Buenos días y muchas gracias por reunirse.

# Caso ideal

* Qué esperan obtener del producto?
  + Solución rápida y tirar alertas
  + No estar revisando constantemente
  + Automático
  + Amigable
  + Pueda hacerlo cualquier persona
* Cómo esperan que funcione?
  + Caso ideal de uso (interfaz de usuario)
  + Resultado esperado (formato del output)
    - Cómo le funciona más?
      * Variables principales
    - Variables primarias discutidas en la sesión pasada
  + Esperan que la predicción sea automática o manual?
    - La generación de la base de datos también necesitaría ser automática.
      * Caer directamente.
      * Genera un ACL desde Oracle. Se generan las columnas a partir de una JSON.
      * Se hace desde un API, antes era en un CSV.
      * Cae en la base de datos.
    - Automático añade grado de complejidad
    - Cada cuanto tiempo generarían nuevas predicciones?
      * A cualquier hora.
      * Más de un día de duración.
      * La lluvia lo publican cada día.
      * Con fallos pueden llegar a durar cada día.
      * Café no es tan crítico que cambie
      * Correrlo a las 6pm para tener las predicciones en la mañana.
  + Modelos para estación meteorológica
    - Un solo modelo para todas las estaciones
    - Un modelo para cada lugar
    - Múltiples modelos para cada lugar
      * Pueden ser mejores predicciones pero más espacio almacenado, no hay problema con eso?
        + Ilimitado.

# Requerimientos actuales

Asumiendo que es de inicio automático

1. Se ejecuta el archivo principal
2. El programa solicita un archivo de configuración (en este momento, se usa JSON como formato) desde la consola.
   1. El archivo posee configuraciones sobre la localización del archivo CSV a usar
   2. Las columnas por utilizar para la predicción
      1. Formato de estas columnas
   3. La cantidad de días a predecir
3. El programa filtra la set de datos y brinda formato para la siguiente etapa
   1. Si existe información faltante el los últimos N días, se utiliza la predicción generada el día pasado como sustituto.
4. El programa aprende a realizar predicciones usando los datos brindados
   1. Si no es la primera vez:
      1. Busca por un archivo con información sobre predicciones
      2. Carga la información aprendida en veces pasadas
      3. Modifica el criterio de predicción usando la información del día que acaba de pasar (asumiendo que se genera una predicción cada día)
5. Se genera una tabla con los datos predichos para los próximos N días
6. Se guarda la información aprendida para predecir
7. Se guarda en el mismo lugar que el CSV original pero en un nuevo archivo
8. El programa se queda esperando al siguiente día
   1. Cuando el siguiente día empieza, se espera que el sistema principal haya exportado la información meteorológica antes de que el programa realice la nueva predicción.

Notas:

* Se están usando, como mínimo, horas entre lecturas. La tabla con la información predicha estaría también en horas
  + Por día
  + Número de horas mojadas en un día.
* Problemas como un sensor que reporta la misma lectura a través de varios días, cuando por lógica no debería ser así, no se pueden remover automáticamente. Se necesitaría un análisis previo.
* Historial de predicciones?
* Se almacena cada 15 minutos.
* Intelligit
* Eclipse

# Especificaciones del equipo

* Qué tipo de equipos tienen en las instalaciones?
  + CPU, GPU, versión de Windows/Linux
    - Linux y Windows. Windows server virtualized
    - Virtualbox
    - No tiene GPU. Analizar CPU
  + Acceso a python
* Existe algo que pueda limitar los recursos usados?
  + Memoria de almacenamiento
  + Programas que ocupen recursos que puedan limitar la ejecución del producto
    - Estación meteorológica es lo más pesado
    - Usan Oracle
    - Ancho de banda podría ser limitante

## Formato/Uso de archivos

* Cómo se generan los datos (CSV) de la estación meteorológica?
  + De manera automática?
    - Usando el API, se genera automática
  + Cómo funcionaría el producto dentro del sistema actual?
  + Tienen algún estándar para el formato a la hora de generar bases de datos?
    - Salida que tiene por defecto
  + Reglas de negocio? Protocolos, estándares o reglas del gobierno que deban seguir que creen que tengan impacto en el proyecto
    - Existiría algún problema con usar programas de terceros con patentes?
    - Existiría algún problema con usar servicios en la nube?
      * Oracle, aplicaciones con Oracle Linux y base de datos
      * Azure, respalda sobre el Windows server
* Se necesita compatibilidad con otro programa?
  + Los datos generados se piensan pasar por otro programa para el análisis/generar más información?
    - Mostrar tal cual para los productores
    - Viendo las predicciones en crudo

