

# Tabla de contenido

1. Introducción .....	1
1.1. Antecedentes de la empresa .....	1
1.2. Justificación .....	2
1.3. Objetivos .....	2
1.4. Metodología.....	3
1.5. Estructura del Documento .....	4
2. Estado del arte.....	5
2.1. PostgreSQL.....	5
2.2. Amazon Web Services .....	5
2.2.1. S3 .....	7
2.2.2. DynamoDB .....	7
2.2.3. Athena .....	7
2.2.4. Redshift .....	8
2.2.5. EC2.....	8
2.2.6. Aurora PostgreSQL.....	8
2.3. Selenium.....	9
2.3.1. Google Chrome .....	9
2.3.2. Mozilla Firefox .....	10
2.3.3. PhantomJS .....	10
2.4. Documentos tributarios .....	10
2.5. VNC.....	12
2.6. Power BI.....	13
2.6. Conexión a servidor remoto .....	13
2.6.1. WinSCP.....	13
2.6.2. Cygwin.....	14
2.6.3. VNC Viewer .....	15
2.7. Modelo de datos .....	15
2.7.1. Modelo estrella .....	15
2.7.1. Modelo copo de nieve .....	16
3. Diseño .....	18
3.1. Introducción.....	18
3.2. Visión General y unidades funcionales.....	18
3.2.1. Obtención de datos .....	19
3.2.2. Modelamiento de datos.....	19
3.3.3. Visualización de datos.....	19
3.3. Arquitectura del sistema .....	19
3.3.1. Primera capa: Fuente de datos .....	20
3.3.2. Segunda capa: Extracción, transformación y carga de datos .....	23
3.3.3. Tercera capa: Modelo de datos .....	26
3.3.4. Cuarta capa: Análisis de datos .....	26
3.3.5. Quinta capa: Propuesta de servicio .....	29
3.4. Servidor remoto .....	29

3.4.1. EC-2.....	30
3.4.2. VNC.....	31
4. Implementación .....	33
4.1. Introducción.....	33
4.2. Preparación del entorno de trabajo.....	33
4.2.1. Instalación navegadores.....	33
4.2.2. Instalación de Python y librerías .....	33
4.2.3. Instalación certificado digital.....	34
4.2.4. Permisos.....	36
4.2.5. Herramientas para comunicación con servidor remoto.....	38
4.2.6. Interfaz gráfica con VNC .....	38
4.2.7. Preparación de entorno de trabajo remoto .....	42
4.3. Arquitectura de cinco capas .....	43
4.3.1. Primera capa: Fuente de datos.....	43
4.3.2. Segunda capa: Extracción, transformación y carga de datos .....	44
4.3.3. Tercera capa: Modelo de datos .....	46
4.3.4. Cuarta capa: Análisis de datos .....	52
4.3.5. Quinta capa: Propuesta de servicio .....	53
4.4. Automatización de proceso en servidor remoto .....	53
5. Pruebas y resultados.....	56
5.1. Introducción.....	56
5.2. Descarga de datos .....	56
5.3. Caso de análisis para clientes .....	58
5.3.1. Caso de análisis: cliente_2 .....	58
5.3.2. Caso de análisis: cliente_3 .....	60
6. Conclusiones.....	63
6.1. Resultados .....	63
6.2. Importancia del tamaño de datos.....	64
6.3. Futuras líneas de trabajo en empresa .....	66
6.4. Alternativas a AWS.....	67
6.4.1. Google Cloud.....	68
6.4.2. Microsoft Azure.....	69
6.4.3. Comparación con AWS .....	69
7. Bibliografía .....	71

# Lista de tablas

Tabla 1: Archivos necesarios para distintos navegadores web en Windows. ....	9
Tabla 2: Detalles de instancias M5 (Amazon AWS, 2019).....	30
Tabla 3: Análisis de resultados de web scraping para clientes de empresa .....	43
Tabla 4: Resultados de descarga manual de facturas.....	44
Tabla 5: Resultados de descarga automática de facturas.....	44
Tabla 6: Tiempo de ejecución para segunda capa en servidor local.....	46
Tabla 7: Tiempo de ejecución para segunda capa en servidor remoto. ....	46
Tabla 8: Estructura de tabla doc_datos_documento.....	47
Tabla 9: Estructura de tabla doc_detalle .....	48
Tabla 10: Estructura de tabla doc_emisor .....	48
Tabla 11: Estructura de tabla doc_fecha_emision.....	49
Tabla 12: Estructura de tabla doc_receptor.....	49
Tabla 13: Estructura de tabla doc_referencia.....	49
Tabla 14: Estructura de tabla doc_transaccion.....	50
Tabla 15: Estructura de tabla registro_descarga_xml .....	51
Tabla 16: Campos necesarios de Cron en Linux/Unix. (González, 2019).....	54
Tabla 17: Cantidad de documentos correspondientes a clientes .....	56
Tabla 18: Tiempo de ejecución de arquitectura de descarga en servidor remoto .....	57
Tabla 19: Reporte general como emisor para cliente_2.....	58
Tabla 20: Reporte específico como emisor para cliente_2.....	59
Tabla 21: Reporte específico como emisor para cliente_3 .....	60
Tabla 22: Reporte general como emisor para cliente_3 .....	61
Tabla 23: Reporte general como receptor para cliente_3.....	62
Tabla 24: Comparación de herramientas de servicios de computación en la nube .....	70

# Lista de figuras

Figura 1: Ejemplo de dispositivo POS. Xeumior SM-V2.....	1
Figura 2: Mapa de la infraestructura global de AWS. (Amazon AWS, 2019) .....	6
Figura 3: Programas en internet según servidores. ....	6
Figura 4: Factura electrónica MiPyme. (SII, 2019) .....	11
Figura 5: Extracto de XML de factura electrónica. ....	12
Figura 6: Ejemplo de dashboard en Power BI. (Microsoft, 2019) .....	13
Figura 7: Interfaz de WinSCP. ....	14
Figura 8: Terminal de Cygwin. ....	14
Figura 9: VNC Viewer en ejecución. ....	15
Figura 10: Ejemplo de modelo de datos en estrella. (Wikipedia, 2019) .....	16
Figura 11: Ejemplo de modelo de datos en copo de nieve. (Wikipedia, 2019) .....	17
Figura 12: Arquitectura de descarga. ....	20
Figura 13: Pasos para ingresar a sistema de descarga de información del SII .....	21
Figura 14: Plataforma de descarga de datos emitidos para un cliente sin aplicar filtros de búsqueda.....	22
Figura 15: Plataforma de descarga de datos recibidos para un cliente sin aplicar filtros de búsqueda.....	22
Figura 16: Mockup de reporte general.....	28
Figura 17: Mockup de reporte específico.....	28
Figura 18: Logo of the Xfce project. (Xfce Team, 2008).....	31
Figura 19: A screenshot of XFCE 4.10-1 with panel, Settings window, default file manager and xfce theme. (Xfce Development Team, 2012) .....	32
Figura 20: Pasos necesarios para dar permisos a puertos en AWS.....	37
Figura 21: Acceso a puertos necesarios en AWS. ....	37
Figura 22: Primer mensaje mostrado en máquina virtual tras primer acceso exitoso mediante VNC Viewer.....	40
Figura 23: Estructura de diccionario para encabezado.....	45
Figura 24: Estructura de diccionario de diccionarios para detalles. ....	45
Figura 25: Modelo estrella para información de documentos descargados. Elaboración propia .....	50
Figura 26: Reporte general para clientes .....	52
Figura 27: Reporte específico para clientes .....	53