

Interacting with Bulk Data in Force.com



Richard Seroter

VICE PRESIDENT OF PRODUCT, CENTURYLINK

@rseroter www.seroter.com



Overview



Key concepts in the Bulk API

Lifecycle of a Bulk API interaction

Authenticating Bulk API users

Job lifecycle

About batches

Using transformation specifications

Reviewing error codes

Executing bulk queries

Best practices and limits

Summary



Bulk API

Programmatic access to quickly load into, and retrieve data from, Salesforce.



Key Concepts in the Bulk API

**Jobs | Batches |
Records**

**Independent
batch processing**

**Multiple file
formats accepted**

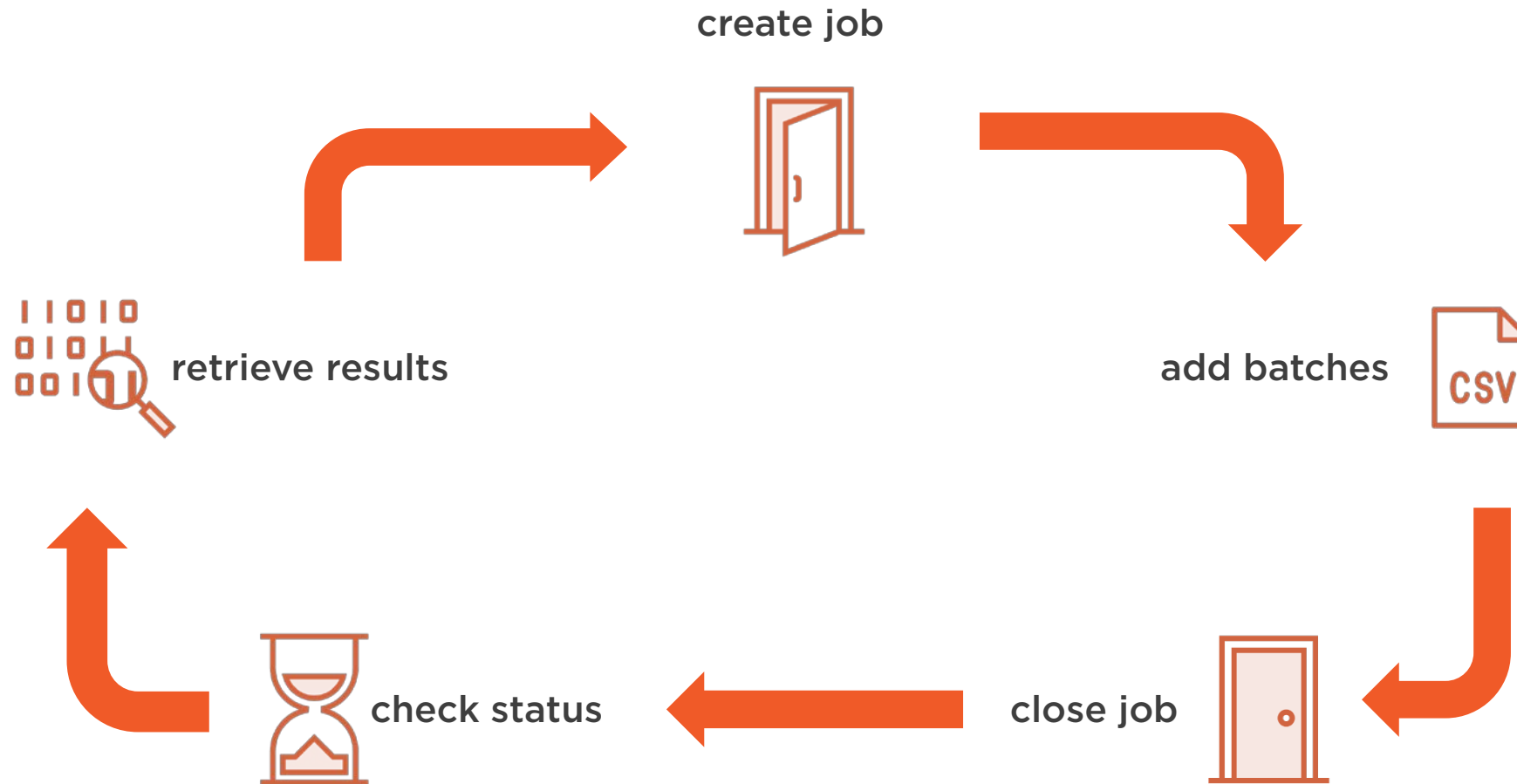


When Do You Use It?



