Games-R-Us

Team 5:
Felice Gardner
Sumit Gupta
Calvin W. Lew
Muhammad Nasution
Kevin Tsuei
ISMT E-599 Capstone Seminar in Digital Enterprise

1 Executive Summary

1.1 Company Overview

Games-R-Us is a global gaming company, which incorporated in 2010, that provides massively multiplayer online (MMO) games through a monthly subscription targeted at individuals aged 13+ with games delivered on electronic devices. In addition to revenue generated from subscriptions, players can make in-game purchases (i.e. character skins, powers, hints, etc.) for use during game play. The company was a leader in the online gaming space until 2014 when increased competition in the gaming space eroded Games-R-Us' market share, revenues, and customer base. With the customer base and revenues in decline, company executives and the Information System and Marketing Team (ISMT) team determined that an investment in Machine Learning (ML) could help the company retain the customer base and grow revenues. Machine learning is a part of artificial intelligence (AI) in which a system automatically learns and improves from experience or vast amounts of data without being explicitly programmed.

1.2 ISMT Group Overview

To retain the customer base and grow revenues, company executives mandated that the ISMT team research and implement technologies that could assist in the retention and expansion of the customer base and revenues. The Information Systems group's responsibilities included the oversight of the installed technology as well as to make recommendations about and install new technologies that will improve the business processes. IS interfaces with the Marketing Team (MT) to ensure the technologies address their needs to retain existing and trial customers and drive more in-game purchases through marketing efforts.

1.3 Business Problem and Objectives

In 2014, as other gaming companies grew their online presence and market shares, Games-R-Us' customer base and market share began to erode. As the customer base declined, executives explored options to stem the losses, which included the increased use of business analytics (BA) to enhance its marketing campaign. Though BA did improve retention by about 15%, it did not provide a significant improvement.

Games-R-Us' main objectives are to retain the existing customer base by minimizing churning, and grow the base of our new subscribers. By leveraging the data analyzed by Yokozuna Data's ML algorithm, our product managers will be able to target trial gamers (free trial subscribers) most likely to convert to pay-for-play as well as players most likely to make in-game purchases and provide targeted marketing campaigns.

1.4 Proposed Solution

After a review of emerging technologies in the online gaming industry, discussions with Machine Learning (ML) experts, consultation with peers in top online gaming companies, ISMT recommended that the company expand their business analytics (BA) capabilities as well as engage with Yokozuna Data. Yokozuna Data is an analytics company focused on ML for the online gaming industry to prevent player churn and increase player engagement. The ISMT determined that the wealth of data generated from gamers' play could be captured for internal BA and exported to the Yokozuna Data platform to develop targeted marketing campaigns based on ML results. Yokozuna's ML algorithm focuses on identifying players most likely to churn. The data platforms could also be used in the future to improve existing games and develop engaging new games based on predictive analytics, business analytics, and machine learning.

ISMT conducted interviews with top online gaming companies to understand their technological solutions. After reviewing this information, ISMT recommended the incorporation of ML models to process gamer generated data. The ML algorithms will enable the company to predict with 80 percent accuracy whether a current subscriber is likely to renew and whether a trial subscriber will convert to a paid subscription. If a trial subscriber is flagged as a potential churner, a customized in-game message will be sent with a coded promotion to entice them to convert to a subscription. The promotion is either for a reduced subscription rate, extended trial (new subscribers), or free in-game items, and is based on the ML algorithm's assessment of churning probabilities. Similar marketing offers can be sent to existing subscribers who are at risk of not renewing (churning) based on ML predictions.
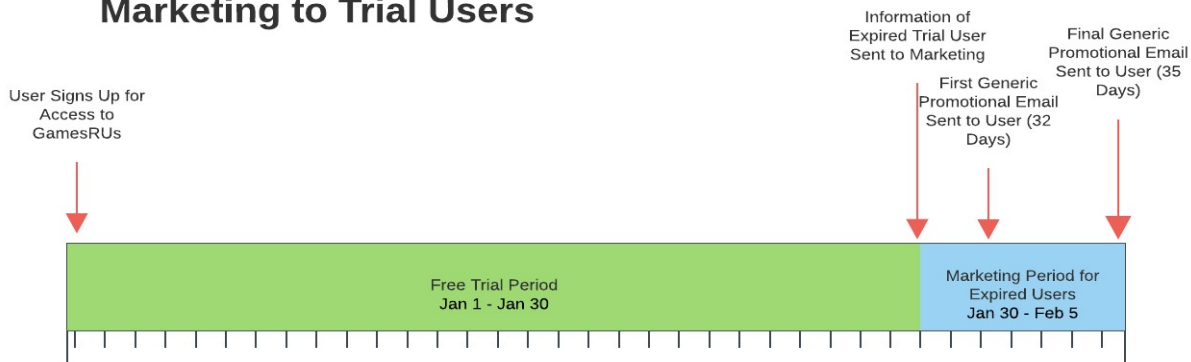
2 Business Requirements

2.1 Business Process

2.1.1 As-Is

The current subscription model for Games-R-Us is based on monthly fees to provide access to all games offered on the platform. Subscribers can purchase additional in-game items such as game characters, character skins, hints, power, potions, and special abilities to use during game play.

The system now tracks trial subscriptions to the expiration date. At the end of the trial period (30 days), expired subscriptions are gathered in a spreadsheet by the marketing team and manually entered into another system to generate a marketing campaign with a generic discount code towards the subscription rate. The code is not associated with a specific player.

Though we collect various gamer metrics, only rudimentary business analytics are applied to the data. Currently, the BA are only used to generate reports to company executives and the marketing team to track revenue and subscribers.

As-Is Workflow for Marketing to Trial Users

User Signs Up for Access to GamesRUs

Information of Expired Trial User Sent to Marketing

First Generic Promotional Email Sent to User (32 Days)

Final Generic Promotional Email Sent to User (35 Days)

Free Trial Period Jan 1 - Jan 30

Marketing Period for Expired Users Jan 30 - Feb 5

Note: Sample timeline for user that signs up on January 1st.

As-Is Platform Architecture
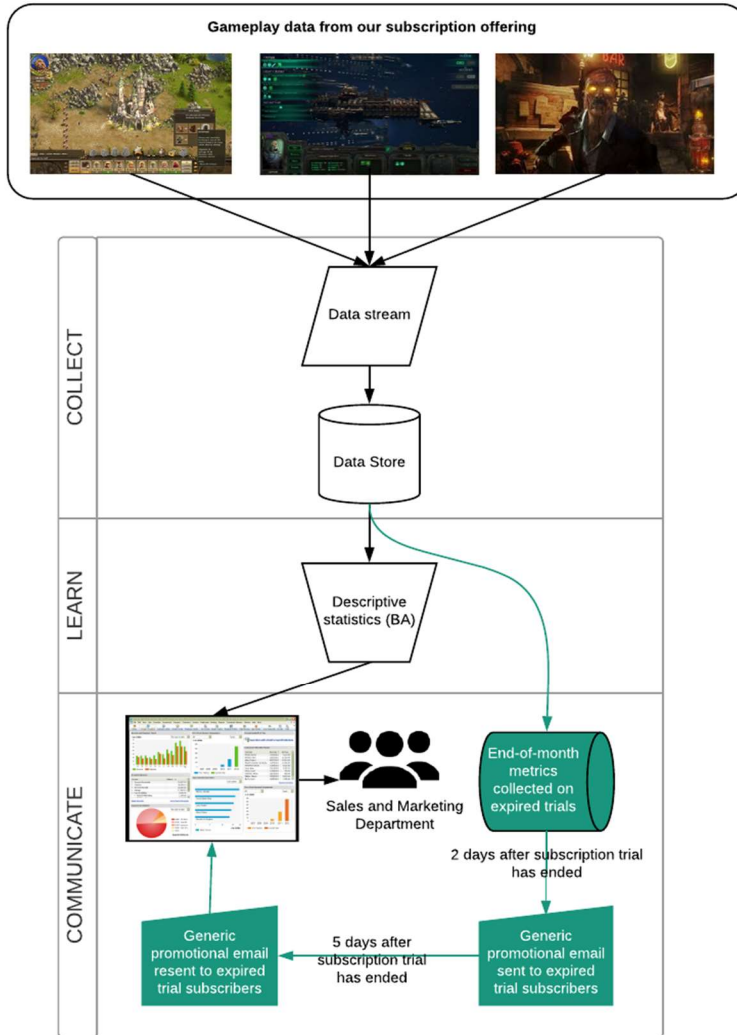
2.1.2 To-Be

The To-Be revenue model remains the same: monthly subscription to access all game offerings on our platform and additional fees for in-game purchases. However, data collection and analytics will begin at the trial player's account creation and follow all of the player's movements. Instead of a manual process to track expired subscriptions and manually generate generic marketing campaigns to expired trial subscribers, the To-Be workflow will use machine learning to automate the process. Marketing managers will have the ability to identify and monitor potential churners.

The To-Be workflow for marketing managers will start as soon as the player starts playing out games. The ML model, based on data collected from account initiation, will determine whether the trial and existing subscriber is likely or unlikely to depart from the game (churn). If the model indicates that a subscriber is likely to churn, an in-game message is initiated with an extended trial or discounted paid subscription. Marketing managers can customize and automate the message before it is sent.

In addition to extended trial or discounted paid subscription, marketing managers can also send in-game items to prevent users from churning. The offer is based on the ML algorithm's assessment of which offer will be most effective in conversion. The campaign can be customized and automated by marketing managers.

We recognize that for our machine learning algorithm to perform these functions, we need to include additional metrics in our To-Be architecture. The table below includes the In-Game As-Is and To-Be metrics and a separate table for additional As-Is and To-Be metrics:

As-Is In-Game Metrics Collected	To-Be In-Game Metrics Collected
- Per-game offering: - Total Number of Active Users - Total Number of New Users - Total Number of Trial Users	- Length of time each subscriber spent per game - Subscriber engagement with other subscribers - Time of day when subscriber plays - Time spent in each level / stage - Types of in-game purchases per game

As-Is Metrics Collected	To-Be Additional Metrics Collected
- Number of games played by each subscriber - Purchase/Subscription History - Length of subscriber relationship	- Lifetime values based on purchase / subscription history - Estimated financial status and other socio-cultural information of subscriber - Analytics from previous customized offers

Though the As-Is metrics include user information and financial data, this data was not previously used to generate detailed business analytical reports. For example, while we have purchase / subscription history information per gamer in our finance database, this data was not previously used to calculate the lifetime value of each gamer. Also, customer demographic data, which is collected, has never been used to estimate the gamer's financial or socio-cultural status. This data, if fed to a ML algorithm, can provide additional venues to engage users and promote more play time and purchases.

