KPMG

cutting through complexity

Hadoop WorkFlow Scheduler / Automation Engine

Azkaban & Oozie

Praveen Thirukonda
Senior Associate
Data & Analytics
Orange County, CA
09/11/2014

What is a workflow?

- **A workflow is a Directed Acyclic Graph (DAG) of “jobs” where each job has one or more inputs and outputs.**
- **A workflow scheduler helps us manage the co ordination among the various jobs.**

When do we need a workflow scheduler?

- **In a Data Pipeline, Batch jobs need to be scheduled to run periodically.**
- **They also typically have intricate dependency chains—for example, dependencies on various data extraction processes or previous steps.**
- **Larger processes might have 50 or 60 steps, of which some might run in parallel and others must wait for the output of earlier steps.**

Azkaban

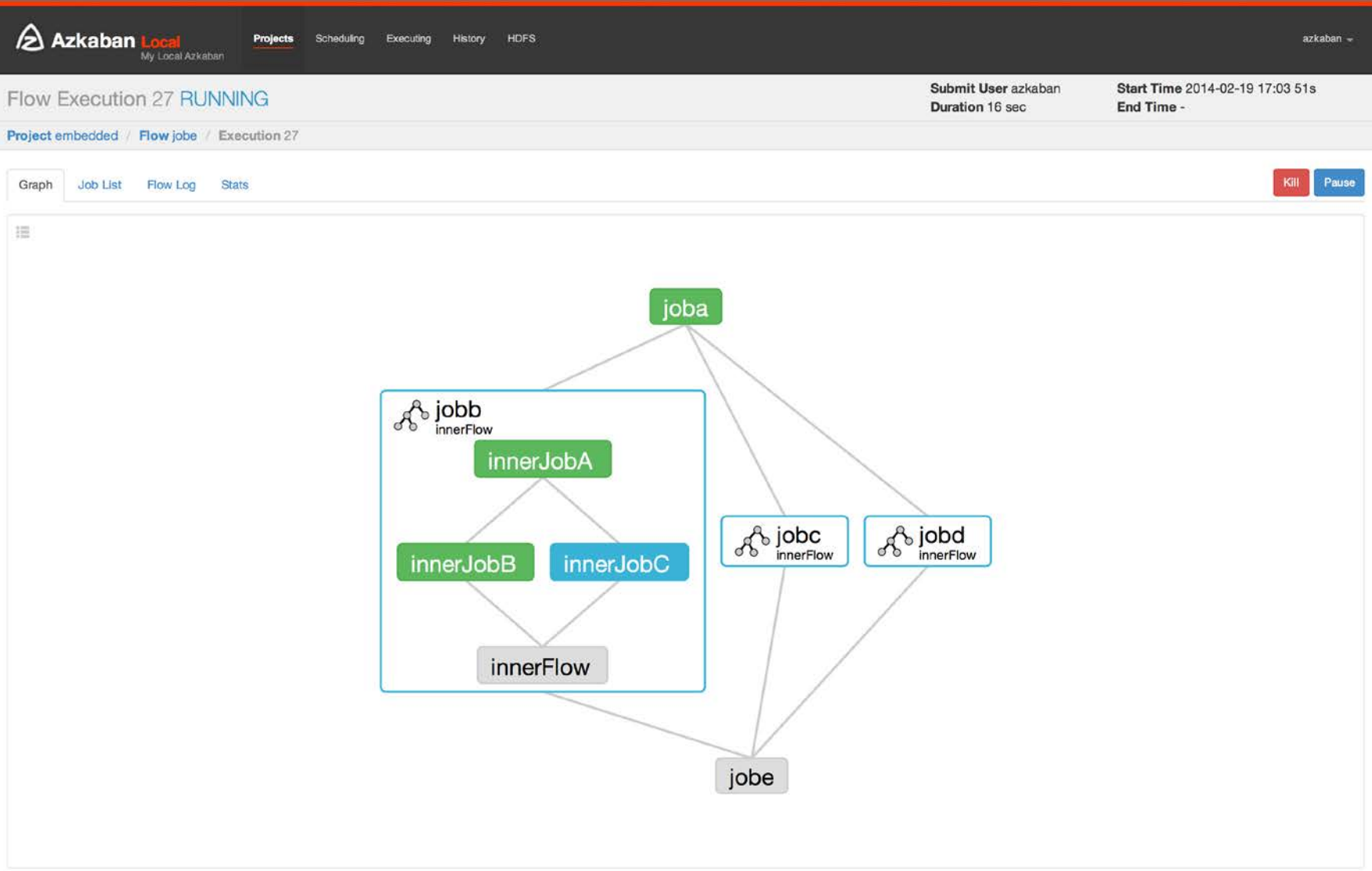
What is Azkaban?

- **“cron on steroids”**
- **A workflow scheduler can be seen as a combination of the cron and make Unix utilities combined with a friendly UI.**

What is Azkaban?

- **Azkaban was implemented at LinkedIn to solve the problem of Hadoop job dependencies.**
- **Azkaban resolves the ordering through job dependencies and provides an easy to use web user interface to maintain and track your workflows.**

An Image is worth a 1000 words..



Apache Oozie

What is Apache Oozie?

- **Similar to Azkaban.**
- **Whereas Azkaban uses a series of Properties files, Oozie uses an XML file.**
- **Oozie supports Java API, command line methods for workflow submission in addition to Browser interface/REST API.**
- **Oozie is part of our Hortonworks environment in our cluster.**

Advantages of using a workflow scheduler

- **Let’s you easily manage dependencies within the various tasks.**
- **Scheduling of workflows**
- **Monitor the progress of your workflow with nice interface.**
- **Email alerts on failure and successes**
- **Retrying of failed jobs.**

- **Real Life example of how and where you might use a workflow scheduler in your Big Data System architecture?**

Thank you

Presentation by Praveen Thirukonda