



PLURALSIGHT

DataOps

Week 6



Proprietary and confidential

DataOps

AWS DDK

The **AWS DataOps Development Kit (DDK)** is an open-source development framework that helps you build data workflows and modern data architecture on AWS. It is based on the AWS Cloud Development Kit (CDK) and offers high-level abstractions to build pipelines that manage data flows on AWS, driven by DevOps best practices.

AWS DDK

Key features of the AWS DDK include:

Build Data Pipelines

- DDK allows creating Data Pipelines composed of chained stages that are automatically "wired" together using AWS EventBridge Rules
- It comes with a library of stages but users can also create their own based on their use cases

Extensible and Modular

- The DDK Core is a library of CDK constructs that can be used to build data workflows following AWS best practices
- It is modular and extensible - users can update and share their own versions of the DDK Core with the rest of the organization

Compose into DDK Apps

- Constructs from the DDK Core can be composed into a DDK App
- DDK Apps can also contain constructs from the CDK Framework or the AWS Construct Library

Regular Expressions

What is Regular Expression (RegEx)

A regular expression, often abbreviated as regex or regexp, is a sequence of characters that defines a search pattern. This pattern is primarily utilized for string searching and manipulation tasks, such as finding or replacing text, validating input, and parsing data.

Regular expressions serve as a powerful tool in programming and text processing. They allow users to specify complex search criteria in a concise manner, enabling operations like "find" or "find and replace" within strings. Regex patterns can match specific sequences of characters, making them invaluable in various applications, from text editors to programming languages and data validation tools.

GREP

Grep is a useful command to search for matching patterns in a file. grep is short for "global regular expression print".

grep stands for "**G**lobal **R**egular **E**xpression **P**rint."

It's a command-line utility for searching plain-text data for lines that match a regular expression. Commonly used for filtering and searching through large sets of text.