2.2 User Stories

User Story	Acceptance Criteria
As a Marketing Manager, I need to customize the promotional campaign based on subscriber activities on our gaming platform.	Based on Machine Learning algorithms predictions, Marketing Managers will be provided with the best promotional offer for specific groups of subscribers, which can be viewed in the dashboard.
As a Marketing Manager, I need to objectively measure the success of our customized marketing campaigns.	Marketing Managers will be able to analyze the results of customized marketing campaigns using the Tableau dashboard. Data includes both financial results and churn rate.
As a Marketing Manager, I need a dashboard that provides various customer metrics	Through a dashboard, Marketing Managers will be able to view - Daily Active Users and Average Playtime - Time of day when subscriber plays (based on their individual time zone - User counts by date (new, active, and total users) - Overall churn risk assessment - Churn analysis on specific users - Cost of customer acquisition and retention

	● ARPU - average revenue per user
As the Marketing Manager, I need a list of existing customers who are most likely to churn.	The Marketing Manager will be able to view a list of customers with an **80%** probability to churn along with customized marketing campaigns based on their hours of play, engagement with other players, and purchase of in-game tokens.
As the Product Manager, I need to understand the gamer's usage pattern.	Engineers will instrument the application to collect the required metrics to be displayed in a dashboard. The following metrics will be collected and displayed: ● Monthly/Daily Active Users ● Number of hours played ● Number of games played ● Purchase/Subscription History ● Length of time each subscriber plays a game ● Time of day when subscriber plays ● Subscribers engagement with other subscribers in game play ● Time spent in each level / stage ● Lifetime values based on purchase / subscription history ● Results from previous customized offers ● Number of multiplayers in game per hour ● Average number of hours spent ● Platform Performance Data such as frame rate execution on gamer's platform and game server stability

2.3 Functional Requirements

- Instrument the gaming application to collect additional event data
- Collect and ingest the following data in the data lake via the Informatica Cloud platform:
  - Financial Data
  - Gaming Data
  - Customer Data
- Determine best machine learning algorithm for churn prediction
- Machine learning predictions available every 24 hours as downloaded from Yokozuna Data ML results
- Machine learning predictions to provide best in-game subscription or items to entice disengaged subscribers to play
- Provide dashboards for the Chief Marketing Officer (CMO) and marketing managers that enable:
  - View campaign metrics and statistics on various marketing campaigns
  - Preview / edit / approve marketing campaigns
- Dashboards for company executives, marketing managers, and product managers that provide:

- View sales — daily, weekly, monthly
- View potential churners — daily, weekly, monthly
- View playtime by game — hourly, daily, weekly, monthly
- View new, active, and total players by game — hourly, daily, weekly, monthly
- View marketing budgets and advertising budget expenditures

2.4 Non-Functional Requirements

The new platform should incorporate the following non-functional requirements:
- Comply with data security and subscriber privacy laws and regulations
- Be scalable as user base increases
- Maintain a 99% uptime for gaming platform
- 80% Machine Learning model accuracy

2.5 Financial Benefits and Costs

2.5.1 Benefits

We have seen a decrease in active customers by about 400,000 from 2015 to 2017, and our revenue has also decreased by more than $100 million in the same period. The benefits from implementing this machine learning solution will be an estimated increase in active customers to previous levels by 2019, and a projected growth to 1.65 million users by the year 2020.

The churn rate has unfortunately increased from 6.32 percent in 2015 to 8.12 percent in 2017. After implementation of the project, it is estimated that the rate will decrease to below 2015-levels and projected to decrease to 4 percent.

The most important benefits will be increases in revenue to back to near-2015 levels of $300 million in 2019 and projected revenues of $375 by the year 2020.

EOY	2015	2016	2017		2018	2019	2020
Active Customers	1,250,000	1,100,000	850,000		1,000,000	1,250,00	1,650,000
Avg Customer Spending per year	$200	$160	$140		$175	$200	$225
Conversion Rate (Trial to Paid Subscriber)	25%	23%	15%		20%	25%	30%
Churn Rate	6.32%	7.45%	8.12%		7.5%	6.25%	4%
Revenue ($ million)	$317.5	$279.4	$215.9		$250	$300	$375
Operating Costs (million)	$12.6	$13.1	$13.7		$15.1	$14	$14.5

2.5.2 Costs

The projected development and operation costs for this machine learning project is relatively minimal compared to the projected benefits and returns mentioned previously. The estimated total project cost is only an added $1.4 million, with an estimated yearly cost of $1 million. When these costs are compared with the estimated increases in revenue of $35 million in 2018, $50 million in 2019, and $75 million in 2020, there will be significant return on investment with relatively little initial and annual investment.

Initial Investment Estimated at $100,000/month	Additional Yearly Operational Costs	Estimated Increase in Revenue
1. Integration with Oracle (Finance Data) and Salesforce (Customer Data) using Informatica: $200,000 2. Integrate Game A to C with Informatica and Google Firebase Cloud Messaging: $300,000 Machine Learning Platform - $600,000 3. Deploy Yokozuna Machine Learning SaaS - $300,000 Estimated Total Project Cost: $1.4 million	Infrastructure - $600,000 / year Maintenance - $400,000 / year Estimated Yearly Cost: $1 million	EOY 2018 $35 million EOY 2019 $50 million EOY 2020 $75 million

2.5.3 Metrics for Success

We will measure the success of our project using the following criteria:

Metrics	Current State	Planned State
Increase conversion of trial to paid subscriptions	15% conversion	25% conversion
Decrease churn of paid subscribers	8.4% churn rate	4% churn rate
Increase in in-game purchases	$140 per user per year	$220 per user per year
Minimal increase in operating costs	$13.7 million per year	$14 million per year
Minimal increase in marketing costs	$4.4 million per year	$5.6 million per year

We expect an increase in marketing costs, primarily due to the additional discounted offers and promotions, which will be treated as an expense for the marketing team. Our marketing team will still continue to play an integral part of our business and will represent about 14.5 percent of the company's overall budget, up from 11.4 percent. With the use of our machine learning algorithms and automation, the team will be able to drive our revenues up further, offsetting the increase in marketing costs by increasing our return on investment.

3 Technical Specification and Prototype

3.1 Architectural Overview

ISMT selected Yokozuna Data to provide the machine learning processes. Data collected from existing sources (gaming events, Salesforce data, Oracle database data) will be directed into Hadoop as our system of records. Scripts will gather the specified data entries into a CSV file from Hadoop (via Apache Hive) and uploaded to Yokozuna Data to be processed. Once processed, the CSV file of results from the Yokozuna platform will be uploaded back to Hadoop via Apache Hive. This will be used to create different dashboards using Tableau for marketing campaigns.

Informatica Cloud is being adopted as our data integration layer to unify the different sources of data, perform data transformations into formats required by other applications, and loaded into and stored in Apache Hadoop.

ISMT selected Yokozuna Data to provide the machine learning processes. The platform was selected for the following reasons:

1. Out-of-the-Box Machine Learning Algorithm
2. Easy data upload to Yokozuna Data in CSV format with scripts
3. 80% accuracy with machine learning predictions

Detailed technical specifications for this project are covered in Section 3.2.

3.1.1 System Architecture

The diagram below illustrates the overall architecture of this project:
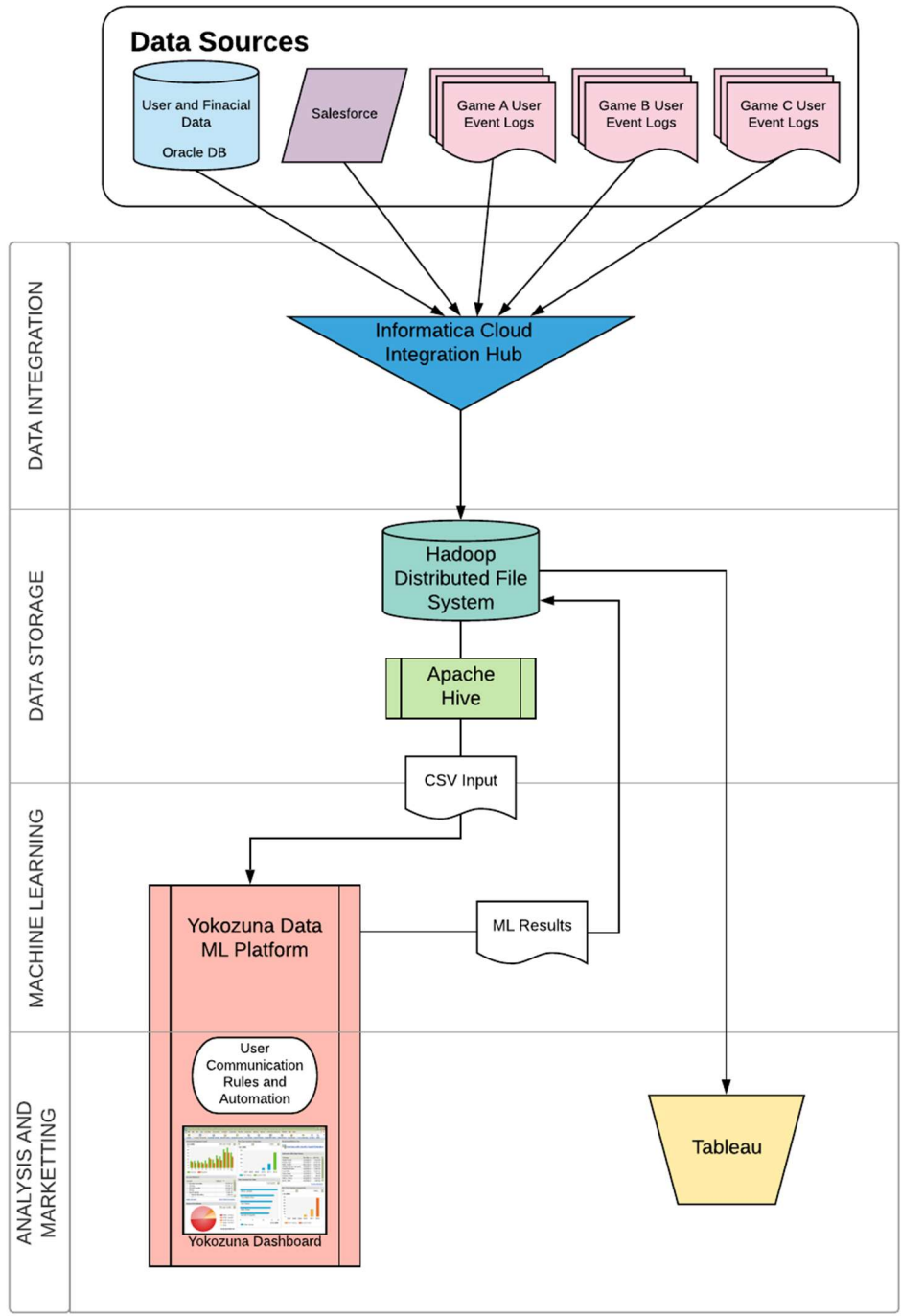
3.1.2 Project Data Flow

The high-level data flow of our project has five major components.

