweck der Daten	Inaktivierung des Cookies
InfoSphere DataStage	IBM InfoSphere DataStage and QualityStage Operations Console	Sitzung	Keine personbezogenen Daten	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> </ul>	Kann nicht inaktiviert werden
InfoSphere Data Click	InfoSphere Information Server-Webkonsole	<ul style="list-style-type: none"> <li>• Sitzung</li> <li>• Persistent</li> </ul>	Benutzername	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> </ul>	Kann nicht inaktiviert werden
InfoSphere Data Quality Console		Sitzung	Keine personbezogenen Daten	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> <li>• SSO-Konfiguration</li> </ul>	Kann nicht inaktiviert werden
InfoSphere QualityStage Standardization Rules Designer	InfoSphere Information Server-Webkonsole	<ul style="list-style-type: none"> <li>• Sitzung</li> <li>• Persistent</li> </ul>	Benutzername	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> </ul>	Kann nicht inaktiviert werden
InfoSphere Information Governance Catalog		<ul style="list-style-type: none"> <li>• Sitzung</li> <li>• Persistent</li> </ul>	<ul style="list-style-type: none"> <li>• Benutzername</li> <li>• Interne IDs</li> <li>• Status der Baumstruktur</li> </ul>	<ul style="list-style-type: none"> <li>• Sitzungsmanagement</li> <li>• Authentifizierung</li> <li>• SSO-Konfiguration</li> </ul>	Kann nicht inaktiviert werden
InfoSphere Information Analyzer	Stage 'Data Rules' im InfoSphere DataStage and QualityStage Designer-Client	Sitzung	Sitzungs-ID	Sitzungsmanagement	Kann nicht inaktiviert werden

Wenn die für dieses Softwareangebot genutzten Konfigurationen Sie als Kunde in die Lage versetzen, personenbezogene Daten von Endbenutzern über Cookies und andere Technologien zu erfassen, müssen Sie sich zu allen gesetzlichen Bestimmungen in Bezug auf eine solche Datenerfassung, einschließlich aller Mitteilungspflichten und Zustimmungsanforderungen, rechtlich beraten lassen.

Weitere Informationen zur Nutzung verschiedener Technologien, einschließlich Cookies, für diese Zwecke finden Sie in der "IBM Online-Datenschutzerklärung, Schwerpunkte" unter <http://www.ibm.com/privacy>, in der "IBM Online-Datenschutzerklärung" unter <http://www.ibm.com/privacy/details> im Abschnitt "Cookies, Web-Beacons und sonstige Technologien" und in "IBM Software Products and Software-as-a-Service Privacy Privacy Statement" unter <http://www.ibm.com/software/info/product-privacy>.

## Marken

IBM, the IBM Logo und [ibm.com](http://www.ibm.com) sind Marken oder eingetragene Marken der International Business Machines Corporation. Weitere Produkt- und Servicennamen können Marken von IBM oder anderen Unternehmen sein. Eine aktuelle Liste der IBM Marken finden Sie im Web unter <http://www.ibm.com/legal/copytrade.shtml>.

Die folgenden Namen sind Marken oder eingetragene Marken anderer Unternehmen:

Adobe ist eine eingetragene Marke der Adobe Systems Incorporated in den USA und/oder anderen Ländern.

Intel und Itanium sind Marken oder eingetragene Marken der Intel Corporation oder ihrer Tochtergesellschaften in den USA und/oder anderen Ländern.

Linux ist eine Marke von Linus Torvalds in den USA und/oder anderen Ländern.

Microsoft, Windows und Windows NT sind Marken der Microsoft Corporation in den USA und/oder anderen Ländern.

UNIX ist eine eingetragene Marke von The Open Group in den USA und anderen Ländern.

Java<sup>™</sup> und alle auf Java basierenden Marken und Logos sind Marken oder eingetragene Marken der Oracle Corporation und/oder ihrer verbundenen Unternehmen.

United States Postal Service ist Inhaber der folgenden Marken: CASS, CASS Certified, DPV, LACS<sup>Link</sup>, ZIP, ZIP + 4, ZIP Code, Post Office, Postal Service, USPS und United States Postal Service. Die IBM Corporation ist ein nicht ausschließlicher Lizenznehmer für DPV und LACS<sup>Link</sup>.

Weitere Unternehmens-, Produkt- oder Servicenamen können Marken anderer Hersteller sein.