- Data synchronization between systems**
- One-time data loads**
- Transferring significant number of records**

Lifecycle of a Bulk API Interaction



Authenticating Bulk API Users



Acquire session ID
from SOAP endpoint



User, object, field
security applies



Set X-SFDC-Session
header

Create a job

Get job details

Monitor a job

Close a job

Abort a job

Precursor to loading data

Called with HTTP POST

Specify object, operation

Optional concurrency parameter

Represented by JobInfo resource



Create a job

Get job details

Monitor a job

Close a job

Abort a job



Return job info, state in XML

Called with HTTP GET

Shows processing statistics



Create a job

Get job details

Monitor a job

Close a job

Abort a job

Visible in Salesforce Setup site

Shows usage quota

Lists in-progress, completed jobs

View job request, response



Create a job

Get job details

Monitor a job

Close a job

Abort a job

Prevents adding new batches

Called with HTTP POST

Batch processing doesn't wait for closure

Jobs deleted after 7 days, regardless of status



Create a job

Get job details

Monitor a job

Close a job

Abort a job



No more records are processed

Called with HTTP POST

No rollback of previous changes



About Batches

**Set of records
sent via HTTP
POST**

**Text or text +
binary content**

**Monitor via UI or
API**

**Multiple states in
lifecycle**

**Consider batch
sizes**

**Partial success
possible**



Batches Take CSV, XML, Binary Content

CSV and XML files

Need to find full field names

First row of CSV contains field names

XML uses SObject and field names

Records all of same object type

Include all required fields

If updating, missing fields ignored

Relationships supported

No upfront validation

Binary content

Zip + request.txt + binary attachments

request.txt is CSV or XML manifest

Supports nested structures



Demo



Review XML batch of Voter records

Create job for inserting Voters

Add batch to job

Check details of job

Monitor progress in Salesforce

Close job



data.csv

Voter Name,Party,Precinct

April Ludgate,Democrat,3rd Precinct

Jerry Gergich,Independent,4th Precinct

Chris Traeger,Republican,3rd Precinct

spec.csv

Salesforce Field,Csv Header,Value,Hint

Name,Voter Name,,

Political_Party__c,Party,Independent,

Precinct__c,Precinct,,

◀ Can take non-standard CSV

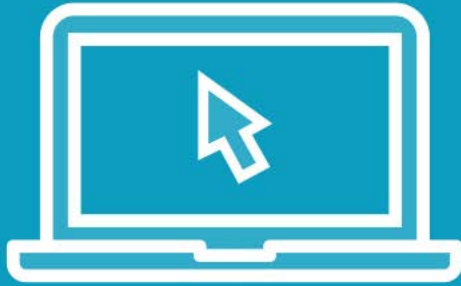
◀ Create spec scoped to job
HTTP POST to URI

◀ Defines mapping

◀ Can provide default value



Demo



Review non-standard batch file of Voters

Review mapping spec

Create job

Submit spec

Add batch

Close job

Review import results



Handling Failures



ExceededQuota

InvalidBatch

InvalidJob

InvalidUser

Timeout

TooManyLockFailure

Unknown

Executing Bulk Queries



Reduce API requests



Retrieve up to 15GB



Put SOQL in HTTP
request



Get back result ID



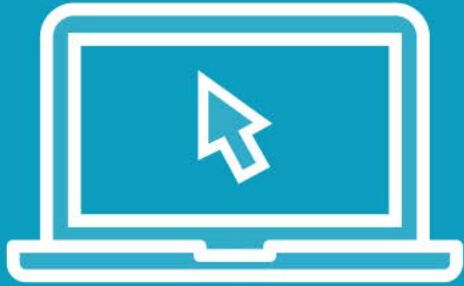
Use result ID to return
response body



Subset of SOQL
supported



Demo



Create job for query operation

Submit query batch to retrieve all Voters

Monitor batch status

Retrieve result key

Retrieve result CSV



Best Practices for Bulk API Usage

**Use parallel mode
when possible**

**Proactively avoid
lock contention**

**Know which
operations cause
lock contention**

**Minimize field
count**

**Minimize post-
processing
activities**



Bulk API Limits

10

Minutes to process a batch

5,000

Batches per rolling 24 hour period

10,000

Maximum record count per batch

10MB

Max size for single CSV or XML file



Summary



Overview

Key concepts in the Bulk API

Lifecycle of a Bulk API interaction

Authenticating Bulk API users

Job lifecycle

About batches

Using transformation specifications

Reviewing error codes

Executing bulk queries

Best practices and limits