## Prevención de errores

* Las instalaciones sufren alguna falla?
  + Apagones recurrentes
    - Datacenter. Tienen una planta eléctrica en caso de apagón
    - Se redirige y se balancea la carga para usar siempre los servicios en la nube
  + Reinicios del sistema por falla o mantenimiento
    - Cuando hay tipo hackeo (DDOS). Reiniciar servidores
    - Transacciones anómala en base de datos, ocurre un bloqueo en la base de datos.
* Hacen respaldos de la información?
  + Sí, ver información anteriores

# Organización de la empresa/Stakeholders

* Se tiene todo el análisis se realiza en un mismo lugar o existen divisiones?
  + Físicas: varias instalaciones en distintos lados del país
    - Sólo una. Todo llega al datacenter
  + En grupos de trabajo: el grupo A se encarga del sector de Barva, mientras el B vigila el sector de Poás
    - Por lo menos la parte de clima. Otros roles con las regionales
    - Seguridad segmentada y módulo para otros empleados si ocupan utilizar el sistema.
* Quienes son posibles involucrados en el uso del programa
  + Analistas con los datos del resultado
    - Análisis utilizando de manera visual
  + Informáticos para mantenimiento
  + TI generaría los resultados y se pasan luego a los analistas
    - Generan los datos parseados
* Tienen experiencia usando productos/programas similares?
  + Entre TI y analistas trabajarían en conjunto para este proyecto
  + Con Luis, cuestiones para predicciones de enfermedades usando datos climáticos
    - Input con clima y datos de enfermedad y genera la predicción. El trabajo era manual.
* Cuál es la rutina actual de los analistas?
  + Cuanto tiempo atrás ven los datos?
    - Enfermedades: comportamiento de las últimas 3 semanas para ver el siguiente. Se hace todos los días
    - EN general, se ven los datos del último mes. Para las plagas
    - En otros casos son los últimos 15 días
    - Para riego, se maneja el acumulado como desde octubre hasta mayo
  + En qué frecuencias (horas, días semanas) ven los datos para analizarlos (cuanto ven)?
    - Semana, se toma un día para revisar información y cosas generales
    - A veces al mes porque no son tan cr´tiicas por cuestiones agronómicas
    - Inicios de año detrás del riego, floración, inicio de enfermedades y lluvias hasta mayo y junio.
      * Octubre con picos de cosecha por maduración temprana
  + Con cuanta frecuencia revisan la información generada por la estación meteorológica para analizar la situación (cuando ven)?
    - Cada día, semana, 15 días?
      * Por día

# Confidencialidad

* Confidencialidad en los datos?
  + Base de datos
    - Sí
  + Datos en crudo son confidenciales
* No hay problema con que se publique. Mencionar al ICAFE con el trabajo en conjunto e imagen.
* Propiedad intelectual. Es libre para el ICAFE y no se limita el mantenimiento

# Medidas de éxito/rendimiento

* Algún margen de error para los datos predichos?
  + Qué tanto error generaría problemas? Ejemplos
  + Para arriba de 80% de certeza, a veces con un 75%
* Cuanto tiempo creen que puede durar?
  + Cuanto es el tiempo máximo que están dispuestos a esperar?
    - En este momento se tiene como margen 1h.
* Tienen pensado una etapa de prueba para cuando el producto ya esté terminado?
  + Qué tanto considerarán las predicciones en su toma de decisiones?
  + Periodo de revisión con bastantes usuarios
* Portal con información que publican
  + Productores
  + Técnicos
  + Exportador
  + Técnicos analistas

# Consideraciones extras

* Se tendrían que ejecutar varias instancias para la predicción de cada estación, independientemente de la cantidad de modelos a utilizar
  + Una instancia analizaría la estación de Poás, mientras otra instancia la de Barva, pues cada archivo de datos genera una predicción diferente y cada estación genera un archivo distinto en cada día.

# Comentarios/Dudas