1. User and financial data, Salesforce data, and event data logs generated from user play are collected every 24-hour period to be extracted and transformed by the Informatica Data Integration platform and loads the data into the Hadoop data lake.
2. Hive will be used to extract CSV files for the Yokozuna Data platform where machine learning algorithms are run on the data.
3. The marketing team will use Yokozuna Data's dashboard to review the machine learning results and automate in-game promotions to potential churners.
4. Machine learning results from the Yokozuna Data platform are loaded back into Hadoop (our system of records) which will be connected to Tableau for Business Analytics.

3.2 System Details

3.2.1 Data Sources

Data from four systems provide the information that is fed to the Yokozuna Data machine learning platform.

1. Semi-structured data generated from user gaming activities in the form of event logs.
2. Structured gaming data from the existing backend databases (Oracle database): user and financial data.
3. Structured customer, past marketing data (email and social media), and customer service management statistics from Salesforce.

3.2.2 Games-R-Us Existing Information Management Systems

Games-R-Us current information management systems include:

1. Salesforce: Marketing cloud (Email and Social studios) and customer service management.
2. Oracle Database: Each game will store user information and financial transactions in a relational database.
3. Hadoop Distributed File System: System of records/data lake to store event logs from user game play.
4. R Project: Performing descriptive statistics.
5. Tableau: Dashboard and data visualization.

Financial Data Collected (Oracle DB)	User Data Collected (Salesforce)
- User ID - Purchase History - Purchase Total - Date Subscription Began	- User ID - User Age - User Gender - User Email - User Zip Code - User Country - User Promotion History - User Promotion Conversion

3.2.3 Games-R-Us New Additions to Information Management Systems

New additions to Games-R-Us current information management system includes:

1. Informatica Cloud Integration Hub to extract and transform data from sources and load into the data lake.
2. Yokozuna Data platform to perform Machine Learning algorithm on user data.
3. Apache Hive to load CSV files from our data lake to Yokozuna Data

3.3 Software Solution

3.3.1 Yokozuna Data

For the machine learning component, Games-R-Us chose Yokozuna Data as their Software as a Service (SaaS) provider. While there are many machine learning SaaS providers, most notably Microsoft Azure Machine Learning and Amazon SageMaker, these providers offer only classical machine learning methods such as regression and binary classification. However, through the ISMT's own testing against Yokozuna's research, they have found that these classical methods do not meet their business requirements. In order to apply regression or binary classification to predict churn rate, all players must leave the game to produce a complete dataset. Without a complete dataset to predict churn rate, the accuracy of the machine learning model is greatly reduced.

Yokozuna Data has solved this issue by using Survival Ensembles for their machine learning model. The ISMT has considered expanding the on-premise R Project server (currently used for descriptive statistics) and build the same model, but the amount of ingress data will require heavy in-house investments into hardware, software, system engineers, and data scientists. As such, the ISMT has chosen to outsource this function to Yokozuna Data to help meet Games-R-Us's requirements.

On a daily basis through a batch process, CSV files with user login data, purchases, action, level changes, and event changes will be uploaded to Yokozuna Data. Using this data, Yokozuna Data will analyze and classify each user into the following categories: Alive, Churned, and Potential Churners. The platform also allows us to use Google Firebase Cloud Messaging so that we can automate in-game notifications to offer subscription discounts per section 2.1.2. This is an improvement compared to using email-based marketing campaigns through Salesforce. In addition, these churn results can then be exported back into Hadoop, our data lake for further analysis.

3.3.2 Model Training and Overfitting

In regression analysis, pruning techniques are used to solve the overfitting issue. Pruning is simply removing sections of the decision tree model. Therefore, it prevents the ML model to perform analyses that are too close to the data provided. However, pruning can often affect the accuracy rate of the model as well.

To combat this, Yokozuna uses conditional inference survival ensembles. Survival ensemble is more robust because every tree is used to estimate and differentiate the survival characteristics on every sample (aka Kaplan-Meier Estimator). In addition, linear statistics is applied separate on each node, so that the survival difference is maximized. As a result of this, accuracy rate is not impacted because pruning is not done on the dataset being analyzed.

In order to continue training the Machine Learning model effectively, Yokozuna uses parallel computation where each core/machine trains a subset of the tree, and the results are combined into a single model.

3.3.3 Model effectiveness and strengths

The problem with traditional binary classification and regression is that:
- Binary classification cannot predict time. In our case, when exactly will the player stop playing our game?
- Regression cannot predict how much time is left until the player churns. As mentioned earlier, this is because there is incomplete information on every player because they have not stopped participating in Games-R-Us' gaming platform.

Survival analysis first began as statistical techniques to predict lifetime expectancy in medical and biological research. This is because survival analysis focuses on time until an event occurs. For medical research, the event is either a failure in mechanical systems or death. For Games-R-Us' use, we would like to predict when a player will leave their games, so that retention marketing can be performed.

The key to survival analysis (when compared to binary classification and regression) is censoring, which the algorithm accounts for incomplete information on the occurrence of the event. As a result of that, the survival function is simply trying to predict the likelihood a player will survive. The survival function uses Kaplan-Meier to estimate churn probability from recorded censored survival times, or in layman's term, cumulative survival probability.

In order to evaluate the effectiveness of Survival Ensemble on churn prediction, ISMT tested both binary classification and survival models, similar to the research performed by Yokozuna. The algorithms used included Support Vector Machine (SVM, naive Bayesian classifier, and decision tree). ISMT then compared the AUC (Area under ROC Curve) to determine the accuracy rate of these models (higher AUC = more accurate):

Model	AUC
Survival Ensemble	0.960
Support Vector Machines	0.940
Decision Tree	0.934
Naive Bayesian	0.900

In addition, further evaluation was done between other Survival models such as Cox regression. ISMT concluded that Cox regression had trouble capturing the time between levels, which is not uniform in the population. Survival Ensemble fare much in that there are less outliers when compared against prediction and actual observation in the figure below.
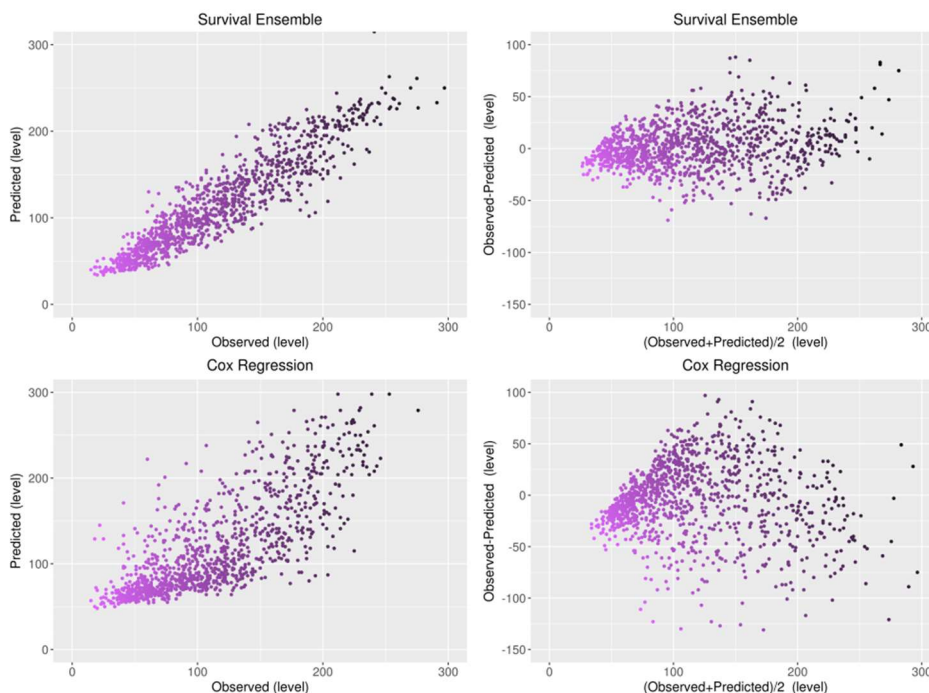
Fig. 2. Predicted median survival level vs. observed level (left) and relative deviation (right) for churned players, using the survival ensemble and Cox regression models

11
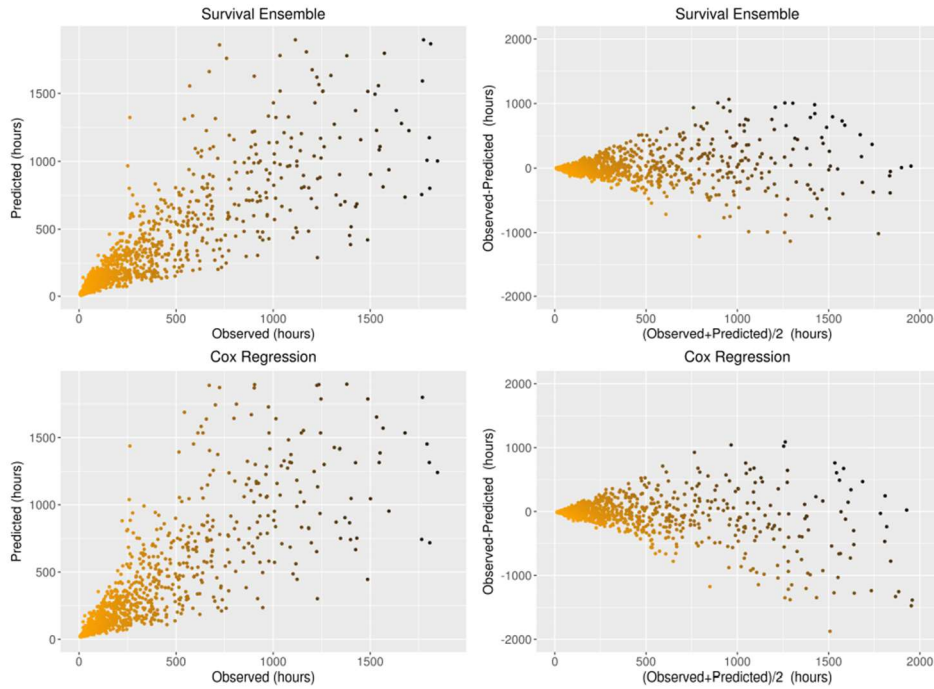
Fig. 3. Predicted median survival playtime vs. observed playtime (left) and relative deviation (right) for churned players, using the survival ensemble and Cox regression models.

