

Architecture Design

Entertainer Data Analysis

| | |
|-------------------|---------------|
| Written By | Avs Jagannath |
| Document Version | 0.3 |
| Last Revised Date | |

DOCUMENT CONTROL

Change Record:

| VERSION | DATE | AUTHOR | COMMENTS |
|---------|--------------|---------------|--|
| 0.1 | 07-July-2024 | Avs Jagannath | Introduction & Architecture define |
| 0.2 | 07-July-2024 | Avs Jagannath | Architecture & Architecture description appended & updated |
| | | | |
| | | | |

Reviews:

| VERSION | DATE | REVIWERS | COMMENTS |
|---------|--------------|---------------|-----------------------------|
| 0.2 | 08-JULY-2024 | Avs Jagannath | Unit test cases to be added |

Approval Status:

| VERSION | REVIEW DATE | REVIEWED BY | | APPROVED BY | COMMENTS |
|---------|-------------|-------------|--|-------------|----------|
| | | | | | |

Contents:

| | |
|---|-----|
| Document Version Control | 2 |
| 1. Introduction 1.1 What is Architecture Design Document? 1.2 Scope | 4 |
| 2. Architecture 2.1 Overall System Architecture 2.2 Data Processing Architecture 2.3 Visualization Tool Architecture 2.4 Deployment Architecture | 5-6 |
| 3. Deployment Description 3.1 Deployment Options 3.2 Single Node Architecture 3.3 Multi-Node Architecture | 7-8 |

1. Introduction

1.1 What is Architecture Design Document?

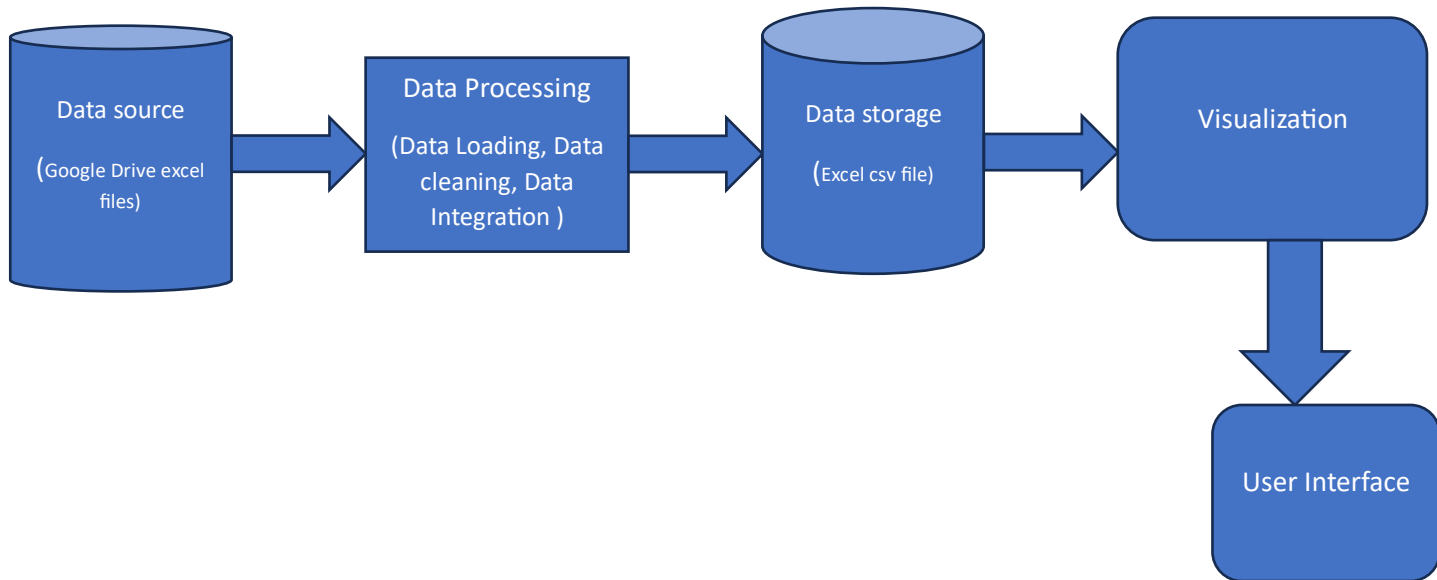
This Architecture Design Document (ADD) represents the structure of the Entertainer Data Analysis system. It defines the hardware and software components, their interfaces, and the framework for system development. The document describes the system categories, components, connectors, and integration methods.

1.2 Scope

This ADD covers the architectural design process for the Entertainer Data Analysis project, including data structures, software architecture, and performance considerations. It provides a step-by-step refinement of the design principles defined during the requirement analysis phase.

