

Loan Risk Analyses in Databricks

Presented By:

Jon Gore
Tom Weinandy
Colby Ford





Jon Gore

Account Lead – Financial Services



Tom Weinandy

Data Scientist



Colby Ford

AI Architect



BLUEGRANITE

Solution Areas

cloud scale analytics



Data Platform

Cloud-based, scalable solutions



BI & Analytics

Self-service and enterprise



AI & Machine Learning

Leading edge data science

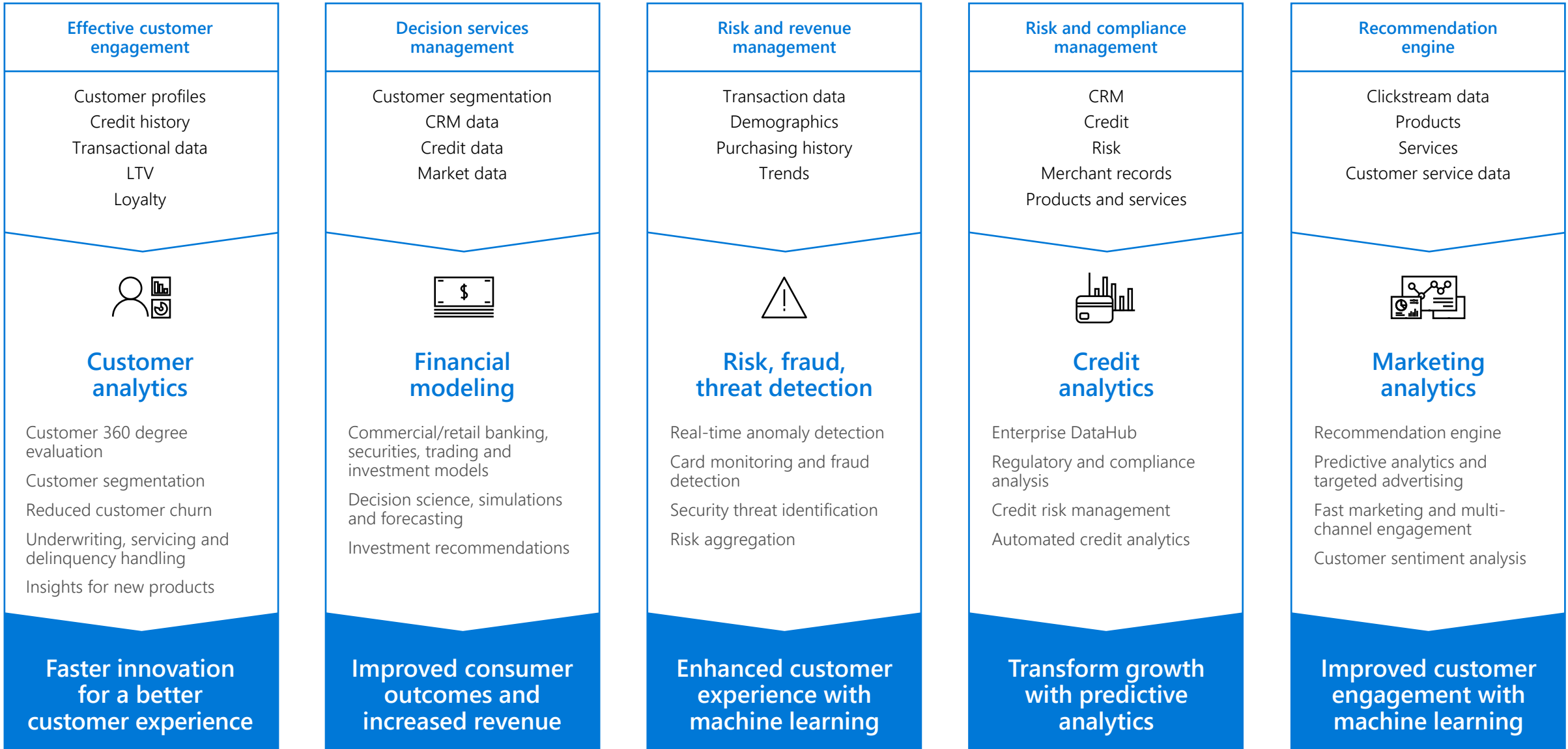


BLUEGRANITE



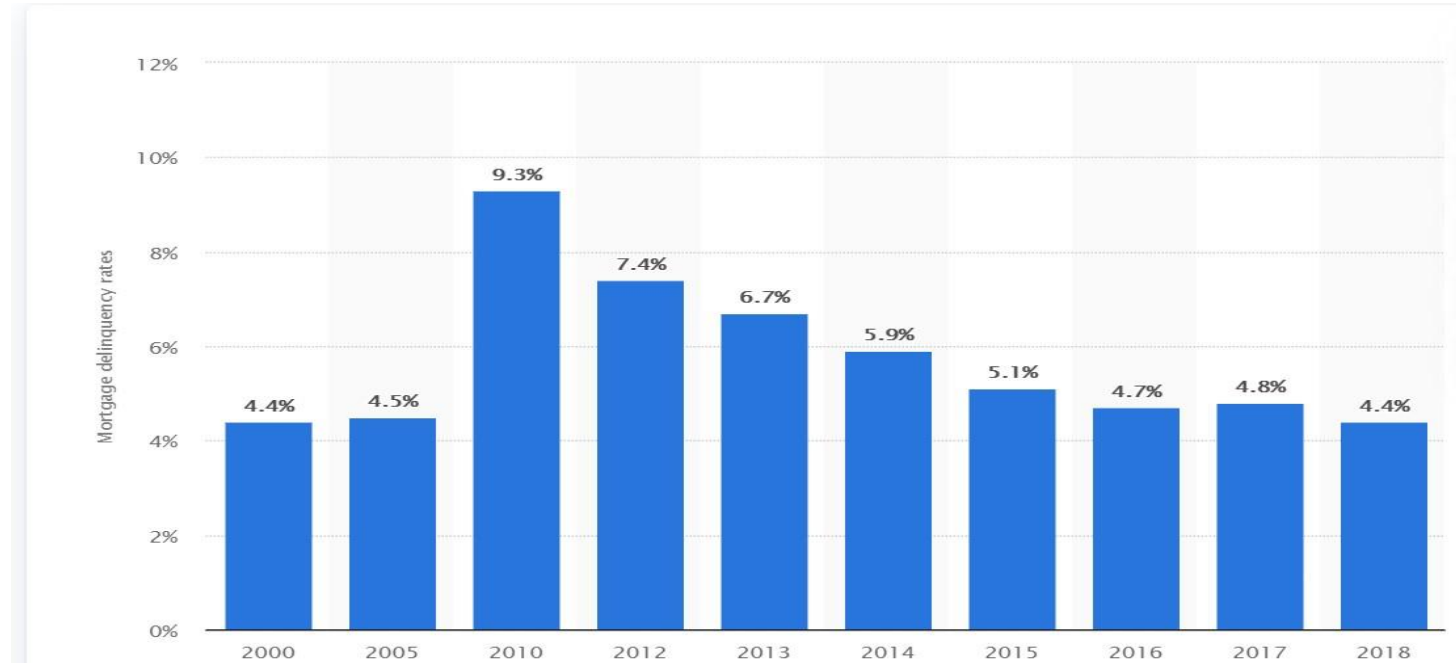
FINANCIAL SERVICES

Use cases

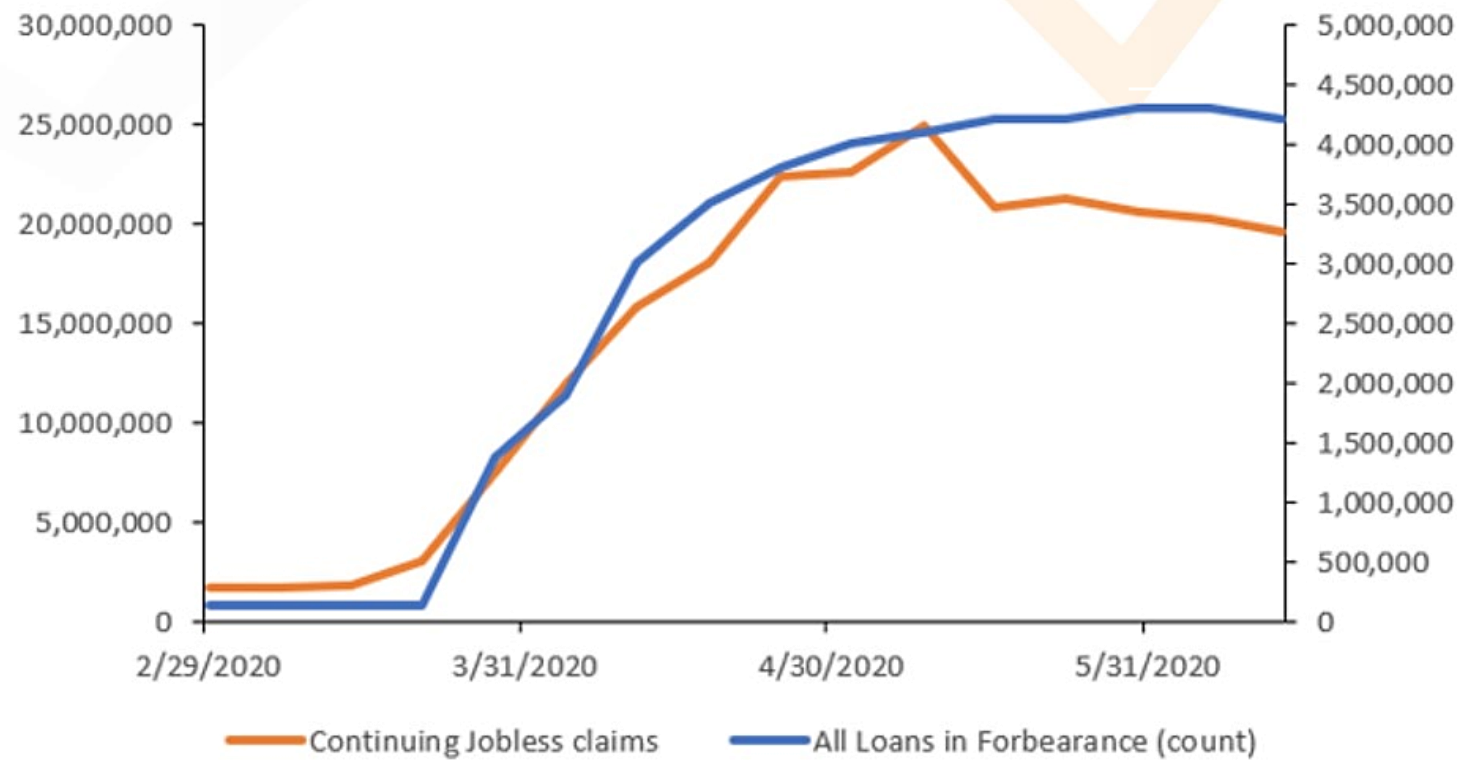


Los Angeles Times

Home lenders brace for up to 15 million mortgage defaults



As many as 30% of Americans with home loans — about 15 million households — could stop paying if the U.S. economy remains closed through the summer or beyond, according to an estimate by Mark Zandi, chief economist for Moody's Analytics.



(Source: Federal Reserve Economic Data and Mortgage Bankers Association)

1. "Americans Are Less Anxious About Paying Their Bills"
2. "Half of U.S. homeowners struggle with mortgage due to COVID-19, consider selling home"

So, according to the first source (the Federal Reserve):

“ About 12.6% of Americans say they were worried about being able to make a minimum debt payment in May, down from a seven-year high in April (16.2%), according to a survey of consumer expectations released by the Federal Reserve. The 12.6% figure is in line with percentages from December, prior to the pandemic that sent a shock wave through the economy.

But according to the second source (NAR and OnePoll):

“ A survey of 2,000 American homeowners found that 52 percent are constantly concerned about making their mortgage payment on time. Forty-seven percent of the poll say they're considering selling their home because they can't afford their mortgage anymore.

