

FACHPRÄSENTATION

Aufgabenstellung zum Kurs: DLMDWWDE01 – Data Engineering

INHALTSVERZEICHNIS

1. Aufgabenstellung.....	2
1.1. Aufgabenstellung 1: Konzepterstellung und Design einer Datensystem-Architektur zur Verarbeitung großer Datenmengen	2
1.2. Aufgabenstellung 2: Aufbau der Infrastruktur für eine Anwendung zur Verarbeitung großer Datenmengen unter Verwendung von DataOps und Microservices	3
1.3. Aufgabenstellung 3: Auswahl eines Cloud-Anbieters mit Schwerpunkt auf Data Governance und Sicherheit	3
2. Zusatzinformation zur Bewertung der Fachpräsentation	4
3. Betreuungsprozess	4

1. AUFGABENSTELLUNG

Für die Fachpräsentation stehen verschiedene Aufgabenstellungen zur Auswahl. Bitte entscheide Dich für eine davon, die Du in Deiner Präsentation bearbeiten möchtest.

Hinweis zum Urheberrecht:

Es wird darauf hingewiesen, dass der IU Internationale Hochschule GmbH das Urheberrecht der Prüfungsaufgaben/Aufgabenstellungen obliegt. Einer Veröffentlichung der Aufgabenstellungen auf Drittplattformen wird ausdrücklich widersprochen. Im Falle einer Zuwiderhandlung stehen der Hochschule u.a. Unterlassungsansprüche zu.

Anwendungsfall: Du bist damit beauftragt, die Architektur für eine groß angelegte Anwendung zu entwickeln, die Daten von Lebensmittelgeschäften eines weltweit tätigen Unternehmens speichert und verarbeitet. Das System sollte in der Lage sein, Lese- und Schreibvorgänge von mehreren Lebensmittelgeschäften auf der ganzen Welt zu verarbeiten, wobei die Latenzzeit Vorrang vor der Datenkonsistenz haben sollte. An den Kassen der Lebensmittelläden werden während der Öffnungszeiten kontinuierlich Daten erzeugt. Vorverarbeitete Daten müssen kontinuierlich für die Berichterstattung an die Manager der Lebensmittelgeschäfte vor Ort sowie in regelmäßigen Abständen für Data Scientist zur Optimierung der Nachfrageprognosemodelle zur Verfügung gestellt werden. Da es sich um ein geschäftskritisches System handelt, sollten eine automatische Ausfallsicherung sowie Datensicherheits- und Governance-Grundsätze implementiert werden.

1.1. Aufgabenstellung 1: Konzepterstellung und Design einer Datensystem-Architektur zur Verarbeitung großer Datenmengen

Dein Architekturentwurf sollte die Standardprinzipien für Zuverlässigkeit (reliability), Skalierbarkeit (scalability) und Wartbarkeit (maintainability) berücksichtigen. Du solltest auch berücksichtigen, dass die Daten sowohl in Echtzeit (für die Manager der lokalen Lebensmittelgeschäfte) als auch für die Batch-Verarbeitung (für die Nachfrageprognose) verfügbar sein müssen.

1. Beschreibe die wichtigsten Grundsätze, die in Deinem System implementiert werden: Erkläre, was die Begriffe "Zuverlässigkeit", "Skalierbarkeit" und "Wartbarkeit" im Zusammenhang mit einer datenintensiven Anwendung bedeuten. Zeige außerdem, wie Du diese Grundsätze in der vorgestellten Systemarchitektur anwendest.
2. Dein System soll Echtzeitdaten bereitstellen, die an den Kassen der Lebensmittelgeschäfte erzeugt und in Berichten für die Manager der lokalen Lebensmittelgeschäfte verwendet werden. Darüber hinaus versorgt es ein Team von zentralisierten internen Data Scientists regelmäßig mit Daten zur Optimierung von Nachfrageprognosemodellen. Erläutere, wie Deine entworfene Architektur diese Anforderungen erfüllen wird. Erläutere außerdem, an welchen Stellen der Architektur OLAP- oder OLTP-Prozesse ausgeführt werden sollen.
3. Erörtere, inwieweit Dein Architekturentwurf für die absehbaren zukünftigen Entwicklungen in diesem Bereich geeignet ist. Gehe kurz auf mindestens drei aktuelle Trends im Data Engineering ein und konzentriere Dich auf einen Aspekt, der von deinem Entwurf besonders gut abgedeckt wird.

Einführende Literaturhinweise:

Kleppmann, M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. Sebastopol, CA: O'Reilly.