ISMT also used Brier score to calculate the prediction error between Cox Regression and Survival Ensembles. Overall, as playtime or level increases, the prediction error is higher with Cox Regression:

Fig. 4. Playtime model (top) and level model (bottom) IBS error curves.

3.3.4 Why Games-R-Us chose Yokozuna

Overall, Yokozuna Data is the product of choice because:

- Their Machine Learning algorithm, survival ensemble machine learning algorithm, was developed and trained on data specifically for the online gaming industry.
- Yokozuna Data's model accuracy – 80% – is acceptable for the ISMT.
- Easy upload of gaming data for analysis to the Yokozuna platform as CSV files.
- Cloud based so that there is no need to develop our own machine learning model.
- Games-R-Us will not need to hire more data scientists to create and train a ML algorithm.
- The Yokozuna platform is customizable where we can add features to the model.
- Additional business analytics variables can be generated from the Yokozuna machine learning results, which help enhance our Tableau reports.
- Yokozuna platform provides a dashboard to manage marketing campaigns. Thus, additional products like Salesforce Email Studio are not needed.
- Yokozuna Data is less expensive than expanding our R Project platform and infrastructure.
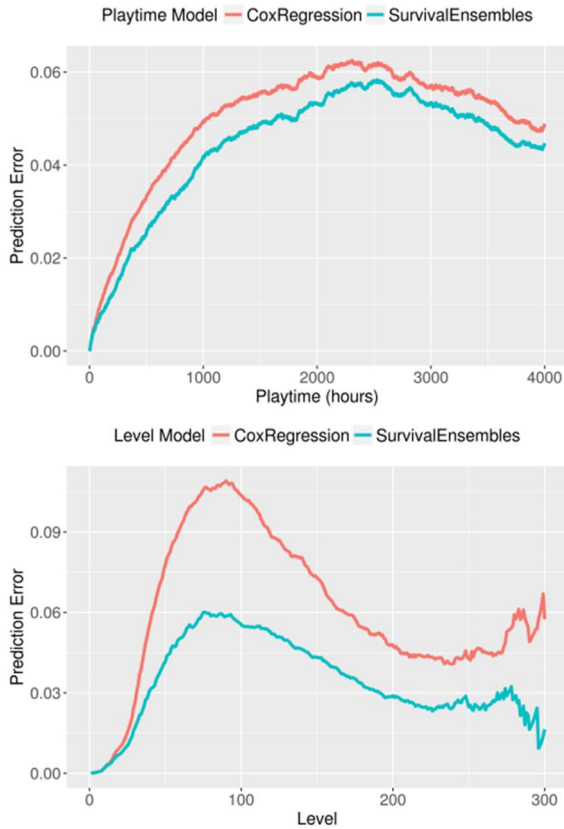
Yokozuna Data's Machine Learning Model has been constantly improving since 2016, using data collected with mobile games hosted by their parent company, Silicon Studio. The dataset used in development goes as far back as 2014. Since then, Yokozuna has acquired additional customers similar to Games-R-Us and used their data to further develop their machine learning model. The amount of data available to Yokozuna far exceeds the limited dataset that Games-R-Us has in-house.

The other group of users that Yokozuna Data will help us identify are the Potential Purchase Churners. The categories include Stopped Spending, Potential Purchase Churners, and Alive. While in-game purchases are not the biggest source of Games-R-Us revenue, Yokozuna Data allows their marketing team to access this list of Potential Purchase Churners and execute actions. For example, sending in-game items, offering in-game purchase discount, etc. However, such action will require integration with our gaming platform.

In order to comply with various data security requirements in various countries. Games-R-Us will adopt the strictest regulations for our environment, which is the European Union's GDPR. In order to comply with GDPR requirements, we plan to upload only hashed values of our user IDs and email addresses to Yokozuna. In addition, while purchase history will be shared with Yokozuna, actual payment information (credit card and bank) and contact information (name, address, and telephone) will remain under our data lake only.

3.4 Integration with Other Applications and Data Sources

3.4.1 Informatica Integration Hub

The new data integration platform interfaces with Salesforce (CRM), Oracle databases, Apache Hadoop Hive and events logs from games. The data sources from all of these applications – structured data from Salesforce, Oracle databases, and semi-structured data from event logs from users' game play – are extracted from each application, transformed, cleansed, and structured in the format found in the Appendix, Section A, and loaded with Informatica Integration Hub to Hadoop Hive. If a transformation rule fails the issue will be escalated with the engineers responsible for instrumenting the events.

Through a rules-based and permission system within Informatica's Data Integration Hub, ISMT and business unit teams can access a virtualized version of the cleaned and transformed data stored in Hadoop. A catalogue of data is available for data scientists and business units in marketing, finance, and sales to select from and use.

The user's System of Record is maintained in Hadoop Hive (see 3.1.1 System Architecture). The data in Hive can be queried directly by the ISMT team and other developers using SQL. The Hadoop data

store will also be connected to Salesforce and Tableau to create dashboards for our Marketing and Business Analytics teams. The platforms and fields that comprise the user System of Record are shown in the diagram below.
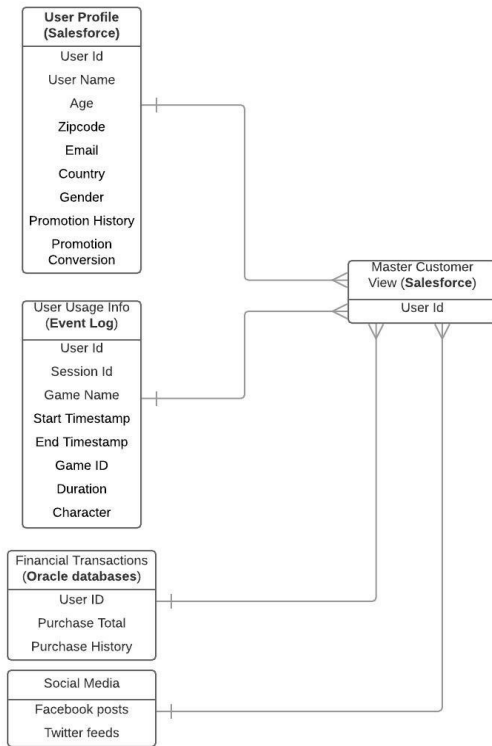
3.4.2 Yokozuna Data Integration

A custom script will query the data from Hadoop Hive and generate a CSV file that will be uploaded to Yokozuna Data. The results from Yokozuna Data's machine learning algorithm is accessible to the marketing department on Yokozuna's platform to generate marketing campaigns. Additionally, the results are downloaded to Hadoop for data scientists to further analyze with business analytics. Yokozuna provides a REST API to upload and download data to their platform. The command below can be used to download the prediction results:

```
curl https://yokozuna.siliconstudio.co.jp/GameAnalytics/api/VipClass \
  -H "Authorization: Key abcdef0123" \
  -d '{"product": "1", "startDate": "2017-04-01"}'
```

To extract data from our Hadoop data store in CSV format, the script will use a Hive command like the one below:

```
hive -e 'select * from your_Table' | sed 's/[\t]/,/g' > /home/yourfile.csv
```

Apache Oozie or other job orchestration frameworks can be used to schedule this process every 24-hours.

3.5 Data Design and Management

3.5.1 Data Entities

The main data entities relevant to our Machine Learning solution to predict customer churn that will be fed into the Yokozuna Machine Learning model are usage and customer data including the demographics of who plays the games (age, gender, location, etc.), number of games played by each subscriber, what types of games they play or do not choose to play, the frequency of their gameplay, their length of play, their social interactions within the games, the level(s) or stage(s) of the games

14

where players stop playing the game, number and type of in-game purchases, the type of subscription the user pays for (monthly, annual, etc.), how long the user has been a subscriber, if their subscription had lapsed in the past, and other relevant player information.

In addition to the user data mentioned above from user input and event logs, the inputs and sources of these data entities are also collected from various sources including social media data sources (i.e. Facebook, Instagram, Twitter, LinkedIn, YouTube, etc.) and customer feedback through our Customer Relationship Management (CRM) system. These various unstructured, semi-structured, and structured data sources will be fed into the Informatica cloud data integration platform, and on to Hadoop.

3.5.2 Data Management

Hadoop Distributed File System (HDFS) and YARN form the data management layer of Hadoop. HDFS provides the scalable, fault-tolerant, cost-efficient storage for our big data, and YARN, as the architectural center of Hadoop, provides the resource management framework that enables our enterprise to process the data in multiple ways simultaneously – for batch, interactive, and real-time data workloads on one shared dataset.

The raw data will be extracted and transformed (cleansed and formatted based on data structures found in the Appendix, Section A) using Informatica for storage into the Hadoop data lake. The data management layer will be handled by Hadoop Hive, which has built-in support for the input and output of CSV files. Using Hive will allow us to create the input files necessary for Yokozuna's ML platform. The machine learning and business analytics results will ultimately be presented on dashboards for marketing with useful information such as users' probability to churn, costs of customer acquisition and retention, average revenue per user, daily active users, and other useful user statistics.

3.5.3 ETL Data Flow
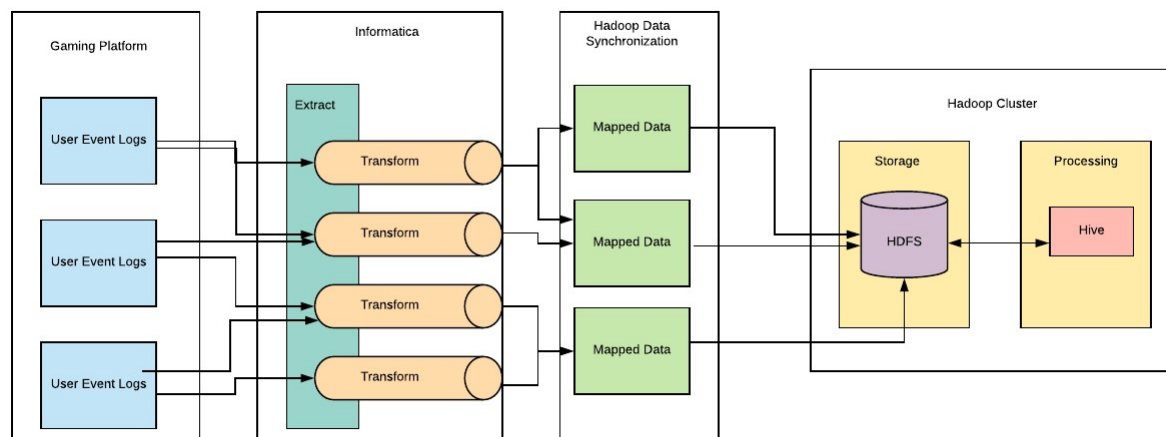
3.5.4 Data Dictionary

The fields contained in the input files for the Yokozuna ML platform are defined in Appendix Section A.