Azure Services for Machine Learning



Databricks

Apache Spark platform for
distributed computing

Includes: Databricks Runtime for
Machine Learning



Machine Learning Service

Familiar Jupyter or RStudio
environment for interactive
analyses and operationalization



Azure Databricks

Fast, easy, and collaborative Apache Spark™ based analytics service.

Microsoft Azure

SnEffAnnotationPipeline (Scala)

Detached | File | Edit | View: Standard | Permissions | Run All | Clear

exportVCF: false | inputVariants: dbfs:/databricks-datasets/ | output: /tmp/genomics | replayMode: skip

Cmd 1

databricks

SnEff Annotation Pipeline

Test data to run this pipeline can be located here:

dbfs:/databricks-datasets/genomics/1kg-vcfs/ALL.chr22.phase3_shapeit2_mvncall_integrated_v5a.20130502.genotypes.vcf.gz

Run the cell below to show all parameters that are currently available for configuration.

Cmd 2

```
1 dbutils.fs.mkdirs("/tmp/genomics")
2 display(dbutils.fs.ls("/tmp/"))
```

	path	name	size
1	dbfs:/tmp/attributeanalysis/	attributeanalysis/	0
2	dbfs:/tmp/ccdaingest/	ccdaingest/	0
3	dbfs:/tmp/dataframe_sample.csv	dataframe_sample.csv	272
4	dbfs:/tmp/dbarnhart@blue-granite.com/	dbarnhart@blue-granite.com/	0
5	dbfs:/tmp/genomics/	genomics/	0
6	dbfs:/tmp/hive/	hive/	0
7	dbfs:/tmp/jcarlson@blue-granite.com/	jcarlson@blue-granite.com/	0

Showing all 21 rows.

Command took 0.32 seconds -- by cford@blue-granite.com at 6/1/2020, 4:53:33 PM on genomics_cluster

Cmd 3

```
1 import com.databricks.hls.pipeline.dnaseq._
2 import com.databricks.hls.pipeline._
3 import com.databricks.hls.pipeline.dnaseq.annotation.AnnotationPipeline
4
5 val pipeline = AnnotationPipeline
```

- Familiar notebook-style IDE
- Massively scalable
- Collaborative
- Python, R, Scala, and SQL



Azure Machine Learning Service

Enterprise-grade machine learning service to build and deploy models faster.

The screenshot shows the Microsoft Azure Machine Learning studio interface. The left sidebar contains navigation links: Home, Author (Notebooks, Automated ML (preview), Designer (preview)), Assets (Datasets, Experiments, Pipelines, Models, Endpoints), and Manage (Compute, Datastores, Data Labeling). The main content area is titled 'Welcome to the studio!' and features four cards: 'Create new' (with a plus icon), 'Notebooks' (Code with Python SDK and run sample experiments), 'Automated ML (preview)' (Automatically train and tune a model using a target metric), and 'Designer (preview)' (Drag-and-drop interface from prepping data to deploying models). Below these is a section 'My recent resources' with a table for 'Compute'.

Name	Type	Provisioning state	Created on
cford-1-ci	Compute instance	Succeeded	May 14, 2020 3:23 PM

At the bottom, there is a 'Tutorials' section with six links: 'What is Azure Machine Learning?', 'Train your first ML model with Notebook', 'Create, explore and deploy Automated ML experiments', 'What is Azure Machine Learning designer?', 'What are compute targets in Azure Machine Learning?', and 'Deploy models with Azure Machine Learning'.

- Familiar JupyterLab, Jupyter, and RStudio IDEs
- Python and R SDKs
- Easy operationalization of code as APIs

About LendingClub

- Headquartered in San Francisco
- World's largest peer-to-peer lending platform of unsecured, personal loans.
- Launched as a Facebook app in 2007, spun off as its own company the next year.
- Amounts range from \$1,000 to \$40,000.
- Lenders earn money from interest; platform earns money from service fees and origination fees.



Largest tech IPO in 2014

(Source: CNBC)



Demo

Thank you!

Modern Data Platform and Analytics Strategy, Governance, and Adoption

FREE One-on-One Q&A Session

<https://www.bluegranite.com/azure-data-architect-offer>



Jon Gore

jgore@bluegranite.com