**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | Netflix Movie Service | |
| **Vertical (area)** | | Commercial Cloud Consumer Services | |
| **Author/Company/Email** | | Geoffrey Fox, Indiana University gcf@indiana.edu | |
| **Actors/Stakeholders and their roles and responsibilities** | | Netflix Company (Grow sustainable Business), Cloud Provider (Support streaming and data analysis), Client user (Identify and watch good movies on demand) | |
| **Goals** | | Allow streaming of user selected movies to satisfy multiple objectives (for different stakeholders) -- especially retaining subscribers. Find best possible ordering of a set of videos for a user (household) within a given context in real-time; maximize movie consumption. | |
| **Use Case Description** | | Digital movies stored in cloud with metadata; user profiles and rankings for small fraction of movies for each user. Use multiple criteria – content based recommender system; user-based recommender system; diversity. Refine algorithms continuously with A/B testing. | |
| **Current**  **Solutions** | **Compute(System)** | | Amazon Web Services AWS |
| **Storage** | | Uses Cassandra NoSQL technology with Hive, Teradata |
| **Networking** | | Need Content Delivery System to support effective streaming video |
| **Software** | | Hadoop and Pig; Cassandra; Teradata |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Add movies institutionally. Collect user rankings and profiles in a distributed fashion |
| **Volume (size)** | | Summer 2012. 25 million subscribers; 4 million ratings per day; 3 million searches per day; 1 billion hours streamed in June 2012. Cloud storage 2 petabytes (June 2013) |
| **Velocity**  **(e.g. real time)** | | Media (video and properties) and Rankings continually updated |
| **Variety**  **(multiple datasets, mashup)** | | Data varies from digital media to user rankings, user profiles and media properties for content-based recommendations |
| **Variability (rate of change)** | | Very competitive business. Need to aware of other companies and trends in both content (which Movies are hot) and technology. Need to investigate new business initiatives such as Netflix sponsored content |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues)** | | Success of business requires excellent quality of service |
| **Visualization** | | Streaming media and quality user-experience to allow choice of content |
| **Data Quality** | | Rankings are intrinsically “rough” data and need robust learning algorithms |
| **Data Types** | | Media content, user profiles, “bag” of user rankings |
| **Data Analytics** | | Recommender systems and streaming video delivery. Recommender systems are always personalized and use logistic/linear regression, elastic nets, matrix factorization, clustering, latent Dirichlet allocation, association rules, gradient boosted decision trees and others. Winner of Netflix competition (to improve ratings by 10%) combined over 100 different algorithms. |
| **Big Data Specific Challenges (Gaps)** | | Analytics needs continued monitoring and improvement. | |
| **Big Data Specific Challenges in Mobility** | | Mobile access important | |
| **Security & Privacy**  **Requirements** | | Need to preserve privacy for users and digital rights for media. | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | Recommender systems have features in common to e-commerce like Amazon. Streaming video has features in common with other content providing services like iTunes, Google Play, Pandora and Last.fm | |
| **More Information (URLs)** | | <http://www.slideshare.net/xamat/building-largescale-realworld-recommender-systems-recsys2012-tutorial> by Xavier Amatriain  <http://techblog.netflix.com/> | |
| **Note:** <additional comments> | | | |