3.6 Solution Demonstration

3.6.1 ETL process demonstration

Through Informatica Cloud platform, we connect to various structured and semi-structured data from Oracle database, Salesforce CRM, Yokozuna Data, and our system of records stored in Hadoop Hive:

3.6.2 Yokozuna Dashboard

After machine learning modeling is performed by Yokozuna, Games-R-Us' marketing team gets to see the following results on Yokozuna's website. These are some of the dashboards which have been created for the marketing/sales team to see the overall health of the business.

Daily Active Users and Average Playtime:
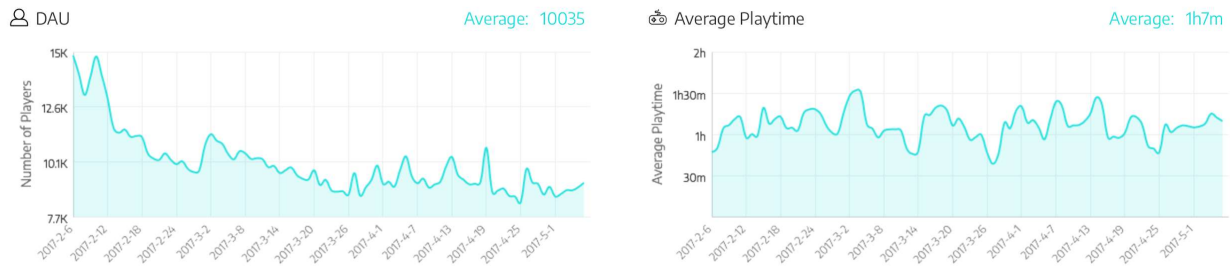
Time of day when subscriber plays:
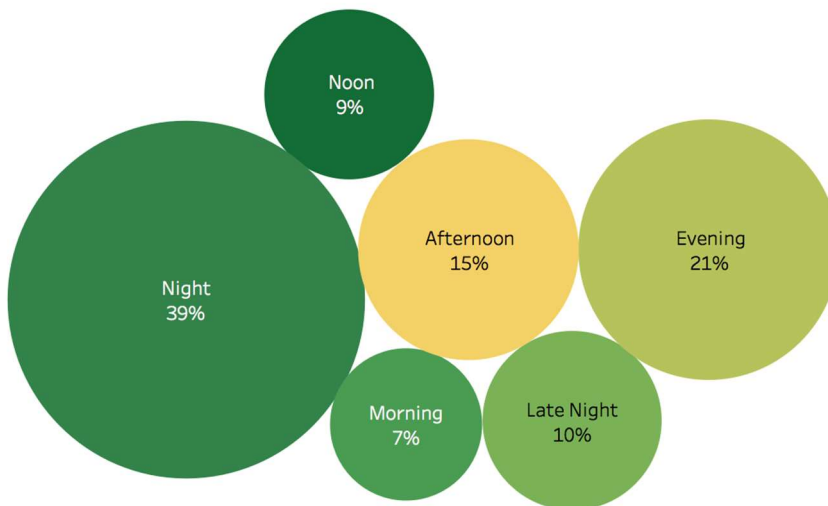
Cumulative Users

📊 Total Users
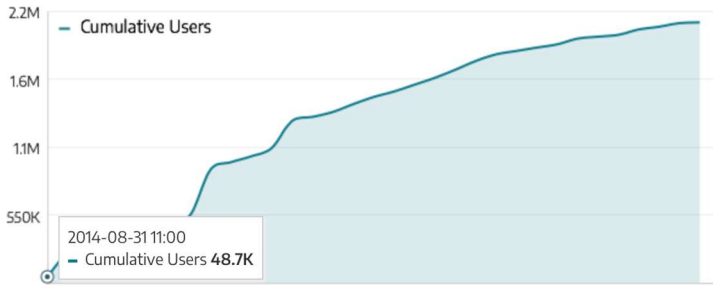
👤 USER COUNT

Date	New User	MAU	Total Users
2017-04	5.8K	16.7K	2.1M
2017-03	29.4K	43.2K	2.1M
2017-02	22K	37.6K	2.1M
2017-01	43.5K	58.3K	2M
2016-12	14.9K	31.5K	2M
2016-11	15.7K	33.3K	2M
2016-10	46.4K	63.5K	2M
2016-09	26.2K	42.8K	1.9M

Survival (Churn) Analysis

The following chart provides information about a particular user, showing the survival probability (churn rate):

2103173	2103173 ID	2017-05-03 Register Date	2017-05-07 Last Login Date	5 days Lifetime	17 days Predicted LifeSpan	level 17 Last Level	level 47 Predicted LevelSpan

17
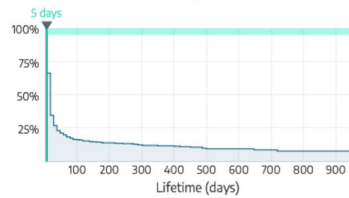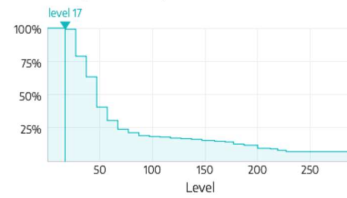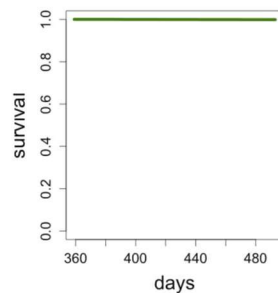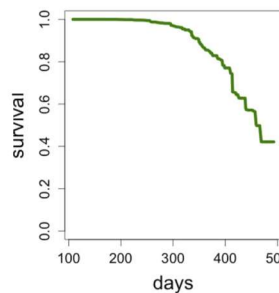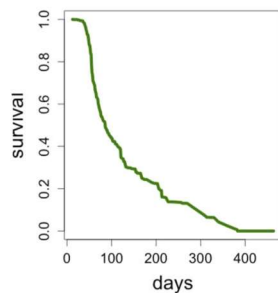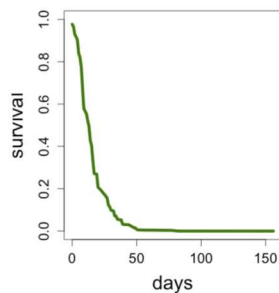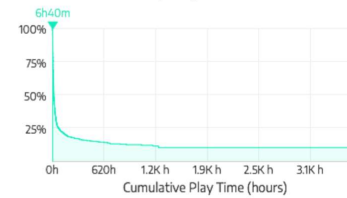
Churn risk assessment:

RISK OF LEAVING THE GAME

☑ High Risk
⤓ 🔔 ⬇

13.86%
333

☑ Mid Risk
⤓ 🔔 ⬇

14.07%
338

☑ Low Risk
⤓ 🔔 ⬇

19.02%
457

☑ No Risk
⤓ 🔔 ⬇

53.06%
1275
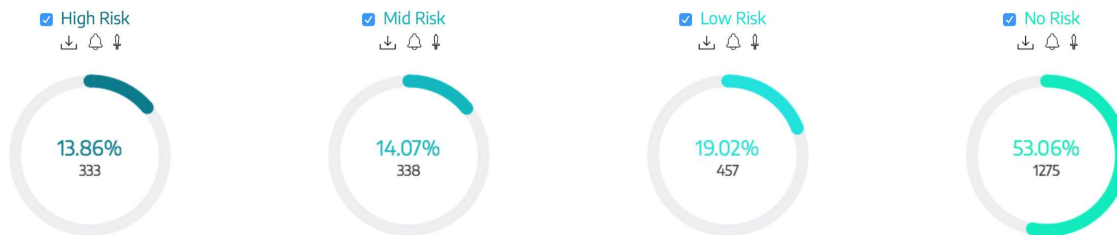
Rules Creation Screen

The marketing team can send the promotional offer either manually or in an automated fashion using the Yokozuna platform. It provides the ability to create these triggers as per the following screenshot.

Automated Campaign

Execute New Action	✕
Target	May Churn (Churn Prediction) (60) ⬍
Date	2018-03-20 ▦
Options	At each player optimal time slot ☑
	Time Since Last Message 7 Days ⬍
Action	Send Item Normal ⬍
Custom Parameters	{"items":[{"count":2,"id":"itemA"},{"count":3,"id":"itemB"}]}
Validate Server Key	

Execute   Cancel

Campaign Management

User Id	Click to Send	
1111123242		No promotion
1111123243		No promotion
1111123244	Click to Send	20% Email Discount
1111123245	Click to Send	35% Email Discount
1111123246		No promotion
1111123247	Click to Send	In app Free Character
1111123248	Click to Send	In app Free Character
1111123249	Click to Send	In app Free Character
1111123250		No promotion
1111123251		No promotion
1111123252	Click to Send	50% Email Discount
1111123253		No promotion
1111123254	Click to Send	2 months free

