

Workflow

Grundidee

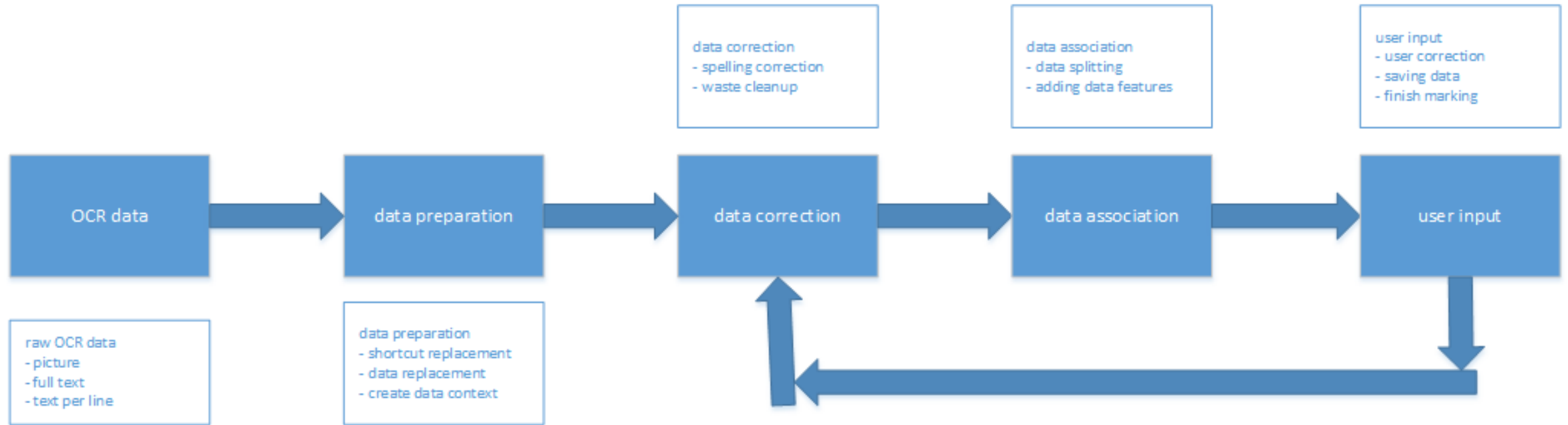
Jede Aktion wird asynchron von unabhängigen Workern ausgeführt

Jede Aktion ist Eindeutig, aber nicht einmalig

Aktionen müssen in fester Reihenfolge durchlaufen werden, sofern nicht anders definiert

Aktionen dürfen pro Karte nur einfach ausgeführt werden

workflow



ORM

Durch ORM verarbeitete Daten

Werden in Ordnerverzeichnisse abgelegt

- Struktur? – Klärung

Karte

Eine Karte enthält alle OCR Daten in gekapselten Objekten

- Bilder
 - Orginal
 - Thumbnails
 - Line
- OCR
 - Full Text
 - Lines
 - Text pro Line

Status – Karten Status „inProgress, finished, locked“ ect.

Job Status - letzter durchlaufender Job Status

Processed Data – selbe wie OCR mit verarbeiteten Daten

- Unabhängig von OCR

Queues und Jobs

Aktionen sind Jobs untergeordnet

Jobs können mehrere Aktionen beinhalten, müssen aber nicht

Queues werden befüllt und durch Worker bearbeitet

Job

Eindeutige Kennung

Besitzt Karte als Objekt

Keine weiteren Aktionen oder Werte

	Job			
Prozesskennung	ID	Handler	Beschreibung	Anmerkung
10	100	DataCrawler	Durchläuft das Directory für OCR Daten und stößt neue Prozesse für die erfassung der Karten an	Erzeugt CardFactory Job selbstständig
10	200	CardFactory	Generiert neues Kartenobjekt mit allen OCR Daten	Erzeugt Karte mit Init-Values
20	300	ShortcutReplacer	Ersetzt Shortcuts durch bekannte Synonyme	
20	400	ExternalInterpreterLookup	Lookup auf Interpreterdaten für spezifische Karte mit Daten der Staatsbibliothek	
20	500	WorkLookup	Lookup auf Werkdaten für spezifische Karte mit Daten der Staatsbibliothek	
	Prozess			
	ID	Name	Beschreibung	
	10	OCR data	Vorverarbeitung der OCR Daten	
	20	data preparation	Anreichern der Kartendaten mit Hilfsinformationen vor dem primären Prozess	
	30	data correction		
	40	data association		
	50	user input		

