Customer Viewing Data Sessionization

Team Size: 3-5 (Absolute Maximum 5)
Location: Primarily Offsite (on campus)

Company Background

Comcast is one of the largest Telecommunication companies in the US. We provide customers with a wide range of products, ranging from an incredible cable platform to the best Internet and a great home security system product.

The team that you would be working with is a data engineering team on the IP video side of the company. We build data pipelines for the customer IP video player data.

Desired Skills

- Comfortable with programming in Java
- Familiar with Scala/Spark
- Comfortable installing and configuring software on personal computer
- Some understanding of Hadoop/HDFS
- Familiar with SQL
- Work well as team

Project Description

Say you were in your apartment and you wanted to watch the newest episode of Silicon Valley on your Computer, with your Xfinity subscription, you would probably search for the episode you want and hit the play button. At this point you initiate what we call the IP Player. The player is the Application plug-in that allows you to watch your episode of Silicon Valley.

Once the player is initialized, it begins to send several events that correspond to the health of the player that is allowing you to enjoy Silicon Valley. We refer to the entire experience of watching Silicon Valley as one session.

What do we need from you? We are looking for a way to take all of these player events and combine them into one record that describes each individual session. You will be given a large amount of events that will make up several unique sessions.

Ultimately we are looking for the team to utilize Scala/Spark functionality in order to "sessionize" IP Player Events into a single viewing session. We will also want a report that provides valuable session metrics.