The same message can be sent multiple times by defining the interval in the rule.
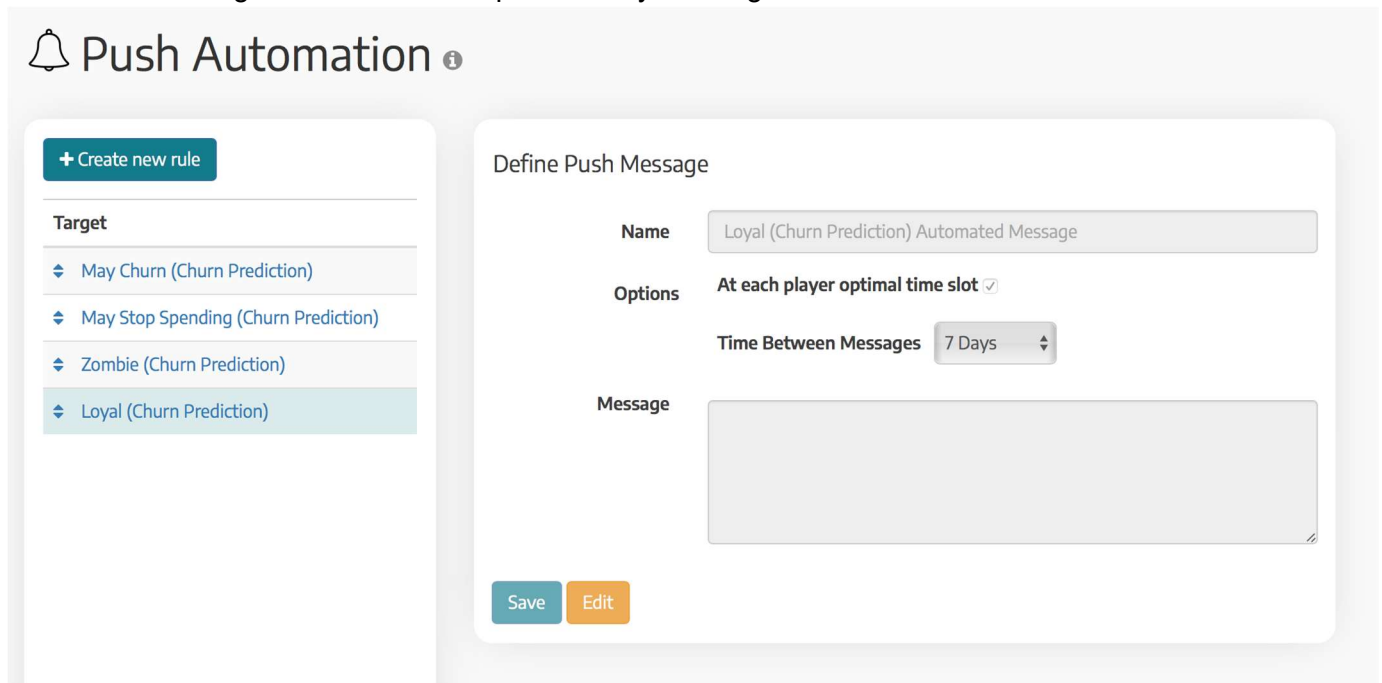
4 Solution Development

4.1 General

The ISMT department will use agile development methodologies to manage the solution deployment. Agile development processes will be employed with daily Scrums to assess each team's progress. ISMT will work with our vendors such as Yokozuna to ensure that their Statement of Work align with our epics schedule and delivery date. Progress will be assessed at the end of each week and the schedule adjusted based on progress made. To kick-off the project, ISMT will have a total of 43 tasks and nine epics to deliver with each epic delivered after a three-week sprint except for the following epics which take two sprints to finish:

- Epic #4 - Integrate data injection pipeline from Game A using Informatica
- Epic #5 - Create automated upload process from Apache Hive to Yokozuna
- Epic #8 and #9 - Integrate Game B and C to our infrastructure

ISMT will incorporate system and user acceptance testing where applicable with each epic. Below are the nine epics planned for delivery:

Epics	Teams	Deliverables	Testing
Epic #1	DBA, ISMT, Data Scientists, Yokozuna	Provide training to Data Scientists and Developers so data model can be built for Oracle and Salesforce. Test connectivity and data model from Informatica to Hadoop using Apache Hive.	System
Epic #2	DBA, ISMT, Data Scientists, Yokozuna	Capture Salesforce data to Hadoop for Yokozuna ML analysis using Informatica.	System
Epic #3	Team 1: DBA, ISMT, Data Scientists Team 2: ISMT, Yokozuna	Capture Oracle DB to Hadoop for Yokozuna ML analysis using Informatica.	System

Epic #4	Team 1: DEV team 1, DBA, Data Scientists Team 2: DEV team 2, Yokozuna, ISMT	Integrate data injection pipeline from Game A using Informatica	System
Epic #5	DBA, DEV, Yokozuna, Data Scientists, Marketing	Create automated upload process from Apache Hive to Yokozuna Data; Yokozuna to run quality control algorithm to ensure data compatibility with their Machine Learning Model	System and UAT
Epic #6	DEV, ISMT	Integrate Game A with Google Firebase Cloud Messaging. Yokozuna platform is live with Game A.	UAT
Epic #7	DBA, DEV, Yokozuna, Marketing, ISMT, Data Scientists	Create automated download process from Yokozuna Data to Apache Hadoop through Hive	UAT
Epic #8	DEV team, Yokozuna, DBA, Data Scientists	Integrate Game B with Informatica, Yokozuna, and Google Firebase Cloud Messaging. Yokozuna platform is live with Game B.	System and UAT
Epic #9	DEV team, Yokozuna, DBA, Data Scientists	Integrate Game C with Informatica, Yokozuna, and Google Firebase Cloud Messaging. Yokozuna platform is live with Game C.	System and UAT

In Epic #1, training will be provided to data scientists and developers so a data model can be built for Oracle and Salesforce. Connectivity and the data model between Informatica and Hadoop will be tested by data scientists using Apache Hive.

With Epics #2 and #3, Games-R-Us database administrators will create and implement the Informatica workflow which will include cleansing the data and applying business rules on Salesforce and Oracle data. Simultaneously, ISMT, and Yokozuna developers will work together to develop the features necessary to analyze the data from these data sources as part of Yokozuna's machine learning model. Data scientists will perform system testing by sampling data integrity between Salesforce and Oracle databases, and Hadoop.

In Epic #4, two development teams will work in parallel. Team 1 will migrate the old game data stream to Informatica Cloud. Team 2 will retrofit Game A to capture all the necessary data and events for the Yokozuna machine learning model. Parallel testing will be used for the old data stream and Informatica Cloud. The goal is to retire the old data stream once ISMT confirms that the new Informatica data injection pipeline is functioning correctly.

In Epic #5, the database administrators, development team, and Yokozuna will work together to create scripts to automate the daily export of data in a CSV format using Apache Hive to the Yokozuna platform with their API. Yokozuna will load the data and run a quality control algorithm to ensure data integrity. Data scientists will perform sample testing between the Games-R-Us data lake and Yokozuna's QC data to ensure integrity is not compromised. In addition, the Marketing department will perform User Acceptance Testing on Yokozuna's dashboard using existing data and new gaming data from Game A.

In Epic #6, the Marketing department will begin to monitor the churn rate while the Development Team will add the Google Firebase Cloud Messaging to Game A. This will allow the Marketing department to begin sending in-game discounts and in-game items towards the end of this epic.

In Epic #7, the database administrator, Development Team, and Yokozuna will also create scripts to export the machine learning results from Yokozuna back to Hadoop for storage. Our Marketing Team will verify that the Tableau dashboard is collecting this new machine learning data from Hadoop to track the results.

Epics #8 and #9 focus on patching Game B and C, one after the other, in a process similar to that employed for Game A. Data scientists will ensure the proper data is captured and analyzed, and the Marketing department will monitor the churn and send in-game subscription discounts or items. The results will be reviewed and tracked on the Tableau dashboard for management reporting.

4.1.2 Project Management/Timeline

Appendix, Section B contains the timeline for the integration of the Informatica Cloud and Yokozuna Data platforms with Games-R-Us' existing infrastructure. To view the timeline, sprint schedules with Epics and Stories, please go to the Appendix. Games-R-Us will begin the project in June 2018 and they will complete by Q1, 2019.

4.2 Operationalization

4.2.1 Service Level Agreement - Yokozuna Data

The agreement covers the provision and support of the services provided by Yokozuna Data.

4.2.1.1 Service Description

Yokozuna Data will provide Games-R-Us with machine learning analysis, as well as a dashboard to view results and manage marketing communications for our customers.

4.2.1.2 Processing Time

Yokozuna Data will begin processing the data sent from Games-R-Us daily at 9:00 AM EST (1:00 PM UTC).

4.2.1.3 Service Availability

The Yokozuna Data dashboard will have a required availability of 99.5 percent uptime, not including planned maintenance times. Maintenance of the dashboard that will require downtime should be performed during the agreed upon maintenance window of 6:00 AM to 8:00 AM on the first Sunday of every month.

The machine learning service provided by Yokozuna Data will have a required availability of 99.995 percent during the agreed upon processing times. If scheduled maintenance will delay processing times by more than one hour, Yokozuna Data must notify Games-R-Us of the possible delay at least one week in advance.

4.2.1.4 Service Performance

The initial analysis of our entire customer base should take no longer than 17 hours to process for one million users. Subsequent analysis that will be conducted daily during our agreed upon processing time should take no longer than 60 minutes for 10 million users. The service should scale linearly as our customer base grows.

4.2.1.5 Data Privacy and Protection

The user data provided to Yokozuna will be kept anonymized throughout all layers of processing. Yokozuna may use the data provided by Games-R-Us to improve the accuracy of their machine-learning models going forward. Yokozuna Data agrees to never share or sell our user data to any external parties.

4.2.1.6 Change Management Procedures

Any proposed changes by Yokozuna Data must be submitted to the Change Advisory Board for review. If the change will have an impact on applications or services owned by Games-R-Us, a reasonable timeframe shall be agreed upon to support the change.

4.2.2 Operational Level Agreement

4.2.2.1 Production Support

The point of contact for users of applications in scope of this project will be through the Games-R-Us Service Desk. Incident Management tickets should be created for issues encountered in production.

Internal Site: https://gamesrus.servicenow.com

The Games-R-Us production support team will escalate incidents to the appropriate application team as necessary.

4.2.2.2 Support Hours

Users can expect support to be available during all regular business hours, as well as extended hours during Yokozuna Data's processing times. Support cannot be expected on weekends or holidays.

Additionally, the hours of 6:00AM to 8:00 AM on the first Sunday of every month are reserved for system maintenance, and prior notification will occur if any services will be out during those times.

4.2.3 Third-Party Licenses and Support Agreements

For the successful implementation of this project, Games-R-Us will have licensing agreements with the third-parties outlined in this section for the use and deployment of their services or applications. Contracts between Games-R-Us and the third-parties may be renegotiated six-months prior to contract expiration. Any third-party systems or services not mentioned in this section, but utilized for this project, are considered open-source, and Games-R-Us agrees to the open-source licensing agreements of each codebase. The links to each agreement can be viewed in Appendix Section C.