JobCreator

Permanent laufender Prozess

Erzeugt aus Karten nächsten Job

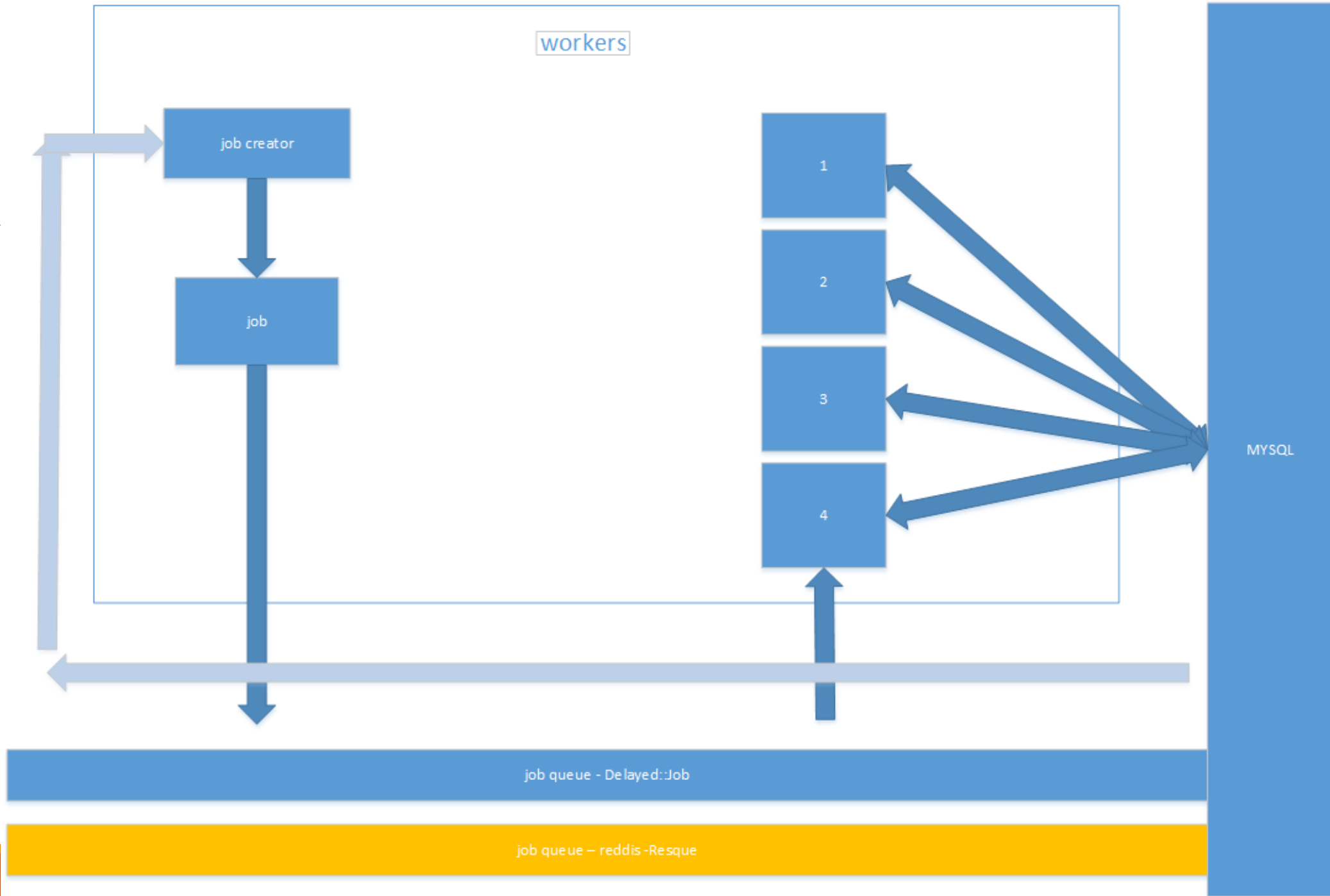
Besitzt eine Reihenfolge für Jobs

Worker

Einzelner unabhängiger Systemprozess

Jeder Worker kann jeden Job bearbeiten

Teilt keine Systemressourcen mit anderen Workern außer der Queue(die symbolisch sein kann)



JobHandler

Verarbeiten eindeutigen Job

Reichert spezifische Daten seines Jobs zur Karte an

Kann Vorstufe für andere Jobs sein

