POWERS: Welcome to This Week on developerWorks, I'm your host Calvin Powers. This Week on developerWorks we've got an excellent tutorial on moving data between noSQL environments and traditional database environments. But first, let's look at what else is new on developerWorks this week.

NARRATOR: Joern Klauke and Martin Jungfer have published a new how-to guide for using the new scripted interface for DB2 Advanced Copy Services. This new feature introduced in DB2 10.5 allows administrators to invoke the Advanced Copy Services from shell scripts instead of native code. This new feature enables IT shops to incorporate new storage hardware into their backup infrastructure more quickly and consistently. This how-to guide will be useful for anyone managing a DB2 backup environment.

Enzo Cialini, Ian D. M. Hakes, Richard Lubell and Paul McInerney have published a new how-to guide on Implementing disaster recovery in IBM PureData System for Transactions. This article outlines the setup and operation of a disaster recovery solution for DB2 V10.5 databases on the IBM PureData System for Transactions, Fix Pack 3 or later.

The solution is based on the DB2 High Availability and Disaster Recovery feature, and includes additional elements that are not handled directly by the product. This article will be of interest to anyone who needs to make sure their mission critical data can always be restored.

Neal Ford has published part two in his Java.next series. In this latest installment, he discusses Java's lack of inheritance mechanisms that are commonly found in Groovy, Scala and Clojure. This installment further explores Clojure's use of protocols as an extension mechanism. If you're interested in the evolution of Java, you'll want to read about these extension mechanisms.

POWERS: Don't forget that you can get links to all those items as ibm.com/developerWorks/thisweek.

My guest this week on developerWorks is Surajit Paul. He's a developer and consultant with IBM in Bangalore and he specializes in supply chain management and he's an [Epics] certified supply chain professional. Surajit, welcome to This Week on developerWorks. Tell us briefly what [Sqoop] is a how it fits into the Hadoop ecosystem.

PAUL: Thanks, Calvin. Sqoop is essentially used for transferring unstructured data from RDBMS to a Hadoop environment and vice versa. In order to transport the data, Sqoop launches MapReduce Java which accomplishes that activated with the help of several concurrently executed tasks.

In order to transport that data, Sqoop supports various data formats -- for example, text files, et cetera. Once the data is analyzed and processed in the Hadoop environment, the data can be exported back to the RDBMS platform for SQL Purposes purpose.

POWERS: Thanks Surajit, now tell us what aspects of Sqoop you cover in your article?

PAUL: Several technical articles, blogs are available online discussing about Sqoop command line interface. Although the Sqoop is developed using Java programming language, Sqoop Java API is hardly discussed anywhere, which developers are usually eager to learn about. In my article, I have discussed about Sqoop command line interface with an additional emphasis on Sqoop Java APIs.
POWERS: Tell us about the example dataset that you used in your article and what you show the readers.

PAUL: In my article, I have collected the data from the Bombay Stock Exchange website. I have taken the data off BSE index. My goal was to compare the market dynamics, and specifically in the [INAUDIBLE] market. So I have taken the data from our RDBMS system to hadoop environment and then I did some manipulation of the data to summarize by month and after analyzing and enrichment of the data, I summarize the original data instead of building RDBMS tables. Once a month I take the summarized data from the Hadoop environment and I export it back to the RDBMS environment.

POWERS: Thanks Surajit, that's a great example use case to show data both going into and out of the Hadoop environment. Don't forget that you can get a link to Surajit's article at ibm.com/developerWorks/thisweek. That's all the time we have for this episode. We'll see you next week on developerWorks.

[ MUSIC ] [END OF SEGMENT]