4.3 User Enablement

4.3.1 Implementation Plan

The implementation plan includes testing, monitoring, automation, training, and continual adjustments, modifications, and improvement. First, the data scientists and developers will use past and current user profile and gaming data with the Yokozuna Data machine learning platform. While the goal is to have automated marketing emails, initially the marketing team and our own data scientists will monitor Yokozuna's Machine Learning AI models' results and predictions, and will manually send out emails. Also to be monitored is the integration with the Informatica Cloud data

integration hub/layer. Once the results and the individual player behavioral and churn predictions are deemed satisfactory and accurate, the system can be fully automated, significantly increasing savings in hours spent on emailing tasks. The system's effectiveness will be monitored, documenting bottlenecks, problems, and areas for improvement.

4.3.2 User Acceptance, Training & Knowledge Base

4.3.2.1 User Acceptance

User acceptance will be included in the implementation plan by appointing representative testers among the user community running test scenarios according to the corresponding critical user acceptance criteria described in the User Stories Acceptance Criteria in Section 2.2. The implementation plan beyond the testing phase will also include extensive training and continuous improvement through monitoring every step of implementation.

4.3.2.1.1 Marketing Managers

The Marketing Managers' role is to use marketing campaigns and send promotional offers to prevent customer churn. With the old system, Marketing Managers simply sent out mailers and promos according to a set schedule to all trial users. Using the new system utilizing the Yokozuna dashboard, Marketing Managers are able to view a list of customers (trial or existing) who are most likely to churn, and based on customer metrics, they will customize their marketing campaigns with improved, user-specific promotional offers. With continual improvement, Marketing Managers will test if the predictions provided by Yokozuna platform are accurate or not.

4.3.2.1.2 Product Managers

The Product Managers' role will be to test the Tableau dashboard functionality and will approve if the dashboard data is correct. Before using Tableau, useful information from data and visualizations had to be created manually from the Oracle database.

4.3.2.2 Training & Knowledge Base

The ISMT team members that implement the new system will train team members to be "experts" on the system which includes the Informatica Integration Hub user interface and the Yokozuna dashboard. There will always be "experts" on hand on every shift, 24/7 to provide support and answer questions. Representatives from Informatica, Yokozuna, the ISMT implementation team, and "experts" will conduct training sessions described below. In addition to formal training, a knowledge base will be created to document any problems and issues, and as a source of FAQs and answers.

4.3.2.2.1 Data Scientists and Developers

Data scientists and developers will have the most extensive initial training and will be required to attend two weeks of training on the Informatica data integration software and Yokozuna data platform during Solution Development Epics 1 and 2. Informatica training will include curating the data; leveraging the data lake through data processing and storage on Hadoop with the data structure; governing the data integration; managing the hybrid workflows and integration across cloud apps, analytics, big data, and on-premise systems; and the user interface. Yokozuna training will include optimizing what-if scenarios; the sequence, frequency, and timing of game events; optimizing item-purchase recommendations within games; recognizing playing habits within games; and assessing the churn prediction model using Survival Ensemble.

Marketing managers will be required to attend 2-3 days of training to use the Yokozuna data predictive game analytic software dashboard to predict customer LifeSpan and churn, and coordinate and optimize the best marketing efforts based on gamer behavior. They will also be trained to use the Informatica user interface for self-service access to data.

4.3.2.2.3 Product Managers

Product managers will be required to attend one week of training on the use of Yokozuna, the dashboard, as well as ways that the software can be customized and improved to help the marketing team, as well as using Informatica.

4.3.2.2.4 Service Desk (Help Desk) Operators

Service desk (help desk) operators will be required to attend 2-3 days of training on the Yokozuna and Informatica platforms, user support, the types of questions that users may have, the types of problems that may arise, the differences between how to do something on the old system versus the new one, and how to maintain the knowledge base.

4.3.3 User Access

User access will be given according to one's role.

4.3.3.1 Data Scientists and Developers

The data scientists and developers will have full administrative access to the Informata and Yokozuna systems including reading and writing to the data and the dashboards to monitor effectiveness and make adjustments and modifications to optimize gamer playtime and retention, and to enhance player-focused data-driven development.

4.3.3.2 Marketing Team

The marketing team will be registered users of the Yokozuna and Informatica dashboards with full user rights to data analysis and marketing information. They will also have read rights to user financial and game-play data.

4.3.3.3 Service Desk (Help Desk) Operators

Although Service Desk operators are not marketing, finance, or sales, they should be experts in the user interface and dashboards of Salesforce, Yokozuna, and Tableau to be able to help users and troubleshoot. Thus, they will have the same rights as all of the users that they assist.

4.3.3.4 Business Units in Marketing, Finance & Sales

Business units in marketing, finance, and sales will have read access to the catalogue of data to select from to use and analyze, including increasing game monetization. The Yokozuna dashboard as well as the Informatica user interface will provide these users with full user-rights self-service access to data to find topics, and self-subscribe to data they need on the schedule they need. Through proper training and guidance of experts and the Service Desk, these business unit users can use the dashboards to organize access to the data into useful analytics and actionable information.

User Access

	Marketing Team	**Data Scientists & Developers**	**Service Desk (Help Desk) Operators**	**Business Units in Finance & Sales**
User & Financial Data	Read	Read, Write &	Read	Read, Write,

		Execute		Execute
Salesforce Data	Read, Write	Read, Write & Execute	Read	Read, Write
Game User Event Logs	Read	Read, Write & Execute	Read	Read
Informatica Cloud Integration Data	Read	Read, Write & Execute	Read	Read
Hadoop Distributed File System Data	---	Read, Write & Execute	---	---
Apache Hive Data	---	Read, Write & Execute	---	---
Machine Learning Data	Read	Read, Write & Execute	Read	Read
Analysis & Marketing Data in Yokozuna	Read, Write, Execute	Read, Write & Execute	Read, Write & Execute	Read, Write & Execute
Analysis & Marketing Data in Tableau	Read, Write, Execute	Read, Write & Execute	Read, Write & Execute	Read, Write & Execute

4.4 Success Metrics

4.3.1 Customer Churn

A dashboard is used to track the customer churn which is created by the ISMT team. The target is to reduce the customer churn of existing subscribers. A decrease of two percentage points is considered a success of the project.

4.3.2 Incentive Program Success

To measure if the incentive program is working, the Yokozuna platform, which provides various reports to track this, will be used. To start, the target is to increase the conversion rate of new customers by 5 percent. Slowly this target will be increased to measure the success of the project. The other way this will be measured is to check whether the in-app purchase recommendation increases the conversion rate. The target is to increase the in-app purchase by ten percent.

4.3.3 Model Accuracy

Model accuracy is provided by Yokozuna. It is measured by the total number of correct conversions predicted by the model divided by the sum of total number of correct and incorrect conversions predicted by the model. Accuracy of 80 percent or greater is considered a success. It is expected that the accuracy will increase as we collect more data and tune the model going forward. The dataset is divided into a training set and test set. The model is built on the training set and then the test set is used to test the trained model using the test data. Confusion matrix is used to measure the accuracy.

Appendix

Section A

<u>usersLoggedIn-yyyy-MM-dd.csv:</u>

Field	Type	Description
userId	text	Unique user ID per game title
name	text	User Name
level	int	User's level (or rank, progression, account level)
clanID	text	Clan/guild/organization the user belongs to
registerDate	datetime	Registration date in format yyyy-MM-dd HH:mm:ss
lastLogin	datetime	Login date in the format yyyy-MM-dd HH:mm:ss
platform	text	Platform user is using, e.g. "Google", "Apple"
advertisingID	text	User's device advertising ID, IDFA (iOS)/AAID (Android)
regToken	text	Token for push notifications (Firebase, OneSignal..)
email	text	User's email address
originID	text	Network ID of the place where user acquired
userTimeZone	timezone	User Timezone relative to UTC
userLanguage	text	User Language Settings, e.g. "en_US", "en-US",
location	json	User Location with optional keys "Latitude", "Longitude"
demographics	json	User Demographics with optional keys "Age", "Gender"

purchases-yyyy-MM-dd.csv:

Field	Type	Description
userId	text	Unique user ID per game title
characterID	text	Unique character ID in case there are multiple characters
date	datetime	Purchase date format yyyy-MM-dd HH:mm:ss
itemID	text	IAP (In-App Purchase) item name, type of purchase made, or gacha ID, including converting to in-game money.
price	double	Price of the item

platform	text	Platform user is using, e.g., "Google", "Apple"

actionLog-yyyy-MM-dd.csv:

Field	Type	Description
userId	text	Unique user ID per game title
characterID	text	Unique character ID in case there are multiple characters
date	datetime	Action date format yyyy-MM-dd HH:mm:ss
actionID	text	ID of the general action performed
actionName	text	General action name, e.g, "LoggedIn", "obtainedItem", "startDuel", "endDuel"
actionDetail	json	Json with the action details, with keys e.g. "targetID", "targetLevel", "targetType"

levelInfoChanges-yyyy-MM-dd.csv:

Field	Type	Description
userId	text	Unique user ID per game title
characterID	text	Unique character ID in case there are multiple characters
date	datetime	Level change date format yyyy-MM-dd HH:mm:ss
level	int	New level after level-up