1.2. Aufgabenstellung 2: Aufbau der Infrastruktur für eine Anwendung zur Verarbeitung großer Datenmengen unter Verwendung von DataOps und Microservices

Um eine schnelle Wiederherstellung nach Systemausfällen und Flexibilität in der Wahl der verwendeten Frameworks zu garantieren, entwickelst Du ein System, welches den Prinzipien von Infrastructure as Code (IaC) und Continuous Integration/Continuous Deployment (CI/CD) folgt. Dabei werden einzelne Teile des Systems aus Gründen der Sicherheit, des Ausfallschutzes und der Wartbarkeit als Microservices (MS) voneinander isoliert.

1. Erläutere die wichtigsten Prinzipien von DataOps in Bezug auf MS, IaC und CI/CD. Lege dar, wie diese Software-Engineering-Konzepte auf daten-intensive Systeme übertragen werden können.
2. Zeige anhand des vorgestellten Anwendungsfalls, wie die DataOps-Prinzipien für CI/CD, IaC und MS praktisch auf ein Produktivsystem angewendet werden können.
3. Diskutiere die Vor- und Nachteile verschiedener Anwendungsframeworks für die vorgestellten Prinzipien. Wähle mindestens zwei DevOps/DataOps-Plattformen aus und vergleiche diese hinsichtlich ihrer Eignung, CI/CD, IaC und MS für den beschriebenen Anwendungsfall zu implementieren.

Einführende Literaturhinweise:

Kleppmann, M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. Sebastopol, CA: O'Reilly.

Farcic, V. (2016). *The DevOps 2.0 toolkit: Automating the continuous deployment pipeline with containerized microservices*. Scotts Valley, CA: CreateSpace Independent Publishing Platform.

1.3. Aufgabenstellung 3: Auswahl eines Cloud-Anbieters mit Schwerpunkt auf Data Governance und Sicherheit

Da im beschriebenen Anwendungsfall weltweit große Datenmengen erzeugt und verarbeitet werden müssen, wählst Du einen Cloud-Anbieter für das Hosting des verteilten Systems. Da die zu implementierende Anwendung geschäftskritisch sein wird, konzentrierst Du Dich bei der Bewertung auf die folgenden Hauptkriterien: Datenschutz, Datensicherheit und Data Governance.

1. Erläutere die Hauptprinzipien des Datenschutzes, der Datensicherheit und der Data Governance. Zeige anhand des vorgestellten Anwendungsfalles, wie diese Prinzipien praktisch angewendet werden können.
2. Diskutiere die Vor- und Nachteile von drei großen Cloud-Anbietern sowie einem kleineren Anbieter Deiner Wahl im Hinblick auf Datenschutz, Datensicherheit und Data Governance. Lege dar, welche Funktionen für die verschiedenen Dienste innerhalb der einzelnen Cloud-Plattformen im Hinblick auf diese Prinzipien verfügbar sind.
3. Mache einen fundierten Vorschlag für einen geeigneten Cloud-Anbieter für das System des Anwendungsfalles. Berücksichtige dabei neben den oben genannten Aspekten auch die Kosten und die absehbaren zukünftigen Entwicklungen in diesem Bereich.

Einführende Literaturhinweise:

Kleppmann, M. (2017). *Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems*. Sebastopol, CA: O'Reilly.

Hille, M., Klemm, D., & Lemmermann, L. (2018). *Cloud computing vendor & service provider comparison*. Crisp Vendor Universe. (URL: https://www.reply.com/Documents/Crisp_Vendor_Universe_Cloud%20Computing_250118_REPLY_englischeVersion_FINAL.pdf)

2. ZUSATZINFORMATION ZUR BEWERTUNG DER FACHPRÄSENTATION

Bei der Konzeption und Erstellung der Fachpräsentation sollten die im Prüfungsleitfaden aufgeführten Bewertungskriterien und Erläuterungen berücksichtigt werden.

3. BETREUUNGSPROZESS

Für die Betreuung der Fachpräsentation stehen grundsätzlich mehrere Kanäle offen. Die jeweilige Inanspruchnahme liegt dabei im eigenen Verantwortungsbereich. Die Tutor:innen stehen für fachliche Rücksprachen zur Themenwahl einerseits sowie für formale und allgemeine Fragen zum wissenschaftlichen Arbeiten andererseits zur Verfügung. Eine Abnahme von Gliederungen, Textteilen oder –entwürfen durch die Tutor:innen ist hierbei jedoch nicht vorgesehen, da die eigenständige Erstellung Teil der zu erbringenden Prüfungsleistung ist und in die Gesamtbewertung einfließt. Es werden jedoch Hinweise zu Gliederungsentwürfen gegeben, um den Einstieg in die Strukturierung einer wissenschaftlichen Arbeit zu erleichtern.