---

# Index

## B

- Befehle
  - Syntax 57
- Befehlszeilensyntax
  - Konventionen 57
- Beispiel 2: Job erstellen 13
- Beispiel 3: Job erstellen 17
- Beispiel 4: Job erstellen 24
- Beispiel 5: Job erstellen 27, 30
- Beispiel 7: Job erstellen
  - Daten in vorhandene Microsoft Excel-Dateien schreiben 35
- Bemerkungen 67

## C

- CC\_UNST\_JAVA\_HEAP, Umgebungsvariable 53

## D

- Daten aus einem Bereich in einer Microsoft Excel-Datei extrahieren
  - Job erstellen 10
- Daten extrahieren
  - DataStage 2, 3
  - Datenbereich 2
- Datenquellen
  - DataStage 2
- Datentypen
  - DataStage 39, 45
  - Daten laden 39, 45
  - Daten schreiben 39, 45
  - Unstructured Data, Stage 39, 45

## F

- Fehlerbehebung
  - Unstructured Data, Stage 49

## K

- Kundenunterstützung
  - kontaktieren 61

## M

- Mapping
  - Datentypen 39, 45
- Marken
  - Liste 67

## P

- Produktdokumentation
  - Zugriff auf 63
- Produkteingabehilfen
  - Eingabehilfen 55

## S

- Software-Services
  - kontaktieren 61
- Sonderzeichen
  - in Befehlszeilensyntax 57
- Syntax
  - Befehlszeile 57

## U

- Umgebungsvariablen
  - Unstructured Data, Stage 53
- Unstructured Data, Stage 13, 17, 24, 27, 30, 35
  - Ausgabe des Jobs anzeigen 12, 15, 18, 25, 29, 31
  - Daten in vorhandene Microsoft Excel-Datei schreiben 37
- Beispiel 1
  - Daten aus einem Bereich extrahieren 10
  - Daten in vorhandene Microsoft Excel-Dateien schreiben 35
  - Daten schreiben 24
- Beispiel 1: Stage 'Sequential File' konfigurieren 11, 14, 18
- Beispiel 1: Stage 'Unstructured Data' konfigurieren 11
- Beispiel 2 27
  - Daten aus mehreren Microsoft Excel-Arbeitsblättern extrahieren 13
- Beispiel 2: Stage 'Unstructured Data' konfigurieren 13
- Beispiel 3
  - Daten aus mehreren Bereichen extrahieren 16
  - Daten in mehrere Microsoft Excel-Dateien schreiben 30
- Beispiel 3: Stage 'Unstructured Data' konfigurieren 17
- Beispiel 4: Stage 'Sequential File' konfigurieren 25
- Beispiel 4: Stage 'Unstructured Data' konfigurieren 24, 28
- Beispiel 5: Stages 'Sequential File' konfigurieren 28
- Beispiel 6: Stage 'Unstructured Data' konfigurieren 30
- Beispiel 6: Stages 'Sequential File' konfigurieren 31
- Beispiel 7: Stage 'Unstructured Data' konfigurieren
  - Daten in vorhandene Microsoft Excel-Datei schreiben 36
- Beispiele: Daten aus Microsoft Excel-Dateien extrahieren 10
- Beispiele: Daten in Microsoft Excel-Dateien schreiben 24

## Unstructured Data, Stage (Forts.)

- Daten in Microsoft Excel-Datei schreiben 18, 21
    - Jobs entwerfen 19
  - Daten in vorhandene Microsoft Excel-Dateien schreiben 32
  - Daten schreiben 27
  - Datenquelle konfigurieren 6
  - Fehlerbehandlung 8, 9
  - Hinweis zum Ende des Laufs 22
  - installieren und konfigurieren 2
  - Jobabbruchbedingungen 46
  - Jobparameter im Konfigurationsfenster 8, 20, 34
  - Jobs definieren 1, 33
  - Jobs entwerfen 1, 5, 32
  - konfigurieren
    - vorhandene Microsoft Excel-Datei modifizieren 33
  - mehrere Arbeitsblätter 27
  - Nullzeilenbehandlung 8
  - Optionen für das Lesen von Daten aus Microsoft Excel-Dateien 8
  - Referenz
    - unterstützte Mappings 39
  - Spaltendefinition angeben 20
  - Spaltendefinition modifizieren 7
  - Spaltenweitergabe während der Ausführung 9
  - Stage 'Sequential File' konfigurieren
    - Daten in vorhandene Microsoft Excel-Dateien schreiben 36
  - Stage 'Unstructured Data' konfigurieren 19, 21
  - Übersicht 1
  - vorhandenes Microsoft Excel-Arbeitsblatt 32
  - Wert einer Zelle oder angepasste Eigenschaften extrahieren 9
- ## Unterstützung
- Kundenunterstützung 61

## W

- Websites
  - nicht von IBM 59









SC43-1098-00