ID	Task Name	Duration
1	Epic #1: Deploy Apache Hive and test connectivity from Informatica to Apache Hive	15 days
2	Deploy Apache Hive	7 days
3	Connect Informatica to Apache Hive	3 days
4	Load test data from Informatica to Apache Hive	5 days
5	ISMT to test to ensure data analyzed is correctly loaded on Apache Hive	2 days
6	Epic #2: Capture Salesforce data to Hadoop for Yokozuna ML analysis	16 days
7	Work with Yokozuna to use Salesforce data as one the features in Yokozuna's ML model	16 days
8	Connect Salesforce to Informatica Cloud	5 days
9	Create Business rules in Informatica for Salesforce data to ensure no duplicate records or missing data	8 days
10	ETL data from Salesforce to Informatica	5 days
11	Sample records between Salesforce and Hadoop test data integrity	2 days
12	Epic #3: Capture Oracle DB data to Hadoop for Yokozuna ML analysis	15 days
13	Work with Yokozuna to use Oracle data as one the features in Yokozuna's ML model	15 days
14	Connect Oracle DB to Informatica Cloud	3 days
15	Create Informatica rules for Oracle DB data	7 days
16	ETL data from Oracle DB to Informatica	3 days
17	Sample records to test data integrity between Oracle and Hadoop	2 days
18	Epic #4: Integrate data injection pipeline from Game A using Informatica	50 days
19	Retrofit Game A to capture data/events through Informatica for Yokozuna ML model	10 days
20	ISMT to perform testing on Game A with Yokozuna to ensure intended data is being captured using Informatica Cloud	5 days
21	Migrate old stream and Informatica Cloud with parallel testing	10 days
22	Create Business Rules in Information for Game A data/events	7 days
23	ETL data from Game A using Informatica	3 days
24	Perform Parallel Testing with both the old data stream and Informatica Cloud	9 days
25	Epic #5: Create automated upload process from Apach Hive to Yokozuna Data	24 days
26	Create scripts to automate daily export of CSV from Apache Hive to Yokozuna	11 days
27	Load required data in CSV format from Apache Hive to Yokozuna Data	3 days
28	Run quality control algorithm to ensure compatibility with machine learning algorithm (Yokozuna)	4 days
29	ISMT to perform sample testing after Yokozuna's quality control to ensure integrity was not compromised during Yokozuna's QC process.	6 days
30	Epic #6: Integrate Game A with Google Firebase Cloud Messaging	15 days
31	Connect Game A to Google Firebase Cloud Messaging to deliver in-game subscription offer	10 days
32	Marketing department to perform Google Firebase Cloud Messaging feature on Game A	4 days
33	Deploy Google Firebase Cloud Messaging Patch For Game A	1 day
34	Epic #7: Create automated download process from Yokozuna Data to Apache HADOOP through Hive	15 days
35	Create scripts to automate daily export of machine learning results from Yokozuna back to Apache Hive	5 days
36	Create scripts to load machine learning results from Apache Hive to Hadoop for Storage	5 days
37	Update Tableau dashboard to incorporate machine learning results as part of management reporting	3 days
38	Marketing Team to review Tableau to ensure reporting meets team's user stories	2 days
39	Epic #8: Integrate Game B with Informatica, Yokozuna, and Google Firebase Cloud Messaging	50 days
40	Retrofit Game B to capture data/events through Informatica for Yokozuna ML model	10 days
41	Setup data injection pipeline from Game B using Informatica	1 day
42	Connect Game B to Google Firebase Cloud Messaging to deliver in-game subscriptions	10 days
43	ISMT to perform testing on Game B with Yokozuna to ensure intended data is being captured using Informatica Cloud	4 days
44	Marketing department to perform Google Firebase Cloud Messaging feature on Game B	4 days
45	Deploy Patch for Game B	1 day
46	Epic #9: Integrate Game C with Informatica, Yokozuna, and Google Firebase Cloud Messaging	30 days
47	Retrofit Game C to capture data/events through Informatica for Yokozuna ML model	10 days
48	Setup data injection pipeline from Game C using Informatica	1 day
49	Connect Game C to Google Firebase Cloud Messaging to deliver in-game subscriptions	10 days
50	ISMT to perform testing on Game C with Yokozuna to ensure intended data is being captured using Informatica Cloud	4 days
51	Marketing department to perform Google Firebase Cloud Messaging feature on Game C	4 days
52	Deploy Patch for Game C	1 day

Epic number	Epics title/Deliverables	1st set of stories (simultaneous)	2nd set of stories (simultaneous)	3rd set of stories (simultaneous)	System Testing	User Acceptance Testing	Deployment
Epic #1	Deploy Apache Hive and test connectivity from Informatica to Apache Hive	Train Developers and Data Scientists (DBA, Data Scientists, and Yokozuna 3 days)	Deploy Apache Hive (DBA and ISMT team / 7 days)	Connect Informatica to Apache Hive & load test data from Informatica to Apache Hive (DBA / 3 days)	Data Scientists to load test data to ensure data mapping is correct from Informatica to Hive (Data Scientists / 2 days)		
Epic #2	Capture Salesforce data to Hadoop for Yokozuna ML analysis	Connect Salesforce to Informatica Cloud (DBA – 3 days)	Create Business rules in Informatica for Salesforce data to ensure: no duplicate records and no missing data (DBA – 7 days)	ETL data from Salesforce using Informatica (DBA – 3 days)	Sample records between Salesforce and Hadoop to test data integrity (Data Scientists / 2 days)		
		Work with Yokozuna to use Salesforce data as one of the features in Yokozuna's ML model (ISMT and Yokozuna - 15 days)					
Epic #3	Capture Oracle DB to Hadoop for Yokozuna ML analysis	Connect Oracle DB to Informatica Cloud (DBA - 3 days)	Create Business rules in Informatica Oracle DB to ensure: no duplicate records and no missing data (DBA – 7 days)	ETL data from Oracle DB using Informatica (DBA – 3 days)	Sample records to test data integrity between Oracle and Hadoop (Data Scientists / 2 days)		
		Work with Yokozuna to use Oracle data as one of the features in Yokozuna's ML model (ISMT and Yokozuna - 15 days)					
Epic #4	Integrate data injection pipeline from Game A using Informatica	Migrate old data stream using Informatica Cloud (Dev team / 10 days)	Create Business Rules in Informatica for Game A data (Dev Team and DBA 7 days**)**	ETL data from Game A using Informatica (DBA - 3 days)	Perform Parallel Testing with both the old data stream and Informatica Cloud (ISMT / 9 days**)**		Old data stream to be retire once ISMT confirm that data injection pipeline from Game A to Informatica is functioning (ISMT / 1 days)
		Retrofit Game A to capture data/events through Informatica for Yokozuna ML model (Dev team and			ISMT to perform testing on Game A with Yokozuna to ensure intended data is being captured		Deploy Patch for Game A (Dev team / 1 day)

1

		Yokozuna / 9 days)			using Informatica Cloud. (ISMT team / 5 days)		
Epic #5	Create automated upload process from Apache Hive to Yokozuna Data	Create scripts to automate daily export of CSV from Apache Hive to Yokozuna (DBA, Dev team, Yokozuna / 11 days)	Load required data in CSV format from Apache Hive to Yokozuna Data (DBA, Yokozuna / 3 days)	Yokozuna to run quality control algorithm to ensure compatibility with their ML model (Yokozuna / 4 days)	ISMT to perform sample testing after Yokozuna's quality control to ensure integrity was not compromised during Yokozuna's QC process (Data Scientists / 6 days)	Marketing department to perform UAT on Yokozuna's dashboard using uploaded existing data and new gaming data from Game A (Marketing Team / 6 days)	Yokozuna dashboard is now in production for Game A
Epic #6	Integrate Game A with Google Firebase Cloud Messaging	Connect Game A to Google Firebase Cloud Messaging to deliver in-game subscription offer (Dev team / 10 days)				Marketing department to perform Google Firebase Cloud Messaging feature on Game A (ISMT / 4 days)	Deploy Google Firebase Cloud Messaging Patch for Game A (Dev team / 1 day)
Epic #7	Create automated download process from Yokozuna Data to Apache HADOOP through Hive	Create scripts to automate daily export of machine learning results from Yokozuna back to Apache Hive (DBA, Dev team, Yokozuna / 5 days)	Create scripts to load machine learning results from Apache Hive to Hadoop for storage (DBA, Dev team, Yokozuna / 5 days)	Update Tableau dashboard to incorporate machine learning results as part of management reporting (Marketing team / 3 days)		Marketing Team to review Tableau to ensure reporting is meeting team's user stories (ISMT / 2 days)	See Tableau dashboard in 3rd set of stories/tasks
Epic #8	Integrate Game B with Informatica, Yokozuna, and Google Firebase Cloud Messaging	Retrofit Game B to capture data/events through Informatica for Yokozuna ML model (Dev team and Yokozuna / 10 days)	Setup data injection pipeline from Game B using Informatica (DBA / 1 day)	Connect Game B to Google Firebase Cloud Messaging to deliver in-game subscription (Dev team / 10 days)	ISMT to perform testing on Game B with Yokozuna to ensure intended data is being captured using Informatica Cloud. (Data Scientists / 4 days)	Marketing department to perform Google Firebase Cloud Messaging feature on Game B (ISMT / 4 days)	Deploy Patch for Game B (Dev team / 1 day)
Epic #9	Integrate Game C with Informatica, Yokozuna, and Google Firebase Cloud Messaging	Retrofit Game C to capture data/events through Informatica for Yokozuna ML model (Dev team and Yokozuna / 10 days)	Setup data injection pipeline from Game C using Informatica (DBA / 1 day)	Connect Game C to Google Firebase Cloud Messaging to deliver in-game subscription (Dev team / 10 days)	ISMT to perform testing on Game C with Yokozuna to ensure intended data is being captured using Informatica Cloud. (Data Scientists / 4 days)	Marketing department to perform Google Firebase Cloud Messaging feature on Game C (ISMT / 4 days)	Deploy Google Firebase Cloud Messaging Patch for Game C (Dev team / 1 day)

Section C

Informatica	https://www.informatica.com/content/dam/informatica-com/global/amer/us/collateral/other/informatica-license-services-agreement.pdf
Oracle	https://docs.oracle.com/cd/E11882_01/license.112/e47877.pdf
Salesforce	https://www.salesforce.com/content/dam/web/en_us/www/documents/legal/salesforce_MSA.pdf
Tableau	https://onlinehelp.tableau.com/current/server/en-us/license_server_overview.htm

References:

Bertens, Paul, Anna Guitart, and África Periáñez. "Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model." Silicon Studio, 2017. http://www.yokozunadata.com/research/MultiDimChurnPrediction.pdf

Informatica. "Industry's Next-Generation Integration Platform as a Service (iPaaS)." Informatica, 2018. https://www.informatica.com/products/cloud-integration.html#fbid=NT5gtm1Rv8L

Leverenz, Lefty. "Apache Hive." Apache, January 20, 2018. https://cwiki.apache.org/confluence/display/Hive/Home

Periáñez, África, Alain Saas, Anna Guitart, and Colin Magne. "Churn Prediction in Mobile Social Games: Towards a Complete Assessment Using Survival Ensembles." Silicon Studio, 2016. http://www.yokozunadata.com/research/churn.pdf

Wikipedia. "Kaplan-Meier Estimator." Wikipedia, April 29, 2018. https://en.wikipedia.org/wiki/Kaplan%E2%80%93Meier_estimator

Wikipedia. "Overfitting." Wikipedia, May 2, 2018. https://en.wikipedia.org/wiki/Overfitting

Wikipedia. "Pruning (Decision Trees)." Wikipedia, February 22, 2018. https://en.wikipedia.org/wiki/Pruning_(decision_trees)

Yokozuna. "Democratizing Behavioral Data Science: Yokozuna Data, a State-of-the-Art Machine Learning Engine that Predicts the Behavior of Each Individual User." Silicon Studio, 2018. http://www.yokozunadata.com

Yokozuna. "Player Performance." Silicon Studio, 2018. https://yokozuna.siliconstudio.co.jp/docs/features-guide/analytics/player-performance/