2. Architecture

2.1 Overall System Architecture



The system consists of the following main components:

- Data Sources (Google Drive Excel files)
- Data Processing Layer (Python)
- Data Storage (CSV/Database)
- Visualization Layer (Power BI)
- User Interface (Web Browser/Desktop Application)

2.2 Data Processing Architecture

The data processing layer uses Python to handle:

- Data loading from multiple Excel files
- Data cleaning and transformation
- Feature engineering
- Data integration

ARCHITECTURE DESIGN 6

Key components:

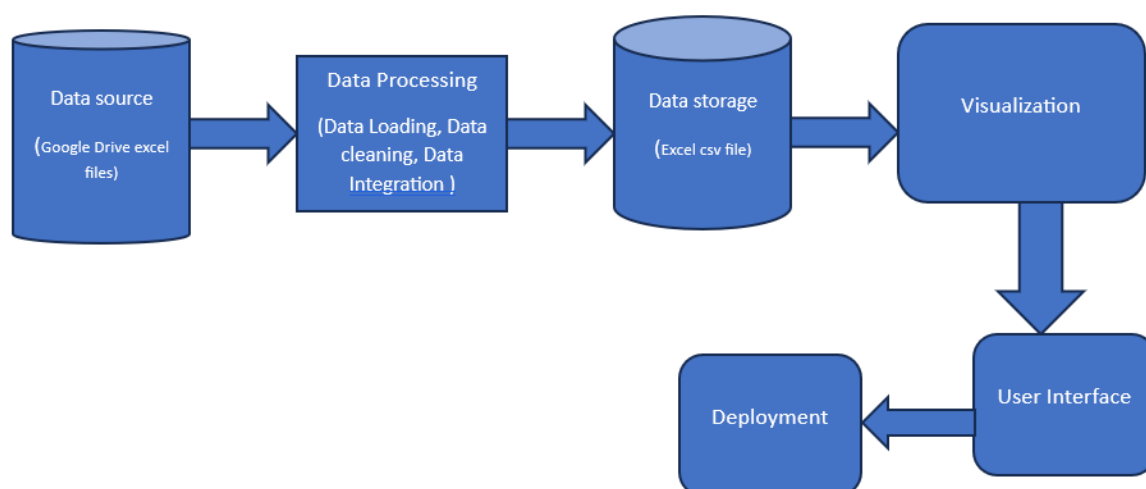
- Pandas for data manipulation
- NumPy for numerical operations
- SQL Alchemy for database interactions (if using a database)

2.3 Visualization Tool Architecture

PowerBi Server Architecture:

- Gateway/Load Balancer: Entry point for client requests
- Application Server: Handles user authentication and permissions
- VizQL Server: Processes visualization requests
- Data Server: Manages data source connections
- Backgrounder: Handles scheduled tasks and extract refreshes
- Data Engine: Stores and queries data extracts
- Repository: Stores Tableau Server metadata

2.4 Deployment Architecture



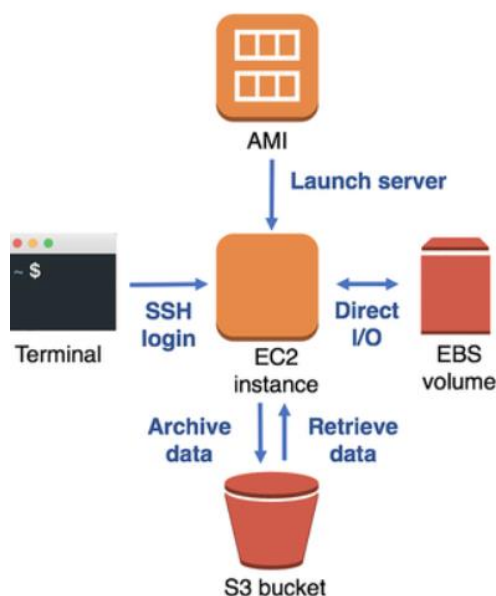
3. Deployment Description

3.1 Deployment Options

1. Local Deployment: Install visualization tool on individual machines for development and testing.
2. Server Deployment: Deploy on a centralized server for multi-user access.
3. Cloud Deployment: Utilize cloud services for scalability and accessibility.

3.2 Single Node Architecture

(a) Single-node workflow

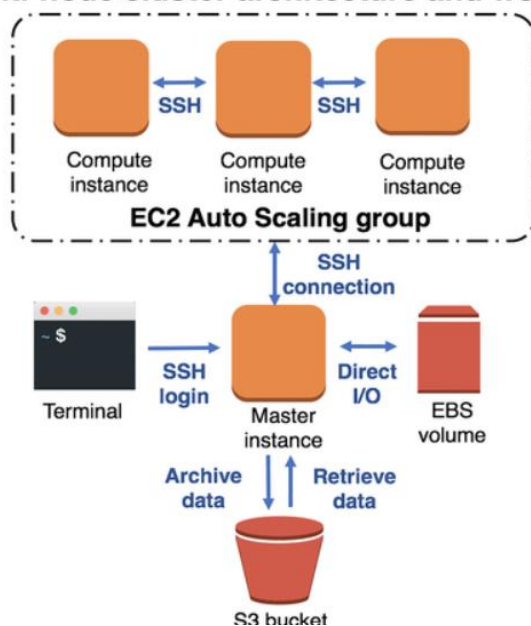


Description:

- All components run on a single machine
- Suitable for small-scale deployments or testing environments
- Limited in terms of scalability and performance for large datasets or multiple concurrent users

3.3 Multi-Node Architecture

Multi-node cluster architecture and workflow



Description:

- Distributes components across multiple machines
- Improves performance and handles concurrent requests more effectively
- Provides options for high availability and failover
- Scalable to accommodate growing data volumes and user bases

Key considerations for multi-node setup:

- Load balancing between nodes
- Data synchronization and consistency
- Network latency and security between nodes
- Monitoring and management of distributed components

This architecture design provides a comprehensive overview of the Entertainer Data Analysis system, covering data processing, visualization, and deployment aspects. It can be further refined based on specific project requirements and chosen technologies.