主题：怎么使用Spark ALS 算法提高电商产品推荐的效率，对比SVD算法处理更高效。同时使用HDFS分布式存储数据，提高数据处理效率。

presentation:

Presentation Component Marks (indicative)  
1、Problem identification and motivation /10  
Motivation or “hook” to establish importance of topic

ALS处理大数据集比SVD更有优势  
2、Define scope /10  
Statement of objective, its context, and its limitations

data preparation cleaning processing—Miller Susie

ALS

SVD

结果分析最后写

3、State-of-the-art of topic area /30  
Appropriate presentation of content in depth and breadth using  
additional sources

Spark： MLIB---Hisky

HDFS--David  
ALS --Susie  
SVD—Susie

算法优化-Hisky

最终结果展现

4、Future Issues /10  
Discussion of future issues, challenges and their implications  
查一下资料

David

5、Examples /20  
Use of examples to illustrate concepts/problems effectively

代码程序演示  
Presentation style /20  
Communication is clear, and engaging  
Use of appropriate and effective media  
Accurate and complete citation and referencing

代码：

1、找大数据集

2、数据用HDFS分布存储

3、数据清洗

4、算法优化

1、Problem identification and motivation:

2、Define scope

**4、Project** **Execution Outline**

**4.1 Project Objective**

Predict and recommend products that users may be interested in, based on their historical behaviors such as clicks, ratings, and purchases.

**4.2 Core Methodology**

Collaborative Filtering using Spark MLlib’s ALS (Alternating Least Squares)

**4.3 Data Preparation**

Datasets: Amazon Product Dataset,

Data Processing:

* Data loading with HDFS
* Data cleaning and type casting
* Filter out invalid ratings

**4.4 Feature Engineering**

Interaction Features:

* Explicit feedback: direct rating scores
* Implicit feedback: behavioral weights

**4.5 Model Training**

Train ALS model with:

* Explicit or implicit feedback
* Hyperparameter tuning (rank, regParam, maxIter)

Handle cold-start problem via:

* Popularity-based fallback
* Content-based recommendations

**4.6 Evaluation & Optimization**

**Metric**: RMSE on split validation sets

**Cross-validation**: Grid search over ALS parameters

**4.7 Generating** **Recommendations**

**Recommend top-N items for:**

* **Individual users**
* **All users (batch processing)**

**Display or export results for further usage**